Research Article

Zhaohan Sun*, Yeying Zhu, and Joel A. Dubin

Estimation of network treatment effects with non-ignorable missing confounders

https://doi.org/10.1515/jci-2022-0079
received December 01, 2022; accepted January 31, 2024

Abstract: In causal inference, interference takes place when the intervention on one unit affects the outcome of other units. Most of the previous methods for estimating network causal effects assume that the covariate information is complete, which may lead to biased estimates when missingness exists. In this study, we consider the *partial and direct interference* setting. Specifically, the whole population can be divided into different clusters. Within each cluster, the outcome of each unit is dependent on the intervention received by other units, but not dependent on the confounders or outcomes of other units within the same cluster or of those in different clusters. We also assume that the confounders are subject to non-ignorable missingness, and a confounder is considered as missing if any component of it is missing. We propose three consistent estimators for the direct, indirect, total, and overall effect of the intervention on the outcome, and derive the asymptotic results accordingly. A comprehensive study is carried out as well to investigate the finite sample properties of the proposed estimators. We illustrate the proposed methods by analyzing the dataset collected from an acid rain program, which was launched to reduce air pollution in the United States by encouraging the scrubber's installation on power plants, where the records of some operating characteristics of the power generating facilities are subject to missingness.

Keywords: causal inference, missing data, partial interference, air pollution

MSC 2020: 62P12

1 Introduction

In the causal inference literature, the problem of inferring the treatment effect when data are subject to missingness has drawn a great amount of attention in recent years. According to Rubin [1], there are two types of missingness: ignorable and non-ignorable missingness. Ignorable missingness refers to missingness that is independent of the missing values, and non-ignorable missingness refers to the missingness that is dependent on the missing values. The inference for non-ignorable missingness is more challenging than ignorable missingness because the full data distribution is not fully identifiable without any assumptions, sometimes very restrictive. Following Yang et al. [2], we consider the *outcome independent missingness assumption*, which is plausible when the covariates are collected at the beginning of the study, and the outcome is collected long after the covariates are measured.

In most of the aforementioned work in handling missing data in the literature, methods rely on the stable unit treatment value assumption (SUTVA) [3]. SUTVA states that (i) the potential outcome of each unit is unaffected by the treatment assignment of any other unit, and (ii) there are no different versions of each

*** Corresponding author: Zhaohan Sun,** Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada, e-mail: z227sun@uwaterloo.ca
Yeying Zhu: Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada, e-mail: yeying.zhu@uwaterloo.ca
Joel A. Dubin: Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada, e-mail: jdubin@uwaterloo.ca

treatment level. However, the first assumption, which is known as the **no interference assumption** [4], can be violated in some scenarios. For example, in 1990, the acid rain program (ARP) was launched to reduce ambient PM2.5 (atmospheric particulate matter (PM) that has a diametre of less than 2.5 μm) by assigning power plants to install scrubber facilities [5]. The monitored reduction of SO_2 emission data at the location of one power plant not only depends on its own scrubber installation but may also be affected by the intervention on power plants upwind. Another example comes from the nationally representative US Population Assessment of Tobacco and Health (PATH) Study [6], where researchers were interested in evaluating the influence of Electronic Nicotine Delivery Systems (ENDS) and pharmaceutical cessation aids on persistent abstinence from cigarette smoking and reduced cigarette consumption, and in this study, it has been revealed that one individual's marital satisfaction and family members smoking status can affect this individual's smoking cessation, that is, the smoking cessation of one individual may be affected by the intervention of other family members. More examples can also be found in biomedical research, health sciences, and social networking studies.

Various identification and estimation methods have been proposed in the scenario when interference exists, but the confounders and outcome are fully observed. Generally, there are two types of interference: *full interference* and *partial interference*. The full interference happens when the outcome of a unit is affected by the intervention on other units that interfere with this unit. The network interference structure can be represented by an adjacency matrix: the entries of which take the value on $\{0, 1\}$ (e.g., if the unit i is affected by the intervention on the individual j, then the entry of the matrix in ith row and jth column equals one; otherwise, the value of the entry equals zero). The partial interference is a special case of the full interference where the adjacency matrix follows block diagonal structure, the entry of the matrix is equal to zero if the unit of the corresponding row and the unit of the corresponding column are not in the same block, that is, the interference may happen between units in the same block but not between units in different blocks. In this article, we consider the latter type of interference and focus on the semiparametric estimation of four network treatment effects: the direct effect, indirect effect, total effect, and overall effect (Tchetgen Tchetgen and VanderWeele [7], Hudgens and Halloran [8], Liu et al. [9], Papadogeorgou et al. [10]). The definitions and illustrations of these treatment effects are presented in Section 2.

There is limited work discussing the estimation of network treatment effect when confounders are subject to non-ignorable missingness. Sun and Liu [11] proposed a doubly robust estimator when data are subject to non-ignorable missingness, where the data are assumed to be independent and the interference was ignored in data application. Unlike Sun and Liu, we study the partial interference setting and propose three pairs of semiparametric estimators: inverse probability weighting (IPW), regression, and doubly robust (DR) estimators for the four types of network treatment effects. The proposed IPW estimator requires correct specification of the propensity score for treatment selection and the missingness mechanism. The regression estimator requires a correctly specified outcome regression model, and the doubly robust estimator is shown to be consistent if either set of the IPW models (propensity score of missingness and propensity score of treatment) or the regression model, but not necessarily both, are correctly specified. The IPW and DR estimators require the product of unit-level propensity scores, where the existence of extreme probabilities may lead to high-variant estimators. To circumvent the problem of varying cluster sizes, we propose self-normalized IPW and DR estimators, which can be viewed as stabilized versions of and have smaller variances than their respective conventional estimators.

The rest of this article is organized as follows. We start by introducing the notation and assumptions in Section 2. In Section 3, we propose three pairs of semiparametric network treatment effect estimators. The performance of the proposed methods is further illustrated via simulation studies in Section 4. We analyze real data from the ARP in Section 5 and leave the discussion of our methodology and potential extensions as future work in Section 6.

2 Notation and assumptions

Following Perez-Heydrich et al. [12] and Tchetgen Tchetgen and VanderWeele [7], we consider that K clusters are randomly sampled from an infinite superpopulation of clusters. Assume the total number of units is N, and each group i has N_i units, where $1 \leq i \leq K$ and $1 \leq N_i \leq N - K + 1$. Let $X_{ij} = (X_{1ij}, X_{2ij}, \dots X_{pij})$ denote p-dimensional confounders of unit j in group i, the values of which may be subject to missingness, and let $X_i = (X_{i1}, X_{i2}, \dots X_{iN_i})$ be the confounders of all units in group i. Let $R_{ij} = 1$ if X_{ij} is complete and $R_{ij} = 0$ if X_{ij} is missing. In this article, we consider the single missingness pattern, that is, $R_{ij} = 0$ if any components of the confounders of unit j in group i are missing. We leave the extension to multiple missingness patterns as a future research direction. Let Y_{ij} and A_{ij} denote the observed outcome and treatment status for unit j in group i, respectively, and $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{iN_i})$ and $A_i = (A_{i1}, A_{i2}, \dots, A_{iN_i})$ denote the vectors of observed outcome and treatment indicators, respectively, for all units in group i. Assume $A_{i(-j)} = (A_{i1}, \dots, A_{ij-1}, A_{ij+1}, \dots A_{iN_i})$ is the vector of treatment status for all units in group i except for individual j. Let a_{ij}, $a_{i(-j)}$, and a_i denote possible values of A_{ij}, $A_{i(-j)}$, and A_i, respectively. Suppose $\mathcal{A}(n)$ is the set of vectors of all possible treatment assignments of length n. Then, there are 2^{N_i} possible treatment assignments in group i, and $a_i \in \mathcal{A}(N_i)$.

We consider the scenario when only X is missing, while the other variables are fully observed. Suppose the treatment is assigned with an α-strategy where every unit in a group is assigned to the treatment with average treatment allocation probability α. In a randomized trial, the probability of being treated is fully determined by the treatment allocation probability. However, in observational studies, whether an individual receives the treatment assignment is not only determined by the randomized treatment allocation strategy but dependent on his/her choice of participation in the study. To avoid the additional modelling of participation (probability of the participation status given confounders), we will model the probability of the treatment indicator conditional on confounders directly. Let $Y_{ij}(a_i)$ and $Y_{ij}(a_{ij}, a_{i-j})$ denote the potential outcome for unit j in group i under treatment allocation a_i, where $a_i = (a_{i1}, \dots a_{i,j-1}, a_{ij}, a_{i,j+1}, \dots a_{i,N_i})$, and let denote $Y_i(a_i)$ as the vector of potential outcomes for all units in group i. Let $\bar{Y}_i(a, \alpha) = N_i^{-1}\sum_{j=1}^{N_i}\sum_{a_{i(-j)} \in \mathcal{A}(N_i - 1)} Y_{ij}(a, a_{i(-j)})\pi(a_{i(-j)}; \alpha)$ denote the average potential outcome for group i, where $\pi(a_{i(-j)}; \alpha) = \Pr(A_{i(-j)} = a_{i(-j)} | A_{ij} = a_{ij}, X_i)$ is the probability of $A_{i(-j)} = a_{i(-j)}$ in group i. Let $\bar{Y}_i(\alpha) = N_i^{-1}\sum_{j=1}^{N_i}\sum_{a_i \in \mathcal{A}(N_i)} Y_{ij}(a_i)\pi(a_i; \alpha)$ denote the marginal average potential outcome for group i.

