Biometrics, 2024, 80(4), ujae154
https://doi.org/10.1093/biomtc/ujae154
Biometric Practice

Semi-parametric sensitivity analysis for trials with irregular and informative assessment times

Bonnie B. Smith [1,*], Yujing Gao [2], Shu Yang [2], Ravi Varadhan [3], Andrea J. Apter [4], Daniel O. Scharfstein [5]

[1]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, United States, [2]Department of Statistics, North Carolina State University, Raleigh, NC 27695, United States, [3]Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD 21205, United States, [4]Pulmonary Allergy Critical Care Division, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States, [5]Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, UT 84108, United States

[*]Corresponding author: Bonnie B. Smith, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Baltimore, MD 21205, United States (bsmit179@jhmi.edu).

ABSTRACT

Many trials are designed to collect outcomes at or around pre-specified times after randomization. If there is variability in the times when participants are actually assessed, this can pose a challenge to learning the effect of treatment, since not all participants have outcome assessments at the times of interest. Furthermore, observed outcome values may not be representative of all participants' outcomes at a given time. Methods have been developed that account for some types of such irregular and informative assessment times; however, since these methods rely on untestable assumptions, sensitivity analyses are needed. We develop a sensitivity analysis methodology that is benchmarked at the explainable assessment (EA) assumption, under which assessment and outcomes at each time are related only through data collected prior to that time. Our method uses an exponential tilting assumption, governed by a sensitivity analysis parameter, that posits deviations from the EA assumption. Our inferential strategy is based on a new influence function-based, augmented inverse intensity-weighted estimator. Our approach allows for flexible semiparametric modeling of the observed data, which is separated from specification of the sensitivity parameter. We apply our method to a randomized trial of low-income individuals with uncontrolled asthma, and we illustrate implementation of our estimation procedure in detail.

KEYWORDS: asthma; explainable assessment; influence function; inverse intensity weighting; semi-parametric estimation.

1 INTRODUCTION

Many randomized trials are designed to collect outcome information at or around certain pre-specified times after randomization. In practice, however, there can be substantial variability in the times when participants' outcomes are actually assessed. Such *irregular assessment times* pose a challenge to learning the effect of treatment, similar to that posed by missing data. While the goal is to learn population mean outcomes and treatment effects at certain target times, not all participants are assessed at those times, and the observed outcomes may not be representative. For example, participants may miss or postpone data collection appointments at times when their outcome is worse, such that outcomes in the study data tend to be better compared to the population distribution. In other studies, participants may tend to have assessments at times when their outcome is worse—for example, if the study collects data at "as-needed" appointments. We say the assessment times are *informative* if the distribution of observed outcomes at a given time differs from the population distribution of outcomes at that time.

A number of inferential methods have been developed for prospective studies with informative assessment times. All approaches impose untestable assumptions about the joint distribution of the outcome and assessment time processes. Lin and Ying (2001) posited semi-parametric regression models with time-varying covariates for the outcome and assessment time processes, and they used an assumption that the 2 processes are conditionally independent given these time-varying covariates to construct estimating equations for the outcome regression parameters. Their approach was generalized by several authors to allow for dependence between the outcome and assessment time processes through latent variables (eg, random effects and frailty terms) in addition to covariates in the outcome regression model; see, for example, Sun et al. (2007), Sun et al. (2011a), Sun et al. (2011b), Liang et al. (2009). Lin et al. (2004) instead developed an inverse intensity weighting approach, also within an estimating equations framework, under which the outcome and assessment time processes can be associated through past observed outcomes and time-varying covariates that are not included in the outcome model. Therefore, their approach allows inference for the marginal mean of the outcome process. Inverse intensity weighting approaches have also been developed by Bůžková and Lumley (2007, 2009), Pullenayegum and Feldman (2013), and Sun et al. (2016). Other authors have used likelihood-based approaches coupled

Received: October 28, 2022; Revised: October 22, 2024; Accepted: November 27, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of The International Biometric Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

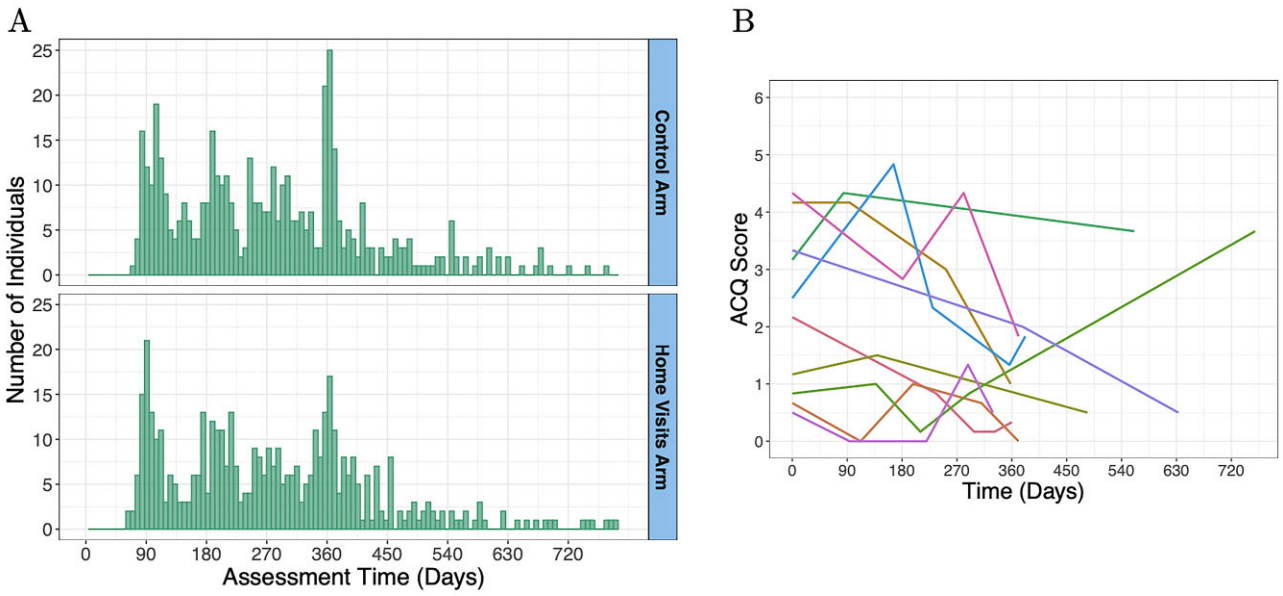

FIGURE 1 Assessment times and outcomes in the Asthma Research for the Community (ARC) study. Panel A: assessment times by treatment arm. The protocol called for assessments at 3, 6, 9, and 12 months after randomization, but there was substantial variability in the actual times of assessment around these targeted times in each arm. Panel B: outcome trajectories for a sample of participants, showing their score on the Asthma Control Questionnaire (ACQ) at each assessment time. Outcomes fluctuated considerably over time for some participants.

with assumptions that obviate the need for modeling the assessment time process: Lipsitz et al. (2002) used a parametric approach, while Chen et al. (2015) and Shen et al. (2019) used composite likelihoods conditioned on order statistics to express the conditional density of observed outcomes in terms of the outcome density of interest. As noted above, the key caveat for all of these approaches is that untestable assumptions are needed; therefore, sensitivity analysis would be a valuable addition to each method. This is analogous to methods for trials with missing data, which require untestable assumptions such as missing at random. There, sensitivity analysis has been recognized as an important component of the analysis; see, for example, the report, *The Prevention and Treatment of Missing Data in Clinical Trials* (National Research Council, 2010).

Inverse intensity weighting approaches rely on the assumption that assessment and outcomes at each time t are related only through study data observed before time t, such as baseline covariates, treatment assignment, times of earlier assessments, and outcomes and time-varying covariates observed at those earlier assessments. We refer to this assumption as *explainable assessment*. While it is less restrictive than assuming that outcomes and assessment times are unrelated or related only through baseline variables, the explainable assessment assumption may not hold in some studies. For example, some participants could have a new downturn in their health that also prevents them from attending a data collection appointment. Therefore, it is important to assess how inference changes under departures from this assumption.

Here, we develop a sensitivity analysis methodology, anchored at the explainable assessment assumption, for estimating the population mean of the (possibly unobserved) outcome values at a fixed time after randomization. Our method accounts for the possibility that participants with worse outcomes at a given time may be more (or less) likely than other participants to have assessments at that time, even after controlling for variables observed earlier in the study. Our estimation approach uses a new influence function-based *augmented inverse intensity-weighted* estimator, which allows for flexible semi-parametric modeling while allowing for root-n rates of convergence for our estimator. Additionally, all modeling of the observed data is separate from the sensitivity parameter.

We apply our methodology to the Asthma Research for the Community (ARC) study (Apter et al., 2019), a pragmatic randomized trial of 301 low-income participants with uncontrolled asthma. Participants in the active 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 in the intervention group received home visits by community health workers to promote care coordination and help with the use of the patient portal, in addition to usual care and portal training. The primary outcome was the score on the Asthma Control Questionnaire (ACQ) (Juniper et al., 1999), reflecting symptoms over the week prior to assessment. 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 actual times of assessments. Additionally, in the intervention (control) arm, 4 (10) participants had 0 post-baseline assessments, 9 (8) had only 1, 24 (29) had only 2, and 34 (27) had only 3 post-baseline assessments.

