Received: 6 November 2019 | Revised: 27 November 2020 | Accepted: 12 December 2020

DOI: 10.1111/sjos.12511

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Semiparametric analysis of interval-censored failure time data with outcome-dependent observation schemes

Yayuan Zhu[1] | **Ziqi Chen**[2] | **Jerald F. Lawless**[3]

[1]Department of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario, Canada

[2]Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, Shanghai, P.R. China

[3]Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada

Correspondence
Yayuan Zhu, Department of Epidemiology and Biostatistics, University of Western Ontario, London, ON N6A 3K7, Canada.
Emails: yayuan.zhu@uwo.ca; yayuanzhu@gmail.com

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Numbers: DGECR-2020-00354, RGPIN-2017-04055, RGPIN-2020-05803

Abstract

Disease progression is often monitored by intermittent follow-up "visits" in longitudinal cohort studies, resulting in interval-censored failure time outcomes. Furthermore, the timing and frequency of visits is often found related to a person's history of disease-related variables in practice. This article develops a semiparametric estimation approach using weighted binomial regression and a kernel smoother to analyze interval-censored failure time data. Visit times are allowed to be subject-specific and outcome-dependent. We consider a collection of widely used semiparametric regression models, including additive hazards and linear transformation models. For additive hazards models, the nonparametric component has a closed-form estimator and the estimators of regression coefficients are shown to be asymptotically multivariate normal with sandwich-type covariance matrices. Simulations are conducted to examine the finite sample performance of the proposed estimators. A data set from the Toronto Psoriatic Arthritis (PsA) Cohort Study is used to illustrate the proposed methodology.

KEYWORDS

additive hazards model, dependent visit times, interval censoring, inverse-intensity-of-visit weight, linear transformation models, semiparametric estimation

1 | INTRODUCTION

In longitudinal cohort studies from disease clinics, disease progression or the occurrence of an adverse clinical event is often assessed at periodical follow-up visits. It results in interval-censored failure times. That is, the event of interest is known only to occur between two consecutive visits instead of being observed exactly. For example, Finkelstein (1986) studied early breast cancer patients who were monitored for the cosmetic effects of adjuvant therapy. Patients were seen at the clinic every 4–6 months. The event of interest is the appearance of breast retraction (dimpling) which only can be ascertained to lie in a 4–6-month interval between two appointments. Another example is the Toronto Psoriatic Arthritis (PsA) Cohort Study (Gladman & Chandran, 2010), where patients' joint activities and damage are assessed by physician evaluation or radiographical examination at clinic visits. Thus failure time outcomes, such as the time to damage reaching a certain level, are subject to interval censoring. Patients are scheduled to visit the clinic every 6–12 months, but actual intervisit times vary substantially both within and across subjects. Moreover, the times between visits are related to factors associated with the progression of joint damage. Similar situations arise in many other settings involving clinical trials and observational follow-up studies.

Several approaches have been developed to conduct semiparametric analysis with interval-censored data. Many of them focus on the Cox proportional hazards model specifically, for example, Finkelstein (1986), Goetghebeur and Ryan (2000), Zhang et al. (2010), among others. Lin et al. (1998) and Zeng et al. (2006) considered additive hazards models and Zhang, Sun, Zhao, and Sun (2005) discussed linear transformation models. However, most existing methods in the literature assume that visit times at which failure status may be observed are independent of the failure time, conditional on covariates in the model of interest. Interval censoring has received much less attention in settings where the visit times are not conditionally independent in this sense. It was considered by van der Laan and Robins (1998) for case-I interval-censored data (current status data); they developed an inverse-probability-of-censoring weighted (IPCW) estimator. In the general case (case-II) the most common approach has been joint modeling of the failure time and visit time processes using shared or correlated random effects; see, for example, Zhang, Sun, and Sun, (2005), Zhang et al. (2007), and Chen et al. (2012). However, in joint models, the interpretation of the estimates of regression coefficients is conditional on both observed covariates and unobservable random effects; moreover, time-invariant random effects that are usually adopted in practice may not be plausible in dynamic processes where the mechanism connecting visits and failure varies over time.

Our objective is to develop a marginal modeling approach, based on a variety of semiparametric regression models, that avoids conditioning on too many covariates and on latent variables while allowing for dependent visit times. The marginal or partly conditional effects of specified factors on event risk are often of interest in biomedical research and public health studies. For example, Finkelstein (1986) compared the effect of treatment on the rate of deterioration of the cosmetic state between patients who received adjuvant chemotherapy and those who received radiotherapy alone. In the PsA example, one point of interest is to compare the risk of joint damage between PsA patients who have certain genomic characteristics and those who do not (as we discuss in Section 4).

We extend methods in Zhu et al. (2017, 2018) that establish weighted estimating equations for the survivor status observed at visits, based on the framework of direct binomial regression (Azarang et al., 2017; Fine, 1999; Scheike & Zhang, 2007). Inverse-intensity-of-visit (IIV)

weighting was introduced for general longitudinal data processes observed at irregular, intermittent times (Buzkova, 2010; Buzkova & Lumley, 2007, 2009; Lin et al., 2004; Pullenayegum & Feldman, 2013). It works by assuming a visit time model based on the observed past history of outcomes, visit times, and risk factors. In this article, we extend the IIV weighting approach to some commonly used semiparametric regression models for the case-II interval-censored data. This involves kernel smoothing for nonparametric estimation of the baseline survivor function and regression coefficients estimated by solving an IIV-weighted profile estimating equation. Large sample theory is developed, based on profile likelihood theory. Finite sample properties of the estimators are examined in simulation studies and compared with methods that assume independent visit times. We first focus on additive hazards models and then extend the approach to a more flexible semiparametric model family which has a linear transformation form and includes the Cox proportional hazards model as a special case. Finally, we fit a variety of semiparametric regression models to data from the Toronto PsA Cohort Study as illustrations.

The remainder of the article is organized as follows. Section 2 introduces some semiparametric failure time regression models, reviews the IIV weighting approach, presents the methodological results for additive hazards models and then discusses the extension to linear transformation models. Simulation studies on the finite sample performance of the proposed estimators are presented in Section 3. The methods are applied to the Toronto PsA Cohort Study in Section 4. Section 5 addresses conclusions and discussion. Some theoretical results are for convenience provided in appendices and/or in online supporting information.

2 | METHODOLOGY

2.1 | Generalized linear failure time models

We consider semiparametric regression models for failure time T_i which can be written as

$$S(t|\boldsymbol{A}_i) = g\{h(t) + \boldsymbol{A}_i^T \boldsymbol{\mathcal{B}}(t)\}, \tag{1}$$

where $S(t|\boldsymbol{A}_i) = P(T_i > t|\boldsymbol{A}_i)$ denotes the survivor function of T_i conditional on a vector of time-fixed covariates $\boldsymbol{A}_i$, $g(\cdot)$ is a known monotone differentiable link function (Cook & Lawless, 2018), $h(t)$ is completely unspecified, and $\boldsymbol{\mathcal{B}}(t)$ denotes a vector of time-invariant or possibly time-varying coefficients. Our focus here is on time-fixed covariates $\boldsymbol{A_i}$, which are of frequent interest in clinical practice. The model can be extended to handle external time-dependent covariates (Kalbfleisch & Prentice, 2002), but it must usually be severely restricted when covariates can be measured only at the intermittent visit times.

Aalen (1980) proposed a general linear model which specifies hazard functions with an additive form,

$$\lambda(t|\boldsymbol{A}_i) = \lambda_0(t) + \boldsymbol{A}_i^T \boldsymbol{\beta}(t), \tag{2}$$

where $\lambda_0(t)$ is an unspecified baseline hazard function, $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_q(t))$ is a $q \times 1$ vector of time-varying coefficients, and $\boldsymbol{A}$ is a $q \times 1$ vector of time-fixed covariates. It is noted that the most attractive feature of Aalen's additive hazards model is that it naturally allows time-varying coefficients, so it provides a flexible alternative to the Cox model (Cox, 1972) when the assumption of proportional hazards does not hold. Pros and cons of additive models are discussed by

Martinussen and Scheike (2006) and Aalen et al. (2008). Previously, Lin et al. (1998) have proposed martingale estimating equations to analyze interval-censored data based on additive hazards models in the case with independent observation times, and for the case of dependent visit times, joint random effects models have been considered by Wang et al. (2010, 2018) and Zhao et al. (2015), among others. The methodology we introduce in Section 2.3 adjusts for dependent visit times and results in consistent estimates of the marginal effects of $\boldsymbol{A}_i$ on T_i. Note that model (2) can be rewritten as

$$S(t|\boldsymbol{A}_i) = \exp\{-\Lambda_0(t) - \boldsymbol{A}_i^T \boldsymbol{\mathcal{B}}(t)\},$$

where $\boldsymbol{\mathcal{B}}(t) = (\int_0^t \beta_1(u)du, \ldots, \int_0^t \beta_q(u)du)^T$ and $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ is the cumulative hazard function at t. Thus, it can be viewed as a special case of (1) with $g(x) = \exp(-x)$.

In addition, (1) with time-fixed coefficients, that is, $\boldsymbol{\mathcal{B}}(t) \equiv \boldsymbol{\beta}$, can be considered as a linear transformation model discussed in Kalbfleisch (1978), Dabrowska and Doksum (1988a, 1988b), Cheng et al. (1995), and Chen et al. (2002). It is equivalently written as

$$h(T_i) = -\boldsymbol{A}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{3}$$

where ϵ_i is a random error with distribution function $F = 1 - g$ and g is a known decreasing function mentioned in (1); $h(t)$ is an unspecified strictly increasing function, which maps the positive half-line onto the whole real line. This class includes many popular regression models in survival analysis. For example, $g(x) = \exp\{-\exp(x)\}$ and $g(x) = 1/\{1 + \exp(x)\}$ correspond to proportional hazards and proportional odds models, respectively. If $h(t)$ is specified, it also covers the parametric accelerated failure time models (Dabrowska & Doksum, 1988b).

Model (1) has been applied to competing risks or multistate problems in which $S(\cdot)$ denotes the probability of being in a certain state or death from a certain cause; it is then referred to as direct binomial regression (Azarang et al., 2017; Cook & Lawless, 2018; Fine, 1999; Scheike et al., 2008; Scheike & Zhang, 2007). Since we observe a binary survival indicator $Y_i(t) = I(T_i > t)$ at each visit time t, the model (1) provides a flexible and convenient framework for parameter estimation with failure time data. We call (1) a generalized linear failure time model in the sense that it can be considered as a generalized linear model defined for binary outcome $Y_i(t)$ across all t. Model (1) is also used in the analysis of interval-censored data, primarily for current status data (e.g., Jewell & Van Der Laan, 2004; Shiboski, 1998; among others), under the assumption of independent censoring. In a subsequent section, we develop a novel semiparametric estimation method using the direct binomial regression framework for case-II interval-censored data where failure is known only to have occurred between two consecutive visits, that is, $T_i \in (t_{ij-1}, t_{ij}]$. Inverse-intensity-of-visit (IIV) weights which we introduce in the next section will be adopted to adjust for the informative visit times t_{ij} when the assumption of independent censoring is not satisfied.

2.2 | Review of inverse-intensity-of-visit (IIV) weighting methods

First of all, we review briefly the inverse-intensity-of-visit (IIV) weighting method initially introduced by Lin et al. (2004) for analyzing longitudinal data. It considers visits as a recurrent event process (Cook & Lawless, 2007) and incorporates the inverse of the visit intensity function as a weight into the analysis of the irregularly observed longitudinal data.

We define the counting process $\{N_i(t), t \geq 0\}$ for visits; $N_i(t)$ is the cumulative number of visits for subject i up to time t. We also write $N_i(t) = \int_0^t dN_i(s)$, where $dN_i(s)$ indicates whether individual i has a visit at time s. It is assumed for simplicity that there is a common end of follow-up time τ for all individuals, and we let $C_i(\leq \tau)$ be a random drop-out time for subject i. Then, $C_i(t) = I\{t < C_i\}$ indicates that the individual is still being followed at time t and $dN_i^*(t) = C_i(t)dN_i(t)$ indicates an observed visit at time t; $0 \leq t_{i1} < t_{i2} < \ldots < t_{im_i} < C_i$ denote the m_i random visit times observed for subject i. Since our focus in this article is the discussion about intermittent visit times, in the following development we assume that drop-out time C_i is independent of the failure time and visits processes, conditional on covariates in the failure time model. The case with dependent drop-out times is discussed at the end of Appendix A.3.

