Statistics in Medicine

WILEY

Statistics in Medicine

RESEARCH ARTICLE

A Simple Sensitivity Analysis Method for Unmeasured Confounders via Linear Programming With Estimating Equation Constraints

Chengyao Tang[1] | Yi Zhou[1,2] | Ao Huang[3] | Satoshi Hattori[1,4]

[1]Department of Biomedical Statistics, Graduate School of Medicine, Osaka University, Osaka, Japan | [2]Beijing International Center for Mathematical Research, Peking University, Beijing, China | [3]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany | [4]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives (OTRI), Osaka University, Osaka, Japan

Correspondence: Satoshi Hattori (hattoris@biostat.med.osaka-u.ac.jp)

Received: 7 December 2023 | **Revised:** 3 November 2024 | **Accepted:** 5 November 2024

Funding: This research was partly supported by Grant-in-Aid for Challenging Exploratory Research (16K12403) and for Scientific Research (16H06299, 18H03208) from the Ministry of Education, Science, Sports, and Technology of Japan.

Keywords: average treatment effect | linear programming | sensitivity analysis | unmeasured confounders

ABSTRACT

In estimating the average treatment effect in observational studies, the influence of confounders should be appropriately addressed. To this end, the propensity score is widely used. If the propensity scores are known for all the subjects, bias due to confounders can be adjusted by using the inverse probability weighting (IPW) by the propensity score. Since the propensity score is unknown in general, it is usually estimated by the parametric logistic regression model with unknown parameters estimated by solving the score equation under the strongly ignorable treatment assignment (SITA) assumption. Violation of the SITA assumption and/or misspecification of the propensity score model can cause serious bias in estimating the average treatment effect (ATE). To relax the SITA assumption, the IPW estimator based on the outcome-dependent propensity score has been successfully introduced. However, it still depends on the correctly specified parametric model and its identification. In this paper, we propose a simple sensitivity analysis method for unmeasured confounders. In the standard practice, the estimating equation is used to estimate the unknown parameters in the parametric propensity score model. Our idea is to make inferences on the (ATE) by removing restrictive parametric model assumptions while still utilizing the estimating equation. Using estimating equations as constraints, which the true propensity scores asymptotically satisfy, we construct the worst-case bounds for the ATE with linear programming. Differently from the existing sensitivity analysis methods, we construct the worst-case bounds with minimal assumptions. We illustrate our proposal by simulation studies and a real-world example.

1 | Introduction

In observational studies, it is always crucial to adjust for the influence of confounders in estimating the average treatment effect (ATE). If all the confounders are observed and satisfy the strongly ignorable treatment assignment (SITA) assumption [1, 2], one can adjust the effects of confounders by using the propensity score. With the propensity score, the inverse probability weighting (IPW) [3, 4] is a popular approach. The IPW method constructs weights on the observations of each subject, and then the ATE can be identified by comparing the weighted outcomes of two groups [5]. In practice, the propensity score is

unknown. Then, the estimation of the propensity score usually relies on a parametric model such as the logistic regression under the SITA assumption. In most observational studies, it is untestable and implausible that there are no unmeasured confounders, and then the SITA assumption may fail to hold. Using the outcome-dependent propensity score is an option to make inference without the SITA assumption [6, 7]. By incorporating the outcome variable in the model of the propensity score, we can make inferences on the ATE without the SITA assumption. In general, the outcome-dependent propensity score is estimated by a parametric logistic regression model with the observed confounders and the outcome as explanatory variables. Thus, model misspecification is still of concern in the estimation of the outcome-dependent propensity score. Moreover, it has an unidentifiability issue [8, 9]; that is, the estimating equation cannot determine the unknown parameters uniquely in the outcome-dependent propensity score. Then, the outcome-dependent propensity score cannot solve the issue of unmeasured confounders completely.

Sensitivity analysis is a useful tool to assess the potential impact of unmeasured confounders, and many sensitivity analysis methods have been developed. With the substantially increasing applications of the propensity score methods in the analysis of observational studies, there is a growing interest in employing sensitivity analysis methods in real-data analyses. Typical sensitivity analysis approaches involve formulating additional assumptions with regard to the relationships among unmeasured confounders, treatment assignments, and outcomes. These assumptions often take the form of plausible values for parameters that cannot be directly estimated from the observed data and must be set by analysts. Rosenbaum and Rubin [10] and Lin et al. [11] modeled the mechanism of confounding with both the measured and unmeasured confounders and then estimated the treatment effect parameter of interest. Alternatively, Cornfield et al. [12] and Ding and Vanderweele [13] developed methods to construct the bounds for the treatment effects to quantify the magnitude of the unmeasured confounders. These bounds were designed to elucidate the extent to which unmeasured confounders could influence observed causal estimates. Particularly, when the sensitivity parameters were expressed as risk ratios, the E-value [14] was introduced and has become a pivotal quantity in the realm of causal inference in observational studies. While the E-value can provide a bound without any model specification, the estimand is restrictive, and the bound is likely to be wide, which can lead to inefficiency in sensitivity analysis.

For the sensitivity analysis approaches based on the IPW method to estimate the ATE, Li et al. [15] modeled the mean between-group differences of potential outcomes to correct bias in the presence of unmeasured confounders. Shen et al. [16] proposed an IPW-based sensitivity analysis method by using two parameters, the variance of the multiplicative errors in the estimated propensity score and its correlations with the potential outcomes, to quantify the bias due to unmeasured confounders. Lu and Ding [17] extended the method of Li et al. [15] into a more flexible sensitivity analysis framework, which can handle the IPW, outcome regression, and doubly robust estimators. In addition, Zhao et al. [18] constructed bounds for the ATE based on the IPW estimators by incorporating a marginal sensitivity model [19]. Dorn and Guo [20] further refined this method and gave sharper bounds. These sensitivity analysis methods can address the impacts of violation of the SITA assumption by quantifying potential biases; however, they rely on untestable parametric assumptions on the departure from the SITA assumption, and it is practically difficult to set a relevant magnitude of the departure.

In this paper, a simple sensitivity analysis framework for unmeasured confounders is proposed. In the standard process of the confounder adjustment with the outcome-dependent propensity score, a parametric model for the propensity score is assumed, and an estimating equation is introduced to estimate its unknown parameters. Instead of determining a unique model for the outcome-dependent propensity score, we construct bounds for the ATE by considering possible propensity scores. We realize it by removing the parametric model for the propensity score but still relying on the estimating equation. We introduce an optimization problem constrained by the estimating equation, which the true propensity score asymptotically satisfies. The worst-case bounds for the ATE can be obtained by solving a linear programming problem. Different from the existing sensitivity analysis methods, the proposed worst-case bounds do not rely on strong assumptions. By increasing the dimension of the estimating equations involving many covariates, one can make the bounds further narrow. Compared with existing sensitivity analysis methods, the proposed method offers the following advantages. First, the proposed method can provide worst-case bounds with minimal assumptions. Second, since the proposed method is free from the estimated propensity score under the SITA assumption, its misspecification does not matter. Finally, the proposed method exhibits computational efficiency as the optimization problem can be solved by linear programming.

The rest of this paper is organized as follows. In Section 2, we introduce the basic notations and the standard methods with the parametric propensity score. In Section 3, some existing sensitivity analysis methods for the IPW estimator are reviewed. In Section 4, the proposed method for sensitivity analysis is introduced. We investigate the performance of the proposed method on simulated datasets in Section 5, and illustrate it on a real-world example in Section 6. In Section 7, we provide a concluding discussion to summarize the main findings and contributions of this paper.

2 | Estimation With the Parametric Propensity Score

2.1 | Notations and the Standard Propensity Score Analysis

In this paper, we consider estimating the ATE for the overall mean over the population in an observational study with two treatment groups. Let Z be the treatment assignment: $Z = 1$ if the subject is in the treated (exposed) group and $Z = 0$ if in the control group. Let X be a vector of baseline covariates and Y be the observed outcome. We follow Rubin's causal model framework [21]. Let $Y^{(1)}$ and $Y^{(0)}$ be the potential outcomes if the subjects were assigned to the treated group ($Z = 1$) and the control group ($Z = 0$), respectively. Suppose the observational study enrolls n subjects, and the observed data (Y_i, Z_i, X_i) for subject i ($i = 1, 2, \ldots, n$) are available, which are independent and

identically distributed copies of (Y, Z, X). Denote $\mu_1 = E[Y^{(1)}]$ and $\mu_0 = E[Y^{(0)}]$. The ATE, which is of our primary interest to estimate, is defined by

$$\psi = \mu_1 - \mu_0 = E[Y^{(1)}] - E[Y^{(0)}].$$

In observational studies, owing to the absence of randomization, the potential influence of confounders should be carefully handled in estimating the ATE. The propensity score is widely used to adjust the bias due to confounding. The propensity score is defined by $e(X_i) = P(Z_i = 1|X_i)$. Various methods, such as stratification, matching, and IPW [3, 4, 22], can be employed to adjust for confounding with the propensity score. The standard propensity analysis is conducted under the following assumptions:

Assumption 1. Consistency: $Y = ZY^{(1)} + (1 - Z)Y^{(0)}$.

Assumption 2. Positivity: there exists a small positive parameter δ such that $0 < \delta \leq e(X) \leq 1 - \delta$, for each value of the covariate X in the population.

Assumption 3. SITA: $(Y^{(1)}, Y^{(0)}) \perp\!\!\!\perp Z|X$.

