

RESEARCH ARTICLE

Model misspecification and bias for inverse probability weighting estimators of average causal effects

Ingeborg Waernbaum¹  | Laura Pazzagli² 

¹Department of Statistics, Uppsala University, Sweden and Institute for Evaluation of Labour Market and Education Policy, IFAU, Uppsala, Sweden

²Centre for Pharmacoepidemiology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

Correspondence

Ingeborg Waernbaum, Department of Statistics, Uppsala University, Sweden and Institute for Evaluation of Labour Market and Education Policy, IFAU, Uppsala, Sweden.

Email:

ingeborg.waernbaum@statistics.uu.se

Funding information

Royal Swedish Academy of Sciences; Swedish Research Council, Grant/Award Number: 2016-00703



This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract

Commonly used semiparametric estimators of causal effects specify parametric models for the propensity score (PS) and the conditional outcome. An example is an augmented inverse probability weighting (IPW) estimator, frequently referred to as a doubly robust estimator, because it is consistent if at least one of the two models is correctly specified. However, in many observational studies, the role of the parametric models is often not to provide a representation of the data-generating process but rather to facilitate the adjustment for confounding, making the assumption of at least one true model unlikely to hold. In this paper, we propose a crude analytical approach to study the large-sample bias of estimators when the models are assumed to be approximations of the data-generating process, namely, when all models are misspecified. We apply our approach to three prototypical estimators of the average causal effect, two IPW estimators, using a misspecified PS model, and an augmented IPW (AIPW) estimator, using misspecified models for the outcome regression (OR) and the PS. For the two IPW estimators, we show that normalization, in addition to having a smaller variance, also offers some protection against bias due to model misspecification. To analyze the question of when the use of two misspecified models is better than one we derive necessary and sufficient conditions for when the AIPW estimator has a smaller bias than a simple IPW estimator and when it has a smaller bias than an IPW estimator with normalized weights. If the misspecification of the outcome model is moderate, the comparisons of the biases of the IPW and AIPW estimators show that the AIPW estimator has a smaller bias than the IPW estimators. However, all biases include a scaling with the PS-model error and we suggest caution in modeling the PS whenever such a model is involved. For numerical and finite sample illustrations, we include three simulation studies and corresponding approximations of the large-sample biases. In a dataset from the National Health and Nutrition Examination Survey, we estimate the effect of smoking on blood lead levels.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Biometrical Journal* published by Wiley-VCH GmbH.

KEYWORDS

average causal effects, comparing biases, outcome model, propensity score

1 | INTRODUCTION

Identifying an average causal effect of a treatment with observational data requires adjustment for background variables that affect both the treatment and the outcome under study. Often parametric models are assumed for parts of the joint distribution of the treatment, outcome, and background variables (covariates) and large-sample properties of estimators are derived under the assumption that the parametric models are correctly specified.

Inverse probability weighting (IPW) estimators use the difference between the weighted means of the outcomes for the treatment groups as an estimator of the average causal effect. See, for example, the early paper by Hirano et al. (2003) for a nonparametric implementation of standard IPW estimators of the average causal effect. Under an assumption of no unmeasured confounding, IPW estimators reweight the observed outcomes to represent a full sample of potential outcomes, missing and observed, by letting each observed outcome account for itself and other individuals with similar characteristics. IPW estimators can be found in applied literature (see Chang et al., 2017; Kwon et al., 2015 for examples) and their properties have been generalized by Robins, Rotnitzky, and others (Robins & Rotnitzky, 1994; Robins et al., 1995, 2000) to address both confounding bias in observational studies and bias due to missing data. Vansteelandt et al. (2010) and Seaman and White (2013) provide reviews.

Properties of IPW estimators for estimating the average causal effect under the assumption that a parametric propensity score (PS) model is correctly specified have been described (il Kim et al., 2019; Lunceford & Davidian, 2004; Yao et al., 2010). Other studies of IPW estimators includes investigating properties when using different weights, for example, stabilized (Hernán et al., 2000; Hernán & Robins, 2006), normalized (Busso et al., 2014; Hirano & Imbens, 2001), or trimmed (Ma & Wang, 2020).

To decrease reliance on the choice of a parametric model and increase the efficiency of *augmented IPW* (AIPW) estimators, Robins and Rotnitzky (1994) proposed including a combination of regression adjustment with weighting based on the PS. See also the review by Seaman and Vansteelandt (2018). The AIPW estimators are referred to as doubly robust (DR) estimators (Bang & Robins, 2005; Tsiatis, 2007) because they are consistent estimators of the average causal effect if either a model for the PS or the outcome regression (OR) model is correct (Scharfstein et al., 1999). The efficiency of different DR estimators is a key property and the variances of the estimators have been described under correct specification of at least one of the models (Cao et al., 2009; Tan, 2010). When both models are correct, the estimator reaches the semiparametric efficiency bound described in Robins and Rotnitzky (1994). The large-sample properties of IPW estimators with standard, normalized and variance minimized weights, together with an AIPW estimator were studied and compared in Lunceford and Davidian (2004) under correct specification of the PS and OR models. Multiply robust estimators allow several candidate models for the PS and OR, respectively. The property of multiple robustness means that the estimators are consistent for the true average treatment effect if any one of the multiple models is correctly specified (Han & Wang, 2013).

There are few studies of doubly or multiply robust estimators under misspecification of both (all) the PS and the OR models. Kang and Schafer (2007) studied and compared the performance of an AIPW estimator for missing data under misspecification of both the PS and OR model. They concluded that many DR methods perform better than simple IPW. However, a regression-based estimator under a misspecified model was not improved upon. The paper was commented on and the relevance of the results was discussed by several authors. See, for example, Tsiatis and Davidian (2007), Tan (2007), and Robins et al. (2007). In Waernbaum (2012), a matching estimator was compared to IPW and AIPW estimators under misspecification of both the PS and OR models. Here, a robustness class for the matching estimator under misspecification of the PS model was described. Formulated in the missing data framework, Tan (2010) evaluated several semiparametric estimators, including IPW and AIPW estimators. In the evaluation, additional criteria were proposed describing robustness classes of the estimators. Vermeulen and Vansteelandt (2015) proposed a bias-reduced AIPW estimator that locally minimizes the squared first-order asymptotic bias under misspecification of both working models. One of the difficulties in the estimation of PSs occurs when the treatment groups have substantially different covariate distributions resulting in some PSs being close to zero or one. This lack of overlap raises issues with respect to model specification. Parametric binary response models, such as the commonly used probit and logit models, are similar in the middle areas of their arguments. However, for probabilities closer to zero or one, they tend to differ more resulting in the specified parametric model being more influential. In Zhou et al. (2020), misspecification of the PS linked to limited overlap is investigated for causal

effect estimators using balancing weights (Li et al., 2018). Comparing IPW estimators with and without trimming with overlap weights, matching weights, and entropy weights, they find in extensive simulations that the latter three methods outperform the former (IPW and trimmed IPW) with respect to bias, root-mean-squared error and coverage (Zhou et al., 2020).

In this paper, we describe two commonly used IPW estimators and a prototypical AIPW estimator of the average causal effect under the assumption that none of the working models is correctly specified. For this purpose, we study the difference between the probability limit of the estimator under model misspecification and the true average causal effect. The purpose of this definition of the bias is that the estimators under study converge to a well-defined limit, that is not, however, necessarily consistent for the true average causal effect. We study the biases of the (A)IPW estimators and compare them under the same misspecification of the PS model. The three estimators contain an error involving the ratio of the true PS and the limiting misspecified PS; however, the error affects the estimators in different ways. As the biases for the three estimators can be in different directions, we describe sufficient and necessary conditions using inequalities involving the absolute value of the biases. For a simple and a normalized IPW estimator, we show that the normalization in general moderates the bias due to the PS model misspecification. Comparing the IPW estimators to the AIPW estimator, the biases provide a means to describe when two wrong models are better than one, which would normally be the case for a moderate misspecification of the outcome model. Three simulation studies are performed to investigate the biases for finite samples. The data-generating processes and the misspecified models from the simulation designs are also used for numerical approximations of the large-sample properties derived in the paper.

The paper proceeds as follows. Section 2 presents the model and theory together with the estimators and their properties when the working models are correctly specified. Section 3 presents a general approach and associated assumptions to study model misspecification. In Section 4, the generic biases are derived and comparisons between the estimators are performed. We present three simulation studies in Section 5 containing both finite sample properties of the estimators and numerical large-sample approximations. We apply the estimators under study on an observational dataset in Section 6 where we evaluate the effect of smoking on blood lead levels, and thereafter we conclude with a discussion.

2 | MODEL AND THEORY

The potential outcome framework defines a causal effect as a comparison of potential outcomes that would be observed under different treatments (Rubin, 1974). Let X be a vector of pretreatment variables, referred to as covariates, T a binary treatment, with realized value $T = 1$ if treated and $T = 0$ if control. Under SUTVA (Rubin, 1980), the causal effect of the treatment is defined as a contrast between two potential outcomes, for example, the difference, $Y(1) - Y(0)$, where $Y(1)$ is the potential outcome under treatment and $Y(0)$ is the potential outcome under the control treatment. The observed outcome Y is assumed to be the potential outcome for each level of the treatment $Y = TY(1) + (1 - T)Y(0)$, so that the data vector that we observe is (T_i, X_i, Y_i) , where $i = 1, \dots, n$ are assumed independent and identically distributed copies. In the remainder of the paper, we will drop the subscript i for the random variables when not needed. Since each individual only can be subject to one treatment, either $Y(1)$ or $Y(0)$ will be missing. If the treatment is randomized, the difference between the sample averages of the treated and controls will be an unbiased estimator of the average causal effect $\Delta = E[Y(1) - Y(0)]$, the parameter of interest. In the following, we will use the notation $\mu_1 = E[Y(1)]$, $\mu_0 = E[Y(0)]$ for the marginal expectations and $\mu_1(X) = E(Y(1)|X)$, $\mu_0(X) = E(Y(0)|X)$ for their conditional counterparts. We denote the probability of being treated conditional on the covariates, the PS by $e(X) = P(T = 1|X)$. When the treatment is not assigned randomly, common identification criteria include assumptions of no unmeasured confounding and overlap:

Assumption 1: (No unmeasured confounding). $Y(t) \perp T|X, t = 0, 1$,

Assumption 2: (Overlap). $\eta < P(T = 1|X = x) < 1 - \eta, \forall x$ and some $\eta > 0$,

where the assumption that $P(T = 1|X = x)$ is bounded away from zero and one guarantees the existence of a consistent estimator (Khan & Tamer, 2010). Under Assumptions 1 and 2, we can estimate the average causal effect by weighting the observed outcomes with the inverse of the PSs, because

$$\Delta = E \left[\frac{TY}{e(X)} - \frac{(1 - T)Y}{1 - e(X)} \right],$$

leading to an estimator $\hat{\Delta}_{IPW_1}$ defined by:

$$\hat{\Delta}_{IPW_1} = \frac{1}{n} \sum_{i=1}^n \frac{T_i Y_i}{\hat{e}(X_i)} - \frac{1}{n} \sum_{i=1}^n \frac{(1-T_i) Y_i}{1 - \hat{e}(X_i)}. \quad (1)$$

A common version of the simple IPW estimator in (1) is an IPW estimator $\hat{\Delta}_{IPW_2}$ with normalized weights

$$\hat{\Delta}_{IPW_2} = \left(\sum_{i=1}^n \frac{T_i}{\hat{e}(X_i)} \right)^{-1} \sum_{i=1}^n \frac{T_i Y_i}{\hat{e}(X_i)} - \left(\sum_{i=1}^n \frac{1-T_i}{1 - \hat{e}(X_i)} \right)^{-1} \sum_{i=1}^n \frac{(1-T_i) Y_i}{1 - \hat{e}(X_i)}. \quad (2)$$

Using parametric IPW, we assume a finite-dimensional model for $e(X) = e(X, \beta)$, $\beta \in R^p$.

Assumption 3: (PS model). The PS $e(X)$ follows a model $e(X, \beta)$ parameterized by $\beta = (\beta_1, \dots, \beta_p)$, and $\hat{e}(X)$ is the estimated PS $e(X, \hat{\beta})$, with $\hat{\beta}$ a consistent estimator of β .

Under Assumptions 1–3, the IPW estimators are consistent estimators of the average causal effect Δ with asymptotic distribution $\sqrt{n}(\hat{\Delta}_{IPW_k} - \Delta) \sim N(0, \sigma^2 IPW_k)$, $k = 1, 2$. Asymptotic properties of (1) and (2) are described in Lunceford and Davidian (2004) under an assumption of a logistic regression model for the treatment assignment.

Similar to the modeling of the PS, it can be assumed that the OR follows a parametric model $\mu_t(X) = \mu_t(X, \alpha_t)$, $\alpha_t \in R^p$, $t = 0, 1$.

Assumption 4: (OR model). The conditional expectation, $\mu_t(X) = E(Y(t)|X)$, $t = 0, 1$, follows a model $\mu_t(X, \alpha_t)$, $t = 0, 1$ parameterized by $\alpha_t = (\alpha_{t1}, \dots, \alpha_{tq_t})$, and $\hat{\mu}_t(X)$ is the estimated OR $\mu_t(X, \hat{\alpha}_t)$, with $\hat{\alpha}_t$ a consistent estimator of α_t .

