

# Doubly robust omnibus sensitivity analysis of externally controlled trials with intercurrent events

Chenyin Gao<sup>1,3</sup>, Xiang Zhang<sup>2</sup>, Shu Yang<sup>1,\*</sup>

<sup>1</sup>Department of Statistics, North Carolina State University, Raleigh, U.S. 27695, <sup>2</sup>Medical Affairs and HTA Statistics, CSL Behring, King of Prussia, U.S. 19406, <sup>3</sup>Department of Biostatistics, Harvard University, Boston, U.S. 02115

\*Corresponding author: Shu Yang (Department of Statistics, North Carolina State University, Raleigh, 27695, U.S., [syang24@ncsu.edu](mailto:syang24@ncsu.edu)).

## ABSTRACT

Externally controlled trials are crucial in clinical development when randomized controlled trials are unethical or impractical. These trials consist of a full treatment arm with the experimental treatment and a full external control arm. However, they present significant challenges in learning the treatment effect due to the lack of randomization and a parallel control group. Besides baseline incomparability, outcome mean non-exchangeability, caused by differences in conditional outcome distributions between external controls and counterfactual concurrent controls, is infeasible to test and may introduce biases in evaluating the treatment effect. Sensitivity analysis of outcome mean non-exchangeability is thus critically important to assess the robustness of the study's conclusions against such assumption violations. Moreover, intercurrent events, which are ubiquitous and inevitable in clinical studies, can further confound the treatment effect and hinder the interpretation of the estimated treatment effects. This paper establishes a semi-parametric framework for externally controlled trials with intercurrent events, offering doubly robust and locally optimal estimators for primary and sensitivity analyses. We develop an omnibus sensitivity analysis that accounts for both outcome mean non-exchangeability and the impacts of intercurrent events simultaneously, ensuring root-n consistency and asymptotic normality under specified conditions. The performance of the proposed sensitivity analysis is evaluated in simulation studies and a real-data problem.

**KEYWORDS:** data heterogeneity; missing not at random; semi-parametric estimation; tilting models.

## 1 INTRODUCTION

### 1.1 Why consider external controls

In medical research, the gold standard for evaluating new treatments has been *randomized controlled trials* (RCTs). Regulatory bodies often require these rigorously controlled clinical studies to validate the effectiveness and safety of these treatments for specific patient groups. Despite the high regard for randomized, double-blind trials, they may not always be viable or ethical, particularly in rare or severe diseases with limited treatment options. Recruiting enough participants for such trials is often difficult, and using placebos in situations when alternative effective treatments are available may be unethical or impractical.

In these cases, *single-arm trials* (SATs) can be a practical alternative, though they come with limitations. Without direct comparison data for untreated subjects, researchers are compelled to infer these outcomes from external sources such as previous studies or real-world databases. These are referred to as *externally controlled trials* (ECTs). However, data from past trials might not always be relevant due to changes in the treatment landscape and patient demographics, reducing its evidence capacity compared with RCTs. Recently, real-world data have gained popularity for external control arms due to its accessibility and contemporaneity with treatment groups.

### 1.2 Estimand considerations with external controls

The Food and Drug Administration's (FDA) latest draft guidelines for natural history studies in rare disease drug development emphasized 5 critical concerns with using external controls (Food and Drug Administration, 2023). These concerns, outlined in Table 1, may introduce biases in research findings when real-world data are utilized as a control arm in ECTs. For a more structured analysis of these biases, we classified them into 2 principal categories: baseline incomparability and outcome mean non-exchangeability. Each category is characterized by unique mechanisms for introducing bias, as detailed in Table 1.

Besides the challenges of non-exchangeability of the external controls in clinical trials, managing intercurrent events such as participant dropout, non-compliance, or premature termination of therapy adds more complexities. Typically, the missingness at random (MAR) framework is adopted after the absence of outcomes following these events. While the MAR assumption is often deemed plausible, it remains unverifiable and may lack practical applicability since it assumes all participants persist with the study medication, addressing only a theoretical aspect of the treatment effect, as noted in the guidelines (ICH, 2021). A more plausible assumption would be that the treatment effect may quickly fade away, leading to the missing not at ran-

**TABLE 1** Key considerations of using external controls.

<b>(Observed) baseline incomparability</b>	
Covariate distribution shift	Systematic differences exist in the baseline characteristics of patients in external studies compared to those in SATs
<b>(Unobserved) outcome mean non-exchangeability</b>	
Unmeasured confounding	External control data might not capture the same detailed patient information as SATs, leading to biases from unknown or unmeasured factors
Lack of concurrency/ temporal bias	External control data and SATs may be collected in different time periods or under varying healthcare settings
Measurement error	There may exist potential inconsistencies in how patient information is collected and recorded, termed here as measurement errors in covariates
Outcome validity	Methods of measuring outcomes in external data sources might differ from those used in SATs, or the outcomes might not be clearly defined or reliable
Proper estimands	
Intercurrent events	Intercurrent events, such as stopping medication and/or adding rescue therapy, can confound the causal effects of the randomized treatment

dom (MNAR) pattern for the intercurrent events. The MNAR assumption can also be put forward for the sensitivity analysis, suggesting that any treatment effect observed, while a participant was active in the study is nullified upon their discontinuation. To evaluate this effect, the implementation of control-based imputation (CBI) models has been proposed, offering a nuanced and realistic assessment of treatment outcomes in clinical trials.

### 1.3 Primary analysis and doubly robust omnibus sensitivity analysis

Numerous statistical methodologies have been developed to address the potential biases from using external controls in ECTs. Most of these methods utilize techniques such as propensity score stratification, matching, or weighting to mitigate selection bias by balancing baseline covariates between external controls and SATs. Nonetheless, the outcome exchangeability of external controls cannot be verified with observed data due to the absence of concurrent controls. Therefore, sensitivity analyses become essential to evaluate the impact of potential violations of outcome non-exchangeability and to understand the effects of intercurrent events under different realistic scenarios.

In this context, we introduce a sensitivity analysis framework tailored for ECTs with intercurrent events. This framework encompasses models such as the tilting models, which control the degree of outcome mean non-exchangeability in external controls and alterations in outcome distributions after intercurrent events. Central to our framework is the establishment of identification results, the development of efficient influence functions (EIFs), and EIF-motivated tilting estimators under sensitivity models, which jointly capture the effect of changes due to outcome mean non-exchangeability and intercurrent events. These EIF-motivated estimators have several advantageous statistical properties, such as local efficiency, double robustness, and asymptotic normality. By analytically establishing the conditions for desirable asymptotic properties, our estimator allows flexible models for nuisance parameters while maintaining root-n consistency (Theorem 4). Therefore, our major contribution is to derive the locally efficient estimator for evaluating the treatment effect under the sensitivity models and to jointly assess the

robustness and reliability of multiple assumptions in ECTs with intercurrent events in a more efficient and flexible manner.

Our paper is organized as follows: Section 2 presents a brief overview of sensitivity analysis. Section 3 introduces the notation and develops the semi-parametric efficient estimator for the primary analysis. Section 4 details the tilting sensitivity models and the efficient inference for the sensitivity analysis. Another efficient estimator for the CBI sensitivity analysis under the jump-to-reference (J2R) model is presented in the Supplementary Materials. Section 5 discusses one practical method for choosing the sensitivity parameters by bounding their impacts. Extensive simulation studies for both primary and sensitivity analyses are presented in Section 6. Section 7 illustrates our approach with an antidepressant trial, and Section 8 concludes with a discussion.

## 2 RELATED WORKS

Before we delve into the proposed sensitivity analysis framework, we provide a review of sensitivity analysis methods. Causal inference involving observational studies usually requires no unmeasured confounding assumption, that is, the treatment assignment is ignorable conditional on a set of covariates (Rosenbaum and Rubin, 1983b; Imbens and Rubin, 2015). Yet, claiming the absence of confounders in the treatment-outcome relationship is untestable and often implausible in practice. Thus, it is advised to conduct a series of sensitivity analyses assessing how robust the causal findings are against the plausible violations of the unconfoundedness (Faries et al., 2024). The problem of sensitivity analyses has been studied in a variety of fields with the earliest work in Cornfield et al. (1959), which is later extended in Imbens (2003), Rosenbaum (1987), and Rosenbaum and Rubin (1983a). However, one concerning issue regarding this framework is that it demands a specific parametric assumption on the unmeasured confounder  $U$ . In many cases, sensitivity analyses can be quite fragile against the model misspecification of  $U$  as shown in Zhang and Tchetgen (2019).