There are different treatment allocation strategies that can be deployed. For example, in the cholera vaccine study [8], $\pi(a_i; \alpha) = \prod_j \alpha^{a_{ij}}(1 - \alpha)^{1 - a_{ij}}$ and $\pi(a_{i(-j)}; \alpha) = \prod_{j' \neq j} \alpha^{a_{ij'}}(1 - \alpha)^{1 - a_{ij'}}$, that is, the treatment allocation strategy does not depend on individuals' characteristics. Once the individuals choose to participate in the trial, they are classified into different groups. For each group, the individuals are assigned with the vaccine randomly with probability α. On the other hand, in the ARP study, the intervention of scrubber installation is encouraged by federal regulations. However, the adoption of intervention is also affected by the characteristics of power plants such as size and heat input. To account for the influence of confounders on the treatment allocation probability, Papadogeorgou et al. [10] models $P_{\alpha,x}(a_{ij})$ as $\operatorname{logit}\{P_{\alpha,x}(a_{ij} = 1 | X_{ij})\} = \xi_i^{\alpha} + \delta X_{ij}$, where $\xi_i^{(\alpha)}$ satisfying $(N_i^{-1})\sum_{i=1}^{N_i} \operatorname{expit}(\xi_i^{(\alpha)} + \delta X_{ij}) = \alpha$ is a parameter to be estimated, and δ is some pre-fixed value. $P_{\alpha,x}(a_i)$ represents treatment allocation for all units in group i, while $P_{\alpha,x}(a_{i(-j)})$ denotes the treatment assignment in group i excluding unit j. More details can be found in Section 5. We adopt the same modelling strategy in this article.

The average potential outcome and the marginal average potential outcome are defined as $\mu_{a\alpha} = E(\bar{Y}_i(a, \alpha))$ and $\mu_{\alpha} = E(\bar{Y}_i(\alpha))$, respectively. Then, the direct effect (or the population average treatment effect) is defined as $\overline{\mathrm{DE}}(\alpha) = E(\bar{Y}_i(1, \alpha) - \bar{Y}_i(0, \alpha)) = \mu_{1\alpha} - \mu_{0\alpha}$. For the ARP study, the direct effect represents the difference in the amount of PM2.5 emissions when a power generating facility is equipped with scrubbers compared to when scrubbers are not installed. For policies α_0 and α_1, the indirect effect is defined as $\overline{\mathrm{IE}}(\alpha_0, \alpha_1) = E(\bar{Y}_i(0, \alpha_1) - \bar{Y}_i(0, \alpha_0)) = \mu_{0\alpha_1} - \mu_{0\alpha_0}$, where only the units that are not equipped with scrubbers are considered. The total effect is denoted by $\overline{\mathrm{TE}}(\alpha_1, \alpha_1) = E(\bar{Y}_i(1, \alpha_1) - \bar{Y}_i(0, \alpha_0)) = \mu_{1\alpha_1} - \mu_{0\alpha_0}$, which corresponds to the combination of both the direct effect and the indirect effect. The overall effect is defined as $\overline{\mathrm{OE}}(\alpha_0, \alpha_1) = E(\bar{Y}_i(\alpha_1) - \bar{Y}_i(\alpha_0)) = \mu_{\alpha_1} - \mu_{\alpha_0}$, which represents the difference in the amount of PM2.5 emissions for units under one coverage probability of scrubbers' installation compared to units with another level of coverage probability. In Section 3, we focus on estimating the direct effect. The estimation of the other network causal effects can be approached in a similar way, and we present the results of four types of network causal effects in the data application. For the purpose of estimation with incomplete confounders, we assume the type of interference to be direct interference (see more details in Ogburn et al. [13]), where the interference happens only through the effect of the treatment assignment on other units in the same clusters. We leave more sophisticated interference types such as contagion interference and allocation interference as a future research direction. Throughout this article, we also make the following assumptions:

(1) **Causal consistency assumption**: A subject's potential outcome under their observed treatment assignment is equal to the outcome that will actually be observed, that is, $Y_{ij} = \sum_{a_i \in \mathcal{A}(N_i)} 1(A_i = a_i) Y_{ij}(a_i)$.

(2) **Exchangeability assumption**: For each group, the treatment vector A_i is assumed to be conditionally independent with potential outcomes given confounders X_i, that is, $A_i \perp\!\!\!\perp Y_i(a_i)|X_i$.

(3) **Positivity assumption**: $\Pr(A_i|X_i) > 0$ and $\Pr(R_{ij} = 1|A_i, X_i) > 0$ for all A_i and X_i.

(4) **Outcome-independent missingness assumption**: $R_i \perp\!\!\!\perp Y_i|A_i, X_i$.

It has been shown in Ding and Geng [14] that without any assumptions, the joint distribution of (A_i, Y_i, X_i) is not identifiable, and they provided the lower and upper bounds for the causal effects, which may be wide. Yang et al. [2] also considered non-parametric identification and non-parametric estimation of the causal effect with an auxiliary variable Z. Here, we propose the IPW estimator, regression, and DR estimator for the defined network causal effects, and discuss the intuition behind the construction of these estimators.

In this article, we consider the non-ignorable missing confounders. The missingness mechanism is ignorable if the absence of the confounders is only dependent on the observed data, that is, $R_i \perp\!\!\!\perp X_i|A_i, Y_i$, and the confounders are non-ignorable if the missingness depends on the missing values. Under non-ignorable missingness, the full data distribution is not identifiable without additional assumptions. Therefore, for identification purposes, we assume that the missingness mechanism for confounders X_i satisfies the group-level outcome independent missingness assumption, which is modified from the outcome-independent missingness assumption in [2], that is, $R_i \perp\!\!\!\perp Y_i|A_i, X_i$. The associated causal diagram is shown in Figure 1. The assumption is plausible if the confounders are measured long before the outcome data are collected. For example, as mentioned in Yang et al. [2], the potentially exposed children and their neighbourhoods were more carefully measured than those that were not under the risk of exposure in the water crisis study in Flint, Michigan U.S., which implies that the missingness may depend on both the measured confounders and the exposure status. In addition, the health status of the children was tested long after the confounders (e.g., age) were collected. Thus, the missingness of confounders is independent of the outcome conditional on all the other relevant information including observed confounders and exposure status. Miao and Tchetgen Tchetgen [15] proposed using an additional auxiliary variable in the missing data setting and proposed various semiparametric and DR estimators. In our case, the outcome Y serves as the auxiliary variable for the confounder missing indicator R. Our estimators here are counterpart estimators of those.

3 Estimation

3.1 Inverse probability weighting

In this section, we propose an IPW estimator for network treatment effects when confounders are subject to non-ignorable missingness mechanism. The idea of constructing the IPW estimator is to weigh each individual

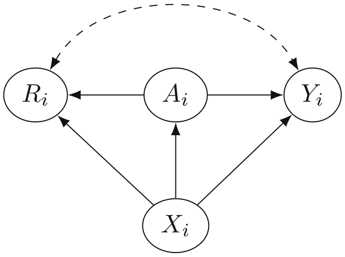

Figure 1: The causal diagram for the group-level outcome independent missingness assumption, where the dashed line represents the conditional independence between Y_i and R_i, $1 \le i \le N_i$.

with the inverse of probability of receiving the treatment, such that the association between confounders and treatment assignment can be removed. However, since X is not fully observed, $\Pr(A|X)$ is not estimable without further adjustment. Besides, even if the data are fully observed, we cannot directly apply the IPW as the data are not independent within the same group. To address this, we utilize the inverse of the group-level joint propensity score of treatment assignment and missingness mechanism as the weight for each subject, that is, $1/\Pr(A_i, R_{ij}|X_i),\ i = 1, 2, \ldots K,\ j = 1, 2, \ldots N_i$, and replace the number of individuals in each group by the sum of inverse joint propensity scores within the group.

To obtain the group-joint propensity score modelling for the treatment and missingness, we assume $P(R_{ij} = 1|A_{ij}, X_{ij})$ is correctly specified as $P(R_{ij} = 1|A_{ij}, X_{ij};\ \gamma)$, and $P(A_{ij} = 1|R_{ij} = 1, X_{ij})$ is correctly specified as $P(A_{ij} = 1|R_{ij} = 1, X_{ij};\ \delta)$. The unknown parameter γ can be estimated using generalized methods of moments, and δ can be estimated via maximum likelihood approach. After obtaining the estimates of $P(R_{ij} = 1|A_{ij}, X_{ij})$ and $P(A_{ij} = 1|R_{ij} = 1, X_{ij})$, the joint density of treatment A and missingness R conditional on confounders X can be parameterized as follows [16]:

$$\Pr(a, r|x) = \frac{\psi(a, a_0, r, r_0|x)\Pr(r|a_0, x)\Pr(a|r_0, x)}{\sum_{r=0}^{1}\sum_{a=0}^{1}\psi(a, a_0, r, r_0|x)\Pr(r|a_0, x)\Pr(a|r_0, x)}, \tag{1}$$

where $r_0 = 1$, $a_0 = 1$, and

$$\psi(a, a_0, r, r_0|x) = \frac{\Pr(r|a, x)\Pr(r_0|a_0, x)}{\Pr(r|a_0, x)\Pr(r_0|a, x)}.$$