We use an overbar to denote the history of a variable process, for example, $\overline{V}_i(t) = \{V_i(s), 0 \leq s \leq t\}$ denotes the history of the covariate $V_i(s)$ through time t for subject i. We also define $\boldsymbol{H_i}(t^-) = \{\overline{C}_i(t^-), \overline{N}_i(t^-), \overline{Y}_i(t^-), \boldsymbol{A_i}, \overline{\boldsymbol{V_i}}(t^-)\}$ as the history of past outcomes, visits, and covariates, where $\boldsymbol{A_i}$ represents a vector of time-fixed covariates of primary interest in the outcome model and $\boldsymbol{V_i}(t)$ denotes a vector of time-varying variables which could be related to the event time outcome process and/or the visit time process. Following Lin et al. (2004), it is assumed that there exists a vector $\boldsymbol{Z_i}(t^-)$ which depends only on the observed components of $\boldsymbol{H_i}(t^-)$ such that

$$E\{dN_i^*(t)|\boldsymbol{H_i}(t^-), Y_i(t)\} = E\{dN_i^*(t)|\boldsymbol{Z_i}(t^-)\} = C_i(t)\lambda_N\{t|\boldsymbol{Z_i}(t^-)\}dt. \tag{4}$$

The assumption (4) is also known as a sequentially missing at random (SMAR) condition (Hogan et al., 2004; Robins et al., 1995). We assume a positivity condition that $\lambda_N\{t|\boldsymbol{Z_i}(t^-)\}$ is positive and bounded for all $\boldsymbol{Z_i}(t^-)$ and $t \in [0, \tau]$.

IIV weights are generally defined by

$$w_i(t) = 1/\lambda_N\{t|\boldsymbol{Z_i}(t^-)\}, \quad \forall\, t \geq 0. \tag{5}$$

The denominator, $\lambda_N\{t|\boldsymbol{Z_i}(t^-)\}$, is the visit intensity at time t given the observed history. Models used in the analysis of recurrent events can be applied to model the visit time process and thus to estimate the intensity. We give a brief review of this in Appendix A.1.

The IIV weighting approach can be extended for analyzing dependently interval-censored failure time data using a weighted version of binomial pseudo-likelihood approach. Zhu et al. (2017, 2018) respectively considered parametric regression models and nonparametric estimation of marginal failure time distributions. In this article, we aim to extend the IIV weighting approach for estimating marginal or partly conditional regression effects in semiparametric regression models (1).

2.3 | Semiparametric estimation for additive hazards models

We consider first an additive hazards model with time-invariant coefficients $\boldsymbol{\beta}$. That is, conditional on a vector of time-fixed covariates $\boldsymbol{A_i}$, failure time T_i has hazard function

$$\lambda(t|\boldsymbol{A}_i) = \lambda_0(t) + \boldsymbol{A}_i^T\boldsymbol{\beta}.$$

For time-varying coefficients, it is noted that $\boldsymbol{\beta}(t)$ can be parametrized by a set of specified functions or smooth basis functions such as B-splines or regression splines. We illustrate this

scenario in our real data application in Section 4 and provide relevant theoretical results in Appendix A.2.

The survivor function of T_i is specified by

$$S(t|A_i) = S_0(t)\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}t), \tag{6}$$

where $S_0(t) = \exp\{-\Lambda_0(t)\}$ is the baseline survivor function which is unspecified. We will use an IIV weighted version of binomial estimation (e.g., Scheike et al., 2008) based on the fact that $E\{Y_i(t)|\boldsymbol{A}_i\} = S(t|\boldsymbol{A}_i)$. Extension from (2.2) in Zhu et al. (2017) and (5) in Zhu et al. (2018) leads to the following IIV-weighted estimating equations for baseline survivor function $S_0(t)$ at any given t and for regression coefficients $\boldsymbol{\beta}$. To obtain a smooth estimate of $S_0(t)$ given the subject-specific intermittent visit times t_{ij}, we use the local smoothing method of Fan (1993) in (7). We consider

$$\begin{aligned}\Psi_{1n}\{S_0(t),\boldsymbol{\beta}\} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} K_h(t-u)w_i(u)\left\{Y_i(u) - S_0(t)\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}u)\right\}dN_i^*(u)\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\left\{Y_i(t_{ij}) - S_0(t)\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}t_{ij})\right\}\\ &= 0,\end{aligned} \tag{7}$$

$$\begin{aligned}\Psi_{2n}\{S_0(t),\boldsymbol{\beta}\} &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\frac{w_i(u)u\boldsymbol{A}_i\left\{Y_i(u) - S_0(u)\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}u)\right\}}{\left\{1 - S_0(u)\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}u)\right\}}dN_i^*(u)\\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{w_i(t_{ij})t_{ij}\boldsymbol{A}_i\left\{Y_i(t_{ij}) - S_0(t_{ij})\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}t_{ij})\right\}}{\left\{1 - S_0(t_{ij})\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}t_{ij})\right\}}\\ &= \mathbf{0},\end{aligned} \tag{8}$$

where $w_i(t_{ij})$ is the IIV weight at t_{ij} for subject i defined in (5); $K_h(u) = K(u/h_n)/h_n$, where $h_n \to 0^+$, as $n \to \infty$, is a positive bandwidth sequence and $K(\cdot)$ denotes a symmetric kernel density function that satisfies the conditions we state later. These estimating equations can be obtained from a weighted pseudo-likelihood function given by summing binomial log likelihood functions based on $Y_i(t_{ij})$.

It can be seen that for a given value $\boldsymbol{\beta}$ the first estimating equation (7) leads to a closed-form estimator of the baseline survivor function, that is,

$$\hat{S}_0(t;\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})Y_i(t_{ij})}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}t_{ij})}. \tag{9}$$

Then, $\hat{\boldsymbol{\beta}}$ is obtained by solving the following profile estimating equation with $\hat{S}_0(t_{ij};\boldsymbol{\beta})$ substituted by (9).

$$\Psi_n\{\boldsymbol{\beta}\} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{w_i(t_{ij})t_{ij}\boldsymbol{A}_i\left\{Y_i(t_{ij}) - \hat{S}_0(t_{ij};\boldsymbol{\beta})\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}t_{ij})\right\}}{\left\{1 - \hat{S}_0(t_{ij};\boldsymbol{\beta})\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}t_{ij})\right\}} = \mathbf{0}. \tag{10}$$

Once we obtain $\hat{\boldsymbol{\beta}}$ from (10), the estimate of baseline survivor function $S_0(t)$ is obtained by $\hat{S}_0(t;\hat{\boldsymbol{\beta}})$ using (9).

The estimating equation (10) has the form of an IIV-weighted version of a generalized estimating equation (Liang & Zeger, 1986) with independence working correlation matrix. It leads to a consistent estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, provided that assumption (4) is true and weight $w_i(t)$ is consistently estimated. We describe the asymptotic properties of $\hat{\boldsymbol{\beta}}$ in Theorem 1 below, under the following conditions:

- Condition 1. As $n \to \infty$, the number of visits m_i is finite for all i with probability one.
- Condition 2. The true baseline survivor function $S_0(t)$ is twice continuously differentiable for all $t \in [0, \tau]$, and we assume that $\sup_{t\in[0,\tau]} E\left|Y_i(t) - S_0(t;\boldsymbol{\beta})\exp(-\boldsymbol{A}_i^T\boldsymbol{\beta}t)\right|^3 < \infty$ for any $\boldsymbol{\beta}$.
- Condition 3. Kernel function $K(\cdot)$ is a bounded twice continuously differentiable symmetric probability density function vanishing outside of a compact set, with $\int u^2K(u)du < \infty$ and $\int K^2(u)du < \infty$.
- Condition 4. Weight $w_i(t) = 1/\lambda_N\{t|\boldsymbol{Z_i}(t^-)\}$ is bounded in $[0, \tau]$ and is consistently estimated.

Theorem 1. *Let $\boldsymbol{\beta}_0 \in int(\mathscr{B})$ be the true value of $\boldsymbol{\beta}$, where $\mathscr{B}$ is a compact subset of R^q. We also let $\hat{\boldsymbol{\beta}}$ be the solution of $\Psi_n(\boldsymbol{\beta}) = \mathbf{0}$ from (10). Under Conditions 1–4, it can be shown that if $h_n \to 0^+$ and $nh_n^4 \to 0$ as $n \to \infty$, $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_0$, and it has the following asymptotic distribution,*

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{D} N(0, D^{-1}VD^{-1}), \tag{11}$$

where

$$V = E\left[\sum_{j=1}^{m_i}\left\{\frac{w_i(t_{ij})t_{ij}\boldsymbol{A}_i}{1-\mu_{ij}} - w_i(t_{ij})Q(t_{ij})\right\}\{Y_i(t_{ij}) - \mu_{ij}\}\right]^{\otimes 2},$$

$$D = E\left[\sum_{j=1}^{m_i}\frac{w_i(t_{ij})t_{ij}\boldsymbol{A}_i\left\{\overline{S}_0^T(t_{ij};\boldsymbol{\beta}_0)\exp\{-\boldsymbol{A}_i^T\boldsymbol{\beta}_0t_{ij}\} - \mu_{ij}t_{ij}\boldsymbol{A}_i^T\right\}}{1-\mu_{ij}}\right],$$

and

$$\mu_{ij} = S_0(t_{ij};\boldsymbol{\beta}_0)\exp\{-\boldsymbol{A}_i^T\boldsymbol{\beta}_0t_{ij}\},$$

$$H(t_{ij}) = E\left[\exp\{-\boldsymbol{A}_i^T\boldsymbol{\beta}_0t_{ij}\}\right],$$

$$Q(t_{ij}) = E\left[\frac{t_{ij}\boldsymbol{A}_i\exp\{-\boldsymbol{A}_i^T\boldsymbol{\beta}_0t_{ij}\}}{1-\mu_{ij}}\right]H^{-1}(t_{ij}),$$

$$\overline{S}_0^T(t_{ij};\boldsymbol{\beta}_0) = S_0(t_{ij})E\left[\boldsymbol{A}_i^T\exp\{-\boldsymbol{A}_i^T\boldsymbol{\beta}_0t_{ij}\}\right]t_{ij}H^{-1}(t_{ij}),$$

where $a^{\otimes 2} = aa^T$ for any vector a.

We remark that the covariance structure of $\hat{\boldsymbol{\beta}}$ presented in Theorem 1 is consistent with the result in theorem 1 in Fan et al. (2007) and theorem 1 in Chen et al. (2018) in the framework of profile likelihood. We obtain variance estimates for $\hat{\boldsymbol{\beta}}$ by replacing expectations in expressions for

V and D with observed sample means evaluated at $\boldsymbol{\beta_0} = \hat{\boldsymbol{\beta}}$, $S_0(t) = \hat{S}_0(t; \hat{\boldsymbol{\beta}})$. The proof of Theorem 1 and discussion about other theoretical results, for example, the asymptotic distribution of $\hat{S}_0(t; \hat{\boldsymbol{\beta}})$, is provided in the appendices A.2–A.5.

The estimator of baseline survivor function $\hat{S}_0(t; \hat{\boldsymbol{\beta}})$ is smooth and defined for any $t \geq 0$, but it is not monotone in general. Similar to theorem 1 in Mammen (1991), it can be shown that $\hat{S}_0(t; \hat{\boldsymbol{\beta}})$ is asymptotically monotone with the use of an optimal bandwidth, if the true function $S_0(t)$ is sufficiently smooth and strictly decreasing. To achieve monotony for finite samples, we suggest an isotonic regression (Barlow et al., 1972) applied to the cumulative distribution function estimator $\hat{F}_0(t; \hat{\boldsymbol{\beta}}) = 1 - \hat{S}_0(t; \hat{\boldsymbol{\beta}})$, similar to the smooth-isotonic (SI) estimator discussed in Mammen (1991). The monotonized estimator $\hat{F}_0^{\dagger}(t; \hat{\boldsymbol{\beta}})$ can be given by the max–min formula (Barlow et al., 1972) or by the pool-adjacent-violators algorithm (PAVA) (Miles, 1959). We use R function *isoreg* to implement the PAVA in subsequent simulation and real data analysis.

So far, we propose the IIV-weighted estimators of baseline survivor function $S_0(t)$ and regression coefficients $\boldsymbol{\beta}$ for additive hazards models. Next, we show how to extend this estimation method and algorithm to the class of linear transformation models.

2.4 | Extension to linear transformation models

In this section, we consider the linear transformation models in the form

$$S(t|\boldsymbol{A}_i) = g\{h(t) + \boldsymbol{A}_i^T \boldsymbol{\beta}\}. \tag{12}$$

Semiparametric estimation of $\boldsymbol{\beta}$ based on model (12) has been studied by Cheng et al. (1995), Fine et al. (1998), Chen et al. (2002), and Zeng and Lin (2006), among others, for right-censored time-to-event data. Here, we introduce an iterative algorithm that incorporates the IIV weights for dependently interval-censored data.

