Semiparametric estimation of structural nested mean models with irregularly spaced longitudinal observations

Shu Yang

Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA

Correspondence
Shu Yang, Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC, USA,
Email: syang24@ncsu.edu

Funding information
National Institute of Environmental Health Sciences, Grant/Award Number: 1R01ES031651; National Cancer Institute, Grant/Award Number: P01 CA142538; Division of Mathematical Sciences, Grant/Award Number: 1811245; National Institute on Aging, Grant/Award Number: 1R01AG066883

Abstract
Structural nested mean models (SNMMs) are useful for causal inference of treatment effects in longitudinal observational studies. Most existing works assume that the data are collected at prefixed time points for all subjects, which, however, may be restrictive in practice. To deal with irregularly spaced observations, we assume a class of continuous-time SNMMs and a martingale condition of no measured confounding (NUC) to identify the causal parameters. We develop the semiparametric efficiency theory and locally efficient estimators for continuous-time SNMMs. This task is nontrivial due to the restrictions from the NUC assumption imposed on the SNMM parameter. In the presence of ignorable censoring, we show that the complete-case estimator is optimal among a class of weighting estimators including the inverse probability of censoring weighting estimator, and it achieves a double robustness feature in that it is consistent if at least one of the models for the potential outcome mean function and the treatment process is correctly specified. The new framework allows us to conduct causal analysis respecting the underlying continuous-time nature of data processes. The simulation study shows that the proposed estimator outperforms existing approaches. We estimate the effect of time to initiate highly active antiretroviral therapy on the CD4 count at year 2 from the observational Acute Infection and Early Disease Research Program database.

Keywords
causality, counting process, discretization, g-estimator, martingale

1 INTRODUCTION

The gold standard to draw causal inference of treatment effects is designing randomized experiments. However, randomized experiments are not always feasible due to practical constraints or ethical issues. In these cases, observational studies are useful. In observational studies, confounding presents a unique challenge to drawing valid causal inferences of treatment effects. For example, sicker patients are more likely to take the active treatment, whereas healthier patients are more likely to take the control treatment. Consequently, it is not fair to compare the outcome from the treated group and the control group directly. Moreover, in longitudinal observational studies, confounding is likely to be time-dependent, in the sense that time-varying prognostic factors of the outcome affect the treatment assignment at each time, and thereby distort the association between treatment and outcome over time.
In these cases, traditional regression methods are biased even adjusting for the time-varying confounders (Robins and Hernán, 2009).

Structural nested models (SNMs; Robins, 1994) have been proposed to overcome the challenges for causal inference with time-varying confounding. We focus on a class of SNMs for continuous outcomes, namely, structural nested mean models (SNMMs). Most existing works on SNMMs assume discrete-time data-generating processes and require all subjects to be followed at the same prefixed time points, such as months. The literature of discrete-time SNMMs is fruitful; see, for example, Robins (1994), Lok and DeGruttola (2012), and Yang and Lok (2016; 2018). However, observational data are often collected by user-initiated visits to clinics, hospitals, and pharmacies, and data are more likely to be measured at irregularly spaced time points, which are not necessarily the same for all subjects. Such data sources are now commonplace, such as electronic health records, claims databases, disease data registries, and so on. The existing causal framework does not directly apply in such situations, requiring some (possibly arbitrary) discretization of the timeline (Neugebauer et al., 2010). Data preprocessing is quite standard and routine to practitioners, but leads to many unresolved problems: the treatment process depends transparently on the discretization, and therefore the interpretation of SNMMs depends on the definition of time interval (Robins, 1998). Moreover, after discretization, the data may need to be recreated at certain time points. Consider monthly data for example. If a subject had multiple visits within the same month, a common strategy is to take the average of the multiple measures as the observation for a given variable at that month. If a subject had no visit for a given month, one may need to impute the missing observation. Because of such distortions, the resulting data may not satisfy the standard causal consistency or no unmeasured confounding (NUC) assumptions. Consequently, model parameters may not have a causal interpretation.

Our interest is motivated by the observational AIEDRP (Acute Infection and Early Disease Research Program) study (Hecht et al., 2006). This study included a cohort of HIV-infected patients diagnosed during the acute or early stage of disease. Patients initiated highly active antiretroviral therapy (HAART) at various times of the follow-up. Among all AIEDRP patients, 64% of patients initiated the treatment before year 2 with the observed time to treatment initiation ranging continuously from 12 to 282 days. The hypothesis was that deferring therapy may have an increased risk of permanent immune system damage but also a decreased risk of developing drug resistance. Thus, the interest was to estimate the effect of time to initiate HAART on disease progression. However, the complex data also present novel challenges for statistical analysis as summarized below. By study protocol, follow-up visits were scheduled at weeks 2, 4, and 12, and then every 12 weeks thereafter, through week 96. During each follow-up visit, various variables can be measured such as CD4 count and viral load, for which lower CD4 count and higher viral load indicate worse immunological function and disease progression. Treatment initiation can then be determined by the discretion of physicians at follow-up visits. This raises a major concern of time-varying confounding that may obscure the causal effect of time to treatment initiation on disease progression; for example, patients with worse disease progression tend to initiate HAART earlier. Moreover, although the study protocol set visit times in advance, these fixed visit times were not adhered to perfectly in practice. Figure 1 shows the visit patterns from five random patients from the AIEDRP study. Importantly, both the number and the timings of visits differ from one patient to the next. Finally, 45% of patients dropped out of the study before year 2, resulting in right-censored data. It is arguable that censoring due to dropout may depend on the patient’s status, which necessitates proper adjustment of censoring.

With irregularly spaced observations, it is more reasonable to assume that the data are generated from continuous-time processes. The work for causal models in continuous-time processes is somewhat sparse; exceptions include, for example, Robins (1998), Lok (2008), Zhang et al. (2011), Lok (2017), and Yang et al. (2020). Extending the existing causal models with discrete-time processes to continuous-time processes is not trivial. An important challenge lies in time-dependent selection bias or confounding; for example, in a health-related study, sicker patients may visit the doctor more frequently and are more likely to initiate the treatment. To overcome this challenge, following Lok (2008), we treat the observed treatment assignment process as a counting process \( N_T(t) \) and assume a martingale condition of NUC on \( N_T(t) \) to identify the SNMM parameter. Specifically, the NUC assumption entails that the jumping rate of \( N_T(t) \) at \( t \) does not depend on future potential outcomes, given the past treatment and covariate history up to \( t \). A practical implication is that the covariate set should be rich enough to include all predictors of outcome and treatment so that we can distinguish the treatment effect and the confounding effect. This assumption was also adopted in Zhang et al. (2011), Yang et al. (2018), and Yang et al. (2020). Lok (2017) provided a strategy of constructing unbiased estimating equations exploiting the relationship between the potential outcome and treatment processes, which leads to a large class of estimators. While this strategy provides unbiased estimators, there is no guidance on how to construct an efficient estimator.
We establish the new semiparametric efficiency theory for continuous-time SNMMs with irregularly spaced observations. In our problem, the SNMM and NUC assumptions constitute the semiparametric model for the data. Although the NUC assumption does not have any testable implications on the observed-data likelihood, it imposes conditional independence restrictions on the counterfactual outcomes and treatment processes, given the past history, and hence restrictions for the SNMM parameter. To circumvent this complication, we use the variable transformation technique and translate the restrictions into the new variables, which lead to the unconstrained observed-data likelihood. This step allows us to characterize the semiparametric efficiency score (SES) for the SNMM parameter and construct locally efficient estimators that achieve the semiparametric efficiency bound. The estimator requires two nuisance models:

(a) the model of the potential outcome mean function conditional on time-varying covariates, and
(b) the model of the treatment process conditional on the history of the treatment and covariates.

The proposed estimator of the SNMM parameter is doubly robust in that it is consistent and asymptotically Normal if at least one of the models for the potential outcome mean function and the treatment process is correctly specified. In the AIEDRP database, a large portion of patients dropped out of the study before year 2. Under an ignorable censoring mechanism given the observed history, we show that the complete-case (CC) estimator, that is, the locally efficient estimator applied to the uncensored subjects, remains doubly robust and is optimal among a class of weighting estimators including the inverse probability of censoring weighting (IPCW) estimator (Rotnitzky et al., 2007).