For a complete proof of equation (1), please refer to Web Appendix A of Chen [16]. We then construct the IPW estimator for the population average potential outcome with estimated parameters as follows:

$$\hat{\mu}_{a\alpha}^{\text{ipw}} = \frac{1}{K}\sum_{i=1}^{K}\hat{Y}_i^{\text{IPW}}(a, \alpha) = \frac{1}{K}\sum_{i=1}^{K}\frac{\sum_{j=1}^{N_i}\hat{w}_{ij}^{a\alpha}Y_{ij}}{\sum_{j=1}^{N_i}\hat{w}_{ij}^{a\alpha}}; \tag{2}$$

and the marginal average potential outcome is defined as follows:

$$\hat{\mu}_{\alpha}^{\text{ipw}} = \frac{1}{K}\sum_{i=1}^{K}\hat{Y}_i^{\text{IPW}}(\alpha) = \frac{1}{K}\sum_{i=1}^{K}\frac{\sum_{j=1}^{N_i}\hat{w}_{ij}^{\alpha}Y_{ij}}{\sum_{j=1}^{N_i}\hat{w}_{ij}^{\alpha}}, \tag{3}$$

where

$$\hat{w}_{ij}^{a\alpha} = \sum_{j=1}^{N_i}\frac{1(A_{ij} = a)1(R_{ij} = 1)P_{\alpha,x}(A_{i(-j)})}{\hat{\Pr}(A_i, R_{ij}|X_i;\ \hat{\delta}, \hat{\gamma})}, \tag{4}$$

$$\hat{w}_{ij}^{\alpha} = \sum_{j=1}^{N_i}\frac{1(R_{ij} = 1)P_{\alpha,x}(A_i)}{\hat{\Pr}(A_i, R_{ij}|X_i;\ \hat{\delta}, \hat{\gamma})}, \tag{5}$$

where $\hat{Y}_i^{\text{IPW}}(a, \alpha) = \sum_{j=1}^{N_i}\hat{w}_{ij}^{a\alpha}Y_{ij}/\sum_{j=1}^{N_i}\hat{w}_{ij}^{a\alpha}$, and $Y_i^{\text{IPW}}(\alpha) = \sum_{j=1}^{N_i}\hat{w}_{ij}^{\alpha}Y_{ij}/\sum_{j=1}^{N_i}\hat{w}_{ij}^{\alpha}$. Assume the estimation equations for δ and γ are $\sum_{i=1}^{K}\psi_{\delta}(O_i;\ \delta) = 0$ and $\sum_{i=1}^{K}\psi_{\gamma}(O_i;\ \gamma) = 0$, respectively, where $O_i = (X_i, Y_i, A_i, R_i)$. Let $\psi_a^{\text{IPW}}(O_i;\ \mu_{1\alpha}, \gamma, \delta) = \hat{Y}_i^{\text{IPW}}(1, \alpha) - \mu_{1\alpha}$, and $\psi_a^{\text{IPW}}(O_i;\ \mu_{0\alpha}, \gamma, \delta) = \hat{Y}_i^{\text{IPW}}(0, \alpha) - \mu_{0\alpha}$. Then the estimated parameter $\hat{\boldsymbol{\theta}}^{\text{IPW}} = (\hat{\delta}, \hat{\gamma}, \hat{\mu}_{1\alpha}^{\text{IPW}}, \hat{\mu}_{0\alpha}^{\text{IPW}})$ is a solution of the following estimation equations:

$$\boldsymbol{\psi}_K(\boldsymbol{\theta}) = \sum_{i=1}^{K}\boldsymbol{\psi}^{\text{IPW}}(O_i;\ \boldsymbol{\theta}) = \boldsymbol{0}, \tag{6}$$

where $\boldsymbol{\psi}^{\text{IPW}}(O_i;\ \boldsymbol{\theta}) = \{\psi_{\delta}(O_i;\ \delta), \psi_{\delta}(O_i;\ \gamma), \psi_a^{\text{IPW}}(O_i;\ \mu_{1\alpha}, \gamma, \delta), \psi_a^{\text{IPW}}(O_i;\ \mu_{0\alpha}, \gamma, \delta)\}^T$. The true values of unknown parameter $\boldsymbol{\theta} = (\delta, \gamma, \mu_{1\alpha}, \mu_{0\alpha})$ is the solution to $\boldsymbol{\psi}(\boldsymbol{\theta}) = \int\boldsymbol{\psi}^{\text{IPW}}(O_i;\ \theta)dF(o_i) = 0$, where F denotes the cumulative function of O_i. Note that the random observations O_i are i.i.d., and it can be shown that, for the last two of the true parameters $\boldsymbol{\theta}$, $\mu_{a\alpha} = E(\sum_{a_{i(-j)}\in\mathcal{A}(N_i-1)}Y_{ij}(a, a_{i(-j)})\pi(a_{i(-j)};\ \alpha)/N_i)$, $a = 0, 1$. By the strong law of large numbers, and the uniqueness of the solution to $\psi_a^{\text{IPW}}(O_i;\ \mu)$, we have that $\hat{Y}^{\text{IPW}}(a;\ \alpha)$ is consistent for $\mu_{a\alpha}$. The asymptotic normality follows by the M-estimation theory [17]. We show an example below of using equations (2) and (3) to obtain the IPW estimators. In this article, we utilize the direct interference pattern, where the interference happens between the units only through their treatment assignment. The treatment assignment is dependent on its own characteristics and the group random effect. For the missingness mechanism, we assume that missingness of each unit is only dependent on its own treatment assignment and characteristics. The outcome of each unit is dependent on its own characteristics and the treatment assignment of the whole group. For example, we may assume:

- $\text{logit}\{\Pr(A_{ij} = 1|X_{ij} = x_{ij}, R_{ij} = 1, b_i; \delta)\} = \delta_0 + \delta_1 x_{ij} + b_i$,
- $\text{logit}\{\Pr(R_{ij} = 1|A_{ij} = a_{ij}, X_{ij} = x_{ij}; \gamma)\} = \gamma_0 + \gamma_1 a_{ij} + \gamma_2 x_{ij}$, $f(y_{ij}|a_i, x_i, r_{ij} = 1) = \mathcal{N}(\beta_0 + \beta_1 + \boldsymbol{\beta}_2 \cdot (a_{ij}, f_a(a_{i(-j)}))^T + \beta_3 x_{ij}, \zeta_i)$, where $\mathcal{N}(\mu, \sigma^2)$ is a normal density function with mean μ and variance σ^2,
- $\text{logit}\{P_{\alpha,x}(a_{ij})\} = \xi_i^{(\alpha)} + \boldsymbol{\delta x}_{ij}$,

where $f_a(\cdot)$ is a summary function that represents the other units effect on the same cluster, $\zeta_i \sim \mathcal{N}(0, \sigma_\zeta^2)$ and $b_i \sim \mathcal{N}(0, \sigma_b^2)$ represent the random effects that introduce the dependency between units in the same cluster. Assume $\hat{p}_{ar}(X_i) = \Pr(A_{ij'}, R_{ij}|X_{ij}, b_i, \hat{\delta}, \hat{\gamma})$. The estimate of the joint probability of $\Pr(A_i, R_{ij}|X_i)$ can be obtained by

$$\Pr(A_i, R_{ij}|X_i, \hat{\delta}, \hat{\gamma}) = \int_{-\infty}^{\infty} \prod_{j'=1}^{N_i} \hat{p}_{ar}(X_i)^{A_{ij}}(1 - \hat{p}_{ar}(X_i))^{1-A_{ij}} f(b_i) db_i, \tag{7}$$

and $\hat{p}_{ar}(X_i)$ is obtained by equation (1). The consistency and asymptotic normality of the IPW estimator are presented in Theorem 1.

Theorem 1. *If $P(A_{ij} = 1|R_{ij} = 1, X_{ij}; \delta)$ and $P(R_{ij} = 1|A_{ij}, X_{ij}; \gamma)$ are correctly specified, then IPW estimator $\hat{\mu}_{\alpha\alpha}^{\text{IPW}}$ is consistent for $\mu_{\alpha\alpha}$, and*

$$\sqrt{K}(\hat{\boldsymbol{\theta}}^{\text{IPW}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Sigma^{\text{IPW}}),$$

as K goes to infinity, where $\Sigma^{\text{IPW}} = U(\boldsymbol{\theta})^{-1} V(\boldsymbol{\theta}) \{U(\boldsymbol{\theta})^{-1}\}^T$, $U(\boldsymbol{\theta}) = E\{-\partial \boldsymbol{\psi}^{\text{IPW}}(O_i; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T\}$, and $V(\boldsymbol{\theta}) = E\{\boldsymbol{\psi}^{\text{IPW}}(O_i; \boldsymbol{\theta}) \boldsymbol{\psi}^{\text{IPW}}(O_i; \boldsymbol{\theta})^T\}$.

The proof of Theorem 1 is given in the Supplementary materials.

3.2 Regression

In this section, we first introduce the idea of constructing the estimation equation for the regression estimator. Notice that by exchangibility assumption, we can estimate the causal effect by regressing Y on A and X, and then marginalize over X if confounders are fully observed. However, if confounders are subject to non-ignorable missingness, we cannot directly estimate the causal effect because we do not know the distribution of the confounders in the whole population. To recover the full data distribution, we need to obtain the joint probability density of X and Y conditional on A and $R = 0$ for each group, which requires estimation of the odds ratio model and the group level joint probability density of X and Y conditional on A and $R = 1$. First, we model $f(Y_{ij}|A_i, X_{ij}, R_{ij} = 1)$ as $f(Y_{ij}|A_i, X_{ij}, R_{ij} = 1; \beta)$, and model $f(X_{ij}|A_i, R_{ij} = 1)$ as $f(X_{ij}|A_i, R_{ij} = 1; \beta)$. We then define the odds ratio function $OR(X, Y|A)$ as follows:

$$\begin{aligned} OR(X_{ij}, Y_{ij}|A_i) &= \log \frac{f(X_{ij}, Y_{ij}|A_i, R_{ij} = 0) f(X_{ij} = x_0, Y_{ij}|A_i, R_{ij} = 1)}{f(X_{ij}, Y_{ij}|A_i, R_{ij} = 1) f(X_{ij} = x_0, Y_{ij}|A_i, R_{ij} = 0)} \\ &= \log \frac{P(R_{ij} = 0|A_i, X_{ij}, Y_{ij}) P(R_{ij} = 1|A, X_{ij} = x_0, Y_{ij} = 0)}{P(R_{ij} = 1|A, X_{ij}, Y_{ij}) P(R_{ij} = 0|A_i, X_{ij} = x_0, Y_{ij} = 0)} \\ &= \log \frac{P(R_{ij} = 0|A_i, X_{ij}) P(R_{ij} = 1|A_i, X_{ij} = x_0)}{P(R_{ij} = 1|A_i, X_{ij}) P(R_{ij} = 0|A_i, X_{ij} = x_0)} \\ &= \log \frac{f(X_{ij}|A_i, R_{ij} = 0) f(X_{ij} = x_0|A_i, R_{ij} = 1)}{f(X_{ij}|A_i, R_{ij} = 1) f(X_{ij} = x_0|A_i, R_{ij} = 0)}, \end{aligned} \tag{8}$$

where $1 \le i \le K, 1 \le j \le N_i$, and x_0 is an arbitrary fixed constant. For simplicity, we let $x_0 = 0$ in the subsequent sections. Since the last equation does not depend on Y, we can simplify the notation $OR(X_{ij}, Y_{ij}|A_i)$ as $OR(X_{ij}|A_i)$, which we model as $OR(X_{ij}|A_i;\ \zeta)$. Thus, we can parameterize the joint probability density of X and Y conditional on A and $R = 0$ as follows:

$$f(X_{ij}, Y_{ij}|A_i, R_{ij} = 0;\ \beta, \zeta) = \frac{\exp\{OR(X_{ij}|A_i;\ \zeta)\}f(X_{ij}, Y_{ij}|A_i, R_{ij} = 1;\ \beta)}{E[\exp\{OR(X_{ij}|A;\ \zeta)\}|A_i, R_{ij} = 1;\ \beta]}. \tag{9}$$