In addition, we study a prototypical AIPW estimator (Lunceford & Davidian, 2004; Tsiatis, 2007)

$$\begin{aligned} \hat{\Delta}_{AIPW} = & \frac{1}{n} \sum_{i=1}^n \frac{T_i Y_i - (T_i - \hat{e}(X_i)) \hat{\mu}_1(X_i)}{\hat{e}(X_i)} \\ & - \frac{1}{n} \sum_{i=1}^n \frac{(1-T_i) Y_i + (T_i - \hat{e}(X_i)) \hat{\mu}_0(X_i)}{1 - \hat{e}(X_i)}. \end{aligned} \quad (3)$$

Under Assumptions 1–4 and regularity conditions, the large-sample distribution is $\sqrt{n}(\hat{\Delta}_{AIPW} - \Delta) \sim N(0, \sigma_{AIPW}^2)$.

For example, a model assumption for the treatment assignment could be a logistic regression model $e(X, \beta) = [1 + \exp(-X'\beta)]^{-1}$ with $\hat{e}(X)$ the fitted values of the PS when $\hat{\beta}$ is a maximum likelihood estimator of β . The OR model could be a linear model $\mu_t(X, \alpha_t) = X'\alpha_t$ where $\hat{\mu}_t(X)$, $t = 0, 1$ are the fitted values when $\hat{\alpha}_t$ is the ordinary least squares estimator.

3 | MODEL MISSPECIFICATION

Our interest lies in the behaviors of the estimators when the PS and the OR models are misspecified. For this purpose, we replace Assumptions 3 and 4 with two other assumptions defining the probability limit of the estimators under a general misspecification. The misspecifications will further be used to define a general bias of the IPW and DR estimators. When the PS is misspecified an estimator, for example, a quasi-maximum likelihood estimator (QMLE) is not consistent for β in Assumption 3. However, a probability limit for an estimator under model misspecification exists under general conditions, see, for example, White (1982, Theorem 2.2) for QMLE or Wooldridge (2010, Section 12.1) and Boos and Stefanski (2013, Theorem 7.1) for estimators that can be written as a solution of an estimating equation (M-estimators).

In the following, and as an alternative to Assumptions 1 and 4, we will assume that such limits exist. Below we define an estimator $\hat{e}^*(X)$ of the PS under a misspecified model $e(X, \beta^*)$.

Assumption 5 (Misspecified PS-model parameters). Let $\hat{\beta}^*$ be an estimator under model misspecification, $e(X, \beta^*)$, then $\hat{\beta}^* \xrightarrow{P} \beta^*$.

Under model misspecification, the probability limit of $\hat{\beta}^*$ is generally well defined; however, $e(X, \beta^*)$ is not equal to the true PS $e(X)$. In the following, we use the notation $\hat{e}^*(X)$ as the estimated PS under model misspecification and $e^*(X) = e(X, \beta^*)$ under Assumption 5. Below we give an example for true and misspecified parametric models, however, for Assumption 5, we do not need the existence of a true parametric model. We use the concept of quasi-maximum likelihood used for maximum likelihood estimators when parts of the distribution are misspecified.

Example 1. For one confounder X and a true PS model, $e(X, \beta) = [1 + \exp(-\beta_0 - \beta_1 X - \beta_2 X^2)]^{-1}$ assume that we misspecify the PS with a probit model $e(X, \beta^*) = \Psi(-\beta_0^* - \beta_1^* X)$, that is, we misspecify the link function and omit a second-order term. Let $\hat{\beta}^* = (\hat{\beta}_0^*, \hat{\beta}_1^*)$ be the QML estimator of the parameters in $e(X, \beta^*)$ obtained by maximizing the quasi-likelihood

$$\ln \mathcal{L} = \sum_{i=1}^n (T_i \ln e(X_i, \beta^*) + (1 - T_i) \ln(1 - e(X_i, \beta^*))).$$

Then $\hat{e}^*(X) = \Psi(-\hat{\beta}_0^* - \hat{\beta}_1^* X)$, $\hat{\beta}^* = (\hat{\beta}_0^*, \hat{\beta}_1^*) \xrightarrow{P} \beta^* = (\beta_0^*, \beta_1^*)$ under Assumption 5 and $e^*(X) = \Psi(-\beta_0^* - \beta_1^* X)$, where $\Psi(\cdot)$ is the CDF of a standard normal random variable.

When considering the existence of true and misspecified parametric models, as illustrated in Example 1, the parameters in β and the limiting parameters β^* under the misspecified model need not to be of the same dimension. For instance, the true model could contain higher order terms and interactions that are not present in the estimation model.

The next assumption concerns overlap under model misspecification.

Assumption 6: (Overlap under misspecification). $\nu < e^*(X) < 1 - \nu$, for some $\nu > 0$.

In addition to the PS model, we also consider misspecified OR models, $\mu_t(X, \alpha_t^*)$, $t = 0, 1$. Denote by $\hat{\alpha}_t^*$, $t = 0, 1$ the estimator of the parameters in $\mu_t(X, \alpha_t^*)$.

Assumption 7: (Misspecified OR model parameters). Let $\hat{\alpha}_t^*$ be an estimator under model misspecification $\mu_t(X, \alpha_t^*)$, $t = 0, 1$, then $\hat{\alpha}_t^* \xrightarrow{P} \alpha_t^*$, $t = 0, 1$.

In the following, we use the notation $\hat{\mu}_t^*(X)$ as the estimated OR under model misspecification and $\mu_t^*(X) = \mu_t(X, \alpha_t^*)$ under Assumption 7 and μ_t^* for the expected value $E[\mu_t^*(X)]$, $t = 0, 1$.

Assumptions 5 and 7 are defined for misspecified PS and OR models for the purpose of describing their influence on the estimation of Δ . The estimators (1)–(3) can be written by estimating equations where the equations solving for the PS and OR parameters are set up below the main equation for the (A)IPW estimators. See, for example, Lunceford and Davidian (2004) and Williamson et al. (2014). Assuming parametric PS and OR models, the IPW estimators correspond to solving $2 + p$ estimating equations $\sum_{i=1}^n \psi(\theta, Y_i, T_i, X_i) = 0$ for the parameters $\theta_{\text{IPW}_k} = (\mu_1, \mu_0, \beta)$, $k = 1, 2$, and for the AIPW estimator $2 + p + q_1 + q_0$ estimating equations for the parameters $\theta_{\text{AIPW}} = (\mu_1, \mu_0, \beta, \alpha_1, \alpha_0)$. Using the notation for the misspecified models in Assumptions 5 and 7, the estimating equations change according to the dimensions of the parameters β^* and α_t^* , $t = 0, 1$. A key condition for Assumptions 5 and 7 to hold is that the misspecification of the PS and/or OR provides estimating equations that uniquely define the parameters β^* and α_t^* , $t = 0, 1$, although, as a consequence of the misspecification, the resulting (A)IPW estimators will be biased. In the next section, we present the asymptotic bias for the (A)IPW estimators under study with general expressions including the limits of the misspecified PS and OR models.

4 | BIAS RESULTING FROM MODEL MISSPECIFICATION

4.1 | General biases

We study the large-sample bias of $\hat{\Delta}_{\text{IPW}_1}$, $\hat{\Delta}_{\text{IPW}_2}$, and $\hat{\Delta}_{\text{AIPW}}$ under model misspecification and define the estimators $\hat{\Delta}_{\text{IPW}_1}^*$, $\hat{\Delta}_{\text{IPW}_2}^*$, and $\hat{\Delta}_{\text{AIPW}}^*$ by replacing $\hat{e}(X)$ in Equations (1)–(3) with $\hat{e}^*(X)$. For the AIPW estimator, we additionally replace $\hat{\mu}_t(X)$ with $\hat{\mu}_t^*(X)$, $t = 0, 1$.

To assess the properties of the estimators, we assume 1, 2, 5, 6, and 7 and regularity conditions (see Appendix A.1). We evaluate the difference between the probability limits of the estimators under model misspecification and the average causal effect Δ for the (A)IPW estimators.

Under Assumptions 1–2 and 5–6, the biases under model misspecification for $\hat{\Delta}_{IPW_1}^*$ and $\hat{\Delta}_{IPW_2}^*$ are

$$\hat{\Delta}_{IPW_1}^* - \Delta \xrightarrow{p} E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - E \left[\frac{1-e(X)}{1-e^*(X)} \mu_0(X) \right] - (\mu_1 - \mu_0), \quad (4)$$

$$\hat{\Delta}_{IPW_2}^* - \Delta \xrightarrow{p} \frac{E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right]}{E \left[\frac{e(X)}{e^*(X)} \right]} - \frac{E \left[\frac{1-e(X)}{1-e^*(X)} \mu_0(X) \right]}{E \left[\frac{1-e(X)}{1-e^*(X)} \right]} - (\mu_1 - \mu_0). \quad (5)$$

Under Assumptions 1–2 and 5–7, the bias under model misspecification for $\hat{\Delta}_{AIPW}^*$ is

$$\begin{aligned} \hat{\Delta}_{AIPW}^* - \Delta \xrightarrow{p} & E \left[\frac{(e(X) - e^*(X))(\mu_1(X) - \mu_1^*(X))}{e^*(X)} \right] \\ & + E \left[\frac{[e(X) - e^*(X)](\mu_0(X) - \mu_0^*(X))}{(1 - e^*(X))} \right]. \end{aligned} \quad (6)$$

We refer to Equations (4)–(6) as the biases of the respective estimators, that is, $\text{Bias}(\hat{\Delta}_{IPW_1}^*)$, $\text{Bias}(\hat{\Delta}_{IPW_2}^*)$, and $\text{Bias}(\hat{\Delta}_{AIPW}^*)$ although they are the difference between the probability limits of the estimators and the true Δ and not the difference in expectations. The double robustness property of $\hat{\Delta}_{AIPW}^*$ is displayed by Equation (6) because if either $e(X) = e^*(X)$ or $\mu_t(X) = \mu_t^*(X)$, $t = 0, 1$, we have that $\hat{\Delta}_{AIPW}^* \xrightarrow{p} \Delta$.

4.2 | Comparisons

The consequences of model misspecification for the estimators, respectively, can further be investigated from the general biases in Equations (4)–(6).

To study the role of the model misspecification, we compare the biases in Section 4.1 for two separate parts of the estimator. The first part concerns the bias with respect to μ_1 and the second part with respect to μ_0 . The motivation behind this component-wise comparison is that if each of the estimators of μ_1, μ_0 is unbiased, then the resulting estimator of $\Delta = \mu_1 - \mu_0$ is also unbiased. The inverse relationship between the model errors $e(X)/e^*(X)$ and $(1 - e(X))/(1 - e^*(X))$ has the result that the contribution to the overall bias from μ_1 and μ_0 may, in general, be of the same sign (see Appendix A.3). We define $\text{Bias}_1(\hat{\Delta}_{IPW_1}^*)$, $\text{Bias}_1(\hat{\Delta}_{IPW_2}^*)$, and $\text{Bias}_1(\hat{\Delta}_{AIPW}^*)$ as

$$\text{Bias}_1(\hat{\Delta}_{IPW_1}^*) = E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1, \quad (7)$$

$$\text{Bias}_1(\hat{\Delta}_{IPW_2}^*) = \frac{E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right]}{E \left[\frac{e(X)}{e^*(X)} \right]} - \mu_1, \quad (8)$$

$$\text{Bias}_1(\hat{\Delta}_{AIPW}^*) = E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) (\mu_1(X) - \mu_1^*(X)) \right]. \quad (9)$$

For $\hat{\Delta}_{IPW_1}^*$, the expression in (7) shows that the bias consists of a scaling between the model error $e(X)/e^*(X)$ and the conditional outcome $\mu_1(X)$. If the distribution of $e(X)/e^*(X)$ is positively skewed resulting in over estimation of the weighted

mean $E\left[\frac{e(X)}{e^*(X)}\mu_1(X)\right]$, we see that for $\hat{\Delta}_{IPW_2}^*$, the bias is mitigated because $E[e(X)/e^*(X)] > 1$. A similar effect is obtained for a negatively skewed distribution of $e(X)/e^*(X)$ where $E[e(X)/e^*(X)] < 1$. There is no correspondence to this bias reduction for $\text{Bias}_1(\hat{\Delta}_{IPW_1}^*)$.

The sign of the two biases in (7) and in (8) depends on the covariance of the PS-model errors and the conditional outcome $\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]$, because

$$\text{Bias}_1(\hat{\Delta}_{IPW_1}^*) = \text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right] + E\left[\frac{e(X)}{e^*(X)} - 1\right]\mu_1,$$

and

$$\text{Bias}_1(\hat{\Delta}_{IPW_2}^*) = \frac{\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]},$$

implying that the biases can be in different directions for the same model misspecification. Here, $\text{Bias}_1(\hat{\Delta}_{IPW_2}^*)$ depends only on the sign of the covariance, $\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]$, whereas this is not the case for $\text{Bias}_1(\hat{\Delta}_{IPW_1}^*)$. It is not surprising that the covariance of $e(X)/e^*(X)$ and $\mu_1(X)$ plays a role for the bias of the estimators. If $\mu_1(X)$ was a constant, it could be taken out of the expectations of the first terms in (7) and (8) and the PS-model ratio, $e(X)/e^*(X)$, would be canceled by the denominator $E[e(X)/e^*(X)]$ in (8). In this case, the bias for $\hat{\Delta}_{IPW_2}^*$ would be 0.