2 | SNMMs IN DISCRETE-TIME PROCESSES

2.1 | Setup, models, and assumptions

We first describe the SNMM in discrete-time processes. We assume that $n$ subjects are followed at prefixed discrete times $t_0 < \cdots < t_{K+1}$ with $t_0 = 0$ and $t_{K+1} = \tau$. We assume that the subjects are simple random samples from a larger population. For simplicity, we suppress the subscript $i$ for subjects. Let $L_m$ be a vector of covariates at time $t_m$. Let $A_m$ be the treatment indicator at $t_m$; that is, $A_m = 1$ if the subject was on treatment at $t_m$ and $A_m = 0$ otherwise. We use the overline notation to denote a variable’s history; for example, $\overline{A}_m = (A_0, \ldots, A_m)$. We assume that once treatment is initiated, it is never discontinued, so each treatment regime corresponds to one treatment initiation time. Let $T$ be the time to treatment initiation, and let $T = \infty$ if the subject never initiated the treatment during the follow-up. Let $\Gamma$ be the indicator that the treatment initiation time is less than $\tau$; that is, $\Gamma = 1$ if the subject initiated the treatment before $\tau$ and $\Gamma = 0$ otherwise. Let $Y(\tau)$ be the potential outcome at the end of the study $\tau$, had the subject initiated the treatment at $t_m$, and let $Y(\infty)$ be the potential outcome at $\tau$ had the subject never initiated the treatment during the study follow-up. Let $V_m = (A_{m-1}, L_m)$ be the vector of treatment and covariate for $0 \leq m \leq K$, where $A_{-1}$ is defined as null. Let $Y$ be the continuous outcome measured at $\tau$. Finally, the subject’s full record is $F = (A_K, L_K, Y)$.
Following Robins (1994) and Lok and DeGruttola (2012), we describe the discrete-time SNMM for the treatment effect as follows.

**Assumption 1** (Discrete-time SNMM). For \( m \geq 0 \), the discrete-time SNMM for the effect of the treatment initiation time is

\[
\gamma_m(\mathbf{L}_m) = \mathbb{E}\left\{ Y^{(m)} - Y^{(\infty)} \mid \mathbf{A}_{m-1} = \mathbf{0}, \mathbf{L}_m \right\}
= \gamma_m(\mathbf{L}_m; \psi^*);
\]

that is, \( \gamma_m(\mathbf{L}_m; \psi) \) with \( \psi \in \mathbb{R}^p \) is a correctly specified model for \( \gamma_m(\mathbf{L}_m) \) with the true parameter value \( \psi^* \).

The general SNMMs in Robins (1994) specify the treatment effects for \( \mathbf{a}_K \) with general patterns. In particular, Robins (1994) focused on modeling \( \mathbb{E}\{Y(\mathbf{a}_m, \bar{0}) - Y(\bar{a}_m, \mathbf{0}) \mid \mathbf{A}_{m-1} = \mathbf{0}, \mathbf{L}_m \} \), where \( (\mathbf{a}_m, \bar{0}) \) is the treatment regime of following \( \mathbf{a}_m \) from \( t_0 \) to \( t_m \) and \( 0 \) onwards. This class of models describes one episode of treatment \( a_m \) on the shift of the outcome means at \( t_m \) given subject’s observed treatment and covariates history \( (\mathbf{A}_{m-1} = \mathbf{0}, \mathbf{L}_m) \). Model (1) is another class of SNMMs that specify the effects of treatment initiation times (Lok and DeGruttola, 2012). This model characterizes the conditional expectation of the treatment contrasts \( Y^{(m)} - Y^{(\infty)} \), given subject’s observed history \( (\mathbf{A}_{m-1} = \mathbf{0}, \mathbf{L}_m) \). Intuitively, it states that the conditional mean of the outcome is shifted by \( \gamma_m(\mathbf{L}_m; \psi^*) \) had the subject initiated the treatment at \( t_m \) comparing to never starting. Therefore, the parameter \( \psi^* \) has a causal interpretation. To help understand the model, consider \( \gamma_m(\mathbf{L}_m; \psi^*) = (\psi_1^* + \psi_2^* t_m)(\tau - t_m)^+ \), where \( \psi^* = (\psi_1^*, \psi_2^*) \) and \( c^+ = \max(c, 0) \) for a real number \( c \). This model entails that on average, the treatment would increase the mean of the outcome at \( \tau \) had the subject initiated the treatment at \( t_m \) by \( (\psi_1^* + \psi_2^* t_m)(\tau - t_m)^+ \), and the magnitude of the increase depends on the duration of the treatment and the treatment initiation time. If \( \psi_1^* + \psi_2^* t_m > 0 \) and \( \psi_2^* < 0 \), it indicates the treatment is beneficial and earlier initiation is better. Although we use the same notation \( \psi^* \) for the SNMM parameter as Robins (1994), it is important to keep in mind that interpretation of model parameters is tied to the class of SNMMs: Robins (1994) defines the effect of “blipping off” treatment at a single time point, whereas Model (1) defines the effect of removing treatment across all time points.

The following consistency assumption links the observed data to the potential outcomes.

**Assumption 2** (Consistency). The observed outcome is equal to the potential outcome under the actual treatment received; that is, \( Y = Y^{(\mathbf{T})} \).

If all potential outcomes were observed for each subject, we can directly compare these outcomes to infer the treatment effect; however, the fundamental problem in causal inference is that we cannot observe all potential outcomes for a specific subject (Holland, 1986). In particular, we can observe \( Y^{(\infty)} \) only for the subjects who did not initiate the treatment during the follow-up. To overcome this issue, define

\[
H(\psi^*) = Y - Y_T(\mathbf{L}_T; \psi^*).
\]

Intuitively, \( H(\psi^*) \) subtracts the treatment effect \( Y_T(\mathbf{L}_T; \psi^*) \) from the observed outcome \( Y \), so it mimics the potential outcome \( Y^{(\infty)} \) had the treatment never been initiated. We provide the formal statement as proved in Lok and DeGruttola (2012).

**Proposition 1** (Mimicking \( Y^{(\infty)} \)). Under Assumption 2, \( H(\psi^*) \) mimics \( Y^{(\infty)} \), in the sense that \( \mathbb{E}\{H(\psi^*) \mid \mathbf{A}_{m-1} = \mathbf{0}, \mathbf{A}_m, \mathbf{L}_m\} = \mathbb{E}\{Y^{(\infty)} \mid \mathbf{A}_{m-1} = \mathbf{0}, \mathbf{A}_m, \mathbf{L}_m\} \) for \( 0 \leq m \leq K \).

We cannot fit the SNMM by a regression model pooled over time, because the model involves the unobserved potential outcomes. Parameter identification requires the NUC assumption (Robins et al., 1992).

**Assumption 3** (No unmeasured confounding). \( \mathbf{A}_m \perp Y^{(\infty)} \mid (\mathbf{A}_{m-1}, \mathbf{L}_m) \) for \( 0 \leq m \leq K \).

Assumption 3 holds if \( (\mathbf{A}_{m-1}, \mathbf{L}_m) \) contains all prognostic factors for \( Y^{(\infty)} \) that affect the treatment decision at \( t_m \) for \( 0 \leq m \leq K \). Under this assumption, the observational study can be conceptualized as a sequentially randomized experiment. Proposition 1 implies that under Assumption 3, for \( 0 \leq m \leq K \),

\[
\mathbb{E}\left\{ H(\psi^*) \mid \mathbf{A}_{m-1} = \mathbf{0}, \mathbf{A}_m, \mathbf{L}_m \right\} = \mathbb{E}\left\{ H(\psi^*) \mid \mathbf{A}_{m-1} = \mathbf{0}, \mathbf{A}_m, \mathbf{L}_m \right\};
\]

see, for example, Robins et al. (1992). Equation (3) also poses restrictions for \( \psi^* \).