The parameter ζ can be estimated from the equation $\sum_{i=1}^{K}\psi_{\zeta}(O_i;\ \zeta) = 0$, where $\psi_{\zeta}(O_i;\ \zeta)$ has the following expression:

$$(1 - R_i)^T\{l(A_i, Y_i) - E\{l(A_i, Y_i)|A_i, R_i = 0;\ \hat{\beta}, \zeta\}\}, \tag{10}$$

where $l(A, Y)$ are pre-defined vectorized differentiable functions, and

$$E\{l(A_i, Y_{ij})|A_i, R_{ij} = 0;\ \hat{\beta}, \zeta\} = \frac{\int \exp\{OR(X_{ij}|A_i;\ \zeta)\}f(X_{ij}, Y_{ij}|A_i, R_{ij} = 1;\ \hat{\beta})l(A_i, Y_{ij})\mathrm{d}y}{E[\exp\{OR(X_{ij}|A_i;\ \zeta)\}|A_i, R_{ij} = 1;\ \hat{\beta}]}. \tag{11}$$

Once we have $f(X_{ij}, Y_{ij}|A_i, R_{ij} = 0;\ \beta, \zeta)$, we can obtain $f(X_{ij}, Y_{ij}|A_i;\ \beta, \zeta) = \sum_{r=0,1}f(X_{ij}, Y_{ij}|A_i, R_{ij} = r_{ij};\ \beta, \zeta)P(R= r_{ij}|A_i)$, where $P(R_{ij} = r_{ij}|A_{ij})$ can be directly estimated by maximum likelihood estimation (MLE) because R and A are fully observed. Note that $f(Y_{ij}|A_i, X_{ij};\ \beta, \zeta) \propto f(X_{ij}, Y_{ij}|A_i;\ \beta, \zeta)$; thus, we can obtain the regression estimators. More specifically, let $g_{ij}(a_i, x_i) = E(Y_{ij}|A_i = a_i, X_i = x_i, R_{ij} = 1)$, then the regression estimators for the average potential outcome and the marginal average potential outcome have the following expressions:

$$\begin{aligned}\hat{\mu}_{a\alpha}^{\mathrm{reg}} &= \frac{1}{K}\sum_{i=1}^{K}\hat{Y}_i^{\mathrm{reg}}(a, \alpha)\\ &= \frac{1}{K}\sum_{i=1}^{K}\Bigg[\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{a_{i(-j)}}1(R_{ij} = 1)\hat{g}_{ij}(a_i, X_i;\ \hat{\beta})P_{\alpha,x}(a_{i(-j)})\\ &\quad + \frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{a_{i(-j)}}1(R_{ij} = 0)P_{\alpha,x}(a_{i(-j)})E\{\hat{g}_{ij}(a_i, X_i;\ \hat{\beta})|A_i = a_i, R_{ij} = 0;\ \hat{\beta}, \hat{\zeta}\}\Bigg],\end{aligned} \tag{12}$$

$$\begin{aligned}\hat{\mu}_{\alpha}^{\mathrm{reg}} &= \frac{1}{K}\sum_{i=1}^{K}\hat{Y}_i^{\mathrm{reg}}(\alpha)\\ &= \frac{1}{K}\sum_{i=1}^{K}\Bigg[\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{a_i}1(R_{ij} = 1)\hat{g}_{ij}(a_i, X_i;\ \hat{\beta})P_{\alpha,x}(a_i)\\ &\quad + \frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{a_i}1(R_{ij} = 0)P_{\alpha,x}(a_i)E\{\hat{g}_{ij}(a_i, X_i;\ \hat{\beta})|A_i = a_i, R_{ij} = 0;\ \hat{\beta}, \hat{\zeta}\}\Bigg],\end{aligned} \tag{13}$$

where $\hat{\beta}$ is the MLE of β and $\hat{\zeta}$ is an estimator of ζ by equation (10). Assume the estimation equation for β is $\sum_{i=1}^{K}\psi_{\beta}(O_i;\ \beta) = 0$. Let $\psi_a^{\mathrm{reg}}(O_i;\ \mu_{1\alpha}, \zeta, \beta) = \hat{Y}_i^{\mathrm{reg}}(1, \alpha) - \mu_{1\alpha}$, and $\psi_a^{\mathrm{reg}}(O_i;\mu_{0\alpha}, \zeta, \beta) = \hat{Y}_i^{\mathrm{reg}}(0, \alpha) - \mu_{0\alpha}$ denote the estimation equations for $\mu_{1\alpha}$ and $\mu_{0\alpha}$, respectively. Then the estimated parameter $\hat{\boldsymbol{\theta}}^{\mathrm{reg}} = (\hat{\beta}, \hat{\zeta}, \hat{\mu}_{1\alpha}^{\mathrm{reg}}, \hat{\mu}_{0\alpha}^{\mathrm{reg}})$ is a solution of the following estimation equations:

$$\sum_{i=1}^{K}\psi^{\mathrm{reg}}(O_i;\ \boldsymbol{\theta}) = \boldsymbol{0}. \tag{14}$$

The consistency and normality of the regression estimator are presented in Theorem 2.

Theorem 2. *If the baseline outcome regression model $f(Y_{ij}|A_i, X_{ij}R_{ij} = 1;\ \beta)$ and the distribution of observed confounders $f(X_{ij}|A_i, R_{ij} = 1)$ are correctly specified, then the regression estimator $\hat{\mu}_{aa}^{\text{reg}}$ is consistent for μ_{aa},* and

$$\sqrt{K}(\hat{\boldsymbol{\theta}}^{\text{reg}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \Sigma^{\text{reg}}),$$

as K goes to infinity, where $\Sigma^{\text{reg}} = U(\boldsymbol{\theta})^{-1}V(\boldsymbol{\theta})\{U(\boldsymbol{\theta})^{-1}\}^T$, $U(\boldsymbol{\theta}) = E\{-\partial\psi^{\text{reg}}(O_i;\ \boldsymbol{\theta})/\partial\boldsymbol{\theta}^T\}$, and $V(\boldsymbol{\theta}) = E\{\boldsymbol{\psi}^{\text{reg}}(O_i;\ \boldsymbol{\theta})\boldsymbol{\psi}^{\text{reg}}(O_i;\ \boldsymbol{\theta})^T\}$.

The proof of Theorem 2 is given in the Supplementary materials.

3.3 Doubly robust estimation

In this section, we aim to propose a DR estimator $\hat{\mu}_{aa}$ in the sense that it is consistent if either the propensity score models or the baseline regression models, but not necessarily both, are correctly specified. The typical DR estimator involves two parts, where the first part is the regression term, and the second part is the inverse probability weighted residuals of the regression estimator. When confounders are missing not at random, both the propensity score of missingness and the outcome regression model contain the odds ratio model which can be seen in the definition in equation (8). Therefore, the specification of propensity score models and the regression models are not variational independent. Hence, the specification of the propensity score models and the regression model cannot be fully seperated. More specifically, in our setting, the specification of both the propensity score of missingness and the probability $f(x_{ij}, y_{ij}|a_i, r_{ij} = 0)$ contains the odds ratio $OR(x_{ij}|a_i)$, that is, $OR(x_{ij}|a_i)$ lies in the intersection of the IPW estimator and the regression estimator. Thus, to construct the DR estimator, we assume $OR(x|a;\ \zeta)$ is always correctly specified, and the proposed DR estimator is consistent if either the IPW models, $\Pr(a_{ij} = 1|r_{ij} = 1, x_{ij};\ \delta)$ and $\Pr(r_{ij} = 1|a_{ij} = 1, x_{ij};\ \gamma')$, are correctly specified or the outcome regression model conditional on the observed values, $f(y_{ij}|a_i, x_{ij}, r_{ij} = 1, \beta)$ and $f(x_{ij}|a_i, r_{ij} = 1, \beta)$, are correctly specified, where $\gamma = (\zeta, \gamma')$. ζ is obtained by solving the equation $\sum_{i=1}^{K}\psi_\zeta(O_i;\ \zeta) = 0$, where $\psi_\zeta(O_i;\ \zeta) = \mathbf{v}^T(A_i, R_i, X_i;\ \delta, \xi)\boldsymbol{\phi}(A_i, Y_i;\ \hat{\beta}, \hat{\xi})$,

$$v(A_i, R_i, X_i;\ \delta, \xi) = \begin{bmatrix} \dfrac{R_{i1}}{\Pr(R_{i1} = 1|A_{i1}, X_{i1}, \hat{\delta}, \hat{\xi})} - 1 \\ \dfrac{R_{i2}}{\Pr(R_{i2} = 1|A_{i2}, X_{i2}, \hat{\delta}, \hat{\xi})} - 1 \\ \cdots \\ \dfrac{R_{iN_i}}{\Pr(R_{iN_i} = 1|A_{iN_i}, X_{iN_i}, \hat{\delta}, \hat{\xi})} - 1 \end{bmatrix}, \quad \text{and} \tag{15}$$

$$\phi(A_i, Y_i;\ \hat{\beta}, \hat{\xi}) = \begin{bmatrix} l(A_{i1}, Y_{i1}) - \mathbf{E}\{l(A_{i1}, Y_{i1})|A_i, R_{i1} = 0;\ \hat{\beta}, \hat{\xi}\} \\ l(A_{i2}, Y_{i2}) - \mathbf{E}\{l(A_{i2}, Y_{i2})|A_i, R_{i2} = 0;\ \hat{\beta}, \hat{\xi}\} \\ \cdots \\ l(A_{iN_i}, Y_{iN_i}) - \mathbf{E}\{l(A_{iN_i}, Y_{iN_i})|A_i, R_{iN_i} = 0;\ \hat{\beta}, \hat{\xi}\} \end{bmatrix}. \tag{16}$$