To circumvent modeling the conditional (or marginal) distribution for the unmeasured confounder, a plethora of sensitivity approaches have been proposed, which preserve the critical elements in Imbens (2003) and Rosenbaum and Rubin

(1983a) and allow for flexible strategies to model the distribution of  $U$ . Zhang and Tchetgen (2019) leverage the modern semi-parametric theory to obtain a consistent estimation in a model while placing no distributional assumptions on  $U$ . The idea of partial misspecification of the nuisance parameters endows the framework with an unrestricted law of latent variables and has been similarly considered in the context of measurement error (Tsiatis and Ma, 2004), mixed models (Garcia and Ma, 2016), and statistical genetics (Allen et al., 2005). Another line of work to tackle this problem is based on the idea of “omitted-variable” bias, which can be computed easily without needs to specify the parametric form of the potential unobserved confounding. The idea of general bias formula is introduced in VanderWeele and Arah (2011), and is further extended in Cinelli and Hazlett (2020) with more flexibility and robustness.

As illustrated so far, most of the work blurs the line between sensitivity analysis and model checking by introducing sensitivity parameters under stringent parametric assumptions. For example, Little and Rubin (2019) suggest that the ignorability assumption in Heckman (1979) can be tested as the result of their Gaussian parametric assumption, thereby inducing testable implications of the untestable ignorability assumption. Moreover, the lack of such “clean” separation requires that the observed model to be refit for each setting of the sensitivity parameters, which will be an onerous task as modern non-parametric models may be adopted to fit the potential outcomes. To address this concern, Robins et al. (2000) propose and extended by Franks et al. (2020) and Nabi et al. (2024) to use the “tilting,” or “selection” function to decouple the sensitivity analyses from the observed data model. Typically, such sensitivity analysis specification does not impose any parametric assumptions on the distribution of the observed data or the unmeasured confounder, but only on a relaxed version of the unconfoundedness assumption:

$$f\{Y(a), A = 1 - a | X\} = f\{Y(a), A = a | X\} \frac{f\{A=1-a|Y(a), X; \psi\}}{f\{A=a|Y(a), X; \psi\}}, \quad (1)$$

where  $Y(a)$  is the potential outcome under treatment  $a$ ,  $A$  is the treatment assignment, and  $X$  is the baseline covariates. Here, the first term constitutes the observed data density, while the second term is the selection function governed by the sensitivity parameter  $\psi$ . Other approaches in Blackwell (2014) and Yang and Lok (2018) take a different track by representing the confounding as a function of the observed covariates, and describe the conditional potential outcomes difference varied by the treatment assignment as  $q(a, X; \alpha) = \mathbb{E}\{Y(a) | A = a, X\} - \mathbb{E}\{Y(a) | A = 1 - a, X\}$ , where the confounding function  $q$  is characterized by the single sensitivity parameter  $\alpha$ . As the observed data distribution is free of the sensitivity parameters ( $\psi$  or  $\alpha$ ), it achieves the “clean” factorization of the identified and unidentified parts of the sensitivity analysis framework. Veitch and Zaveri (2020), extending from Imbens (2003), posit a probabilistic model to bypass the need to specify any distributional assumption on  $U$ , which also decouples the sensitivity analyses from the observed data and leads to tractable bias calculations.

### 3 SINGLE-ARM TRIAL DATA WITH EXTERNAL CONTROLS: PRIMARY ANALYSIS

To ground ideas, we first focus on cross-sectional studies. Let  $S = 1$  denote trial participation, and the trial data consist of  $\{V_i = (X_i, A_i, R_i, Y_i, S_i = 1) : i \in \mathcal{R}\}$ , where  $R = 1$  indicates the absence of intercurrent events (eg, treatment discontinuation) during follow-up and 0 otherwise. Let  $S = 0$  denote the external participation, and the external data consist of  $\{V_i = (X_i, A_i, R_i, Y_i, S_i = 0) : i \in \mathcal{E}\}$ . Assume  $(X_i, A_i, R_i, Y_i, S_i)$  are independent and identically distributed, and we omit the subscript  $i$  for the simplicity of notation. Let  $V = (X, A, R, Y, S)$  be the random vector of all observed variables and follow the observed data distribution  $P_0$ . The *treatment policy* strategy designates a treatment effect estimand that measures the total effect of the treatment assignment and the intercurrent event on the outcome. Therefore, to define the estimand unambiguously, we extend the causal framework in Lipkovich et al. (2020) and introduce the potential outcomes framework for  $R$  and  $Y$ . Denote  $R(a)$  as the potential indicator for the absence of intercurrent events under treatment  $a$ ,  $Y(a, r)$  as the potential outcome under treatment  $a$  with intercurrent event status  $r$ , and  $Y(a) = Y\{a, R(a)\}$ .

The estimand of interest is the average treatment effect (ATE) for the trial, defined as

$$\begin{aligned} \tau &= \mathbb{E}[Y\{1, R(1)\} - Y\{0, R(0)\} | S = 1] \\ &= \mathbb{E}\{Y(1) - Y(0) | S = 1\}, \end{aligned}$$

which is a *treatment policy* estimand, as the occurrence of intercurrent events is considered irrelevant. However, due to the missing outcomes following these events, the *treatment policy* strategy cannot be implemented for intercurrent events that are terminal, such as treatment discontinuation. Table 2(A) outlines several key causal assumptions to identify the treatment effect for the primary analysis.

Assumptions 1 and 2 are standard causal assumptions for identification (Rosenbaum and Rubin, 1983b). Assumption 1 maps the potential outcomes to the observed data, and Assumption 2 ensures that each participant has a positive probability of being recruited into the SAT or external controls, and not having intercurrent events. Assumption 3 states that the conditional expectation of  $Y(0)$  is the same for the trial and the external controls. Assumptions 4 and 5 imply that the intercurrent events occur at random for SAT and the external controls, respectively. These assumptions are satisfied if the covariates  $X$  capture all the confounding variables. Take a weight-loss trial for an example, where the intercurrent events are the non-compliance with the prescribed diet. If we assume that the covariates, such as age, baseline weight, and other lifestyle factors, capture all the confounding variables, it follows that given the same covariates for 2 participants, they have the same likelihood of being non-compliant with the diet, regardless of their weight loss. Therefore, we can conclude that the weight loss will not be affected by the occurrence of the non-compliance conditional on these covariates, and the post-intercurrent event outcomes are exchangeable to the observed outcomes.

TABLE 2 Lists of (A) key assumptions for primary analysis (B) necessary notation.

(A) Assumptions	Details
1. Causal consistency	$R = R(A)$ , and $Y = Y\{A, R(A)\}$
2. Positivity	$P(S = 1   X) > 0$ and $P(R = 1   X, S = s) > 0$ for $s = 0, 1$
3. Outcome exchangeability of the external controls	$\mathbb{E}\{Y(0)   X, S = 1\} = \mathbb{E}\{Y(0)   X, S = 0\} = \mu_0(X)$
4. Ignorability of intercurrent event for SAT data	$R(a) \perp Y(a, r)   X, S = 1$ , for all $a, r$
5. Ignorability of intercurrent event for external controls	$R(a) \perp Y(a, r)   X, S = 0$ , for all $a, r$
(B) Formula	Details
$\pi_S(X)$	participation propensity, defined as $\pi_S(X) = P(S = 1   X)$
$q_S(X)$	participation propensity density ratio, defined as $q_S(X) = \pi_S(X) / \{1 - \pi_S(X)\}$
$\pi_{R_s}(X)$	propensity of not having intercurrent event for $s = 0, 1$ , defined as $\pi_{R_s}(X) = P(R = 1   X, S = s)$
$q_{R_s}(X)$	propensity density ratio of intercurrent event for $s = 0, 1$ , defined as $q_{R_s}(X) = \{1 - \pi_{R_s}(X)\} / \pi_{R_s}(X)$
$\mu_s(X)$	outcome means for $s = 0, 1$ , defined as $\mu_s(X) = \mathbb{E}(Y   X, S = s, R = 1)$
$c(X; \gamma_{R_0}), c(X; \gamma_{R_1}), c(X; \gamma_S), c(X; \gamma_S + \gamma_{R_0})$	Normalizing terms, defined as $c(X; \gamma_{R_0}) = \mathbb{E}[\exp\{\gamma_{R_0} Y(0)\}   X, S = 0, R = 1]$ , $c(X; \gamma_{R_1}) = \mathbb{E}[\exp\{\gamma_{R_1} Y(1)\}   X, S = 1, R = 1]$ , $c(X; \gamma_S) = \mathbb{E}[\exp\{\gamma_S Y(0)\}   X, S = 0, R = 1]$ , $c(X; \gamma_S + \gamma_{R_0}) = \mathbb{E}[\exp\{(\gamma_S + \gamma_{R_0}) Y(0)\}   X, S = 0, R = 1]$
$b(X; \gamma_{R_0}), b(X; \gamma_{R_1}), b(X; \gamma_S), b(X; \gamma_S + \gamma_{R_0})$	Tilted outcome means, defined as $b(X; \gamma_{R_0}) = \mathbb{E}[Y \exp\{\gamma_{R_0} Y(0)\}   X, S = 0, R = 1]$ , $b(X; \gamma_{R_1}) = \mathbb{E}[Y \exp\{\gamma_{R_1} Y(1)\}   X, S = 1, R = 1]$ , $b(X; \gamma_S) = \mathbb{E}[Y \exp\{\gamma_S Y(0)\}   X, S = 0, R = 1]$ , $b(X; \gamma_S + \gamma_{R_0}) = \mathbb{E}[Y \exp\{(\gamma_S + \gamma_{R_0}) Y(0)\}   X, S = 0, R = 1]$
$d(X; \gamma_{R_0}, \gamma_S), e(X; \gamma_{R_0}, \gamma_S)$	$d(X; \gamma_{R_0}, \gamma_S) = \pi_{R_0}(X)b(X; \gamma_S)c(X; \gamma_{R_0}) + \{1 - \pi_{R_0}(X)\}b(X; \gamma_S + \gamma_{R_0})$ , $e(X; \gamma_{R_0}, \gamma_S) = \pi_{R_0}(X)c(X; \gamma_S)c(X; \gamma_{R_0}) + \{1 - \pi_{R_0}(X)\}c(X; \gamma_S + \gamma_{R_0})$