Assumption 3 implies that the bias due to confounding can be adjusted by using X in principle. The SITA assumption is corresponding to the Missing At Random (MAR) in the missing data analysis context. In this paper, we handle situations in which the SITA is violated, which is corresponding to the concept of the Missing Not At Random (MNAR) in the missing data problem. We use the terminologies SITA and MAR interchangeably. In practice, the propensity score is unknown, and then some parametric models, such as the logistic regression model, are usually assumed. Let $\text{logit}(e(X_i; \theta, \alpha)) = \theta + \alpha^\top X_i$. The unknown parameters are usually estimated by solving the following score equation:

$$\sum_{i=1}^{n} \begin{pmatrix} 1 \\ X_i \end{pmatrix} \left(Z_i - \frac{exp(\theta + \alpha^\top X_i)}{1 + exp(\theta + \alpha^\top X_i)} \right) = 0 \tag{1}$$

Let the solution to the score equation for (θ, α) be denoted by $(\hat{\theta}, \hat{\alpha})$, and $\hat{e}(X_i) = e(X_i; \hat{\theta}, \hat{\alpha})$. Then, we can determine the unique set of propensity scores for all subjects. In this paper, the propensity score estimated under the SITA assumption is called the MAR-based propensity score to avoid confusion; another type of the propensity score is introduced in a later section, which is called the outcome-dependent propensity score. The IPW estimator for μ_1 is defined by

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i Y_i}{\hat{e}(X_i)} \tag{2}$$

Similarly, we can estimate μ_0 with

$$\hat{\mu}_0 = \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - Z_i) Y_i}{1 - \hat{e}(X_i)} \tag{3}$$

and then the ATE ψ is estimated with

$$\hat{\psi} = \hat{\mu}_1 - \hat{\mu}_0.$$

The aforementioned IPW estimator has an unstabilized form, which may suffer from extremely large weights when some propensity scores are very close to one or zero and then can cause instability in the estimation. The stabilized IPW (SIPW) estimator introduces a stabilization term to the weights, which helps mitigate the impact of extreme weights. Specifically, the SIPW estimator for μ_1 is defined by

$$\hat{\mu}_{1,SIPW} = \left(\sum_{i=1}^{n} \frac{Z_i}{\hat{e}(X_i)} \right)^{-1} \sum_{i=1}^{n} \frac{Z_i Y_i}{\hat{e}(X_i)} \tag{4}$$

Similarly, we can estimate μ_0 with

$$\hat{\mu}_{0,SIPW} = \left(\sum_{i=1}^{n} \frac{1 - Z_i}{1 - \hat{e}(X_i)} \right)^{-1} \sum_{i=1}^{n} \frac{(1 - Z_i) Y_i}{1 - \hat{e}(X_i)} \tag{5}$$

and then the ATE ψ is estimated with

$$\hat{\psi}_{SIPW} = \hat{\mu}_{1,SIPW} - \hat{\mu}_{0,SIPW}.$$

In this paper, we focus on the SIPW estimator.

If the SITA assumption holds and the model of the propensity score is correctly specified, the ATE is consistently estimated. However, the SITA assumption does not hold in the presence of unmeasured confounders.

2.2 | Estimation With the Outcome-Dependent Propensity Score

In this section, suppose that the SITA assumption does not necessarily hold in the presence of unmeasured confounder U. The estimation of the ATE using the method in Section 2.1 is no longer valid.

To address the issue of unmeasured confounders, the outcome-dependent propensity score approach [23–25] has been successfully introduced. We define the outcome-dependent propensity scores by $o^1(X_i, Y_i^{(1)}) = P(Z_i = 1|X_i, Y_i^{(1)})$ and $o^0(X_i, Y_i^{(0)}) = P(Z_i = 1|X_i, Y_i^{(0)})$ for subjects in the treated and control groups, respectively.

One may consider the logistic regression models for $o^1(X_i, Y_i^{(1)})$ and $o^0(X_i, Y_i^{(0)})$. Let us consider the models $\text{logit}(o^1(X_i, Y_i^{(1)}; \theta^1, \alpha^1, \beta^1)) = \theta^1 + \alpha^{1\top} X_i + \beta^1 Y_i^{(1)}$ and $\text{logit}(o^0(X_i, Y_i^{(0)}; \theta^0, \alpha^0, \beta^0)) = \theta^0 + \alpha^{0\top} X_i + \beta^0 Y_i^{(0)}$. The score Equation (1) does not work for estimation of the unknown parameters in these models, since $Y_i^{(z)}$ is observed only for subjects with $Z_i = z$. The unknown parameters in the model of $o^1(X_i, Y_i^{(1)}; \theta^1, \alpha^1, \beta^1)$ can be estimated by solving the following estimating equation:

$$\sum_{i=1}^{n} g(X_i) \left(1 - \frac{Z_i}{o^1(X_i, Y_i; \theta^1, \alpha^1, \beta^1)} \right) = 0 \tag{6}$$

where $g(X)$ is a vector of the same dimensions as $(\theta^1, \alpha^1, \beta^1)$ and the solution to the estimating Equation (6) is denoted by $(\hat{\theta}^1, \hat{\alpha}^1, \hat{\beta}^1)$. Similarly, the unknown parameters in the model of $o^0(X_i, Y_i^{(0)}; \theta^0, \alpha^0, \beta^0)$ can be estimated by solving the following estimating equation:

$$\sum_{i=1}^{n} g(X_i)\left(1-\frac{1-Z_i}{1-o^0(X_i,Y_i;\theta^0,\alpha^0,\beta^0)}\right)=0 \tag{7}$$

The dimension of $g(X)$ should be equal to that of $(\theta^0,\alpha^0,\beta^0)$ to obtain a solution. The solution to the estimating Equation (7) is denoted by $(\hat{\theta}^0,\hat{\alpha}^0,\hat{\beta}^0)$. Denote $\hat{o}^1(X_i,Y_i^{(1)})=o^1(X_i,Y_i;\hat{\theta}^1,\hat{\alpha}^1,\hat{\beta}^1)$ and $\hat{o}^0(X_i,Y_i^{(0)})=o^0(X_i,Y_i;\hat{\theta}^0,\hat{\alpha}^0,\hat{\beta}^0)$, respectively. We can then estimate μ_1 under the MNAR with

$$\hat{\mu}_{1,SIPW}^{MNAR}=\left(\sum_{i=1}^{n}\frac{Z_i}{\hat{o}^1(X_i,Y_i^{(1)})}\right)^{-1}\sum_{i=1}^{n}\frac{Z_iY_i}{\hat{o}^1(X_i,Y_i^{(1)})}.$$

Similarly, we can estimate μ_0 under the MNAR with

$$\hat{\mu}_{0,SIPW}^{MNAR}=\left(\sum_{i=1}^{n}\frac{1-Z_i}{1-\hat{o}^0(X_i,Y_i^{(0)})}\right)^{-1}\sum_{i=1}^{n}\frac{(1-Z_i)Y_i}{1-\hat{o}(X_i,Y_i^{(0)})},$$

and then the ATE is estimated with

$$\hat{\psi}_{SIPW}^{MNAR}=\hat{\mu}_{1,SIPW}^{MNAR}-\hat{\mu}_{0,SIPW}^{MNAR}.$$

The SIPW estimator with the outcome-dependent propensity score can consistently estimate the ATE without the SITA assumption as long as the parametric models for the outcome-dependent propensity score are correctly specified. However, estimations with (6) and (7) often encounter an unidentifiability issue, wherein the model coefficients obtained through solving the estimating equations may not be uniquely determined. Miao et al. [9] pointed out that even if the model for the propensity score has a known parametric form, the model is not identifiable without specifying a parametric outcome distribution. A unique solution to the estimating equations is only achieved when both the outcome model and the propensity score model are appropriately specified. Specifically, without additional restrictions or assumptions, solely solving the estimating Equations (6) is not sufficient to determine the coefficients $(\hat{\theta}^1,\hat{\alpha}^1,\hat{\beta}^1)$ uniquely. Therefore, the outcome-dependent propensity score cannot solve the issue of the unmeasured confounder completely.

3 | Existing Sensitivity Analysis Methods

In this section, we briefly review some existing sensitivity analysis methods for the IPW estimator.

3.1 | Modeling the Mean Difference of the Potential Outcomes

Along with the lines of the work by Robins et al. [26, 27], Brumback et al. [28] proposed to quantify the impact of the unmeasured confounders by modeling the mean between-group difference of the potential outcomes, conditional on all observed covariates. The sensitivity function is defined by $c(z,X)=E[Y^{(z)}|Z=1,X]-E[Y^{(z)}|Z=0,X]$. If the SITA assumption holds, $c(z,X)$ equals zero. Thus, the sensitivity function can describe the magnitude of the departure from the SITA assumption or the impact of the unmeasured confounders. Once we specify the sensitivity function $c(z,X)$, one can predict the mean function of the counterfactual variables conditional on X and then estimate the ATE without the SITA assumption. Li et al. [15] criticized a technical difficulty in defining the sensitivity function when covariates X contain multiple dimensions. Of note, in practical sensitivity analysis, if X is multi-dimensional, not only the functional form but also the specific coefficients for each covariate are required to be specified. Such specifications were criticized to be unlikely to accurately reflect the relationship between the departure from the SITA assumption and the potential outcomes. Li et al. [15] proposed a refinement by defining the sensitivity function as a function of the MAR-based propensity score: $c(z,e(X))=E[Y^{(z)}|Z=1,e(X)]-E[Y^{(z)}|Z=0,e(X)]$. The MAR-based propensity score is a one dimension summary of observed covariates, and this refinement made the specification of the sensitivity function much simpler.

However, in reality, even with the simplification by Li et al. [15], it is not an easy task to define a plausible range of sensitivity functions. Furthermore, their method still relies on the estimation of the MAR-based propensity score. Misspecification of the parametric model for the MAR-based propensity score may result in difficulty in interpreting the results of the sensitivity analysis.

3.2 | The Marginal Sensitivity Model

Tan [19] proposed the marginal sensitivity model, which describes a relaxation of the SITA assumption. The model assumes a single sensitivity parameter, which permits the presence of the unmeasured confounders U but restricts the extent of selection bias that can be attributed to these confounders. One can specify a parameter λ, and then the following inequality is supposed to hold:

$$1/\lambda\le\frac{e(X_i,U_i)/(1-e(X_i,U_i))}{\hat{e}(X_i)/(1-\hat{e}(X_i))}\le\lambda,$$

where $e(X_i,U_i)$ refers to the true propensity score measuring all covariates, and $\hat{e}(X_i)$ refers to the estimated MAR-based propensity score. The single parameter λ, that is, the odds ratio (OR) between true propensity score and estimated propensity score, can control the degree of unconfoundedness. When $\lambda=1$, the inclusion of additional confounders has no effect on the treatment odds. This implies that the allocation of the treatment is not influenced by confounding factors. That is, the SITA assumption holds. Increasing λ represents the allowance for a stronger extent to which the SITA assumption is violated. Tan [19] proposed a sensitivity analysis method to assess how the estimates based on the nonparametric likelihood change under the violation of the SITA assumption.