Since the confounders are not fully observed, we replace the weights in the residuals of the regression estimator in the traditional DR estimator by the joint probability of A and R conditional on X, and construction of the regression estimator term is similar as that in Section 3.2. Hence, the DR estimator has the following representation:

$$\begin{aligned} \hat{\mu}_{aa}^{dr} = \frac{1}{K}\sum_{i=1}^{K}\Bigg[\sum_{j=1}^{N_i}\Bigg[& E\{\hat{h}_{ij}^{a}(A_i, X_i, Y_i)|A_i = a_i, R_{ij} = 0;\ \hat{\delta}, \hat{\gamma}, \hat{\beta}\} + \frac{1(R_{ij} = 1)}{\Pr(R_{ij}|A_i = a_i, X_i;\ \gamma)} \\ & \times \{\hat{h}_{ij}^{a}(A_i, X_i, Y_i) - E\{\hat{h}_{ij}^{a}(A_i, X_i, Y_i)|A_i = a_i, R_{ij} = 0;\ \hat{\delta}, \hat{\gamma}, \hat{\beta}\}\}\Bigg]\Bigg] \end{aligned} \tag{17}$$

and

$$\hat{\mu}_\alpha^{dr} = \frac{1}{K}\sum_{i=1}^{K}\left[\sum_{j=1}^{N_i}\left[E\{\hat{h}_{ij}(A_i, X_i, Y_i)|A_i = a_i, R_{ij} = 0; \hat{\delta}, \hat{\gamma}, \hat{\beta}\} + \frac{1(R_{ij} = 1)}{\Pr(R_{ij}|A_i, X_i; \gamma)}\right.\right.$$
$$\left.\left.\{\hat{h}_{ij}(A_i, X_i, Y_i) - E\{\hat{h}_{ij}(A_i, X_i, Y_i)|A_i, R_{ij} = 0; \hat{\delta}, \hat{\gamma}, \hat{\beta}\}\}\right]\right], \tag{18}$$

where $h_{ij}^a(A_i, X_i, Y_i; \delta, \gamma, \beta)$, $h_{ij}(A_i, X_i, Y_i; \delta, \gamma, \beta)$ have the following expressions:

$$h_{ij}^a(A_i, X_i, Y_i; \delta, \gamma, \beta) = w_{ij}^{a\alpha}(Y_{ij} - g_{ij}(A_i, X_i; \beta)) + \frac{1}{N_i}\sum_{a_{i(-j)}} P_{\alpha,x}(a_{i(-j)})g_{ij}(a, X_i, \beta), \tag{19}$$

$$h_{ij}(A_i, X_i, Y_i; \delta, \gamma, \beta) = w_{ij}^{\alpha}(Y_{ij} - g_{ij}(A_i, X_i; \beta)) + \frac{1}{N_i}\sum_{a_i} P_{\alpha,x}(a_i)g_{ij}(a, X_i, \beta), \tag{20}$$

$$w_{ij}^{a\alpha} = \frac{1(A_{ij} = a)P_{\alpha,x}(A_i)}{\Pr(A_i|X_i; \delta, \gamma)} \Bigg/ \sum_{j=1}^{N_i} \frac{1(A_{ij} = a)P_{\alpha,x}(A_i)}{\Pr(A_i|X_i; \delta, \gamma)}, \tag{21}$$

and

$$w_{ij}^{\alpha} = \frac{P_{\alpha,x}(A_i)}{\Pr(A_i|X_i; \delta, \gamma)} \Bigg/ \sum_{j=1}^{N_i} \frac{P_{\alpha,x}(A_i)}{\Pr(A_i|X_i; \delta, \gamma)}. \tag{22}$$

The basic idea of the construction of $h_{ij}^a(A_i, X_i, Y_i; \delta, \gamma, \beta)$ lies in that equation (19) is a doubly robust estimator of $\mu_{a\alpha}$, in the sense that the expectation of $h^a(A, X, Y)$ converges to $\mu_{a\alpha}$ when data are fully observed, that is, $E\{\sum_{j=1}^{N_i} h_{ij}^a(A_i, X_i, Y_i; \delta, \gamma, \beta)\} = \mu_{a\alpha}$, if either the IPW models or regression models are correctly specified. When the propensity score models are correctly specified,

$$\begin{aligned} &E\left\{\sum_{j=1}^{N_i} h_{ij}^a(A_i, X_i, Y_i; \delta, \gamma, \beta)\right\} \\ &\quad = E\left[\sum_{j=1}^{N_i}\left\{w_{ij}^{a\alpha}Y_{ij} + \sum_{a_{i(-j)}} P_{\alpha,x}(A_{i(-j)})g_{ij}(a, X_i, \beta) - w_{ij}^{a\alpha}g_{ij}(A_i, X_i; \beta)\right\}\right] \\ &\quad = \mu_{aa}^{\text{ipw}}. \end{aligned}$$

Similarly, when the baseline regression models have been correctly specified, it is obvious to see that the expectation of the first part of equation (19) is equal to zero, and the expectation of the second part is the regression estimator, that is, $E\{\sum_{j=1}^{N_i} h_{ij}^a(A_i, X_i, Y_i; \delta, \gamma, \beta)\} = \mu_{aa}^{\text{reg}}$. Therefore, to obtain the doubly robust estimator when confounders are subject to non-ignorable missingness, we can replace the observed outcomes and regression estimator in the conventional residuals weighted DR estimator by $h_{ij}^a(A_i, X_i, Y_i; \delta, \gamma, \beta)$ and $E\{h_{ij}(X_i, Y_i; \delta, \gamma, \beta)|A_i, R_{ij} = 0\}$, respectively. Hence, the constructed estimator can be viewed as a new residuals weighted DR estimator, where the weights are the inverse of the estimated propensity score of missingness, and the residuals come from the constructed $h_{ij}^a(A_i, X_i, Y_i; \delta, \gamma, \beta)$ instead of the regression estimator. By allowing either one of the set of models to be correctly specified, the doubly robustness property can also be achieved accordingly. For estimating γ, δ, β, the estimators can be obtained by maximum likelihood approach using observed data when $R = 1$. While our current study primarily focuses on the assumption of homogeneous parameters, β, γ, and δ, it is worth considering the potential implications of allowing these parameters to vary across clusters. In such a scenario, the parameters of the group average potential outcome would exhibit variation across different clusters. The estimation equation for γ, δ, β can be written as $\sum_{i=1}^{K}\psi_\alpha(O_i; \gamma) = 0$, $\sum_{i=1}^{K}\psi_\delta(O_i; \delta) = 0$, and $\sum_{i=1}^{K}\psi_\beta(O_i; \beta) = 0$, respectively. Let $\psi_a^{dr}(O_i; \mu_{1\alpha}, \zeta, \beta) = \hat{Y}_i^{dr}(1, \alpha) - \mu_{1\alpha}$, and $\psi_a^{dr}(O_i; \mu_{0a}, \zeta, \beta) = \hat{Y}_i^{dr}(0, \alpha) - \mu_{0a}$ denote the estimation equations for μ_{1a} and μ_{0a}, respectively. Then the estimated parameter $\hat{\boldsymbol{\theta}}^{dr} = (\hat{\alpha}, \hat{\delta}, \hat{\beta}, \hat{\zeta}, \hat{\mu}_{1a}^{dr}, \hat{\mu}_{0a}^{dr})$ is a solution of the following estimation equations:

$$\sum_{i=1}^{K} \psi^{dr}(O_i; \boldsymbol{\theta}) = \mathbf{0}. \tag{23}$$

The consistency and asymptotic normality of the proposed DR estimator are summarized in the following theorem.

Theorem 3. *Assume $OR(X_{ij}|A_i; \zeta)$ is correctly specified, if either* (*a*) $\mathrm{Pr}(R_{ij} = 1|A_{ij} = 1, X_i; \delta)$ *and* $\mathrm{Pr}(A_{ij} = 1|R_{ij} = 1, X_{ij}; \gamma)$ *or* (ii) $f(Y_{ij}|A_i, X_{ij}, R_{ij} = 1; \beta)$ *and* $f(X_{ij}|A_i, R_{ij} = 1)$ *are correctly specified; then the DR estimator* $\hat{\mu}_{aa}^{dr}$ *is consistent for* μ_a, *where* $\hat{\mu}_{aa}^{dr}$ *is obtained from equation* (17), *and*

$$\sqrt{K}(\hat{\boldsymbol{\theta}}^{dr} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma^{dr}),$$

as K goes to infinity, where $\Sigma^{dr} = U(\boldsymbol{\theta})^{-1}V(\boldsymbol{\theta})\{U(\boldsymbol{\theta})^{-1}\}^T$, $U(\boldsymbol{\theta}) = E\{-\partial\psi^{dr}(O_i; \boldsymbol{\theta})/\partial\boldsymbol{\theta}^T\}$, *and* $V(\boldsymbol{\theta}) = E\{\psi^{dr}(O_i; \boldsymbol{\theta})\psi^{dr}(O_i; \boldsymbol{\theta})^T\}$.

The proof of Theorem 3 is given in the Supplementary materials.