Next, we investigate inequalities involving the absolute values of the biases in Equations (7)–(9). (All derivations are given in Appendix A.3). The results can be directly applied for μ_0 by replacing $e(X)/e^*(X)$, with $(1 - e(X))/(1 - e^*(X))$ and $\mu_1(X)$ with $\mu_0(X)$, see Appendix A.3.

First, to study the role of normalization for the IPW estimators, we compare $\text{Bias}_1(\hat{\Delta}_{IPW_1}^*)$ and $\text{Bias}_1(\hat{\Delta}_{IPW_2}^*)$.

A sufficient and necessary condition for

$$\left|\text{Bias}_1(\hat{\Delta}_{IPW_2}^*)\right| < \left|\text{Bias}_1(\hat{\Delta}_{IPW_1}^*)\right|, \tag{10}$$

is that

$$\left|\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right| < \left|E\left[\frac{e(X)}{e^*(X)}\right]E\left[\frac{e(X)}{e^*(X)}\mu_1(X)\right] - \mu_1\right|. \tag{11}$$

That is, the absolute value of the covariance between the PS-model ratio and the conditional outcome is smaller than the absolute value of $\text{Bias}_1(\hat{\Delta}_{IPW_2}^*)$ scaled with the PS-model ratio.

To study the issue of misspecifying two models instead of one, we investigate the difference between the bias of the IPW estimators (7) and (8) and the bias of the AIPW estimator (9). We give a necessary condition for the bias of the AIPW estimator to be smaller than the bias of the simple IPW estimator:

If

$$\left|\text{Bias}_1(\hat{\Delta}_{AIPW}^*)\right| < \left|\text{Bias}_1(\hat{\Delta}_{IPW_1}^*)\right|,$$

then

$$\left|E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1^*(X)\right]\right| < 2 \cdot \left|E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right]\right|. \tag{12}$$

By (12), we see that if the AIPW estimator improves upon the simple IPW estimator under misspecification of both the PS and the OR model, then the absolute value of the misspecified outcome model is less than double the absolute value of the true conditional mean under the same scaling of the PS-model error, $e(X)/e^*(X) - 1$.

A sufficient condition for the AIPW estimator to have a smaller bias than the simple IPW estimator can be expressed as a comparison between the misspecified OR model and the true conditional outcomes under the same PS-model error.

If

$$\left| E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1^*(X) \right] \right| < 2 \cdot \left| E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1(X) \right] \right|, \quad (13)$$

and $E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1^*(X) \right]$ and $E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1(X) \right]$,

are either both positive or both negative, then $|\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)|$.

To provide a numerical example, we assume a second-order model in one variable and obtain the misspecified models' limits by omitting the second order term in both the PS (logistic regression) and the OR model (linear model). We use numerical approximations to provide values for the parameters in $e^*(X)$ and $\mu_t^*(X)$, $t = 0, 1$ under the given true and misspecified models.

Example 2. For a covariate $X \sim \text{Uniform}(-2, 2)$ and a binary treatment $T \sim \text{Bernoulli}(e(X))$, we assume a logistic PS model and a linear conditional outcome and misspecified nonlinearities in the models

$$e(X) = [1 + \exp[(1 + 0.5X - 0.1X^2)]^{-1}], \quad e^*(X) = [1 + \exp(0.87 + 0.52X)]^{-1},$$

$$\mu_1(X) = 3.5 + X + 0.7X^2, \quad \mu_1^*(X) = 4.42 + 0.76X.$$

In this example, we have that the inequality in (11) holds

$$\text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right] = 0.12,$$

$$E \left[\frac{e(X)}{e^*(X)} \right] E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1 = 0.16.$$

We expect based on previous calculations that $|\text{Bias}_1(\hat{\Delta}_{\text{IPW}_2}^*)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)|$. Here, we have that $\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*) = 0.16$, $\text{Bias}_1(\hat{\Delta}_{\text{IPW}_2}^*) = 0.12$, confirming the result. To compare $\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)$ with $\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)$, we check to see that

$$E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1^*(X) \right] = 0.07,$$

$$E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1(X) \right] = 0.16,$$

are both positive, which is consistent with the sufficient conditions for

$|\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)|$ of Equation (13). To confirm, $\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*) = 0.09$ that is smaller than 0.16.

Comparing the biases between the AIPW estimator and the IPW estimator with normalized weights, we investigate the inequality $|\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_2}^*)|$. Here, we see that for $|\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_2}^*)|$ to be true, the misspecified OR model is included in the necessary condition:

$$E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) (\mu_1(X)) \right] - \frac{\left| \text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right] \right|}{E \left[\frac{e(X)}{e^*(X)} \right]} < E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1^*(X) \right]$$

$$< E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) [\mu_1(X)] \right] + \frac{\left| \text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right] \right|}{E \left[\frac{e(X)}{e^*(X)} \right]}. \quad (14)$$

By (14), we see that in order for the AIPW estimator to improve upon the normalized IPW estimator, the (PS-error scaled) outcome misspecification must lie within an interval defined by the true conditional outcome and the absolute value of the covariance of $e(X)/e^*(X)$ and $\mu_1(X)$. This means that the smaller the covariance, the greater the accuracy of the outcome model for $\hat{\Delta}_{\text{AIPW}}^*$ to be less biased than $\hat{\Delta}_{\text{IPW}_2}^*$.

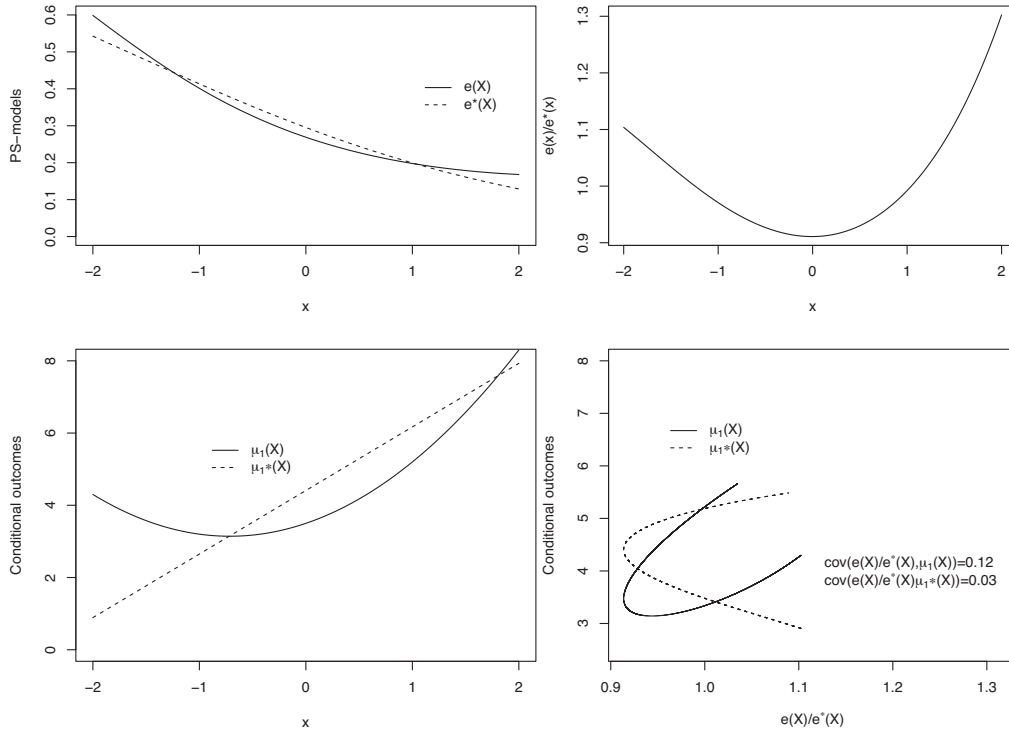


FIGURE 1 Illustration of the components of the biases using the data-generating process from Example 2. Top left: $e(X)$ and $e^*(X)$ by X ; top right: $e(X)/e^*(X)$ by X ; bottom left: $\mu_1(X)$ and $\mu_1^*(X)$ by X ; and bottom right: $\mu_1(X)$ and $\mu_1^*(X)$ by $e(X)/e^*(X)$

For a sufficient condition for $|\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_2}^*)|$, we have that if

$$E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1(X) \right] \text{ and } \text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right]$$

are either both positive or both negative, and

$$\left| E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1^*(X) \right] \right| < \left| E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1(X) \right] + \frac{\text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right]}{E \left[\frac{e(X)}{e^*(X)} \right]} \right|, \tag{15}$$

then

$$\left| E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) (\mu_1(X) - \mu_1^*(X)) \right] \right| < \left| \frac{E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right]}{E \left[\frac{e(X)}{e^*(X)} \right]} - \mu_1 \right|.$$

Illustrating the sufficient conditions with the data-generating process in Example 2

$$E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1(X) \right] = 0.16, \text{ cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right] = 0.12$$

are both positive, and

$$\left| E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) \mu_1^*(X) \right] \right| < \left| E \left[\left(\frac{e(X)}{e^*(X)} - 1 \right) (\mu_1(X)) \right] + \frac{\text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right]}{E \left[\frac{e(X)}{e^*(X)} \right]} \right|,$$

$$0.07 < 0.28,$$

in agreement with (15).

TABLE 1 Simulation Designs A–C

Models	True model		MISSPECIFIED MODEL	
	Class	Linear predictor and parameter values	Class	Linear predictor
<i>Simulation 1</i>				
		$\beta = (-1, 0.6, 0.1, -0.9, 0.1, 0.7), \alpha_0 = (3, 0.5, 0.2, 0.5, 0.2, 0.2), \alpha_1 = (4, 1.1, 0.1, 0.5, 0.3, 0.2)$		
<i>Design A</i>				
PS	Binomial, logit	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Binomial, logit	$1, X_1, X_2, X_2^2, X_3$
OR	Gaussian, identity	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Gaussian, identity	$1, X_1, X_2, X_2^2, X_3$
<i>Design B</i>				
PS	Binomial, logit	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Binomial, logit	$1, X_1, X_2, X_3$
OR	Gaussian, identity	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Gaussian, identity	$1, X_1, X_2, X_3$
<i>Design C</i>				
PS	Binomial, cauchit	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Binomial, logit	$1, X_1, X_2, X_3$
OR	Gamma, identity	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Gaussian, identity	$1, X_1, X_2, X_3$
<i>Simulation 2</i>				
		$\beta = (-0.9, 1, 1.4, 1, 0.2, 0.3), \alpha_0 = (0, 1, 1, 2, 0.5, 1), \alpha_1 = (1.5, 1.5, 2, -0.8, 0.9, 0.4)$ $1, Z_1 = X_1 + X_2 + X_3, Z_2 = X_1 + X_3 + X_4,$ $1, Z_3 = X_1 + X_2, Z_4 = X_1 + X_3$		
<i>Design A</i>				
PS	Binomial, logit	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Binomial, logit	$1, X_1, X_2, X_3$
OR	Gaussian, identity	$1, X_1, X_3, X_1^2, X_3^2, X_4$	Gaussian, identity	$1, X_1, X_3, X_4$
<i>Design B</i>				
PS	Binomial, logit	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Binomial, logit	$1, Z_1$
OR	Gaussian, identity	$1, X_1, X_3, X_1^2, X_3^2, X_4$	Gaussian, identity	$1, Z_2$
<i>Design C</i>				
PS	Binomial, logit	$1, X_1, X_2, X_1^2, X_2^2, X_3$	Binomial, logit	$1, Z_3$
OR	Gaussian, identity	$1, X_1, X_3, X_1^2, X_3^2, X_4$	Gaussian, identity	$1, Z_4$

Summarizing the results of the comparisons of the bias conditions, we note that the expected values of the product of the PS-model error and the true and misspecified conditional outcomes play important roles. Here, the covariances of the PS-model ratio and the true and misspecified conditional outcomes are two of their respective components. In Figure 1, we illustrate these parts with the data-generating processes from Example 2. The PS-model ratio deviates from 1 for both small and large values of X , but more so for larger values of X . Since both conditional outcomes $\mu_1(X)$ and $\mu_1^*(X)$ are strictly increasing, both covariances are positive ($\text{cov}[e(X)/e^*(X), \mu_1(X)] = 0.12$ and $\text{cov}[e(X)/e^*(X), \mu_1^*(X)] = 0.03$) owing to the PS-model ratio being larger for larger values of X . The interval characterization of the described conditions implies that if the two covariances are of the same magnitude, the bias of $\hat{\Delta}_{\text{AIPW}}^*$ will often be smaller than the biases of $\hat{\Delta}_{\text{IPW}_1}^*$ and $\hat{\Delta}_{\text{IPW}_2}^*$.