Data on specific reasons for delays were not collected; however, investigators believe that difficulties in reaching participants were largely due to factors such as participants' competing work obligations and other demands on their time, where

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

participants may have paid less close attention to requests for follow-up assessments during times when they were functioning well. There were also some delays when participants waited to return contact from project staff because they were not feeling well enough or were seeking treatment or were hospitalized. Figure 1 also shows outcome trajectories for a sample of participants, with substantial increases and decreases in ACQ Score over time for some individuals. Given all these factors, it is possible that assessment at time t may be associated with the outcome at time t, even after adjusting for variables such as previous outcome values. The distribution of assessment times in Figure 1 is similar in both arms, as is the distribution of inter-assessment times (not shown); however, this does not indicate that treatment effect estimation would remain valid if we failed to account for informative assessment times in the analysis. For example, the direction or strength of informativeness could be differential across treatment arms.

The rest of the paper is organized as follows: in Section 2, we introduce notation and define explainable assessment. In Section 3, we present our sensitivity analysis framework and model assumptions. Section 4 details our estimation procedure. In Section 5, we discuss calibration of the sensitivity parameter. A reanalysis of the ARC study is provided in Section 6. A simulation study is presented in Section 7, and Section 8 concludes with a discussion. A tutorial illustrating implementation of our estimator on simulated data is provided in Web Appendix A.1 of the Supplementary Materials, along with all code and the simulated dataset.

2 BACKGROUND

2.1 Setting and notation

We consider a trial with a continuous outcome in which participants are randomized to either treatment or control. Each participant's outcome is assessed at baseline and at some number of subsequent times, where the timing and possibly the number of post-baseline assessments vary by participant. The goal of the trial is to learn the population mean outcome under treatment versus control at one or more fixed follow-up times. For simplicity, we suppose that there is some time interval $[t_1, t_2]$ that includes all of these target follow-up times, and such that assessments take place throughout this interval in each arm; see Web Appendix A.2 for trials with gaps in time when few assessments occur. We assume that no participants drop out of the study (though they may have fewer assessments than the protocol specifies).

Let A be the treatment assignment for a random individual, with $A = 1$ if the individual is assigned to treatment and $A = 0$ if they are assigned to control. Let τ be the end of follow-up; note that $[0, \tau]$ is the maximal period of follow-up, while the interval $[t_1, t_2]$ described above is a possibly smaller interval (with $0 < t_1 < t_2 \leq \tau$) selected by the analyst over which inference will be drawn; see also the positivity requirement in Assumption 2. For each $t \in [0, \tau]$, let $Y(t)$ be the (possibly unobserved) value of the participant's outcome at time t. If fixed and/or time-varying auxiliary covariates are collected, let $\boldsymbol{X}(t)$ be the value of the participant's covariates at time t. Let $N(t)$ be the number of assessments that the individual has had up through time t, and let $\Delta N(t) = N(t) - N(t-)$ be the indicator that the individual has an assessment at time t. Let T_k be the time of the individual's kth post-baseline assessment. We refer to $\{Y(t) : t \in [0, \tau]\}$ as the *outcome process* and $\{N(t) : t \in [0, \tau]\}$ as the *assessment process*. For each $t \in [0, \tau]$, let $\overline{\boldsymbol{O}}(t)$ denote all of the participant's study data observed before time t, including baseline data, treatment assignment, times of assessments prior to t, and data collected at each assessment prior to t. We call $\overline{\boldsymbol{O}}(t)$ the participant's *observed past* before time t, with $\boldsymbol{O} = \overline{\boldsymbol{O}}(\tau)$ the participant's observed data over the entire study. Finally, for each $t \in [0, \tau]$, let $Y^1(t)$ and $Y^0(t)$ be the outcomes that the participant would have at time t under assignment to treatment and control, respectively. Let $\mu_1(t) = E\left\{Y^1(t)\right\}$ and $\mu_0(t) = E\left\{Y^0(t)\right\}$, the population mean outcome at time t were all individuals assigned to treatment or to control, respectively. For the effect of treatment, we focus on the difference $\delta(t) = \mu_1(t) - \mu_0(t)$.

2.2 Explainable assessment

The assumption of explainable assessment says (informally) that any relationship between assessment at time t and the outcome $Y(t)$ is accounted for by the observed past $\overline{\boldsymbol{O}}(t)$. This assumption has been referred to as sequential ignorability (Lin et al., 2004), visiting at random (Pullenayegum and Lim, 2016), or assessment at random (Pullenayegum and Scharfstein, 2022), and is analogous to the sequential exchangeability assumption that has been used in the longitudinal missing data literature, for example, in Vansteelandt et al. (2007). To define explainable assessment formally, here we use the *intensity function for the assessment process given the observed past*:

$$\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\} = \lim_{\epsilon \to 0^+}\left[P\left\{N(t+\epsilon) - N(t-) = 1 \mid \overline{\boldsymbol{O}}(t)\right\}/\epsilon\right], \quad (1)$$

where $N(t+\epsilon) - N(t-)$ is the indicator that the participant has an assessment during the time interval $[t, t+\epsilon]$. Consider a time t with $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\} > 0$. We define

$$dF\left\{y(t) \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right\} = \lim_{\epsilon \to 0^+} dF\left\{y(t) \mid N(t+\epsilon) - N(t-) = 1, \overline{\boldsymbol{O}}(t)\right\}$$

and

$$dF\left\{y(t) \mid \Delta N(t) = 0, \overline{\boldsymbol{O}}(t)\right\} = \lim_{\epsilon \to 0^+} dF\left\{y(t) \mid N(t+\epsilon) - N(t-) = 0, \overline{\boldsymbol{O}}(t)\right\},$$

the distributions of $Y(t)$ among those who were, and who were not, assessed at time t, given $\overline{\boldsymbol{O}}(t)$.

Definition 1 (Explainable assessment) *We say that assessment is explainable (by the observed past) if*

$$dF\left\{y(t) \mid \Delta N(t)=1, \overline{\boldsymbol{O}}(t)\right\} = dF\left\{y(t) \mid \Delta N(t) = 0, \overline{\boldsymbol{O}}(t)\right\}$$

for all t with $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\} > 0$.

That is, within strata of the observed past, under explainable assessment the distribution of $Y(t)$ is the same among those who were, and who were not, assessed at time t.

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

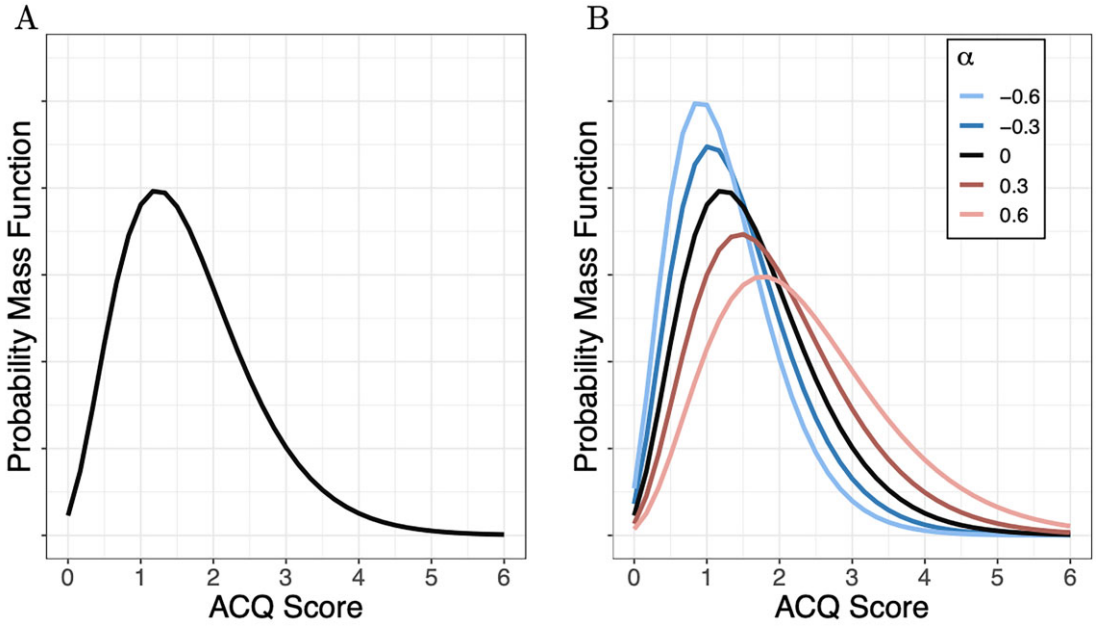

FIGURE 2 Illustration of the tilt assumption (Assumption 1) in the context of the Asthma Research for the Community (ARC) trial. Panel A: distribution of $Y(t)$ at 6 months among participants with a certain observed past who were assessed at 6 months. Panel B: posited distributions of $Y(t)$ at 6 months among participants who had the same observed past but were not assessed at 6 months. Under the explainable assessment assumption ($\alpha = 0$), the distribution for assessed versus non-assessed participants would be the same. Under a positive (negative) value of α, the distribution for non-assessed participants would be tilted with more weight on higher (lower) values of $Y(t)$. Here, we show smoothed depictions of the probability mass functions for this outcome.