Theorem 1 provides 3 identification formulas for  $\tau$  under the assumptions in Table 2(A), and Table 2(B) summarizes the necessary models for the identification.

**Theorem 1 (Identification):** Under the assumptions in Table 2,  $\tau$  is identifiable by

(a) trial participation propensity and outcome means:

$$\tau = \frac{\mathbb{E}\{\pi_S(X)\mu_1(X) - \pi_S(X)\mu_0(X)\}}{P(S = 1)},$$

(b) trial participation propensity and response propensity:

$$\tau = \frac{1}{P(S = 1)} \mathbb{E} \left\{ \frac{SRY}{\pi_{R_1}(X)} - \frac{(1 - S)Rq_S(X)Y}{\pi_{R_0}(X)} \right\},$$

(c) response propensity and outcome mean:

$$\tau = \frac{1}{P(S=1)} \mathbb{E} \{ S\pi_{R_1}(X)Y + S\{1 - \pi_{R_1}(X)\}\mu_1(X) - S\mu_0(X) \}.$$

We give some intuitions behind these identification formulas. Theorem 1(a) describes that the individual treatment effect given the covariates  $X$  is  $\mu_1(X) - \mu_0(X)$ . Taking the expectation over the trial population yields the identification for ATE. Theorem 1(b) can be understood as a transportability problem. The first term, corresponding to  $\pi_S(X)\mu_1(X)$ , adjusts the outcomes from SAT  $SY$  by  $R/\pi_{R_1}(X)$ , which weights the observed subjects by their response propensity. The second term, corresponding to  $\pi_S(X)\mu_0(X)$ , adjusts the outcomes  $(1 - S)Y$  from the external controls by  $Rq_S(X)/\pi_{R_0}(X)$ , which trans-

ports the external controls to the trial via the density ratio  $q_S(X)$  after response propensity weighting. In Theorem 1(c), it predicts the outcomes for the treatment group by  $\pi_{R_1}(X)Y + \{1 - \pi_{R_1}(X)\}\mu_1(X)$ , which imputes the post-intercurrent event outcomes by  $\mu_1(X)$ . Similarly, it predicts the outcomes of the external controls by  $\mu_0(X)$ . The difference between these 2 predictions marginalized over the trial population quantifies the ATE.

Based on these identification formulas, infinitely many estimators can be constructed. To develop a more principled estimator, we derive the EIF for  $\tau$ , and the resulting EIF-motivated tilting estimator achieves the rate double robustness, local efficiency, and asymptotic normality. The details of the estimator and these properties are relegated to Theorems S1 and S2 in the Supplementary Materials.

## 4 SENSITIVITY ANALYSIS UNDER TILTING MODELS

### 4.1 Assumptions and a graphical representation

Assumptions 3-5 in Table 2 are critical for the identification of  $\tau$ . However, these assumptions may be subject to violations in practice and are unverifiable based on the observed data. Here, we develop the tilting sensitivity models to assess whether the primary analysis result is sensitive to the violation of these assumptions.

*Model 1 (Tilting sensitivity models):* Assume that the tilting models for EC outcome mean non-exchangeability and the effects of intercur-

rent events are

$$dF\{Y(s, 0) | X, S = s, R = 0\} = dF\{Y(s, 1) | X, S = s, R = 1\} \frac{\exp\{\gamma_{R_1} Y(s)\}}{c(X; \gamma_{R_1})},$$

$$dF\{Y(0) | X, S = 1\} = dF\{Y(0) | X, S = 0\} \frac{\exp\{\gamma_S Y(0)\}}{\mathbb{E}[\exp\{\gamma_S Y(0)\} | X, S = 0]},$$

for  $s = 0, 1$ , where the normalizing terms  $c(X; \gamma_{R_i})$  are defined in Table 2(B).

Model 1 posits that each unobserved outcome distribution is a “tilted version” of the observed outcomes, where  $\gamma_S$ ,  $\gamma_{R_0}$ , and  $\gamma_{R_1}$  are treated as the sensitivity parameters, entailing the level of EC outcome non-exchangeability and the effect of intercurrent events within each arm; see Figure 1 for an illustration when Assumptions 3-5 in Table 2 are violated due to unmeasured confounders  $U_S$ ,  $U_{R_0}$ , and  $U_{R_1}$  under the tilting sensitivity models.

Given negative (or positive) sensitivity parameters, the unobserved outcome distribution is tilted to the left (or right) relative to the distribution of observed outcomes, with smaller (or larger) values receiving greater weight. For example, if  $\gamma_S$  is smaller (or larger) than 0, Model 1 implies that trial participants, if untreated, tend to have smaller (or larger) outcomes compared to the external controls, given the same covariates. Similarly, if  $\gamma_{R_0}$  and  $\gamma_{R_1}$  are both smaller (or larger) than 0, Model 1 implies that participants with intercurrent events tend to have smaller (or larger) outcomes compared to those without such events, given the same covariates. In particular,  $\gamma_S = 0$  leads to Assumption 3,  $\gamma_{R_0} = 0$  leads to Assumption 4, and  $\gamma_{R_1} = 1$  leads to Assumption 5.

**Remark 1 (Logistic selection):** The tilting sensitivity models are motivated by the logistic model for the binary indicators  $S$  and  $R$ . Assume the log odds of being in the trial are linear in  $Y(0)$  and  $X$  under the logistic selection specification:

$$P(S = 1 | Y(0), X) = \text{logit}^{-1}\{\alpha_S(X) + \gamma_S Y(0)\}, \quad (2)$$

where  $\text{logit}^{-1}(x) = \{1 + \exp(-x)\}^{-1}$ , and  $\alpha_S(X)$  can be identified by the observed data once  $\gamma_S$  is specified. Using the Bayes rule, the unobserved outcome distribution  $f(Y(0) | S = 1, X)$  is a “tilting” version of the observed outcomes:

$$f(Y(0) | S = 1, X) = f(Y(0) | S = 0, X) \frac{P(S = 0 | X)}{P(S = 1 | X)} \times \frac{P(S = 1 | Y(0), X)}{P(S = 0 | Y(0), X)}$$

$$= f(Y(0) | S = 0, X) \frac{\exp\{\gamma_S Y(0)\}}{\mathbb{E}[\exp\{\gamma_S Y(0)\} | X, S = 0]},$$

which is free of  $\alpha_S(X)$ ; similar logistic selection specifications can be applied to model the indicator of the intercurrent event within the SAT and external controls as well.

Remark 1 implies that our tilting sensitivity model, also known as the exponential tilting model, is connected to the logistic selection specification (2) with flexible formulation, including many non-parametric models, such as sieve approximation, Dirichlet process mixtures, and Bayesian additive regression trees. The logistic selection model, despite its drawbacks noted in Copas and Li (1997), is widely used to assess selection bias in missing data (Robins et al., 2000; Dahabreh et al., 2023) and

to conduct sensitivity analyses for unmeasured confounding in causal inference (Franks et al., 2020; Nabi et al., 2024).