5 | SIMULATION STUDIES

5.1 | Design

To investigate the asymptotic biases described in Section 4 and also the finite-sample performance of $\hat{\Delta}_{\text{IPW}_1}^*$, $\hat{\Delta}_{\text{IPW}_2}^*$, and $\hat{\Delta}_{\text{AIPW}}^*$ under model misspecification, we perform three simulation studies with three different designs A–C. The first part of the simulations evaluates the finite-sample performance of the estimators and consist of 1000 replications of sample sizes 500, 1000, and 5000. In addition to the simulation results, we also give numerical approximations to the asymptotic biases by fitting the misspecified models with a large sample $N = 1,000,000$. The simulations and numerical approxi-

TABLE 2 Simulation Designs A–C

Models	True model		MISSPECIFIED MODEL	
	Class	Linear predictor and parameter values	Class	Linear predictor
<i>Simulation 3</i>				
		$\beta = (-0.20, 0.30, 0.15, 0.22, 0.15, -0.15), \alpha_0 = (9, -0.05, 0.05, -0.05), \alpha_1 = (10, -0.05, 0.05, -0.05)$ $1, M_1 = \exp(0.10X_1), M_2 = X_2(1 + X_1) + 10, M_3 = (0.04X_3 + 0.60)^2, M_4 = (X_4 + 20)^2$		
<i>Design A</i>				
PS	Binomial, logit	$1, X_1, X_2, X_3, X_4, X_5 = X_1^2$	Binomial, logit	$1, M_1, M_2, M_3, M_4$
OR	Poisson, log	$1, X_1, X_2, X_3, X_4$	Gaussian, identity	$1, M_1, M_2, M_3, M_4$
		$\beta = (-0.40, 0.60, 0.30, 0.44, 0.30, -0.30), \alpha_0 = (9, -0.05, 0.05, -0.05), \alpha_1 = (10, -0.05, 0.05, -0.05)$ $1, M_1 = \exp(0.10X_1), M_2 = X_2(1 + X_1) + 10, M_3 = (0.04X_3 + 0.60)^2, M_4 = (X_4 + 20)^2$		
<i>Design B</i>				
PS	Binomial, logit	$1, X_1, X_2, X_3, X_4, X_5 = X_1^2$	Binomial, logit	$1, M_1, M_2, M_3, M_4$
OR	Poisson, log	$1, X_1, X_2, X_3, X_4$	Gaussian, identity	$1, M_1, M_2, M_3, M_4$
		$\beta = (-0.60, 0.90, 0.45, 0.66, 0.45, -0.45), \alpha_0 = (9, -0.05, 0.05, -0.05), \alpha_1 = (10, -0.05, 0.05, -0.05)$ $1, M_1 = \exp(0.10X_1), M_2 = X_2(1 + X_1) + 10, M_3 = (0.04X_3 + 0.60)^2, M_4 = (X_4 + 20)^2$		
<i>Design C</i>				
PS	Binomial, logit	$1, X_1, X_2, X_3, X_4, X_5 = X_1^2$	Binomial, logit	$1, M_1, M_2, M_3, M_4$
OR	Poisson, log	$1, X_1, X_2, X_3, X_4$	Gaussian, identity	$1, M_1, M_2, M_3, M_4$

mations are carried out using R (R Core Team, 2020). The misspecified models are fitted with the `glm()` function and were selected in order to be well-known simple models that could have been chosen in practice by a data analyst. The link functions together with the true parameter values are given in Tables 1 and 2, which also contain the details for the misspecified models.

5.1.1 | Simulation 1

The covariates (X_1, X_2, X_3) are generated $X_1 \sim \text{Uniform}(1,4), X_2 \sim \text{Poisson}(3)$ and $X_3 \sim \text{Bernoulli}(0.4)$. Generalized linear models are used to generate a binary treatment T and potential outcomes $Y(t), t = 0, 1$ with second-order terms of X_1 and X_2 in both the PS and OR models, see Table 1. The PS distributions for the treated and controls are bounded away from 0 and 1 under the true models and under the model misspecifications (see Figure 2). The PS and OR models (for the AIPW estimator) are stepwise misspecified in the three designs (A, B, C).

- A:** a quadratic term X_1^2 is omitted in the PS and OR models;
- B:** two quadratic terms, X_1^2 and X_2^2 , are omitted in the PS and OR models;
- C:** two quadratic terms are omitted and both the OR and PS link functions are misspecified.

5.1.2 | Simulation 2

The design is inspired by the simulation study of Funk et al. (2011). We generate the same covariates (X_1, X_2, X_3, X_4) where $X_1 \sim \text{Normal}(0,1), X_2 \sim \text{Normal}(0,1), X_3 \sim \text{Uniform}(0,1)$, and $X_4 \sim \text{Normal}(0,1)$. The treatment and outcomes are generated with second-order terms of X_1 and X_2 in both the PS and OR models given in Table 1. In Figure 4, we see that

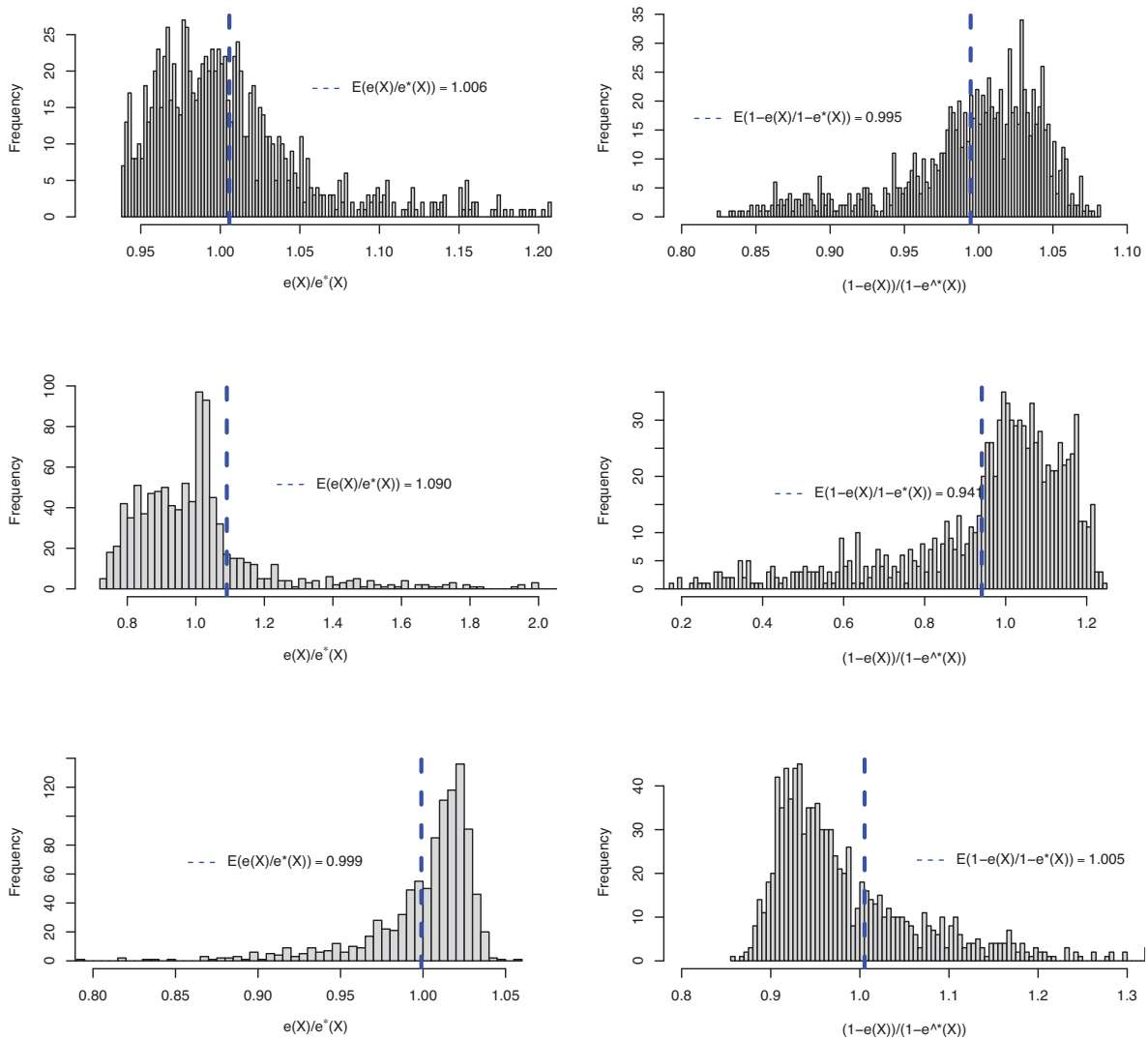


FIGURE 2 Illustration of the bias reduction in $\hat{\Delta}_{IPW_2}^*$ of the means $E[e(X)/e^*(X)]$ and $E[(1 - e(X))/(1 - e^*(X))]$ of the PS errors in Designs A, Simulation 1 (top), 2 (middle), and 3 (bottom)

the PS distributions, under both the true and misspecified models, have poorer overlap and values that are close to 0 and 1. The PS and OR models (for the AIPW estimator) are stepwise misspecified in three designs where:

- A:** two quadratic terms, X_1^2 and X_2^2 , are omitted in the PS model and X_1^2 and X_3^2 in the OR models;
- B:** two quadratic terms, X_1^2 and X_2^2 , are omitted in the PS model and X_1^2 and X_3^2 in the OR models; and transformations of the first-order terms, $Z_1 = X_1 + X_2 + X_3$ and $Z_2 = X_1 + X_3 + X_4$, are applied in the PS and the OR models, respectively;
- C:** two quadratic terms, X_1^2 and X_2^2 , are omitted in the PS model and X_1^2 and X_3^2 in the OR models, X_3 and X_4 are omitted in the PS and the OR models, respectively; and transformations of the first-order terms, $Z_3 = X_1 + X_2$ and $Z_4 = X_1 + X_3$, are applied in the PS and the OR models, respectively.

5.1.3 | Simulation 3

The design replicates the covariates and PS models of Zhou et al. (2020), in the setting referred to as medium treatment prevalence and PS distributions with good, moderate, and poor overlap (see Figure 5). In our design, the PS and OR models (for the AIPW estimator) are misspecified using the following variable transformations: $M_2 = X_2(1 + X_1) + 10$,

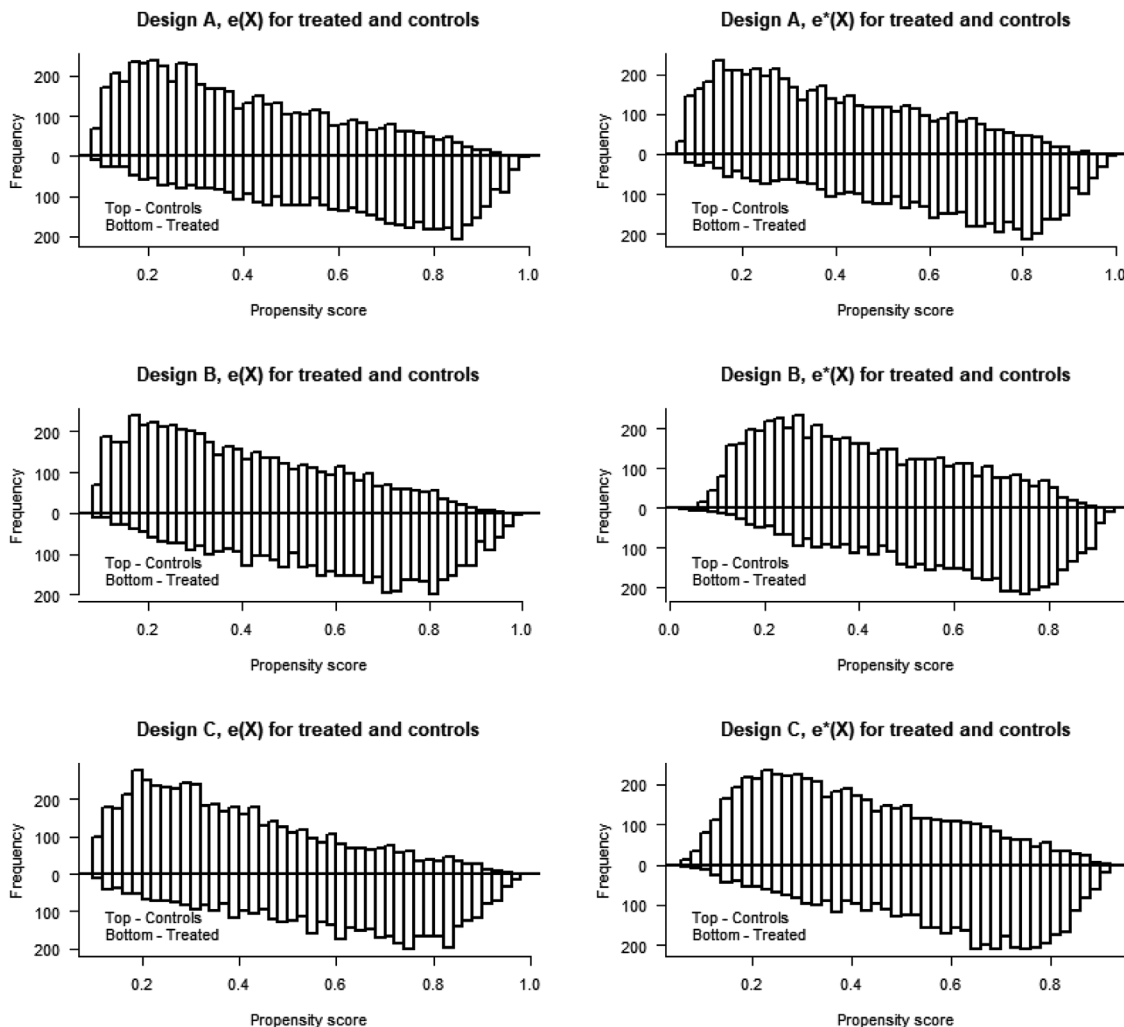


FIGURE 3 Overlap plots for the propensity score distributions, $e(X)$ and $e^*(X)$ for treated and controls for Design A (top), B (middle), and C (bottom) in Simulation 1

$M_3 = (0.04X_3 + 0.06)^2$, and $(M_1, M_4) = (\exp(0.10X_1), (X_4 + 20)^2)$. Similarly, we generate $\mathbf{X} = (X_1, \dots, X_5)$ such that $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

$$\sim \text{Normal} \left[\begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \right], \begin{pmatrix} X_3 \\ X_4 \end{pmatrix} \sim \text{Normal} \left[\begin{pmatrix} 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix} \right]. X_5 = X_1^2$$

The three simulation designs are:

- A:** good PS distribution overlap, variable transformation, and misspecified link function for the OR model;
- B:** moderate PS distribution overlap, variable transformation, and misspecified link function for the OR model;
- C:** poor PS distribution overlap, variable transformation, and misspecified link function for the OR model.