Similar to Section 2.3, we construct two IIV-weighted estimating equations to estimate $h(t)$ for any t and $\boldsymbol{\beta}$

$$\Psi_{1n}\{h(t), \boldsymbol{\beta}\} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t - t_{ij}) w_i(t_{ij})[Y_i(t_{ij}) - g\{h(t) + \boldsymbol{A}_i^T \boldsymbol{\beta}\}] = 0, \tag{13}$$

$$\Psi_{2n}\{h(t), \boldsymbol{\beta}\} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{w_i(t_{ij})\boldsymbol{A}_i g'(\mu_{ij})\{Y_i(t_{ij}) - g(\mu_{ij})\}}{g(\mu_{ij})\{1 - g(\mu_{ij})\}} = \boldsymbol{0}, \tag{14}$$

where $\mu_{ij} = h(t_{ij}) + \boldsymbol{A}_i^T \boldsymbol{\beta}$ and $g'(x) = dg(x)/dx$ for any variable x.

Unlike (7) for additive hazards models, the local-smoothed estimating equation for nonparametric component $h(t)$, given in (13), does not have a closed-form solution. We propose an iterative estimation procedure that alternates between updating estimates of $\boldsymbol{\beta}$ and $h(t)$, similar to that described in Chen et al. (2002). It is summarized as follows.

Step 1: specify an initial value for $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_{(1)}$, which can be given by a naive estimate based on model (12) with dependent interval censoring not adjusted for.

Step 2: given $\hat{\boldsymbol{\beta}}_{(k)}$ for any $k \geq 1$, solve the estimating equation $\Psi_{1n}\{h(t), \hat{\boldsymbol{\beta}}_{(k)}\} = 0$ based on (13) to obtain the estimate $\hat{h}(t)$.

Step 3: substitute $\hat{h}(t)$ in $\Psi_{2n}\{h(t), \boldsymbol{\beta}\} = \mathbf{0}$ given in (14) to obtain an updated estimate of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_{(k+1)}$.

Step 4: increase k by one and go back to steps 2 and 3; repeat until specified convergence criteria are satisfied, for example, $||\hat{\boldsymbol{\beta}}_{(k+1)} - \hat{\boldsymbol{\beta}}_{(k)}|| < 10^{-4}$.

The asymptotic distribution of $\hat{\boldsymbol{\beta}}$ could be developed similarly as in Theorem 1, but the absence of a closed-form estimator $\hat{h}(t)$ will lead to additional technical challenges; we consider this as future work. Alternatively, we propose a nonparametric bootstrap to estimate the standard error of $\hat{\boldsymbol{\beta}}$ and examine the performance of our method based on a Cox model via simulation in Section 3.2.

3 | SIMULATION STUDIES

3.1 | Simulation study for an additive hazards model

In this section we examine the finite sample performance of the proposed method for an additive hazards model. We consider a setting where models involving a single covariate A are of interest, but there exists another observed covariate V which is probably related both to T and to the visit process. Martinussen and Vansteelandt (2013) discussed the collapsibility of additive hazards models, which facilitates the simulation scenarios we consider. We simulate the exposure variable A_i from a Bernoulli(0.5) distribution. We then generate an auxiliary variable V_i from $N(A_i, 1)$. It is assumed that both A_i and V_i are risk factors of failure time T_i and may also be predictors of patients' visit times. We generate T_i from an exponential distribution with hazard function given by $\lambda(t|V_i, A_i) = \beta_0^* + \beta_1^* V_i + \beta_2^* A_i$. Following Martinussen and Vansteelandt (2013), it can be shown that the distribution of T_i conditional on A_i alone is still of additive hazards form with hazard function

$$\lambda(t|A_i) = \lambda_0(t) + \beta_2 A_i, \tag{15}$$

where $\lambda_0(t) = \beta_0^* - \beta_1^{*2} t$, and $\beta_2 = \beta_2^* + \beta_1^*$. We let $\beta_0^* = 0.5$ for nonnegative β_2 and 0.8 for negative β_2, $\beta_1^* = 0.2$, and $\beta_2^* = -0.6, -0.4, -0.2, 0$, or 0.2 so that true values of β_2 range from -0.4 to 0.4 with an increment of 0.2 to represent various scenarios. Model (15) gives the survivor function

$$S(t|A_i) = S_0(t) \exp(-\beta_2 A_i t), \tag{16}$$

where $S_0(t)$ is treated nonparametrically and regression coefficient β_2 is of primary interest for estimation and inference.

To generate random and irregular visit times, we discretize the time scale using a very fine grid (100 per unit of time) so that visit times simulated approximate those from a continuous time process; this represents what occurs in practice, where time is usually recorded in discrete units like days, though continuous time models are typically used for analysis. That is, the visit process is discrete on the grid $0 = a_0 < a_1 < \ldots < a_M = 5$ with $a_k = 0.01k$ for $k = 1, \ldots, M$ with $M = 500$. At each a_k, we generate a binary visit indicator dN_{ik} based on model $P(dN_{ik} = 1|A_i, V_i) = \exp(\gamma_0 + \gamma_1 V_i + \gamma_2 A_i)$ with $\gamma_0 = -4.5$, $\gamma_1 = 0$ or 0.8 to indicate an independent visit scheme or dependent visit scheme, respectively, and $\gamma_2 = 0.1$. As a result, the times between visits have a geometric distribution, with about 8 visits per person for the group with $A_i = 0$ and 18 visits per person for the group with $A_i = 1$. This mimics a scenario common in clinical practice where A can be viewed as a treatment indicator and patients who receive more aggressive treatment are expected to be

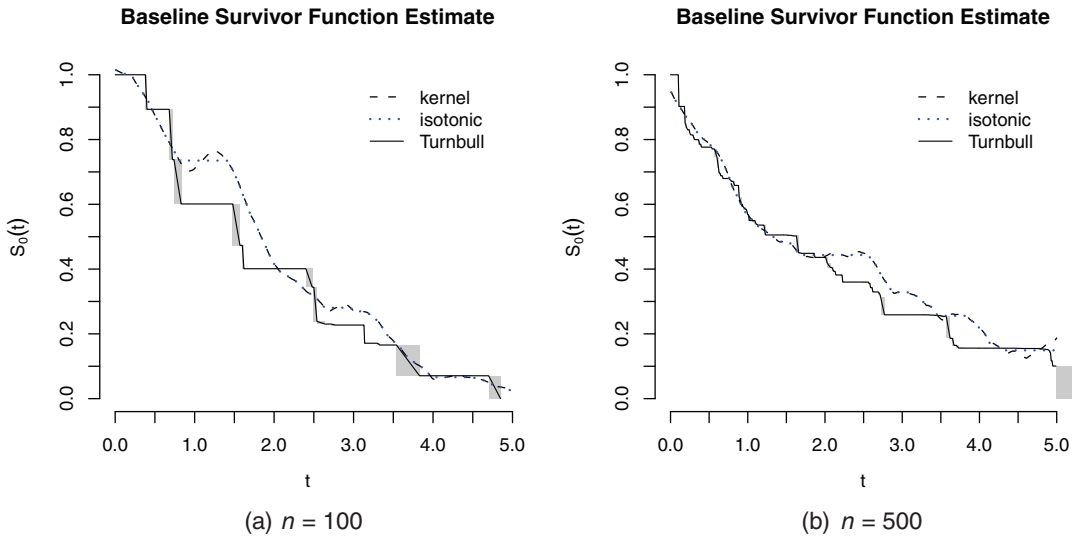

FIGURE 1 Kernel estimator by (9), monotonized by an isotonic estimator, and the Turnbull's estimator of baseline survivor function $S_0(t)$ from one simulation for $n = 100$ or $n = 500$, where $\beta_2 = 0.4$, $\gamma_1 = 0.8$

observed more frequently. It is noted that when $\gamma_1 \neq 0$, V_i is a shared risk factor for failure and visit times but is not included in the target model (16). This renders visit times outcome-dependent with respect to that model. On the other hand, when $\gamma_1 = 0$ the visit process is independent of failure time. We set administrative end of follow-up for all individuals as $\tau = 5$; this results in about 80% of the subjects with $A_i = 0$ and 95% of those with $A_i = 1$ experiencing failure by then. For simplicity we do not allow premature loss to follow-up, so all $C_i(t) = 1$.

We perform estimation using the procedure in Section 2.3. We treat the visit time process as though it was continuous and estimate the IIV weights $w_i(t)$ by fitting model (A3) with $\boldsymbol{Z_i}(t^-) = (A_i, V_i)^T$ and cut-points (0.40, 0.75, 1.0). That is,

$$\lambda_N\{t|A_i, V_i; \psi_1, \psi_2\} = \sum_{j=1}^{4} \rho_j I_j\{G_i(t)\} \exp(\psi_1 A_i + \psi_2 V_i). \tag{17}$$

This family of models contains a close approximation to the discrete process used to simulate visit times. We solve score function (A4) using R function *phreg* in package *eha*. A marginal visit intensity, estimated by the increments of the Nelson–Aalen estimator of the cumulative rate function, that is, $\hat{s}(t) = \sum_{i=1}^{n} dN_i^*(t) / \sum_{i=1}^{n} C_i(t)$, is used to stabilize the IIV weights.

Simulation results about regression parameter β_2 in model (15) or (16) are summarized in Table 1. Estimating equation (10) is solved by R function *nleqslv*. Standard errors of $\hat{\beta}_2$ are obtained as described following Theorem 1. To obtain a monotone estimate of the baseline survivor function, as discussed at the end of Section 2.3, an isotonic regression is applied to the kernel smooth estimator $\hat{S}_0(t; \hat{\beta}_2)$, shown in Figures 1 and 2. The isotonic estimator of cumulative distribution function, that is, $\hat{F}_0^{\dagger}(t; \hat{\beta}_2)$, is obtained by using R function *isoreg*, and then $\hat{S}_0^{\dagger}(t; \hat{\beta}_2) = 1 - \hat{F}_0^{\dagger}(t; \hat{\beta}_2)$. Our simulation is conducted based on R version 3.6.2. Sample sizes considered were $n = 100$ and $n = 500$, and $N = 500$ simulated samples were generated for each setting.

TABLE 1 Simulation summary of parameter β_2 in model (15)

		Unstabilized weights				Stabilized weights				Midpoint approx			
	TRUE	Bias	SSE	ASE	ECP	bias	SSE	ASE	ECP	Bias	SSE	ASE	ECP
$n=100$	−0.4	−0.011	0.156	0.143	0.944	−0.015	0.164	0.156	0.942	0.053	0.112	0.112	0.912
	−0.2	−0.006	0.166	0.156	0.946	−0.010	0.171	0.159	0.948	0.036	0.124	0.126	0.934
$\gamma_1=0$	0	0.002	0.110	0.109	0.954	0.001	0.113	0.112	0.958	0.006	0.091	0.095	0.964
	0.2	0.009	0.133	0.134	0.950	0.008	0.136	0.136	0.960	−0.014	0.105	0.111	0.958
	0.4	0.020	0.168	0.162	0.950	0.020	0.175	0.165	0.946	−0.049	0.117	0.127	0.954
$\gamma_1=0.8$	−0.4	−0.003	0.163	0.152	0.940	−0.005	0.166	0.156	0.946	0.048	0.115	0.117	0.934
	−0.2	−0.005	0.174	0.164	0.938	−0.008	0.177	0.168	0.944	0.043	0.134	0.133	0.936
	0	0.001	0.115	0.114	0.952	0.000	0.117	0.117	0.950	−0.005	0.097	0.102	0.966
	0.2	0.008	0.140	0.137	0.948	0.007	0.143	0.140	0.946	−0.009	0.116	0.121	0.962
	0.4	0.019	0.172	0.164	0.938	0.019	0.177	0.166	0.944	−0.026	0.134	0.141	0.944
$n=500$	−0.4	0.005	0.065	0.063	0.938	0.005	0.066	0.064	0.936	0.123	0.044	0.044	0.220
	−0.2	0.000	0.072	0.069	0.942	0.000	0.074	0.070	0.938	0.070	0.050	0.049	0.692
$\gamma_1=0$	0	−0.001	0.051	0.049	0.932	−0.001	0.051	0.049	0.936	0.001	0.039	0.038	0.940
	0.2	0.000	0.061	0.059	0.938	−0.001	0.062	0.060	0.932	−0.055	0.044	0.044	0.754
	0.4	0.004	0.074	0.073	0.946	0.004	0.075	0.074	0.942	−0.129	0.048	0.049	0.240
$\gamma_1=0.8$	−0.4	0.009	0.073	0.070	0.934	0.009	0.073	0.071	0.942	0.117	0.049	0.046	0.306
	−0.2	0.003	0.079	0.075	0.944	0.003	0.080	0.076	0.944	0.073	0.054	0.052	0.692
	0	0.000	0.053	0.051	0.930	0.000	0.054	0.052	0.930	−0.007	0.043	0.041	0.924
	0.2	0.000	0.063	0.062	0.938	0.000	0.064	0.062	0.930	−0.054	0.049	0.047	0.778
	0.4	0.004	0.076	0.074	0.950	0.004	0.077	0.075	0.948	−0.116	0.055	0.054	0.410

Note: Parameter γ_1 is set to be 0 for an independent visit process and to be 0.8 for a dependent visit process. SSE denotes the sampling standard error; ASE is the sampling mean of asymptotic standard errors; ECP is the empirical coverage probability calculated for 95% CIs. Bandwidth h is specified to be $h=0.5$ for sample size $n=100$ and $h=0.3$ for sample size $n=500$. Simulation results are summarized based on 500 replicates.