By introducing the marginal sensitivity model, the sensitivity analysis for unmeasured confounders can be applied to the IPW estimator under the MNAR. If U was observed, one can estimate μ_1 with

$$\left(\sum_{i=1}^{n}\frac{Z_i}{e(X_i,U_i)}\right)^{-1}\sum_{i=1}^{n}\frac{Z_iY_i}{e(X_i,U_i)} \tag{8}$$

In practice, U is unobserved, and $\hat{\mu}_1$ in Equation (8) actually makes no sense. However, under the marginal sensitivity model,

λ can link the unobserved true propensity score and the estimated MAR-based propensity score, so that it is possible to evaluate bounds of (8) under some constraints. That is

$$\begin{aligned} &\max \text{or} \min \left(\sum_{i=1}^{n} \frac{Z_i}{e(X_i, U_i)}\right)^{-1} \sum_{i=1}^{n} \frac{Z_i Y_i}{e(X_i, U_i)} \\ &\text{subject to} \quad 1/\lambda^1 \leq \frac{e(X_i, U_i)/(1-e(X_i, U_i))}{\hat{e}(X_i)/(1-\hat{e}(X_i))} \leq \lambda^1 \end{aligned} \tag{9}$$

where λ^1 is the pre-specified constant, which describes the upper and lower bounds of the discrepancy of the true propensity score from the estimated MAR-based propensity score for the estimation of μ_1. As long as the true propensity score for all the subjects satisfies the constraint, the true μ_1 should be bounded by the minimum and maximum of (9) asymptotically. It is possible to have an interval for μ_0 in a similar way. This method under the marginal sensitivity model was proposed firstly by Zhao [18]. In this method, the sensitivity parameter λ^1 quantifies the extent to which the SITA assumption is violated. However, it still suffers from defining a plausible range for the sensitivity parameter and reliance on correct specification of the MAR-based propensity score model.

It was criticized that the interval obtained by (9) may not be tight, and the interval was asymptotically conservative [20]. Dorn and Guo [20] proposed the quantile balancing method, a refinement based on the marginal sensitivity model. Let $F(y|x,z) = P(Y \leq y|X = x, Z = z)$, and the quantile function is defined by $Q_t(x,z) = \inf\{q : F(q|x,z) \geq t\}$. For bounding μ_1, the quantile balancing method solves the following optimization problem:

$$\max \text{or} \min \left(\sum_{i=1}^{n} \frac{Z_i}{e(X_i, U_i)}\right)^{-1} \sum_{i=1}^{n} \frac{Z_i Y_i}{e(X_i, U_i)} \tag{10}$$

$$\text{subject to} \quad \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \hat{Q}_\tau(X_i, 1) \end{pmatrix} \left(\frac{Z_i}{e(X_i, U_i)} - \frac{Z_i}{\hat{e}(X_i)}\right) = 0 \tag{11}$$

$$1/\lambda^1 \leq \frac{e(X_i, U_i)/(1-e(X_i, U_i))}{\hat{e}(X_i)/(1-\hat{e}(X_i))} \leq \lambda^1 \tag{12}$$

where $\tau = \frac{\lambda^1}{1+\lambda^1}$ and $\hat{Q}_\tau(X_i, 1)$ is estimated with some quantile regression models [20]. Bounding μ_0 and the ATE can be achieved in a similar way. For μ_0, we may use an alternative value λ^0 for λ^1 in the constraint corresponding to (12). Throughout this paper, we suppose $\lambda^1 = \lambda^0$ and the common value is denoted by λ. The quantile balancing method refined Zhao's sensitivity analysis method [18] by adding the quantile function to balance the treatment assignment Z over the true propensity score at the population level. This additional constraint based on the estimated quantile function ensured the asymptotic optimality of the interval obtained by solving (10). Although it solves asymptotic conservativeness in Zhao's method [18], it still suffers from the misspecification of the estimated MAR-based propensity score. Moreover, the quantile function also requires specifying some parametric models or machine learning-related methods.

4 | The Proposed Sensitivity Analysis Method

4.1 | Bounds for ATE With the Estimating Equation Constraints

We begin with the bound for μ_1. Let $e^1(X_i, U_i)$ denote the true propensity score for subjects in the treated group. Let us consider constructing the upper bound of μ_1 by solving the following optimization problem:

$$\overline{\mu}_1^+ = \max \quad \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i Y_i}{e^1(X_i, U_i)} \tag{13}$$

$$\text{subject to} \quad \delta \leq e^1(X_i, U_i) \leq 1 - \delta \tag{14}$$

$$\sum_{i=1}^{n} g(X_i) \left(1 - \frac{Z_i}{e^1(X_i, U_i)}\right) = 0 \tag{15}$$

In a similar way, to obtain the lower bound of μ_1, let us consider the following problem:

$$\overline{\mu}_1^- = \min \quad \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i Y_i}{e^1(X_i, U_i)} \tag{16}$$

$$\text{subject to} \quad \delta \leq e^1(X_i, U_i) \leq 1 - \delta \tag{17}$$

$$\sum_{i=1}^{n} g(X_i) \left(1 - \frac{Z_i}{e^1(X_i, U_i)}\right) = 0 \tag{18}$$

The constraints (14) and (17) come from the positivity assumption (Assumption 2), which is a fundamental assumption in causal inference. We regard δ in (14) and (17) as a sensitivity parameter. The constraints (15) and (18) come from the estimating equation for the outcome-dependent propensity score (6). As mentioned, the estimating Equation (6) cannot necessarily identify the true propensity score model uniquely from a parametric model. However, according to the law of large numbers, it holds that

$$\sum_{i=1}^{n} g(X_i) \left(1 - \frac{Z_i}{e^1(X_i, U_i)}\right) \xrightarrow{p} E\left[g(X)\left(1 - \frac{Z}{e^1(X, U)}\right)\right] = 0 \tag{19}$$

Then, the true propensity scores should satisfy the constraints (15) and (18) asymptotically, and therefore, μ_1 should be included in the interval $[\mu_1^-, \mu_1^+]$ asymptotically.

Let us consider the inverse of the true propensity score, denoted by $w_i^1 = (e^1(X_i, U_i))^{-1}$, as the decision variable. Then, optimization problems (13) and (16) become a linear programming problem with linear constraints:

$$\min \text{or} \max \quad \frac{1}{n} \sum_{i=1}^{n} Z_i Y_i w_i^1 \tag{20}$$

$$\text{subject to} \quad \frac{1}{1-\delta} \leq w_i^1 \leq \frac{1}{\delta} \tag{21}$$

$$\sum_{i=1}^{n} g(X_i)(1 - Z_i w_i^1) = 0 \tag{22}$$

Compared to the quantile balancing method, which is nonlinear optimization and requires estimation of the quantile

functions, our proposal can be solved time-efficiently with the interior-point method or the simplex algorithm for linear programming and then tractable with standard software for mathematical programming.

The bound for μ_0 can be constructed in a similar way as follows. Let $e^0(X_i, U_i)$ denote the true propensity score for subjects in the control group, and similarly consider the weight $w_i^0 = (1 - e^0(X_i, U_i))^{-1}$ as the decision variable. Then the interval $[\overline{\mu_0^-}, \overline{\mu_0^+}]$ can be obtained by solving the following linear programming problem:

$$\text{min or max} \quad \frac{1}{n}\sum_{i=1}^{n}(1 - Z_i)Y_i w_i^0 \tag{23}$$

$$\text{subject to} \quad \frac{1}{1-\delta} \leq w_i^0 \leq \frac{1}{\delta} \tag{24}$$

$$\sum_{i=1}^{n} g(X_i)(1 - (1 - Z_i)w_i^0) = 0 \tag{25}$$

We obtain bounds for ψ by $[\overline{\mu_1^-} - \overline{\mu_0^+}, \overline{\mu_1^+} - \overline{\mu_0^-}]$.

Generally, in the estimation of the propensity score, the dimension of $g(X_i)$ should be equal to the number of unknown parameters in the parametric model for the propensity score. In the proposed sensitivity analysis method, one can impose more constraints by increasing the dimension of $g(X_i)$, thereby yielding a narrower bound obtained by the linear programming problems (20) and (23). $g(X_i)$ can be any function of X_i. Suppose that there are K covariates: $X^\top = (X_{i,1}, X_{i,2}, \ldots, X_{i,K})$. Then, $g(X_i)$ can be like

$$g(X_i) = \begin{pmatrix} 1 \\ X_{i,1} \\ \vdots \\ X_{i,K} \end{pmatrix} \quad \text{or} \quad g(X_i) = \begin{pmatrix} 1 \\ X_{i,1} \\ \vdots \\ X_{i,K} \\ X_{i,1}^2 \\ \vdots \\ X_{i,K}^2 \end{pmatrix} \quad \text{or} \quad g(X) = \begin{pmatrix} 1 \\ X_{i,1} \\ \vdots \\ X_{i,K} \\ X_{i,1}^2 \\ \vdots \\ X_{i,K}^2 \\ \vdots \end{pmatrix} \tag{26}$$

As long as the resulting constraints give us feasible solutions to the optimization problems, the proposed method is expected to narrow the bound by simply increasing the dimension of $g(X_i)$, since greater flexibility on the choice of $g(X_i)$ is allowed.

The IPW estimators (2) and (3) do not satisfy the population boundedness property: the IPW estimator can be beyond the range of the outcome [29, 30]. On the other hand, the SIPW estimators (4) and (5) satisfy it. The objective functions (13) and (16) have the form of the IPW estimator. If we set the first element of $g(X_i)$ to be 1, as seen in (26), the IPW estimator agrees with the SIPW estimator. Consequently, it is sufficient to consider using a more computationally tractable, unstabilized form as the objective function and suggested to consider 1 as the first element of $g(X_i)$.