A stronger version of Assumption 3 is \( \mathbf{A}_m \perp Y^{(k)} \mid (\mathbf{A}_{m-1}, \mathbf{L}_m) \) for \( 0 \leq m \leq K \) and \( m \leq k \), requiring the independence between the treatment assignment at \( t_m \) and all potential outcomes \( Y^{(k)} \) for \( k \geq m \), given \( (\mathbf{A}_{m-1}, \mathbf{L}_m) \). Under this assumption, for \( k \leq T \), one may construct the mimicking potential outcome \( H(\psi^*) + Y^{(k)}(\mathbf{L}_k; \psi^*) \) for \( Y^{(k)} \). The induced restriction for \( \psi^* \) from \( \mathbb{E}\{Y^{(k)} \mid \mathbf{A}_{m-1} = \mathbf{0}, \mathbf{A}_m, \mathbf{L}_m \} = \mathbb{E}\{Y^{(k)} \mid \mathbf{A}_{m-1} = \mathbf{0}, \mathbf{L}_m \} \) is the same as (3). Therefore, it is not necessary to make the stronger version of Assumption 3 in our context.
2.2 | SES for discrete-time SNMMs

The semiparametric model is characterized by the discrete-time SNMM (1) and restriction (3), where the parameter of primary interest is \( \psi^* \). Robins (1994) established the semiparametric efficiency theory for the discrete-time SNMM, following the geometric approach of Bickel et al. (1993) by characterizing the nuisance tangent space, its orthogonal complementary space, where all influence functions of regular asymptotically linear (RAL) estimators belong to, and lastly the SES for the SNMM parameter.

Proposition 2 characterizes all influence functions of RAL estimators for \( \psi^* \).

**Proposition 2.** For the semiparametric model characterized by the discrete-time SNMM (1) and restriction (3), the influence function space for \( \psi^* \) is

\[
\Lambda^* = \left\{ G(\psi^*; F, c) : \text{for all } c(m, \tilde{V}_m) \in \mathbb{R}^p \right\},
\]

where \( \tilde{V}_m = (\tilde{A}_{m-1}, \tilde{T}_m) \) and \( G(\psi; F, c) = \sum_{m=1}^{K} c(m, \tilde{V}_m) (A_m - \operatorname{pr}(A_m = 1 \mid \tilde{V}_m))(H(\psi) - \mathbb{E}[H(\psi) \mid \tilde{V}_m]) \) indexed by \( c \). To make the notation accurate, the abbreviation \( c \) in \( G(\psi; F, c) \) means \( c(m, \tilde{V}_m) \).

The SES, that is, the most efficient one among the class in (4), often does not have a closed-form expression. To fix ideas, consider \( \gamma_m(\tilde{T}_m; \psi^*) = (\psi^*_i + \psi^*_m t_m)(\tau - t_m)^+ \). We now make a working assumption, which extends restriction (3) and allows us to derive an analytical expression of the SES of \( \psi^* \).

**Proposition 3** (Discrete-time SES). Suppose Assumptions 1–3 hold. Suppose further that for \( 0 \leq m \leq K \), \( \operatorname{var}[H(\psi^*) \mid \tilde{A}_m, \tilde{T}_m] = \operatorname{var}[H(\psi^*) \mid \tilde{V}_m] \). Then, the SES of \( \psi^* \) is

\[
S_{\text{eff}}(\psi^*; F) = G(\psi^*; F, c_{\text{eff}}),
\]

where

\[
c_{\text{eff}}(m, \tilde{V}_m) = \begin{cases} 
(\tau - t_m) - \mathbb{E}\{\text{dur}(t_m) \mid \tilde{A}_m = \tilde{0}, \tilde{T}_m\} & \\
t_m(\tau - t_m) - \mathbb{E}\{\text{dur}(t_m) \mid \tilde{A}_m = \tilde{0}, \tilde{T}_m\} & \\
\operatorname{var}\{H(\psi^*) \mid \tilde{V}_m\}^{-1} &
\end{cases}
\]

and \( \text{dur}(t_m) = \sum_{i=m}^{K-1} A_i(t_{i+1} - t_i) \) is the observed treatment duration from \( t_m \) to \( \tau \).

3 | SNMMs in continuous-time processes

3.1 | Setup, models, and assumptions

We now extend discrete-time SNMMs in Section 2 to continuous-time SNMMs. We assume that the variables can change their values at any real time between 0 and \( \tau \). We assume that all subjects are followed until \( \tau \) and consider censoring in Section 3.4. Each subject has multiple visit times. Let \( N(t) \) be the counting process for the visit times. Let \( L_t \) be the multidimensional covariate process. In contrast to the setting with discrete-time data processes, \( L_t \) is a vector of covariates at \( t \) and additional information of the past visit times up to but not including \( t \). This is because the past visit pattern, for example, the number and frequency of the visit times may be important confounders for the treatment and outcome processes. Let \( A_t \) be the binary treatment process. In our motivating application, the treatment can only be initiated at the follow-up visits; that is, if \( A_t = 1 \), then \( N(t) = 1 \). We will model the treatment process directly, although one can model first the visit time process and then treatment assignment at the visit times. Define \( Y(t) \) as the potential outcome at \( \tau \) had the subject initiated the treatment at \( t \), and define \( Y^*(\infty) \) as the potential outcome at \( \tau \) had the subject never initiated the treatment before \( \tau \). Let \( Y \) be the continuous outcome measured at \( \tau \). For regularity, we assume that the processes are càdlàg processes, that is, the processes are right-continuous with left limits. Let \( V_t = (A_{t-}, L_t) \) be the combined treatment and covariate process, where \( A_{t-} \) is the treatment information right before \( t \). We use the overline notation to denote a variable’s observed history; for example, \( \overline{A}_t = \{A_u : 0 \leq u \leq t, dN(u) = 1\} \). The subject’s full record is \( F = \{\overline{V}_t, Y(t) : 0 \leq t \leq \tau\} \).

The observed data for a subject through \( \tau \) is \( D = (\overline{V}_\tau, Y) \).

We assume the continuous-time SNMM as follows.

**Assumption 4** (Continuous-time SNMM). For \( t \geq 0 \), the continuous-time SNMM for the effect of the treatment initiation time is

\[
Y_t(L_t) = \mathbb{E}\left\{ Y(t) - Y^*(\infty) \mid L_t, T \geq t \right\} = \gamma_t(L_t; \psi^*);
\]

that is, \( \gamma_t(L_t; \psi) \) with \( \psi \in \mathbb{R}^p \) is a correctly specified model for \( \gamma_t(L_t) \) with the true parameter value \( \psi^* \). Moreover, \( Y(t) \sim Y^*(\infty) + \gamma_t(L_t; \psi^*) \) given \( (L_t, T \geq t) \), where \( \sim \) means “is (conditionally) distributed as.”

In the continuous-time SNMM (6), \( \psi^* \) can be interpreted as the treatment effect rate for the outcome. For the continuous-time SNMM, we assume that given \( (L_t, T \geq t) \),
the treatment effect only changes the location of the distribution of the outcome but not on other aspects of the distribution such as the variance. This assumption is stronger than the discrete-time SNMM in Assumption 1. But this assumption is weaker than the rank-preserving assumption of $Y^{(t)} = Y^{(\infty)} + \gamma(t; \psi^*)$ considered in Zhang et al. (2011). However, the rank preservation may be restrictive in practice, because it implies that for two subjects $i$ and $j$ with the same treatment and covariate history, $Y_i > Y_j$ must imply $Y_i^{(t)} > Y_j^{(t)}$. We relax this restriction by imposing a biological assumption. To link the observed outcome to the potential outcomes, we assume that $Y = Y^{(T)}$. Define the mimicking outcome for $Y^{(\infty)}$ as $H(\psi^*) = Y - \gamma(\infty; \psi^*)$. By Assumption 4, $H(\psi^*) \sim Y^{(\infty)}$, given $(\overline{L}_t, T \geq t)$.