We want to compare the proposed estimators with existing methods. However, the literature on semiparametric additive hazards models with interval-censored data is very limited, and we did not find any method which is easily implementable. Instead, we use the midpoint of $(t_{ij-1}, t_{ij}]$ where T_i fell as the approximate occurrence time and then analyze the mimicked right-censored data by R function *aalen* in package *timereg*. This midpoint approximation method is widely adopted in practice to analyze interval-censored data especially when visit gaps are not very wide. In Table 1, it is seen that for each scenario, our proposed estimators (stabilized or nonstabilized) have very small bias and moderate variance. The average asymptotic standard errors are close to the sampling standard errors. Empirical coverage probabilities of the 95% confidence intervals (CIs) for β_2 are close to the nominal level. On the other hand, midpoint approximation method leads to nonnegligible bias, and bias increases with sample size, making coverage probabilities of 95% CIs far away from the nominal level for $n=500$, while it is hard to differentiate whether bias comes from the ignorance of informative visit times or from approximation.

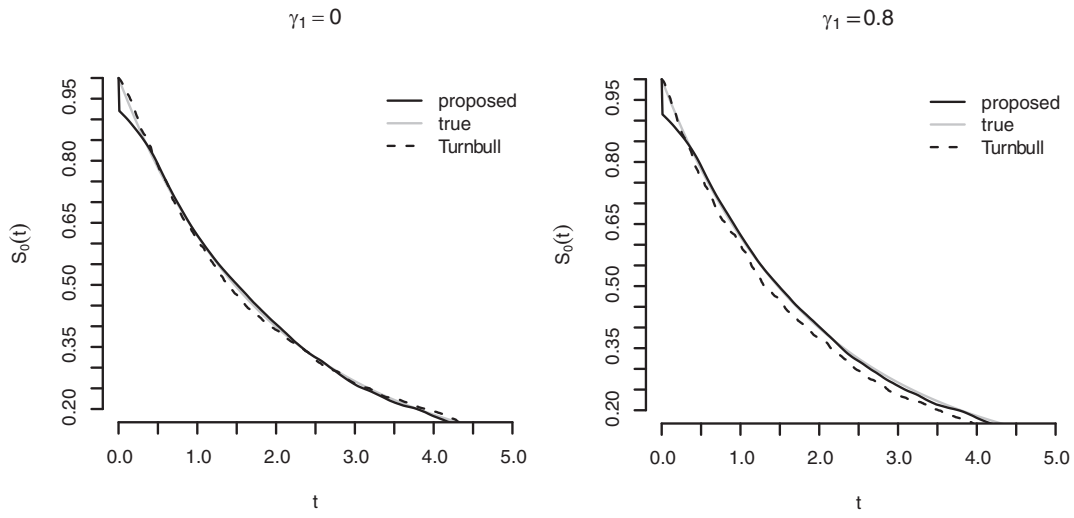

FIGURE 2 Estimates of baseline survivor function $S_0(t)$ averaged over 100 simulations for $n = 100$ and $\beta_2 = 0.4$. Independent visit scheme and dependent visit scheme are simulated by specifying $\gamma_1 = 0$ or $\gamma_1 = 0.8$

In Table 1, we note that stabilized estimators have slightly larger variances than the unstabilized estimators. This is because our baseline visit intensity $\lambda_{N0}\{G_i(t)\}$ in (A2) is defined with the elapsed time from the most recent previous visit related to an individual's visit history, so it is not cancellable by $\hat{s}(t)$. This is different from the cases in Lin et al. (2004) and Buzkova and Lumley (2007, 2009) where their denominator and numerator are both based on Markov models so baseline intensities are canceled or exempted. Although our stabilized weights lead to a bit more variation of $\hat{\beta}_2$, we found in additional simulations (not given here) and in Zhu et al. (2018) that it can improve the normal approximation for $\hat{\beta}_2$ when weights have extremely large variability. Therefore, we recommend stabilized weights for real data analysis where intervisit times vary widely, as in the analysis of the PsA data in Section 4.

We employ the Epanechnikov kernel $K(u) = (3/4)(1 - u^2)$ for $-1 \le u \le 1$ for calculating $\hat{S}_0(t; \beta_2)$ in (9) throughout this article. Bandwidth selection should also satisfy the conditions in Theorem 1. In our simulation, we selected bandwidths which minimize the mean squared error of estimates, that is, $\frac{1}{500}\sum_{s=1}^{500}(\hat{\beta}_{2s} - \beta_{2s})^2$, among $\{0.1, 0.2, \ldots, 1.0\}$. As a result, $h = 0.5$ was selected for $n = 100$ and $h = 0.3$ was selected for $n = 500$, though simulation results differ very slightly with different bandwidths. A bias-variance tradeoff was noticed in nonparametric estimation in Zhu et al. (2018), but it seems that semiparametric estimators of regression coefficients are not very sensitive to bandwidth selection. In practice, bandwidth is often selected data adaptively via cross-validation while adopting some assessment metrics (e.g., minimization of prediction error). Optimal choices of kernel bandwidths may improve the asymptotic properties and finite sample performance of estimators; this will be considered in future work.

Figure 1 shows the kernel estimate $\hat{S}_0(t; \hat{\beta}_2)$ and the isotonic estimate $\hat{S}_0^{\dagger}(t; \hat{\beta}_2)$ from one simulation sample for each of $n = 100$ and $n = 500$. It is seen in Figure 1 that isotonic regression corrects segments of the kernel estimate where it is increasing, while for a larger sample size (e.g., $n = 500$), our baseline survivor estimator (9) is further smoothed and closer to be monotonic. Turnbull's (Turnbull, 1976) estimator is a well-known nonparametric estimator of the survivor function for interval-censored data. However, like many other standard methods, it requires that censoring is independent. We plot Turnbull's estimates in Figures 1 and 2 for comparison with our estimates.

TABLE 2 Investigation of the estimation of parameter β_2 in model (15) under misspecified visit time model

	Wrong weight model				Correct weight model			
TRUE	**Bias**	**SSE**	**ASE**	**ECP**	**Bias**	**SSE**	**ASE**	**ECP**
−0.4	−0.012	0.066	0.062	0.942	0.005	0.067	0.065	0.942
−0.2	−0.015	0.072	0.068	0.930	0.001	0.073	0.070	0.944
0	−0.016	0.050	0.048	0.926	−0.002	0.050	0.049	0.932
0.2	−0.015	0.060	0.058	0.930	−0.001	0.060	0.059	0.944
0.4	−0.012	0.072	0.070	0.932	0.002	0.072	0.071	0.952

Note: Parameter γ_1 is set to be 0.8 to simulate a dependent visit process. Sample size is $n = 500$. Bandwidth is given by $h = 0.3$. Simulation results are summarized based on 500 replicates.

In Figure 1, we see that Turnbull's estimate has regions where it is not uniquely defined, shown by gray areas in the plot. Our kernel estimator defined in (9) does not have that problem as long as at least one visit falls in the window $[t-h, t+h]$ for any t. In addition, Turnbull's estimator does not agree with our proposed estimator in some places. Because visit times are dependent when $\gamma_1 = 0.8$, this may cause Turnbull's estimator to be biased. This is demonstrated in Figure 2, where we plot our estimates $\hat{S}_0^{\dagger}(t; \hat{\beta}_2)$ and Turnbull's estimates averaged over 100 simulations in comparison with the true curves. We see that both estimators agree with the true curve when visit times are independent (i.e., $\gamma_1 = 0$), but when $\gamma_1 = 0.8$, Turnbull's estimator shows some bias.

As suggested by reviewers, we examine the robustness of our IIV-weighted method under a misspecified weight model. In particular, we simulate visit indicators using the model $P(dN_{ik} = 1|A_i, V_i) = \exp\{-4 + 0.8V_i + 0.1A_i - 0.05(V_i + 2)^2\}$ but still fit the same working model for the estimation of IIV weights using (17). In Table 2, it can be seen that under the misspecified weight model, the proposed estimator has similar variance but a little larger bias, and coverage probabilities of 95% CIs of regression coefficient β_2 are slightly lower than the nominal level. In addition, we note that models for the visit process can be checked using known methods (e.g., Cook & Lawless, 2007), so significant misspecification of the model used to obtain weights is avoidable.

So far, we focus on the discussion of the case where visits continue to occur after a failure event; that is, the visit process does not discontinue following the occurrence of failure. This is the case in many settings in clinical practice where nonterminal events are of research interest, for example, joint damage progression for patients with PsA. The case where visits stop after a terminal failure event such as death is briefly discussed in Section 5. The methods proposed here can readily be modified to deal with that case.

3.2 | Simulation study for a Cox proportional hazards model

Examination of an arbitrary generalized linear failure time model in Section 2.1 by simulation is difficult since data generation models in which integrating (collapsing) over some covariates results in a marginal model of the desired form may not exist. For example, a Cox model with some covariates marginalized does not in general produce a marginal Cox model; the collapsibility of Cox models is discussed in Martinussen and Vansteelandt (2013). We consider a special case of the Cox model related to a normal model for covariates, and then derive the marginal model by using the normal distribution's collapsibility. It can be shown that the marginal model in this case retains the form of proportional hazards. More details can be found in Zhu et al. (2017); below we briefly describe how data were simulated based on a Cox model.

TABLE 3 Simulation summary of parameter θ_1 based on a Cox model

		IIV-weighted method				Semiparametric MLE			
		Bias	**SSE**	**BSE**	**ECP**	**Bias**	**SSE**	**BSE**	**ECP**
$n = 100$	$\gamma_1 = 0$	0.084	0.249	0.301	0.964	−0.086	0.288	0.378	0.984
	$\gamma_1 = 0.8$	0.060	0.511	0.467	0.912	0.059	0.305	0.412	0.956
$n = 500$	$\gamma_1 = 0$	0.037	0.119	0.127	0.950	−0.029	0.120	0.118	0.930
	$\gamma_1 = 0.8$	<0.001	0.285	0.273	0.940	0.127	0.121	0.120	0.806

Note: Parameter γ_1 is set to be 0 for an independent visit process and to be 0.8 for a dependent visit process. SSE denotes the sampling standard error; BSE is the sampling mean of bootstrapped standard errors estimated by 100 bootstrap samples; ECP is the empirical coverage probability calculated for 95% CIs. Bandwidth h is specified to be 0.5 when $\gamma_1 = 0$ and 0.8 when $\gamma_1 = 0.8$ for sample size $n = 100$ and 0.3 for sample size $n = 500$. Simulation results are summarized based on 500 replicates.

We still generate A_i which is of primary interest from a Bernoulli(0.5) distribution and then generate a time-varying auxiliary variable $V_i(a_k)$ from $N(-2A_i, 3^2)$. To simulate the failure time outcome T_i, we assume that the discrete time hazard of T_i at time a_k is $P(Y_{ik} = 0|\overline{V}_{ik-1}, A_i, \overline{Y}_{ik-1} = 1) = \Phi(\eta_0 + \eta_1 V_{ik-1} + \eta_2 A_i)$, where Φ denotes the distribution function of a standard normal distribution. The marginal model controlling for A_i alone can be shown to have $\lambda_{ik} \approx P(Y_{ik} = 0|A_i, \overline{Y}_{ik-1} = 1) = \Phi\{c\,\{\eta_0 + (\eta_2 + \eta_1\beta_1)A_i\}\}$, where $c = 1/\sqrt{1 + 9\eta_1^2}$; it can also be rewritten in proportional hazards form as $P(Y_{ik} = 0|A_i, \overline{Y}_{ik-1} = 1) = e^{\theta_0}\exp(\theta_1 A_i)$, where $\theta_0 = \log\{\Phi(c\,\eta_0)\}$, $\theta_1 = \log\left\{\frac{\Phi[c\,(\eta_0+\eta_2+\eta_1\beta_1)]}{\Phi(c\,\eta_0)}\right\}$. Then Y_{ik}, $k = 1, \ldots, M$, are generated from Bernoulli(λ_{ik}) distributions until a zero occurs. Our objective in this simulation study is to use the iterative algorithm described in Section 2.4 to estimate the marginal effect of A on T, that is, θ_1. Parameters are specified by $\eta_0 = -7$, $\eta_1 = 1.2$, and $\eta_2 = 0.5$. The true value of θ_1 is −1.26 in this setting. The visit process is simulated the same as in Section 3.1; we let $\gamma_0 = -3.5$, $\gamma_1 = 0$, and $\gamma_2 = -0.5$ for independent visit times and $\gamma_0 = -6$, $\gamma_1 = 0.8$, and $\gamma_2 = 0.5$ for dependent visit times.