Dorn and Guo [20] also considered the condition (19), but criticized that $g(X_i)$ should involve infinitely many moment conditions. Coupled with the constraint (12), they showed that the infinitely many constraints can be replaced with a single constraint of the quantile balancing (11). This simplification with the quantile balancing is realized with the OR-based constraint (12). In practical sensitivity analysis of observational studies, the bounds for the ATE obtained by optimizing (13) and (16) are generally compared with a specific threshold, such as zero, to ensure the robustness of the results. Therefore, there is no need to introduce an infinite number of constraints, as it is sufficient to increase the dimension of $g(X_i)$ to ensure the robustness of the causal inference in an observational study. In addition, the quantile function must be estimated and then be subject to assumptions in modeling and estimation, although Dorn and Guo [20] tried to minimize the risk of misspecification by introducing flexible models. The authors provide several machine learning-related estimation methods, which might yield notably different bounds from each other in their simulation study [20]. One advantage of the proposed method is that it does not rely on messy estimation in the quantile regression. Simply by increasing the dimension of $g(X_i)$, we can try to make the bound narrower. It is important that the proposed method can provide bounds without relying on the condition (12). We must note that the proposed method is not completely free from specification of sensitivity parameters; we need to specify δ in (14) and (17) to ensure boundedness of the maximum and the minimum of (13) and (16), respectively. Although the proposed method shares the challenge of specifying sensitivity parameters with Li's method and the quantile balancing method, it differs in that the sensitivity function $c(z, e(X))$ in Li's method and the parameter λ in the quantile balancing method quantify the magnitude of violation from the SITA assumption, which are very difficult to specify, whereas δ should be set sufficiently small, and some insights on its specification can be obtained by the MAR-based propensity scores estimated with some parametric model.

4.2 | Bounds for ATE With the Additional OR-Based Constraints

The bounds may be too wide to give any meaningful information without (12). If this is the case, the constraint (12) can be incorporated into the proposed method. For optimization problem (20), we consider:

$$\text{min or max} \quad \frac{1}{n}\sum_{i=1}^{n} Z_i Y_i w_i^1 \tag{27}$$

$$\text{subject to} \quad \sum_{i=1}^{n} g(X_i)(1 - Z_i w_i^1) = 0$$

$$\frac{1 + (\lambda^1 - 1)\hat{e}(X_i)}{\lambda^1 \hat{e}(X_i)} \leq w_i^1 \leq \frac{1 + \hat{e}(X_i)(1/\lambda^1 - 1)}{\hat{e}(X_i)(1/\lambda^1)} \tag{28}$$

The additional constraint (28) is derived from the marginal sensitivity model (9), in which $\hat{e}(X_i)$ refers to the estimated MAR-based propensity score depending merely on measured covariates, and λ^1 refers to the OR between true propensity score and the estimated MAR-based propensity score for the estimation of μ_1. Note that, after introducing the constraint (28), the optimization problem remains a linear programming problem. The constraint (28) defines the range for w_i^1 and can replace the constraint (21) once λ^1 is specified. The positivity assumption is

inherently satisfied, allowing the constraint (21) to be omitted without affecting the optimization problem's feasibility or structure as long as the MAR-based propensity score satisfies the positivity assumption. With $g(X_i)$ fixed, the addition of the constraint (28) can further narrow the bound obtained by solving the linear programming (27). The bound for μ_0 can be obtained in a similar way:

$$\text{min or max} \quad \frac{1}{n}\sum_{i=1}^{n}(1-Z_i)Y_i w_i^0 \tag{29}$$

$$\text{subject to} \quad \sum_{i=1}^{n} g(X_i)(1-(1-Z_i)w_i^0) = 0$$

$$\frac{\hat{e}(X_i)}{\lambda^0(1-\hat{e}(X_i))} + 1 \le w_i^0 \le \frac{\lambda^0\hat{e}(X_i)}{1-\hat{e}(X_i)} + 1 \tag{30}$$

where λ^0 refers to the OR between true propensity score and the estimated MAR-based propensity score for the estimation of μ_0. Similarly, the constraint (30) adequately defines the range for w_i^0 and can replace the constraint (24) without violating the positivity assumption.

Although the specifications of λ^1 and λ^0 may require additional assumptions and prior knowledge, the OR-based constraints (28) and (30) can be compatible with the estimating equation constraints of the proposed method and further narrow the bound for μ_1 and μ_0. The proposed method offers flexibility. First, the proposed method avoids introducing the additional assumptions, and it provides a worst-case bound that can be narrowed by increasing the dimension of $g(X_i)$ and including more covariates. Secondly, when the ranges of λ^1 and λ^0 can be reasonably determined, a further narrower bound can be achieved. We may specify different values for λ^1 and λ^0. In this paper, for simplicity, we use the common value $\lambda = \lambda^1 = \lambda^0$. Furthermore, as done by Dorn and Guo [20], estimation error for the bounds can be accounted for by using the bootstrap confidence intervals of the lower and upper bounds. One may hope to make the bounds tighter by introducing $g(X_i)$ of the higher dimension. A concern is that putting more variables may lead to unreliable bounds of less stability. The bootstrap samples would also be useful to evaluate how stable the resulting bounds are: if the number of the bootstrap samples of feasible solutions of the linear programming is small, the resulting bounds should be carefully interpreted.

5 | Simulation Study

5.1 | Data Generation

In this section, we investigate the performance of the proposed method and compare it with Dorn and Guo's method [20] based on the marginal sensitivity model and quantile balancing over several simulated datasets. The simulation settings followed Morikawa and Kim's framework [31], allowing us to evaluate the performance of the proposed method when encountering unidentifiability issues. In our simulation, we considered generating five covariates $\overline{X}_i^\top = (X_{i,1}, X_{i,2}, X_{i,3}, X_{i,4}, X_{i,5})$ from the normal distribution with

$$X_{i,1} \sim \mathcal{N}(0,1)$$

$$X_{i,k+1}|X_{i,k} = x_{i,k} \sim \mathcal{N}\left(\frac{-x_{i,k}}{3}, 1\right), \quad k = 1,2,3,4.$$

Here, $\{X_{i,1}, X_{i,2}, X_{i,3}, X_{i,4}\}$ were regarded as measured covariates, while $X_{i,5}$ was regarded as an unmeasured confounder. By setting different $a_1^{[s]}$ in two scenarios ($s = 1, 2$), where $a_1^{[1]} = 0.0775$ and $a_1^{[2]} = 0.998$, the outcome was generated as follows:

$$\mu^{(1)}(\overline{x}_i) = a_1^{[s]} + 0.4x_{i,1} + 0.4x_{i,2} + 0.6x_{i,1}x_{i,2} + 0.5x_{i,3} - 0.7x_{i,4} + 0.2x_{i,5},$$

$$\mu^{(0)}(\overline{x}_i) = 0.0654 + 0.2x_{i,1} + 0.1x_{i,2} + 1.2x_{i,1}x_{i,2} + 0.2x_{i,3} - 0.3x_{i,4} + 0.6x_{i,5},$$

$$Y_i^{(1)}|(\overline{X}_i = \overline{x}_i) \sim \mathcal{N}\left(\mu^{(1)}(\overline{x}_i), \frac{1}{4}\right),$$

$$Y_i^{(0)}|(\overline{X}_i = \overline{x}_i) \sim \mathcal{N}\left(\mu^{(0)}(\overline{x}_i), \frac{1}{4}\right).$$

The treatment assignment $Z_i \in \{0,1\}$ was generated by the Bernoulli distribution with

$$P(Z_i = 1|\overline{X}_i = \overline{x}_i, Y_i^{(1)} = y_i^{(1)}) = \frac{1}{1+\exp(-0.904 + 0.5x_{i,1} + 0.5x_{i,2} + 0.5x_{i,3} - 0.2x_{i,4} - x_{i,5} + 0.3y_i^{(1)})}.$$

We simulated 1,000 observational studies with $n = 1,000$ subjects for each scenario. For each simulated study, the bounds for the ATE were calculated by the proposed method, as well as the quantile balancing method [20]. In applying the quantile balancing method, the linear quantile regression on $\{X_{i,1}, X_{i,2}, X_{i,3}, X_{i,4}\}$ was applied. For the constraint (12), we estimated the MAR-based propensity score $\hat{e}(X_i)$ by the logistic regression model with $\{X_{i,1}, X_{i,2}, X_{i,3}, X_{i,4}\}$ and applied 5-fold cross-fitting in the estimation of the quantile function. For Scenario 1, the mean of MAR-based propensity score was 0.6682, with values ranging from 0.0016 to 0.9997. For Scenario 2, the mean of MAR-based propensity score was 0.6219, ranging from 0.0341 to 0.9950. The application of the quantile balancing method was conducted by utilizing the R package provided by Dorn and Guo [20]. The proposed method with (20) and (23) did not involve any estimation of the MAR-based propensity score. The OPTMODEL Procedure of SAS (SAS Institute Inc, Cary, North Carolina) was used for solving the linear programming problems to obtain the bounds for ATE in the proposed method. We considered four settings of the specification of $g(X)$:

1. D1 includes 1, and the linear and quadratic terms for all the observed covariates;
2. D2 includes D1 plus all two-variable interactions;
3. D3 includes D1 plus the cubic terms for all the observed covariates and all interactions;
4. D4 includes D3 plus the quartic and quintic terms for all the observed covariates, $X_{i,4}^3(X_{i,3} - X_{i,2})(X_{i,1} + 3X_{i,2}/2)$, and $X_{i,4}^3(X_{i,3} - X_{i,2})/(X_{i,1} - X_{i,3})(X_{i,1} + 3X_{i,2}/2)$.