The continuous-time SNMM (6) can model the treatment effect flexibly. For example, the two-parameter model $\gamma_i(\overline{L}_t; \psi^*) = (\psi_1^* + \psi_2^* t)(\tau - t)^+ + \psi_3^* t + \psi_4^* t + \psi_5^* t + \psi_6^* t$, where $l_1$ and $CD_4$ are the log viral load and CD4 count at $t$. We discuss effect modification and model selection in Section 6.

An important issue with data from user-initiated visits and treatment initiation is the potential selection bias and confounding, for example, sicker patients may visit the doctor more frequently and are likely to initiate treatment earlier. To overcome this issue, we impose the NUC assumption on the treatment process.

Assumption 5 (No unmeasured confounding). The hazard of treatment initiation is

$$
\lambda_T(t \mid F) = \lim_{h \to 0} h^{-1}P(t \leq t + h, \Gamma = 1 \mid \overline{V}_t, Y^{(\infty)}, T \geq t) = \lim_{h \to 0} h^{-1}P(t \leq t + h, \Gamma = 1 \mid \overline{V}_t, T \geq t),
$$

(7)
denoted by $\lambda_T(t \mid \overline{V}_t)$. Because treatment is never discontinued once it is initiated, we impose the condition of $\lambda_T(t \mid \overline{V}_t) = 0$ for $t > T$.

Assumption 5 implies that the hazard of treatment initiation at $t$ depends only on the observed treatment and covariate history $\overline{V}_t$ but not on the future observations and potential outcomes. This assumption holds if the set of historical covariates contains all prognostic factors for the outcome that affect the patient’s decision of visiting the doctor and initiating treatment. As an example, in the motivating application, time-invariant characteristics such as age at infection, gender, race, and whether ever used injection drugs are important confounders for the treatment and outcome processes. Moreover, time-varying CD4 and viral load are important confounders. Often, poor disease progression necessitates more frequent follow-up visits and earlier treatment initiation.

The treatment process $A_t$ can also be represented in terms of the counting process $N_T(t)$ and the at-risk process $R_T(t)$ of observing treatment initiation. Let $\sigma(V_t)$ be the $\sigma$-field generated by $V_t$, and let $\sigma(\overline{V}_t)$ be the $\sigma$-field generated by $U_{\leq t} \sigma(V_u)$. Under the standard regularity conditions for the counting process, $M_T(t) = N_T(t) - \int_0^t \lambda_T(u \mid \overline{V}_u)R_T(u)du$ is a martingale with respect to the filtration $\sigma(\overline{V}_t)$. Assumption 5 entails that the jumping rate of $N_T(t)$ at $t$ does not depend on $Y^{(\infty)}$, given $\overline{V}_t$. Because $H(\psi^*) \sim Y^{(\infty)}$ in the sense that it has the same distribution as $Y^{(\infty)}$ given $\overline{V}_t$, Assumption 5 also implies that the jumping rate of $N_T(t)$ at $t$ does not depend on $H(\psi^*)$, given $\overline{V}_t$.

To be formal, we show in the online supporting information that

$$
\lambda_T(t \mid \overline{V}_t, H(\psi^*)) = \lambda_T(t \mid \overline{V}_t).
$$

(8)

Therefore, under the standard regularity conditions, $M_T(t)$ is a martingale with respect to the filtration $\sigma(\overline{V}_t, H(\psi^*))$.

### 3.2 SES for continuous-time SNMMs

To estimate the causal parameter precisely, we establish the new semiparametric efficiency theory for the continuous-time SNMMs in parallel to that for the discrete-time SNMMs. We defer all proofs to the online supporting information.

Theorem 1. For the semiparametric model characterized by the continuous-time SNMM (6) and Assumption 5, the influence function space for $\psi^*$ is $\Lambda^\perp = \{G(\psi^*; F, c) :$ for all $c(u, \overline{V}_u) \in R^p\}$, where

$$
G(\psi; F, c) = \int_0^T c(u, \overline{V}_u) \left[ H(\psi) - E\left\{ H(\psi) \mid \overline{V}_u, T \geq u \right\} \right] \times R_T(u) dM_T(u).
$$

(9)

From Theorem 1, we can construct a wide class of estimating equation for $\psi^*$ based on $G(\psi; F, c)$ by varying the choice of $c(u, \overline{V}_u)$. The existence of a large number of estimators calls for a principled way to choose $c(u, \overline{V}_u)$ that leads to efficient estimators. Toward this end, we derive the SES for $\psi^*$ by $S_{\text{eff}}(\psi^*; F) = \prod\{S(\psi^*; F) \mid \Lambda^\perp\}$, where
$S(\psi^*; F)$ is the score function of $\psi^*$. This result motivates efficient estimators of $\psi^*$ in the next subsection.

**Theorem 2 (Continuous-time SES).** Under the semiparametric model characterized by the continuous-time SNMM (6) and Assumption 5, the SES of $\psi^*$ is

$$S_{\text{eff}}(\psi^*; F) = G(\psi^*; F, c_{\text{eff}}),$$

where $G(\psi; F, c)$ is defined in (9), $H_{\mu}(\psi) = H(\psi) - E[H(\psi) \mid \overline{V}_u, T \geq u]$, and

$$c_{\text{eff}}(u, \overline{V}_u) = \left[ E[\delta H_{\mu}(\psi^*)/\delta \psi \mid \overline{V}_u, T = u] - E[\delta H_{\mu}(\psi^*)/\delta \psi \mid \overline{V}_u, T \geq u] \right] \times \left[ \var[H(\psi^*) \mid \overline{V}_u, T \geq u] \right]^{-1}.$$  

(10)

To illustrate the theorem, we provide the explicit expression of the SES using an example.

**Example 1.** Consider $\gamma_i(T; \psi) = (\psi_1 + \psi_2 t)(\tau - t)^+$.

Suppose Assumption 5 holds. The SES of $\psi^*$ is $S_{\text{eff}}(\psi^*; F) = G(\psi^*; F, c_{\text{eff}})$, where

$$c_{\text{eff}}(u, \overline{V}_u) = \left( (\tau - u)^+ - E[\{\tau - T\}^+ \mid \overline{V}_u, T \geq u] \right) \times \left[ \var[H(\psi^*) \mid \overline{V}_u, T \geq u] \right]^{-1}.$$  

(12)

The proposed continuous-time SES contains the discrete-time SES as a special case. If the processes take observations at discrete times $\{t_0, \ldots, t_K\}$, then (i) the conditioning event $(\overline{V}_u, T \geq u)$ at $t_m$ is the same as $(\overline{A}_m = \overline{0}, \overline{L}_m)$, (ii) $M_I(t) = N_I(t) = \int_0^t \lambda_I(u \mid \overline{V}_u) R_I(u) du$ at $t = t_m$ becomes $A_m - pr(A_m = 1 \mid \overline{A}_{m-1} = \overline{0}, \overline{L}_m)$, and $E[\delta H_{\mu}(\psi^*)/\delta \psi \mid \overline{V}_u, T = t]$ at $t = t_m$ becomes

$$E[\delta H_{\mu}(\psi^*)/\delta \psi \mid \overline{V}_u, T = t_m]$$

$$= \left\{ \begin{array}{ll} (\tau - t_m)^+ - E\left\{ \text{dur}(t_m) \mid \overline{A}_m = \overline{0}, \overline{L}_m \right\} & \text{if } t_m(\tau - t_m)^+ - E\left\{ T \times \text{dur}(t_m) \mid \overline{A}_m = \overline{0}, \overline{L}_m \right\} & \text{if } t = t_m \end{array} \right.$$

Therefore, the continuous-time SES (10) reduces to the discrete-time SES (5).

### 3.3 Doubly robust and locally efficient estimators

We first construct a general class of estimators based on the estimating function $G(\psi^*; F, c)$. Because $E[G(\psi^*; F, c)] = 0$, we obtain the estimator of $\psi^*$ by solving

$$P_n[G(\psi; F, c)] = 0.$$  

(13)

In particular, Equation (13) with $c_{\text{eff}}$ provides the semiparametric efficient estimator of $\psi^*$.