Weight estimation is the same as in Section 3.1. Model components are estimated by solving (13) and (14) following the algorithm described in Section 2.4. In Table 3, we compare our proposed method with a semiparametric maximum likelihood estimation (MLE) algorithm implemented by R function *ic_sp* in package *icenReg* (Anderson-Bergman, 2017). Using our selection criteria, for $n = 100$, the bandwidth h was chosen to be 0.5 for the case of independent visit times ($\gamma_1 = 0$) and be 0.8 for dependent visit times ($\gamma_1 = 0.8$); for $n = 500$, h was chosen to be 0.3 regardless the specification of γ_1. Nonparametric bootstrap standard errors are estimated by the standard errors of the estimates from 100 bootstrap samples for each method. It is seen in Table 3 that when sample size is small the MLE based on an assumption of independent visit times is not very biased and it performs similarly as our proposed method. However, for larger sample size, for example, $n = 500$, it leads to substantially larger bias than our proposed method and to coverage probabilities lower than the 95% nominal level when $\gamma_1 = 0.8$.

4 | APPLICATION TO A PSA COHORT

The University of Toronto Psoriatic Arthritis (PsA) Clinic registry was established in 1978 by Dr. Dafna Gladman (Gladman & Chandran, 2010). So far about 1200 patients have been recruited to the clinic, making it one of the largest cohorts of PsA in the world. PsA may, over time,

lead to damage in a patient's joints. At intermittent visits to the clinic the joints are examined either physically or radiologically, and a large number of other measurements are also made. Visits are suggested by protocol 6–12 months apart, but their actual timing varies considerably over time and across individuals. In fact, visit times are related to prior disease history, past visit history, and a variety of disease-related variables. For example, patients who have stronger disease activity or receive more aggressive treatment are found more likely to visit frequently and regularly.

Human leukocyte antigens (HLA) have been discovered to be strongly predictive of the progression of rheumatoid arthritis. We have compared the risks of arthritis mutilans, a severe stage of the disease which is defined as having five or more joints (out of 64) with the highest grade of damage, with respect to HLA-B27, and estimated the distribution of time to arthritis mutilans for each group separately (Zhu et al., 2018). Our objective in this analysis is to assess the association between HLA-B27 and the risk of first joint damage. This can be accomplished by fitting a semiparametric regression model with $A_i = I(\text{HLA-B27 is positive})$ as a covariate, where $I(\cdot)$ is the indicator function.

Our analysis focuses on 407 patients who did not have clinical joint damage at their time of enrolment (340 with HLA-B27 negative and 67 with HLA-B27 positive). We study the time to the first presence of joint damage up to 15 years after enrolment for these patients. Among them, it is found that 192 (154 with HLA-B27 negative, 38 with HLA-B27 positive) developed joint damage by a later follow-up visit, and 215 still had no joint damage at the end of 15 years follow-up. Joint damage is only assessed at visits to the clinic, so the occurrence time of first joint damage is subject to interval censoring.

These patients enrolled over a substantial period of time, which affects their follow-up length and number of visits. The B27 negative group had an average of 10 visits and the B27 positive group nine visits per person, with the average time between visits approximately 1.5 years. Thus, we choose the bandwidth for our kernel estimators of baseline survivor function to be 1 year. For estimating the IIV weights, visit gap times are fitted with model (A3) with cut-points selected according to a nonparametric estimate of the baseline intensity. Covariates included in the visit gap time model are gender, age*, psoriasis (PS) duration and PsA duration at the time of enrolment, family history of PS (yes/no), family history of PsA (yes/no), and the status for human leukocyte antigen HLA-B27* (positive/negative) as time-fixed variables and erythrocyte sedimentation rate* (ESR), total number of active (inflamed) joints, use of NSAIDs (nonsteroidal anti-inflammatory drugs), DMARDs* (disease-modifying antirheumatic drugs), or biologics* medications, year* (of most recent visit), and median length* (of past visit gaps) as time-varying variables. Time-varying covariate values are measured at the most recent visit time except for medication status (NSAIDs, DMARDs, and biologics) which could change between visits. Factors significant at $\alpha = .05$ are annotated by *. The Nelson–Aalen (NA) estimate of the marginal visit intensity is employed to stabilize the IIV weights; the value of $\hat{s}(t_{ij})$ was taken as the increment in the NA estimate at time t_{ij}. The stabilized weights have first, second, and third quartiles 0.57, 1.65, and 3.33.

We consider and compare here a variety of outcome models in the generalized linear family; the models are seen to agree quite closely, and we then compare the additive hazards models with nonparametric estimates. We first consider the linear transformation model (3) where the distribution function of ϵ has the form

$$F_\epsilon(t) = \begin{cases} 1 - \{1 + r\exp(t)\}^{-1/r} & \text{if } r \neq 0 \\ 1 - \exp\{-\exp(t)\} & \text{if } r = 0 \end{cases}.$$

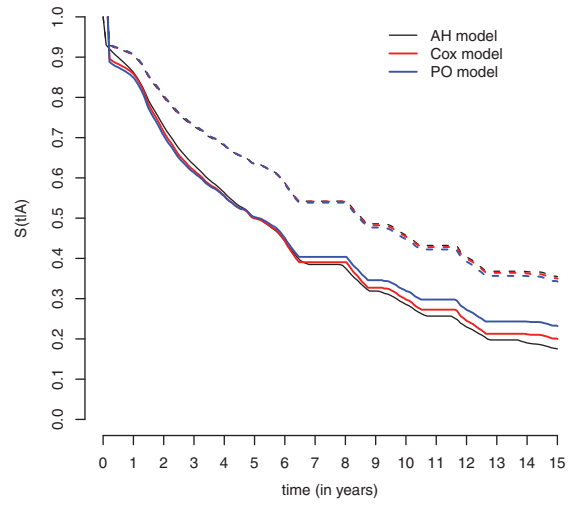

FIGURE 3 Estimates of survivor functions $S(t|A)$, $A = 0$ or 1, based on additive hazards (AH) model, Cox model and proportional odds (PO) model; regression coefficient is assumed to be time-invariant. Solid curves denote that HLA-B27 is positive ($A = 1$) and dashed curves denote that HLA-B27 is negative ($A = 0$)

This family has been discussed by Dabrowska and Doksum (1988a), Chen et al. (2002), and Zeng and Lin (2006), among others; it covers the Cox model with $r = 0$ and proportional odds (PO) model with $r = 1$, respectively. Figure 3 displays the fitted survivor functions with $r = 0$ or 1 and compares them with the time-invariant coefficient additive hazards model (6). It is seen that all these models give very similar estimates of the survivor functions up to 6 years from enrolment. The Cox model agrees closely with the additive hazards model beyond 6 years; the PO model differs gradually from the other two after 6 years for the HLA-B27 positive group. However, there are only six patients who have the failure event occurring after 6 years for the HLA-B27 positive group. Each model results in estimates with large variance in that region. The additive hazards model gives the regression coefficient estimate 0.047 with a standard error of 0.025, giving an approximate 95% confidence interval $(-0.002, 0.096)$, so the p-value for a test of no difference in the risks of first joint damage between patients with and without HLA-B27 is just over 5%.

Next, we consider an additive hazards model with time-varying coefficient (2) in addition to the constant coefficient model. Different formulations of $\beta(t)$ (t is in years) are considered as follows.

$$\text{Model 1: } \beta(t) = \beta_1, \quad \text{Model 3: } \beta(t) = \beta_1 I(t \le 6) + \beta_2 I(t > 6),$$

$$\text{Model 2: } \beta(t) = \beta_1 + \beta_2 e^{0.01t}, \quad \text{Model 4: } \beta(t) = \beta_1 + \beta_2 t + \sum_{j=3}^{5} \beta_j C_j(t).$$

Model 4 uses restricted cubic splines with five knots, where $C_j, j = 3, 4, 5$, denote three piecewise cubic basis functions. About the number of knots, three to five knots are often selected in practice, and it is found that five knots are enough to provide a good fit to most real data (Harrell Jr., 2015; Stone, 1986). We chose the five knots as equally spaced percentiles (5%, 20%, 35%, 50%, and 65%) of the marginal distribution of time to first joint damage using Turnbull's estimator up to the last failure time in the whole sample, so as to have an equal number of events contributing to the estimation of $\beta(t)$ at a given time t. The resulting estimates of coefficients in Model 4 are $\beta_1 = -0.153$, $\beta_2 = 0.001$, $\beta_3 = -0.012$, $\beta_4 = 0.025$, and $\beta_5 = -0.022$.

To assess model fit, we compare the estimates from Models 1 to 4 with the IIV-weighted nonparametric estimates following the method introduced in Zhu et al. (2018) by plotting the

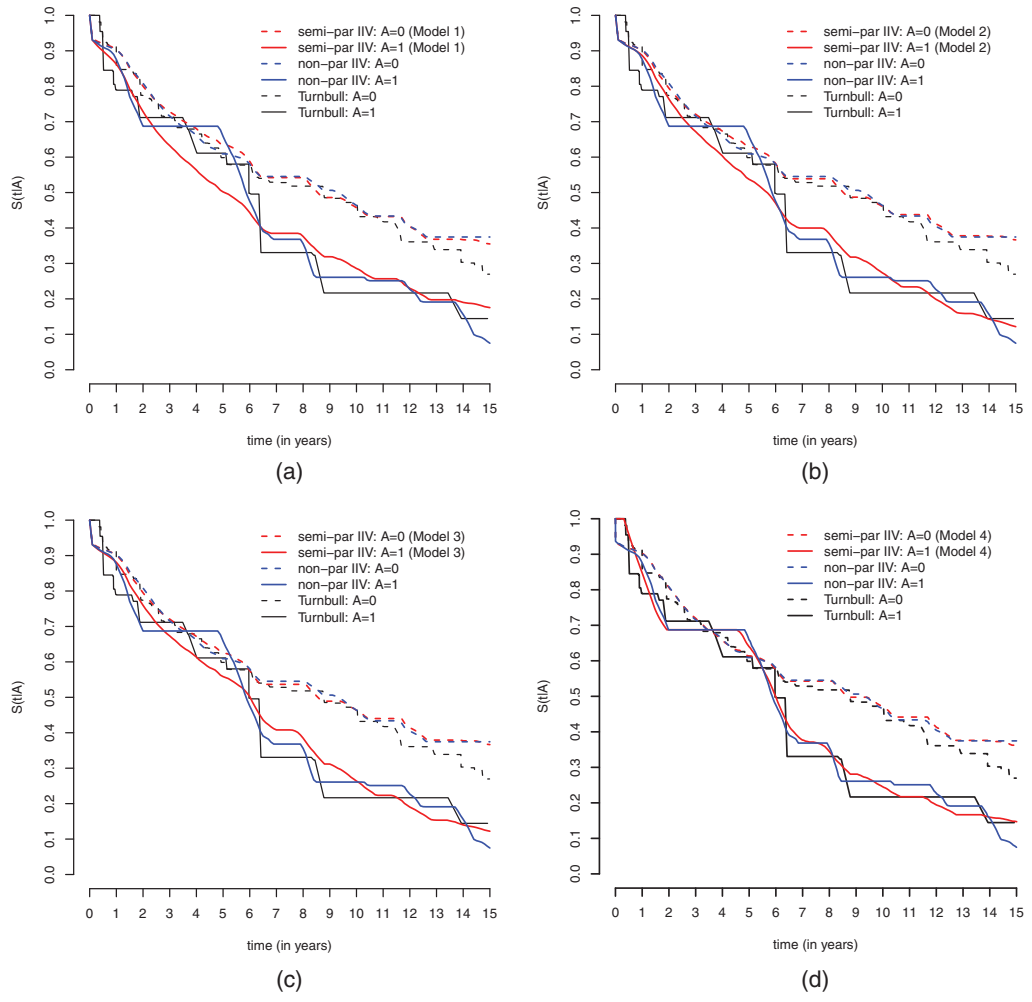

FIGURE 4 Estimates of survivor functions $S(t|A)$ based on time-invariant and different forms of time-varying coefficient AH models with $A = I$(HLA-B27 is positive)

estimated survivor curves $S(t|A)$ for each group in Figure 4. It is seen that Model 4 with cubic splines for coefficients is in close agreement with the nonparametric estimate for each group which is attributed to its flexibility. The other three models produce similar results but agree slightly less well with the nonparametric estimate for group $A = 1$. We also plot Turnbull's estimates as an unadjusted method for comparison. We can see from Figure 4 that they do not differ much from the proposed estimator except before 2 years for the positive group and after 10 years for the negative group.

5 | DISCUSSION

Irregular outcome-related visit times are common in longitudinal studies of chronic disease and in other settings. We propose an IIV-weighted semiparametric estimation approach to analyze

interval-censored failure time data when the assumption of independent censoring does not hold. This approach adjusts for dependent interval censoring and allows estimation in marginal and partly conditional failure time models, which are often of interest in clinical studies.