5.2 | Performance of the Proposed Method With the Estimating Equation Constraints

In the proposed method, we considered estimating the bounds for μ_1 under the constraints (21) and (22) and for μ_0 under the constraints (24) and (25). Thus, the validity of the proposed method would be dependent on the choice of δ. To discuss this point, we checked the distributions of the true propensity score in the simulated datasets. In the datasets under Scenario 1, with $\delta = 0.1, 0.01$, and 0.001, 18.39%, 0.29%, and 0.01% true propensity scores did not satisfy the conditions (21) and (24), respectively. The corresponding proportions for the datasets under Scenario 2 were 2.47%, 0.01% and 0.00%, respectively. Thus, with $\delta = 0.1$, the constraints seemed not to hold, whereas setting $\delta = 0.01$ or less was relevant. Table 1 shows the averages of the upper and the lower bounds and the coverage probability of the bounds for the ATE in the proposed method with several settings of δ and $g(X_i)$, where the coverage probability was defined as the proportion that the lower and the upper bounds covered the true ATE. The left panels of Figures 1 and 2 show the boxplots of the lower and upper bounds for the ATE when δ was set to 0.01 in the two scenarios, respectively. We computed the averages of $Y^{(1)}$ and $Y^{(0)}$ over all simulated datasets and regarded their subtraction as the true ATE, which is depicted by a solid horizontal line in Figures 1 and 2. In Scenario 1, as shown in Table 1 and the left panel of Figure 1, the proposed method demonstrated excellent performance in terms of the coverage probability when δ was set to 0.01 or smaller. For Scenario 2, as shown in Table 1 and the left panel of Figure 2, the coverage probability of the bounds was even excellent when δ was set to 0.1. In addition, the bounds (D2, D3, and D4) in Scenario 2 effectively excluded the null. For the quantile balancing method, the right panels of Figures 1 and 2 present the boxplots of the bound for the ATE based on different ORs in the two scenarios, respectively; the detailed estimates are summarized in Table 2. As the decrease of OR, the quantile balancing method gave a narrower bound with sacrificing the coverage probability. The proposed method provided feasible solutions for all the 1,000 simulated datasets with different settings of δ and $g(X_i)$, except for one setting (Scenario 2, D4, and $\delta = 0.1$), in which the coverage probability was almost 1 among 952 simulated datasets with feasible solutions and the bounds for ATE were the narrowest. In this case, the proposed method gave the worst-case bounds with an average length of 1.09, which was less than the length of the bound obtained by assuming OR to be 2 in the quantile balancing method. In words, by increasing the dimension of the function alone, we could narrow down the length of the bound obtained by the proposed method to the level of the quantile balancing method with OR specified as 2. The results complied with our expectations that (1) the worst-case bound obtained by the proposed method could cover the true ATE without any additional assumptions and (2) by increasing the dimension of $g(X_i)$, narrowing of our bound could be achieved.

5.3 | Performance of the Proposed Method With the Additional OR-Based Constraints

To evaluate the effects of extra OR-based constraints in the proposed method, we provided the averages of the lower and upper bounds for the ATE in the proposed method with several settings of the OR and $g(X_i)$ in Table 3. The introduction of the OR-based constraint (28) narrowed the bounds for ATE. Under the same OR specified as the quantile balancing, for instance, when λ was set to be 2, even in the simplest setting D1, the proposed method outperformed the quantile balancing method in terms of the length

TABLE 1 | Summary of the proposed method with several settings of δ and $g(X)$ over 1 000 simulated datasets in two scenarios.

		Scenario 1 (True ATE: 0.21)				Scenario 2 (True ATE: 1.13)			
$g(X)$	δ	Bound*	Length**	Coverage***	Feasibility****	Bound	Length	Coverage	Feasibility
D1	0.1[a]	[−0.44,1.41]	1.85	1.00	1 000	[0.32,2.41]	2.09	1.00	1 000
	0.01[b]	[−1.36,2.12]	3.48	1.00	1 000	[−0.60,3.09]	3.68	1.00	1 000
	0.001[c]	[−1.43,2.18]	3.61	1.00	1 000	[−0.93,3.18]	4.11	1.00	1 000
D2	0.1	[0.10,1.08]	0.97	0.93	1 000	[0.90,2.07]	1.17	1.00	1 000
	0.01	[−0.38,1.56]	1.94	1.00	1 000	[0.32,2.53]	2.21	1.00	1 000
	0.001	[−0.41,1.59]	2.00	1.00	1 000	[0.05,2.57]	2.53	1.00	1 000
D3	0.1	[0.14,1.04]	0.91	0.69	1 000	[0.91,2.05]	1.14	1.00	1 000
	0.01	[−0.33,1.50]	1.83	1.00	1 000	[0.34,2.47]	2.13	1.00	1 000
	0.001	[−0.35,1.53]	1.88	1.00	1 000	[0.09,2.50]	2.41	1.00	1 000
D4	0.1	[0.21,0.97]	0.75	0.45	1 000	[0.93,2.02]	1.09	0.99	952
	0.01	[−0.19,1.37]	1.56	0.95	1 000	[0.36,2.43]	2.07	1.00	1 000
	0.001	[−0.21,1.38]	1.59	0.97	1 000	[0.13,2.45]	2.32	1.00	1 000

*The averages of the lower and upper bounds.
**The difference between the averages of the lower and upper bounds.
***The proportion of inclusion of the true ATE between the lower and upper bounds.
****The number of the resampled datasets in the proposed method, in which feasible solution of the linear programming can be obtained.
[a]18.39% and 2.47% of the true propensity scores did not satisfy the constraints (21) or (24), respectively, in scenarios 1 and 2.
[b]0.29% and 0.01% of the true propensity scores did not satisfy the constraints (21) or (24), respectively, in scenarios 1 and 2.
[c]0.01% and 0.00% of the true propensity scores did not satisfy the constraints (21) or (24), respectively, in scenarios 1 and 2.

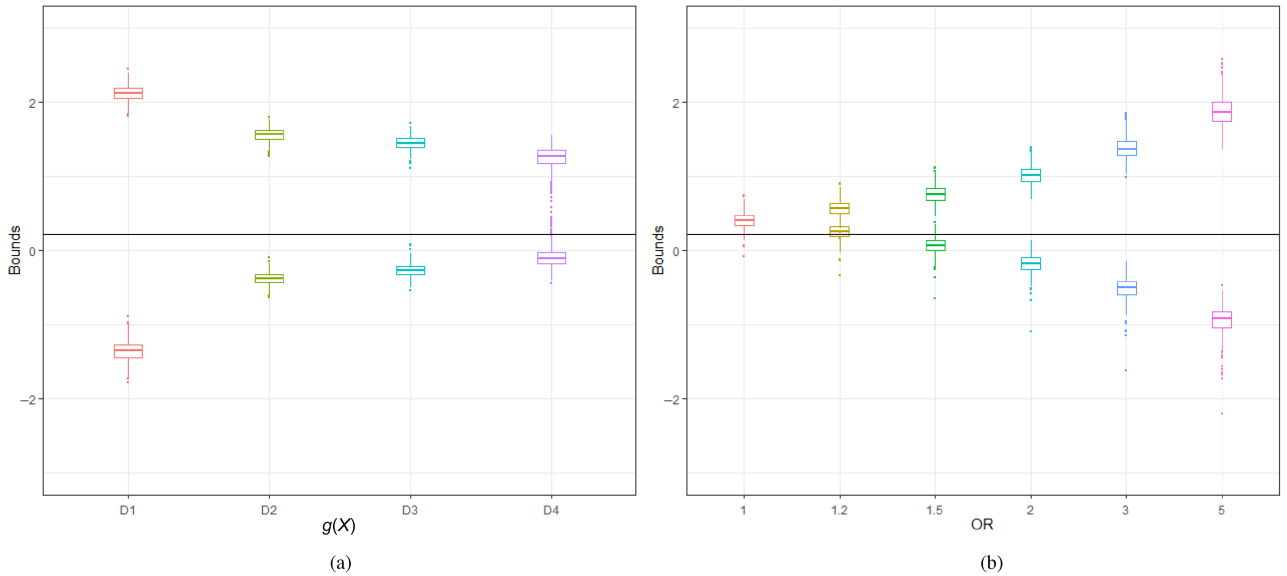

FIGURE 1 | Boxplots of the lower and upper bounds for ATE obtained by the proposed method (left panel) and the quantile balancing method (right panel) over 1 000 simulated datasets in Scenario 1. (a) The proposed method ($\delta = 0.01$). (b) The quantile balancing method.

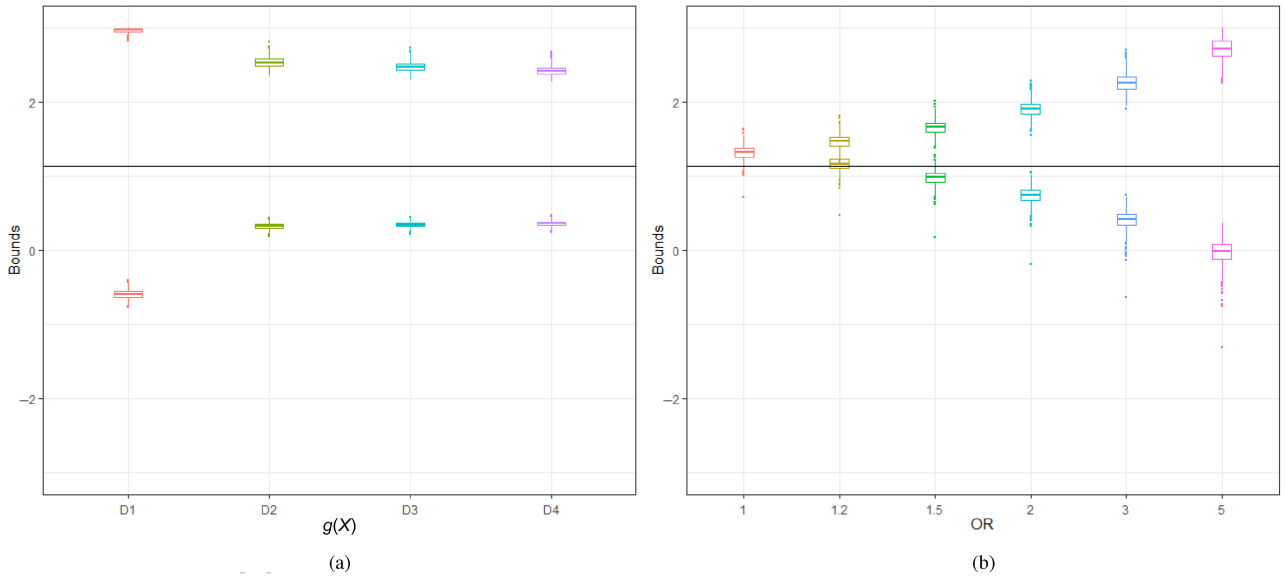

FIGURE 2 | Boxplots of the lower and upper bounds for ATE obtained by the proposed method (left panel) and the quantile balancing method (right panel) over 1,000 simulated datasets in Scenario 2. (a) The proposed method ($\delta = 0.01$). (b) The quantile balancing method.