Lin et al. (2004) developed the method of *inverse intensity weighting* for studies with explainable assessment, extending weighting methods to the continuous-time setting by using weights based on the intensity function in Equation (1). The weights create a pseudo-population in which assessment times and outcomes are no longer related if assessment is explainable.

3 SENSITIVITY ANALYSIS FRAMEWORK AND MODELS

In some studies, dependence between assessment and outcomes at time t may not be fully explained by variables from earlier assessments. For example, in studies that collect outcomes at "as-needed" appointments, a sudden downturn in health may lead participants to seek care. Unfortunately, whether assessment is explainable cannot be determined from the study data, which contain no information about the distribution of unobserved outcomes, $dF\left\{y(t) \mid \Delta N(t) = 0, \overline{\boldsymbol{O}}(t)\right\}$. In particular, the current outcome $Y(t)$ could impact assessment at t even if earlier outcomes do not impact assessment at t strongly, particularly in studies where outcomes tend to fluctuate over time. There may be alternate assumptions that are equally as plausible as explainable assessment, which could yield different inferences about the treatment effect. Our sensitivity analysis provides an inferential strategy for the treatment effect $\delta(t)$ under a range of different plausible assumptions.

3.1 Sensitivity analysis framework

Here, we draw inference for $\mu_a(t) = E\left\{Y^a(t)\right\}$ separately for each treatment assignment $a = 0, 1$. We leverage the fact that, by randomization, $\mu_a(t) = E\{Y(t) \mid A = a\}$, the mean outcome at time t among participants assigned to treatment arm a, and we work separately by treatment arm. That is, all assumptions, distributions, and estimators are treatment arm-specific. For ease of notation, we suppress dependence on the treatment arm until Section 4.4. In addition to explainable assessment, we include assumptions under which outcomes among participants who are not assessed at a given time t tend to be larger, or smaller, than outcomes among similar participants who are assessed at time t. Specifically, $dF\left\{y(t) \mid \Delta N(t) = 0, \overline{\boldsymbol{O}}(t)\right\}$ is assumed to be some "tilted version" of the distribution $dF\left\{y(t) \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right\}$, with the magnitude and direction of the tilt determined by an arm-specific sensitivity parameter α. We assume that $E\left[\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right]$ exists for all α in some neighborhood of $\alpha = 0$. Then, for each value of α in a range around $\alpha = 0$ to be specified by the analyst (see Section 5) and lying within this neighborhood:

Assumption 1 (Tilting assumption) *For each time t with $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\} > 0$,*

$$dF\left\{y(t) \mid \Delta N(t) = 0, \overline{\boldsymbol{O}}(t)\right\} = dF\left\{y(t) \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right\} \exp\{\alpha y(t)\}/c\{\overline{\boldsymbol{O}}(t); \alpha\},$$

where $c\left\{\overline{\boldsymbol{O}}(t); \alpha\right\} = E\left[\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right]$ and we assume $c\left\{\overline{\boldsymbol{O}}(t); \alpha\right\} < \infty$.

For a negative (positive) value of α, the distribution of unobserved outcomes in the given arm is tilted to the left (right) relative to the distribution of observed outcomes in

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

that arm, with smaller (larger) values of $Y(t)$ receiving greater weight. A value of $\alpha = 0$, or no tilt, is the explainable assessment assumption. Figure 2 illustrates tilting for values of $\alpha = -0.6, -0.3, 0, 0.3, 0.6$ in the context of the ARC trial. More generally, an alternate version of the tilting assumption can be used, with a different choice of function $q\{t, Y(t); \alpha\}$ in place of $\alpha Y(t)$; here we use $\alpha Y(t)$ for interpretability. The construction that we use here, known as *exponential tilting* (Barndorff-Nielsen and Cox, 1994), has been used by Rotnitzky et al. (2001), Birmingham et al. (2003), Vansteelandt et al. (2007), and Scharfstein and McDermott (2019), among others, in sensitivity analyses for trials with missing or censored data. It was also used by Franks et al. (2020) for observational studies with possible unmeasured confounding and for trials with irregular and informative assessment times by Wang (2020). Wang (2020) developed a discrete-time framework with estimation carried out using g-computation with fully parametric models. In contrast, our influence function-based approach allows more flexible modeling while maintaining $\sqrt{n}$ rates of convergence.

Proposition 1 *For each time t with $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\} > 0$, the mean $\mu(t)$ in each arm is identified under Assumption 1, as*

$$\mu(t) = E\left(\frac{E\left[Y(t)\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right]}{E\left[\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right]}\right).$$

The proof is shown in Web Appendix B.1 in the Supplementary Materials and also shows the following:

Corollary 1 *For each t with $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\} > 0$, the conditional mean outcome given the observed past, $E\left\{Y(t) \mid \overline{\boldsymbol{O}}(t)\right\}$, is identified from the observed data as*

$$E\left\{Y(t) \mid \overline{\boldsymbol{O}}(t)\right\} = \frac{E\left[Y(t)\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right]}{E\left[\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right]}.$$

3.2 Inverse intensity weighting under the tilting assumption

Our approach extends inverse intensity-weighting to our sensitivity analysis framework. Since assessment at time t can depend on the current outcome $Y(t)$ under Assumption 1, we use weights based on the following intensity function:

$$\rho\left\{t \mid Y(t), \overline{\boldsymbol{O}}(t)\right\} = \lim_{\epsilon \to 0^+}\left[P\left\{N(t+\epsilon) - N(t-) = 1 \mid Y(t), \overline{\boldsymbol{O}}(t)\right\}/\epsilon\right]. \quad (2)$$

Assumption 2 (Positivity assumption) *There is some $c > 0$ such that, for all t in $[t_1, t_2]$, $\rho\left\{t \mid Y(t), \overline{\boldsymbol{O}}(t)\right\} > c$ for all values of $Y(t)$ and $\overline{\boldsymbol{O}}(t)$.*

The intensity function $\rho\left\{t \mid Y(t), \overline{\boldsymbol{O}}(t)\right\}$ is related to the intensity function $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\}$ in Equation (1) through the following:

Proposition 2 *Under Assumptions 1 and 2, for each t in $[t_1, t_2]$,*

$$\rho\left\{t \mid Y(t), \overline{\boldsymbol{O}}(t)\right\} = \lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\} E\left[\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right] / \exp\{\alpha Y(t)\}.$$

The proof is given in Web Appendix B.1. We leverage this relationship in Section 4 to keep observed data modeling separated from sensitivity parameters. Proposition 2 also gives an interpretation of α as the log of the ratio of the intensities at time t for participants who have the same observed past and whose outcomes $Y(t)$ differ by one unit:

$$\log\left[\frac{\rho\left\{t \mid Y(t) = y(t), \overline{\boldsymbol{O}}(t)\right\}}{\rho\left\{t \mid Y(t) = y(t) + 1, \overline{\boldsymbol{O}}(t)\right\}}\right] = \alpha.$$

3.3 Additional assumptions

Proposition 1 shows that, under Assumption 1, $\mu(t)$ at each time t would theoretically be estimable from infinite data. In order to estimate $\mu(t)$ from finite data, we also make the following smoothing assumption that allows us to borrow information across different times.

Assumption 3 (Marginal mean assumption) $\mu(t) = \boldsymbol{B}(t)'\boldsymbol{\beta}$ *for all $t \in [t_1, t_2]$, for some specified vector-valued basis function $\boldsymbol{B}(t) = (B_1(t), \ldots, B_p(t))'$ with $\boldsymbol{V} = \int_{t=t_1}^{t_2} \boldsymbol{B}(t)\boldsymbol{B}(t)'dt$ invertible, and $\boldsymbol{\beta} \in \mathbb{R}^p$ a parameter vector.*

Note that Assumption 3 uses an identity link appropriate for a continuous outcome.