4 Simulation

In this section, we conduct a simulation study to illustrate the proposed IPW, regression, and DR estimator. First, under setting 1 (S1), we generate a population of size $N = 10{,}000$, and randomly classify the whole population into $K = 100$ mutually exclusive groups with $N_i = 100$ individuals in each group. Our model generation process contains the generation of (X, A, R, Y) in order. First, we generate the confounders X, followed by generating A using the specified model for the propensity score of treatment. Next, we generate the missingness indicator using a model for the missingness mechanism with A and X. Finally, we generate the outcome Y using an outcome regression model that incorporates A and X. More specifically, for each individual, the covariate X_1 is generated from Bernoulli distribution with probability 0.5, and the covariate X_2 is generated from the standard normal distribution. To generate treatment and missing indicators, we assume two logistic models. We assume a mixed-effect logistic model $\mathrm{logit}\{\mathrm{Pr}(A_{ij} = 1|R_{ij} = 1, X_{ij} = x_{ij}, b_i)\} = 0.1 + 0.2x_{1ij} - 0.1x_{2ij} + b_i$ for the propensity score of treatment, where b_i is group level random effect terms generated from $\mathcal{N}(0, \sigma^2)$, i.e., the normal distribution with mean 0 and variance σ^2. We assume a logistic model $\mathrm{logit}\{\mathrm{Pr}(R_{ij} = 1|A_{ij} = a_{ij}, X_{ij} = x_{ij})\} = 1 + a_{ij} + 1.2x_{1ij} - 0.5x_{2ij}$ for the propensity score model of missingness. Then the joint propensity score for treatment and missingness with random effects $\mathrm{Pr}(a_{ij}, r_{ij} = 1|x_{ij}, b_i)$ is calculated according to equation (3) by Chen [16], and the missingness indicator R_{ij} and the treatment indicator A_{ij} are generated from the joint propensity score, which is given by the following equation:

$$\mathrm{Pr}(A_i, R_{ij} = 1|X_i) = \int \prod_{j'=1}^{n_i} \{h_{ij'}^{11}(b_i)\}^{A_{ij}}\{h_{ij'}^{01}(b_i)\}^{(1-A_{ij})} f_b(b_i; \sigma^2) db_i,$$

where $h_{ij}^{ar}(b_i) = \mathrm{Pr}(a_{ij} = a, r_{ij} = r|x_{ij}, b_i)$. Finally, we generate the outcome Y_{ij} from $Y_{ij} = 1 + 2x_{1ij} - 3x_{2ij} + 0.5x_{1ij}x_{2ij} + 2a_{ij} + 3p_i(a_i) + \varepsilon_{ij}$, where $p_i(a_i)$ is the proportion of units in group i that receive the treatment, and $\{\varepsilon_{ij}\}_{1\le i\le K, 1\le j\le N_i}$ are i.i.d. random noise terms generated by $\mathcal{N}(0, \sigma^2)$. We consider four settings for the number of observations and the variance of the group effect:

(S1) $N = 10{,}000$, $K = 50$, and $\sigma^2 = 0.25$;
(S2) $N = 10{,}000$, $K = 50$, and $\sigma^2 = 0.16$;
(S3) $N = 12{,}000$, $K = 200$, and $\sigma^2 = 0.25$;
(S4) $N = 12{,}000$, $K = 200$, and $\sigma^2 = 0.16$.

Second, letting $\alpha = 0.5$, the propensity score models are correctly specified as $\Pr(a_{ij} = 1 | r_{ij} = 1, x_{ij}; \delta) = \text{logit}^{-1}(\delta_1 + \delta_2 x_{1ij} + \delta_3 x_{2ij} + b_i)$ and $\Pr(r = 1 | a, x; \gamma) = \text{logit}^{-1}(\gamma_1 + \gamma_2 a + \gamma_1 x_{1ij} + \gamma_2 x_{2ij})$. The parameter $\delta = (\delta_1, \delta_2, \delta_3)$ and $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ are estimated with *lme4* [18] package in *R* [19]. The regression model is correctly specified as $E(y_{ij} | a_i, x_i, r_{ij} = 1) = \beta_0 + \beta_1 a_{ij} + \beta_2 p_i(a_i) + \beta_3 x_{1ij} + \beta_4 x_{2ij} + \beta_5 x_{1ij} x_{2ij}$. The parameters β are estimated by MLE. Then, the IPW estimator, regression estimator, and doubly robust estimators (DR-TT: when all models are correctly specified), i.e., $\hat{\mu}_{aa}^{\text{ipw}}$, $\hat{\mu}_{aa}^{\text{reg}}$, $\hat{\mu}_{aa}^{dr}$, are calculated according to equations (2), (12), and (17), respectively.

Third, in the scenario when there exist model misspecification, DR estimator DR-TF is calculated when the outcome model is misspecified as $E(y_{ij} | a_i, x_i, r_{ij} = 1) = \beta_0 + \beta_1 a_{ij} + \beta_3 x_{1ij} + \beta_4 x_{2ij} + \beta_5 x_{1ij} x_{2ij}$, and the propensity score models are correctly specified.

The DR estimator DR-FT is calculated when the propensity score of treatment is incorrectly specified as $\Pr(a_{ij} = 1 | r_{ij} = 1, x_{ij}; \delta) = \text{logit}^{-1}(\delta_1 + \delta_2 x_{1ij} + b_i^{(1)})$, and the the outcome model is correctly specified.

Finally, the DR estimator DR-FF is calculated when the propensity score for treatment is incorrectly specified as $\Pr(a_{ij} = 1 | r_{ij} = 1, x_{ij}; \delta) = \text{logit}^{-1}(\delta_1 + \delta_2 x_{1ij} + b_i^{(1)})$, and the outcome model is misspecified as $E(y_{ij} | a_i, x_i, r_{ij} = 1) = \beta_0 + \beta_1 a_{ij} + \beta_3 x_{1ij} + \beta_4 x_{2ij} + \beta_5 x_{1ij} x_{2ij}$.

The simulation study is repeated 1,000 times, and the estimators of α, γ, β, $\mu_{1\alpha}$, $\mu_{0\alpha}$, and $\bar{D}E(\alpha)$ are summarized in the following tables.

The bias and empirical coverages of the proposed estimators are shown in Tables 1, 2, and Figure 2, where the 95% Wald-type confidence intervals are constructed according to the asymptotic distribution of the

Table 1: Bias, empirical standard error (se), and standard deviation [in brackets] for IPW estimators, regression estimators, and doubly robust estimators under S1,S2, S3, and S4

		IPW	REG	DR-TT	DR-TF	DR-FT	DR-FF
S1	$\mu_{1,0.5}$	0.005[0.051]	−0.037[0.013]	−0.062[0.073]	0.043[0.072]	0.081[0.046]	−0.102[0.191]
	$\mu_{0,0.5}$	0.003[0.059]	−0.003[0.047]	−0.023[0.082]	0.023[0.080]	0.030[0.067]	0.054[0.072]
	$\bar{\mathrm{D}}\mathrm{E}(0.5)$	0.002[0.082]	−0.033[0.130]	−0.040[0.072]	0.020 [0.068]	0.050[0.091]	−0.156[0.104]
	$se(\mu_{1,0.5})$	0.152	0.030	0.078	0.103	0.102	0.181
	$se(\mu_{0,0.5})$	0.110	0.032	0.102	0.110	0.085	0.138
	$se(\bar{\mathrm{D}}\mathrm{E}(0.5))$	0.114	0.069	0.080	0.075	0.113	0.166
S2	$\mu_{1,0.5}$	0.035[0.109]	0.034[0.059]	−0.018[0.042]	0.017[0.046]	−0.023[0.047]	0.060[0.121]
	$\mu_{0,0.5}$	0.018[0.122]	−0.008[0.057]	−0.005[0.042]	−0.007[0.042]	0.041[0.061]	−0.796[0.092]
	$\bar{\mathrm{D}}\mathrm{E}(0.5)$	0.016[0.118]	0.042[0.077]	0.020[0.052]	0.019 [0.058]	−0.019[0.080]	0.860[0.142]
	$se(\mu_{1,0.5})$	0.302	0.079	0.054	0.052	0.041	0.084
	$se(\mu_{0,0.5})$	0.111	0.073	0.057	0.057	0.069	0.094
	$se(\bar{\mathrm{D}}\mathrm{E}(0.5))$	0.234	0.069	0.067	0.064	0.088	0.171
S3	$\mu_{1,0.5}$	−0.005[0.180]	0.019[0.046]	−0.009[0.041]	0.037[0.027]	−0.004[0.031]	−0.105[0.081]
	$\mu_{0,0.5}$	−0.012[0.041]	0.021[0.042]	−0.014[0.029]	−0.020[0.028]	0.028[0.032]	0.049[0.081]
	$\bar{\mathrm{D}}\mathrm{E}(0.5)$	0.007[0.099]	−0.002[0.010]	0.006[0.035]	0.057 [0.032]	−0.03[0.041]	−0.155[0.072]
	$se(\mu_{1,0.5})$	0.154	0.036	0.041	0.037	0.029	0.090
	$se(\mu_{0,0.5})$	0.045	0.038	0.042	0.037	0.018	0.110
	$se(\bar{\mathrm{D}}\mathrm{E}(0.5))$	0.137	0.089	0.041	0.051	0.033	0.080
S4	$\mu_{1,0.5}$	0.110[0.162]	0.022[0.046]	−0.030[0.018]	0.010[0.022]	−0.011[0.032]	0.053[0.085]
	$\mu_{0,0.5}$	−0.018[0.035]	0.013[0.038]	−0.002[0.017]	−0.001[0.017]	0.028[0.034]	0.053[0.085]
	$\bar{\mathrm{D}}\mathrm{E}(0.5)$	0.128[0.185]	0.008[0.019]	−0.028[0.024]	0.011 [0.033]	−0.040[0.042]	0.160[0.074]
	$se(\mu_{1,0.5})$	0.225	0.048	0.037	0.027	0.037	0.103
	$se(\mu_{0,0.5})$	0.149	0.048	0.037	0.027	0.028	0.120
	$se(\bar{\mathrm{D}}\mathrm{E}(0.5))$	0.249	0.015	0.041	0.040	0.038	0.090