TABLE 2 | Summary of the quantile balancing method with several settings of OR over 1 000 simulated datasets in two scenarios.

	Scenario 1 (True ATE: 0.21)			Scenario 2 (True ATE: 1.13)		
λ	Bound*	Length**	Coverage***	Bound	Length	Coverage
1	[0.01,0.01]	/	/	[1.31,1.31]	/	/
1.2	[−0.13,0.14]	0.27	0.48	[1.16,1.47]	0.31	0.36
1.5	[−0.29,0.31]	0.61	0.89	[0.98,1.66]	0.68	0.96
2	[−0.51,0.54]	1.04	1.00	[0.74,1.90]	1.17	1.00
3	[−0.81,0.87]	1.68	1.00	[0.40,2.26]	1.85	1.00
5	[−1.22,1.35]	2.57	1.00	[−0.03,2.74]	2.77	1.00

*The averages of the lower and upper bounds.
**The difference between the averages of the lower and upper bounds.
***The proportion of inclusion of the true ATE between the lower and upper bounds.

TABLE 3 | Summary of the proposed method with several settings of OR and $g(X)$ over 1 000 simulated datasets in two scenarios.

		Scenario 1 (True ATE: 0.21)				Scenario 2 (True ATE: 1.13)			
λ[*]	$g(X)$	Bound[**]	Length[***]	Coverage[****]	Feasibility[*****]	Bound	Length	Coverage	Feasibility
1.5	D1	[−0.05,0.27]	0.32	0.71	273	[1.18,1.57]	0.39	0.33	739
	D2	[0.15,0.29]	0.15	0.89	9	[1.41,1.58]	0.17	0.00	94
	D3	/	/	/	/	[1.41,1.56]	0.15	0.00	1
	D4	/	/	/	/	/	/	/	/
2	D1	[−0.24,0.46]	0.70	0.97	903	[0.96,1.79]	0.83	0.95	994
	D2	[0.11,0.40]	0.29	0.93	385	[1.31,1.68]	0.37	0.01	861
	D3	[0.16,0.34]	0.18	0.43	7	[1.35,1.63]	0.27	0.00	170
	D4	/	/	/	/	/	/	/	/
3	D1	[−0.53,0.76]	1.29	1.00	994	[0.66,2.05]	1.39	1.00	1000
	D2	[−0.02,0.57]	0.59	0.99	961	[1.14,1.85]	0.71	0.45	999
	D3	[0.07,0.50]	0.43	0.95	461	[1.21,1.78]	0.57	0.15	897
	D4	[0.19,0.49]	0.29	1.00	2	[1.26,1.70]	0.44	0.03	37
5	D1	[−0.83,1.05]	1.88	1.00	999	[0.35,2.33]	1.98	1	1000
	D2	[−0.17,0.77]	0.94	1.00	996	[0.97,2.02]	1.05	1.00	1000
	D3	[−0.06,0.70]	0.76	0.99	928	[1.02,1.96]	0.94	0.95	995
	D4	[0.05,0.60]	0.55	0.89	96	[1.13,1.84]	0.71	0.57	284

[*] λ represents that λ^1 and λ^0 take the same value.
[**] The averages of the lower and upper bounds.
[***] The difference between the averages of the lower and upper bounds.
[****] The proportion of inclusion of the true ATE between the lower and upper bounds.
[*****] The number of the resampled datasets in the proposed method, in which a feasible solution of the linear programming can be obtained.

of the bounds. The corresponding lengths of the bounds for ATE were 1.29 and 1.39 in Scenario 1 and Scenario 2, respectively, while their counterparts in the quantile balancing method were 1.68 and 1.85. The results shown in Table 3 consistently complied with our expectations that (1) the OR-based constraint ensured that the positivity assumption was inherently satisfied and (2) the introduction of the OR-based constraint could further narrow the bounds for ATE achieved by increasing the dimension of $g(X_i)$. We observed that the number of the simulated datasets of feasible solutions was very small with complicated specification of $g(X)$ and the OR-based constraint of small λ. It suggested that the OR-based constraints could conflict with the estimating equation constraints. This implies that tight OR-based constraints may be replaced with the estimating equation constraints, which are free from sensitivity parameters.

5.4 | Performance of the Proposed Method When the Positivity Assumption is Violated

We further considered the simulation in a supplementary scenario where practical violation of the positivity assumption exists to investigate the performance of the proposed method. Since the only difference between Scenario 1 and Scenario 2 lies in the intercept of the outcome Y_1, we only consider the modification of Scenario 1 in generating datasets. The data generation process for the covariates $\overline{X}_i^\top = (X_{i,1}, X_{i,2}, X_{i,3}, X_{i,4}, X_{i,5})$ and the outcome was the same with it was in Scenario 1.The difference lay in the treatment assignment. Specifically, when the value of $X_{i,1}$ was greater than or equal to 1, the corresponding treatment Z_i was 1 with probability 1, indicating a violation of the positivity assumption. This circumstance may arise in practice; for example, when a subject's age exceeds a certain threshold, he/she will be assigned to the treatment group certainly. In all other cases, the treatment assignment was generated following Scenario 1. We considered the same settings of the specification of $g(X)$ as in Section 5.1. The modified one is referred to as Scenario 3.

Table 4 illustrates the results of Scenario 3 by the proposed method. With D1 and D2, the proposed method had 100% coverage with a reasonable length of the bounds. On the other hand, with D3 and D4, the proposed method did not have feasible solutions, and therefore the corresponding results were not presented in Table 4. These results indicated that the proposed method may not work well when the positivity assumption was violated.

6 | Application

In this section, we apply the proposed method to real-world data from the TONE study [32]. This study aimed to evaluate the effectiveness of a designated exercise program in preventing dementia among the elderly. In this study, scores in five cognitive domains (attention, memory, visuospatial function, language, and reasoning) were used to quantify the level of cognition. We considered estimating the effectiveness of the exercise program on the attention domain, which was regarded as a continuous variable. The confounders included age, sex, education level (1/0: high/low), and attention scores at the baseline.

TABLE 4 | Summary of the proposed method in a supplementary scenario where the positivity assumption is violated.

		Scenario 3 (True ATE: 0.21)			
$g(X)$	δ	Bound*	Length**	Coverage***	Feasibility****
D1	10^{-2}[a]	[−1.14,1.95]	3.09	1.00	1 000
	10^{-10}[b]	[−1.25,2.04]	3.29	1.00	1 000
	10^{-30}[c]	[−2.74,3.67]	6.41	1.00	1 000
D2	10^{-2}	[−0.23,1.38]	1.60	1.00	1 000
	10^{-10}	[−0.27,1.42]	1.68	1.00	1 000
	10^{-30}	[−1.21,2.35]	3.56	1.00	1 000

*The averages of the lower and upper bounds.
**The difference between the averages of the lower and upper bounds.
***The proportion of inclusion of the true ATE between the lower and upper bounds.
****The number of the resampled datasets in the proposed method, in which a feasible solution of the linear programming can be obtained.
[a]16.17% of the true propensity scores did not satisfy the constraints (21) or (24) in Scenario 3.
[b]15.89% of the true propensity scores did not satisfy the constraints (21) or (24) in Scenario 3.
[c]0% of the true propensity scores did not satisfy the constraints (21) or (24) in Scenario 3.

In the primary analysis, a total of 935 participants were included, in which 234 were in the exercise program group and 701 in the control group. We utilized the IPW estimator to adjust imbalances in covariates between the exercise program and control groups. In the primary analysis, we used logistic regression to estimate MAR-based propensity scores, and the unknown parameters were estimated by the maximum likelihood method with the above variables as explanatory variables. The mean of estimated MAR-based propensity score was 0.2503, with values ranging from 0.0116 to 0.8210. Significantly large between-group imbalances were observed in age, attention scores at baseline, and education level before weighting inversely by the estimated MAR-based propensity scores. The between-group imbalances were effectively eliminated after weighting, indicating that IPW significantly enhanced balance across the two groups. The IPW point estimate of the ATE was 4.09 with a 95% confidence interval of [2.97, 5.22]. Therefore, the result of the primary analysis indicated a significantly positive effect of the exercise program on the improvement of the attention level. This finding was consistent with some previous randomized controlled trials [33, 34]. However, a meta-analysis of observational studies [35] reported an insignificantly positive result, suggesting that the robustness of the positive effects needs further confirmation. Then, a sensitivity analysis was necessary.

In the sensitivity analysis, we estimated the bounds of the ATE both without and with the OR-based constraints. When there were no OR-based constraints, considering the range of the estimated MAR-based propensity scores [0.0116, 0.8210], δ was specified as 0.01. We applied the proposed method with four settings of $g(X_i)$:

1. E1 includes 1 and the linear term of all the covariates;
2. E2 includes 1 and the linear and quadratic terms of all the covariates;
3. E3 includes E2 plus the interaction between age and attention score at baseline;
4. E4 includes E2 plus all two-variable interactions.