## 4.2 Identification and EIF

The following theorem establishes the non-parametric identification of  $\tau$  when the sensitivity parameters  $\gamma_S$ ,  $\gamma_{R_0}$ , and  $\gamma_{R_1}$  are fixed.

**Theorem 2 (Identification under tilting sensitivity models):** Under Assumptions 1 and 2 in Table 2, and Model 1 with fixed  $\gamma_{R_0}$ ,  $\gamma_{R_1}$ , and  $\gamma_S$ , the following identification formula holds for  $\tau$ :

$$\tau = \frac{1}{P(S = 1)} \mathbb{E} \left[ \pi_S(X) \pi_{R_1}(X) \mu_1(X) + \pi_S(X) \{1 - \pi_{R_1}(X)\} \frac{b(X; \gamma_{R_1})}{c(X; \gamma_{R_1})} \right]$$

$$- \frac{1}{P(S = 1)} \mathbb{E} \left[ \frac{\pi_S(X) d(X; \gamma_{R_0}, \gamma_S)}{e(X; \gamma_{R_0}, \gamma_S)} \right],$$

where the tilted outcomes  $b(X; \gamma_{R_1})$  and  $d(X; \gamma_{R_0}, \gamma_S)$  and the normalizing terms  $c(X; \gamma_{R_1})$  and  $e(X; \gamma_{R_0}, \gamma_S)$  are defined in Table 2(B).

The identification formula in Theorem 2 is derived under the same logic as Theorem 1. Under the tilting sensitivity model, we have

$$b(X; \gamma_{R_1})/c(X; \gamma_{R_1}) = \mathbb{E}\{Y(1) | X, S = 1, R = 0\}, \quad (3)$$

$$d(X; \gamma_{R_0}, \gamma_S)/e(X; \gamma_{R_0}, \gamma_S) = \mathbb{E}\{Y(0) | X, S = 1\}, \quad (4)$$

where (3) reduces to  $\mu_1(X)$  when  $\gamma_{R_1} = 0$  as the intercurrent events occur at random for SAT, and (4) reduces to  $\mu_0(X)$  when  $\gamma_{R_0} = \gamma_S = 0$  as external controls are exchangeable to SAT and the intercurrent events occur at random for external controls. Similarly, we can derive the EIF for  $\tau$  under Model 1 to motivate the semi-parametric efficient estimator.

**Theorem 3 (EIF under tilting sensitivity models):** Under the assumptions in Theorem 2, the EIF for  $\tau$  with fixed  $\gamma_{R_0}$ ,  $\gamma_{R_1}$ , and  $\gamma_S$  is

$$\phi_{\text{eff}}^t(V; P_0, \gamma_S, \gamma_{R_0}, \gamma_{R_1}) = \frac{SRY}{P(S=1)}, \quad (5)$$

$$+ \frac{S}{P(S=1)} \left\{ (1 - R) \frac{b(X; \gamma_{R_1})}{c(X; \gamma_{R_1})} + Rq_{R_1}(X)g(V; \gamma_{R_1}) \right\}, \quad (6)$$

$$- \frac{1}{P(S = 1)} \left\{ S \frac{d(X; \gamma_{R_0}, \gamma_S)}{e(X; \gamma_{R_0}, \gamma_S)} + (1 - S)q_S(X)h(V; \gamma_{R_0}, \gamma_S) \right\}$$

$$- \frac{S\tau}{P(S = 1)}, \quad (7)$$

where the augmentation terms  $\mathbb{E}\{g(V; \gamma_{R_1})\} = 0$  and  $\mathbb{E}\{h(V; \gamma_{R_0}, \gamma_S)\} = 0$  with detailed definitions in the Supplementary Materials.

The EIF in Theorem 3 is constituted by 3 parts. The first part (5) is contributed by the SAT with no intercurrent events (ie,  $S = 1, R = 1$ ); the second part (6) is contributed by the SAT with intercurrent events. When  $\gamma_{R_1} = 0$ , indicating the intercurrent event occurs at random for SAT,  $g(V; \gamma_{R_1})$  equals to  $Y - \mu_1(X)$ , and part (6) reduces to (S10) for primary analysis; the third part (7) is contributed by the external controls. When  $\gamma_{R_0} = \gamma_S = 0$ , indicating external controls are exchangeable to SAT and the intercurrent event occurs at random for external controls,  $h(V; \gamma_{R_0}, \gamma_S)$  equals to  $R\{Y - \mu_0(X)\}/\pi_{R_0}(X)$ , and

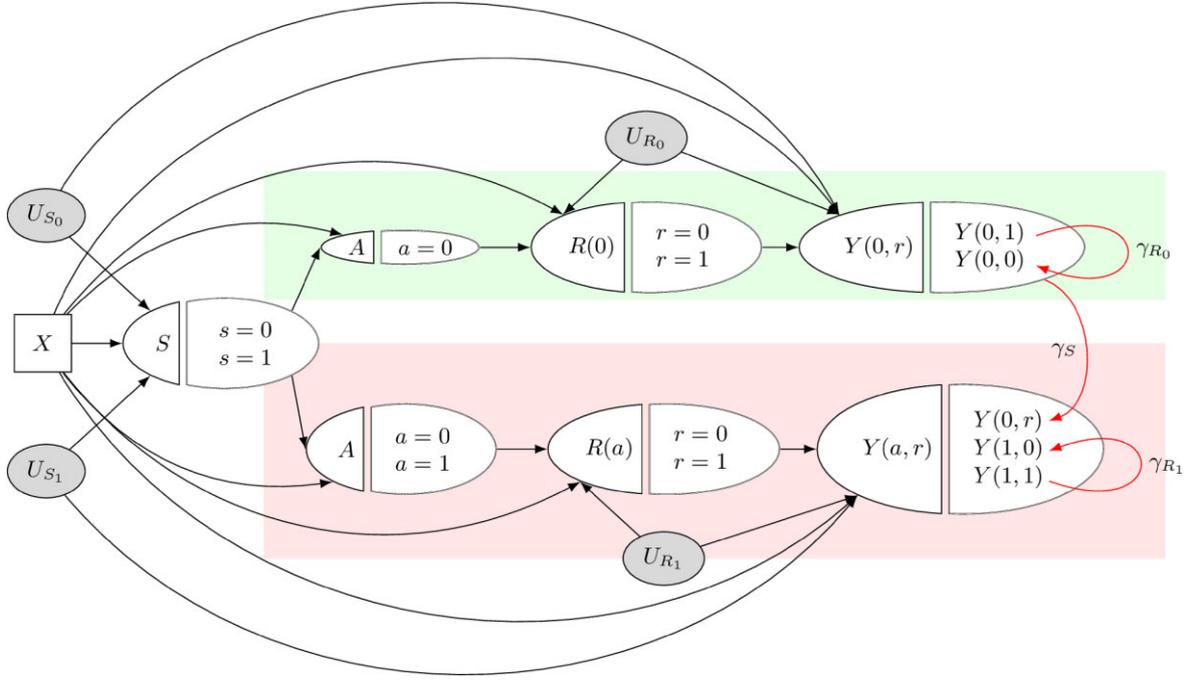


FIGURE 1 Schematic plot of the tilting sensitivity models subject to unmeasured confounders  $U_S, U_{R_0}$ , and  $U_{R_1}$ .

the part (7) reduces to (S11). Next, we construct an estimator for  $\tau$  by solving the empirical mean of  $\phi_{\text{eff}}^t(V; P_0, \gamma_S, \gamma_{R_0}, \gamma_{R_1})$  with  $P_0$  replaced by its estimated counterpart, and present the asymptotic properties for  $\hat{\tau}^t$  in Theorem 4.