5.2 | Results

In Tables 3 and 4, we give the simulation bias, standard error, and mean squared error (MSE) of the three estimators. Tables 5–7 give numerical approximations for $\text{Bias}(\hat{\Delta}_{IPW_1}^*)$, $\text{Bias}(\hat{\Delta}_{IPW_2}^*)$, and $\text{Bias}(\hat{\Delta}_{AIPW}^*)$ using a sample size of $n = 1,000,000$. When using the true models, that is, when studying the estimators $\hat{\Delta}_{IPW_1}$, $\hat{\Delta}_{IPW_2}$, and $\hat{\Delta}_{AIPW}$, the bias is small and decreases as the sample size increases. In Simulations 1 and 2, the standard errors follow the expected order with the smallest for $\hat{\Delta}_{AIPW}$ followed by $\hat{\Delta}_{IPW_2}$ and $\hat{\Delta}_{IPW_1}$ (Lunceford & Davidian, 2004). In Simulation 3, the standard

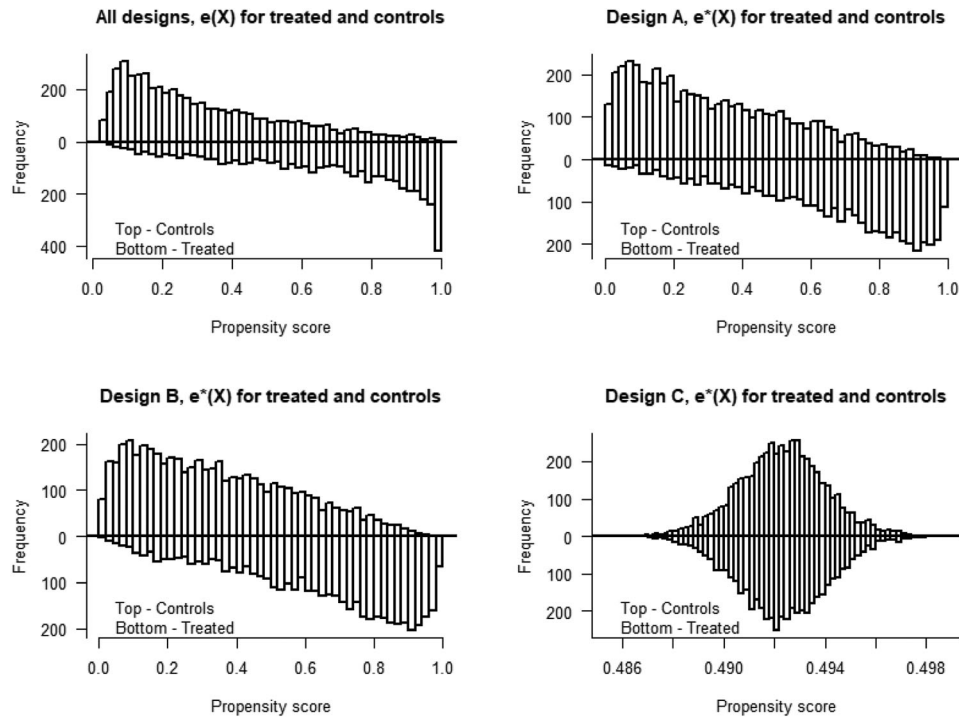


FIGURE 4 Overlap plots for the propensity score distributions, $\hat{e}(X)$ and $\hat{e}^*(X)$ for treated and controls for Design A (top), B (middle), and C (bottom) in Simulation 2

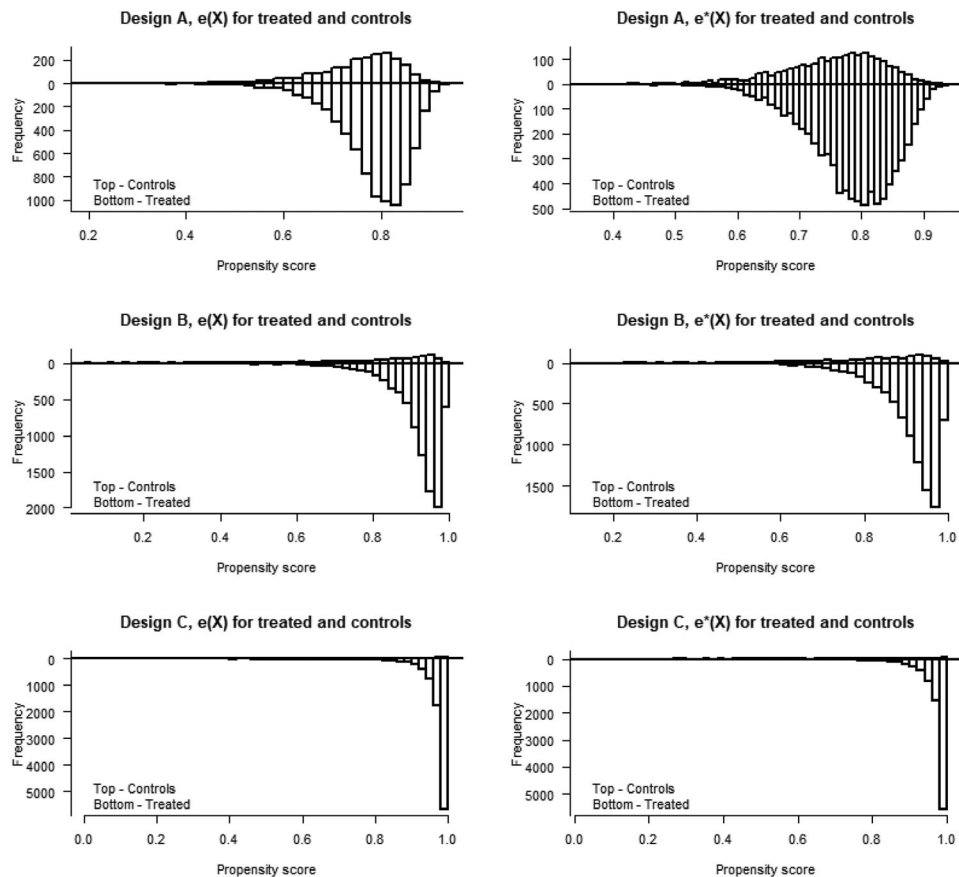


FIGURE 5 Overlap plots for the propensity score distributions, $\hat{e}(X)$ and $\hat{e}^*(X)$ for treated and controls for Design A (good overlap), B (moderate overlap), and C (poor overlap) in Simulation 3

TABLE 3 Results for Simulations 1 and 2 for sample sizes 500, 1000, and 5000. In Simulation 1, Designs A and B share the same true models. In Simulation 2, the true models are the same in Designs A–C. All true models are given in Table 1.

			Estimators								
			Simulation 1								
n	Models	Design	$\hat{\Delta}_{IPW_1}^*$			$\hat{\Delta}_{IPW_2}^*$			$\hat{\Delta}_{AIPW}^*$		
			Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
500	True	A B	0.030	0.396	0.158	0.010	0.143	0.020	0.003	0.110	0.012
	False	A	0.127	0.410	0.184	0.017	0.142	0.020	0.025	0.118	0.014
	False	B	0.316	0.372	0.238	0.044	0.124	0.017	0.026	0.117	0.014
	True	C	0.011	0.422	0.178	0.010	0.153	0.024	0.009	0.111	0.012
	False	C	0.235	0.336	0.168	-0.051	0.140	0.022	0.047	0.118	0.016
1000	True	A B	0.010	0.254	0.065	0.002	0.098	0.010	-0.003	0.075	0.006
	False	A	0.112	0.270	0.085	0.008	0.097	0.009	0.018	0.079	0.007
	False	B	0.283	0.253	0.144	0.033	0.085	0.008	0.019	0.078	0.006
	True	C	-0.003	0.278	0.077	0.005	0.106	0.011	0.003	0.078	0.006
	False	C	0.219	0.223	0.098	-0.055	0.095	0.012	0.040	0.083	0.008
5000	True	A B	0.002	0.110	0.012	0.001	0.044	0.002	-0.000	0.035	0.001
	False	A	0.108	0.120	0.026	0.009	0.044	0.002	0.022	0.037	0.002
	False	B	0.291	0.107	0.096	0.036	0.039	0.003	0.023	0.036	0.002
	True	C	-0.003	0.123	0.015	0.000	0.044	0.002	-0.000	0.033	0.001
	False	C	0.219	0.098	0.057	-0.060	0.041	0.005	0.038	0.036	0.003

			Estimators								
			Simulation 2								
n	Models	Design	$\hat{\Delta}_{IPW_1}^*$			$\hat{\Delta}_{IPW_2}^*$			$\hat{\Delta}_{AIPW}^*$		
			Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE
500	True	A B C	0.006	0.853	0.727	0.033	0.393	0.155	0.010	0.213	0.045
	False	A	0.635	0.650	0.826	0.236	0.334	0.167	0.403	0.708	0.664
	False	B	0.440	0.732	0.728	0.084	0.381	0.152	0.217	0.582	0.385
	False	C	0.608	0.607	0.737	0.271	0.361	0.203	0.597	0.708	0.858
1000	True	A B C	0.011	0.662	0.438	0.029	0.292	0.086	0.013	0.154	0.024
	False	A	0.698	0.724	1.012	0.248	0.286	0.143	0.450	0.824	0.880
	False	B	0.455	0.681	0.669	0.084	0.340	0.123	0.230	0.531	0.335
	False	C	0.619	0.413	0.554	0.266	0.257	0.137	0.606	0.461	0.579
5000	True	A B C	0.004	0.307	0.094	0.010	0.175	0.031	0.010	0.094	0.009
	False	A	0.669	0.217	0.494	0.238	0.129	0.073	0.434	0.266	0.259
	False	B	0.430	0.231	0.238	0.078	0.142	0.026	0.217	0.176	0.078
	False	C	0.646	0.327	0.524	0.282	0.183	0.113	0.650	0.422	0.601

errors of $\hat{\Delta}_{AIPW}$ and $\hat{\Delta}_{IPW_2}$ have similar magnitude. Figure 3 gives an illustration of the suppressing effect obtained by the normalization in $\hat{\Delta}_{IPW_2}^*$. For Design A in Simulations 1–3, the figure gives histograms for the simulation model errors $e(X)/e^*(X)$ and $(1 - e(X))/(1 - e^*(X))$ together with vertical lines for the corresponding means.

Under misspecification, the bias in all three simulations and all designs are close to the asymptotic approximations, at least for the largest sample size. For example, in Design C in Simulation 3 with poor overlap, the bias of $\hat{\Delta}_{IPW_2}^*$ is smaller than $\hat{\Delta}_{AIPW}^*$ when $N = 1000$ although for $N = 5000$, the bias of $\hat{\Delta}_{AIPW}^*$ is smaller than $\hat{\Delta}_{IPW_2}^*$ that is what we see in the asymptotic approximation. For the simulations with poor overlap (both in Simulation 2 and Simulation 3, Design C), the bias is very large because the model misspecification for $\hat{\Delta}_{IPW_1}$, and both $\hat{\Delta}_{IPW_2}$ and $\hat{\Delta}_{AIPW}^*$ have substantially smaller biases.

