Trials with Irregular and Informative Assessment Times: A Sensitivity Analysis Approach

Bonnie Smith  
Department of Biostatistics,  
Johns Hopkins Bloomberg School of Public Health  
Shu Yang  
Department of Statistics, North Carolina State University  
Andrea J. Apter  
Department of Medicine,  
University of Pennsylvania Perelman School of Medicine  
and  
Daniel O. Scharfstein  
Department of Population Health Sciences,  
University of Utah School of Medicine

Abstract

Many trials are designed to collect outcomes at pre-specified times after randomization. However, in practice, there is often substantial variability in the times at which participants are actually assessed, which poses a challenge to learning about the treatment effects at the targeted assessment times. Any method for analyzing a trial with such irregular assessment times relies on untestable assumptions. Therefore, conducting a sensitivity analysis is an important step that can strengthen conclusions from such trials. However, no sensitivity analysis methodology has been developed. We develop a methodology that accounts for possibly informative assessment times, where assessment at time $t$ may be related to the outcome at that time, even after accounting for observed past history. We implement our methodology using a new augmented inverse-intensity-weighted estimator, and we apply it to a trial of low-income participants with uncontrolled asthma. We also evaluate the performance of our estimation procedure in a realistic simulation study.

Keywords: Assessment at random, Augmented inverse intensity weighting, Influence function, Asthma
1 Introduction

Randomized trials are often designed to collect outcome information at certain pre-specified times after randomization. In practice, there is often substantial variability in the times at which participants are actually assessed. This poses a challenge to learning the treatment effect at each targeted assessment time, i.e., the difference in the treatment-specific mean outcomes that we would see were all participants assessed exactly on-schedule. Such irregular assessment times can be due to a number of obstacles, which may or may not be related to participants’ outcomes. When they are related, we say that the assessment times are informative. For example, in an asthma trial, a participant may be less likely to attend a data collection appointment during a period when they are having an asthma exacerbation.

An ad hoc approach that is commonly used in this situation is to form windows around each targeted assessment time. Assessments that occur within a given window are treated as though they had occurred at the targeted time itself, while any observations that fall outside of the windows are excluded from the analysis. This approach can lead to biased estimation of treatment effects. One reason is that participants’ outcomes at their actual assessment times may be systematically different than their outcomes at the targeted times. This can occur if the true mean in one or both arms changes with time, or if assessment times are informative. A second reason is that the excluded, out-of-window observations may induce selection bias. A number of more principled approaches have been developed in the literature (see Section 2.3). However, any approach must rely on untestable assumptions about the relationship between assessment times and outcomes. This is analogous to trials with missing data, in which it is widely recognized that conducting a sensitivity analysis is crucially important (see, for example, the National Research Council report, *The Prevention and Treatment of Missing Data in Clinical Trials* [10]). While sensitivity analysis is also important for trials with irregular assessment times, no formal methodology has been developed.

Scharfstein and colleagues [32] have described three types of sensitivity analysis. These range from the ad hoc type, in which results are compared across a small number of different inferential methods, to the more comprehensive global type. A global sensitivity analysis conducts inference across a broad family of assumptions, within a range guided by subject-matter expertise. In this paper, we develop a global sensitivity analysis methodology anchored around the assessment at random (AAR) assumption, which posits that outcomes and assessments at time $t$ are independent after accounting for study data observed before time $t$. Our methodology evaluates inferences under departures from AAR, where the outcome at time $t$ is allowed to directly impact whether participants are assessed at that time, after accounting for past history. Such dependence can be expected in, for example, asthma or pain trials, where more severe pain or difficulty breathing at a given time may keep participants from coming to appointments. We derive a new augmented inverse-intensity-weighted estimator which we show is consistent and asymptotically normal under our non-AAR models.

The paper is organized as follows. In Section 2, we define notation, review existing methods, and introduce the Asthma Research for the Community (ARC) trial [4]. We
specify our models and sensitivity analysis framework in Section 3. In Section 4, we develop influence function-based inference for our target parameters. In Section 5, we describe our approach for constraining the range of assumptions to be included in a sensitivity analysis. We analyze the ARC trial in Section 6. In Section 7, we evaluate our method in a realistic simulation study. Section 8 concludes with a discussion.

2 Background

2.1 Notation

Let \( \tau \) be the end of follow-up, and let \( [a, b] \) be the time period over which we wish to conduct inference, where \( t = a \) is a time before any post-baseline assessments occur and \( b \leq \tau \). Let \( Y(t) \) denote the outcome of a random individual at time \( t \), and let \( N(t) \) be their number of assessments from time \( a \) up through and including time \( t \). We refer to the counting process \( \{N(t) : a \leq t \leq \tau\} \) as the assessment process, and we refer to \( \{Y(t) : 0 \leq t \leq \tau\} \) as the outcome process. We also use the shorthand notation \( L \) for the outcome process.

Let \( A(t) := N(t) - \lim_{s \to t^-} N(s) \) be the indicator that the individual has an assessment at time \( t \), and, for a given \( \epsilon > 0 \), let \( A[t, t + \epsilon) := \lim_{s \to t^-} N(t + s) - \lim_{s \to t^-} N(s) \) be the indicator that they have an assessment in the time window \( [t, t + \epsilon) \). The observed data for a random individual is \( O := \{N(t) : a \leq t \leq \tau\} \cup \{Y(t) : A(t) = 1, 0 \leq t \leq \tau\} \). We let \( \bar{O}(t) \) denote their observed data up to, but not including, time \( t \); we refer to this as their observed past.

Information about an individual’s rate of assessment at time \( t \) is captured by the conditional intensity function of their assessment process. Our approach will use both the intensity function given the participant’s observed past, \( \lambda(t, \bar{O}(t)) := \lim_{\epsilon \to 0^+} \{P(A[t, t + \epsilon) = 1|\bar{O}(t))/\epsilon\} \), and the intensity function given their full outcome process \( L \) and their observed past, \( \rho(t, \bar{O}(t), L) := \lim_{\epsilon \to 0^+} \{P(A[t, t + \epsilon) = 1|\bar{O}(t), L)/\epsilon\} \).

2.2 Assessment at random (AAR) assumption

In order to learn about the mean outcome at a targeted time from a trial with irregular assessments, some assumption about the relationship between assessment times and outcomes is needed. In practice, it is commonly assumed that the assessment times are completely independent of outcomes; however, in many trials, such a strong assumption may not be tenable. The weaker assessment at random (AAR) assumption is often more realistic. Under AAR, for each time \( t \) and within each stratum of the observed past \( \bar{O}(t) \), the distribution of \( Y(t) \) is assumed to be the same among participants who were, and who were not, assessed at time \( t \). That is, \( dF(y(t)|A(t) = 0, \bar{O}(t)) = dF(y(t)|A(t) = 1, \bar{O}(t)) \). Importantly, AAR is not testable, since the observed data provide no information about \( dF(y(t)|A(t) = 0, \bar{O}(t)) \), the conditional distribution of outcomes at time \( t \) among participants who were not assessed at that time.
2.3 Review of existing methods

Previous methods for analyzing trials with irregular assessment times have mainly used one of two approaches: joint modeling, or inverse weighting. We review each below.

2.3.1 Joint modeling approach

Methods in this category rely on strong assumptions on the joint distribution of the assessment process and the outcome process. These include untestable assumptions, such as conditional independence of the two processes given baseline covariates and latent variables. For example, Lin and Ying [20] posited a semiparametric model for the outcomes and a semiparametric model for the assessment times, and assumed that assessments and outcomes were independent after conditioning on baseline covariates common to both models. Lipsitz et al. [22] assumed a parametric Gaussian model for the outcome process. They assumed that the assessment process was ancillary, with the likelihood factorizing such that parameters of the outcome process model could be estimated without having to account for the distribution of assessment times. Liang et al [19] used a mixed effects model for the mean of the outcome and a frailty model for the assessment times. They assumed that the assessment and outcome processes were independent given \( Z_1 \) and \( Z_2 \), and assumed that the assessment and outcome processes were independent given \( Z_1 \), \( Z_2 \), and baseline covariates. Further work has allowed shared random effects to impact the outcome model through an unspecified time-varying function [7]; and relaxed assumptions on \( Z_1 \) and \( Z_2 \), while also allowing for a dependent terminal event [34]. Additionally, Pullenayegum [26] developed a method combining the joint modeling approach with multiple oututation. This allows dependence between the outcomes and the assessment times due to time-varying covariates, as well as baseline covariates and correlated random effects.

2.3.2 Inverse weighting approach

Robins, Rotnitzky, and Zhao [30] proposed inverse probability weighted estimators for longitudinal data with missing outcomes. Their method extended generalized estimating equations (GEE’s) to account for settings where the observed past may impact the probability of missing an assessment. For trials with irregular assessment times, Lin, Scharfstein, and Rosenheck [21] introduced a method of inverse-intensity weighting: working under the AAR assumption, they weighted a participant’s outcome \( Y(t) \) by the inverse of the intensity function given their observed past, for each time \( t \) at which the participant had an assessment. A participant who was assessed at time \( t \), but who was unlikely to have been assessed at that time based on their observed past, would have a low intensity \( \lambda(t, O(t)) \). Such a participant can be thought of as representing a large number of like participants whose outcomes at time \( t \) were not observed, and weighting \( Y(t) \) by \( 1/\lambda(t, O(t)) \) accounts for this. Lin, Scharfstein, and Rosenheck [21] incorporated inverse-intensity weighting into an estimating equation approach, and showed that the resulting estimators were consistent and asymptotically normal for the regression parameters of a structural model for \( E(Y(t)) \). Bůžková and Lumley [6] adapted this method to allow an intensity function with discontinuities, as could occur when many participants are assessed exactly at the targeted...
assessment times. Pullenayegum and Feldman [27] combined inverse-intensity weighting with a model for the increments of the outcome process to form a doubly robust estimator, while Sun et al. [35] extended inverse-intensity weighting to the setting of quantile regression.

Building on the inverse-intensity-weighting approach, our method uses augmented inverse-intensity weighting, which allows for flexible modeling of the intensity function and makes more complete use of the data compared to inverse-intensity weighting alone (see Section 4). Another important difference to previous methods is that the weighting function that we use accounts for informative assessments where AAR does not hold.

2.4 The Asthma Research for the Community trial

The Asthma Research for the Community (ARC) study [4] was a pragmatic randomized trial of 301 low-income participants with uncontrolled asthma. Participants randomized to the control group received usual care plus access to and training in a web-based portal designed to improve communication between participants and their healthcare providers. Participants randomized to the intervention group received home visits by community health workers, in addition to usual care and portal training. At these visits, the community health workers created plans of action for asthma care and helped participants use the portal to schedule appointments and send messages to their doctors. The primary outcome was Asthma Control score, reflecting symptoms over the week prior to assessment on a scale from 0 (completely controlled) to 6 (extremely uncontrolled). The study protocol called for outcome data to be collected at 3, 6, 9, and 12 months after randomization. However, research coordinators were often unable to schedule data collection appointments until substantially later than these targeted times. Figure 1 shows the distribution of the actual times at which assessments took place.

3 Models and sensitivity analysis framework

All of our assumptions are treatment arm-specific. For ease of notation, we suppress dependence on treatment group.

3.1 Sensitivity analysis framework

In a global sensitivity analysis, inferences for the target parameters of interest are made under a broad family of assumptions (anchored around a reasonable benchmark assumption) that is considered plausible by subject-matter experts. The sensitivity analysis allows researchers to judge whether inferences are robust to deviations from the benchmark, or whether an analysis relying on the (untestable) benchmark assumption should be viewed with caution. We consider a family of assumptions anchored around the AAR assumption. Members of this family are indexed by a sensitivity parameter $\alpha$. For a fixed value of the observed past $O(t)$, let Subgroup 1 (Subgroup 0) be the group with this observed
Figure 1: Assessment times in the ARC study. The four panels show the distributions of the actual times of participants’ first, second, third, and fourth post-baseline assessments. The protocol called for assessments at 3, 6, 9, and 12 months, but there was substantial spread in the actual times of assessment around these targeted times.
Figure 2: Illustration of the tilt assumption in the context of the ARC trial. Here we consider the distribution of the Asthma Control outcome $Y(t)$ at $t = 6$ months among participants with a certain observed past. The distribution of $Y(6)$ among participants who were assessed at 6 months is shown in the left panel. Under the AAR assumption ($\alpha = 0$), the distribution of $Y(6)$ for participants who had the same observed past but were not assessed at 6 months would be the same; under a positive (negative) value of $\alpha$, the distribution for non-assessed participants would be tilted with more weight on higher (lower) values of $Y(6)$, as shown in the right panel. Shown are smoothed depictions of the probability mass functions.