**Theorem 4:** Under the assumptions in Theorem 3 and other regularity conditions in Assumption S1, we have

$$\begin{aligned} \hat{\tau}^t &= \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R}} R_i Y_i + \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R}} \left[ (1 - R_i) \frac{\hat{b}(X_i; \gamma_{R_1})}{\hat{c}(X_i; \gamma_{R_1})} + R_i \hat{q}_{R_1}(X_i) \hat{g}(V_i; \gamma_{R_1}) \right] \\ &\quad - \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R}} \frac{\hat{d}(X_i; \gamma_{R_0}, \gamma_S)}{\hat{c}(X_i; \gamma_{R_0}, \gamma_S)} - \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{E}} \hat{q}_S(X_i) \hat{h}(V_i; \gamma_{R_0}, \gamma_S). \\ &= \tau + \frac{1}{N} \sum_{i \in \mathcal{R} \cup \mathcal{E}} \phi_{\text{eff}}^t(V_i; P_0, \gamma_S, \gamma_{R_0}, \gamma_{R_1}) + \|\text{Rem}^t(\hat{P}, P)\|_{L_2} + o_p(N^{-1/2}), \end{aligned}$$

where  $\|\text{Rem}^t(\hat{P}, P_0)\|_{L_2}$  is bounded by

$$\begin{aligned} &\left\{ \|\hat{q}_{R_1}(X) - q_{R_1}(X)\|_{L_2} + \|\hat{c}(X; \gamma_{R_1}) - c(X; \gamma_{R_1})\|_{L_2} \right\} \times \left\{ \|\hat{c}(X; \gamma_{R_1}) - c(X; \gamma_{R_1})\|_{L_2} + \|\hat{b}(X; \gamma_{R_1}) - b(X; \gamma_{R_1})\|_{L_2} \right\} \\ &+ \sum_{\gamma \in \{\gamma_{R_0}, \gamma_S, \gamma_S + \gamma_{R_0}\}} \|\hat{b}(X; \gamma) - b(X; \gamma)\|_{L_2} \\ &\times \left\{ \|\hat{q}_{R_0}(X) - q_{R_0}(X)\|_{L_2} + \|\hat{q}_S(X) - q_S(X)\|_{L_2} + \sum_{\gamma \in \{\gamma_{R_0}, \gamma_S, \gamma_S + \gamma_{R_0}\}} \|\hat{c}(X; \gamma) - c(X; \gamma)\|_{L_2} \right\} \\ &+ \left\{ \|\hat{q}_S(X) - q_S(X)\|_{L_2} + \sum_{\gamma \in \{\gamma_{R_0}, \gamma_S, \gamma_S + \gamma_{R_0}\}} \|\hat{c}(X; \gamma) - c(X; \gamma)\|_{L_2} \right\} \\ &\times \left\{ \sum_{\gamma \in \{\gamma_{R_0}, \gamma_S, \gamma_S + \gamma_{R_0}\}} \|\hat{c}(X; \gamma) - c(X; \gamma)\|_{L_2} + \sum_{\gamma \in \{\gamma_{R_0}, \gamma_S, \gamma_S + \gamma_{R_0}\}} \|\hat{b}(X; \gamma) - b(X; \gamma)\|_{L_2} \right\} \end{aligned}$$

up to some multiplicative constants.

Theorem 4 shows that  $\widehat{\tau}^t$  is root- $n$  consistent and asymptotically normal for fixed sensitivity parameters  $\gamma_S$ ,  $\gamma_{R_0}$ , and  $\gamma_{R_1}$  when the remainder term  $N^{1/2} \|\text{Rem}^t(\widehat{P}, P_0)\|_{L_2}$  is  $o_{\mathbb{P}}(1)$ . Intuitively, when  $\gamma_S = \gamma_{R_1} = \gamma_{R_0} = 0$ , we have  $\|\widehat{c}(X; \gamma) - c(X; \gamma)\|_{L_2} = 0$  for any  $\gamma$ , and

$$\begin{aligned} \|\widehat{b}(X; \gamma_{R_1}) - b(X; \gamma_{R_1})\|_{L_2} &= \|\widehat{\mu}_1(X) - \mu_1(X)\|_{L_2}, \\ \sum_{\gamma \in \{\gamma_{R_0}, \gamma_S, \gamma_{R_1}, \gamma_{R_0}\}} \|\widehat{b}(X; \gamma) - b(X; \gamma)\|_{L_2} &= \|\widehat{\mu}_0(X) - \mu_0(X)\|_{L_2}. \end{aligned}$$

Thus, the remainder term reduces to

$$\begin{aligned} \|\text{Rem}^t(\widehat{P}, P_0)\|_{L_2} &\lesssim \|\widehat{q}_{R_1}(X) - q_{R_1}(X)\|_{L_2} \times \|\widehat{\mu}_1(X) - \mu_1(X)\|_{L_2} \\ &+ \|\widehat{\mu}_0(X) - \mu_0(X)\|_{L_2} \times \{\|\widehat{q}_{R_0}(X) - q_{R_0}(X)\|_{L_2} + \|\widehat{q}_S(X) - q_S(X)\|_{L_2}\}, \end{aligned}$$

which is at the same convergence rate as  $\|\text{Rem}(\widehat{P}, P_0)\|_{L_2}$  for the primary analysis in Theorem S2. The remainder term  $\|\text{Rem}^t(\widehat{P}, P_0)\|_{L_2}$  suggests that the error of  $\widehat{\tau}^{[t]}$  is only affected by the estimation errors of the nuisance models in second-order terms. Therefore,  $\widehat{\tau}^{[t]}$  is more robust and remains consistent when flexible machine learning methods used for nuisance estimation converge at rates faster than  $N^{-1/4}$ . This condition is satisfied by some machine learning methods (Kennedy, 2016; Bradic et al., 2019), which is the so-called rate double robustness (Chernozhukov et al., 2018). Note that the conditional expectations  $b(X; \gamma)$  and  $c(X; \gamma)$  are critical to obtain accurate tilting estimates for the sensitivity analysis. In general, these conditional expectations require heavy computation or strong restrictions on the outcome model. Fortunately, these terms are analytically tractable when the observed outcome model belongs to the class of exponential family mixtures, for example, the Dirichlet processes mixture models (Dorie et al., 2016).

**Remark 2 (Exponential family mixtures):** Let the observed outcome follow  $Y(0) | S = 0, R = 1, X \sim \sum_k \pi_k \mathcal{N}(\mu_{0k}(X), \sigma_{0k}^2(X))$ . Under the tilting sensitivity models, we can show that  $c(X; \gamma_{R_0}) = \sum_k \pi_k \exp\{\mu_{0k}(X)\gamma_{R_0} + \gamma_{R_0}^2 \sigma_{0k}^2(X)/2\}$ , and  $b(X; \gamma_{R_0}) = \sum_k \pi_k \{\mu_{0k}(X) + \gamma_{R_0} \sigma_{0k}^2(X)\} \exp\{\mu_{0k}(X)\gamma_{R_0} + \gamma_{R_0}^2 \sigma_{0k}^2(X)/2\}$ . Analogously, these conditional expectations are analytically obtainable if an invertible function of  $Y$  follows the exponential family mixtures (eg, Box-Cox transformation); other advanced methods are also available to compute the conditional expectations in exchange for heavy computation, for example, modeling the conditional distribution of the observed data (Chiang and Huang, 2012).

### 5 CALIBRATING SENSITIVITY PARAMETERS

The magnitude of the sensitivity parameters indicates the strength of the non-ignorability of the indicators ( $R, S$ ) given the covariates, which is commonly caused by the existence of unmeasured confounders. However, it is practically infeasible to identify the sensitivity parameters with the observed data. Furthermore, assessing whether a confounder with such strength plausibly exists, given the prior knowledge and domain expertise, is arguably challenging as well. While sensitivity parameters are not directly identifiable from the data, it is

reasonable to bound their relative strength using the observed data.

Following the calibration approach proposed by Franks et al. (2020), we introduce a method to determine the plausible quantities for sensitivity analysis based on the observed data. Assume the logistic selection specification outlined in Remark 1 holds. Next, we assume that the relative strength of the unmeasured confounder cannot exceed that of the observed covariates, meaning it should not account for more variation of the indicators as the most important covariate. To measure the relative strength, we adopt the ‘‘implicit  $R^2$ ’’ concept from Imbens (2003), which generalizes variance-explained measures to the case of binary outcomes. For example, the partial variances  $\rho_{Y(0)|X}^2$  explained by  $Y(0)$  given  $X$  is

$$\begin{aligned} \rho_{Y(0)|X}^2 &= \frac{\sigma_Y^2 \gamma_S^2}{\text{var}\{m_S(X)\} + \pi^2/3 + \sigma_Y^2 \gamma_S^2}, \\ P(S = 1 | X) &= \text{logit}^{-1}\{m_S(X)\}, \end{aligned}$$

where  $\sigma_Y^2 = \mathbb{E}[\text{var}\{Y(0) | X, S = 0\}]$ . Then, we propose a target value  $(\rho^*)^2$  for the unidentified  $\rho_{Y(0)|X}^2$  using the observed data. In specific, we compute the partial variance explained by each covariate  $X_j$  given all other covariates  $X_{-j}$ , and set  $(\rho^*)^2 = \max_j \rho_{X_j|X_{-j}}^2 / (1 - \max_j \rho_{X_j|X_{-j}}^2)$ . Here,  $(\rho^*)^2$  represents the maximum partial variance explained by adding one covariate  $X_j$  to the others, relative to the baseline variance that needs to be explained, referred to as the partial Cohen’s  $f$  in Cinelli and Hazlett (2020). Setting  $\rho_{Y(0)|X}^2 = (\rho^*)^2$  allows us to calibrate  $\gamma_S$  by  $\gamma_S^*$ , which implies that the information gained by adding  $Y(0)$  to  $X$  as a predictor of  $S$  is comparable to the maximum information gain by the most important covariate. To calibrate the sensitivity parameter  $\gamma_S$ , the following one-to-one mapping is adopted:

$$|\gamma_S^*| = \frac{1}{\sigma_Y} \sqrt{\frac{(\rho^*)^2}{1 - (\rho^*)^2} [\text{var}\{m_S(X)\} + \pi^2/3]}. \quad (8)$$

Similar bounding procedures apply to the calibration of  $\gamma_{R_0}$  and  $\gamma_{R_1}$ .