In (13), we assume that the models for the potential outcome mean function $E[H(\psi^*) \mid \overline{V}_u, T \geq u]$ and the treatment process are known. In practice, they are often unknown and must be modeled and estimated from the data. We posit a working model $E[H(\psi^*) \mid \overline{V}_u, T \geq u; \hat{\beta}]$, such as a linear regression model, where $\hat{\beta}$ is a vector of unknown parameters. We also posit a proportional hazards model with time-dependent covariates for the treatment process; that is, $\lambda_I(t \mid \overline{V}_u; \alpha) = \lambda_{I, \alpha}(t) \exp[\alpha^t W_I(t, \overline{V}_u)]$, where $\lambda_{I, \alpha}(t)$ is an unknown baseline hazard function, $W_I(t, \overline{V}_u)$ is a prespecified function of $t$ and $\overline{V}_u$, and $\alpha$ is a vector of unknown parameters. Under Assumption 5, we can estimate $\alpha$ and $\lambda_{I, \alpha}(t)$ from the standard software such as “coxph” in R (R Development Core Team, 2012).

Fitting the time-dependent proportional hazards model to the data $\{\overline{V}_{I,j}, T_j, Y_j\} : i = 1, \ldots, n$, where $Y_i = 0(T_i \leq \tau)$, treating the treatment initiation as the failure event, we obtain the estimators $\hat{\alpha}$ and $\hat{\lambda}_{I, \alpha}(t)$. Then, we obtain $\hat{\lambda}_I(u \mid \overline{V}_u) = \exp[\hat{\alpha}^t W_I(u, \overline{V}_u)] \hat{\lambda}_{I, \alpha}(u)$ and $\hat{M}_I(t) = N_I(t) - \int_0^t \hat{\lambda}_I(u \mid \overline{V}_u) R_I(u) du$. As we show below, the resulting estimator $\hat{\psi}$ achieves the double robustness property.

**Theorem 3 (Double robustness).** Suppose the continuous-time SNMM (6) in Assumption 4, and Assumption 5 hold. The estimator $\hat{\psi}$ solving the estimating equation (13) based on the class of $G(\psi; F, c)$ in (9) by varying $c(u, \overline{V}_u)$ is doubly robust in that it is consistent if at least one of the models for the potential outcome mean function $E[H(\psi^*) \mid \overline{V}_u, T \geq u]$ and the treatment process is correctly specified.

The choice of $c$ does not affect the double robustness but the efficiency of the resulting estimator. For efficiency consideration, we consider $c_{\text{eff}}$ in (11). The resulting estimator solving the estimating equation (13) with $c_{\text{eff}}$ is locally efficient, in the sense that it achieves the semiparametric efficiency bound if the working models for the treatment process and the potential outcome mean function $E[H(\psi^*) \mid \overline{V}_u, T \geq u]$ are correctly specified. Because $c_{\text{eff}}$ depends on the unknown distribution, we require additional models for $E[\{\tau - T\}^+ \mid \overline{V}_u, T \geq u]$ and $E[T(\tau - T)^+ \mid \overline{V}_u, T \geq u]$ to approximate $c_{\text{eff}}$. For example, we can approximate $E[\{\tau - T\}^+ \mid \overline{V}_u, T \geq u]$ by $P(T \leq \tau \mid \overline{V}_u, T \geq u) \times E[\{\tau - T\} \mid \overline{V}_u, u \leq T \leq \tau]$ and each approximated by (logistic) linear models. For $\var[H(\psi^*) \mid \overline{V}_u, T \geq u]$, we consider the following options: (i) assume $\var[H(\psi^*) \mid \overline{V}_u, T \geq u]$ to be a constant, and (ii) approximate $\var[H(\psi^*) \mid \overline{V}_u, T \geq u]$
by the sample variance of \( H(\hat{\psi}_p) \) among subjects with \( T \geq u \), where \( \hat{\psi}_p \) is a preliminary estimator. We compare the two options via simulation. Although option (ii) provides a slight efficiency gain in the estimation, for ease of implementation we recommend option (i). Option (i) is common in the generalized estimating equation framework. From here on, we use this option for \( c \) and suppress the dependence on \( c \) for estimating functions.

**Remark 2.** Double robustness has appeared for estimators in other contexts of causal inference, such as the augmented inverse probability weighting estimator (AIPW) of the average treatment effect (e.g., Lunceford and Davidian, 2004; Bang and Robins, 2005). Specifically, the AIPW estimator is consistent if either the potential outcome mean function or the propensity score is correctly specified, similar to the requirement for the nuisance functions in Theorem 3. However, the double robustness result in Theorem 3 requires the SNMM to be correctly specified at the outset. Thus, our result requires an additional modeling assumption on the SNMM compared to the typical result for the AIPW estimator.

### 3.4 Censoring

As in the AIEDRP study, in most longitudinal observational studies, subjects may drop out of the study prematurely before the end of the study, which renders the data censored at the time of dropout. If the censoring mechanism depends on time-varying prognostic factors, for example, sicker patients drop out of the study with a higher probability than healthier patients, the patients remaining in the study is a biased sample of the full population. We now introduce \( C \) to be the time to censoring. Let \( X = \min(C, \tau) \) be time to censoring or the end of the study, whichever came first. Let \( \delta_C = I(C \geq \tau) \) be the indicator of not censoring before \( \tau \). The observed data are \( D = (X, \delta_C, \delta_C Y) \).

In the presence of censoring, the estimating equation (13) is not feasible. We assume an ignorable censoring mechanism as follows.

**Assumption 6** (Ignorable censoring). The hazard of censoring is

\[
\lambda_C(t \mid F) = \lim_{h \to 0} h^{-1} P(t \leq C < t + h \mid F, C \geq t) = \lim_{h \to 0} h^{-1} P(t \leq C < t + h \mid \delta_C = 0)
\]

(14)

denoted by \( \lambda_C(t \mid \delta_C = 0) \).

Assumption 6 states that \( \lambda_C(t \mid F) \) depends only on the past treatment and covariate history until \( t \), but not on the future variables and potential outcomes. This assumption holds if the set of historical covariates contains all prognostic factors for the outcome that affect the possibility of loss to follow up at \( t \). Under this assumption, the missing data due to censoring are missing at random (Rubin, 1976). From \( \lambda_C(t \mid \delta_C = 0) \), we define \( K_C(t \mid \delta_C = 0) = \exp(- \int_0^t \lambda_C(u \mid \delta_C = 0) du) \), which is the probability of the subject not being censored before \( t \).

We discuss important implications of Assumption 6 on the mimicking potential outcome and the treatment process. First, Assumptions 4 and 6 yield \( H(\psi^*) \sim Y^{(\infty)} \), given \( (\delta_C, \tau, T \geq t, C \geq t) \). Second, under Assumption 6, the hazard of treatment initiation in (7) is equal to \( \lim_{h \to 0} h^{-1} P(t \leq T < t + h, \Gamma = 1 \mid \delta_C = 0) \). Redefining \( T \) to be the time to treatment initiation, or censoring, or the end of the study, whichever came first, (7) can be estimated by conditioning on \( T \geq t \) with the new definition of \( T \). Therefore, the estimating equation (13) restricted to the uncensored subjects remain unbiased. This leads to the CC estimator \( \hat{\psi} \) solving the following equation:

\[
P_n\{\delta_C G(\psi; F)\} = 0.
\]

In fact, one can obtain a class of weighting estimators by solving

\[
P_n\{\delta_C g(\delta_C) G(\psi; F)\} = 0
\]

(16)

for any weight function \( g(\delta_C) \in \mathcal{R} \). In particular, choosing \( g(\delta_C) \) to be \( \{K_C(\tau \mid \delta_C)\}^{-1} \) leads to the IPCW estimator (Rotnitzky et al., 2007). We show that the CC estimator \( \hat{\psi} \) is optimal among all estimators solving the estimating equation (16). This is because using varying weights reduces the effective sample size compared to constant weights, a classical result in survey sampling (Kish, 1992). Theorem 4 summarizes the asymptotic properties of the CC estimator.