Proposition 3 *The parameter $\boldsymbol{\beta}$ is identified under Assumptions 1, 2, and 3.*

Proof. Under Assumptions 1 and 2, $\mu(t)$ is identified from the observed data for each $t \in [t_1, t_2]$, and under Assumption 3, $\boldsymbol{\beta} = \boldsymbol{V}^{-1}\int_{t=t_1}^{t_2} \boldsymbol{B}(t)\mu(t)dt$. □

Finally, we assume that assessment depends on future values of the outcome and covariates, the current value of covariates, and past unobserved values of the outcome and covariates only through past observed data and the current value of the outcome:

Assumption 4 (Non-future dependence assumption) *Let $\boldsymbol{L} = \{Y(t), \boldsymbol{X}(t) : 0 \le t \le \tau\}$. Then*

$$\lim_{\epsilon \to 0^+}\left[P\left\{N(t+\epsilon) - N(t-) = 1 \mid \overline{\boldsymbol{O}}(t), \boldsymbol{L}\right\}/\epsilon\right] = \lim_{\epsilon \to 0^+}\left[P\left\{N(t+\epsilon) - N(t-) = 1 \mid \overline{\boldsymbol{O}}(t), Y(t)\right\}/\epsilon\right].$$

Similar non-future dependence assumptions have been used in longitudinal settings with missing data (Kenward et al., 2003; Wang and Daniels, 2011). Assumption 4 aids in derivation of an influence function for $\boldsymbol{\beta}$. However, investigators should consider whether it is tenable in their study. An example where Assumption 4 would likely not hold is a study where assessments occur at doctors' visits when participants are receiving care, which then impacts future outcomes. In this case, after adjusting for the observed past and $Y(t)$, there could be dependence between future

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

outcomes and assessment at time t, since both are related to receiving care at time t. (Note that here receipt of care at time t is not conditioned on, since our approach does not accommodate adjustment for variables that occur at the time of assessment t, except for $Y(t)$.)

4 ESTIMATION

4.1 Observed data modeling

To implement our approach, researchers must fit 2 types of models. First, in each arm the intensity function $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\}$ is modeled using an Andersen-Gill model (Andersen and Gill, 1982), or a stratified Andersen-Gill model stratified by assessment number $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\} = \lambda_{0,k}(t)\exp\left\{\boldsymbol{\gamma}'\boldsymbol{Z}(t)\right\}D_k(t)$. Here, $\boldsymbol{Z}(t)$ is a specified (possibly vector-valued) function of the participant's observed past $\overline{\boldsymbol{O}}(t)$ containing key baseline covariates and time-varying factors that impact both assessment time and outcome, such as outcomes at previous assessments. The function $\lambda_{0,k}(t)$ is an unspecified baseline intensity function for stratum k, $\boldsymbol{\gamma}$ is a parameter vector, and $D_k(t)$ is an indicator that the participant is at risk for having the kth assessment at time t. The baseline intensity functions $\lambda_{0,k}(t)$ are estimated by kernel smoothing the Breslow estimator of the cumulative baseline intensity functions (Breslow, 1972).

Second, the conditional distribution of observed outcomes in each arm given the observed past, $dF\left\{y(t) \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right\}$, is modeled using a single index model (Chiang and Huang, 2012). In the single index model, the conditional cumulative distribution function of $Y(t)$ given a vector of predictors, say $\boldsymbol{W}(t)$, is assumed to depend on $\boldsymbol{W}(t)$ only through a scalar $\boldsymbol{\theta}'\boldsymbol{W}(t)$. Thus, this function is modeled as $G\{\cdot, \boldsymbol{\theta}'\boldsymbol{W}(t); \boldsymbol{\theta}\}$, where $G\{y, u; \boldsymbol{\theta}\}$ is a cumulative distribution function in y for each u, and $\boldsymbol{\theta}$ is a vector of unknown parameters. The estimator of G is a step function in y that is kernel smoothed with respect to $\boldsymbol{\theta}'\boldsymbol{W}(t)$ via a bandwidth parameter h; $\boldsymbol{\theta}$ and h are jointly estimated by minimizing a pseudo sum of integrated squares (Chiang and Huang, 2012). Specifically, in our context, for each t and each value of $\overline{\boldsymbol{o}}(t)$, $\widehat{F}\left\{y(t) \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t) = \overline{\boldsymbol{o}}(t)\right\}$ is a step function with jumps at all outcome values observed in the data.

4.2 Estimation of the mean outcome $\boldsymbol{\mu(t)}$ under $\boldsymbol{\alpha}$

The following result provides a way of constructing estimators for $\boldsymbol{\beta}$ and $\mu(t)$ that incorporate the flexible models fit in Section 4.1, yet converge at fast parametric rates.

Theorem 1 *Under Assumptions 1-4, an influence function for* $\boldsymbol{\beta}$ *is given by*

$$\boldsymbol{\varphi}(\boldsymbol{O}) = \int_{t=t_1}^{t_2} \boldsymbol{V}^{-1}\boldsymbol{B}(t)\frac{\left[Y(t) - E\left\{Y(t) \mid \overline{\boldsymbol{O}}(t)\right\}\right]}{\rho\left\{t \mid Y(t), \overline{\boldsymbol{O}}(t)\right\}}dN(t) + \int_{t=t_1}^{t_2} \boldsymbol{V}^{-1}\boldsymbol{B}(t)E\left\{Y(t) \mid \overline{\boldsymbol{O}}(t)\right\}dt \; - \; \boldsymbol{\beta},$$

where $\boldsymbol{B}(t)$ *and* $\boldsymbol{V}$ *are given in Assumption 3.*

The proof of Theorem 1 is given in Web Appendix B.3.

Suppose that we have data for n independent individuals. We construct estimators $\widehat{\beta}$ and $\widehat{\mu}(t)$ using the following steps, where we use a subscript i to denote data for individual i. For each individual i:

1. For each assessment k with T_{ik} in the interval $[t_1, t_2]$, compute

$$\widehat{E}\left[\exp\left\{\alpha Y(T_{ik})\right\} \mid \Delta N(T_{ik}) = 1, \overline{\boldsymbol{O}}_i(T_{ik})\right] = \int_{y\in\mathcal{Y}} \exp(\alpha y)\, d\widehat{F}\left\{Y(T_{ik}) = y \mid \Delta N(T_{ik}) = 1, \overline{\boldsymbol{O}}_i(T_{ik})\right\},$$

where $\mathcal{Y}$ is the set of all outcome values occurring in the data; the estimated conditional mean (see Corollary 1)

$$\widehat{E}\left\{Y(T_{ik}) \mid \overline{\boldsymbol{O}}_i(T_{ik})\right\} = \frac{\widehat{E}[Y(T_{ik})\exp\{\alpha Y(T_{ik})\} \mid \Delta N(T_{ik}) = 1, \overline{\boldsymbol{O}}_i(T_{ik})]}{\widehat{E}[\exp\{\alpha Y(T_{ik})\} \mid \Delta N(T_{ik}) = 1, \overline{\boldsymbol{O}}_i(T_{ik})]};$$

and the estimated intensity (see Proposition 2)

$$\widehat{\rho}\left\{T_{ik} \mid Y_i(T_{ik}), \overline{\boldsymbol{O}}_i(T_{ik})\right\} = \widehat{\lambda}_{0,k}(T_{ik})\exp\{\widehat{\boldsymbol{\gamma}}'\boldsymbol{Z}_i(T_{ik})\}\exp\{-\alpha Y_i(T_{ik})\} \times \widehat{E}[\exp\{\alpha Y(T_{ik})\} \mid \Delta N(T_{ik}) = 1, \overline{\boldsymbol{O}}_i(T_{ik})].$$

2. For each time t in $[t_1, t_2]$, compute the predicted mean outcome at time t given their observed past before time t:

$$\widehat{E}\left\{Y(t) \mid \overline{\boldsymbol{O}}_i(t)\right\} = \frac{\widehat{E}[Y(t)\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}_i(t)]}{\widehat{E}[\exp\{\alpha Y(t)\} \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}_i(t)]}.$$

3. Compute $\widehat{\boldsymbol{\Psi}}(\boldsymbol{O}_i) =$

$$\sum_{k\in S_i}\left\{\boldsymbol{V}^{-1}\boldsymbol{B}(T_{ik})\frac{\left[Y_i(T_{ik}) - \widehat{E}\left\{Y(T_{ik}) \mid \overline{\boldsymbol{O}}_i(T_{ik})\right\}\right]}{\widehat{\rho}\left\{T_{ik} \mid Y_i(T_{ik}), \overline{\boldsymbol{O}}_i(T_{ik})\right\}}\right\} + \int_{t=t_1}^{t_2} \boldsymbol{V}^{-1}\boldsymbol{B}(t)\widehat{E}\left\{Y(t) \mid \overline{\boldsymbol{O}}_i(t)\right\}dt,$$

where $S_i = \{k : T_{ik} \in [t_1, t_2]\}$.

Our *augmented inverse intensity-weighted estimators* are $\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{\Psi}}(\boldsymbol{O}_i)$ and $\widehat{\mu}(t) = \widehat{\beta}'\boldsymbol{B}(t)$.

4.3 Large-sample distribution of $\widehat{\boldsymbol{\beta}}$

If the models for $\lambda\left\{t \mid \overline{\boldsymbol{O}}(t)\right\}$ and $dF\left\{y(t) \mid \Delta N(t) = 1, \overline{\boldsymbol{O}}(t)\right\}$ are both correctly specified, and if Assumptions 1-4 and additional regularity conditions hold, then $\sqrt{n}\left(\widehat{\beta} - \boldsymbol{\beta}\right) \Rightarrow N\left[\boldsymbol{0}, Var\{\boldsymbol{\varphi}(\boldsymbol{O})\}\right]$. See Web Appendix B.4; there we derive the second-order remainder term in an expansion of $\sqrt{n}\left(\widehat{\beta} - \boldsymbol{\beta}\right)$ following Kennedy (2016), and we show conditions under which this remainder term is asymptotically negligible.