Modeling failure time data and longitudinal data jointly by specifying submodels for each process and employing random effects to account for interprocess correlation is an alternative way to handle informative visit times. As we discussed in Section 1, interpretation of estimated coefficients in the model for failure time data is different in that case, and depends on unobservable random effects. Our method, on the other hand, gives direct marginal effects of covariates in the failure time model of interest.

We focus on situations where visits do not discontinue after failure occurs. This is the case for the PsA study we consider, and many other examples involving nonterminal event as outcome. The approach here can also be applied to the case where visits stop after the occurrence of failure. Pseudo-visit times can be imputed based on the assumed visit time model, for example, (A2), or we can discretize the time scale with a fine grid and assume visits occur at designated time points following failure. Event status, $Y_i(t_{ij})$, at such pseudo-visit times is obviously known, and are still included in estimating equations such as (9) and (10) to estimate model components. More details about this can be found in Zhu et al. (2017), where simulation studies for both terminal and nonterminal failure time events are conducted.

In addition, we use isotonic regression to correct the kernel-smoothed estimator of nonparametric component and obtain a monotone estimator, for example, $\hat{S}_0^{\dagger}(t;\hat{\boldsymbol{\beta}})$ or $\hat{h}^{\dagger}(t;\hat{\boldsymbol{\beta}})$. Our reviewers suggest that we could employ an isotonic regression intermediately between (13) and (14) to monotonize the nonparametric estimator. We tried this in the simulation for the Cox model in Section 3.2 and found that results are very close to those we presented in Table 3. In fact, we plot the kernel estimator of baseline survivor and the kernel-isotonic baseline survivor estimator for additive hazards model in Figure 1. It is seen that even for $n = 100$ the kernel estimate is fairly close to the kernel-isotonic estimate, and they further merge as sample size increases. To incorporate monotone constraints into the profile likelihood framework is an interesting topic, but the theoretical derivation remains an open question and we consider it as future work.

We compared our fitted models with the nonparametric estimator proposed in Zhu et al. (2018) to evaluate model fitting for each HLA-B27 group separately in Section 4. Regarding general strategies for model selection and evaluation, when there is a single categorical covariate A, comparison with nonparametric estimates are viable options. Another possibility is to assess each model's predictive performance in terms of calibration or discrimination. For example, Wu and Cook (2020) developed an inverse probability weighting approach to compute the area under a receiver operating characteristic curve (ROC) for general interval-censored data. However, like most other existing methods, they assumed that censoring is independent, so further extension is required to apply to the case we discuss here.

Finally, we reiterate our remark in Section 1 that the IIV weighting approach requires that visit times depend only on past observed history of visits and disease-related factors. Violations of this assumption can occur; for example, a change in a disease-related factor since the last visit may influence when the next visit occurs. This cannot be addressed without supplementary data being collected about the reasons for visits, or on outcome-related variables between visits, though sensitivity analysis can be conducted by making unverifiable assumptions. In recent work, Cook and Lawless (2019) discuss situations where the conditions on visits made here may be violated, and where visits that are related to recent disease history (termed "disease-driven visits") can be identified. We mention also that long periods where no visits occur

before an administrative end-of-follow-up time are sometimes seen; this may be due to a person becoming lost-to-follow-up earlier, with no date for this being recorded. There is no reliable method for dealing with such long visit gap times unless auxiliary data that help to determine whether there has been loss-to-follow-up can be obtained. It is recommended that longitudinal studies be conducted so as to minimize such situations; this can be done by having subjects adhere as closely as possible to scheduled visit times, so that premature losses to follow-up become known. It is also helpful if information concerning events between visits, or the reasons for a visit, can be collected.

ACKNOWLEDGMENTS

The authors would like to thank Drs. Dafna Gladman and Vinod Chandran from the Toronto Western Hospital for sharing the PsA data. The authors also thank Dr. Richard J. Cook from the University of Waterloo for valuable advice on data and clinical issues. This research was supported in part by The Natural Sciences and Engineering Research Council of Canada Grant RGPIN-2017-04055 to JF Lawless, and Grants RGPIN-2020-05803 and DGECR-2020-00354 to Y Zhu.

ORCID

Yayuan Zhu https://orcid.org/0000-0003-1349-3038
Ziqi Chen https://orcid.org/0000-0002-4128-2986
Jerald F. Lawless https://orcid.org/0000-0002-3192-0470

REFERENCES

Aalen, O. (1980). *A model for nonparametric regression analysis of counting processes*. In *Mathematical statistics and probability theory. Lecture notes in statistics* (Vol. *2*, pp. 1–25). Springer.

Aalen, O., Borgan, O., & Gjessing, H. (2008). *Survival and event history analysis: A process point of view*. Springer Science & Business Media.

Anderson-Bergman, C. (2017). ICENREG: Regression models for interval censored data in R. *Journal of Statistical Software*, *81*(12), 1–23.

Azarang, L., Scheike, T., & de Uña-Álvarez, J. (2017). Direct modeling of regression effects for transition probabilities in the progressive illness–death model. *Statistics in Medicine*, *36*(12), 1964–1976.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions: The theory and application of isotonic regression*. Wiley.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, *30*(1), 89–99.

Buzkova, P. (2010). Panel count data regression with informative observation times. *The International Journal of Biostatistics*, *6*(1), 30.

Buzkova, P., & Lumley, T. (2007). Longitudinal data analysis for generalized linear models with follow-up dependent on outcome-related variables. *Canadian Journal of Statistics*, *35*(4), 485–500.

Buzkova, P., & Lumley, T. (2009). Semiparametric modeling of repeated measurements under outcome-dependent follow-up. *Statistics in Medicine*, *28*(6), 987–1003.

Chen, C.-M., Lu, T.-F. C., Chen, M.-H., & Hsu, C.-M. (2012). Semiparametric transformation models for current status data with informative censoring. *Biometrical Journal*, *54*(5), 641–656.

Chen, K., Jin, Z., & Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika*, *89*(3), 659–668.

Chen, Z., Tang, M.-L., & Gao, W. (2018). A profile likelihood approach for longitudinal data analysis. *Biometrics*, *74*(1), 220–228.

Cheng, S. C., Wei, L. J., & Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika*, *82*(4), 835–845.

Cook, R. J., & Lawless, J. F. (2007). *The statistical analysis of recurrent events*. Springer.

Cook, R. J., & Lawless, J. F. (2018). *Multistate models for the analysis of life history data*. Chapman & Hall/CRC Press.

Cook, R. J., & Lawless, J. F. (2019). Independence conditions and the analysis of life history studies with intermittent observation. *Biostatistics*.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika*, *62*(2), 269–276.

Dabrowska, D. M., & Doksum, K. A. (1988a). Estimation and testing in a two-sample generalized odds-rate model. *Journal of the American Statistical Association*, *83*(403), 744–749.

Dabrowska, D. M., & Doksum, K. A. (1988b). Partial likelihood in transformation models with censored data. *Scandinavian Journal of Statistics*, *15*(1), 1–23.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, *21*(1), 196–216.

Fan, J., Huang, T., & Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of the American Statistical Association*, *102*(478), 632–641.

Fine, J. P. (1999). Analysing competing risks data with transformation models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*(4), 817–830.

Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, *42*(4), 845–854.

Gladman, D. D., & Chandran, V. (2010). Observational cohort studies: Lessons learnt from the university of toronto psoriatic arthritis program. *Rheumatology*, *50*(1), 25–31.

Goetghebeur, E., & Ryan, L. (2000). Semiparametric regression analysis of interval-censored data. *Biometrics*, *56*(4), 1139–1144.

Harrell, F. E., Jr. (2015). *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Hogan, J. W., Roy, J., & Korkontzelou, C. (2004). Handling drop-out in longitudinal studies. *Statistics in Medicine*, *23*(9), 1455–1497.

Jewell, N. P., & Van Der Laan, M. (2004). Case-control current status data. *Biometrika*, *91*(3), 29–541.

Fine, J. P., Ying, Z., & Wei, L. G. (1998). On the linear transformation model for censored data. *Biometrika*, *85*(4), 980–986.

Kalbfleisch, J. D. 1978. Censoring and the immutable likelihood. *Technical Report* (pp. 78–109).

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). John Wiley & Sons, Inc.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Lin, D. Y., Oakes, D., & Ying, Z. (1998). Additive hazards regression with current status data. *Biometrika*, *85*(2), 289–298.

Lin, H., Scharfstein, D. O., & Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 791–813.

Mack, Y.-p., & Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *61*(3), 405–415.

Mammen, E. (1991). Estimating a smooth monotone regression function. *The Annals of Statistics*, *19*(2), 724–740.

Martinussen, T., & Scheike, T. H. (2006). *Dynamic regression models for survival data*. Springer Science & Business Media.

Martinussen, T., & Vansteelandt, S. (2013). On collapsibility and confounding bias in Cox and Aalen regression models. *Lifetime Data Analysis*, *19*(3), 279–296.

Miles, R. E. (1959). The complete amalgamation into blocks, by weighted means, of a finite set of real numbers. *Biometrika*, *46*(3/4), 317–327.

Pullenayegum, E. M., & Feldman, B. M. (2013). Doubly robust estimation, optimally truncated inverse-intensity weighting and increment-based methods for the analysis of irregularly observed longitudinal data. *Statistics in Medicine*, *32*(6), 1054–1072.

Robins, J. M., & Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, *56*(3), 779–788.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, *90*(429), 106–121.

Scheike, T. H., & Zhang, M.-J. (2007). Direct modelling of regression effects for transition probabilities in multistate models. *Scandinavian Journal of Statistics*, *34*(1), 17–32.

Scheike, T. H., Zhang, M.-J., & Gerds, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, *95*(1), 205–220.

Shiboski, S. C. (1998). Generalized additive models for current status data. *Lifetime Data Analysis*, *4*(1), 29.

Stone, C. J. (1986). Generalized additive models: Comment. *Statistical Science*, *1*(3), 310–312.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *38*(3), 290–295.

van der Laan, M. J., & Robins, J. M. (1998). Locally efficient estimation with current status data and time-dependent covariates. *Journal of the American Statistical Association*, *93*(442), 693–701.

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.

Wang, S., Wang, C., Wang, P., & Sun, J. (2018). Semiparametric analysis of the additive hazards model with informatively interval-censored failure time data. *Computational Statistics & Data Analysis*, *125*, 1–9.

Wu, Y., & Cook, R. J. (2020). Assessing the accuracy of predictive models with interval-censored data. *Biostatistics*.

Zeng, D., Cai, J., & Shen, Y. (2006). Semiparametric additive risks model for interval-censored data. *Statistica Sinica*, *16*(1), 287–302.

Zeng, D., & Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, *93*(3), 627–640.

Zhang, Y., Hua, L., & Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the cox model with interval-censored data. *Scandinavian Journal of Statistics*, *37*(2), 338–354.

Zhang, Z., Sun, J., & Sun, L. (2005). Statistical analysis of current status data with informative observation times. *Statistics in Medicine*, *24*(9), 1399–1407.

Zhang, Z., Sun, L., Sun, J., & Finkelstein, D. M. (2007). Regression analysis of failure time data with informative interval censoring. *Statistics in Medicine*, *26*(12), 2533–2546.

Zhang, Z., Sun, L., Zhao, X., & Sun, J. (2005). Regression analysis of interval-censored failure time data with linear transformation models. *Canadian Journal of Statistics*, *33*(1), 61–70.

Zhao, S., Hu, T., Ma, L., Wang, P., & Sun, J. (2015). Regression analysis of interval-censored failure time data with the additive hazards model in the presence of informative censoring. *Statistics and Its Interface*, *8*(3), 367–377.

Zhu, Y., Lawless, J. F., & Cotton, C. A. (2017). Estimation of parametric failure time distributions based on interval-censored data with irregular dependent follow-up. *Statistics in Medicine*, *36*(10), 1548–1567.

Zhu, Y., Lawless, J. F., & Cotton, C. A. (2018). Nonparametric analysis of dependently interval-censored failure time data. *Statistics in Medicine*, *37*(21), 3091–3105.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhu Y, Chen Z, Lawless JF. Semiparametric analysis of interval-censored failure time data with outcome-dependent observation schemes. *Scand J Statist*. 2022;49:236–264. https://doi.org/10.1111/sjos.12511

APPENDIX

A.1 Discussion about the estimation of IIV weight

To model the visit time process and estimate the IIV weight defined in (5), recurrent event models can be considered. For example, Lin et al. (2004) and later authors used a multiplicative intensity model based on Markov assumption

$$\lambda_N\{t|\boldsymbol{Z_i}(t^-)\} = \lambda_{N0}(t)\exp\{\boldsymbol{\alpha}^T\boldsymbol{Z_i}(t^-)\}, \tag{A1}$$

where $\lambda_{N0}(t)$ is an unspecified baseline visit intensity. Parameters $\boldsymbol{\alpha}$ can be estimated by partial likelihood for Cox models (Cox, 1975) and the nonparametric baseline intensity can be estimated by the Breslow estimator (Breslow, 1974). Stabilized weights can be adopted to mitigate potential adverse effects caused by the variability of weight estimates. In that case, the IIV weight $w_i(t)$ in (5) is replaced by $s_i(t)/\lambda_N\{t|\boldsymbol{Z_i}(t^-)\}$, where $s_i(t)$ is a marginal or partly conditional visit intensity given only time-fixed covariates in the failure time model (Lin et al., 2004; Robins & Finkelstein, 2000).