In Simulation 1, $\text{Bias}(\hat{\Delta}_{IPW_1}^*)$ is the largest. $\text{Bias}(\hat{\Delta}_{IPW_2}^*)$ and $\text{Bias}(\hat{\Delta}_{AIPW}^*)$ are similar, although $\text{Bias}(\hat{\Delta}_{AIPW}^*)$ is slightly smaller in most cases. In Simulation 2, with poorer overlap, $\text{Bias}(\hat{\Delta}_{IPW_2}^*)$ is the smallest for all three study designs, demon-

TABLE 4 Results for Simulation 3 for sample sizes 500, 1000, and 5000. The true and false models for Designs A–C are described in Table 2.

n	Models	Design	Estimators								
			Simulation 3								
			$\hat{\Delta}_{IPW_1}^*$			$\hat{\Delta}_{IPW_2}^*$			$\hat{\Delta}_{AIPW}^*$		
Bias	SD	MSE	Bias	SD	MSE	Bias	SD	MSE			
500	True	A	−0.011	0.350	0.123	−0.003	0.328	0.108	−0.003	0.329	0.108
	False	A	−0.091	0.367	0.143	0.002	0.332	0.110	−0.001	0.329	0.108
	True	B	−0.050	1.133	1.284	−0.000	0.655	0.429	−0.019	0.697	0.486
	False	B	−1.388	2.402	7.691	0.045	0.778	0.607	0.004	0.856	0.731
	True	C	0.059	3.546	12.566	−0.018	1.260	1.587	−0.033	1.430	2.043
	False	C	−10.041	23.663	660.216	0.077	1.579	2.496	0.031	4.416	19.487
1000	True	A	0.010	0.250	0.063	0.012	0.239	0.057	0.012	0.239	0.057
	False	A	−0.078	0.257	0.072	0.015	0.242	0.059	0.010	0.239	0.057
	True	B	−0.016	0.703	0.493	−0.006	0.440	0.194	−0.008	0.445	0.198
	False	B	−1.379	1.473	4.068	0.031	0.542	0.294	−0.001	0.558	0.311
	True	C	−0.139	2.347	5.522	−0.000	0.952	0.906	−0.002	1.028	1.055
	False	C	−12.065	28.245	942.510	0.118	1.364	1.872	−0.159	6.912	47.752
5000	True	A	0.008	0.110	0.012	0.009	0.107	0.012	0.009	0.107	0.012
	False	A	−0.076	0.114	0.019	0.014	0.108	0.012	0.010	0.107	0.012
	True	B	0.001	0.291	0.085	0.005	0.206	0.042	0.005	0.206	0.043
	False	B	−1.356	0.598	2.197	0.042	0.260	0.069	0.006	0.255	0.065
	True	C	−0.033	0.988	0.976	0.008	0.449	0.202	0.007	0.462	0.213
	False	C	−12.955	12.913	334.401	0.166	0.933	0.898	0.010	3.233	10.442

strating the stabilizing effect of the normalization. Here, in Designs A and B, $\text{Bias}(\hat{\Delta}_{AIPW}^*)$ is smaller than $\text{Bias}(\hat{\Delta}_{IPW_1}^*)$ but for Design C, they are similar. In Simulation 3, for the largest sample size, $\text{Bias}(\hat{\Delta}_{AIPW}^*)$ is the smallest for all three Designs A–C.

For the MSE in Simulation 1, $\hat{\Delta}_{AIPW}^*$ is the smallest; however, the difference with regard to $\hat{\Delta}_{IPW_2}^*$ is not large. In Simulation 2, Designs A and B, the MSE of $\hat{\Delta}_{IPW_2}^*$ is the smallest, followed by the MSE of $\hat{\Delta}_{AIPW}^*$ and $\hat{\Delta}_{IPW_1}^*$. For Design C, the MSE of $\hat{\Delta}_{AIPW}^*$ is greater than the MSE of $\hat{\Delta}_{IPW_1}^*$. In Simulation 3, the MSE of $\hat{\Delta}_{IPW_2}^*$ and $\hat{\Delta}_{AIPW}^*$ is similar for Designs A and B. For Design C, with poor overlap, the MSE is much smaller for $\hat{\Delta}_{IPW_2}^*$.

Studying the two different parts of the biases illustrate how-not only the variances, but also the biases, get inflated from the lack of overlap in the PS distribution. In Simulation 3 in the design with poor overlap (see Table 7), we have that $\text{Bias}_1(\hat{\Delta}_{IPW_1}^*)$ is small but $\text{Bias}_2(\hat{\Delta}_{IPW_1}^*)$ is very large. This result is owing to the ratio $(1 - e(X))/(1 - e^*(X))$ being instable because of sparse data for values close to 0. Correspondingly, we see that the mean $E(1 - e(X))/(1 - e^*(X))$ is far away from 1. In the previous section, we described the stabilizing properties of the normalized IPW, counteracting the model misspecification error. The standard errors under misspecification follow the same pattern as under the true models, which can be expected under regularity conditions from semiparametric theory, see, for example, Boos and Stefanski (2013, Chapter 7.2). In Simulation 2, the standard errors of $\hat{\Delta}_{IPW_2}^*$ are the smallest followed by $\hat{\Delta}_{AIPW}^*$ for Designs A and B and $\hat{\Delta}_{IPW_1}^*$ for Design C. In Simulation 3, the standard errors of $\hat{\Delta}_{IPW_2}^*$ are the smallest followed by $\hat{\Delta}_{AIPW}^*$ and $\hat{\Delta}_{IPW_1}^*$ for Designs B and C corresponding to moderate and poor PS distribution overlap.

To relate the simulations to the sufficient and necessary conditions derived in Section 4.2, the related expressions from the asymptotic approximations are shown in Tables 5–7. The expectations and covariances that are used for the necessary and sufficient conditions in Equations (11)–(15) can be applied to draw the corresponding conclusions of the estimators. As an example, we see that the necessary condition for the absolute values of $\text{Bias}_t(\hat{\Delta}_{AIPW}^*)$ to be smaller than the absolute value of $\text{Bias}_t(\hat{\Delta}_{IPW_1}^*)$ holds for both $t = 0$, and $t = 1$ in Simulation 1, but the same condition is not satisfied in Simulation 2.

TABLE 5 Asymptotic approximations from Designs A–C in Simulation 1

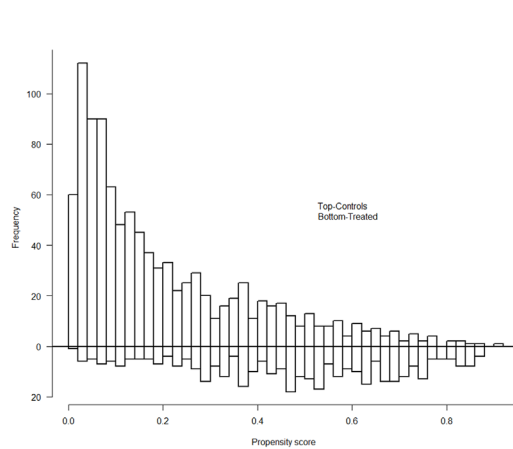
Parameter	Design		
	A	B	C
μ_1	11.128	11.128	12.130
μ_1^*	11.094	11.094	12.098
μ_0	8.630	8.627	9.633
μ_0^*	8.579	8.575	9.582
$\text{Bias}(\hat{\Delta}_{\text{IPW}_1}^*)$	0.106	0.298	0.212
$\text{Bias}(\hat{\Delta}_{\text{IPW}_2}^*)$	0.009	0.037	-0.058
$\text{Bias}(\hat{\Delta}_{\text{AIPW}}^*)$	0.021	0.025	0.037
$\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)$	0.029	0.197	0.130
$\text{Bias}_1(\hat{\Delta}_{\text{IPW}_2}^*)$	-0.028	0.023	-0.089
$\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)$	0.011	0.013	0.029
$E[\frac{e^{(X)}}{e^{*(X)}}]$	1.005	1.016	1.019
$\text{cov}[\frac{e^{(X)}}{e^{*(X)}}, \mu_1(X)]$	-0.028	0.022	-0.094
$\text{cov}[\frac{e^{(X)}}{e^{*(X)}}, \mu_1^*(X)]$	-0.038	0.009	-0.121
$E[\frac{e^{(X)}}{e^{*(X)}} - 1]\mu_1$	0.057	0.175	0.225
$E[(\frac{e^{(X)}}{e^{*(X)}} - 1)\mu_1(X)]$	0.029	0.197	0.132
$E[(\frac{e^{(X)}}{e^{*(X)}} - 1)\mu_1^*(X)]$	0.018	0.184	0.104
$\text{Bias}_2(\hat{\Delta}_{\text{IPW}_1}^*)$	0.078	0.101	0.082
$\text{Bias}_2(\hat{\Delta}_{\text{IPW}_2}^*)$	0.038	0.014	0.031
$\text{Bias}_2(\hat{\Delta}_{\text{AIPW}}^*)$	0.011	0.012	0.008
$E[\frac{1-e^{(X)}}{1-e^{*(X)}}]$	0.996	0.990	0.994
$\text{cov}[\frac{1-e^{(X)}}{1-e^{*(X)}}, \mu_0(X)]$	-0.035	-0.015	-0.034
$\text{cov}[\frac{1-e^{(X)}}{1-e^{*(X)}}, \mu_0^*(X)]$	-0.025	-0.004	-0.024
$E[\frac{1-e^{(X)}}{1-e^{*(X)}} - 1]\mu_0$	-0.037	-0.088	-0.054
$E[(\frac{1-e^{(X)}}{1-e^{*(X)}} - 1)\mu_0(X)]$	-0.072	-0.103	-0.087
$E[(\frac{1-e^{(X)}}{1-e^{*(X)}} - 1)\mu_0^*(X)]$	-0.062	-0.091	-0.078

6 | DATA EXAMPLE

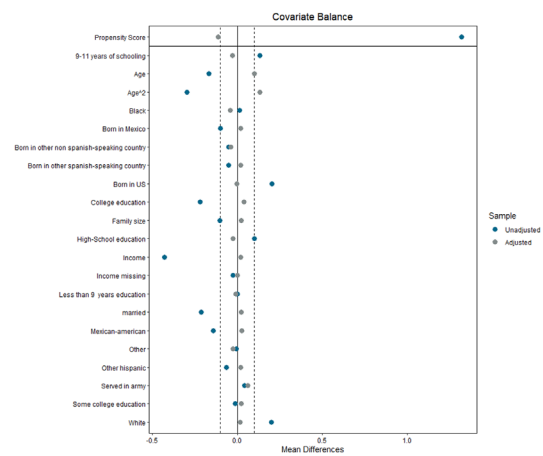
As a motivating example, we analyze data from the National Health and Nutrition Examination Survey (NHANES, 2007–2008) for the purpose of estimating the effect of smoking on blood lead levels. Earlier studies have suggested that increased blood lead levels are associated with chronic kidney disease and peripheral arterial diseases (Muntner et al., 2005). Higher blood lead levels are also associated with mortality in the general U.S. population (Menke et al. 2006). The NHANES dataset studied here is a subset of the data previously analyzed by Hsu and Small (2013) evaluating the relationship between smoking and blood lead levels for the treated population (ATT) with a matching approach. To improve overlap for the estimation of the average causal effect, Δ , we select the study population of males, $N = 1392$. The covariate set is also expanded with four more covariates from the original NHANES demographic data. The treated individuals are defined as daily smokers ($N_1 = 386$) and the controls are individuals who had smoked fewer than 100 cigarettes during their life and no cigarettes in the last 30 days ($N_0 = 1006$). The outcome of interest is blood lead levels (micrograms per deciliter, $\mu\text{g}/\text{dL}$). We control for the covariates, age, army service, marriage, birth country, education, family size, and income-to-poverty level, see Table 8. For this dataset, we apply the three estimators using a logistic propensity score model and for the AIPW estimator, we additionally use a linear OR model with the same covariates. The overlap is displayed in the mirror histogram of Figure 6a together with balance diagnostics in Figure 6b. Here, we see that the balance reduction achieved from the weighing seems satisfactory for most of the covariates, having standardized mean differences within a balance threshold of 0.10. However, for the age squared, army service, college education group, and the group born in Spanish-speaking countries other than Mexico, they have standardized mean differences just exceeding this threshold.

TABLE 6 Asymptotic approximations from Designs A–C in Simulation 2

Parameter	Design		
	A	B	C
μ_1	3.529	3.530	3.535
μ_1^*	3.460	3.953	3.472
μ_0	1.337	1.331	1.335
μ_0^*	1.166	1.398	1.171
$\text{Bias}(\hat{\Delta}_{\text{IPW}_1}^*)$	0.642	0.422	0.622
$\text{Bias}(\hat{\Delta}_{\text{IPW}_2}^*)$	0.216	0.060	0.264
$\text{Bias}(\hat{\Delta}_{\text{AIPW}}^*)$	0.397	0.203	0.631
$\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)$	0.453	0.325	0.451
$\text{Bias}_1(\hat{\Delta}_{\text{IPW}_2}^*)$	0.091	0.024	0.150
$\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)$	0.302	0.147	0.532
$E[\frac{e^{(X)}}{e^{*(X)}}]$	1.102	1.081	1.080
$\text{cov}[\frac{e^{(X)}}{e^{*(X)}}, \mu_1(X)]$	0.102	0.033	0.166
$\text{cov}[\frac{e^{(X)}}{e^{*(X)}}, \mu_1^*(X)]$	-0.202	-0.151	-0.359
$E[\frac{e^{(X)}}{e^{*(X)}} - 1]\mu_1$	0.359	0.285	0.283
$E[(\frac{e^{(X)}}{e^{*(X)}} - 1)\mu_1(X)]$	0.462	0.317	0.449
$E[(\frac{e^{(X)}}{e^{*(X)}} - 1)\mu_1^*(X)]$	0.151	0.168	-0.082
$\text{Bias}_2(\hat{\Delta}_{\text{IPW}_1}^*)$	0.198	0.097	0.171
$\text{Bias}_2(\hat{\Delta}_{\text{IPW}_2}^*)$	0.133	0.036	0.113
$\text{Bias}_2(\hat{\Delta}_{\text{AIPW}}^*)$	0.103	0.056	0.100
$E[\frac{1-e^{(X)}}{1-e^{*(X)}}]$	0.947	0.953	0.952
$\text{cov}[\frac{1-e^{(X)}}{1-e^{*(X)}}, \mu_0(X)]$	-0.124	-0.036	-0.116
$\text{cov}[\frac{1-e^{(X)}}{1-e^{*(X)}}, \mu_0^*(X)]$	-0.031	0.024	-0.017
$E[\frac{1-e^{(X)}}{1-e^{*(X)}} - 1]\mu_0$	-0.071	-0.062	-0.064
$E[(\frac{1-e^{(X)}}{1-e^{*(X)}} - 1)\mu_0(X)]$	-0.195	-0.099	-0.180
$E[(\frac{1-e^{(X)}}{1-e^{*(X)}} - 1)\mu_0^*(X)]$	-0.093	-0.042	-0.073