From this result, influence function-based variance estimators for $\widehat{\beta}$ and $\widehat{\mu}(t)$ are given by $\widehat{Var}\left(\widehat{\beta}\right) =$

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

$\frac{1}{n^2}\sum_{i=1}^{n}\left\{\widehat{\boldsymbol{\Psi}}(\boldsymbol{O}_i)-\widehat{\beta}\right\}\left\{\widehat{\boldsymbol{\Psi}}(\boldsymbol{O}_i)-\widehat{\beta}\right\}'$ and $\widehat{Var}\{\widehat{\mu}(t)\}=\boldsymbol{B}(t)'\widehat{Var}(\widehat{\beta})\boldsymbol{B}(t)$. A Wald confidence interval for $\mu(t)$ can be constructed using this influence function-based variance estimator or using a jackknife variance estimator. In simulations mimicking the ARC data, we found that nonparametric bootstrap was not a feasible of way of constructing confidence intervals; ties in bootstrapped datasets caused estimates of conditional cumulative distribution functions based on the single index model to be undefined.

4.4 Inference for $\delta(t)$

Here, we re-introduce subscripts for each treatment arm, and we also let α_1 and α_0 be sensitivity parameters for the treatment and control arms, respectively. To conduct the sensitivity analysis for $\delta(t)$, the estimation procedure above is repeated in the treatment arm to estimate $\mu_1(t)$ under a range of α_1 values, and separately in the control arm to estimate $\mu_0(t)$ under a range of α_0 values. These results are then combined to estimate the treatment effect $\delta(t)=\mu_1(t)-\mu_0(t)$ over a grid of sensitivity parameters (α_0,α_1).

5 SELECTION OF A RANGE OF SENSITIVITY PARAMETER VALUES

The analyst must decide on a range of sensitivity parameter values to include in the sensitivity analysis. Domain expertise should be used in making this decision, and how best to use such expertise is a key question for all sensitivity analyses. Cinelli and Hazlett (2020) 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 they write that the "bounding procedure we should use depends on which...quantities the investigator prefers and can most soundly reason about in their own research." In keeping with this, we propose the following approach in which domain experts reason about the treatment arm-specific mean outcome: We first query domain experts for extreme values $\mu_{\min}$ and $\mu_{\max}$ such that, in their judgment, a value of $\mu(t)$ outside of the bounds $(\mu_{\min},\mu_{\max})$ at any time t would be implausible. We then treat any value α under which $\mu(t)$ falls outside of $(\mu_{\min},\mu_{\max})$ for some t as implausible, and retain all other values. Other possible approaches could draw on bounding procedures that have been developed for sensitivity analyses for unmeasured confounding in observational studies. Authors including Franks et al. (2020), Sjölander et al. (2022), and Veitch and Zaveri (2020) have developed methods that use the strength of measured covariates' impact on the exposure and outcome to calibrate plausible values for the impact of unmeasured factors, after adjusting for measured covariates. An approach for our setting that borrows from ideas of Sjölander et al. (2022) could be to use some observable quantity that is related to the attenuation in $E\{Y(t)\mid\Delta N(t)=1\}-E\{Y(t)\mid\Delta N(t)=0\}$ obtained by adjusting for the observed past $\overline{\boldsymbol{O}}(t)$. This could in principle be used to calibrate plausible values for $E\{Y(t)\mid\Delta N(t)=1,\overline{\boldsymbol{O}}(t)\}-E\{Y(t)\mid\Delta N(t)=0,\overline{\boldsymbol{O}}(t)\}$, the residual difference due to unmeasured factors $\boldsymbol{U}(t)$ after adjusting for $\overline{\boldsymbol{O}}(t)$. As with other benchmarking approaches, this would use researchers' substantive beliefs that the strength of the impact of $\boldsymbol{U}(t)$ on $\Delta N(t)$ and $Y(t)$, after adjusting for $\overline{\boldsymbol{O}}(t)$, is no more than some factor, say r, times the marginal impact of $\overline{\boldsymbol{O}}(t)$ on $\Delta N(t)$ and $Y(t)$. Careful consideration would be needed in selecting a plausible value of r, taking into account the time scale of the study, since $\boldsymbol{U}(t)$ could include the outcome just before time t, whereas the impact of $\overline{\boldsymbol{O}}(t)$ on $\Delta N(t)$ and $Y(t)$ could be dampened due to the time elapsed since the previous study visit. Development of a method along these lines, and investigations that would guide when and how to implement it in practice, could be a direction for future research.

6 DATA ANALYSIS: ARC TRIAL

Here, we analyze the ARC data using our sensitivity analysis methodology. The ACQ is on a scale from 0 (completely controlled asthma) to 6 (extremely uncontrolled asthma) and takes values in $\{0, 1/6, 2/6, 3/6, \ldots, 6\}$. A positive value of α_a posits that unobserved values of the ACQ Score in treatment arm a tend to be higher (that is, worse) than observed values of the outcome in that arm at each time t, after controlling for variables observed before time t. This could be the case if participants tended to miss or postpone data collection appointments at times when their asthma was worse, so that some of the participants' higher ACQ Score values were not observed, while α_a could be negative if participants in arm a tended to be more engaged with the study at times when their asthma was worse. Since we do not know which, if either, of these is the case, we consider positive, negative, and zero values of α_a. We consider Assumption 4 to be reasonable since the assessment process was not tied to clinical care that might affect future outcomes.

We estimate $\mu_1(t)$ and $\mu_0(t)$ over a time interval of 60-460 days, since this interval contains the target times of 90, 180, 270, and 360 days and assessments occur throughout this period. For each $a=0,1$, we assume that $\mu_a(t)=\boldsymbol{\beta}_a'\boldsymbol{B}(t)$ for $t\in[60,460]$, with $\boldsymbol{B}(t)$ a cubic spline basis with one interior knot at $t=260$ days; this choice of $\boldsymbol{B}(t)$ allows the marginal mean to be a fairly flexible smooth function of time. We fit the models described in Section 4.1, modeling the intensity function $\lambda\left\{t\mid\overline{\boldsymbol{O}}(t)\right\}$ separately for each treatment arm using a stratified Andersen-Gill model with the outcome at the previous assessment as the predictor. We also considered an intensity model that includes lag time since the previous visit as an additional predictor. While lag time was a strong predictor in this model, the resulting inference for $\mu_a(t)$ was extremely similar under both models, and we therefore present the results of the simpler model. The coefficient for the previous outcome is -0.024 (standard error 0.038) in the intervention arm, and 0.042 (standard error 0.036) in the control arm. We estimated the baseline intensity functions using kernel smoothing of the Breslow estimate of the cumulative baseline intensity, with an Epanechnikov kernel and a bandwidth of 30 days. We modeled the conditional distribution of observed outcomes separately for each treatment arm using a single index model with the current time, lag time since the previous assessment, and a natural spline of the outcome at the previous assessment as predictors. We then