A value of $\alpha = 0$ corresponds to no tilt, which is the AAR assumption. Under a negative (positive) value of $\alpha$, the outcome distribution for Subgroup 0 is tilted to the left (right) relative to the distribution for Subgroup 1, with smaller (larger) values of $Y(t)$ receiving greater weight. To conduct the sensitivity analysis, target parameters are estimated under a range of values of $\alpha$. Figure 2 illustrates tilting for values of $\alpha = -0.6, -0.3, 0, 0.3$, and 0.6 in the context of the ARC trial. In what follows, we will suppress the dependence of quantities on $\alpha$. 

Assumption 1. (Tilt assumption) For specified $\alpha$ and for all $t \in [a, b],$

$$
\frac{dF(Y(t) = y|A(t) = 0, \bar{O}(t))}{\text{Subgroup 0 distribution}} = \frac{dF(Y(t) = y|A(t) = 1, \bar{O}(t))}{\text{Subgroup 1 distribution}} \cdot \frac{\exp\{\alpha y\}}{E[\exp\{\alpha Y(t)\}|A(t) = 1, \bar{O}(t)]}.
$$

A value of $\alpha = 0$ corresponds to no tilt, which is the AAR assumption. Under a negative (positive) value of $\alpha$, the outcome distribution for Subgroup 0 is tilted to the left (right) relative to the distribution for Subgroup 1, with smaller (larger) values of $Y(t)$ receiving greater weight. To conduct the sensitivity analysis, target parameters are estimated under a range of values of $\alpha$. Figure 2 illustrates tilting for values of $\alpha = -0.6, -0.3, 0, 0.3$, and 0.6 in the context of the ARC trial. In what follows, we will suppress the dependence of quantities on $\alpha$. 

7
We additionally assume that the intensity function $\rho(t, \bar{O}(t), L)$ is not impacted by unobserved variables other than the current outcome $Y(t)$, or by any future outcome variables occurring after time $t$:

**Assumption 2.** (Intensity function assumption) The intensity function given a participant’s observed past $\bar{O}(t)$ and their full outcome process $L = \{Y(t) : 0 \leq t \leq \tau\}$ depends only on $t$, $Y(t)$, and $\bar{O}(t)$. That is:

$$
\rho(t, \bar{O}(t), L) := \lim_{\epsilon \to 0^+} \frac{P(A[t, t + \epsilon) = 1|\bar{O}(t), L)}{\epsilon}
= \lim_{\epsilon \to 0^+} \frac{P(A[t, t + \epsilon) = 1|\bar{O}(t), Y(t))}{\epsilon} =: \rho(t, Y(t), \bar{O}(t)).
$$

3.2 Structural assumption and target estimand

Since few participants are assessed at exactly the targeted assessment times, we assume a form for the curve of means $E[Y(t)]$, $a \leq t \leq b$, that allows us to borrow information from participants assessed at other times:

**Assumption 3.** (Structural assumption for the curve of means) Let $B(t)$ be a specified spline basis, let $p$ be its dimension, and let $s(\cdot)$ be a specified invertible link function. We assume that $E[Y(t)] = s(\beta' B(t)), a \leq t \leq b$, for some $\beta \in \mathbb{R}^p$.

Rather than making Assumption 3, one could define $\beta$ by taking $\beta' B(t)$ to be the projection of $s^{-1}(E[Y(t)])$ onto the space spanned by $B(t)$. For simplicity, we will assume that Assumption 3 does hold, and we take $\beta$ to be our target parameter. For a known value of $\alpha$, for each time $t$, $E[Y(t)]$ is identified from the observed data by Assumption 1.

$$
E[Y(t)] = \int_w \int_y y(t) \left\{ dF(y(t)|A(t) = 0, \bar{O}(t) = w) P(A(t) = 0|\bar{O}(t) = w) +
\right.
\left. dF(y(t)|A(t) = 1, \bar{O}(t) = w) P(A(t) = 1|\bar{O}(t) = w) \right\} dF(w)
= \int_w \int_y \frac{y(t) \exp\{\alpha y(t)\}}{E[\exp\{\alpha Y(t)\}|A(t) = 1, \bar{O}(t) = w]} dF(y(t)|A(t) = 1, \bar{O}(t) = w) dF(w). \quad (1)
$$

Here we used the fact that, in continuous time, $P(A(t) = 1|\bar{O}(t) = w) = 0$ for all $w$. Therefore $\beta$ is identified, since $\beta = \int_{t=a}^{b} V^{-1}B(t)B(t)' \beta dt = \int_{t=a}^{b} V^{-1}B(t)s^{-1}(E[Y(t)]) dt$ by Assumption 3, where $V := \int_{t=a}^{b} B(t)B(t)' dt$.

4 Inference under a fixed value of $\alpha$

In this section we suppose that Assumptions 1-3 hold for a given known value of $\alpha$. Our inference procedure is based on an influence function for $\beta$ that we derive in Section
4.1 We use this influence function to construct a semiparametric estimator $\hat{\beta}$, and to derive its large-sample distribution. Estimation of $\beta$ will require modeling certain features of the observed data distribution. Importantly, our estimation procedure allows us to use flexible, semiparametric models while still yielding a root-$n$-consistent estimator $\hat{\beta}$. We assume that we observe $n$ i.i.d copies of the observed data $O$, and we use the subscript $i$ to refer to data for the $i$th individual.

4.1 Influence function

Let $P$ denote the true distribution of the observed data. We first consider the case where the link function $s(\cdot)$ in Assumption 3 is any invertible function, then specialize to the case where $s(\cdot)$ is the identity link.

**Theorem 1.** Let $V(\beta) := \int_{t=a}^{b} \left\{ \left( \frac{\partial}{\partial \beta} s(\beta' B(t)) \right) \left( \frac{\partial}{\partial \beta} s(\beta' B(t)) \right) \right\} dt$, and let $W(t; \beta) := V(\beta)^{-1} \frac{\partial}{\partial \beta} s(\beta' B(t))$. Then an influence function for $\beta$ is given by:

$$
\varphi(O; P) = \int_{t=a}^{b} W(t; \beta) \left\{ \frac{1}{\rho(t, Y(t), \bar{O}(t))} \left( Y(t) - E[Y(t)|\bar{O}(t)] \right) \right\} dN(t) + \int_{a}^{b} W(t; \beta) \left( E[Y(t)|\bar{O}(t)] - s(\beta' B(t)) \right) dt.
$$

(2)

**Corollary 1.** When $s(\cdot)$ is the identity link, an influence function for $\beta$ is $\varphi(O; P) = m(O; P) - \beta$, where $m(O; P) :=$

$$
\int_{t=a}^{b} \left\{ V^{-1}(t) \frac{Y(t) - E[Y(t)|\bar{O}(t)]}{\rho(t, Y(t), \bar{O}(t))} \right\} dN(t) + \int_{t=a}^{b} \left\{ V^{-1}(t) E[Y(t)|\bar{O}(t)] \right\} dt
$$

(3)

and $V := \int_{t=a}^{b} B(t) B(t)' dt$.

We prove Theorem 1 using the semiparametric theory for missing data presented in [33, 37]. The proof is given in Section B of the Supplementary materials; Corollary 1 then follows immediately.

Throughout the rest of the paper, we take $s(\cdot)$ to be the identity link. The first term in $m(O; P)$ uses inverse intensity weighting by $\rho(t, Y(t), \bar{O}(t))$; the second term is an augmentation term that allows for more complete use of the data compared to inverse intensity weighting alone.

4.2 Point estimation

The influence function $\varphi(O; P)$ is a mean-zero function of the observed data $O$, the target parameter $\beta$, $\rho(t, Y(t), \bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$. With estimators of $\rho(t, Y(t), \bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$, the influence function can be used as an estimating function for $\beta$. Following the philosophy of [31], our estimators of $\rho(t, Y(t), \bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$ involve models that place no restrictions on the sensitivity parameter $\alpha$. For this, we leverage the following results:
Proposition 1. Under Assumptions $[7]$ and $[2]$:

(a) $\rho(t,Y(t),\bar{O}(t)) = \lambda(t,\bar{O}(t)) \exp\{-\alpha Y(t)\} E\left[\exp\{\alpha Y(t)\} \mid A(t) = 1, \bar{O}(t)\right]$

(b) $E[Y(t)\mid \bar{O}(t)] = \frac{E[Y(t)\exp\{\alpha Y(t)\} \mid A(t) = 1, \bar{O}(t)]}{E[\exp\{\alpha Y(t)\} \mid A(t) = 1, \bar{O}(t)]}.$

The proof of Proposition $[\S]$ (a) is given in Section A of the Supplementary Materials. Part (b) holds by a computation similar to the one used in Equation (1).

By Proposition $[\S]$ modeling $\lambda(t,\bar{O}(t))$ and the conditional distribution of observed outcomes $dF(y(t)|A(t) = 1, \bar{O}(t))$—which places no restrictions on $\alpha$—allows us to obtain estimates for $\rho(t,Y(t),\bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$. We assume that the models for $\lambda(t,\bar{O}(t))$ and $dF(y(t)|A(t) = 1, \bar{O}(t))$ are both correctly specified, and that they are chosen in such a way that the estimators $\hat{\rho}(t,Y(t),\bar{O}(t))$ and $\hat{E}[Y(t)|\bar{O}(t)]$ converge to the true values $\rho(t,Y(t),\bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$ at fast enough rates, as discussed in Section 4.4. In general, using models that are too flexible can result in an estimator for the target parameter that converges to the truth at slower than root-$n$ rates and for which valid confidence intervals cannot be readily obtained $[24]$. A key advantage of using an estimator based on $\varphi(O,P)$ is that $\rho(t,Y(t),\bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$ can be estimated flexibly, at slower than $n^{-1/2}$ rates, while still yielding a root-$n$-consistent estimator for $\beta$ (see Section 4.4).

Using the form of the influence function and estimators of $\rho(t,Y(t),\bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$, the resulting estimator for $\beta$ is:

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{t_k \in S_i} V^{-1} B(t_k) \left( Y_i(t_k) - \hat{E}[Y(t_k)|\bar{O}(t_k)_i] \right) \rho(t_k,Y_i(t_k),\bar{O}(t_k)_i) \right\} + \int_{t=a}^{b} V^{-1} B(t) \hat{E}[Y(t)|\bar{O}(t)] dt,$$

where $S_i$ denotes the set of participant $i$’s assessment times that occur in the interval $[a,b]$.

4.3 Large sample distribution and confidence intervals

Here we derive the large-sample distribution of $\hat{\beta}$ and give confidence intervals for $\beta$ and the target means $\mu_t = E[Y(t)]$, under assumptions on rates for estimating $\rho(t,Y(t),\bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$. Let $\hat{P}$ be an estimate of the distribution $P$ that uses the empirical distribution of $\bar{O}(t)$ and and estimates of $\rho(t,Y(t),\bar{O}(t))$ and $E[Y(t)|\bar{O}(t)]$ based on models for $\lambda(t,\bar{O}(t))$ and $dF(y(t)|A(t) = 1, \bar{O}(t))$ as described in Section 4.2. Following $[15]$,
\(\sqrt{n}(\hat{\beta} - \beta)\) can be expanded as the sum of three terms:

\[
\sqrt{n}(\hat{\beta} - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(O_i; P) + \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\varphi(O_i; \hat{P}) - \varphi(O_i; P)\} - E[\varphi(O; \hat{P}) - \varphi(O; P)] \right) + \sqrt{n} \text{Rem}(\hat{P}, P).
\]

(4)

The second term in (4) is an empirical process term that, under Donsker conditions \[38\], becomes asymptotically negligible. Provided that these conditions are satisfied, the form of the remainder term \(\text{Rem}(\hat{P}, P)\), which we compute below, determines how flexibly \(\rho(t, Y(t), \hat{O}(t))\) and \(E[Y(t)|\hat{O}(t)]\) can be estimated in order for \(\hat{\beta}\) to converge to \(\beta\) at root-\(n\) rates.

**Proposition 2.** The remainder term \(\text{Rem}(\hat{P}, P)\) appearing in (4) is given by:

\[
E\left[ \int_{t=a}^{b} V^{-1} B(t) \left( \frac{\rho(t, Y(t), \hat{O}(t))}{\hat{\rho}(t, Y(t), \hat{O}(t))} - 1 \right) \left( E[Y(t)|\hat{O}(t)] - \hat{E}[Y(t)|\hat{O}(t)] \right) \right] dt.
\]

(5)

The proof of Proposition 2 is given in Section C of the Supplementary Materials.