### 6 SIMULATION STUDY

We first conduct a set of simulations to evaluate the operating characteristics of the proposed estimators under possible model misspecification when Assumptions 3-5 in Table 2 are satisfied. Set the sample sizes of the SAT and EC to be around  $N_{\mathcal{R}} = 200$  and  $N_{\mathcal{E}} = 500$  with total size  $N = 700$ . The covariates  $X \in \mathbb{R}^5$  are generated by  $X_j \sim N(0.25, 1)$  for  $j = 1, \dots, 4$  and  $X_5 \sim \text{Bernoulli}(0.5)$ . Consider a nonlinear transformation of the covariates and denote  $Z_j = \{X_j^2 + 2 \sin(X_j) - 1.5\} / \sqrt{2}$  for  $j = 1, \dots, 4$  and  $Z_5 = X_5$ . We generate the indicator of being selected to SAT or EC by  $S | X \sim \text{Bernoulli}\{\pi_S(X)\}$ , where  $\pi_S(X) = \text{logit}^{-1}(\alpha_S + 0.1 \sum_{j=1}^5 Z_j)$  and  $\alpha_S$  is chosen adaptively to ensure the average of  $S$  is about  $N_{\mathcal{R}}/N$ . Next, we generate the indicators of the intercurrent events and the outcomes

for SAT and EC by

$$R | X, S = s \sim \text{Bernoulli}\{\pi_{R_s}(X)\}, \quad \pi_{R_s}(X) \\ = \text{logit}^{-1}(\alpha_{R_s} + \sum_{j=1}^5 Z_j/6), \quad s = 0, 1,$$

$Y | X, S = 1 \sim \mathcal{N}(\sum_{j=1}^5 Z_j/2, 1)$ , and  $Y | X, S = 0 \sim \mathcal{N}(\sum_{j=1}^5 Z_j/3, 1)$ , where  $(\alpha_{R_{S1}}, \alpha_{R_{S0}})$  are adaptively chosen to ensure the average propensity of  $R$  is around 0.5. With a large sample size of Monte Carlo simulation, we compute true ATE  $\tau = 0.13$ . First, we assess the robustness of the proposed estimator  $\hat{\tau}^t$  when the sensitivity parameters  $\gamma_S = \gamma_{R_0} = \gamma_{R_1} = 0$ . Denote the tilting estimator  $\hat{\tau}^t$  with fixed zero-valued sensitivity parameters as  $\hat{\tau}^{[0]}$ , we consider 2 model specifications of the propensity (PS) of the participation  $\pi_S(X)$  and the intercurrent events  $\pi_{R_s}(X)$ , and the outcome models (OM). In particular, we fit the corresponding parametric models with the covariates  $Z$  as the correctly specified models or with the covariates  $X$  as the misspecified models. We compare our proposed EIF-motivated tilting estimator with other 2 estimators, which are constructed solely based on PS or OM, denoted by  $\hat{\tau}^{PS}$  and  $\hat{\tau}^{OM}$ , respectively. Figure 2A shows the point estimation results based on 500 Monte Carlo experiments. When PS and OM are correctly specified, the considered estimators are all unbiased. However,  $\hat{\tau}^{PS}$  and  $\hat{\tau}^{OM}$  are biased when their required models are misspecified. Our proposed tilting estimator is shown to be doubly robust with fixed zero-valued sensitivity parameters as it is consistent if either PS or OM is correctly specified.

Table 3(A) shows the absolute bias, standard errors (SE), mean squared error (MSE), coverage rates (CRs), and the average CI lengths of each estimator. We construct the corresponding 95% Wald-type CIs for inference, where the variances are estimated by non-parametric bootstrap with size  $B = 50$ . We observe that both  $\hat{\tau}^{[0]}$  and  $\hat{\tau}^{OM}$  exhibit the smallest average confidence interval lengths. However, when the OM is misspecified, the CR for  $\hat{\tau}^{[0]}$  is closer to the nominal level compared to  $\hat{\tau}^{OM}$ . This observation underscores the double robustness of  $\hat{\tau}^{[0]}$ , aligning with the results shown in Figure 2 and supporting our claims in Theorem S2 for the primary analysis.

Next, we assess the performance of our proposed tilting estimators  $\hat{\tau}^t$  when Assumptions 3-5 in Table 2 are violated. We keep the same data-generating process for the  $X$ ,  $Z$ , and  $Y$  but generate the indicators using Bernoulli sampling with different propensities:

$$P\{S = 1 | X, Y(0)\} = \text{logit}^{-1}(\alpha_S + 0.1 \sum_{j=1}^5 Z_j + \gamma_S Y(0)), \\ P\{R = 0 | X, Y(s), S = s\} = \text{logit}^{-1}(-\alpha_{R_s} - \sum_{j=1}^5 Z_j/6 + \gamma_{R_s} Y(s)), \quad s = 0, 1,$$

where  $(\gamma_S, \gamma_{R_1}, \gamma_{R_0})$  control the strength of the unmeasured confounder, describing how the propensity depends on the potential outcome after accounting for the covariates. We consider a range of sensitivity parameters where the indicators are equally confounded by the potential outcomes, that is,  $\gamma = \gamma_S = \gamma_{R_1} = \gamma_{R_0}$ . Under the hypothetically fixed sensitivity parameters  $(\gamma_S, \gamma_{R_1}, \gamma_{R_0})$  and assuming that all the nuisance models are correctly specified, Figure 2(B) shows the point estimation for all 3 estimators, and Table 3(B) presents the finite-sample performances of  $\hat{\tau}^t$  in details. One important finding is that our proposed estimator is always consistent across a range

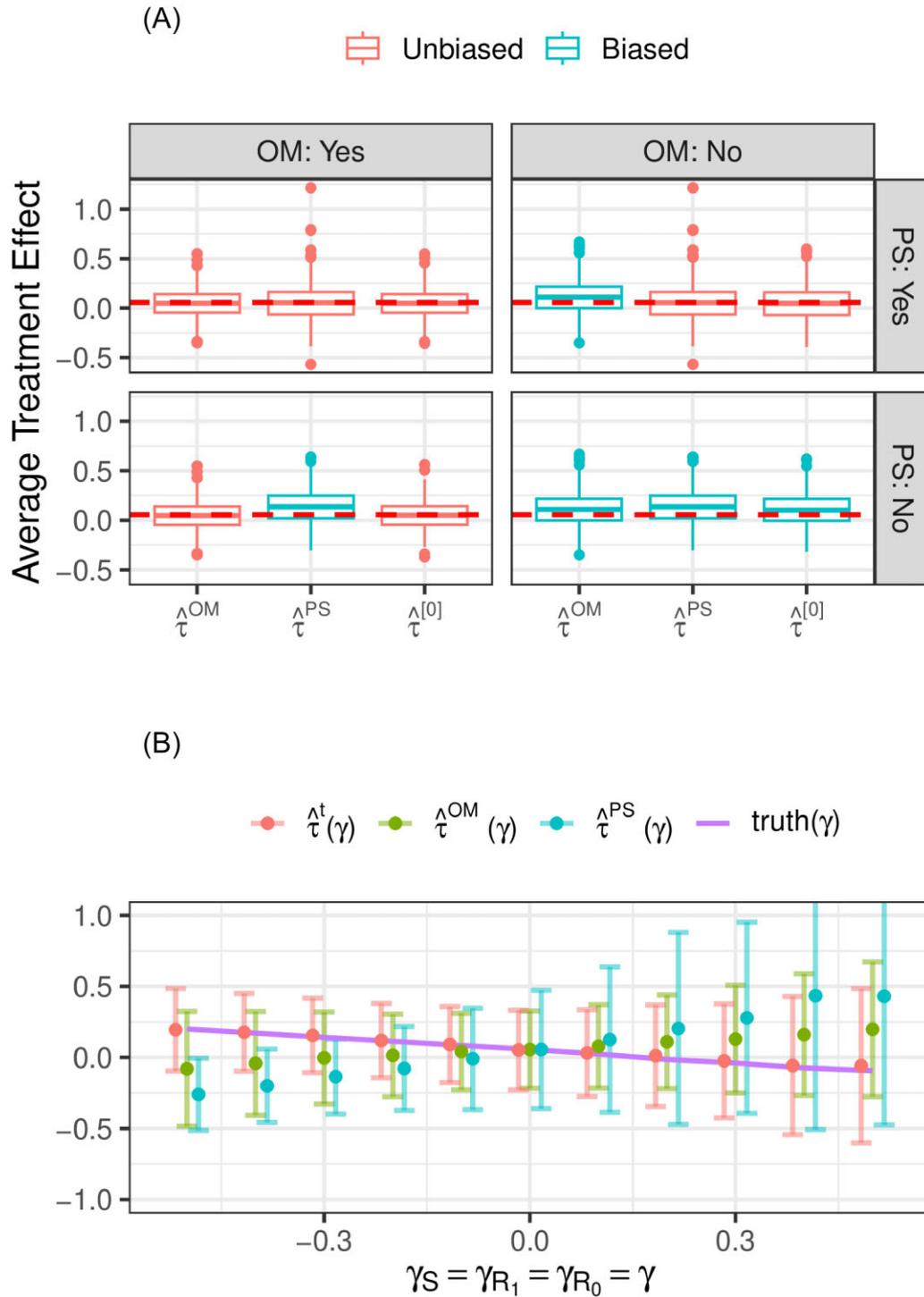
of fixed sensitivity parameters, as indicated by its small bias and approximately correct CRs. In contrast, the other 2 estimators are unable to handle the joint sensitivity analysis for multiple assumptions, suggested by their non-negligible biases for fixed non-zero sensitivity parameters.