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

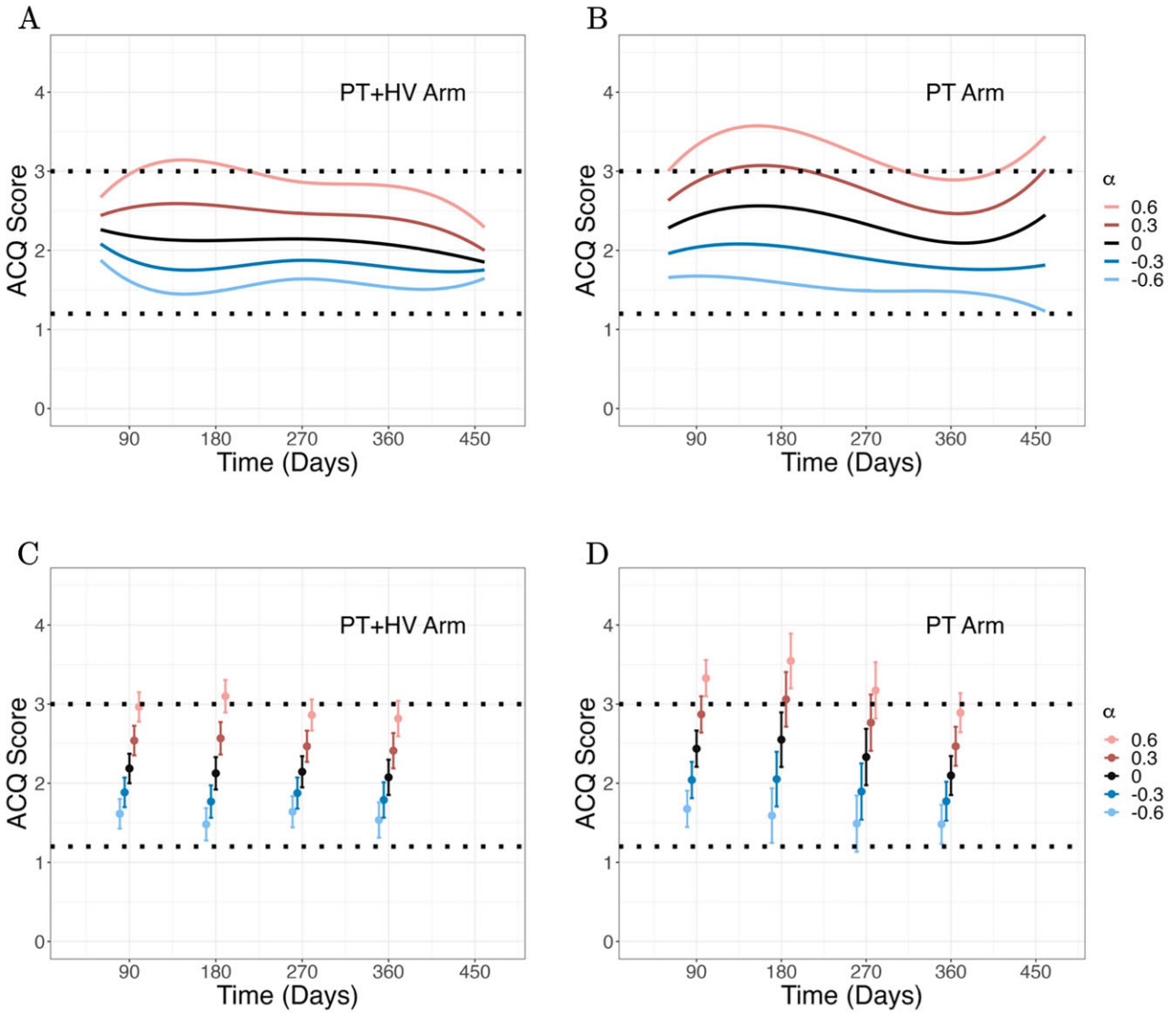

FIGURE 3 Estimated population means of scores on the Asthma Control Questionnaire (ACQ) in the Asthma Research for the Community (ARC) trial under a range of sensitivity parameter values. Estimation is made under values of $\alpha = -0.6, -0.3, 0, 0.3, 0.6$, where a positive (negative) value of α posits that unobserved outcome values at time t tend to be higher (lower) than observed values, after controlling for variables observed before time t. Upper panels: point estimates on the interval 60-460 days after randomization in the intervention (PT+HV) arm (panel A) and the control (PT) arm (panel B). Lower panels: point estimates and 95% Wald confidence intervals using the jackknife variance estimate at each target time of 90, 180, 270, and 360 days in the intervention arm (panel C) and the control arm (panel D). For each arm, only sensitivity parameter values under which the estimated curves lie completely between the dotted lines at $\mu_{\min} = 1.2$ and $\mu_{\max} = 3.0$ are considered plausible based on subject-matter expertise.

constructed the augmented inverse intensity weighted estimators given in Section 4.2.

Figure 3 shows estimates of the curve $\mu_1(t), t \in [60, 460]$, under a range of α_1 values and estimates of $\mu_0(t), t \in [60, 460]$, under a range of α_0 values, with higher $\mu_a(t)$ under higher values of α_a. Estimates and confidence intervals for $\mu_a(t)$ at the target times are also shown. The minimal clinically important difference for the ACQ Score is 0.5; therefore, one way of interpreting the magnitude of α_a in this study is that increasing α_a by 0.3 corresponds to an increase in $\mu_a(t)$ that, at some times t, is approximately as much as the minimal clinically important difference for the outcome. Next, we consider the ranges of α_1 and α_0 values to include. Our clinical collaborator (Author AJA) considered that a mean ACQ Score of 3.0 or higher, or 1.2 or lower, at any time would be extreme in either treatment arm. These bounds are shown in Figure 3. A value of $\alpha_1 > 0.52$ led to a value of $\mu_1(t)$ that was greater than 3.0, and a value of $\alpha_0 > 0.25$ led to a value of $\mu_0(t)$ that was greater than 3.0; therefore we use bounds of $-0.6 \leq \alpha_1 \leq 0.52$ and $-0.6 \leq \alpha_0 \leq 0.25$ in our final sensitivity analysis.

Figure 4 shows estimates and confidence intervals for $\delta(t)$ at 6 and 12 months under selected values of α_0 and α_1. If we assume $\alpha_0 = \alpha_1$ (which includes explainable assessment in each treatment arm), there is not enough evidence to conclude a

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

Month 6		α_0		
		-0.2	0	0.2
α_1	0.2	0.19 (-0.32, 0.71)	-0.14 (-0.71, 0.43)	-0.48 (-1.11, 0.15)
	0	-0.09 (-0.58, 0.40)	-0.43 (-0.97, 0.12)	**-0.77 (-1.38, -0.16)**
	-0.2	-0.34 (-0.81, 0.14)	**-0.67 (-1.20, -0.14)**	**-1.01 (-1.61, -0.41)**

Month 12		α_0		
		-0.2	0	0.2
α_1	0.2	**0.42 (0.05, 0.78)**	0.19 (-0.18, 0.57)	-0.05 (-0.43, 0.33)
	0	0.20 (-0.15, 0.55)	-0.02 (-0.38, 0.33)	-0.26 (-0.63, 0.10)
	-0.2	0.00 (-0.33, 0.34)	-0.22 (-0.56, 0.12)	**-0.46 (-0.81, -0.11)**

FIGURE 4 Treatment effect estimates from the Asthma Research for the Community (ARC) data under selected values of the sensitivity parameters α_1 and α_0. Shown are estimates (95% CI) for the treatment effect $\delta(t) = \mu_1(t) - \mu_0(t)$ at 6 months and 12 months. Highlighted entries in the lower right for Month 6 and Month 12 correspond to values of α_1 and α_0 under which there would be evidence that the home visits intervention reduces (that is, improves) the population mean score on the Asthma Control Questionnaire (ACQ) at that time, compared to portal training alone. The highlighted entry in the upper left for Month 12 corresponds to values under which there would be evidence that the intervention raises the population mean ACQ Score. Confidence intervals are Wald intervals using the jackknife variance estimate.

treatment effect at 6 months or at 12 months. However, if we consider the possibility that informative assessments may operate differently in each arm, then we do have evidence of a treatment effect in some cases. For example, if we assume $\alpha_0 = 0$, but assume that, in the intervention arm, unobserved values of the ACQ Score tend to be lower than observed values by an amount corresponding to $\alpha_1 = -0.2$, then there would be evidence that the home visits intervention improves (that is, reduces) the population mean ACQ Score at 6 months compared to portal training alone, by an estimated six tenths of a point. Estimation of $\delta(t)$ under a finer grid of (α_0, α_1) values is presented via the contour plots in Figure 5, showing point estimates and confidence interval information at 6 and 12 months. Point estimates range between -1.48 and 1.32 at 6 months and between -0.85 and 1.19 at 12 months. If α_1 and α_0 are similar, or if $\alpha_1 > \alpha_0$ and the difference in values no more than 0.5, then there is not enough evidence to conclude a treatment effect at 6 months. On the other hand, if $\alpha_0 > \alpha_1 + 0.2$, then in many cases there would be evidence that the intervention improves (reduces) the population mean ACQ Score relative to portal training alone at 6 months. There are also values of α_0 and α_1 under which there would be evidence that the population mean ACQ Score is higher (that is, worse) under the intervention at 6 months; however, informativeness would have to be strongly differential across treatment arms. Informativeness would also have to be strongly differential across arms to have evidence of either a positive or negative treatment effect at 12 months. In this study, such a large difference in the value of the sensitivity parameters between arms would likely not be plausible.