Importantly, from Proposition 2, \(\hat{\beta}\) can be seen to enjoy product bias: the error in \(\hat{\beta}\) depends on the product of the error in \(\hat{\rho}(t, Y(t), \hat{O}(t))\) and the error in \(\hat{E}[Y(t)|\hat{O}(t)]\). This product decreases at a faster rate than either of the individual errors. Therefore, \(\hat{\beta}\) can achieve a fast, root-\(n\) rate of convergence even if \(\hat{\rho}(t, Y(t), \hat{O}(t))\) and \(\hat{E}[Y(t)|\hat{O}(t)]\) each converge at rates slower than root-\(n\). We make this precise in Section 4.4, where we give sufficient conditions under which \(\sqrt{n} \text{Rem}(\hat{P}, P)\) is \(o_p(1)\). Under these sufficient conditions—along with the Donsker conditions by which the empirical process term is \(o_p(1)\)—\(\hat{\beta}\) has influence function \(\varphi(O; P)\) and is asymptotically normal with asymptotic variance equal to the variance of \(\varphi(O; P)\). Then the variance of \(\hat{\beta}\) is approximately \(\frac{1}{n} E[\varphi(O; P)\varphi(O; P)']\), and a consistent variance estimator is given by:

\[
\hat{\text{Var}}(\hat{\beta}) = \frac{1}{n^2} \sum_{i=1}^{n} \left[ m(O_i; \hat{P}) - \hat{\beta} \right] \left[ m(O_i; \hat{P}) - \hat{\beta} \right]'.
\]

where \(m(O; P)\) is given in Corollary 1. A target mean \(\mu_t = E[Y(t)] = \beta' B(t)\) can be estimated as \(\hat{\mu}_t = \hat{\beta}' B(t)\) with influence function-based variance estimator:

\[
\hat{\text{Var}}(\hat{\mu}_t) = B(t)' \hat{\text{Var}}(\hat{\beta}) B(t).
\]

(6)
The variance estimator in [6] can be used to construct a Wald confidence interval for \( \mu_t \), or it can be used within a bootstrap procedure to form a bootstrap-\( t \) confidence interval. Consider \( B \) bootstrap resamples of the data. For each \( b = 1, \ldots, B \), let \( \hat{\mu}_t^{*b} \) and \( \hat{SE}(\hat{\mu}_t^{*b}) \) be the estimate and the influence function-based estimated standard error from the \( b \)th resample, and set \( T^{*b} = (\hat{\mu}_t^{*b} - \hat{\mu}_t)/\hat{SE}(\hat{\mu}_t^{*b}) \), where \( \hat{\mu}_t \) is the estimate based on the original data. Take \( c^* \) to be the 0.95 quantile of \( \{|T^{*1}|, \ldots, |T^{*B}|\} \). Then, with \( \hat{SE}(\hat{\mu}_t) \) the influence function-based estimated standard error based on the original data, \( (\hat{\mu}_t - c^* \hat{SE}(\hat{\mu}_t), \hat{\mu}_t + c^* \hat{SE}(\hat{\mu}_t) ) \) is a 95% bootstrap-\( t \) confidence interval for \( \mu_t \); confidence intervals for the treatment effect can be found similarly by exploiting independence between treatment arms. Bootstrap-\( t \) confidence intervals are second-order accurate [16; 13] and therefore tend to have improved coverage compared to Wald intervals based on [6].

### 4.4 Estimation details and rates of convergence

Here we discuss the modeling choices for \( \lambda(t, \bar{O}(t)) \) and \( dF(y(t)|A(t) = 1, \bar{O}(t)) \) that we used in our analysis of the ARC data, and we show sufficient conditions under which the remainder term \( \sqrt{n Rem(P, \bar{P})} \) is asymptotically negligible.

We modeled \( \lambda(t, \bar{O}(t)) \) using a stratified Andersen-Gill model [1]:

\[
\lambda(t, \bar{O}(t)) = \lambda_{0,k}(t) \exp \{ \gamma Z(t) \} D_k(t) \tag{7}
\]

with stratification variable \( k = N(t-) \) (i.e., the number of assessments that the participant had prior to time \( t \)). Here \( \lambda_{0,k}(t) \) is an unspecified baseline intensity function for the \( k \)th assessment, \( Z(t) \) is a specified function of \( \bar{O}(t) \), \( \gamma \) is a vector of unknown parameters, and \( D_k(t) \) is the indicator that the participant is at risk for having the \( k \)th assessment at time \( t \). We estimated \( \gamma \) and the cumulative baseline intensity functions \( \Lambda_{0,k}(t) = \int_{s=0}^{t} \lambda_{0,k}(s)ds \), \( k = 1, \ldots, J \) (where \( J \) is the maximum possible number of post-baseline assessments), using the partial likelihood estimator \( \hat{\gamma} \) of Cox [11; 12] and the Breslow estimator [5] \( \hat{\Lambda}_{0,k}(t) \), respectively. Andersen and Gill [1] derived the asymptotic distribution of \( \hat{\gamma} \) and \( \hat{\Lambda}_{0}(t) \) in the univariate \( (J = 1) \) setting. Andersen and Borgan [3] generalized their results to multivariate counting processes with \( J > 1 \), such as our stratified model [7]. (Also see [9] for more discussion of stratified models.) We then used kernel smoothing of \( \hat{\Lambda}_{0,k}(t) \) to estimate the baseline intensity \( \lambda_{0,k}(t) \) as \( \hat{\lambda}_{0,k}(t) = \frac{1}{h} \sum_j K \left( \frac{t - T_j}{h} \right) d\hat{\Lambda}_{0,k}(T_j) \), where \( h \) is a bandwidth, \( K(\cdot) \) is a choice of kernel, \( T_j \) is a time at which one or more participants had their \( k \)th assessment, and \( d\hat{\Lambda}_{0,k}(T_j) \) is the size of the jump in \( \hat{\Lambda}_{0,k}(t) \) at \( T_j \). Ramlau-Hansen [29] proposed this approach for kernel-smoothing the Nelson-Aalen estimator for a univariate counting process, and showed the large-sample distribution of the resulting estimator. Andersen and Borgan [3], Andersen, Borgan, Gill, and Kieding [2], and Wells [40] extended his results to multivariate versions of the Andersen-Gill model such as (7).

Let \( \theta \) denote the parameters of the model for \( dF(y(t)|A(t) = 1, \bar{O}(t)) \). We modeled \( dF(y(t)|A(t) = 1, \bar{O}(t); \theta) \) using a fully parametric model. The asymptotic behavior of the remainder term under these models is shown by the following theorem.
Theorem 2. Suppose that the following conditions hold:

(i) The models for \( \lambda(t, \bar{O}(t)) \) and \( dF(y(t)|A(t) = 1, \bar{O}(t); \theta) \) are correctly specified, and Assumptions 1, 2, and 3 hold.

(ii) \( \lambda(t, \bar{O}(t)) \) is modeled as in (7), \( \gamma \) and \( \lambda_{0,k}(t) \), \( k = 1, \ldots, J \), are estimated as described above, and the bandwidth \( h = h_n \) used for \( \lambda_{0,k}(t) \) is chosen using a procedure such that \( \lim_{n \to \infty} nh_n^5 = d \), for some constant \( d \) with \( 0 < d < \infty \).

(iii) The model for \( dF(y(t)|A(t) = 1, \bar{O}(t); \theta) \) and the estimator \( \hat{\theta} \) are such that

\[
E \left[ \int_{t=a}^{b} \left\{ dF(y(t)|A(t) = 1, \bar{O}(t), \hat{\theta}) - dF(y(t)|A(t) = 1, \bar{O}(t), \theta) \right\}^2 dt \right] = o_p(n^{-1/2}).
\]

(iv) There is some value \( c > 0 \) such that \( \rho(t, Y(t), \bar{O}(t)) > c \) for all \( t \) and all values of \( Y(t) \) and \( \bar{O}(t) \).

Then the remainder term \( \text{Rem}(\hat{P}, P) \) in (5) is \( o_P(n^{-1/2}) \).

In particular, if a fully parametric model is used for \( dF(y(t)|A(t) = 1, \bar{O}(t); \theta) \), then the expectation in (iii) is \( O_P(n^{-1}) \), and therefore (iii) is satisfied; and this also suggests that replacing a fully parametric model with a more flexible model may also be possible. The proof of Theorem 2 is given in Section C of the Supplementary Materials.

5 Selection of a range of sensitivity parameter values

To conduct our sensitivity analysis, we estimate the target parameters under a family of different assumptions, in which the sensitivity parameter \( \alpha \) varies away from the benchmark assumption of \( \alpha = 0 \). Domain expertise should be used to decide on a range of \( \alpha \) values. How best to do this is a question that has been explored in the context of sensitivity analysis for unmeasured confounding in observational studies; see for example [8], [14], and [39]. Cinelli and Hazlett [8] have noted that “perhaps [the] most fundamental obstacle to the use of sensitivity analysis is the difficulty in connecting the formal results to the researcher’s substantive understanding about the object under study,” and that the “bounding procedure we should use depends on which . . . quantities the investigator prefers and can most soundly reason about in their own research.” The authors of [8, 14, 39] have proposed ways for using the strength of the impact of a key covariate or group of covariates \( X_j \) given the remaining covariates \( X_{-j} \) to obtain a posited bound on the strength of unmeasured confounders, and hence obtain bounds for the sensitivity parameters. However, this may not adapt well to our setting: the impact of any group of variables \( Z(t) \) in the observed past on assessment at time \( t \) may actually be weaker—by an amount that would not be known—than the impact of \( Y(t) \) on assessment at time \( t \) (after adjusting for the remaining variables in the observed past in each case).

Instead, we query domain experts for extreme values \( \mu_{\min} \) and \( \mu_{\max} \) such that, in their judgment, a mean outcome \( E[Y(t)] \) outside of the bounds \( (\mu_{\min}, \mu_{\max}) \) at any time \( t \) would
be implausible. We then treat any \( \alpha \) under which \( E[Y(t)] \) falls outside of \((\mu_{\text{min}}, \mu_{\text{max}})\) for some \( t \) as implausible and exclude such values from our sensitivity analysis. We do this separately for each treatment arm. This approach is aligned with Cinelli and Hazlett’s recommendation, as it is based on the treatment arm-specific mean outcome, a quantity about which subject matter experts can provide direct guidance.

6 Data analysis: ARC trial

Here, we use the proposed methodology to estimate the treatment-specific curves for mean Asthma Control score over the period from 60 to 460 days after randomization. The target assessment times of interest are 90, 180, 270, and 360 days. For each treatment, we assumed \( E[Y(t)] = \beta' B(t), 60 \leq t \leq 460 \), where \( B(t) \) is a basis of cubic B-splines with one interior knot at \( t = 260 \) days. We modeled the intensity function \( \lambda(t, \bar{O}(t)) \) using a stratified Andersen-Gill model \cite{1} as in (7), where the stratification variable was the number of previous assessments, and where the number of strata was \( J = 4 \). We took the predictor \( Z(t) \) to be the outcome at the previous assessment. We fit this model in R\cite{28} using the survival package \cite{36}. We then used kernel smoothing with an Epanechnikov kernel and a bandwidth of 30 days to estimate the baseline intensity function. For the outcome regression model, we used negative binomial regression, using current time, time of previous assessment, and outcome at the previous assessment as predictors.

We initially considered a range of \(-0.6 \leq \alpha \leq 0.6\) for the sensitivity parameter in each arm. Figure 3 shows estimates of the curves \( E[Y(t)], 60 \leq t \leq 460 \), for each arm (upper panels), and estimates and 95% confidence intervals for the means at the targeted assessment times (lower panels). Confidence intervals are bootstrap-t confidence intervals based on \( B = 500 \) bootstrap resamples. Our clinical collaborator considered, for each treatment arm, a mean Asthma Control score of 3.0 or higher at any time, or a mean of 1.2 or lower at any time, as extreme. These bounds are shown in Figures 3. Based on this, \( \alpha = 0.6 \) is outside the range that should be considered for the intervention arm, while both \( \alpha = 0.3 \) and \( \alpha = 0.6 \) are implausible in the control arm. We therefore considered the ranges \(-0.6 \leq \alpha \leq 0.55\) and \(-0.6 \leq \alpha \leq 0.25\) for the intervention and control arms respectively. Next we considered the difference of the mean outcome in the intervention arm minus the mean in the control arm at the targeted assessment times. Under the AAR assumption, the estimate and 95% bootstrap-t confidence interval for the treatment effect at 6 months are \(-0.39 (-0.91, 0.13)\). The point estimate under the AAR assumption is consistent with the home visits intervention reducing (improving) mean Asthma Control at six months; however, the evidence is not strong enough to conclude that there is a nonzero treatment effect. For the treatment effect at 12 months, the estimate and confidence interval under the AAR assumption are \(0.05 (-0.28, 0.38)\). Figure 4 shows the sensitivity analysis for these two treatment effects as we deviate from the AAR assumption in each treatment arm, within the range of \( \alpha \) values given above. In panels (b) and (d), the regions in white correspond to sensitivity parameter values under which the confidence interval contains zero. The green regions in the lower left of panels (b) and (d) correspond to values of \( \alpha \) under which the confidence intervals are entirely negative, i.e. for which there would be
Figure 3: Mean outcomes in the ARC trial under a range of sensitivity parameter values. Here we show inference for the mean Asthma Control score in the intervention (PT+HV) and control (PT) arm, under values of $\alpha = -0.6, -0.3, 0, 0.3, 0.6$. The upper panels show the estimated curves of mean Asthma Control scores at time $t$, for $t = 60$ to $460$ days after randomization. The lower panels show estimates and 95% bootstrap-$t$ confidence intervals for the mean at each of the 4 target times. For each arm, only those values of $\alpha$ under which the mean falls between the dotted lines at $\mu_{\text{min}} = 1.2$ and $\mu_{\text{max}} = 3.0$ at all times are considered plausible based on subject-matter expertise.
evidence that the intervention reduces the mean Asthma Control score; while the orange regions in the upper left correspond to values for which there would be evidence that the intervention raises the mean Asthma Control score.