## 7 REAL-DATA APPLICATION

We examine a study designed to estimate an antidepressant drug effect on the scores of the Hamilton Depression Rating Scale for 17 items (HAMD-17). This study was conducted under the Auspices of the Drug Information Association, which collects the data at baseline and weeks 1, 2, 4, 6, and 8 for  $N = 196$  patients with 99 in the control group and 97 in the treatment group. However, some patients may drop out during the study for various reasons. Our primary interest is the ATE on the change of HAMD-17, irrespective of the intercurrent events such as dropout-related missing data. According to the guidelines in ICH (2021), the ATE is defined as the mean difference of the change in the HAMD-17 scores from the baseline to the final time point in week 8. We adhere to the same analysis plan as Liu et al. (2024) with covariates  $X$ , including the investigation sites and baseline HAMD-17 scores. Let  $Y(a)$  and  $R$  be the change of HAMD-17 scores under treatment  $a$  and the indicator of whether a patient stayed in the study at week 8.

In our application, we consider the original trial as a single-arm trial, with the concurrent control group being considered as external controls to illustrate the proposed sensitivity analysis. We assume the potential outcomes  $Y(a)$  follow a Gaussian mixture model fitted using the R package `flexmix`. Next, we bound the magnitude of the sensitivity parameters using the approach described in Section 5. We illustrate this approach with the baseline HAMD-17 score, which is the most important predictors in terms of partial variance explained, with  $(\rho_S^*)^2 \approx 0.02$ ,  $(\rho_{R_1}^*)^2 \approx 0.11$ , and  $(\rho_{R_0}^*)^2 \approx 0.04$ . To map these values to the sensitivity parameters, we apply the one-to-one mapping formula (8), and obtain the calibrated sensitivity parameters  $|\gamma_{R_0}^*| \approx 0.02$ ,  $|\gamma_{R_1}^*| \approx 0.02$ , and  $|\gamma_S^*| \approx 0.01$ . Figure 3 illustrates the ATE estimates across a range of hypothetical sensitivity parameters, adjusting for the potential outcomes as the unmeasured confounder in the logistic selection specification. The shaded area indicates the unmeasured confounder with impacts up to the values of the calibrated sensitivity parameters. Here, “NS” denotes “not significant,” meaning the 95% confidence interval of the ATE contains 0.

When the assumptions are satisfied, that is, all the sensitivity parameters equal 0, the ATE estimates are  $\hat{\tau}^t = -1.42$  with the 95% bootstrap CI as  $(-2.80, -0.05)$ , which is statistically significant. Next, we assume the impact of the confounders acts toward hurting our preferred hypothesis, that is,  $\gamma_S^* < 0$ . Here, a negative value of  $\gamma_S$  suggests that the unobserved change in HAMD-17 scores in the concurrent control group tends to be lower (ie, better) than the observed external controls, reducing the absolute value of the effect size. This could occur if patients are more likely to participate in the single-arm trial when less depressed. When  $\gamma_S = -0.01$  and  $\gamma_{R_0} = \gamma_{R_1} = -0.02$ , the estimated treatment effect of the antidepressant drug becomes  $\hat{\tau}^t = -1.28$ , where the unmeasured confounder is as strong as



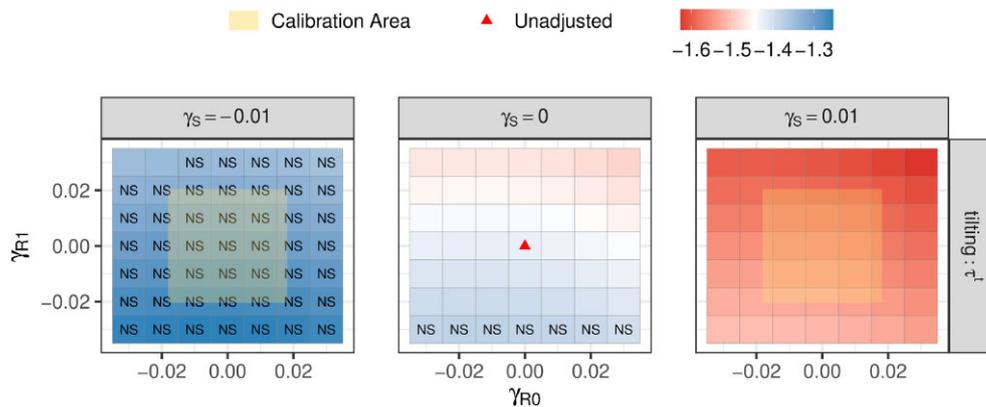
**FIGURE 2** Performance of the PS-based, OR-based, and EIF-motivated tilting estimator with fixed sensitivity parameters: (A) zero-valued under 4 different model specifications, and (B) non-zero-valued when all the nuisance models are correctly specified, based on 500 Monte Carlo simulations.

the baseline HAMD-17 scores. Although the tilted estimate is below 0, suggesting the effectiveness of the drug on the HAMD-17 scores, it is no longer statistically significant, as its 95% CI is  $(-2.65, 0.09)$ . Thus, following our sensitivity analyses, we show that the magnitude of the possible drug effects on the HAMD-17 scores is robust, but the significance of such effects is not robust against unmeasured confounding at the strength of

baseline HAMD-17 scores. However, domain knowledge is still

**TABLE 3** The bias, standard errors (SE), mean squared error (MSE), coverage rates (CR), and the average CI width based on 500 Monte Carlo experiments of (A) the PS-based, OR-based, and EIF-motivated estimator under 4 different model specifications when Assumptions 3-5 in Table 2 hold; (B) the EIF-motivated tilting estimates with the fixed sensitivity parameters when Assumptions 3-5 in Table 2 are violated.

(A)		Bias	SE	MSE	CR	CI width
		PS-based estimator $\hat{\tau}^{PS}$				
PS = yes	OM = yes	0.01	0.22	0.05	95.8%	0.86
PS = yes	OM = no	0.01	0.22	0.05	95.8%	0.86
PS = no	OM = yes	0.07	0.18	0.04	91.4%	0.86
PS = no	OM = no	0.07	0.18	0.04	86.4%	0.86
		OM-based estimator $\hat{\tau}^{OM}$				
PS = yes	OM = yes	0.02	0.14	0.02	95.0%	0.55
PS = yes	OM = no	0.05	0.17	0.03	86.4%	0.55
PS = no	OM = yes	0.02	0.14	0.02	95.0%	0.55
PS = no	OM = no	0.05	0.17	0.03	86.4%	0.55
		EIF-motivated tilting estimator $\hat{\tau}^{[0]}$				
PS = yes	OM = yes	0.02	0.14	0.02	94.2%	0.56
PS = yes	OM = no	0.02	0.18	0.03	93.2%	0.56
PS = no	OM = yes	0.02	0.14	0.02	95.2%	0.56
PS = no	OM = no	0.04	0.17	0.03	89.8%	0.56
(B)		bias	SE	MSE	CR	CI width
$\gamma_S = \gamma_{R_1} = \gamma_{R_0}$		EIF-motivated tilting estimator $\hat{\tau}^t$				
	-0.5	0.01	0.15	0.02	95.2%	0.58
	-0.3	0.01	0.13	0.02	95.4%	0.52
	-0.1	0.01	0.14	0.02	96.2%	0.54
	0.0	0.00	0.14	0.02	94.2%	0.56
	0.1	0.01	0.16	0.02	94.8%	0.61
	0.3	0.02	0.22	0.05	93.6%	0.80
	0.5	0.03	0.31	0.09	93.4%	1.09



**FIGURE 3** Average treatment effects of an antidepressant drug effect on the HAMD-17 scores over a grid of hypothesized sensitivity parameters under the tilting sensitivity models.

required to consider the plausibility of unmeasured confounders of such strength level under this situation.