Table 2: Coverage probability (%) of IPW estimators, regression estimators, and doubly robust estimators under S1, S2, S3, and S4

		IPW	REG	DR-TT	DR-TF	DR-FT	DR-FF
S1	$\mu_{1,0.5}$	97.0	91.0	94.7	92.8	98.5	0
	$\mu_{0,0.5}$	96.0	91.0	91.2	94.0	95.5	0
	$\overline{\text{DE}}(0.5)$	95.5	89.5	90.5	92.5	93.0	0
S2	$\mu_{1,0.5}$	98.5	89.5	98.5	93.5	90.5	0
	$\mu_{0,0.5}$	95.5	92.5	91.5	95.5	90.5	0
	$\overline{\text{DE}}(0.5)$	98.5	92.0	93.0	92.0	88.5	0
S3	$\mu_{1,0.5}$	95.5	93.0	96.2	94.8	96.8	59.8
	$\mu_{0,0.5}$	96.0	93.0	95.0	93.6	94.8	69.8
	$\overline{\text{DE}}(0.5)$	97.0	95.5	97.2	96.0	89.2	56.6
S4	$\mu_{1,0.5}$	93.0	90.0	97.0	92.0	92.5	50.6
	$\mu_{0,0.5}$	97.0	96.0	98.5	96.5	91.5	57.4
	$\overline{\text{DE}}(0.5)$	92.0	92.5	96.5	92.5	85.5	56.0

proposed estimators in Theorems 1–3. When both the IPW models and the regression model are correctly specified, the IPW, regression, and DR estimators all perform well in terms of having small bias and variances, and the DR robust estimator DR-TT has the smallest variance among all the estimators, which supports our theoretical result that the associated DR estimator achieves the semiparametric efficiency bound when confounders are fully observed and all the working models are correctly specified. When the outcome regression model is correctly specified, the regression estimator has the smallest variance among all the proposed estimators. When both the model of propensity score for treatment and the regression model are misspecified, the bias of $\hat{\mu}_{0\alpha}^{dr}$ has the same magnitude with the other estimators when either the IPW or the regression model,

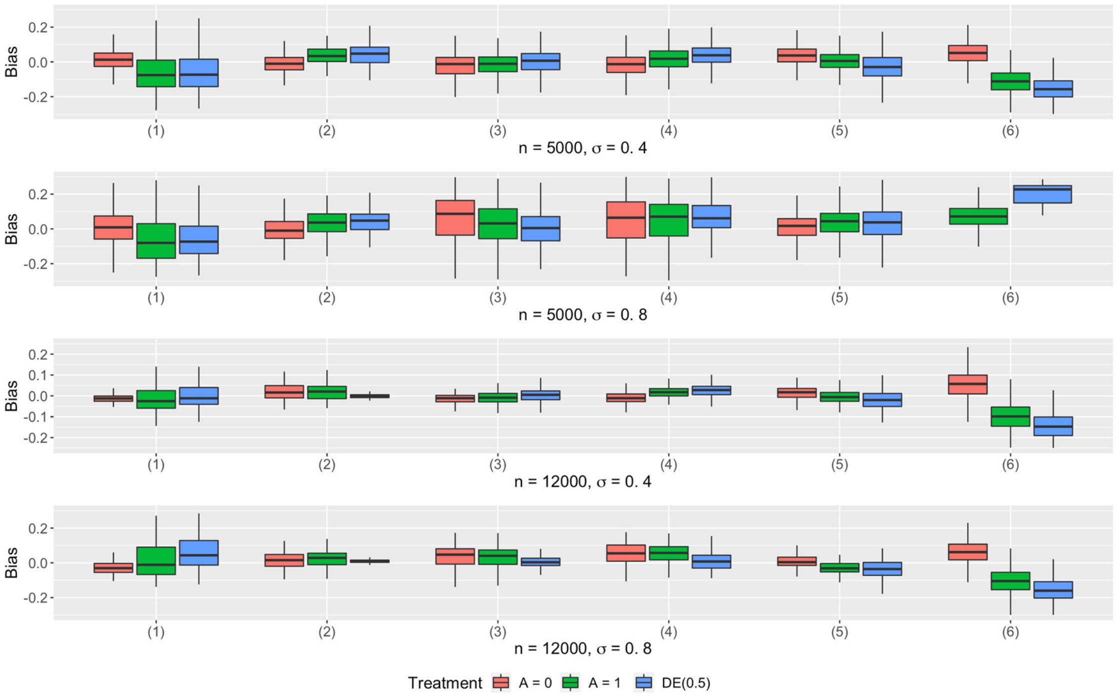

Figure 2: Bias of the (1) IPW, (2) regression, and DR estimators for $\mu_{1,0.5}$ and $\mu_{0,0.5}$ under four scenarios: (3) both propensity and outcome regression models are correctly specified; (4) propensity score models are correctly specified; (5) when only outcome model is correctly specified; (6) when neither the outcome regression model nor the IPW models are correctly specified. (The boxplot of $\hat{\mu}_{0,0.5}^{dr}$ in the second scenario is dropped because the absolute value of the bias is greater than 0.2.).

but not both, is correctly specified. As shown in Table 2, all three estimators achieve nominal levels when corresponding models are correctly specified, which indicates the standard error formulas proposed in Theorem 1, 2, and 3 are valid; when both propensity score and regression models are misspecified, the coverage probability of the DR estimator decreases to zero.

5 Application

The particulate matter 2.5 (PM2.5) refers to tiny particles or droplets in the air that can affect people's short-term or even long-term health conditions such as respiratory issues, increased mortality from lung cancer, and heart disease. A primary strategy to achieve the reduction of ambient PM2.5 is the installation of flue-gas desulfurization equipment or controls ("scrubbers") to reduce the sulfur dioxide (SO_2), nitrogen dioxide (NO_x), and carbon dioxide (CO_2) emissions, which are three main air pollutants that mediate the changes of PM2.5. In 1990, the U.S. Clean Air Act (CAA) amendments launched the ARP to reduce the emissions of the ambient SO_2, NO_x, and CO_2 by regulating those power plants to install scrubbers on coal-fired electricity-generating units (EGUs).

In this section, we apply the proposed estimators on a dataset from the U.S. EPA's Air Quality System (AQS) to estimate the causal effect of installing the scrubbers on the reduction of ambient NO_x emissions. The dataset contains monthly emissions data from 1218 EGUs in the United States in 2004. The 1218 EGUs are classified into 40 clusters through the linkage algorithm by Zigler et al. [5]. Among 1218 EGUs, 913 EGUs installed the scrubbers and 205 EGUs did not install the scrubbers. Six important characteristics of the EGUs and the atmosphere are included in the data analysis: the heat input rate, the capacity of the EGU, the number of scrubbers installed, the sulfur content, the operation time, and the average temperature in the previous year. As these characteristics affect both the emissions of NO_x and the installation of scrubbers, they are treated as confounders of the causal relationship between the outcome and the treatment. For the treatment ("scrubbers" installation) coverage probability, the average of the proportion of treated units among all the groups is 66.79%. For the missingness in confounders, there are 15.76% missingness in sulfur content, 2.25% missingness in operation time, 21.69% missingness in heat input rate, 0.99% in capacity, and 24.90% missingness in total. The missingness can be caused by multiple reasons such as monitor errors (e.g., the operation time exceeds a certain time), the failure of recording, and the changes of filter. Since the units with fewer NO_x emissions are less likely to disclose the baseline characteristics, and records of the baseline characteristics were measured long before the emissions of NO_x took place, it is plausible that the outcome-independence assumption holds.

We assume a linear fixed-effect regression model for the outcome, and assume different mixed-effects logistic regression models for the propensity score of treatment and missingness, respectively. To account for the dependency of the treatment allocation strategy on the baseline characteristics, following Papadogeorgou et al. [10], we assume $P_{\alpha,x}(a_i)$ as the logistic regression model: $\text{logit}\{P_{\alpha,x}(a_{ij})\} = \xi_i^\alpha + \sum_{j=1}^{6}\beta_i X_{ij}$, where ξ is estimated by solving the following equation,

$$\frac{1}{N_i}\sum_{j=1}^{N_i}\text{expit}(\xi_i^\alpha + \beta_j X_{ij}) = \alpha.$$

Since in the dataset there are over 80% of units that lying in the clusters with average proportion of units with scrubbers installed ranges from 0.3 to 0.8, we consider values of α varying from 0.3 to 0.8.

Figure 3 presents the results of IPW, regression, and DR estimators for the direct effect DE(α) across different values of α. When $\alpha = 0.3$, DE(α) of the IPW, regression, and DR estimators are −29.45, −49.25, and −34.72, respectively. When $\alpha = 0.8$, DE(α) of the IPW, regression and DR estimators are −19.55, −21.99, and −17.44, respectively. The estimated DE(α) are negative for all three estimators across different α values, indicating that the intervention of installing scrubbers has a positive effect on reducing the emissions of the tons of NO_x, and there would be approximately 30 tons of NO_x emissions fewer per unit among the units with scrubbers installed than that among the units without scrubbers installed. As α increases, DE(α) has an