7 Simulations

We assessed the finite-sample performance of our estimators in a realistic simulation study based on the ARC data. Simulations were performed separately by treatment arm. To generate simulated data for participant \( i \), we first drew their baseline outcome \( Y_i(0) \) from the empirical distribution of baseline outcomes in the ARC data. We then iterated between generating times of subsequent assessments and outcomes at those assessment times. Given their \( k \)th assessment time \( T_{ik} \) and outcome \( Y_i(T_{ik}) \), we generated participant \( i \)’s \((k + 1)\)st assessment time \( T_{i,k+1} \) using Ogata’s Thinning Algorithm \(^{[25]}\) as follows:

1. Set \( \lambda^* \) to be a value that is greater than or equal to sup \{ \( \hat{\lambda}_{0,k}(t) \) exp \{ \( \hat{\gamma} Y_i(T_{ik}) \) \} : \( t \in [0, \tau] \} \), where \( \hat{\lambda}_{0,k}(t) \) and \( \hat{\gamma} \) are the estimates from the stratified Andersen-Gill model that we fit on the ARC data.

2. Take the start time to be \( T_{ik} \).

3. Draw a potential gap time \( s^* \sim Exp(\lambda^*) \), and set \( t^* := \text{start time} + s^* \). Accept the candidate assessment time \( t^* \) with probability \( \hat{\lambda}_{0,k}(t^*) \) exp \{ \( \hat{\gamma} Y_i(T_{ik}) \) \} / \lambda^* \).

4. If \( t^* \) is accepted, set \( T_{i,k+1} := t^* \). If \( t^* \) is rejected, set the new start time to be \( t^* \) and return to step 3. Iterate until a time \( t^* \) is accepted, or until \( t^* > \tau \), in which case \( T_{ik} \) was participant \( i \)’s final assessment time.

After \( T_{i,k+1} \) was generated, we generated \( Y_i(T_{i,k+1}) \) using the model for \( dF(y(t)|A(t) = 1, \bar{O}(t)) \) that we fit on the ARC data. We obtained participant \( i \)’s predicted distribution \( d\bar{F}(y(T_{i,k+1})|A(T_{i,k+1}) = 1, \bar{O}_i(T_{i,k+1})) \), then drew an outcome \( Y_i(T_{i,k+1}) \) based on this predicted distribution.

Under this data-generating mechanism, the true curve of means \( E[Y(t)] \), \( 60 \leq t \leq 460 \), may not coincide exactly with a spline curve \( \beta' B(t) \). Therefore, we took as the true values the points at \( t = 90, 180, 270, 360 \) on the spline curve \( \gamma' B(t) \) closest to the true curve of means, in the sense of minimizing squared error loss (see equation \([10]\) in the Supplementary Materials). Here \( \gamma \) is the solution to \( \int_{t=a}^{b} E[Y(t)] B(t)' dt = \gamma' \int_{t=a}^{b} B(t) B(t)' dt \). To compute the true value of \( \gamma \) under each \( \alpha = -0.6, -0.3, 0, 0.3, 0.6 \), we simulated a large sample of \( N = 2,500,000 \) participants. For each time \( t \), we obtained the predicted value of \( E[Y(t)|\bar{O}_i(t)] \) using Proposition \([\Box]\)(b) and the model for \( dF(y(t)|A(t) = 1, \bar{O}(t)) \) that we fit on the ARC data, then used the approximation \( E[Y(t)] \approx \frac{1}{N} \sum_{i=1}^{N} \bar{E}[Y(t)|\bar{O}_i(t)] \). We approximated the integrals using intervals of one day.

We simulated data with a sample size of \( n = 200 \) for each treatment arm. We analyzed the simulated data using the true value of \( \alpha \) in each case. For each scenario, we evaluated our estimation approach in terms of empirical bias and confidence interval coverage probabilities across 500 simulations. We considered bootstrap-\( t \) intervals, bootstrap percentile
Figure 4: Sensitivity analysis for the ARC trial. Here we present inference for the difference of the mean Asthma Control score in the intervention (PT+HV) arm minus the mean in the control (PT) arm at 6 months and at 12 months, under a range of different sensitivity parameter values $\alpha$ for each arm. Panels (a) and (c) show the point estimates for these treatment effects under the varying values of $\alpha$. Panels (b) and (d) give information about the confidence intervals: each white region corresponds to sensitivity parameter values for which the confidence interval contains zero. In the green regions in the lower right of these plots, confidence intervals are entirely negative, and the values shown are the upper bound of the confidence interval. In the orange regions in the upper left of these plots, confidence interval are entirely positive, and the values shown in these regions are the lower bound of the confidence interval.
intervals, and Wald intervals, with the bootstrap intervals based on $B = 500$ bootstrap replications per simulated dataset. Results are shown in Tables 1 and 2. Bias was close to zero. As expected, better coverage was generally obtained with the bootstrap-$t$ confidence intervals than with the bootstrap percentile intervals or the Wald intervals. While the bootstrap-$t$ coverage was close to the nominal level in many cases, it undercovered in some situations, particularly those with large values of $\alpha$. Bias and coverage under $\alpha = 0.6$ both improved at a larger sample size of $N = 1,000$ (results not shown).

8 Discussion

In many trials, outcome data are collected at irregular assessment times. This may be by design, as with trials where data are collected at the time of care, which can vary across individuals. It may also occur in practice even if the trial protocol calls for all participants to be assessed at certain targeted times. Nonetheless, the irregularity of assessment times is largely ignored in applied work. Rather, visit windows are typically created and data are analyzed using missing data methods. There is no need to create artificial visit windows. Methods have been developed for analyzing trials with irregular and potentially assessment times. Like missing data methods, they rely on untestable assumptions and therefore sensitivity analysis is important. In this paper, we have developed a sensitivity analysis methodology anchored around the AAR assumption. Our tool will allow researchers to evaluate how robust their conclusions are to departures from this assumption. Importantly, if the sensitivity analysis shows that an intervention is beneficial under AAR, and also under a range of plausible deviations from this benchmark, then researchers will have a firmer basis for recommending the intervention.

Our inference procedure uses estimators for the intensity function and the conditional outcome distribution within an augmented inverse-intensity-weighted estimator. While we focused on the identity link in Assumption 3, the ideas presented here extend naturally to other invertible link functions. One question that arises concerns optimal bandwidth selection for kernel smoothing of intensity functions, in settings where the intensity function is not itself the final target of estimation. Additionally, here we used bootstrap-$t$ confidence intervals constructed using an influence function-based variance estimator. These confidence intervals undercovered at large values of the sensitivity parameter; future work could investigate methods for improving coverage, for example using the double bootstrap [15].

9 Funding

Shu Yang’s research is partially supported by NSF DMS 1811245, and by NIH 1R01AG066883 and 1R01ES031651. The ARC study was funded through a Patient-Centered Outcomes Research Institute (PCORI) award (AS-1307-05218).
| $\alpha$ | Parameter | True Value | Emp Mean | Bias | Boot-$t$ | Percentile | Wald |
|---------|-----------|------------|----------|------|----------|------------|------|
| -0.6    | $E[Y(3)]$ | 1.441      | 1.439    | 0.002| 0.936    | 0.906      | 0.892|
|         | $E[Y(6)]$ | 1.395      | 1.398    | 0.004| 0.958    | 0.942      | 0.942|
|         | $E[Y(9)]$ | 1.360      | 1.358    | 0.002| 0.962    | 0.946      | 0.946|
|         | $E[Y(12)]$| 1.332      | 1.337    | 0.006| 0.936    | 0.934      | 0.918|
| -0.3    | $E[Y(3)]$ | 1.755      | 1.742    | 0.013| 0.950    | 0.934      | 0.918|
|         | $E[Y(6)]$ | 1.692      | 1.690    | 0.002| 0.968    | 0.958      | 0.952|
|         | $E[Y(9)]$ | 1.646      | 1.640    | 0.006| 0.962    | 0.958      | 0.960|
|         | $E[Y(12)]$| 1.608      | 1.608    | 0.001| 0.936    | 0.934      | 0.922|
| 0       | $E[Y(3)]$ | 2.166      | 2.141    | 0.025| 0.944    | 0.926      | 0.918|
|         | $E[Y(6)]$ | 2.081      | 2.073    | 0.008| 0.962    | 0.944      | 0.948|
|         | $E[Y(9)]$ | 2.021      | 2.011    | 0.010| 0.958    | 0.960      | 0.956|
|         | $E[Y(12)]$| 1.971      | 1.966    | 0.005| 0.948    | 0.942      | 0.930|
| 0.3     | $E[Y(3)]$ | 2.675      | 2.641    | 0.034| 0.914    | 0.892      | 0.876|
|         | $E[Y(6)]$ | 2.568      | 2.555    | 0.013| 0.956    | 0.936      | 0.932|
|         | $E[Y(9)]$ | 2.492      | 2.482    | 0.011| 0.952    | 0.944      | 0.944|
|         | $E[Y(12)]$| 2.430      | 2.421    | 0.009| 0.940    | 0.928      | 0.916|
| 0.6     | $E[Y(3)]$ | 3.250      | 3.213    | 0.037| 0.898    | 0.874      | 0.854|
|         | $E[Y(6)]$ | 3.126      | 3.114    | 0.012| 0.936    | 0.922      | 0.918|
|         | $E[Y(9)]$ | 3.041      | 3.034    | 0.007| 0.946    | 0.930      | 0.930|
|         | $E[Y(12)]$| 2.970      | 2.957    | 0.013| 0.932    | 0.904      | 0.908|

Table 1: Simulation results: control arm. Shown are the true values of the four target parameters under each of five different data-generating mechanisms that mimic the control arm of the ARC data with $\alpha = -0.6, -0.3, 0, 0.3, 0.6$; the empirical mean and absolute value of the empirical bias of the estimators across 500 simulations; and the coverage of bootstrap-$t$, bootstrap percentile, and Wald confidence intervals.
Table 2: Simulation results: treatment arm. Shown are the true values of the four target parameters under each of five different data-generating mechanisms that mimic the treatment arm of the ARC data with $\alpha = -0.6, -0.3, 0, 0.3, 0.6$; the empirical mean and absolute value of the empirical bias of the estimators across 500 simulations; and the coverage of bootstrap-$t$, bootstrap percentile, and Wald confidence intervals.
References

[1] P. K. Andersen and R. D. Gill. Cox’s Regression Model for Counting Processes: A Large Sample Study. The Annals of Statistics, 10(4):1100 – 1120, 1982.

[2] Per Kragh Andersen, Ornulf Borgan, Richard D. Gill, and Niels Keiding. Statistical models based on counting processes. Springer-Verlag, New York, 1993.

[3] Per Kragh Andersen, Ørnulf Borgan, Nils Lid Hjort, Elja Arjas, Jon Stene, and Odd Aalen. Counting process models for life history data: A review [with discussion and reply]. Scandinavian Journal of Statistics, 12(2):97–158, 1985.

[4] Andrea J. Apter, A. Russell Localio, Knashawn H. Morales, Xiaoyan Han, Luzmercy Perez, Alyssa N. Mullen, Marisa Rogers, Heather Klusaritz, John T. Howell, Mary-ori N. Canales, and Tyra Bryant-Stephens. Home visits for uncontrolled asthma among low-income adults with patient portal access. Journal of Allergy and Clinical Immunology, 144(3):846–853.e11, 2019.

[5] Norman E. Breslow. Discussion on Professor Cox’s paper. Journal of the Royal Statistical Society: Series B (Methodological), 34(2):216–217, 1972.

[6] Petra Bůžková and Thomas Lumley. Longitudinal data analysis for generalized linear models with follow-up dependent on outcome-related variables. The Canadian Journal of Statistics / La Revue Canadienne de Statistique, 35(4):485–500, 2007.

[7] Na Cai, Wenbin Lu, and Hao Helen Zhang. Time-varying latent effect model for longitudinal data with informative observation times. Biometrics, 68(4):1093–1102, 2012.