7 SIMULATIONS

We generated realistic simulated data based on the ARC data, with a sample size of $N = 200$ in each arm. Details of our data-generating process are given in Web Appendix C. Data were generated to follow our sensitivity analysis assumption (Assumption 1) with true α_1 and α_0 values of -0.6, -0.3, 0, 0.3, and 0.6 and analyzed using our augmented inverse intensity-weighted estimators. We first assessed the finite sample performance of our estimators by analyzing the simulated data using the true values of α_1 and α_0. To demonstrate the benefit of our approach by showing the dangers of not accounting for informative assessment times in the analysis, we also analyze the same simulated data using the explainable assessment assumption that $\alpha_1 = \alpha_0 = 0$ in each case. This explores the performance of an approach that relies on the explainable assessment assumption in cases where that assumption does not hold. Results for the treatment effects at 6 months and 12 months for each of these analyses are shown in Table 1. Results for the mean outcome in each treatment arm at 3, 6, 9, and 12 months (including the true values of each mean) are given in Web Appendix C. In the analysis estimating treatment effects using the true values of α_1 and α_0, the empirical bias over 500 simulations is small, with an absolute value of less than 0.05 in each case. Confidence interval coverage over 500 simulations is close to the nominal level of 0.95, ranging between 0.930 and 0.966. In the analysis assuming $\alpha_1 = \alpha_0 = 0$, in many cases the bias is large and confidence interval coverage is poor, including coverage as low as 0.594 even in cases with $\alpha_1 = \alpha_0$. This highlights the importance of

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

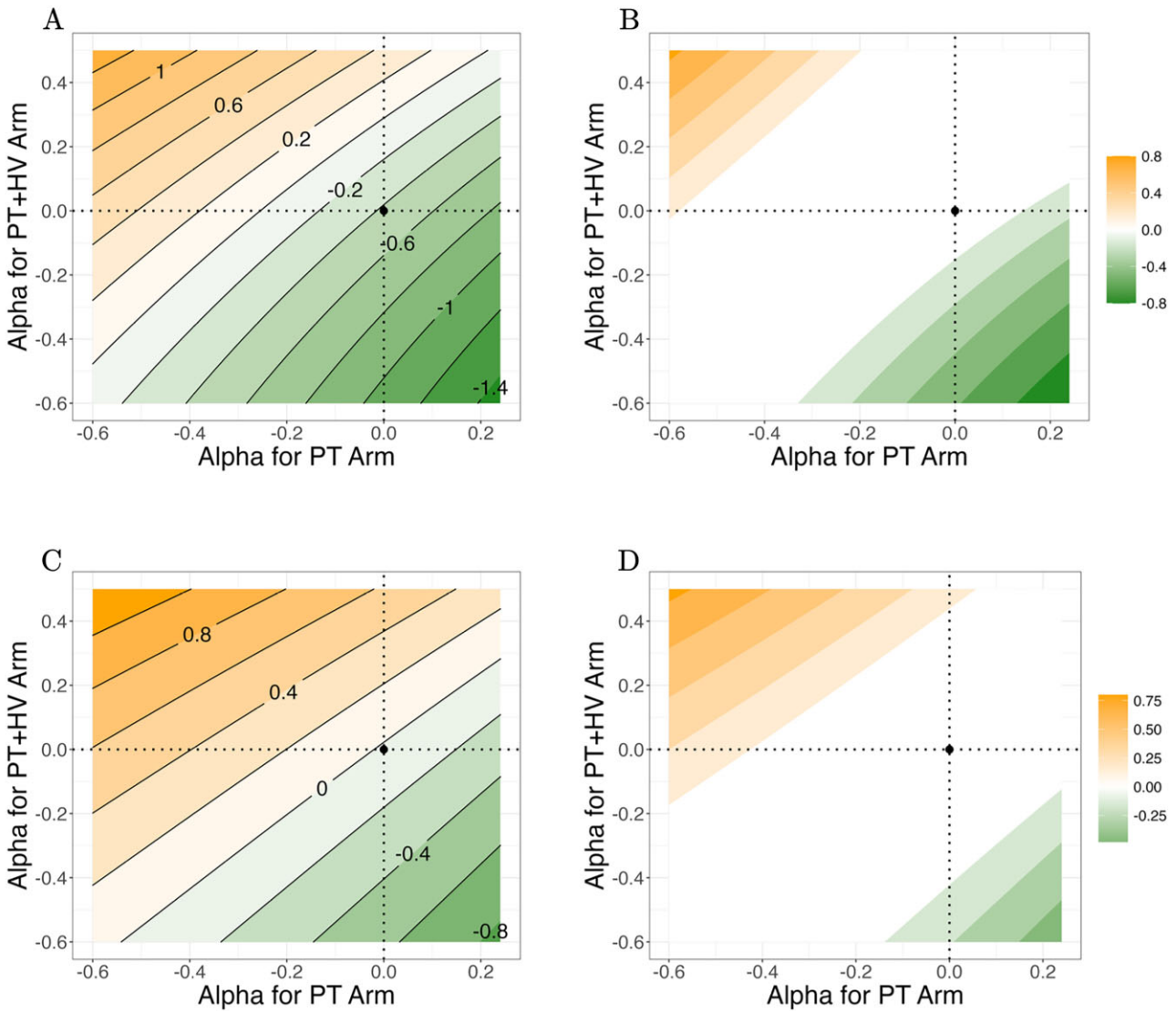

FIGURE 5 Sensitivity analysis for the Asthma Research for the Community (ARC) trial. Panels on the left show the point estimate of the treatment effect $\delta(t)$ at 6 months (panel A) and 12 months (panel C) under each pair of sensitivity parameter values. Point estimates vary between −1.48 and 1.32 at 6 months and between −0.85 and 1.19 at 12 months. Panels on the right display information about 95% confidence intervals for $\delta(t)$ at 6 months (panel B) and 12 months (panel D). The region in white corresponds to sensitivity parameter values under which the confidence interval contains zero, while the shaded region in the lower right (upper left) corresponds to values under which the confidence interval is entirely negative (positive); and the contour height is the value closest to zero that is inside the confidence interval. Confidence intervals are Wald intervals using the jackknife variance estimate.

considering a range of different assumptions through sensitivity analysis.

8 DISCUSSION

In many trials where the timing of outcome assessments varies by participant, assessment times may be related to underlying outcome values. This dependence can give misleading conclusions about the effect of treatment if not correctly accounted for in the analysis. Analysis methods for this setting make an untestable assumption about the informative assessment process; however, many assumptions can be consistent with the study data, and the treatment effect may differ across these assumptions. In this sense, researchers face 2 sources of uncertainty: the usual statistical uncertainty due to sample size and the unknown degree to which assessments may be informative. Our sensitivity analysis methodology provides researchers with a tool that accounts for both of these factors. By presenting inferences for treatment effects under a range of different assumptions, researchers will be able to provide a more accurate representation of the overall uncertainty in their study conclusions.

A detailed tutorial for the methodology presented here is included in the Supplemental Materials. Additionally, an R package implementing this methodology titled SensIAT has been published on CRAN (Redd et al., 2024), and the source code can be found at https://uofuepibio.github.io/SensIAT/.

Unfortunately, the question of whether assessment times are informative in a given study cannot necessarily be determined

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

TABLE 1 Simulation results. Data were generated under the sensitivity analysis assumption (Assumption 1) using values of $\alpha_0, \alpha_1 = -0.6, -0.3, 0, 0.3, 0.6$. The treatment effects at 6 and 12 months were then estimated using augmented inverse intensity weighted estimators: (a) using the true values of α_0, α_1 (rows denoted "S.A."), and (b) under the explainable assessment assumption that $\alpha_0 = \alpha_1 = 0$ (rows denoted "Expl."). Shown are the absolute values of the empirical bias and the confidence interval coverage across 500 simulations. Confidence intervals are Wald confidence intervals using the jackknife variance estimate.

			True α_0									
			−0.6		−0.3		0		0.3		0.6	
			|Bias|	Cov.	|Bias|	Cov.	|Bias|	Cov.	|Bias|	Cov.	|Bias|	Cov.
Month 6	0.6	S.A.	0.021	0.936	0.016	0.940	0.010	0.944	0.005	0.958	0.000	0.954
		Expl.	1.459	0.000	1.162	0.002	0.772	0.008	0.286	0.650	0.272	0.624
	0.3	S.A.	0.015	0.946	0.010	0.946	0.003	0.942	0.001	0.962	0.007	0.954
		Expl.	1.044	0.004	0.747	0.014	0.357	0.462	0.129	0.888	0.687	0.052
	0	S.A.	0.010	0.952	0.005	0.934	0.001	0.942	0.006	0.956	0.012	0.952
		Expl.	0.686	0.022	0.389	0.398	0.001	0.942	0.487	0.208	1.045	0.006
True α_1	−0.3	S.A.	0.005	0.954	0.000	0.944	0.006	0.954	0.011	0.954	0.016	0.938
		Expl.	0.390	0.396	0.093	0.912	0.297	0.576	0.783	0.028	1.341	0.002
	−0.6	S.A.	0.001	0.964	0.004	0.960	0.011	0.956	0.015	0.958	0.021	0.942
		Expl.	0.151	0.876	0.146	0.854	0.536	0.160	1.022	0.008	1.580	0.000
Month 12	0.6	S.A.	0.004	0.954	0.002	0.952	0.010	0.956	0.021	0.960	0.036	0.956
		Expl.	1.356	0.000	1.080	0.000	0.717	0.008	0.257	0.652	0.283	0.594
	0.3	S.A.	0.002	0.964	0.008	0.958	0.017	0.956	0.028	0.956	0.042	0.944
		Expl.	0.958	0.000	0.682	0.018	0.319	0.496	0.141	0.870	0.681	0.016
	0	S.A.	0.006	0.956	0.012	0.948	0.020	0.944	0.032	0.940	0.046	0.930
		Expl.	0.619	0.040	0.343	0.450	0.020	0.944	0.480	0.158	1.020	0.000
True α_1	−0.3	S.A.	0.009	0.960	0.015	0.956	0.023	0.938	0.034	0.938	0.048	0.938
		Expl.	0.340	0.458	0.064	0.938	0.299	0.568	0.759	0.008	1.299	0.000
	−0.6	S.A.	0.011	0.966	0.017	0.960	0.025	0.952	0.036	0.952	0.050	0.942
		Expl.	0.115	0.906	0.161	0.852	0.524	0.092	0.984	0.000	1.524	0.000

from the study data. In particular, assessment times may be informative even in studies without indications such as number of assessments varying by participant; differential timing of assessments across treatment arms; or timing of assessment impacted by the outcome at previous assessments. Substantive knowledge about the study should also be consulted in considering whether participants may be more (or less) likely to have an assessment when their outcome is worse. If investigators anticipate having irregular follow-up times in their study, they can consider conducting participant interviews to learn whether the reasons for missed, delayed, or early appointments were related to participants' outcomes. Participant responses could then be used to help assess whether sensitivity analysis is needed and inform the range of sensitivity parameter values to be included.