### 8 DISCUSSION

In this paper, we develop a semi-parametric efficient framework for sensitivity analysis under the tilting models. Motivated by Tukey’s factorization, this framework effectively separates the model checking from the sensitivity analysis, which does not rely on any modeling assumption and fits perfectly well with the EIF-motivated tilting estimators. Decoupling the

sensitivity analysis from the model fit assessment is crucial and ubiquitous within the model-based sensitivity analysis. However, joint sensitivity analysis for multiple assumptions remains largely unexplored to the best of our knowledge. By simultaneously assessing the EC outcome mean non-exchangeability and the effects of intercurrent events, our framework hopes to shed more light on the advancements of joint modeling for sensitivity analysis.

Future work could extend our framework to longitudinal trials with intercurrent events, particularly those with irregular and informative observation patterns, as discussed by Yang (2021)

and Smith et al. (2024). Such an extension may increase the number of sensitivity parameters across multiple time points, and introduce additional challenges in deriving the EIFs conditioned on the historical information. Another potential extension could focus on the choices of sensitivity parameters, which is profoundly useful in practice. Our approach relies on bounding the magnitude of sensitivity parameters using observed data, and substantive domain expertise should be consulted to examine whether an unmeasured confounder with such strength is plausible. For some hybrid control designs, the sensitivity parameters can be partially identified with the help of concurrent controls; similar ideas have been explored in Gao et al. (2024) to adjust for EC outcome mean non-exchangeability. Thus, the internal validity from the hybrid controls can be leveraged to inform the choices of sensitivity parameters. In summary, our proposed semi-parametric sensitivity analysis is both efficient and flexible as it is rate-doubly robust, locally optimal, and can be incorporated with a range of models with a modern causal inference workflow.

### ACKNOWLEDGMENTS

We thank the editor, associate editor, and 2 anonymous referees for their constructive suggestions and valuable feedback on our manuscript.

### SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biometrics* online.

Web Appendices, tables, figures, and the proofs of all the theorems referenced in Sections 3-5, along with additional details for the J2R sensitivity model, are available with this paper on the Biometrics website at Oxford Academic. An R-package for implementing the proposed method is posted along with this article and available online at <https://github.com/Gaochenyin/SensDR>.

### FUNDING

None declared.

### CONFLICT OF INTEREST

None declared.

### DATA AVAILABILITY

The data that support the findings in this paper are available in the Drug Information Association Missing Data at <https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#dia-missing-data>, collected by Mallinckrodt et al. (2014), and are also provided at <https://github.com/Gaochenyin/SensDR>.

### REFERENCES

- Allen, A. S., Satten, G. A. and Tsiatis, A. A. (2005). Locally-efficient robust estimation of haplotype-disease association in family-based studies. *Biometrika*, 92, 559–571.
- Blackwell, M. (2014). A selection bias approach to sensitivity analysis for causal effects. *Political Analysis*, 22, 169–182.
- Bradic, J., Wager, S. and Zhu, Y. (2019). Sparsity double robust inference of average treatment effects. arXiv, arXiv:1905.00744, preprint: not peer reviewed.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. et al. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal*, 21, C1–C68.
- 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.
- Copas, J. B. and Li, H. (1997). Inference for non-random samples. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 59, 55–95.
- Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B. and Wynder, E. L. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, 22, 173–203.
- Dahabreh, I. J., Robins, J. M., Haneuse, S. J.-P., Saeed, I., Robertson, S. E., Stuart, E. A. et al. (2023). Sensitivity analysis using bias functions for studies extending inferences from a randomized trial to a target population. *Statistics in Medicine*, 42, 2029–2043.
- Dorie, V., Harada, M., Carnegie, N. B. and Hill, J. (2016). A flexible, interpretable framework for assessing sensitivity to unmeasured confounding. *Statistics in Medicine*, 35, 3453–3470.
- Faries, D., Gao, C., Zhang, X., Hazlett, C., Stamey, J., Yang, S. et al. (2024). Real effect or bias? Best practices for evaluating the robustness of real-world evidence through quantitative sensitivity analysis for unmeasured confounding. *Pharmaceutical Statistics*.
- Food and Drug Administration (2023). Considerations for the design and conduct of externally controlled trials for drug and biological products guidance for industry. <https://www.fda.gov/media/164960/download>. [Accessed 23 February 2023].
- Franks, A., 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.
- Gao, C., Yang, S., Shan, M., YE, W., Lipkovich, I. and Faries, D. (2024). Improving randomized controlled trial analysis via data-adaptive borrowing. *Biometrika*, asae069.
- Garcia, T. P. and Ma, Y. (2016). Optimal estimator for logistic model with distribution-free random intercept. *Scandinavian Journal of Statistics*, 43, 156–171.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, 47, 153–161.
- ICH (2021). E9(R1) statistical principles for clinical trials: addendum: estimands and sensitivity analysis in clinical trials. *FDA Guidance Documents*.
- Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, 93, 126–132.
- Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.
- Kennedy, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In: *Statistical Causal Inferences and Their Applications in Public Health Research* (eds He, H., Wu, P. and Chen, D.G.), 141–167. Cham: Springer.
- Lipkovich, I., Ratitch, B. and Mallinckrodt, C. H. (2020). Causal inference and estimands in clinical trials. *Statistics in Biopharmaceutical Research*, 12, 54–67.
- Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*, vol. 793. Wiley.

- Liu, S., Yang, S., Zhang, Y. and Liu, G. (2024). Multiply robust estimators in longitudinal studies with missing data under control-based imputation. *Biometrics*, 80, ujad036.
- Mallinckrodt, C., Roger, J., Chuang-Stein, C., Molenberghs, G., O’Kelly, M., Ratitch, B. et al. (2014). Recent developments in the prevention and treatment of missing data. *Therapeutic Innovation and Regulatory Science*, 48, 68–80.
- Nabi, R., Bonvini, M., Kennedy, E. H., Huang, M.-Y., Smid, M. and Scharfstein, D. O. (2024). Semiparametric sensitivity analysis: unmeasured confounding in observational studies. *Biometrics*, 80, ujae106.
- Robins, J. M., Rotnitzky, A. and Scharfstein, D. O. (2000). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (eds Halloran, M.E. and Berry, D.), 1–94. New York, NY: Springer.
- Rosenbaum, P. R. (1987). Sensitivity analysis for certain permutation inferences in matched observational studies. *Biometrika*, 74, 13–26.
- Rosenbaum, P. R. and Rubin, D. B. (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 45, 212–218.
- Rosenbaum, P. R. and Rubin, D. B. (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55.
- Smith, B. B., Gao, Y., Yang, S., Varadhan, R., Apter, A. J. and Scharfstein, D. O. (2024). Semi-parametric sensitivity analysis for trials with irregular and informative assessment times. *Biometrics*, 80, ujae154.
- Tsiatis, A. A. and Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika*, 91, 835–848.
- VanderWeele, T. J. and Arah, O. A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*, 22, 42–52.
- Veitch, V. and Zaveri, A. (2020). Sense and sensitivity analysis: simple post-hoc analysis of bias due to unobserved confounding. *Advances in Neural Information Processing Systems*, 33, 10999–11009.
- Yang, S. (2021). Semiparametric estimation of structural nested mean models with irregularly spaced longitudinal observations. *Biometrics*, 78, 937–949.
- Yang, S. and Lok, J. J. (2018). Sensitivity analysis for unmeasured confounding in coarse structural nested mean models. *Statistica Sinica*, 28, 1703.
- Zhang, B. and Tchetgen, E. J. T. (2019). A semiparametric approach to model-based sensitivity analysis in observational studies. arXiv:1910.14130, preprint: not peer reviewed.