[8] Carlos Cinelli and Chad Hazlett. Making sense of sensitivity: extending omitted variable bias. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(1):39–67, 2020.

[9] Richard J. Cook and Jerald Lawless. The statistical analysis of recurrent events. Springer, New York, 2007.

[10] National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. The National Academies Press, Washington, DC, 2010.

[11] D. R. Cox. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2):187–202, 1972.

[12] D.R. Cox. Partial likelihood. Biometrika, 62(2):269–276, 1975.

[13] Thomas J. DiCiccio and Bradley Efron. Bootstrap confidence intervals. Statistical Science, 11(3):189 – 228, 1996.
[14] Alexander M. Franks, Alexander D’Amour, and Avi Feller. Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*, 115(532):1730–1746, 2020.

[15] Peter Hall. On the Bootstrap and Confidence Intervals. *The Annals of Statistics*, 14(4):1431 – 1452, 1986.

[16] Peter Hall. Theoretical Comparison of Bootstrap Confidence Intervals. *The Annals of Statistics*, 16(3):927 – 953, 1988.

[17] Frank R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974.

[18] Edward H. Kennedy. Semiparametric theory and empirical processes in causal inference. In Hua He, Pan Wu, and Ding-Geng (Din) Chen, editors, *Statistical Causal Inferences and Their Applications in Public Health Research*, ICSA Book Series in Statistics, pages 141–167. Springer International Publishing, Cham, 2016.

[19] Yu Liang, Wenbin Lu, and Zhiliang Ying. Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, 65(2):377–384, 2009.

[20] D. Y Lin and Z Ying. Semiparametric and nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, 96(453):103–126, 2001.

[21] Haiqun Lin, Daniel O. Scharfstein, and Robert A. Rosenheck. Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):791–813, 2004.

[22] Stuart R. Lipsitz, Garrett M. Fitzmaurice, Joseph G. Ibrahim, Richard Gelber, and Steven Lipshultz. Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, 58(3):621–630, 2002.

[23] Iván Díaz Munoz and Mark J. van der Laan. Super learner based conditional density estimation with application to marginal structural models:. *The International Journal of Biostatistics*, 7(1), 2011.

[24] Ashley I Naimi, Alan E Mishler, and Edward H Kennedy. Challenges in obtaining valid causal effect estimates with machine learning algorithms. *American Journal of Epidemiology*, 07 2021. kwab201.

[25] Y. Ogata. On Lewis’ simulation method for point processes. *IEEE Transactions on Information Theory*, 27(1):23–31, 1981.

[26] Eleanor M. Pullenayegum. Multiple outputation for the analysis of longitudinal data subject to irregular observation. *Statistics in Medicine*, 35(11):1800–1818, 2016.

[27] Eleanor M. Pullenayegum and Brian M. Feldman. 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, 2013.
[28] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[29] Henrik Ramlau-Hansen. Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*, 11(2):453–466, 1983.

[30] James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429):106–121, 1995.

[31] Daniel Scharfstein, Aidan McDermott, Iván Díaz, Marco Carone, Nicola Lunardon, and Ibrahim Turköz. Global sensitivity analysis for repeated measures studies with informative drop-out: A semi-parametric approach. *Biometrics*, 74(1):207–219, 2018.

[32] Daniel Scharfstein, Aidan McDermott, William Olson, and Frank Wiegand. Global sensitivity analysis for repeated measures studies with informative dropout: A fully parametric approach. *Statistics in Biopharmaceutical Research*, 6(4):338–348, 2014.

[33] Daniel O. Scharfstein, Andrea Rotnitzky, and James M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.

[34] Liuquan Sun, Xinyuan Song, Jie Zhou, and Lei Liu. Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association*, 107(498):688–700, 2012.

[35] Xiaoyan Sun, Limin Peng, Amita Manatunga, and Michele Marcus. Quantile regression analysis of censored longitudinal data with irregular outcome-dependent follow-up. *Biometrics*, 72(1):64–73, 2016.

[36] Terry M Therneau. *A Package for Survival Analysis in R*, 2021. R package version 3.2-12.

[37] Anastasios A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, New York, 2006.

[38] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[39] Victor Veitch and Anisha Zaveri. Sense and sensitivity analysis: Simple post-hoc analysis of bias due to unobserved confounding. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 10999–11009. Curran Associates, Inc., 2020.

[40] Martin T. Wells. Nonparametric kernel estimation in counting processes with explanatory variables. *Biometrika*, 81(4):795–801, 12 1994.
SUPPLEMENTARY MATERIAL

A  Intensity functions

Let \( \{\mathcal{F}_t\}_{t=a} \) be the filtration defined by \( \mathcal{F}_t = (Y(0), \{Y(u) : A(u) = 1, a \leq u \leq t\}, \{N(u) : a \leq u \leq t\}) \). Here \( \mathcal{F}_t \) is the observed past up through time \( t \), and \( \mathcal{F}_{t-} = \hat{O}(t) \) is the observed past prior to, but not including, time \( t \). The intensity function \( \lambda(t, \hat{O}(t)) \) is the intensity for the counting process \( \{N(t)\}_{t=a} \) with respect to the filtration \( \{\mathcal{F}_t\}_{t=a} \). Additionally, let \( L = \{Y(t) : 0 \leq t \leq \tau\} \) and \( \{\mathcal{G}_t\}_{t=a} \) be the filtration defined by \( \mathcal{G}_t = (L, \{N(u) : a \leq u \leq t\}) = (L, \mathcal{F}_t) \). Under Assumption [2], \( \rho(t, Y(t), \hat{O}(t)) \) is the intensity function for \( \{N(t)\}_{t=a} \) with respect to the filtration \( \{\mathcal{G}_t\}_{t=a} \).

Proof of Proposition [1]:

Fix a time \( t \) and a value \( y(t) \), and for each \( \epsilon > 0 \) let:

\[
A_{\epsilon} := dF(y(t)\mid A[t, t+\epsilon] = 0, \hat{O}(t)) \quad \text{and} \quad B_{\epsilon} := dF(y(t)\mid A[t, t+\epsilon] = 1, \hat{O}(t)) \exp\{\alpha y\} E[\exp\{\alpha Y(t)\}\mid A[t, t+\epsilon] = 1, \hat{O}(t)].
\]

For each \( \epsilon > 0 \), by Bayes rule we have:

\[
P(A[t, t+\epsilon] = 1\mid Y(t) = y(t), \hat{O}(t)) = \frac{P(A[t, t+\epsilon] = 1\mid \hat{O}(t))dF(y(t)\mid A[t, t+\epsilon] = 1, \hat{O}(t)) + P(A[t, t+\epsilon] = 0\mid \hat{O}(t))A_{\epsilon}}{P(A[t, t+\epsilon] = 1\mid \hat{O}(t)) + P(A[t, t+\epsilon] = 0\mid \hat{O}(t))D_{\epsilon}}, \tag{8}
\]

for \( D_{\epsilon} := A_{\epsilon}/dF(y(t)\mid A[t, t+\epsilon] = 1, \hat{O}(t)) \). Dividing each side of (8) by \( \epsilon \) and taking the limit as \( \epsilon \to 0^+ \) gives:

\[
\lim_{\epsilon \to 0^+} P(A[t, t+\epsilon] = 1\mid Y(t) = y(t), \hat{O}(t)) = \left( \lim_{\epsilon \to 0^+} \frac{P(A[t, t+\epsilon] = 1\mid \hat{O}(t))}{\epsilon} \right) \bigg/ \left( \lim_{\epsilon \to 0^+} P(A[t, t+\epsilon] = 1\mid \hat{O}(t)) + \lim_{\epsilon \to 0^+} P(A[t, t+\epsilon] = 0\mid \hat{O}(t)) \lim_{\epsilon \to 0^+} D_{\epsilon} \right). \tag{9}
\]

The left hand side of (9) equals \( \rho(t, Y(t), \hat{O}(t)) \), while the numerator of the right hand side equals \( \lambda(t, \hat{O}(t)) \). In the denominator, write:

\[
D_{\epsilon} = \frac{B_{\epsilon}}{dF(y(t)\mid A[t, t+\epsilon] = 1, \hat{O}(t))} + \frac{A_{\epsilon} - B_{\epsilon}}{dF(y(t)\mid A[t, t+\epsilon] = 1, \hat{O}(t))} = \frac{\exp\{\alpha y\} E[\exp\{\alpha Y(t)\}\mid A[t, t+\epsilon] = 1, \hat{O}(t)]}{dF(y(t)\mid A[t, t+\epsilon] = 1, \hat{O}(t))} + \frac{A_{\epsilon} - B_{\epsilon}}{dF(y(t)\mid A[t, t+\epsilon] = 1, \hat{O}(t))},
\]

24
and note that \( \lim_{\epsilon \to 0^+}(A_\epsilon - B_\epsilon) = \)

\[
dF(y(t)|A(t) = 0, \tilde{O}(t)) - dF(y(t)|A(t) = 1, \tilde{O}(t)) \frac{\exp\{\alpha y\}}{E[\exp\{\alpha Y(t)\}|A(t) = 1, \tilde{O}(t)]} = 0
\]

by Assumption \[23\]. Therefore, the denominator in \[9\] is equal to:

\[
0 + 1 \times \left( \frac{\exp\{\alpha y\}}{E[\exp\{\alpha Y(t)\}|A(t) = 1, \tilde{O}(t)]} + 0 \right),
\]

which gives the result.

\[ \square \]

**B Influence function derivation**

In this section we prove Theorem \[1\] using the semiparametric theory presented in \[33, 37\]. We first briefly review elements of this theory applied to our context.

We will refer here to \( L = \{Y(t): 0 \leq t \leq \tau\} \) as the full data. We write \( R = \{N(t): a \leq t \leq b\} \), and we refer to \( Z = (L, R) \) as the total data. Let \( \mathcal{H}^L, \mathcal{H}^O, \) and \( \mathcal{H}^Z \) be the Hilbert spaces of mean-zero, finite-variance, \( p \)-dimensional functions of the full, observed, and total data respectively, with the covariance inner product. Full-data influence functions are normalized elements of the orthogonal complement of the full-data nuisance tangent space \( \Lambda(F_L) \). Observed data influence functions are normalized elements of \( \Lambda^{O,\perp} = \Lambda_1^{O,\perp} \cap \Lambda_2^{O,\perp} \), where \( \Lambda_1^O \) is the image of \( \Lambda(F_L) \), and \( \Lambda_2^O \) is the image of the coarsening nuisance tangent space \( \Lambda(F_{R|L}) \), under the map \( A : \mathcal{H}^Z \to \mathcal{H}^O \) given by \( A(\cdot) = E[\cdot | O] \). Therefore, the orthogonal complement \( \Lambda_1^{O,\perp} \) is equal to the null space of the adjoint of the map \( A_1 : \Lambda(F_L) \to \mathcal{H}^O \) given by \( A_1(\cdot) = E[\cdot | O] \), and similarly \( \Lambda_2^{O,\perp} \) is the null space of the adjoint of the map \( A_2 : \Lambda(F_{R|L}) \to \mathcal{H}^O \) given by \( A_2(\cdot) = E[\cdot | O] \). The adjoint of \( A_1 \) is the map \( A_1^*: \mathcal{H}^O \to \Lambda(F_L) \) given by \( A_1^*(\cdot) = \Pi( E[\cdot | L] | \Lambda(F_L)) \). Therefore, \( \Lambda_1^{O,\perp} = \{h(O) \in \mathcal{H}^O : \mathcal{K}(g(O)) \in \Lambda(F_L)^\perp\} \), where \( \mathcal{K}: \mathcal{H}^O \to \mathcal{H}^L \) is the map defined by \( \mathcal{K}(g(O)) = E[g(O)|L] \).

Again using adjoints, \( \Lambda_2^{O,\perp} = \{g(O) \in \mathcal{H}^O : g(O) \in \Lambda(F_{R|L})^\perp\} = \mathcal{H}^O \cap \Lambda(F_{R|L})^\perp \), and therefore, \( \Lambda^{O,\perp} = \Lambda_1^{O,\perp} \cap \Lambda(F_{R|L})^\perp \).

Based on this theory, we use the following procedure to derive an observed-data influence function: We first derive a full-data influence function \( \varphi(L) \). We then obtain elements of \( \Lambda_1^{O,\perp} \) by inverse-weighting \( \varphi(L) \) and adding elements of the augmentation space (defined below). We then identify an element \( h(O) \) of this form which is also orthogonal to \( \Lambda(F_{R|L}) \). By the above theory, \( h(O) \) is in the space \( \Lambda^{O,\perp} \). Furthermore, \( h(O) \) inherits the property of being correctly normalized from \( \varphi(L) \), and therefore \( h(O) \) is an observed data influence function.