In this paper, we opted for an intensity modeling approach, as was used, for example, in Lin et al. (2004), Sun et al. (2007), Bůžková and Lumley (2007), and Liang et al. (2009). It would also be possible to develop a discrete-time version of our approach using pooled logistic regression with smoothing of the time-specific intercepts. Our estimation approach was developed for continuous outcomes and uses a mean model with the identity link (Assumption 3). Future work will generalize our work to other link functions, such as the logit link appropriate for binary outcomes. The ARC study had minimal dropout, and in our approach we have assumed that no participants are censored (though they may have fewer assessments than the study protocol specifies). Future work will relax this assumption. We also used an assumption of non-future dependence (Assumption 4), which may not be appropriate for some types of studies. This assumption is not needed for identification, but is used in our semi-parametric estimation approach. Finally, an important issue is selecting the range of sensitivity parameter values that will be included in the analysis. Here, we have used a bounding approach that uses domain experts' knowledge in a direct way without the need for additional assumptions; however, this approach may result in a wide range of values. A key direction for future research is to develop and study other bounding procedures, including methods that would incorporate participant interviews described above.

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

ACKNOWLEDGMENTS

We express our gratitude to Ming-Yueh Huang for sharing his code for the single index model. We thank Russell Localio for sharing his insights about the ARC trial and Eleanor Pullenayegum for helpful discussions. We are grateful to the anonymous referees and associate editor for valuable comments and suggestions that improved the manuscript.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices referenced in Sections 1, 2, 3, 4, and 7, as well as simulated data and code, are available with this paper at the Biometrics website on Oxford Academic.

FUNDING

Research reported in this article was funded through a Patient-Centered Outcomes Research Institute® (PCORI®) Award ME-2021C3-24972. Shu Yang, Yujing Gao, Andrea Apter and Daniel Scharfstein received funding from this award. The views and statements presented in this article are solely the responsibility of the authors and and do not necessarily represent the views of the PCORI®. The ARC study was funded through Patient-Centered Outcomes Research Institute® award AS-1307-05218. Bonnie Smith's research was partially supported by National Institutes of Health grant EB029977 from the National Institute of Biomedical Imaging and Bioengineering. Shu Yang received funding support from National Science Foundation grant NSF SES 2242776. Andrea Apter received funding support from National Institutes of Health grant NIH R18HL116285 from the National Heart, Lung, and Blood Institute, and Daniel Scharfstein received funding support from National Institutes of Health grant NIH R01DA046534 from the National Institute on Drug Abuse. Ravi Varadhan would like to acknowledge funding support from the National Institutes of Health grant NCI CCSG P30 CA006973 from the National Cancer Institute.

CONFLICT OF INTEREST

None declared.

DATA AVAILABILITY

The data from the Asthma Research for the Community study that support the findings in this paper will be shared on reasonable request to the corresponding author.

REFERENCES

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

Apter, A. J., Localio, A. R., Morales, K. H., Han, X., Perez, L., Mullen, A. N. et al. (2019). Home visits for uncontrolled asthma among low-income adults with patient portal access. *Journal of Allergy and Clinical Immunology*, 144, 846–853.e11.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Boca Raton, FL: Chapman and Hall.

Birmingham, J., Rotnitzky, A. and Fitzmaurice, G. M. (2003). Pattern–mixture and selection models for analysing longitudinal data with monotone missing patterns. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 275–297.

Breslow, N. E. (1972). Discussion on Professor Cox's paper. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34, 216–217.

Bůžková, P. and Lumley, T. (2007). 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, 485–500.

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

Chen, Y., Ning, J. and Cai, C. (2015). Regression analysis of longitudinal data with irregular and informative observation times. *Biostatistics*, 16, 727–739.

Chiang, C.-T. and Huang, M.-Y. (2012). New estimation and inference procedures for a single-index conditional distribution model. *Journal of Multivariate Analysis*, 111, 271–285.

Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82, 39–67.

Franks, A. M., D'Amour, A. and Feller, A. (2020). Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*, 115, 1730–1746.

Juniper, E., O'byrne, P., Guyatt, G., Ferrie, P. and King, D. (1999). Development and validation of a questionnaire to measure asthma control. *European Respiratory Journal*, 14, 902–907.

Kennedy, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In: *Statistical Causal Inferences and Their Applications in Public Health Research, ICSA Book Series in Statistics*(eds He, H., Wu, P. and Chen, D.-G. D.), 141–167. Cham: Springer International Publishing.

Kenward, M. G., Molenberghs, G. and Thijs, H. (2003). Pattern-mixture models with proper time dependence. *Biometrika*, 90, 53–71.

Liang, Y., Lu, W. and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics*, 65, 377–384.

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

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

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R. and Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, 58, 621–630.

National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: The National Academies Press.

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

Pullenayegum, E. M. and Lim, L. S. (2016). Longitudinal data subject to irregular observation: a review of methods with a focus on visit processes, assumptions, and study design. *Statistical Methods in Medical Research*, 25, 2992–3014.

Pullenayegum, E. M. and Scharfstein, D. O. (2022). Randomized trials with repeatedly measured outcomes: handling irregular and potentially informative assessment times. *Epidemiologic Reviews*, 44, 121–137.

Redd, A., Gao, Y., Yang, S., Smith, B., Varadhan, R. and Scharfstein, D. (2024). SensIAT: sensitivity analysis for irregular assessment times, *R package version 0.1.1.9000*. https://uofuepibio.github.io/SensIAT/.

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

Rotnitzky, A., Scharfstein, D., Su, T.-L. and Robins, J. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics*, 57, 103–113.

Scharfstein, D. O. and McDermott, A. (2019). Global sensitivity analysis of clinical trials with missing patient-reported outcomes. *Statistical Methods in Medical Research*, 28, 1439–1456.

Shen, W., Liu, S., Chen, Y. and Ning, J. (2019). Regression analysis of longitudinal data with outcome-dependent sampling and informative censoring. *Scandinavian Journal of Statistics*, 46, 831–847.

Sjölander, A., Gabriel, E. E. and Ciocănea-Teodorescu, I. (2022). Sensitivity analysis for causal effects with generalized linear models. *Journal of Causal Inference*, 10, 441–479.

Sun, J., Sun, L. and Liu, D. (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *Journal of the American Statistical Association*, 102, 1397–1406.

Sun, L., Mu, X., Sun, Z. and Tong, X. (2011a). Semiparametric analysis of longitudinal data with informative observation times. *Acta Mathematicae Applicatae Sinica, English Series*, 27, 29–42.

Sun, L., Song, X. and Zhou, J. (2011b). Regression analysis of longitudinal data with time-dependent covariates in the presence of informative observation and censoring times. *Journal of Statistical Planning and Inference*, 141, 2902–2919.

Sun, X., Peng, L., Manatunga, A. and Marcus, M. (2016). Quantile regression analysis of censored longitudinal data with irregular outcome-dependent follow-up. *Biometrics*, 72, 64–73.

Vansteelandt, S., Rotnitzky, A. and Robins, J. (2007). Estimation of regression models for the mean of repeated outcomes under nonignorable nonmonotone nonresponse. *Biometrika*, 94, 841–860.

Veitch, V. and Zaveri, A. (2020). Sense and sensitivity analysis: simple post-hoc analysis of bias due to unobserved confounding. In: *Advances in Neural Information Processing Systems*(eds Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F. and Lin, H.), Vol. 33, 10999–11009. Red Hook, NY: Curran Associates, Inc.

Wang, C. and Daniels, M. J. (2011). A note on MAR, identifying restrictions, model comparison, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data. *Biometrics*, 67, 810–818.

Wang, Z. (2020). Global sensitivity analysis for randomized trials with informative assessment times: a fully parametric approach. Master's thesis. Johns Hopkins University.

Downloaded from https://academic.oup.com/biometrics/article/80/4/ujae154/7932723 by NC State University Libraries user on 16 October 2025

Received: October 28, 2022; Revised: October 22, 2024; Accepted: November 27, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of The International Biometric Society. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com