However, in many applications the visit intensity depends much more strongly on the elapsed time since the most recent visit than on the process age t. Then, modulated renewal processes (Cook & Lawless, 2007) defined for the gap times between consecutive visits are more plausible in this sense. For example Zhu et al. (2017, 2018) considered a proportional hazards modulated renewal process

$$\lambda_N\{t|\boldsymbol{Z_i}(t^-);\boldsymbol{\alpha}\} = \lambda_{N0}\{G_i(t)\}\exp\{\boldsymbol{\alpha}^T\boldsymbol{Z_i}(t^-)\}, \tag{A2}$$

where $G_i(t) = t - t_{iN_i(t^-)}$ is the elapsed time since the most recent visit prior to t. We remark that process age terms (e.g., t) can be included in $\boldsymbol{Z}_i(t^-)$ as predictors if desired.

Another modeling option is to use flexible parametric baseline intensities $\lambda_{N0}\{G_i(t);\boldsymbol{\rho}\}$ for the visit process. This results in estimates that automatically satisfy the positivity condition, and also reduces the variability of estimated weights. A convenient choice is piecewise-constant intensities

$$\lambda_N\{t|\boldsymbol{Z_i}(t^-);\boldsymbol{\alpha}\} = \sum_{j=1}^{J}\rho_j I_j\{G_i(t)\}\exp\{\boldsymbol{\gamma}^T\boldsymbol{Z_i}(t^-)\}, \tag{A3}$$

where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_J)^T$ are unknown nonnegative constants, $I_j(x) = I\{x \in (d_{j-1}, d_j]\}$, and $0 = d_0 < d_1 < \ldots < d_J = \tau$ are the corresponding cut-points. Parameters $\boldsymbol{\alpha} = (\boldsymbol{\rho}^T, \boldsymbol{\gamma}^T)^T$ can be estimated by solving the likelihood score function (Cook and Lawless 2007, ch. 5))

$$\boldsymbol{U}_n^{\dagger}(\boldsymbol{\alpha}) = \sum_{i=1}^{n}\left\{\sum_{j=1}^{m_i}\frac{\partial\log\lambda_i\{G_i(t_{ij})\}}{\partial\boldsymbol{\alpha}} - \sum_{j=1}^{m_i+1}\int_0^{G_i(t_{ij})}\frac{\partial\lambda_i(s)}{\partial\boldsymbol{\alpha}}ds\right\}, \tag{A4}$$

where we write $\lambda_i\{G_i(t)\} := \lambda_N\{t|\boldsymbol{Z_i}(t^-);\boldsymbol{\alpha}\}$ in (A3) with some abuse of notation and let $t_{im_i+1} = C_i$. We adopt this approach for the estimation of visit intensity through this article.

A.2 Methodological results developed for Model (2) with time-varying coefficients

This section extends the methodology and theoretical results developed in Section 2.3 based on the additive hazards model (2) to time-varying coefficients defined by a set of specified functions. We let $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_q(t))^T$ and assume each coefficient $\beta_\ell(t)$, $\ell = 1, \ldots, q$, could be a constant

β_ℓ, a specified function of process age t such as $\beta_\ell(t) = \beta_\ell e^{0.01t}$, or is represented by smooth basis functions as we did in the analysis of the PsA data in Section 4.

We define $\beta_\ell(t;\boldsymbol{\beta}_\ell) = \boldsymbol{B}_d^T(t)\boldsymbol{\beta}_\ell$ where $\boldsymbol{B}_d(t) = (B_1(t), \ldots, B_d(t))^T$ is a d-dimensional vector of functions of t defined on $[0,\tau]$ (e.g., spline basis functions with d degrees of freedom) and $\boldsymbol{\beta}_\ell$ is a $d\times 1$ vector of unknown coefficients corresponding to the ℓth covariate A_ℓ. Then, the survivor function is given by

$$S(t|A_i) = S_0(t)\exp\{-\boldsymbol{A}_i^T\boldsymbol{\mathcal{B}}(t;\boldsymbol{\beta})\},$$

where $\boldsymbol{\mathcal{B}}(t;\boldsymbol{\beta}) = (\int_0^t \beta_1(u;\boldsymbol{\beta}_1)du, \ldots, \int_0^t \beta_q(u;\boldsymbol{\beta}_q)du)^T$. It can be written as

$$\boldsymbol{\mathcal{B}}(t;\boldsymbol{\beta}) = [\boldsymbol{I}_{q\times q}\otimes \overline{\boldsymbol{B}}_d^T(t)]\boldsymbol{\beta} =: \mathbb{B}_d(t)\boldsymbol{\beta}, \tag{A5}$$

where $\overline{\boldsymbol{B}}_d^T(t) = (\int_0^t B_1(u)du, \ldots, \int_0^t B_d(u)du)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_q^T)^T$, $\otimes$ denotes the Kronecker product and $\boldsymbol{I}_{q\times q}$ is a $q\times q$ identity matrix.

Estimating equations (9) and (10) in Section 2.3 can be accordingly modified as

$$\hat{S}_0(t;\boldsymbol{\beta}) = \frac{\sum_{i=1}^n\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})Y_i(t_{ij})}{\sum_{i=1}^n\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}\}}, \tag{A6}$$

and

$$\Psi_n\{\boldsymbol{\beta}\} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^{m_i}\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i\left[Y_i(t_{ij}) - \hat{S}_0(t_{ij};\boldsymbol{\beta})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}\}\right]}{\left[1-\hat{S}_{0;}(t_{ij};\boldsymbol{\beta})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}\}\right]} = \boldsymbol{0}. \tag{A7}$$

Then, the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ in Theorem 1 can be slightly modified under similar conditions with

$$V = E\left[\sum_{j=1}^{m_i}\left\{\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i}{1-\mu_{ij}} - w_i(t_{ij})Q(t_{ij})\right\}\{Y_i(t_{ij})-\mu_{ij}\}\right]^{\otimes 2},$$

$$D = E\left[\sum_{j=1}^{m_i}\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i\left\{\overline{S}_0^T(t_{ij};\boldsymbol{\beta}_0)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\} - \mu_{ij}\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\right\}}{1-\mu_{ij}}\right],$$

and

$$\mu_{ij} = S_0(t_{ij};\boldsymbol{\beta}_0)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\},$$

$$H(t_{ij}) = E\left[\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right],$$

$$Q(t_{ij}) = E\left[\frac{\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}{1-\mu_{ij}}\right]H^{-1}(t_{ij}),$$

$$\overline{S}_0^T(t_{ij};\boldsymbol{\beta}_0) = S_0(t_{ij})E\left[\boldsymbol{A}_i^T\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]\mathbb{B}_d(t_{ij})H^{-1}(t_{ij}).$$

In the subsequent sections, we discuss the proofs of Theorem 1 under the flexible parametrization

$$S(t|A_i) = S_0(t)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\boldsymbol{\beta}\}.$$

For time-fixed coefficients case (6), similar derivations can be done with $\mathbb{B}_d(t)$ substituted by t.

A.3 Asymptotic consistency

We first discuss the consistency of $\hat{S}_0(t;\boldsymbol{\beta}_0)$ defined in (A6), where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$. Then, for any $t \in [0,\tau]$

$$\begin{aligned}
\hat{S}_0(t;\boldsymbol{\beta}_0) - S_0(t) &= \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \left[Y_i(t_{ij}) - S_0(t) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\} \right]}{\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}} \\
&= \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \left[Y_i(t_{ij}) - S_0(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\} \right]}{\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}} \\
&\quad + \frac{\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \left[\{S_0(t_{ij}) - S_0(t)\} \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\} \right]}{\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}} \\
&=: (A_{1n} + A_{2n}),
\end{aligned} \tag{A8}$$

where A_{1n} is the asymptotical variance term; A_{1n} is of mean zero and converges to zero uniformly with rate $\sqrt{\log(n)/nh}$ (Chen et al., 2018; Fan et al., 2007; Mack & Silverman, 1982).

The asymptotical bias stems from A_{2n}. Specifically, by the literature of kernel regression and Taylor expansion,

$$\begin{aligned}
A_{2n} &= \frac{S_0'(t) \sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}(t_{ij}-t)}{\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}} \\
&\quad + \frac{0.5 S_0''(t) \sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}(t_{ij}-t)^2}{\sum_{i=1}^n \sum_{j=1}^{m_i} K_h(t-t_{ij}) w_i(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}} + o_p(h^2) \\
&=: \frac{S_0'(t) H_{1n}(t)}{H_n(t)} + \frac{0.5 S_0''(t) H_{2n}(t)}{H_n(t)} + o_p(h^2) \\
&\to -S_0'(t) h^2 \left(\int K(u) u^2 du \right) E\left[\boldsymbol{A}_i^T \frac{\partial \mathbb{B}_d(t)}{\partial t} \boldsymbol{\beta}_0 \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t)\boldsymbol{\beta}_0\} \right] \Big/ E\left[\exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t)\boldsymbol{\beta}_0\} \right] \\
&\quad + 0.5 S_0''(t) h^2 \left(\int K(u) u^2 du \right) + o_p(h^2) \\
&=: B_0(t) h^2 + o_p(h^2).
\end{aligned} \tag{A9}$$

The limiting values of $H_n(t)$, $H_{1n}(t)$, and $H_{2n}(t)$ are provided in part (a) of Web-Appendix S.1.

Because $\Psi_n(\boldsymbol{\beta}_0)$ defined in (A7) is a sample average over n and $\hat{S}_0(t;\boldsymbol{\beta}_0) \xrightarrow{P} S_0(t)$ uniformly as $n \to \infty$ as discussed above, by the law of large numbers, we know that as $n \to \infty$, $\Psi_n(\boldsymbol{\beta}_0)$ converges to

$$\begin{aligned}
&\int_0^\tau E\left\{ \frac{w_i(u) \mathbb{B}_d^T(u) \boldsymbol{A}_i \left[Y_i(u) - S_0(u) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(u)\boldsymbol{\beta}_0\} \right]}{\left[1 - S_0(u) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(u)\boldsymbol{\beta}_0\} \right]} dN_i^*(u) \right\} \\
&= \int_0^\tau E\left\{ \frac{w_i(u) \mathbb{B}_d^T(u) \boldsymbol{A}_i \left[Y_i(u) - S_0(u) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(u)\boldsymbol{\beta}_0\} \right]}{\left[1 - S_0(u) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(u)\boldsymbol{\beta}_0\} \right]} E\{dN_i^*(u) | \boldsymbol{H_i}(u^-), Y_i(u)\} \right\}.
\end{aligned}$$

By assumption (4), the above equation equals

$$\begin{aligned}
&\int_0^\tau E\left\{\frac{w_i(u)\mathbb{B}_d^T(u)\boldsymbol{A}_i\left[Y_i(u)-S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}\right]}{\left[1-S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}\right]}C_i(u)\lambda_N\{u|\boldsymbol{Z_i}(u^-)\}du\right\}\\
&=\int_0^\tau E\left\{\frac{w_i(u)\mathbb{B}_d^T(u)\boldsymbol{A}_i\left[Y_i(u)-S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}\right]}{\left[1-S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}\right]}C_i(u)\frac{1}{w_i(u)}du\right\}\\
&=\int_0^\tau E\left\{\frac{\mathbb{B}_d^T(u)\boldsymbol{A}_i\left[E\{Y_i(u)|\boldsymbol{A}_i,C_i(u)\}-S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}\right]}{\left[1-S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}\right]}C_i(u)du\right\}\\
&=\int_0^\tau E\left\{\frac{\mathbb{B}_d^T(u)\boldsymbol{A}_i\left[S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}-S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}\right]}{\left[1-S_0(u)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(u)\boldsymbol{\beta}_0\}\right]}C_i(u)du\right\}\\
&=\mathbf{0}.
\end{aligned}$$

Here, we assume the random drop-out time process $C_i(t)$ is independent of the outcome process $Y_i(t)$ and auxiliary prognostic processes given $\boldsymbol{A}_i$ for simplicity. However, if C_i is observable or can be approximated, this can be relaxed by involving an inverse-probability-of-censoring weight based on an additional sequential ignorability assumption for drop-out times. In the following, for the simplicity of derivation, we consider the case without random censoring, that is, $dN_i^*(t) = dN_i(t)$ and $C_i(t) = 1$, for all i and all $t \geq 0$.