**Proof of Theorem \[1\]**

**Proof.**

**i. A full-data influence function.** We obtain a full-data influence function for \( \beta \) following the approach of \[23\]. We begin by considering the nonparametric model on \( L \)
that does not make Assumption 3. We consider a squared-error loss function for the loss incurred from approximating the curve $E[Y(t)], a \leq t \leq b$, using a smooth function of the form $s(\gamma'B(t))$, for specified $s(\ )$ and $B(t)$, for some parameter $\gamma \in \mathbb{R}^p$. We define the parameter $\gamma_0$ to be the minimizer of this squared-error loss function:

$$
\gamma_0 := \text{argmin}_{\gamma} \left\{ \int_{t=a}^{b} \left( E[Y(t)] - s(\gamma'B(t)) \right)^2 dt \right\},
$$
or equivalently, as the solution of the following system of equations:

$$
\int_{t=a}^{b} \left\{ \left( E[Y(t)] - s(\gamma'B(t)) \right) \frac{\partial s(\gamma'B(t))}{\partial \gamma} \right\} dt = 0.
$$

Below we compute the influence function for $\gamma_0$ in the nonparametric model, which we denote by $\varphi^{NP}(L; \gamma_0)$. Since the nonparametric model is a supermodel of our model, $\varphi^{NP}(L; \gamma_0)$ is also a full-data influence function for $\gamma_0$ in our model. If Assumption 3 holds, then the minimizer of the squared-error loss function is the value for which the loss is zero, i.e., in this case $\gamma_0 = \beta$. Therefore, $\varphi^{NP}(L; \gamma_0)$ is also a valid full-data influence function for $\beta$ in our model.

Let $\mathcal{P}$ be any parametric submodel (parametrized by $\epsilon$ such that $\epsilon = 0$ corresponds to the true distribution of $L$) and let $a(L)$ denote the score vector for $\mathcal{P}$. The influence function for $\gamma_0$ is the mean-zero, $p$-dimensional function of $L$ satisfying: $\frac{\partial s}{\partial \epsilon} \bigg|_{\epsilon=0} = E[\varphi^{NP}(L; \gamma_0) a(L)]$.

For a fixed $a(L)$, we consider the parametric submodel $\mathcal{P}_a = \{ P_a(\epsilon) : \epsilon \in D \subset \mathbb{R} \}$, where $P_a(\epsilon) = dF_0(L)(1 + \epsilon a(L))$, and where the score of $P_a(\epsilon)$ evaluated at $\epsilon = 0$ is $a(L)$. In $\mathcal{P}_a$, equation (10) becomes:

$$
\int_{t=a}^{b} \left\{ \left( \int_{\ell} y(t) \, dF_0(\ell) \right)(1 + \epsilon a(\ell) - s(\gamma(\epsilon)B(t))) \frac{\partial}{\partial \gamma} s(\gamma(\epsilon)B(t)) \right\} dt = 0.
$$

Taking the derivative with respect to $\epsilon$ gives:

$$
\int_{t=a}^{b} \left\{ \left( \int_{\ell} y(t) \, dF_0(\ell) a(\ell) - \frac{\partial }{\partial \gamma} s(\gamma(\epsilon)B(t)) \frac{\partial s(\gamma(\epsilon)B(t))}{\partial \epsilon} \right) \left( \frac{\partial}{\partial \gamma} s(\gamma(\epsilon)B(t)) \right) + \left( \int_{\ell} y(t) \, dF_0(\ell)(1 + \epsilon a(\ell) - s(\gamma(\epsilon)B(t))) \left( \frac{\partial^2}{\partial \gamma \partial \gamma'} s(\gamma(\epsilon)B(t)) \frac{\partial s(\gamma(\epsilon)B(t))}{\partial \epsilon} \right) \right) dt = 0.
$$

Setting $\epsilon = 0$ gives:

$$
\int_{t=a}^{b} \left\{ E[Y(t)a(L)] \frac{\partial}{\partial \gamma} s(\gamma_0'B(t)) \right\} dt = \left( \int_{t=a}^{b} \left\{ \frac{\partial}{\partial \gamma} s(\gamma_0'B(t)) \frac{\partial s(\gamma_0'B(t))}{\partial \gamma'} \right\} dt \right) \times \left( \frac{\partial}{\partial \epsilon} \bigg|_{\epsilon=0} \right).
$$

26
Solving for $\frac{\partial \gamma}{\partial \epsilon} \bigg|_{\epsilon=0}$ and writing: $V(\gamma_0) = \int_{t=a}^{b} \left\{ \frac{\partial}{\partial \gamma} s\left(\gamma'_0 B(t)\right) - \frac{\partial^2}{\partial \gamma \partial \gamma'} s\left(\gamma'_0 B(t)\right) \left( E[Y(t)] - s(\gamma'_0 B(t)) \right) \right\} dt,$ we have:

$$\frac{\partial \gamma}{\partial \epsilon} \bigg|_{\epsilon=0} = V(\gamma_0)^{-1} \int_{t=a}^{b} \left\{ E\left[Y(t) a(L) \right] \frac{\partial}{\partial \gamma'} s(\gamma'_0 B(t)) \right\} dt$$

$$= E \left[ \left( \int_{t=a}^{b} \left\{ V(\gamma_0)^{-1} \left( Y(t) - E[Y(t)] \right) \frac{\partial}{\partial \gamma'} s(\gamma'_0 B(t)) \right\} dt \right) a(L) \right]$$

Therefore:

$$\varphi^{NP}(L; \gamma_0) = \int_{t=a}^{b} \left\{ V(\gamma_0)^{-1} \left( Y(t) - E[Y(t)] \right) \frac{\partial}{\partial \gamma'} s(\gamma'_0 B(t)) \right\} dt.$$

Now making Assumption 3, $\varphi^{NP}(L; \gamma_0)$ simplifies as:

$$\varphi(L) = \int_{t=a}^{b} \left\{ V(\beta)^{-1} \left( Y(t) - s(\beta' B(t)) \right) \frac{\partial}{\partial \beta} s(\beta' B(t)) \right\} dt$$

where

$$V(\beta) = \int_{t=a}^{b} \left\{ \frac{\partial}{\partial \beta} s(\beta' B(t)) \frac{\partial}{\partial \beta'} s(\beta' B(t)) \right\} dt. \quad (11)$$

**ii. An inverse-weighted element and the augmentation space.** Next we find an element $g^*(O) \in \mathcal{H}^O$ which maps to $\varphi(L)$ under the map $\mathcal{K}$. One such element is the following inverse-weighted element:

$$g^*(O) := \int_{t=a}^{b} \left\{ \frac{dN(t)}{\rho(t, Y(t), O(t))} V(\beta)^{-1} \left( Y(t) - s(\beta' B(t)) \right) \frac{\partial}{\partial \beta'} s(\beta' B(t)) \right\}, \quad (12)$$

as we verify below. To show this, we will make use of a counting process martingale whose properties we will exploit throughout this appendix. Define $M(t) := N(t) - \int_{u=a}^{t} \rho(u, Y(u), O(u)) du$. Then $\{M(t)\}_{t=a}^{t}$ is a martingale adapted to the filtration $\{\mathcal{G}_t\}_{t=a}^{t}$ defined in Appendix A. Now let $H(t)$ be any predictable process; a sufficient condition for this is that $H(t)$ is left continuous in $t$, and that for each $t$, $H(t)$ is $\mathcal{G}_t$-measurable. In particular, $H(t)$ could be any function of random variables in $O(t)$ and of $Y(t)$, since we assume that $Y(t)$ is left continuous. Then $U(t) := \int_{u=a}^{t} H(u) dM(u)$ is also a martingale adapted to $\{\mathcal{G}_t\}_{t=a}^{t}$, and therefore $E\left[U(t) \mid \mathcal{G}_s\right] = U(s)$ for all $s < t$. In particular, $U(t)$ has mean zero and, since $\mathcal{G}_a = L$:

$$E \left[ \int_{u=a}^{t} H(u) dM(u) \bigg| L \right] = E\left[U(t) \mid \mathcal{G}_a\right] = U(a) = 0. \quad (13)$$
Therefore, by (13)

\[
E[g^*(O)|L] = E \left[ \left( \int_{t=a}^b \frac{V(\beta)^{-1}(Y(t) - s(\beta'B(t))) \frac{\partial}{\partial \beta} s(\beta'B(t))}{\rho(t,Y(t),O(t))} \left( dN(t) - \rho(t,Y(t),O(t))dt \right) \right) \mid L \right] + E \left[ \left( \int_{t=a}^b \left( V(\beta)^{-1}(Y(t) - s(\beta'B(t))) \frac{\partial}{\partial \beta} s(\beta'B(t)) \right) dt \right) \mid L \right]
\]

\[
= 0 + \int_{t=a}^b \left\{ V(\beta)^{-1}(Y(t) - s(\beta'B(t))) \frac{\partial}{\partial \beta} s(\beta'B(t)) \right\} dt = \varphi(L)
\]

as claimed. The inverse-weighted element \(g^*(O)\) and the augmentation space \(\text{Aug} := \{ g(O) \in \mathcal{H}^O : E[g(O)|L] = 0 \}\) determine the space \(\Lambda^{O,\perp}_1\), as \(\Lambda^{O,\perp}_1 = \{ g^*(O) + h^*(O) : h^*(O) \in \text{Aug} \}\). By (13), a subset of \(\text{Aug}\) is given by:

\[
\widetilde{\text{Aug}} := \left\{ \left( \int_{t=a}^b h(t,Y(t),O(t)) \left[ dN(t) - \rho(t,Y(t),O(t))dt \right] \right) : \text{any } h(t,Y(t),O(t)) \right\}
\]

### iii. The coarsening tangent space.

The coarsening tangent space \(\Lambda(F_{\hat{R}|L})\) is the set of all elements of \(\mathcal{H}^Z\) that are linear combinations of scores of \(R|L\) in a parametric submodel of our model. Following [9], we can compute the likelihood for \(R|L\) by first partitioning the interval \([a, b]\) into a finite number of subintervals \(\{t_k, t_k + \epsilon\}, k = 0, \ldots, K\), each of length \(\epsilon\), and then taking the limit as \(K \to \infty\). Let \(\Delta(t_k)\) denote the number of visits during the interval \([t_k, t_k + \epsilon]\); for small \(\epsilon\), this number of visits is assumed to be 0 or 1. For fixed \(K\), the conditional (on \(L\)) likelihood of \(\hat{R} = (N(t_0), \ldots, N(t_K))\) is the same as the conditional (on \(L\)) likelihood of \((\Delta(t_0), \ldots, \Delta(t_K))\), which can be written as:

\[
P(\Delta(t_0)|L) \prod_{k=1}^{K} P(\Delta(t_k)|\Delta(t_0), \ldots, \Delta(t_{k-1}), L).
\]

For each \(k = 1, \ldots, K\), \(P(\Delta(t_k)|\Delta(t_0), \ldots, \Delta(t_{k-1}), L) =

\[
\rho(t_k, L, \bar{O}(t_k)) + o(\epsilon) = P(\Delta(t_k)|\bar{O}(t_k), L) = \rho(t_k, Y(t_k), \bar{O}(t_k)) + o(\epsilon),
\]

where the last equality holds by Assumption [2]. Similarly \(P(\Delta(t_0)|L) = \rho(t_0, L, \bar{O}(t_0)) + o(\epsilon) = \rho(t_0, Y(t_0), \bar{O}(t_0)) + o(\epsilon)\). Therefore, for fixed \(K\), the conditional (on \(L\)) likelihood of \((\Delta(t_0), \ldots, \Delta(t_K))\) is:

\[
\prod_{k=0}^{K} \left( \rho(t_k, Y(t_k), \bar{O}(t_k)) + o(\epsilon) \right)^{\Delta(t_k)} \left(1 - \rho(t_k, Y(t_k), \bar{O}(t_k)) + o(\epsilon) \right)^{1-\Delta(t_k)}. \]

After dividing by the constant \(\prod_{k=0}^{K} \epsilon^{\Delta(t_k)}\), the conditional likelihood is proportional to:

\[
\prod_{k=0}^{K} \left( \rho(t_k, Y(t_k), \bar{O}(t_k)) + o(\epsilon) / \epsilon \right)^{\Delta(t_k)} \left(1 - \rho(t_k, Y(t_k), \bar{O}(t_k)) + o(\epsilon) \right)^{1-\Delta(t_k)}. \]
The continuous-time conditional (on $L$) likelihood is the limit of (14) as $K \to \infty$. We compute this by viewing (14) as a product of three factors and noting that:

$$\lim_{K \to \infty} \prod_{k=0}^{K} \left( \rho(t_k, Y(t_k), \bar{O}(t_k)) + o(\epsilon) / \epsilon \right)^{\Delta(t_k)} = \prod_{t=a}^{b} \left( \rho(t, Y(t), \bar{O}(t)) \right)^{A(t)}.$$ 