**Theorem 4.** Suppose the continuous-time SNMM (6), Assumptions 5 and 6, and regularity conditions in Assumption S1 hold. The CC estimator \( \hat{\psi} \) solving the estimating equation (15) is doubly robust in that it is consistent and asymptotically Normal if at least one of the models for the potential outcome mean function \( E[H(\psi^*) \mid \delta_C, T \geq u] \) and the treatment process is correctly specified. Moreover, if both nuisance models are correctly specified, \( \hat{\psi} \) achieves the smallest variance among the class of estimators solving the estimating equation (16).

The asymptotic Normal distribution is presented in the online supporting information and is agnostic about
whether the potential outcome mean function \( E[H(\psi^*) \mid \overline{V}_u, T \geq u] \) or the treatment process is correctly estimated. However, it is difficult to use the asymptotic variance formula for variance estimation because it requires approximating additional nuisance functions. From Theorem 4, under the conditions that ensure double robustness, \( \hat{\psi} \) is asymptotically linear with a Normal limiting distribution, and therefore, we can use the nonparametric bootstrap for variance estimation.

### 4 | SIMULATION STUDY

We now evaluate the finite-sample performance of the proposed estimator on simulated data sets with two objectives. First, we assess the double robustness and efficiency of the proposed estimator based on the SES, compared with some preliminary estimator. Second, to demonstrate the impact of data discretization as commonly done in practice, we include the g-estimator applied to the preprocessed data. We simulate 1000 data sets under two settings with and without censoring with sample size \( n = 1000 \). Additional simulation results with \( n = 2000 \) are presented in the online supporting information.

In Setting I, we generate two covariates, one time-independent (\( L_{TI} \)) and one time-dependent (\( L_{TD} \)). The time-independent covariate \( L_{TI} \) is generated from a Bernoulli distribution with mean equal to 0.55. The time-dependent covariate is \( L_{TD,j} = l_1 \times I(0 \leq t < 0.5) + l_2 \times I(0.5 \leq t < 1) + l_3 \times I(1 \leq t < 1.5) + l_4 \times I(1.5 \leq t \leq 2) \), where \((l_1, l_2, l_3, l_4)^T\) is a \( 1 \times 4 \) row vector generated from a multivariate Normal distribution with mean equal to \((0,0,0,0)\) and covariance equal to \(0.7^{i-1}/i\) for \( i, j = 1, \ldots, 4\). We assume that the time-dependent variable remains constant between measurements. The maximum follow-up time is \( \tau = 2 \) (in year). We generate the time to treatment initiation \( T \) with the hazard rate \( \lambda_T(t \mid \overline{V}_T) = \lambda_{T,0}(t) \exp(\alpha_1 L_{TI} + \alpha_2 L_{TD,j}) \) with \( \lambda_{T,0}(t) = 0.4 \), \( \alpha_1 = 0.15 \), and \( \alpha_2 = 0.8 \). We generate \( T \) according to the time-dependent model sequentially. This is because the hazard of treatment initiation in the time interval from \( t_1 = 0 \) to \( t_2 = 0.5 \) differs from the hazard of treatment initiation in the next interval and so on; see the online supporting information for details. We let \( Y(\infty) = L_{TD,j} \) be the potential outcome had the subject never initiated the treatment before \( \tau \). The observed outcome is \( Y = Y(\infty) + \gamma_T(\overline{V}_T; \psi^*) \), where \( \gamma_T(\overline{V}_T; \psi^*) = (\psi^*_1 + \psi^*_2 t)(\tau - t)^+ \) with \( \psi^*_1 = 15 \) and \( \psi^*_2 = -1 \).

We consider the following estimators with details for the nuisance models and their estimation presented in the online supporting information:

a) A preliminary estimator \( \hat{\psi}_p \) solves (13) with \( E[H(\psi^*) \mid \overline{V}_u, T \geq u] \equiv 0 \) and \( c(u, \overline{V}_u) = (1, u)^T(\tau - u)^+ - E[(1, T)^T(\tau - T)^+ \mid \overline{V}_u, T \geq u] \). Therefore, \( \hat{\psi}_p \) corresponds to the proposed estimator with a misspecified model for \( E[H(\psi^*) \mid \overline{V}_u, T \geq u] \).

b) The proposed estimator \( \hat{\psi}_{cont,1} \) solves (13), where we replace \( var[H(\hat{\psi}) \mid \overline{V}_u, T \geq u] \) by a constant.

c) The proposed estimator \( \hat{\psi}_{cont,2} \) solves (13), where we obtain \( \text{var}[H(\hat{\psi}^*) \mid \overline{V}_u, T \geq u] \) by the empirical variance of \( H(\hat{\psi}_p) - E[H(\hat{\psi}_p) \mid \overline{V}_u, T \geq u; \hat{\beta}] \), restricted to subjects with \( T \geq u \).

d) The g-estimator \( \hat{\psi}_{disc,\hat{g}} \) in Section 2 applies to the monthly data after discretization with 24 equally spaced time points from 0 to \( \tau \). For \( m \geq 1 \), at the \( mth \) time point \( t_m \), \( L_m \) is the average of \( L_i \) from \( t_{m-1} \leq t \leq t_m \). \( A_m \) is the indicator of whether the treatment is initiated before \( t_m \), and the time to treatment initiation \( T \) is \( t_m \) if \( A_m = 1 \) and \( \overline{A}_{m-1} = 0 \). The g-estimator solves the estimating equation based on (5), where the nuisance models are estimated similar to what are used for \( \hat{\psi}_{cont,1} \) but with the reshaped data.

To investigate the double robustness in Theorem 3, we consider two models for estimating \( M_{TD} \): the correctly specified proportional hazards model with both time-independent and time-dependent covariates; and the misspecified proportional hazards model with only time-independent covariate. For all estimators, we use the bootstrap for variance estimation with the bootstrap size 100.

Table 1 shows the simulation results in Setting I. Under Scenario (i), when the model for the treatment process is correctly specified, \( \hat{\psi}_p \), \( \hat{\psi}_{cont,1} \), and \( \hat{\psi}_{cont,2} \) show small biases. As a result, the coverage rates are close to the nominal level. Under Scenario (ii), when the model for the treatment process is misspecified, \( \hat{\psi}_p \) shows large biases, but \( \hat{\psi}_{cont,1} \) and \( \hat{\psi}_{cont,2} \) still show small biases. Moreover, the root mean squared errors of \( \hat{\psi}_{cont,1} \) and \( \hat{\psi}_{cont,2} \) decrease as the sample size increases; see the additional simulation results in the online supporting information. This confirms the double robustness of the proposed estimators. The proposed estimator \( \hat{\psi}_{cont,2} \) with \( \text{var}[H(\hat{\psi}) \mid \overline{V}_u, T \geq u] \) produces slightly smaller standard errors; however, this reduction is not large. In practice, we recommend \( \hat{\psi}_{cont,1} \) because of its simpler implementation than \( \hat{\psi}_{cont,2} \). We note large biases in the g-estimator, which illustrates the consequence of data preprocessing for the subsequent analysis.

In Setting II, we further generate the time to censoring \( C \) with the hazard rate \( \lambda_C(t \mid \overline{V}_T) = \lambda_{C,0}(t) \exp(\eta_T L_{TI} + \eta_2 L_{TD,j}) \), with \( \lambda_{C,0}(t) = 0.2 \), \( \eta_1 = 0.15 \), and \( \eta_2 = 0.35 \). In the presence of censoring, we consider the three estimators (a), (b), and (d) considered in Setting I applied to the
uncensored subjects with weighting; that is, the estimators solving the corresponding estimating equations (16) weighted by \( g(V_\tau) \). To investigate the robustness and optimality of \( g(V_\tau) = 1 \) in Theorem 4, we consider \( g(V_\tau) = 1 \) and \( g(V_\tau) = (K_C(\tau | V_\tau))^{-1} \) with two models for estimating \( K_C \): the correctly specified proportional hazards model with \( (L_{T1}, L_{TD}) \), and the misspecified proportional hazards model with \( (L_{T1}, L_{T1}) \).