For the quantile balancing method, the quantile function was estimated using linear quantile regression with 5-fold cross-fitting with the R package by Dorn and Guo [20]. The result by the proposed method is given in Table 5. In addition to the plain bounds, we calculated the confidence intervals of the upper and lower bounds with 1 000 bootstrap samples. BootLower and BootUpper refer to the lower and upper bounds of the 95% bootstrap confidence interval for the lower and upper bounds of ATE, respectively. The column "Feasibility" refers to the number of the resampled datasets in which feasible solutions of the linear programming in the proposed method can be obtained. At first, we examined the worst-case bounds based on the proposed method (20) without the OR-based constraints (Table 5). The resulting bounds with the four settings of $g(X_i)$ are presented in the row with $\lambda = /$ in Table 5. Even with $g(X_i)$ of a higher dimension (E4), the lower bound was less than the null value 0, indicating that the proposed methods did not eliminate concerns on unmeasured confounders. Since the length of the bounds was wide and not necessarily well interpretable, the OR-based constraints were added, and the bounds were calculated with (27) and (29). The results with $\lambda = 2, 3$, and 5 are also shown in Table 5. Our proposal indicated that the worst-case bound could exclude the null if it was supposed that the OR between the true propensity score and estimated MAR-based propensity score was at most 2. For reference, we also applied the quantile balancing method. The corresponding bounds with the OR-based constraints based on the quantile balancing method (10) are presented in Table 6. The bounds were similar to ours, and, of note, when subjected to the same OR-based constraint, our bound was tighter than that of the quantile balancing method. We noticed that the smaller OR and greater complexity of $g(X_i)$ caused fewer feasible solutions of the linear programming in resampled datasets. Without the additional OR-based constraint, the results of the bootstrap kept stable and feasible. However, when a small OR was assumed, occasions to have feasible solutions in resampled datasets drastically decreased. Thus, the estimating equation constraints are not necessarily compatible with the OR-based constraints, in particular when a small λ is set.

Instead of adding the OR-based constraints, by incorporating more covariates, we tried to make the bounds tighter. We further

TABLE 5 | Bounds of the ATE for the attention score in TONE study by the proposed method.

$g(X)$	λ	Lower bound	Upper bound	Length	BootLower*	BootUpper*	Feasibility**
E1	/***	−10.57	18.30	28.87	−11.50	19.27	1 000
	2	0.57	6.61	6.04	−0.41	7.46	981
	3	−1.18	8.59	9.77	−2.36	9.49	1 000
	5	−3.36	10.93	14.29	−4.85	11.78	1 000
E2	/	−9.81	17.75	27.56	−10.83	18.39	1 000
	2	1.66	5.50	3.84	−0.19	6.92	360
	3	−0.92	8.18	9.10	−1.86	8.84	890
	5	−3.19	10.50	13.69	−4.26	11.16	998
E3	/	−9.70	17.23	26.93	−10.46	17.49	1 000
	2	/	/	/	−0.18	6.70	61
	3	−0.40	7.45	7.85	−1.58	8.29	595
	5	−3.09	9.91	13.00	−3.76	10.51	965
E4	/	−8.74	16.54	25.28	−8.26	16.17	991
	2	/	/	/	0.07	6.19	46
	3	−0.25	7.16	7.41	−1.00	7.83	551
	5	−2.67	9.48	12.15	−2.96	10.03	913

*The lower and upper bounds of 95% bootstrap confidence interval for the lower and upper bounds of ATE, respectively.
**The number of the resampled datasets in the proposed method, in which feasible solution of the linear programming can be obtained.
***The rows with/in the λ column show the results without OR-based constraint; in this case, δ is set to 0.01.

TABLE 6 | Bounds of the ATE for the attention score in TONE study by the quantile balancing method.

λ	Lower bound	Upper bound	Length	BootLower*	BootUpper*
1	4.09	4.09	/	3.04	5.30
1.2	3.29	4.92	1.63	2.21	6.14
1.5	2.32	5.97	3.64	1.20	7.21
2	1.13	7.29	6.17	−0.01	8.59
3	−0.61	9.15	9.77	−1.85	10.58
5	−2.80	11.56	14.36	−4.25	13.27

*The lower and upper bounds of the 95% bootstrap confidence interval for the lower and upper bounds of ATE, respectively.

included the baseline scores of four other cognitive domains (memory, visuospatial function, language, and reasoning) as confounders in the sensitivity analysis. The results by the proposed method without the OR-based constraint (20) are shown in Table 7. The worst-case bounds for the ATE in Table 7 achieve great tightness. In some settings (E3 and E4), the worst-case bounds even excluded the null, thus indicating the robustness of the primary analysis without any additional OR-based constraints. For reference, Table 8 presents the results by the quantile balancing method with including all cognitive domains. The proposed method could have bounds of less length than the quantile balancing method when λ was specified as 3 or 5. Nevertheless, we observed comparatively small numbers of bootstrap samples with feasible solutions. Thus, the successful exclusion of the null with E3 and E4 would be subject to instability, and it could not eliminate concerns against unmeasured confounders completely. The instability may come from the conflicts between the increasing dimensions of $g(X_i)$ and the constraint with δ. In this analysis, we further explored the impact of varying the δ on the feasibility of the bootstrap samples, particularly in challenging settings (E3, E4) where feasibility was low with $\delta = 0.01$. For instance, in the settings of E3 and E4, we observed that reducing δ from 0.01 to 10^{-30} resulted in a significant increase in feasibility, from 323 to 807 and 220 to 743, respectively. On the other hand, specification of smaller δ resulted in wider bounds. For example, in E3 (Table 7), the lower bound was estimated to be -12.13 with $\delta = 10^{-30}$ (Feasibility: 807), whereas 0.31 with $\delta = 10^{-2}$ (Feasibility: 323) and 0.16 with $\delta = 10^{-10}$ (Feasibility: 331). With $\delta = 10^{-2}$ and $\delta = 10^{-10}$, the proposed method successfully excluded the null value, but the small number of the bootstrap samples with feasible solutions successfully informed us of caution in interpreting the observed narrow bounds. We also applied the proposed method with the OR-based constraints; we added the OR-based constraint to the method reported in Table 7. We observed there was no feasible solution in almost all the cases of E1–E4. Thus, we employed a simpler specification of $g(X)$ consisting of all the linear terms of variables. With the OR-based constraint of 5.5, we observed a bound on [1.15, 5.47], whereas the quantile balancing method provided a similar bound with the OR-based constraint of 1.2. In practice, this OR value of the constraint was likely to

TABLE 7 | Bounds of the ATE for the attention score in the TONE study by the proposed method with $g(X)$ of additional four domains.

$g(X)$	δ	Lower bound	Upper bound	Length	BootLower*	BootUpper*	Feasibility**
E1	10^{-2}	−6.75	14.93	21.68	−8.84	15.84	1 000
	10^{-10}	−7.08	15.67	22.76	−9.12	16.34	1 000
	10^{-30}	−22.05	29.50	51.54	−23.93	29.60	1 000
E2	10^{-2}	−2.08	11.14	13.22	−4.29	12.44	657
	10^{-10}	−2.23	11.28	13.51	−4.48	12.59	661
	10^{-30}	−14.92	23.75	38.67	−16.58	24.54	931
E3	10^{-2}	0.31	8.32	8.01	−3.14	11.33	323
	10^{-10}	0.16	8.47	8.31	−3.28	11.46	331
	10^{-30}	−12.13	20.88	33.02	−14.34	22.72	807
E4	10^{-2}	0.70	7.82	7.12	−2.45	10.51	220
	10^{-10}	0.50	7.98	7.48	−2.61	10.64	226
	10^{-30}	−11.92	20.84	32.76	−13.49	22.07	743

*The lower and upper bounds of the 95% bootstrap confidence interval for the lower and upper bounds of ATE, respectively.
**The number of the resampled datasets in the proposed method, in which a feasible solution of the linear programming can be obtained.

TABLE 8 | Bounds of the ATE for the attention score in the TONE study by the quantile balancing method with an additional four domains.

λ	Lower bound	Upper bound	Length	BootLower*	BootUpper*
1	3.86	3.86	/	2.55	5.20
1.2	3.02	4.69	1.67	1.76	6.04
1.5	2.04	5.76	3.71	0.80	7.17
2	0.77	7.16	6.39	−0.56	8.67
3	−0.97	9.22	10.19	−2.41	10.71
5	−3.23	11.95	15.17	−5.05	13.25

*The lower and upper bounds of the 95% bootstrap confidence interval for the lower and upper bounds of ATE, respectively.

exceed its appropriateness and reasonableness. Thus, applying the estimating equation constraints would be helpful to obtain a more interpretable and realistic OR-based constraint.

7 | Discussion

Interest in drawing medical evidence from the real-world data has been rapidly growing, and the number of papers reporting results of real-world data analyses with the confounder adjustment has been substantially increasing. The propensity score analysis is now routinely applied in the analysis of observational studies. However, almost all the papers only report the results of the propensity score matching and/or the IPW method by the propensity score and do not address the important issue of the unmeasured confounders. Since the issue of residual confounding is always left as a limitation in the analysis of observational studies, it is very important to develop sensitivity analysis methods that are easily applicable and rely on fewer assumptions. In this paper, we proposed a simple sensitivity analysis method based on the IPW method. To our best knowledge, all the existing sensitivity analysis methods for the IPW estimator rely on some untestable assumptions on the departure from the SITA assumption. Although they provide very useful tools to address potential impacts of unmeasured confounders, there are still concerns with potential violation of the assumption. The proposed method requires only minimal assumptions and can construct the bounds for the ATE, completely free from any quantification of the departure from the SITA assumption. Although it may not give sufficiently informative bounds of small width, showing the bounds based on minimal assumptions would be useful as a basis in addressing the potential impacts of the residual confounding. The proposed method can easily incorporate the OR-based constraints for the departure from the SITA assumption to give tighter bounds. The resulting method with the additional OR-based constraints corresponds to the quantile methods by Dorn and Guo [20]. Moreover, the proposed method can be easily applied with linear programming, avoiding the estimation of the quantile functions, and empirically, it gave likely tighter bounds in our simulation study. Comparing with the elegant theory of the quantile balancing method by Dorn and Guo [20], the proposed method is based on a very simple idea. We believe that the proposed method is practical and our strategy would be useful in addressing the issue of residual confounding.