By a result on product integrals [9],

$$\lim_{K \to \infty} \prod_{k=0}^{K} \left( 1 - \rho(t_k, Y(t_k), \bar{O}(t_k)) \epsilon + o(\epsilon) \right) = \exp \left\{ - \int_{t=a}^{b} \rho(t, Y(t), \bar{O}(t)) dt \right\}.$$ 

Finally, since $\Delta(t_k) = 1$ at only a fixed number of values $t_k$ as $K$ varies, and since $\lim_{\epsilon \to 0^+} \{ 1 - \rho(t_k, Y(t_k), \bar{O}(t_k)) \epsilon + o(\epsilon) \} = 1$ for each $t_k$, we have:

$$\lim_{K \to \infty} \prod_{k=0}^{K} (1 - \rho(t_k, Y(t_k), \bar{O}(t_k)) \epsilon + o(\epsilon))^{\Delta(t_k)} = 1,$$

Therefore, the continuous-time conditional (on $L$) likelihood $L$ is:

$$L = \prod_{t=a}^{b} \rho(t, Y(t), \bar{O}(t))^{A(t)} \exp \left\{ - \int_{t=a}^{b} \rho(t, Y(t), \bar{O}(t)) dt \right\}.$$ 

To compute $\Lambda(F_{R|L})$, we now consider the conditional likelihood in a parametric submodel of our model for $dF(R|L)$. Any such parametric submodel is induced by a model for the intensity function $\rho(t, Y(t), \bar{O}(t))$, say $\{ \rho(t, Y(t), \bar{O}(t); \gamma) : \gamma \in \Gamma \}$ with $\gamma$ a finite-dimensional parameter and $\gamma = 0$ corresponding to the truth. In order to satisfy the relation in Proposition 1(a), we must have $\rho(t, Y(t), \bar{O}(t); \gamma) = r(t, \bar{O}(t); \gamma) \exp\{ -\alpha Y(t) \}$ for some function $r(t, \bar{O}(t); \gamma)$. The conditional likelihood in this parametric submodel is:

$$L(\gamma) = \left( \prod_{t=a}^{b} \rho(t, Y(t), \bar{O}(t); \gamma)^{A(t)} \right) \exp \left\{ - \int_{t=a}^{b} \rho(t, Y(t), \bar{O}(t); \gamma) dt \right\}.$$ 

The log-conditional likelihood is:

$$\ell(\gamma) = \int_{t=a}^{b} dN(t) \log(\rho(t, Y(t), \bar{O}(t); \gamma)) - \int_{t=a}^{b} \rho(t, Y(t), \bar{O}(t); \gamma) dt,$$

and the score is:

$$S_\gamma(O) = \int_{t=a}^{b} dN(t) \left[ \frac{\partial}{\partial \gamma} \rho(t, Y(t), \bar{O}(t); \gamma) \right] - \int_{t=a}^{b} \frac{\partial}{\partial \gamma} \rho(t, Y(t), \bar{O}(t); \gamma) dt$$

$$= \int_{t=a}^{b} \left[ \frac{\partial}{\partial \gamma} \rho(t, Y(t), \bar{O}(t); \gamma) \right] \left[ dN(t) - \rho(t, Y(t), \bar{O}(t); \gamma) dt \right].$$
Since \( \frac{\partial}{\partial \gamma} \rho(t, Y(t), \bar{O}(t); \gamma) = \frac{\partial}{\partial \gamma} r(t, \bar{O}(t); \gamma) \) is a function of \( t, \bar{O}(t) \) only, evaluating the score at the truth \( \gamma = 0 \) shows that:

\[
\Lambda(\mathcal{F}_{RL}) \subseteq \left\{ \left( \int_{t=a}^{b} h(t, \bar{O}(t)) \left[ dN(t) - \rho(t, Y(t), \bar{O}(t)) dt \right] \right) \in \mathcal{H}^2 : \text{any } \left( \begin{array}{c} h(t, \bar{O}(t)) \\ \rho(t, Y(t), \bar{O}(t)) \end{array} \right) \right\}.
\]

**iv. An observed-data influence function.** We are now ready to identify an element of \( \Lambda^{O,\perp} \). Write \( W(t; \beta) := V(\beta)^{-1} \frac{\partial}{\partial \beta} s(\beta^t B(t)) \) where \( V(\beta) \) is as in (11), such that the inverse-weighted element \( g^*(O) \) can be written as:

\[
g^*(O) = \int_{t=a}^{b} W(t; \beta) \left\{ \frac{dN(t)}{\rho(t, Y(t), \bar{O}(t))} \left[ Y(t) - s(\beta^t B(t)) \right] \right\}.
\]

Consider the function:

\[
h^*(O) = \int_{t=a}^{b} W(t; \beta) \left( E[Y(t)|\bar{O}(t)] - s(\beta^t B(t)) \right) \frac{dN(t)}{\rho(t, Y(t), \bar{O}(t))} \left[ dN(t) - \rho(t, Y(t), \bar{O}(t)) dt \right].
\]

Here \( h^*(O) \in \widetilde{Aug} \), and therefore \( g^*(O) - h^*(O) \in \Lambda^{O,\perp} \). We claim that \( g^*(O) - h^*(O) \) is also orthogonal to \( \Lambda(\mathcal{F}_{RL}) \). To show this, we use the martingale covariance property [2], which says that, for predictable processes \( H_1(t) \) and \( H_2(t) \),

\[
E \left[ \left( \int_{t=a}^{b} H_1(t) \left( dN(t) - \rho(t, Y(t), \bar{O}(t)) dt \right) \right) \left( \int_{t=a}^{b} H_2(t) \left( dN(t) - \rho(t, Y(t), \bar{O}(t)) dt \right) \right) \right] = E \left[ \int_{t=a}^{b} H_1(t) H_2(t) \rho(t, Y(t), \bar{O}(t)) dt \right].
\]

(15)

To show that \( g^*(O) - h^*(O) \in \Lambda(\mathcal{F}_{RL}) \), let \( w(Z) \) be any element of \( \Lambda(\mathcal{F}_{RL}) \). Then \( w(Z) = \int_{t=a}^{b} h(t, \bar{O}(t)) \left[ dN(t) - \rho(t, Y(t), \bar{O}(t)) dt \right] \) for some \( h(t, \bar{O}(t)) \). Then by (15):

\[
E[w(Z)h^*(O)] = E \left[ \int_{t=a}^{b} h(t, \bar{O}(t)) W(t; \beta) \left( E[Y(t)|\bar{O}(t)] - s(\beta^t B(t)) \right) \rho(t, Y(t), \bar{O}(t)) dt \right]
\]

\[
= E \left[ \int_{t=a}^{b} h(t, \bar{O}(t)) W(t; \beta) \left( E[Y(t)|\bar{O}(t)] - s(\beta^t B(t)) \right) dt \right].
\]

Next, adding and subtracting to write \( g^*(O) \) as the sum of a martingale plus another term, we have:

\[
E[g^*(O)w(Z)] = E \left[ \left( \int_{t=a}^{b} W(t; \beta) \left( Y(t) - s(\beta^t B(t)) \right) \rho(t, Y(t), \bar{O}(t)) dt \right) \left( \int_{t=a}^{b} W(t; \beta) \left( Y(t) - s(\beta^t B(t)) \right) \rho(t, Y(t), \bar{O}(t)) dt \right) w(Z) \right]
\]

(16)

\[
+ E \left[ \left( \int_{t=a}^{b} W(t; \beta) \left( Y(t) - s(\beta^t B(t)) \right) dt \right) w(Z) \right].
\]
Using iterated expectation conditioning on $L$, the second term in (16) equals:

$$E \left[ \left( \int_{t=a}^{b} W(t; \beta) \left( Y(t) - s(\beta^t B(t)) \right) dt \right) \times \right.$$  

$$E \left[ \left( \int_{t=a}^{b} h(t, \bar{O}(t)) \left[ dN(t) - \rho(t, Y(t), \bar{O}(t)) dt \right] \right) | L \right] = 0.$$

By (15), the first term in (16) is equal to:

$$E \left[ \int_{t=a}^{b} h(t, \bar{O}(t)) W(t; \beta) \left( Y(t) - s(\beta^t B(t)) \right) dt \right]$$

$$= E \left[ \int_{t=a}^{b} h(t, \bar{O}(t)) W(t; \beta) \left( E[Y(t)|\bar{O}(t)] - s(\beta^t B(t)) \right) dt \right].$$

Therefore $E \left[ (g^*(O) - h^*(O)) w(Z) \right] = 0$. This shows that $g^*(O) - h^*(O) \in \Lambda^{O,\perp} = \Lambda^{O,\perp}_1 \cap \Lambda(F_{RL})^\perp$, and therefore $g^*(O) - h^*(O)$ is an influence function if it also satisfies the equation $E[(g^*(O) - h^*(O)) S_{\beta}(O)] = 1_{p \times p}$, where $S_{\beta}(O)$ is the observed-data score for $\beta$ and $1_{p \times p}$ is the $p \times p$ identity matrix. Following [37] Theorem 8.3, we rewrite the left side using iterated expectations and the fact that $S_{\beta}(O) = E[S_{\beta}(L)|O]$, where $S_{\beta}(L)$ is the full-data score for $\beta$:

$$E \left[ (g^*(O) - h^*(O)) S_{\beta}(O) \right]$$

$$= E \left[ (g^*(O) - h^*(O)) E[S_{\beta}(L)|O] \right]$$

$$= E \left[ (g^*(O) - h^*(O)) S_{\beta}(L) \right]$$

$$= E \left[ E[g^*(O)|L] S_{\beta}(L) \right] - E \left[ E[h^*(O)|L] S_{\beta}(L) \right]$$

$$= E \left[ \varphi(L) S_{\beta}(L) \right] = 1_{p \times p}$$

where the last equality holds since $\varphi(L)$ is a full-data influence function. Denoting $g^*(O) - h^*(O)$ now by $\varphi(O; P)$, we have thus shown that one of the observed-data influence functions for $\beta$ in our model is:

$$\varphi(O; P) = \int_{t=a}^{b} W(t; \beta) \left\{ \frac{1}{p(t,Y(t),\bar{O}(t))} \left( Y(t) - E[Y(t)|\bar{O}(t)] \right) \right\} dN(t) + \int_{a}^{b} W(t; \beta) \left( E[Y(t)|\bar{O}(t)] - s(\beta^t B(t)) \right) dt.$$
C  Remainder Term

Proof of Proposition 2

Proof. Let $\mathcal{M}$ be our model for the observed data $O$. In what follows, we will use a $\tilde{P}$ subscript to denote expectations or intensities under a distribution $\tilde{P} \in \mathcal{M}$. Expectations and intensities with no subscript are taken under the true distribution $P$. Let $\psi : \mathcal{M} \rightarrow \mathbb{R}^p$ be the parameter mapping, and for each $\tilde{P} \in \mathcal{M}$, write $\psi(\tilde{P}) = \beta_{\tilde{P}}$. By Assumption 3, $E_{\tilde{P}}[Y(t)] = B(t)\beta_{\tilde{P}}$, so that $\psi(\tilde{P}) = \int_{t=a}^{b} V^{-1}B(t)E_{\tilde{P}}[Y(t)]dt$. Now $\psi(\tilde{P})$ has a von Mises expansion as:

$$\psi(\tilde{P}) = \psi(P) - E[\varphi(O; \tilde{P})] + \text{Rem}(\tilde{P}, P),$$

which we use to compute $\text{Rem}(\tilde{P}, P)$. We have:

$$E[\varphi(O; \tilde{P})] = E\left[\int_{t=a}^{b} \left\{ V^{-1}B(t) \left( \frac{Y(t) - E_{\tilde{P}}[Y(t)|\tilde{O}(t)]}{\rho_{\tilde{P}}(t,Y(t),\tilde{O}(t))} \right) \right\} \left( dN(t) - \rho(t,Y(t),\tilde{O}(t)) \right) \right] = 0$$

$$+ E\left[\int_{t=a}^{b} \left\{ V^{-1}B(t) \left( \frac{Y(t) - E_{\tilde{P}}[Y(t)|\tilde{O}(t)]}{\rho_{\tilde{P}}(t,Y(t),\tilde{O}(t))} \right) \right\} \rho(t,Y(t),\tilde{O}(t))dt \right]$$

$$+ E\left[\int_{t=a}^{b} \left\{ V^{-1}B(t)E_{\tilde{P}}[Y(t)|\tilde{O}(t)] \right\} dt \right] - \beta_{\tilde{P}},$$

$$= E\left[\int_{t=a}^{b} \left\{ V^{-1}B(t) \rho(t,Y(t),\tilde{O}(t)) \frac{\rho(t,Y(t),\tilde{O}(t))}{\rho_{\tilde{P}}(t,Y(t),\tilde{O}(t))} (Y(t) - E_{\tilde{P}}[Y(t)|\tilde{O}(t)]) \right\} dt \right]$$

$$+ E\left[\int_{t=a}^{b} \left\{ V^{-1}B(t)E_{\tilde{P}}[Y(t)|\tilde{O}(t)] \right\} dt \right] - \beta_{\tilde{P}},$$

$$= E\left[\int_{t=a}^{b} \left\{ V^{-1}B(t) \rho(t,Y(t),\tilde{O}(t)) \frac{\rho(t,Y(t),\tilde{O}(t))}{\rho_{\tilde{P}}(t,Y(t),\tilde{O}(t))} \left( Y(t) - E_{\tilde{P}}[Y(t)|\tilde{O}(t)] - E_{\tilde{P}}[Y(t)|\tilde{O}(t)] \right) \right\} dt \right]$$

$$+ E\left[\int_{t=a}^{b} \left\{ V^{-1}B(t)E_{\tilde{P}}[Y(t)|\tilde{O}(t)] \right\} dt \right] - \beta_{\tilde{P}},$$

where in the last equality we used the fact that, by Proposition 1 (a):

$$\frac{\rho(t,Y(t),\tilde{O}(t))}{\rho_{\tilde{P}}(t,Y(t),\tilde{O}(t))} = \frac{\lambda(t,\tilde{O}(t))E[\exp\{\alpha Y(t)|A(t) = 1, \tilde{O}(t)\}]}{\lambda_{\tilde{P}}(t,\tilde{O}(t))E_{\tilde{P}}[\exp\{\alpha Y(t)\}|A(t) = 1, \tilde{O}(t)\]}$$