Table 2 shows the simulation results in Setting II. Under scenarios when either the model for the potential outcome mean function \( E[H(\psi^*) | V_u, T \geq u] \) or the model for the treatment process is incorrectly specified, the estimators show small biases, regardless of the specification of \( g(V_\tau) \). Moreover, if both models for the potential outcome mean function and the treatment process are misspecified, \( \hat{\psi}_p \) shows large biases. Under the same model specification, the CC estimator with \( g(V_\tau) = 1 \) is more efficient than the IPCW estimator with a correctly specified censoring model which is more efficient than the IPCW estimator with a misspecified censoring model. This confirms the optimality of \( g(V_\tau) = 1 \) in Theorem 4. Again, the discretized g-estimator shows large biases across all scenarios.

5 ESTIMATING THE EFFECT OF TIME TO INITIATING HAART

We apply our method to the observational AIEDRP database consisting of 1762 HIV-positive patients diagnosed during acute and early infection (Hecht et al., 2006). This data set was previously used by Lok and DeGruttola (2012) and Yang and Lok (2016, 2018); all these methods were based on the monthly data after discretization. As discussed in the introduction, the observations from the original data are collected by user-initiated visits and are irregularly spaced (Hecht et al., 2006). Figure 1 shows the visit times for five random patients. As can be seen, we have irregular visits, and the number and frequency of visits vary from patient to patient. We aim to estimate the average causal effect of the time to HAART initiation on the mean CD4 count at year 2 after HIV infection directly on the basis of the original data without discretization. The outcome variable \( Y \) is the CD4 count measured by the end of year 2, with the interquantile range from 443 to 794 cells/mm\(^3\). The observed time to treatment initiation ranges continuously from 12 to 282 days. To ensure the NUC and ignorable censoring assumptions hold, we include the following baseline and time-varying covariates: age at infection, gender, race, injection drug ever/never, and measured CD4 count and log viral load at follow-up visits. We assume a continuous-time SNMM \( \gamma(V_u; \psi^*) = (\psi^*_1 + \psi^*_2 t)(\tau - t)^+ \), where \( \psi^*_2 \) quantifies the impact of time to treatment initiation. The rationale for this modeling choice is because the duration of treatment may well be predictive of its effect.

We consider the proposed CC estimators \( \hat{\psi}_{cont,1} \) and \( \hat{\psi}_{cont,2} \) specified in Section 4 applied to the uncensored subjects. The estimation procedure requires specifying and fitting nuisance models, which we discuss below.

Model for the potential outcome mean function. \( E[H(\hat{\psi}_p) | V_u, T \geq u; \beta] \) is a linear regression model where the covariates include age, male, race, injdrug, CD4\(_{u, 3/4}\), lvi\(_u\), CD4\(_{u, 3/4}\) \((\tau - u)\), CD4\(_{u, 3/4}\) \((\tau - u)\) \(\times\) age, CD4\(_{u, 3/4}\) \((\tau - u)\) \(\times\) male, CD4\(_{u, 3/4}\) \((\tau - u)\) \(\times\) race, CD4\(_{u, 3/4}\) \((\tau - u)\) \(\times\) injdrug, CD4\(_{u, 3/4}\) \((\tau - u)\) \(\times\) lvi\(_u\), CD4slope\(_u\) measured, CD4slope\(_u\), \((\tau - u)^{1/2}(6 - u)^{1/2}\), and \((36 - u)^{1/2}\). This model specification is motivated based on the substantive literature; see, for example, May et al. (2009).
TABLE 2  Simulation results in Setting II with censoring based on 1000 simulated datasets: the Monte Carlo bias, standard error, root mean square error of the estimators, and coverage rate of 95% confidence intervals

| Scenario | Model for \( M_T \) (√) and \( g \equiv 1 \) | Model for \( M_T \) (×) | Model for \( M_T \) (√) | Model for \( M_T \) (×) | Model for \( M_T \) (√) | Model for \( M_T \) (×) |
|-----------|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| (\( n = 1000 \)) | Method | \( \psi_1^* \) | \( \psi_2^* \) | \( \psi_1^* \) | \( \psi_2^* \) | \( \psi_1^* \) | \( \psi_2^* \) |
| Scenario (i): Model for \( M_T \) (√) and \( g \equiv 1 \) | Model for POM (×) | \( \hat{\psi}_p \) | 0.4 | -1.4 | 65.4 | 110.9 | 65.4 | 110.9 | 95.7 | 94.6 |
| | Model for POM (√) | \( \hat{\psi}_{cont,1} \) | -0.1 | 3.4 | 65.0 | 106.5 | 65.0 | 106.6 | 95.0 | 94.5 |
| | - | \( \hat{\psi}_{disc,g} \) | 276.2 | 315.6 | 74.5 | 121.5 | 286.0 | 338.2 | 8.7 | 10.9 |
| Scenario (ii): Model for \( M_T \) (×) and \( g \equiv 1 \) | Model for POM (×) | \( \hat{\psi}_p \) | 63.7 | 177.5 | 64.4 | 110.6 | 90.6 | 209.2 | 69.1 | 43.5 |
| | Model for POM (√) | \( \hat{\psi}_{cont,1} \) | 6.5 | 5.3 | 65.0 | 107.5 | 65.4 | 107.7 | 94.2 | 94.5 |
| | - | \( \hat{\psi}_{disc,g} \) | 281.2 | 318.7 | 74.4 | 120.9 | 290.9 | 340.9 | 7.4 | 8.8 |
| Scenario (iii): Model for \( M_T \) (√) and \( g = K^{-1}_C(\psi) \) | Model for POM (×) | \( \hat{\psi}_p \) | -0.3 | -0.3 | 66.8 | 114.6 | 66.8 | 114.6 | 95.7 | 95.5 |
| | Model for POM (√) | \( \hat{\psi}_{cont,1} \) | -0.6 | 5.2 | 65.7 | 109.4 | 65.7 | 109.5 | 95.2 | 94.2 |
| | - | \( \hat{\psi}_{disc,g} \) | 273.4 | 312.4 | 75.0 | 124.1 | 283.5 | 336.2 | 10.3 | 12.0 |
| Scenario (iv): Model for \( M_T \) (×) and \( g = K^{-1}_C(\psi) \) | Model for POM (×) | \( \hat{\psi}_p \) | 63.5 | 160.0 | 63.2 | 109.8 | 89.6 | 194.0 | 63.1 | 48.3 |
| | Model for POM (√) | \( \hat{\psi}_{cont,1} \) | 10.2 | 2.5 | 66.2 | 108.8 | 67.0 | 108.9 | 94.1 | 94.7 |
| | - | \( \hat{\psi}_{disc,g} \) | 283.7 | 311.4 | 76.0 | 124.7 | 293.7 | 335.4 | 11.8 | 13.2 |
| Scenario (v): Model for \( M_T \) (√) and \( g = K^{-1}_C(\psi) \) | Model for POM (×) | \( \hat{\psi}_p \) | 0.2 | -1.5 | 67.3 | 114.0 | 67.3 | 114.0 | 95.2 | 95.6 |
| | Model for POM (√) | \( \hat{\psi}_{cont,1} \) | -0.3 | 3.7 | 66.7 | 109.3 | 66.7 | 109.4 | 94.6 | 94.3 |
| | - | \( \hat{\psi}_{disc,g} \) | 273.4 | 312.4 | 75.0 | 124.1 | 283.5 | 336.2 | 8.4 | 10.2 |
| Scenario (vi): Model for \( M_T \) (×) and \( g = K^{-1}_C(\psi) \) | Model for POM (×) | \( \hat{\psi}_p \) | 72.6 | 191.8 | 66.1 | 113.6 | 98.2 | 223.0 | 61.4 | 32.2 |
| | Model for POM (√) | \( \hat{\psi}_{cont,1} \) | 13.7 | 4.1 | 66.9 | 107.6 | 68.3 | 107.7 | 94.2 | 94.6 |
| | - | \( \hat{\psi}_{disc,g} \) | 283.7 | 311.4 | 76.0 | 124.7 | 293.7 | 335.4 | 7.7 | 9.3 |

Note “POM” means the potential outcome mean function \( E[H(\psi^*) \mid \psi, T \geq u] \); √ (is correctly specified), and × (is misspecified).