By incorporating more constraints with $g(X)$ of higher dimension, one may have tighter bounds with the proposed method. It motivates us to collect as many potential confounders as possible in conducting observational studies. On the other hand, we do not have any clear guidance on how to define $g(X)$. Putting more constraints would be desirable to make the bounds tighter. More specifically, to conduct a comprehensive sensitivity analysis

using the proposed method for the unmeasured confounders in observational studies, we recommend following procedures. One can begin by determining $g(X)$ with simple terms, such as linear terms for all covariates, and gradually consider more complicated terms like quadratic and interaction terms based on the data and study context. Initially, it is recommended to perform the proposed sensitivity analysis method without the OR-based constraint by linear programming to obtain a bound under less restrictive assumptions. If the resulting bounds can exclude the null or provide meaningful information, there is no need to incorporate the additional OR-based constraint but to consider to specify a plausibly small value for δ. When a small δ is specified, the stability of the bounds may increase, probably at the cost of wider bounds. On the other hand, when the bounds are wide and not able to provide meaningful information, the OR-based constraint can be incorporated to narrow the bounds. Comparing with the quantile balancing method, the proposed method could provide narrower bounds with the same OR-based constraints specified. Thus, the proposed method with the OR-based constraint would provide less restrictive and then more interpretable results. With the OR-based constraints incorporated, the positivity assumption is inherently satisfied. However, complex specifications of $g(X)$ and strong OR assumptions may conflict with each other, resulting in relatively low feasibility even after the δ constraint is removed. This observation is intuitively understandable. The quantile balancing method successfully introduced a single constraint (11) to represent infinitely many balancing properties of the propensity score with the support of the OR-based constraint (12). It suggests that the OR-based constraint describes some parts of the balancing properties of the propensity score and the quantile balancing recovered some conditions that cannot be represented as the OR-based constraint. Since our motivation was basically on the removal of the OR-based constraint, which may not be easy to interpret, we represent the balancing property as the estimating equation constraints. We recommend to monitor the number of the bootstrap samples of feasible solutions in the above steps to evaluate whether the bound is sufficiently stable or not. In practice, we guide researchers to apply multiple settings of $g(X)$ and to report the bootstrap results for these settings in the sensitivity analysis. It is highly motivated to develop more formal guidance on the choice of $g(X)$ with certain theoretical bases. With the complexity of the problem, we leave this challenging issue as our future work. In the simulation study, we observed that the proposed method might not work when the positivity assumption was violated. Inference under violation of the positivity assumption is a very important problem. However, it is beyond the scope of our method.

As noted, our idea is simple: we remove any parametric models for the propensity score but still rely on the estimating equation for the propensity score. This simplicity would make us easily extend the idea to more complicated problems in causal inference and missing data analysis. This also warrants addressing in future research.

Acknowledgments

This research was partly supported by Grant-in-Aid for Challenging Exploratory Research (16K12403) and for Scientific Research (16H06299, 18H03208) from the Ministry of Education, Science, Sports, and Technology of Japan.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

References

1. D. B. Rubin, "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics* 6, no. 1 (1978): 34–58.

2. P. R. Rosenbaum and D. B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70, no. 1 (1983): 41–55.

3. K. Hirano and G. W. Imbens, "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization," *Health Services and Outcomes Research Methodology* 2 (2001): 259–278.

4. P. C. Austin and E. A. Stuart, "Moving Towards Best Practice When Using Inverse Probability of Treatment Weighting (IPTW) Using the Propensity Score to Estimate Causal Treatment Effects in Observational Studies," *Statistics in Medicine* 34, no. 28 (2015): 3661–3679.

5. J. K. Lunceford and M. Davidian, "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study," *Statistics in Medicine* 23, no. 19 (2004): 2937–2960.

6. M. A. Hernán, S. Hernández-Díaz, and J. M. Robins, "A Structural Approach to Selection Bias," *Epidemiology* 15, no. 5 (2004): 615–625.

7. M. L. Petersen and J. van der Laan Mark, "Causal Models and Learning From Data: Integrating Causal Modeling and Statistical Estimation," *Epidemiology* 25, no. 3 (2014): 418.

8. D. O. Scharfstein, A. Rotnitzky, and J. M. Robins, "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models," *Journal of the American Statistical Association* 94, no. 448 (1999): 1096–1120.

9. W. Miao, P. Ding, and Z. Geng, "Identifiability of Normal and Normal Mixture Models With Nonignorable Missing Data," *Journal of the American Statistical Association* 111, no. 516 (2016): 1673–1683.

10. P. R. Rosenbaum and D. B. Rubin, "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, no. 2 (1983): 212–218.

11. D. Y. Lin, B. M. Psaty, and R. A. Kronmal, "Assessing the Sensitivity of Regression Results to Unmeasured Confounders in Observational Studies," *Biometrics* 54, no. 3 (1998): 948–963.

12. J. Cornfield, W. Haenszel, E. C. Hammond, A. M. Lilienfeld, M. B. Shimkin, and E. L. Wynder, "Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions," *Journal of the National Cancer Institute* 22, no. 1 (1959): 173–203.

13. P. Ding and T. J. VanderWeele, "Sensitivity Analysis Without Assumptions," *Epidemiology* 27, no. 3 (2016): 368–377.

14. T. J. VanderWeele and P. Ding, "Sensitivity Analysis in Observational Research: Introducing the E-Value," *Annals of Internal Medicine* 167, no. 4 (2017): 268–274.

15. L. Li, C. Shen, A. C. Wu, and X. Li, "Propensity Score-Based Sensitivity Analysis Method for Uncontrolled Confounding," *American Journal of Epidemiology* 174, no. 3 (2011): 345–353.

16. C. Shen, X. Li, L. Li, and M. C. Were, "Sensitivity Analysis for Causal Inference Using Inverse Probability Weighting," *Biometrical Journal* 53, no. 5 (2011): 822–837.

17. S. Lu and P. Ding, "Flexible Sensitivity Analysis for Causal Inference in Observational Studies Subject to Unmeasured Confounding," *arXiv* Preprint arXiv.2305.1764.

18. Q. Zhao, D. S. Small, and B. B. Bhattacharya, "Sensitivity Analysis for Inverse Probability Weighting Estimators via the Percentile Bootstrap," *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 81, no. 4 (2019): 735–761.

19. Z. Tan, "A Distributional Approach for Causal Inference Using Propensity Scores," *Journal of the American Statistical Association* 101, no. 476 (2006): 1619–1637.

20. J. Dorn and K. Guo, "Sharp Sensitivity Analysis for Inverse Propensity Weighting via Quantile Balancing," *Journal of the American Statistical Association* 118, no. 544 (2023): 2645–2657.

21. D. B. Rubin, "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology* 66, no. 5 (1974): 688–701.

22. J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association* 89, no. 427 (1994): 846–866.

23. J. S. Greenlees, W. S. Reece, and K. D. Zieschang, "Imputation of Missing Values When the Probability of Response Depends on the Variable Being Imputed," *Journal of the American Statistical Association* 77, no. 378 (1982): 251–261.

24. R. Andrea, D. Scharfstein, T. L. Su, and J. Robins, "Methods for Conducting Sensitivity Analysis of Trials With Potentially Nonignorable Competing Causes of Censoring," *Biometrics* 57, no. 1 (2001): 103–113.

25. G. Verbeke, G. Molenberghs, H. Thijs, E. Lesaffre, and M. G. Kenward, "Sensitivity Analysis for Nonrandom Dropout: A Local Influence Approach," *Biometrics* 57, no. 1 (2001): 7–14.

26. J. M. Robins, "Association, Causation, and Marginal Structural Models," *Synthese* 121, no. 1/2 (1999): 151–179.

27. J. M. Robins, A. Rotnitzky, and D. O. Scharfstein, "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. M. E. Halloran and D. Berry (New York, NY: Springer, 2000), 1–94.

28. B. A. Brumback, M. A. Hernán, S. J. Haneuse, and J. M. Robins, "Sensitivity Analyses for Unmeasured Confounding Assuming a Marginal Structural Model for Repeated Measures," *Statistics in Medicine* 23, no. 5 (2004): 749–767.

29. Z. Tan, "Bounded, Efficient and Doubly Robust Estimation With Inverse Weighting," *Biometrika* 97, no. 3 (2010): 661–682.

30. J. Robins, M. Sued, Q. Lei-Gomez, and A. Rotnitzky, "Comment: Performance of Double-Robust Estimators When" inverse probability "Weights Are Highly Variable," *Statistical Science* 22, no. 4 (2007): 544–559.

31. K. Morikawa and J. K. Kim, "Semiparametric Optimal Estimation With Nonignorable Nonresponse Data," *Annals of Statistics* 49, no. 5 (2021): 2991–3014.

32. M. Sasaki, C. Kodama, S. Hidaka, et al., "Prevalence of Four Subtypes of Mild Cognitive Impairment and APOE in a Japanese Community," *International Journal of Geriatric Psychiatry* 24, no. 10 (2009): 1119–1126.

33. L. Sanders, T. Hortobágyi, E. Karssemeijer, Z. V. dE, E. Scherder, and M. Van Heuvelen, "Effects of Low-and High-Intensity Physical Exercise on Physical and Cognitive Function in Older Persons With Dementia: A Randomized Controlled Trial," *Alzheimer's Research & Therapy* 12, no. 1 (2020): 28.

34. S. E. Lamb, B. Sheehan, N. Atherton, et al., "Dementia and Physical Activity (DAPA) Trial of Moderate to High Intensity Exercise Training for People With Dementia: Randomised Controlled Trial," *BMJ* 361 (2018): k1675.

35. J. Demurtas, D. Schoene, G. Torbahn, et al., "Physical Activity and Exercise in Mild Cognitive Impairment and Dementia: An Umbrella Review of Intervention and Observational Studies," *Journal of the American Medical Directors Association* 21, no. 10 (2020): 1415–1422.