32
is a function of \( t \) and \( \tilde{O}(t) \) only. Therefore:

\[
\text{Rem}(\tilde{P}, P) = E \left[ \int_{t=a}^{b} \left\{ V^{-1}B(t) \frac{\rho(t, Y(t), \tilde{O}(t))}{\rho_{\tilde{P}}(t, Y(t), \tilde{O}(t))} \left( E[Y(t)|\tilde{O}(t)] - E_{\tilde{P}}[Y(t)|\tilde{O}(t)] \right) \right\} \, dt \right]
\]

\[
+ E \left[ \int_{t=a}^{b} \left\{ V^{-1}B(t) \left( \frac{\rho(t, Y(t), \tilde{O}(t))}{\rho_{\tilde{P}}(t, Y(t), \tilde{O}(t))} - 1 \right) \left( E[Y(t)|\tilde{O}(t)] - E_{\tilde{P}}[Y(t)|\tilde{O}(t)] \right) \right\} \, dt \right].
\]

\[\square\]

**Proof of Theorem 2.**

**Proof.** Let \( \eta = \{ \lambda_{0}(t) : a \leq t \leq b \}, \gamma, \theta \), and let \( \hat{\eta} = \{ \hat{\lambda}_{0}(t) : a \leq t \leq b \}, \hat{\gamma}, \hat{\theta} \) be the estimates in the distribution \( \hat{P} \). By the Cauchy-Schwarz inequality,

\[
(\text{Rem}(\hat{P}, P))^{2} = E \left[ \int_{t=a}^{b} \left\{ V^{-1}B(t) \left( \frac{\rho(t, Y(t), \tilde{O}(t); \eta)}{\rho(t, Y(t), \tilde{O}(t); \hat{\eta})} - 1 \right) \times \right. \]

\[
\left. \left( E[Y(t)|\tilde{O}(t); \hat{\theta}] - E[Y(t)|\tilde{O}(t); \hat{\theta}] \right) \right\} \, dt \right]^{2}
\]

\[
\leq E \left[ \int_{t=a}^{b} \left\{ \left( \frac{V^{-1}B(t)}{\rho(t, Y(t), \tilde{O}(t); \hat{\eta})} \right)^{2} \left( \rho(t, Y(t), \tilde{O}(t); \eta) - \rho(t, Y(t), \tilde{O}(t); \hat{\eta}) \right)^{2} \right\} \, dt \right] \quad (17)
\]

\[
\times E \left[ \int_{t=a}^{b} \left\{ E[Y(t)|\tilde{O}(t); \hat{\theta}] - E[Y(t)|\tilde{O}(t); \hat{\theta}] \right\}^{2} \, dt \right]. \quad (18)
\]

In the factor \( (17) \), we expand \( \rho(t, Y(t), \tilde{O}(t); \eta) - \rho(t, Y(t), \tilde{O}(t); \hat{\eta}) \) as the sum of three
terms, as:

\[
\rho(t, Y(t), \tilde{O}(t); \hat{\eta}) - \rho(t, Y(t), \bar{O}(t); \eta) = \hat{\lambda}_0(t) \exp\{\tilde{\gamma} Z(t)\} \exp\{-\alpha Y(t)\} E\left[ \exp\{\alpha Y(t)\} | A(t) = 1, \bar{O}(t); \hat{\theta} \right] - \\
\lambda_0(t) \exp\{\gamma Z(t)\} \exp\{-\alpha Y(t)\} E\left[ \exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \theta \right]
\]

\[
= \left( \hat{\lambda}_0(t) - \lambda_0(t) \right) \exp\{\tilde{\gamma} Z(t)\} \exp\{-\alpha Y(t)\} E\left[ \exp\{\alpha Y(t)\} | A(t) = 1, \bar{O}(t); \hat{\theta} \right] + \\
\lambda_0(t) \left( \exp\{\tilde{\gamma} Z(t)\} - \exp\{\gamma Z(t)\} \right) \exp\{-\alpha Y(t)\} E\left[ \exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \theta \right] + \\
\lambda_0(t) \exp\{\gamma Z(t)\} \exp\{-\alpha Y(t)\} \left\{ E\left[ \exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \hat{\theta} \right] - \\
E\left[ \exp\{\alpha Y(t)\} | A(t) = 1, \bar{O}(t); \hat{\theta} \right] \right\}.
\]

Using this expansion and the inequality \( rs \leq \frac{1}{2}(r^2 + s^2) \), the rate for (17) will be determined by the rates of three terms, namely the squares of each of the terms in the expansion above.

For the factor (18), using Proposition (b), we write:

\[
E[Y(t)|\bar{O}(t); \theta] - E[Y(t)|\tilde{O}(t); \hat{\theta}]
\]

\[
= \frac{E[Y(t) \exp\{\alpha Y(t)\} | A(t) = 1, \bar{O}(t); \hat{\theta}]}{E[\exp\{\alpha Y(t)\} | A(t) = 1, \bar{O}(t); \hat{\theta}]} - \frac{E[Y(t) \exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \theta]}{E[\exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \theta]}
\]

\[
= \frac{E[\exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \theta] \times \left( E[\exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \hat{\theta}] - E[\exp\{\alpha Y(t)\} | A(t) = 1, \bar{O}(t); \hat{\theta}] \right)}{E[\exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \theta] \times \left( E[Y(t) \exp\{\alpha Y(t)\} | A(t) = 1, \tilde{O}(t); \hat{\theta}] - E[Y(t) \exp\{\alpha Y(t)\} | A(t) = 1, \bar{O}(t); \hat{\theta}] \right)}
\]

The rates for (18) will similarly be determined by the rates for the squares of the two terms.
in this expansion. Combining these results, we therefore have:

\[
\left| \text{Rem}(\hat{P}, P) \right|^2 \leq \left( C_1(\tilde{\eta}) \mathbb{E} \left[ \int_{t=a}^{b} \left( \hat{\lambda}_0(t) - \lambda_0(t) \right)^2 dt \right] + \right.
\]

\[
\left. C_2(\tilde{\eta}) \mathbb{E} \left[ \int_{t=a}^{b} \left( \exp\{\tilde{\gamma}Z(t)\} - \exp\{\gammaZ(t)\} \right)^2 dt \right] + \right.
\]

\[
\left. C_3(\tilde{\eta}) \mathbb{E} \left[ \int_{t=a}^{b} \left\{ \mathbb{E}[\exp\{\alpha Y(t)\} : A(t) = 1, \bar{O}(t); \hat{\theta}] - \right\} \right. \right.
\]

\[
\left. \mathbb{E}[\exp\{\alpha Y(t)\}|A(t) = 1, \bar{O}(t); \theta] \}^2 dt \right) \right)
\]

\[
\left. \times \left( C_4(\tilde{\eta}) \mathbb{E} \left[ \int_{t=a}^{b} \left\{ \mathbb{E}[Y(t) \exp\{\alpha Y(t)\}|A(t) = 1, \bar{O}(t); \hat{\theta}] - \right\} \right. \right.
\]

\[
\left. \mathbb{E}[Y(t) \exp\{\alpha Y(t)\}|A(t) = 1, \bar{O}(t); \theta] \}^2 dt \right) + \right.
\]

\[
\left. C_5(\tilde{\eta}) \mathbb{E} \left[ \int_{t=a}^{b} \left\{ \mathbb{E}[\exp\{\alpha Y(t)\} : A(t) = 1, \bar{O}(t); \hat{\theta}] - \right\} \right. \right.
\]

\[
\left. \mathbb{E}[\exp\{\alpha Y(t)\}|A(t) = 1, \bar{O}(t); \theta] \}^2 dt \right) \right)
\]

where \( C_1(\tilde{\eta}), \ldots, C_5(\tilde{\eta}) \) are terms that are each \( O_P(1) \) under conditions (i) and (iv) of Theorem 2.

We first consider term (19). Let \( X_n(t) = (nb_n)^{1/2} (\hat{\lambda}_0(t) - \lambda_0(t)) \). Wells [40] has shown that, under condition (ii) of Theorem 2, the stochastic processes \( \{X_n(t) : a \leq t \leq b\}_{n=1}^{\infty} \) converges weakly to a process \( \{X(t) : a \leq t \leq b\} \), where \( \{X(t) : a \leq t \leq b\} \) is a Wiener process plus a term depending on \( \lambda_n(t) \). Therefore it follows by the Continuous Mapping Theorem that \( \{X_n^2(t) : a \leq t \leq b\}_{n=1}^{\infty} \) converges weakly to \( \{X^2(t) : a \leq t \leq b\} \).

Again by the Continuous Mapping Theorem, \( \int_{t=a}^{b} X_n^2(t) dt = nh_n \left( \int_{t=a}^{b} \left( \hat{\lambda}_0(t) - \lambda_0(t) \right)^2 dt \right) \)

is \( O_p(1) \). Therefore, the term in (19) is \( O_P((nh_n)^{-1}) \). By assumption (ii), for large \( n \), \( nh_n \) is approximately \( n^{4/5} \), so in particular the convergence rate for (19) is faster than \( n^{1/2} \).

Since \( \tilde{\gamma} \) converges to \( \gamma \) at root-\( n \) rates, the term in (20) is \( O_P(n^{-1}) \).

Next we consider terms (21) and (23). By the Cauchy-Schwarz inequality, for each \( t \),
we can bound the integrand in (21) and (23) as:

\[
\begin{align*}
\left( E[ \exp\{\alpha Y(t)\} : A(t) = 1, \bar{O}(t); \hat{\theta} ] \right. & \left. - E[ \exp\{\alpha Y(t)\} | A(t) = 1, \bar{O}(t); \hat{\theta} ] \right)^2 \\
& = \left( \int_{y(t)} e^{\alpha y(t)} \left( f(y(t)|A(t) = 1, \bar{O}(t); \hat{\theta}) - f(y(t)|A(t) = 1, \bar{O}(t); \hat{\theta}) \right) d\nu(y(t)) \right)^2 \\
& \leq \left( \int_{y(t)} e^{2\alpha y(t)} d\nu(y(t)) \right) \times \\
& \left( \int_{y(t)} \left( f(y(t)|A(t) = 1, \bar{O}(t); \hat{\theta}) - f(y(t)|A(t) = 1, \bar{O}(t); \hat{\theta}) \right)^2 d\nu(y(t)) \right),
\end{align*}
\]

where \(\nu(\cdot)\) is an appropriate dominating measure. Therefore, since \(\int_{y(t)} e^{2\alpha y(t)} d\nu(y(t))\) is uniformly bounded, it follows from assumption (iii) of Theorem 2 that (21) and (23) are \(o_P(n^{-1/2})\).

Therefore, the sum of terms (19), (20), and (21) is \(o_P(n^{-1/2})\), since each of these terms converges at faster than root-\(n\) rates.

Finally we consider term (22). Since \(\int_{y(t)} y(t)^2 e^{2\alpha y(t)} d\nu(y(t))\) is also uniformly bounded, (22) is also \(o_P(n^{-1/2})\) by the same argument as for terms (21) and (23). Therefore, the sum of terms (22) and (23) is \(o_P(n^{-1/2})\). Hence the product that bounds \(\left| Rem(\hat{P}, P) \right| \) is \(o_P(n^{-1})\), so that \(Rem(\hat{P}, P)\) is \(o_P(n^{-1/2})\) as claimed. 

\[\square\]