Model for the treatment process. The model for the treatment process \( (M_T) \) is a time-dependent proportional hazards model adjusting for gender, age (age at infection), race (white non-Hispanic race), injdrug (injection drug ever/never), CD4\(^{1/2}\) (square root of current CD4 count), lvl\(_u\) (log viral load), days from last visit\(_u\) (number of days since the last visit), first visit\(_u\) (whether the visit is the first visit), and second visit\(_u\) (whether the visit is the second visit).

Other nuisance models. \( E(\tau - T \mid \bar{T}_u, T \geq u) \) and \( E(T(\tau - T) \mid \bar{T}_u, T \geq u) \) are linear regression models where the covariates include \( u, (\tau - u), \text{male}(\tau - u), \text{age}(\tau - u), \text{race}(\tau - u), \text{injdrug}(\tau - u), \text{CD4}\(^{1/2}\)(\tau - u), \text{lvl}\(_u\)(\tau - u), \text{days from last visit}\(_u\)(\tau - u), \text{first visit}\(_u\)(\tau - u), \text{and second visit}\(_u\)(\tau - u).\)

We use bootstrap for variance estimation with the bootstrap size 100 and compute the 95% Wald confidence interval. We also include the discretized g-estimator (Lok and DeGruttola, 2012) applied to the monthly data. Table 3 shows the results for the effect of time to HAART initiation on the CD4 count at year 2. We note only slight differences in the point estimates between the proposed estimators. The discretized g-estimator is much larger than the proposed estimators for \( \psi_1^* \), but all estimators have similar
results for $\frac{\psi}{\psi_2}$. The results show that earlier HAART initiation is better in increasing CD4 counts. Our estimators respect the underlying continuous-time nature of data processes. Based on our results, on average, initiation of HAART at the time of infection ($t = 0$) can increase CD4 counts at year 2 by 14.2 cells/mm$^3$ per month $\times$ 24 months $\approx 341$ cells/mm$^3$; while initiation of HAART 3 months after the time of infection can increase CD4 counts at year 2 by $(14.2 - 0.96 \times 3) \times (24 - 3) \approx 238$ cells/mm$^3$.

In the supporting information, we conduct a sensitivity analysis using an elaborated SNMM with possible treatment effect modifiers. The analysis also shows earlier HAART initiation is better in increasing CD4 counts, although the result becomes not significant possibly due to the increased number of the SNMM parameter. Finally, we add a caveat that we require the SNMMs to be correctly specified and a formal goodness-of-fit test for model assessment will be our future work.

6 | DISCUSSION

In this paper, we have developed a new semiparametric estimation framework for continuous-time SNMMs to evaluate treatment effects with irregularly spaced longitudinal observations under the assumptions of NUC and ignorability of censoring. We do not require specifying the full distribution of the covariate, treatment, outcome, and censoring processes. Our approach achieves a double robustness property requiring the correct specification of either the model for the potential outcome mean function or the model for the treatment process, regardless of whether or not the model for the censoring process is correctly specified. As discussed previously, the key assumptions hold if all variables are measured that are related to both treatment and outcome and that are related to both censoring and outcome. Although essential, they are not verifiable based on the observed data but rely on subject matter experts to assess their plausibility.

The proposed framework is also applicable to the analysis of patient-reported outcomes in pragmatic clinical trials. Although trial protocols often require collecting outcomes at prefixed time points after randomization, patients may delay or even miss their visits in practice, resulting in irregular-spaced observations. One important implication of the proposed framework for future trial designs is to collect a sufficiently rich set of variables that predict the treatment and censoring processes and ensure the required assumptions hold.

There are several directions for future work: (i) we will extend the continuous-time SNMMs framework to other types of outcomes such as binary outcomes and survival outcomes; and (ii) we will develop a variable selection procedure for identifying effect modifiers. The insight is that we have a larger number of estimating functions than the number of parameters. The problem for effect modifiers selection falls into the recent work of Chang et al. (2018) on high-dimensional statistical inferences with overidentification; (iii) a goodness-of-fit test using overidentifiability (Yang and Lok, 2016). can also be developed to assess the assumption for the SNMM; and (iv) we will develop sensitivity analyses to assess the impact of possible uncontrolled confounding (Yang and Lok, 2018).

ACKNOWLEDGMENTS

The author would like to thank Anastasios A. Tsiatis for insightful and fruitful discussions. The author is grateful to the patients who volunteered for AIEDRP and the AIEDRP study team. Yang is partially supported by the NSF grant DMS 1811245, NCI grant P01 CA142538, NIA grant 1R01AG066883, and NIEHS grant 1R01ES031651.

DATA AVAILABILITY STATEMENT

The data that support the findings of this article are available from the Acute Infection and Early Disease Research Program (AIEDRP) study team. Data are available from the author with the permission of the AIEDRP study team.

ORCID

Shu Yang https://orcid.org/0000-0001-7703-707X

REFERENCES

Bang, H. and Robins, J.M. (2005) Doubly robust estimation in missing data and causal inference models. Biometrics, 61, 962–973.

Bickel, P.J., Klaassen, C., Ritov, Y. and Wellner, J. (1993) Efficient and Adaptive Inference in Semiparametric Models. Baltimore: Johns Hopkins University Press.

Chang, J., Tang, C.Y. and Wu, T.T. (2018) A new scope of penalized empirical likelihood with high-dimensional estimating equations. Annals of Statistics, 46, 3185–3216.

Hecht, F.M., Wang, L., Collier, A., Little, S., Markowitz, M., Margolick, J. et al. (2006) A multicenter observational study of the potential benefits of initiating combination antiretroviral therapy during acute HIV infection. Journal of Infectious Diseases, 194, 725–733.

Holland, P.W. (1986) Statistics and causal inference. Journal of the American Statistical Association, 81, 945–960.

Kish, L. (1992) Weighting for unequal pi. Journal of Official Statistics, 8, 183–200.

Lok, J.J. (2017) Mimicking counterfactual outcomes to estimate causal effects. Annals of Statistics, 45, 461–499.

Lok, J.J. (2008) Statistical modeling of causal effects in continuous time. Annals of Statistics, 36, 1464–1507.

Lok, J.J. and DeGruttola, V. (2012) Impact of time to start treatment following infection with application to initiating HAART in HIV-positive patients. Biometrics, 68, 745–754.
Lunceford, J.K. and Davidian, M. (2004) Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, 23, 2937–2960.

May, M., Wood, R., Myer, L., Taffé, P., Rauch, A., Battegay, M. et al. (2009) CD4 T-cell declines by ethnicity in untreated HIV-1 infected patients in South Africa and Switzerland. *Journal of Infectious Diseases*, 200, 1729–1735.

Neugebauer, R., Silverberg, M.J. and van der Laan, M.J. (2010) Observational study and individualized antiretroviral therapy initiation rules for reducing cancer incidence in HIV-infected patients. Technical Report, U.C. Berkeley Division of Biostatistics Working Paper Series: Working Paper 272.

Robins, J.M. (1994) Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics – Theory and Methods*, 23, 2379–2412.

Robins, J.M. (1998) Correction for non-compliance in equivalence trials. *Statistics in Medicine*, 17, 269–302.

Robins, J.M., Blevins, D., Ritter, G. and Wulfsohn, M. (1992) G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. *Epidemiology*, 3, 319–336.

Robins, JM, Hernán, MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice, G, Davidian, M, Verbeke, G, Molenberghs, G, editors. *Advances in Longitudinal Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC Press; (2009). 553–599.

Rubin, D.B. (1976) Inference and missing data. *Biometrika*, 63, 581–592.

SUPPORTING INFORMATION
Web Appendices and Tables referenced in Sections 3—5 and R code for implementing the proposed method are available with this article at the *Biometrics* website on Wiley Online Library.

How to cite this article: Yang S. (2022) Semiparametric estimation of structural nested mean models with irregularly spaced longitudinal observations. *Biometrics*, 78, 937–949. [https://doi.org/10.1111/biom.13471](https://doi.org/10.1111/biom.13471)