(a) Overlap plot for estimated propensity score, logistic regression



(b) Covariate balance, SMD in percentage

FIGURE 6 NHANES; data example

TABLE 7 Asymptotic approximations from Designs A–C in Simulation 3

Parameter	Design		
	A	B	C
μ_1	9.903	9.899	9.899
μ_1^*	9.902	9.899	9.898
μ_0	8.897	8.910	8.900
μ_0^*	8.895	8.909	8.904
Bias($\hat{\Delta}_{IPW_1}^*$)	-0.087	-1.349	-12.072
Bias($\hat{\Delta}_{IPW_2}^*$)	0.005	0.046	0.151
Bias($\hat{\Delta}_{AIPW}^*$)	-0.000	0.009	0.035
Bias ₁ ($\hat{\Delta}_{IPW_1}^*$)	-0.009	-0.022	-0.006
Bias ₁ ($\hat{\Delta}_{IPW_2}^*$)	-0.001	-0.001	0.003
Bias ₁ ($\hat{\Delta}_{AIPW}^*$)	-0.000	-0.000	0.003
$E[\frac{e^{(X)}}{e^{*(X)}}]$	0.999	0.998	0.997
$cov[\frac{e^{(X)}}{e^{*(X)}}, \mu_1(X)]$	-0.000	-0.001	-0.001
$cov[\frac{e^{(X)}}{e^{*(X)}}, \mu_1^*(X)]$	-0.000	-0.001	-0.001
$E[\frac{e^{(X)}}{e^{*(X)}} - 1]\mu_1$	-0.008	-0.021	-0.030
$E[(\frac{e^{(X)}}{e^{*(X)}} - 1)\mu_1(X)]$	-0.009	-0.022	-0.031
$E[(\frac{e^{(X)}}{e^{*(X)}} - 1)\mu_1^*(X)]$	-0.009	-0.022	-0.031
Bias ₂ ($\hat{\Delta}_{IPW_1}^*$)	-0.077	-1.328	-12.040
Bias ₂ ($\hat{\Delta}_{IPW_2}^*$)	0.006	0.047	0.153
Bias ₂ ($\hat{\Delta}_{AIPW}^*$)	0.000	0.009	0.036
$E[\frac{1-e^{(X)}}{1-e^{*(X)}}]$	1.009	1.160	2.427
$cov[\frac{1-e^{(X)}}{1-e^{*(X)}}, \mu_0(X)]$	-0.005	-0.050	-0.376
$cov[\frac{1-e^{(X)}}{1-e^{*(X)}}, \mu_0^*(X)]$	-0.005	-0.046	-0.344
$E[\frac{1-e^{(X)}}{1-e^{*(X)}} - 1]\mu_0$	0.083	1.426	12.714
$E[(\frac{1-e^{(X)}}{1-e^{*(X)}} - 1)\mu_0(X)]$	0.077	1.376	12.338
$E[(\frac{1-e^{(X)}}{1-e^{*(X)}} - 1)\mu_0^*(X)]$	0.077	1.379	12.366

Applying the simple IPW estimator, $\hat{\Delta}_{IPW_1}^*$, smoking increases the blood lead levels with 1.10 $\mu\text{g/dL}$ (95% CI: 0.63–1.56), whereas the normalized version, $\hat{\Delta}_{IPW_2}^*$ results in a smaller estimated effect of 0.88 $\mu\text{g/dL}$ and a smaller standard error (95% CI: 0.59–1.18). The AIPW estimator, $\hat{\Delta}_{AIPW}^*$, further reduces the effect estimate to 0.86 $\mu\text{g/dL}$ and the standard error (95% CI: 0.57–1.14, see Table 9). Although applying the estimators to the data does not give information on possible bias due to model misspecification, our results from Section 4 provide guidance to rely on the estimates from $\hat{\Delta}_{IPW_2}^*$ or $\hat{\Delta}_{AIPW}^*$, that is, 0.88 or 0.86 $\mu\text{g/dL}$ with corresponding confidence intervals, rather than the higher value from $\hat{\Delta}_{IPW_1}^*$, of 1.10 $\mu\text{g/dL}$.

7 | DISCUSSION

In this paper, we investigate biases of two IPW estimators and an AIPW estimator under model misspecification. For this purpose, we use a generic probability limit, under misspecification of the PS and OR models, which exists under general conditions. Since the PS enters the estimator in different ways for the IPW estimators under study, the consequences of the model misspecification are not the same. The bias of the IPW estimators depends on the covariance between the PS-model error and the conditional outcome in different ways and the resulting bias can be in opposite directions. For the IPW estimators, normalization has the potential of reducing the bias because it scales the estimator in a mitigating manner. Comparing the bias of the AIPW estimator with a simple IPW estimator, the necessary condition for the AIPW estimator to have a smaller bias is that the expectation of the outcome model under misspecification is less than twice the true conditional outcome, where the expectations include a scaling with the PS-model error. For comparison with the normalized

TABLE 8 Summary statistics for covariates in the NHANES data. Means and sd for lead, age, income-to-poverty level and family size, proportions for education, missing income, race, army service, marriage indicator, and birth country

Variables	Nonsmokers $N_0 = 1006$	Smokers $N_1 = 386$	Overall $N = 1392$
Blood lead level ($\mu\text{g/dL}$)	1.91 (1.70)	2.77 (2.37)	2.15 (1.95)
Age	48.7 (17.5)	46.0 (14.7)	48.0 (16.0)
<i>Education</i>			
Less than 9th grade	131 (13.0%)	51 (13.2%)	182 (13.1%)
9–11th grade	126 (12.5%)	99 (25.6%)	225 (16.2%)
High school graduate	243 (24.2%)	132 (34.2%)	375 (26.9%)
Some college	238 (23.7%)	86 (22.3%)	324 (23.3%)
College	268 (26.6%)	18 (14.7%)	286 (20.5%)
<i>Income</i>			
Income-to-poverty level	2.79 (1.55)	2.14 (1.48)	2.61 (1.56)
Missing	86 (8.5%)	23 (6.0%)	109 (7.8%)
Not missing	920 (91.5%)	363 (94.0%)	1283 (92.2%)
<i>Race</i>			
White	428 (42.5%)	241 (62.4%)	669 (48.1%)
Black	190 (18.9%)	78 (20.2%)	268 (19.3%)
Mexican American	206 (20.5%)	24 (6.2%)	230 (16.5%)
Other Hispanic	120 (11.9%)	22 (5.7%)	142 (10.2%)
Other races	62 (6.2%)	21 (5.4%)	83 (6.0%)
<i>Served in Army</i>			
Yes	192 (19.1%)	90 (23.3%)	282 (20.2%)
No	814 (80.9%)	296 (76.7%)	1100 (79.7%)
<i>Married</i>			
Yes	640 (46.0%)	164 (11.8%)	804 (57.8%)
No	366 (26.3%)	222 (15.9%)	588 (42.2%)
<i>Birth country</i>			
Born in 50 U.S. states or Washington, DC	685 (68.0%)	341 (88.3%)	1026 (73.7%)
Born in Mexico	126 (12.5%)	10 (2.6%)	136 (9.8%)
Born in other Spanish speaking country	97 (9.6%)	17 (4.4%)	116 (8.3%)
Born in other Non-Spanish speaking country	98 (9.7%)	18 (4.7%)	114 (8.2%)
Family size	3.04 (1.67)	2.87 (1.76)	2.99 (1.70)

TABLE 9 Results from (A)IPW estimators, effect of smoking on blood lead levels

Estimator	$\hat{\Delta}$	s.e.	95% CI
$\hat{\Delta}_{IPW_1}^*$	1.10	0.24	(0.63 – 1.56)
$\hat{\Delta}_{IPW_2}^*$	0.88	0.15	(0.59 – 1.18)
$\hat{\Delta}_{AIPW}^*$	0.86	0.14	(0.57 – 1.14)

IPW estimator, the (PS-error scaled) misspecified outcome involves an interval defined by the true conditional outcome adding and subtracting the absolute value of the covariance between the PS-model error and the conditional outcome.

The biases and conditions are exemplified in three simulation studies where the fitted misspecified models fails in specifying nonlinearities, functional form (through misspecified link functions), and covariates. The simulation studies are also accompanied by numerical approximations of the large-sample biases. The third simulation study specifically compares the impact of good, moderate, and poor overlap on the bias due to model misspecification. Here, we see that it is not only the variance that gets inflated from PS values close to 0 or 1, but the bias due to model misspecification also increases rapidly. The normalized IPW and AIPW estimators show a more stable performance. The bias expressions of

the IPW and AIPW estimators suggest that the AIPW estimator has a smaller bias than the IPW estimators even under moderate misspecification of the outcome model. For the AIPW estimator, poor overlap and large differences between $e(X)$ and $e^*(X)$ are compensated for by outcome model assumptions in the area where data are sparse. However, in the simulations, the normalized IPW estimator also performs well due to the implicit stabilization from the PS-model errors. Since all biases include the PS-model error, we suggest that a researcher should be careful when modeling the PS even though an OR model is additionally involved.


ACKNOWLEDGMENT

The authors acknowledge Professor Elena Stanghellini for valuable input. Financial support was received from the Royal Swedish Academy of Sciences and the Swedish Research Council (grant number 2016-00703).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://github.com/IngWae/Bias_AIPW

OPEN RESEARCH BADGES

 This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the [Supporting Information](#) section.

This article has earned an open data badge “**Reproducible Research**” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Ingeborg Waernbaum  <https://orcid.org/0000-0002-4457-5311>

Laura Pazzagli  <https://orcid.org/0000-0002-1908-6073>

REFERENCES

- Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4), 962–973.
- Boos, D. D., & Stefanski, L. (2013). M-estimation (estimating equations). In D. D. Boos & L. A. Stefanski (Eds.), *Essential statistical inference* (Vol. 120, pp. 297–337). Springer.
- Busso, M., DiNardo, J., & McCrary, J. (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics*, 96(5), 885–897.
- Cao, W., Tsiatis, A. A., & Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96(3), 723–734.
- Chang, S.-H., Chou, I.-J., Yeh, Y.-H., Chiou, M.-J., Wen, M.-S., Kuo, C.-T., See, L.-C., & Kuo, C.-F. (2017). Association between use of non-vitamin k oral anticoagulants with and without concurrent medications and risk of major bleeding in nonvalvular atrial fibrillation. *JAMA*, 318(13), 1250–1259.
- Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, 173(7), 761–767.
- Han, P., & Wang, L. (2013). Estimation with missing data: Beyond double robustness. *Biometrika*, 100(2), 417–430.
- Hernán, M. A., Brumback, B., & Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology*, 11(5), 561–570.
- Hernán, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health*, 60(7), 578–586.
- Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology*, 2(3), 259–278.
- Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4), 1161–1189.
- Hsu, J. Y., & Small, D. S. (2013). Calibrating sensitivity analyses to observed covariates in observational studies. *Biometrics*, 69(4), 803–811.
- il Kim, K. (2019). Efficiency of average treatment effect estimation when the true propensity is parametric. *Econometrics*, 7(2), 25.
- Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4), 523–539.
- Khan, S., & Tamer, E. (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica*, 78(6), 2021–2042.
- Kwon, T., Jeong, I. G., Lee, J., Lee, C., You, D., Hong, B., Hong, J. H., Ahn, H., & Kim, C.-S. (2015). Adjuvant chemotherapy after radical cystectomy for bladder cancer: A comparative study using inverse-probability-of-treatment weighting. *Journal of Cancer Research and Clinical Oncology*, 141(1), 169–176.

- Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, *113*(521), 390–400.
- Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, *23*(19), 2937–2960.
- Ma, X., & Wang, J. (2020). Robust inference using inverse probability weighting. *Journal of the American Statistical Association*, *115*(532), 1851–1860.
- Menke, A., Muntner, P., Batuman, V., Silbergeld, E. K., & Guallar, E. (2006). Blood lead below 0.48 $\mu\text{mol/l}$ (10 $\mu\text{g/dl}$) and mortality among us adults. *Circulation*, *114*(13), 1388–1394.
- Muntner, P., Menke, A., DeSalvo, K. B., Rabito, F. A., & Batuman, V. (2005). Continued decline in blood lead levels among adults in the united states: The National Health and Nutrition Examination Surveys. *Archives of Internal Medicine*, *165*(18), 2155–2161.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Robins, J., Hernán, M., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology (Cambridge, Mass.)*, *11*(5), 550–560.
- Robins, J. M., & Rotnitzky, A. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*(427), 846–866.
- Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, *90*(429), 106–121.
- Robins, J., Sued, M., Lei-Gomez, Q., & Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when “inverse probability” weights are highly variable. *Statistical Science*, *22*(4), 544–559.
- Rubin, D. (1980). Discussion of “randomization analysis of experimental data in the fisher randomization test” by D. Basu. *Journal of the American Statistical Association*, *75*, 591–593.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.
- Scharfstein, D. O., Rotnitzky, A., & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, *94*(448), 1096–1120.
- Seaman, S. R., & Vansteelandt, S. (2018). Introduction to double robust methods for incomplete data. *Statistical Science*, *33*(2), 184–197.
- Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*(3), 278–295.
- Tan, Z. (2007). Comment: Understanding OR, PS and DR. *Statistical Science*, *22*(4), 560–568.
- Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika*, *97*(3), 661–682.
- Tsiatis, A. A. (2007). *Semiparametric theory and missing data*. Springer Science & Business Media.
- Tsiatis, A. A., & Davidian, M. (2007). Comment: Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, *22*(4), 569–573.
- Vansteelandt, S., Carpenter, J., & Kenward, M. G. (2010). Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, *6*(1), 37–48.
- Vermeulen, K., & Vansteelandt, S. (2015). Bias-reduced doubly robust estimation. *Journal of the American Statistical Association*, *110*(511), 1024–1036.
- Waernbaum, I. (2012). Model misspecification and robustness in causal inference: Comparing matching with doubly robust estimation. *Statistics in Medicine*, *31*(15), 1572–1581.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, *50*(1), 1–25.
- Williamson, E., Forbes, A., & White, I. (2014). Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Statistics in Medicine*, *33*(1), 721–737.
- Wooldridge, J. (2010). *Econometric analysis of cross section and panel data*. Econometric Analysis of Cross Section and Panel Data. MIT Press.
- Yao, L., Sun, Z., & Wang, Q. (2010). Estimation of average treatment effects based on parametric propensity score model. *Journal of Statistical Planning and Inference*, *140*(3), 806–816.
- Zhou, Y., Matsouaka, R. A., & Thomas, L. (2020). Propensity score weighting under limited overlap and model misspecification. *Statistical Methods in Medical Research*, *29*(12), 3721–3756.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Waernbaum, I., & Pazzagli, L. (2023). Model misspecification and bias for inverse probability weighting estimators of average causal effects. *Biometrical Journal*, *65*, 2100118.

<https://doi.org/10.1002/bimj.202100118>

APPENDIX A

A.1 | Regularity conditions

The convergence in probability of $\hat{\Delta}_{IPW_1}^*$, $\hat{\Delta}_{IPW_2}^*$, and $\hat{\Delta}_{AIPW}^*$ to their corresponding expectations would follow directly from a weak law of large numbers (WLLN) for an iid sample of (T_i, X_i, Y_i) except for the estimated parameters $\hat{\beta}^*$ in $\hat{e}^*(X_i)$ and $\hat{\alpha}_t^*$ in $\mu_t^*(X_i)$, $t = 0, 1$. To justify the biases in Section 4, consider a general representation of a function $g[T, Y, X, \hat{\theta}]$ where $\hat{\theta} \xrightarrow{p} \theta_0$ and

$$\frac{1}{n} \sum g[T_i, X_i, Y_i, \hat{\theta}] \xrightarrow{p} E[g(T, X, Y, \theta_0)]. \quad (A1)$$

The $\hat{\theta}$ in (A1) corresponds to $\hat{\beta}^*$ for $\hat{\Delta}_{IPW_1}^*$, $\hat{\Delta}_{IPW_2}^*$, and $(\hat{\alpha}^*, \hat{\beta}^*)$ for $\hat{\Delta}_{AIPW}^*$ and under Assumptions 5 and 7, the consistency of $\hat{\theta}$ is ensured. Regularity conditions for the function g can be given, see, for example, Boos and Stefanski (2013, Theorem 7.3) who show that (A1) holds for differentiable functions with bounded derivatives (w.r.t. θ). The regularity conditions for $g[X_i, T_i, Y_i, \hat{\theta}]$, for the three estimators, imply conditions on the models $e^*(X, \beta^*)$ and $\mu_t^*(X, \alpha_t^*)$ such that the regularity condition for g is satisfied. Under (A1), we can insert the limiting values β^* and α^* and their corresponding $e^*(X)$ and $\mu_t^*(X)$, $t = 0, 1$ when taking a WLLN.

A.2 | Biases

Under the regularity conditions and Assumptions 5–6 (IPW) and 5–7 (AIPW), we derive the bias in Equation (4). For $\hat{\Delta}_{IPW_1}^*$:

$$\frac{1}{n} \sum_{i=1}^n \frac{T_i Y_i}{\hat{e}^*(X_i)} - \frac{1}{n} \sum_{i=1}^n \frac{(1 - T_i) Y_i}{1 - \hat{e}^*(X_i)} \xrightarrow{p} E \left[\frac{TY}{e^*(X)} \right] - E \left[\frac{(1 - T)Y}{(1 - e^*(X))} \right],$$

and

$$\begin{aligned} E \left[\frac{TY}{e^*(X)} \right] - E \left[\frac{(1 - T)Y}{(1 - e^*(X))} \right] &= E \left[E \left[\frac{TY}{e^*(X)} \middle| X \right] \right] - E \left[E \left[\frac{(1 - T)Y}{[1 - e^*(X)]} \middle| X \right] \right] \\ &= E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - E \left[\frac{(1 - e(X))}{(1 - e^*(X))} \mu_0(X) \right], \end{aligned}$$

subtracting with Δ gives

$$\text{Bias}(\hat{\Delta}_{IPW_1}^*) = E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - E \left[\frac{1 - e(X)}{1 - e^*(X)} \mu_0(X) \right] - (\mu_1 - \mu_0).$$

The biases of Equations (5) and (6) are derived similarly.

A.3 | Comparisons

To study the consequences of model misspecification for the estimators, we compare each difference involving $\mu_1(X)$ and $\mu_0(X)$ separately. For example, we study the bias of $\hat{\Delta}_{IPW_1}^*$

$$\begin{aligned} \hat{\Delta}_{IPW_1}^* - \Delta &\xrightarrow{p} E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - E \left[\frac{1 - e(X)}{1 - e^*(X)} \mu_0(X) \right] - (\mu_1 - \mu_0) \\ &= E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1 + \mu_0 - E \left[\frac{1 - e(X)}{1 - e^*(X)} \mu_0(X) \right]. \end{aligned}$$

Since the errors $e(X)/e^*(X)$ and $(1 - e(X))/(1 - e^*(X))$ are inversely related, we would normally not expect that the biases from the two parts $E[\frac{e(X)}{e^*(X)} \mu_1(X)] - \mu_1$ and $\mu_0 - E[\frac{1 - e(X)}{1 - e^*(X)} \mu_0(X)]$ cancel each other out. For example, if $E[\frac{e(X)}{e^*(X)} \mu_1(X)] - \mu_1 > 0$, then we might expect that $\mu_0 - E[\frac{1 - e(X)}{1 - e^*(X)} \mu_0(X)] > 0$ and similarly for negative values. However, if there is large effect modification on the difference scale, the two components might be in different directions.

Inequalities concerning the biases are made with respect to the absolute values for two parts separately, for example, for $\text{Bias}(\hat{\Delta}_{\text{IPW}_1}^*)$, we investigate

$$|\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)| = \left| E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1 \right|. \quad (\text{A2})$$

Since $|-a| = |a|$, the second part is

$$|\text{Bias}_2(\hat{\Delta}_{\text{IPW}_1}^*)| = \left| E \left[\frac{1 - e(X)}{1 - e^*(X)} \mu_0(X) \right] - \mu_0 \right|, \quad (\text{A3})$$

and similarly for (8). The conditions derived for the first part of the biases, (7) and (8), can be directly applied to the second part of the biases replacing $e(X)/e^*(X)$, with $(1 - e(X))/(1 - e^*(X))$ and $\mu_1(X)$ with $\mu_0(X)$. In a similar manner, we have for $\text{Bias}_2(\hat{\Delta}_{\text{AIPW}}^*)$

$$\left| E \left[\frac{(1 - e(X)) - (1 - e^*(X))(\mu_0(X) - \mu_0^*(X))}{(1 - e^*(X))} \right] \right| = \left| E \left[\frac{(e(X) - e^*(X))(\mu_0(X) - \mu_0^*(X))}{(1 - e^*(X))} \right] \right|,$$

and the conditions derived for the first part of the bias, (9), can be directly applied to (A3) additionally replacing $\mu_1^*(X)$ with $\mu_0^*(X)$.

Necessary and sufficient condition for $|\text{Bias}_1(\hat{\Delta}_{\text{IPW}_2}^)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)|$ in (10)*

$$\left| \frac{E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right]}{E \left[\frac{e(X)}{e^*(X)} \right]} - \mu_1 \right| < E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1.$$

1. Assuming that $E \left[\frac{e(X)}{e^*(X)} \mu_1(X) - \mu_1 \right] > 0$:

We have

$$-E \left[\frac{e(X)}{e^*(X)} \right] \left(E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1 \right) < \text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right] < E \left[\frac{e(X)}{e^*(X)} \right] \left(E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1 \right). \quad (\text{A4})$$

2. Assuming that $E \left[\frac{e(X)}{e^*(X)} \mu_1(X) - \mu_1 \right] < 0$:

$$\left| \frac{E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right]}{E \left[\frac{e(X)}{e^*(X)} \right]} - \mu_1 \right| < -E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1.$$

Here, we have

$$E \left[\frac{e(X)}{e^*(X)} \right] \left(E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1 \right) < \text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right] < -E \left[\frac{e(X)}{e^*(X)} \right] \left(E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1 \right), \quad (\text{A5})$$

by A4 and A5, we have

$$|\text{cov} \left[\frac{e(X)}{e^*(X)}, \mu_1(X) \right]| < |E \left[\frac{e(X)}{e^*(X)} \right] (E \left[\frac{e(X)}{e^*(X)} \mu_1(X) \right] - \mu_1)|.$$

Necessary condition for $|\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)|$ in (12)

1. Assuming that $E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] > 0$:

$$\left| E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)(\mu_1(X) - \mu_1^*(X))\right] \right| < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right].$$

Here, we have

$$-E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)(\mu_1(X) - \mu_1^*(X))\right] < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right],$$

and

$$0 < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1^*(X)\right] < 2 \cdot E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right]. \quad (\text{A6})$$

2. Assuming that $E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] < 0$:

$$\left| E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)(\mu_1(X) - \mu_1^*(X))\right] \right| < -E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right].$$

Here, we have

$$\begin{aligned} E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] &< E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)(\mu_1(X) - \mu_1^*(X))\right] < -E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right], \\ 2 \cdot E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] &< E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1^*(X)\right] < 0, \end{aligned} \quad (\text{A7})$$

and by (A6) and (A7),

$$\left| E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1^*(X)\right] \right| \leq 2 \cdot \left| E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] \right|.$$

Sufficient condition for $|\text{Bias}_1(\hat{\Delta}_{\text{AIPW}}^*)| < |\text{Bias}_1(\hat{\Delta}_{\text{IPW}_1}^*)|$ in (13)

The sufficient condition follows from the above by adding to the necessary condition that

$$E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1^*(X)\right] \text{ and } E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right]$$

are either both positive or both negative, since then either (A6) or (A7) holds.

Necessary condition of Equation 14

1. Assuming that:

$$\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right] > 0 \iff \frac{E\left[\frac{e(X)}{e^*(X)}\mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]} - \mu_1 > 0.$$

Here, we have

$$\begin{aligned} -\frac{E\left[\frac{e(X)}{e^*(X)}\mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]} + \mu_1 &< E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)(\mu_1(X) - \mu_1^*(X))\right] < \frac{E\left[\frac{e(X)}{e^*(X)}\mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]} - \mu_1, \\ -\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right] &< E\left[\frac{e(X)}{e^*(X)}\right]E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)(\mu_1(X) - \mu_1^*(X))\right] < \text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right], \end{aligned}$$

and

$$E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] - \frac{\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]} < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1^*(X)\right] < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \frac{\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]}. \quad (\text{A8})$$

2. Assuming that:

$$\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right] < 0 \iff \frac{E\left[\frac{e(X)}{e^*(X)}\mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]} - \mu_1 < 0,$$

it follows that

$$\begin{aligned} \frac{E\left[\frac{e(X)}{e^*(X)}\mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]} - \mu_1 &< E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)(\mu_1(X) - \mu_1^*(X))\right] < -\frac{E\left[\frac{e(X)}{e^*(X)}\mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]} + \mu_1 \\ \text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right] &< E\left[\frac{e(X)}{e^*(X)}\right]E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)(\mu_1(X) - \mu_1^*(X))\right] < -\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right], \end{aligned}$$

and

$$E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \frac{\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]} < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1^*(X)\right] < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] - \frac{\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]}{E\left[\frac{e(X)}{e^*(X)}\right]}, \quad (\text{A9})$$

by (A8) and (A9), we have that

$$E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] - \frac{\left|\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right|}{E\left[\frac{e(X)}{e^*(X)}\right]} < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1^*(X)\right] < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \frac{\left|\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right|}{E\left[\frac{e(X)}{e^*(X)}\right]},$$

Sufficient condition of (15)

If $E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right]$ and $\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]$ are both positive then,

$$E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \left|\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right| = \left|E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right|$$

and

$$-\left|E\left\{\left[\frac{e(X)}{e^*(X)} - 1\right]\mu_1(X)\right\} + \text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right| < E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \left|\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right|.$$

If $E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right]$ and $\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]$ are both negative, then we have

$$E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \left|\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right| < \left|E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right|,$$

and

$$-\left|E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right| = E\left[\left(\frac{e(X)}{e^*(X)} - 1\right)\mu_1(X)\right] + \left|\text{cov}\left[\frac{e(X)}{e^*(X)}, \mu_1(X)\right]\right|,$$

and the necessary condition from Equation (14) follows.