It has been shown that $\Psi_n(\boldsymbol{\beta}_0) \xrightarrow{P} \mathbf{0}$. Because $\hat{\boldsymbol{\beta}}$ is the solution of $\Psi_n(\boldsymbol{\beta}) = \mathbf{0}$, then we have $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}_0$, as $n \to \infty$.

A.4 Asymptotic distribution and variance estimation of $\boldsymbol{\beta}$

We expand $\Psi_n(\hat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}_0$ by Taylor series, and then get

$$\mathbf{0} = \Psi_n(\boldsymbol{\beta}_0) + \Psi_n'(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (1/2)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\Psi_n''(\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T,$$

where $\tilde{\boldsymbol{\beta}}$ is a point between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$; $\Psi_n'(\boldsymbol{\beta})$ and $\Psi_n''(\boldsymbol{\beta})$ respectively denote the first and second derivative of Ψ_n w.r.t. $\boldsymbol{\beta}$, and we assume that they both have finite expectations for any $\boldsymbol{\beta}$. Then, $\Psi_n'(\boldsymbol{\beta}) = O_p(1)$ and $\Psi_n''(\boldsymbol{\beta}) = O_p(1)$ for all $\boldsymbol{\beta}$ by the law of large numbers, because they are both written as sample averages. Then,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \left\{-\Psi_n'(\boldsymbol{\beta}_0)\right\}^{-1}\left\{\sqrt{n}\Psi_n(\boldsymbol{\beta}_0)\right\} + o_p(1). \tag{A10}$$

First, we write $\sqrt{n}\Psi_n\{\boldsymbol{\beta}_0\}$ as

$$\begin{aligned}
\sqrt{n}\Psi_n\{\boldsymbol{\beta}_0\} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^{m_i}\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i\left[Y_i(t_{ij})-\hat{S}_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]}{1-\hat{S}_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^{m_i}\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i\left[Y_i(t_{ij})-S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]}{1-S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}\\
&\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^{m_i}\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i\left\{\hat{S}_0(t_{ij})-S_0(t_{ij})\right\}\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}{1-S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}+o_p(1)
\end{aligned}$$

$$=: C_{1n} - C_{2n} + o_p(1).$$

Define

$$b(\boldsymbol{\beta}_0) = E\left\{ \sum_{j=1}^{m_i} \frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i B_0(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}{1 - S_0(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}} \right\}, \tag{A11}$$

where $B_0(t_{ij})$ is defined in (A9) for $t = t_{ij}$. By (A8), we have

$$C_{2n} - \sqrt{n} b(\boldsymbol{\beta}_0) h^2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}{nH_n(t_{ij}) \left[1 - S_0(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]}$$
$$\times \sum_{i_1=1}^{n} \sum_{j_1=1}^{m_{i_1}} K_h(t_{ij} - t_{i_1 j_1}) w_i(t_{i_1 j_1}) \left\{ Y_i(t_{i_1 j_1}) - \mu_{i_1 j_1} \right\} + o_p(1),$$

where

$$H_n(t_{ij}) = \frac{1}{n} \sum_{i_1=1}^{n} \sum_{j_1=1}^{m_{i_1}} K_h(t_{ij} - t_{i_1 j_1}) w_i(t_{i_1 j_1}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{i_1 j_1})\boldsymbol{\beta}_0\},$$
$$\mu_{i_1 j_1} = S_0(t_{i_1 j_1}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{i_1 j_1})\boldsymbol{\beta}_0\}.$$

Switching the order of summations, we then get

$$C_{2n} - \sqrt{n} b(\boldsymbol{\beta}_0) h^2 = \frac{1}{\sqrt{n}} \sum_{i_1=1}^{n} \sum_{j_1=1}^{m_{i_1}} w_i(t_{i_1 j_1}) Q_n(t_{i_1 j_1}) \left\{ Y_i(t_{i_1 j_1}) - \mu_{i_1 j_1} \right\} + o_p(1),$$

where

$$Q_n(t_{i_1 j_1}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{K_h(t_{ij} - t_{i_1 j_1}) w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}{H_n(t_{ij}) \left[1 - S_0(t_{ij}) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]}.$$

Thus,

$$\sqrt{n}\Psi_n(\boldsymbol{\beta}_0) + \sqrt{n} b(\boldsymbol{\beta}_0) h^2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left\{ \frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i}{1 - \mu_{ij}} - w_i(t_{ij}) Q_n(t_{ij}) \right\} \{ Y_i(t_{ij}) - \mu_{ij} \} + o_p(1).$$

Second, $-\Psi_n'(\boldsymbol{\beta}_0)$ in (A10) can be written as

$$-\Psi_n'(\boldsymbol{\beta}_0) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\} \left[\partial \hat{S}_0(t_{ij}; \boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}_0^T - \hat{S}_0(t_{ij}; \boldsymbol{\beta}_0)\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\right]}{1 - \hat{S}_0(t_{ij}; \boldsymbol{\beta}_0) \exp\{-\boldsymbol{A}_i^T \mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}$$
$$+ o_p(1).$$

Applying the law of large numbers, as $n \to \infty$, we have $-\Psi_n'(\boldsymbol{\beta}_0)$ converge to

$$D = E\left\{\sum_{j=1}^{m_i} \frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i\left[\overline{S}_0^T(t_{ij};\boldsymbol{\beta}_0)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\} - S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\right]}{1 - S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}\right\},$$

where $\overline{S}_0^T(t;\boldsymbol{\beta}_0)$ is given in the part (b) of Web-Appendix S.1.

Based on (A10) and following the central limit theorem and Slutsky's theorem, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 + D^{-1}b(\boldsymbol{\beta}_0)h^2) \xrightarrow{D} N(0, D^{-1}VD^{-1}),$$

where we assume that D is nonsingular and

$$V = E\left[\sum_{j=1}^{m_i}\left\{\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i}{1-\mu_{ij}} - w_i(t_{ij})Q(t_{ij})\right\}\{Y_i(t_{ij}) - \mu_{ij}\}\right]^{\otimes 2},$$

where $Q(t)$ is the limit value of $Q_n(t)$ which is provided in the part (a) of Web-Appendix S.1. Assuming $nh^4 \to 0$ when $n \to \infty$, we have the asymptotic distribution as stated in Theorem 1.

Now, we discuss how to estimate the variance of $\hat{\boldsymbol{\beta}}$. Components in $\widehat{Var}(\hat{\boldsymbol{\beta}}) = (1/n)\hat{D}_n^{-1}\hat{V}_n\hat{D}_n^{-1}$ are estimated by

$$\hat{D}_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i\left[\hat{\overline{S}}_0^T(t_{ij};\hat{\boldsymbol{\beta}})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\hat{\boldsymbol{\beta}}\} - \hat{\mu}_{ij}\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\right]}{1-\hat{\mu}_{ij}},$$

$$\hat{V}_n = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{m_i}\left\{\frac{w_i(t_{ij})\mathbb{B}_d^T(t_{ij})\boldsymbol{A}_i}{1-\hat{\mu}_{ij}} - w_i(t_{ij})\hat{Q}_n(t_{ij})\right\}\{Y_i(t_{ij}) - \hat{\mu}_{ij}\}\right]^{\otimes 2},$$

where

$$\hat{Q}_n(t) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{B}_d^T(t)\boldsymbol{A}_i\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\hat{\boldsymbol{\beta}}\}}{1-\hat{\mu}(t)}\hat{H}_n^{-1}(t),$$

$$\hat{H}_n(t) = \frac{1}{n}\sum_{i=1}^{n}\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\hat{\boldsymbol{\beta}}\},$$

$$\hat{\overline{S}}_0^T(t;\hat{\boldsymbol{\beta}}) = \frac{\hat{S}_0(t;\hat{\boldsymbol{\beta}})\sum_{i=1}^{n}\left[\boldsymbol{A}_i^T\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\hat{\boldsymbol{\beta}}\}\right]\mathbb{B}_d(t)}{\sum_{i=1}^{n}\left[\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\hat{\boldsymbol{\beta}}\}\right]},$$

$$\hat{\mu}(t) = \hat{S}_0(t;\hat{\boldsymbol{\beta}})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\hat{\boldsymbol{\beta}}\} \quad \text{and} \quad \hat{\mu}_{ij} = \hat{\mu}(t_{ij}).$$

A.5 Asymptotic distribution of $\hat{S}_0(t;\hat{\boldsymbol{\beta}})$

Similar to (A8) and (A9) in Appendix A.3, given $\hat{\boldsymbol{\beta}}$ is the solution of $\Psi_n(\boldsymbol{\beta}) = \mathbf{0}$ from (A7), that is, $\Psi_n(\hat{\boldsymbol{\beta}}) = \mathbf{0}$, it can be shown that

$$\hat{S}_0(t;\hat{\boldsymbol{\beta}}) - S_0(t) = (A_{1n} + A_{2n})(1 + o_p(1)),$$

and

$$A_{2n} = B_0(t)h^2 + o_p(h^2),$$

is the asymptotic bias term. We then discuss the asymptotic normality of A_{1n}:

$$A_{1n} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\left[Y_i(t_{ij}) - S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}}.$$

Recall that

$$H_n(t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\},$$

and it is shown in part (a) in Web-Appendix S.1 that, as $n\to\infty$, we have $H_n(t)$ converges to

$$H(t) = E\left[\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\boldsymbol{\beta}_0\}\right].$$

We now investigate the asymptotic normality of

$$J_n(t) := \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\left[Y_i(t_{ij}) - S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right].$$

It is easily seen that $E\{J_n(t)\}=0$. Thus, we next calculate $\mathrm{Var}\{J_n(t)\}$.

$$\begin{aligned}
\mathrm{Var}\{J_n(t)\} &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left\{\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\left[Y_i(t_{ij}) - S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]\right\}\\
&= \frac{1}{n^2}\sum_{i=1}^{n}E\left\{\sum_{j=1}^{m_i} K_h(t-t_{ij})w_i(t_{ij})\left[Y_i(t_{ij}) - S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]\right\}^2\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}E\left\{K_h(t-t_{ij})w_i(t_{ij})\left[Y_i(t_{ij}) - S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]\right\}^2(1+o(1))\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}E\left\{K_h^2(t-t_{ij})w_i^2(t_{ij})\left[Y_i(t_{ij}) - S_0(t_{ij})\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t_{ij})\boldsymbol{\beta}_0\}\right]^2\right\}(1+o(1))\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\int_0^\tau K_h^2(t-s)E\left\{w_i^2(s)[Y_i(s) - S_0(s)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(s)\boldsymbol{\beta}_0\}]^2 dN_i(s)\right\}(1+o_p(1))\\
&= \frac{1}{n}\int_0^\tau K_h^2(t-s)E\left\{w_i(s)[Y_i(s) - S_0(s)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(s)\boldsymbol{\beta}_0\}]^2\right\}ds(1+o_p(1))\\
&= \frac{1}{nh_n}\int K^2(u)E\left\{w_i(t-uh)[Y_i(t-uh) - S_0(t-uh)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t-uh)\boldsymbol{\beta}_0\}]^2\right\}du\\
&\quad (1+o_p(1))\\
&= \frac{1}{nh_n}\left(\int K^2(u)\left[E\left\{w_i(t)[Y_i(t) - S_0(t)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\boldsymbol{\beta}_0\}]^2\right\} + O(h)u\right]du\right)
\end{aligned}$$

$$
\begin{aligned}
&(1+o_p(1))\\
&=\frac{1}{nh_n}\left(E\left\{\frac{[Y_i(t)-S_0(t)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\boldsymbol{\beta}_0\}]^2}{\lambda_N\{t|\boldsymbol{Z_i}(t^-)\}}\right\}\int K^2(u)du+O(h)\int K^2(u)udu\right)\\
&(1+o_p(1))\\
&=:\frac{P(t)}{nh_n}(1+o_p(t)),
\end{aligned}
$$

where $P(t):=E\left\{\frac{[Y_i(t)-S_0(t)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\boldsymbol{\beta}_0\}]^2}{\lambda_N\{t|\boldsymbol{Z_i}(t^-)\}}\right\}\left\{\int K^2(u)du\right\}$. Note that $\int K^2(u)udu=0$ and it is assumed in condition 3 that $\int K^2(u)du<\infty$.

It follows from Lyapunov central limit theorem that

$$\sqrt{nh_n}J_n(t)\xrightarrow{D}N(0,P(t)),$$

provided that $\sup_{t\in[0,\tau]}E\left|Y_i(t)-S_0(t)\exp\{-\boldsymbol{A}_i^T\mathbb{B}_d(t)\boldsymbol{\beta}_0\}\right|^3<\infty$ and other conditions stated in Theorem 1 are satisfied. Then, by Slutsky's theorem we have the asymptotic normality of $\hat{S}_0(t)$ as

$$\sqrt{nh_n}\{\hat{S}_0(t)-S_0(t)\}\xrightarrow{D}N(0,H(t)^{-2}P(t)),\quad \text{as } n\to\infty.$$