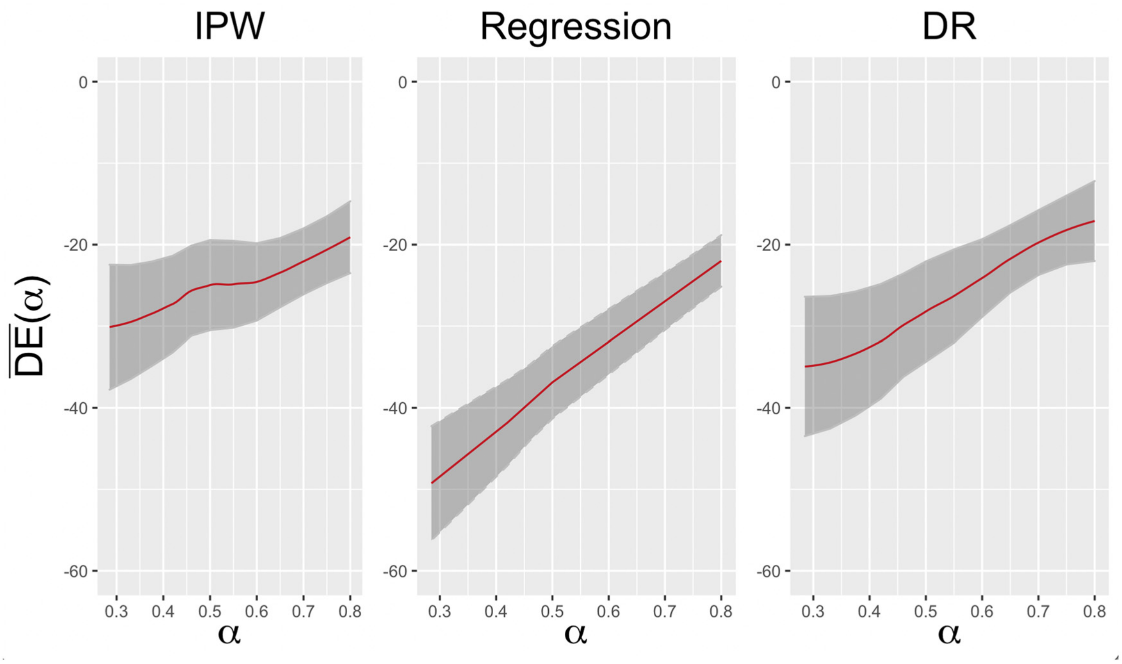

Figure 3: Estimates and 95% Wald-type confidence intervals of $\bar{DE}(\alpha)$ of scrubber installation for (1) IPW, (2) regression, and (3) DR estimator with $\alpha \in (0.3, 0.8)$, where the shadow area represents the pointwise confidence intervals.

increasing trend, which implies that the intervention at one EGU is beneficial for the reduction of the emissions of NO$_x$, but the effect is smaller when the proportion of the units within the same cluster that have scrubbers installed becomes larger.

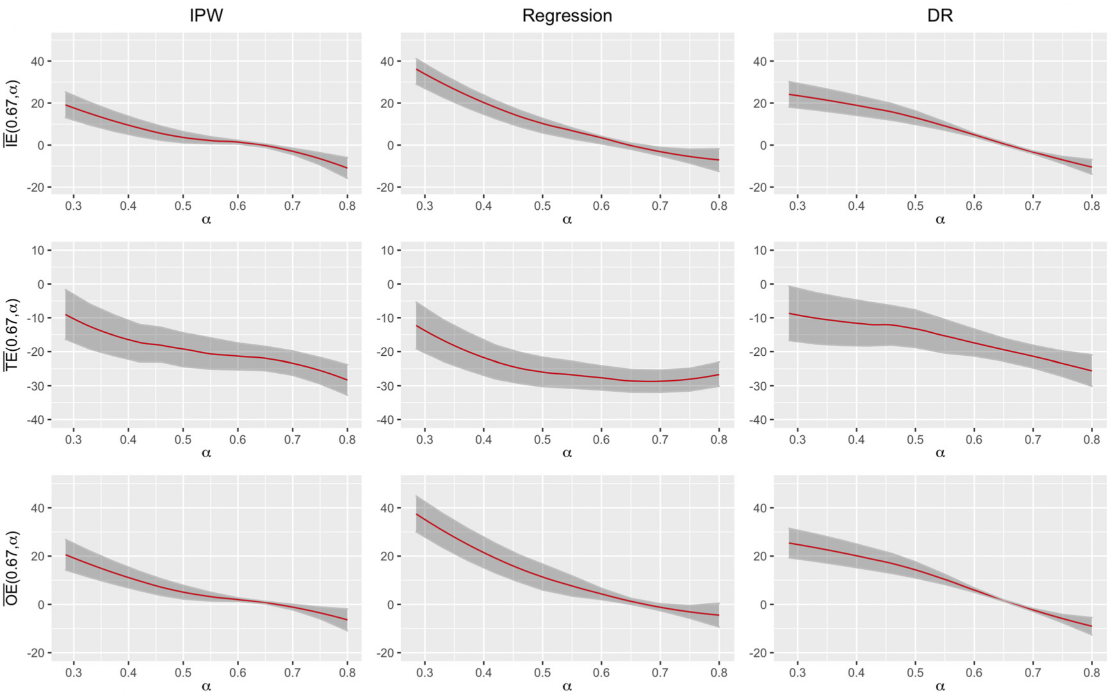

Figure 4: Estimates and 95% Wald-type confidence intervals of $\bar{IE}(\alpha)$, $\bar{TE}(\alpha)$, and $\bar{OE}(\alpha)$ of scrubber installation for (1) IPW, (2) regression, and (3) DR estimator with $\alpha \in (0.3, 0.8)$, where the shadow area represents the pointwise confidence intervals.

Let α be the average treatment coverage probability among 40 clusters, i.e., $\alpha_0 = 0.67$; in this setting, IPW, regression, and DR estimates for the indirect ($\bar{IE}(0.6, \alpha_1)$), total effect ($\bar{TE}(0.6, \alpha_1)$), and overall effect ($\bar{OE}(\alpha_1)$) are given in Figure 4. The indirect effect has a decreasing trend when $\alpha_1 - \alpha_0$ increases, and it is positive when $\alpha_1 < 0.67$ and negative when $\alpha_1 > 0.67$. For example, when $\alpha_1 = 0.46$, the DR estimate is 15.14, and 95% confidence interval (CI) is (11.19, 19.76), which suggests that there would be 15.14 more tons of NO_x emissions in units without scrubbers installed within groups with 46% coverage compared to that with groups with 67% coverage of scrubber installation; when $\alpha_1 = 0.75$, the estimate is −6.84 for DR estimator, and the corresponding 95% CI is (-8.82, -5.40), which implies that we would expect 3.90 tons of NO_x emissions fewer among units without scrubbers installed within groups with 75% coverage compared to groups with average coverage probability. It is also worth noting that the confidence intervals would decrease as the difference between α_0 and α_1 decreases, which indicates that the decrease in the variance of the estimator of $\mu_{0\alpha}$ and $\mu_{1\alpha}$ cannot offset the increase in the correlation between the two estimators, because both the estimators are dependent on the treatment allocation function $\pi(a_{i(-j)}; \alpha)$. The total effect of the proposed estimators, which combine both the direct and the indirect effects, have a slightly decreasing trend when α increases. For example, when $\alpha = 0.5$, the IPW, regression, and DR estimates are −18.15, −25.70, and −16.91, and the corresponding 95% CIs are (−24.38, −14.43), (−30.27, −21.64), and (−18.78, −7.66), respectively; when $\alpha = 0.75$, the IPW, regression, and DR estimates are −25.56, −28.35, and −22.29, and the corresponding 95% CIs are (−29.44, −21.71), (−31.56, −24.85), and (−27.28, −19.83), respectively. Therefore, all three estimators indicate that there would be fewer NO_x emissions per unit with scrubbers within groups with higher coverage probability compared to the unit without scrubbers installed in groups with lower coverage probability. The estimates of the overall effect of scrubber installation, which quantify the difference in the tons of NO_x emissions under two treatment allocation strategies, suggest that there would be fewer NO_x emissions within groups with higher average percentage of the coverage of scrubber installation.

6 Discussion

In this article, we constructed three consistent estimators: IPW, regression, and DR estimators for four types of network causal effects: the direct, indirect, total, and overall effects when the confounders are missing not at random. Under the outcome-independent missingness assumption, the IPW and regression estimators are consistent and asymptotically normal if the corresponding models are correctly specified, and the consistency of the DR estimator requires that either the joint modelling of the propensity score of treatment and missingness or the outcome regression model, but not necessarily both, are correctly specified.

The designed doubly robust estimator is based on the conventional DR estimator proposed by Kang and Schafer [20] in the general setting without interference. In the setting when interference exists, the methodology avoids the assumption of SUTVA and can recover the causal effect in the whole population with observed data. In the real application, we compared the performance of three proposed estimators and showed that the intervention of installing scrubbers has a positive effect on reducing the emissions of NO_x which, in turn, may potentially reduce the ambient PM2.5 and the concentrations of ozone.

One limitation of the proposed methods is that it can cause increased computational burden when the number of units in each group increases, and it may not be suitable to implement the estimators when the confounders are high-dimensional. Adapting the proposed methods to handle high-dimensional data is an interesting direction for future research. Another limitation lies in that the estimators do not work well for the cases when the proportion of missingness is large. In this study, we only consider the setting under the partial interference only; In the scenarios with more general interference patterns, where the outcome of one unit can be affected by the confounders or outcomes of other units, relying solely on linear models for the regression estimator and regression component in the doubly robust estimator may not be sufficient. In such scenarios, alternative methods, such as deep neural networks, may be necessary to obtain consistent estimates of the average potential outcome. These methods can capture the complex dependencies and non-linear relationships that may arise in the presence of complex interference patterns. In this article, we assume a binary missingness indicator for each individual. As pointed out by reviewers, this assumption may not always reflect reality. We leave the investigation of more general missingness patterns for future research.

Funding information: Dr. Zhu's research was supported by Grant Number RGPIN-2017-04064 from the Natural Sciences and Engineering Research Council of Canada. Dr. Dubin's research was supported by Grant Number RGPIN-2020-04382 from the Natural Sciences and Engineering Research Council of Canada.

Author contributions: ZS proposed the methodology and conducted all simulation and data analyses, and did most of the writing of the article, while all authors (ZS, YZ, JAD) contributed modifications to the methods, suggestions for simulation runs and the data analysis, as well as proofread and edited the paper.

Conflict of interest: The author states no conflict of interest.

Data availability statement: The datasets analyzed during the current study are available at https://campd.epa.gov/data/custom-data-download.

References

[1] Rubin DB. Inference and missing data. Biometrika. 1976;63:581–92.

[2] Yang S, Wang L, Ding P. Causal inference with confounders missing not at random. Biometrika. 2019;106(4):875–88.

[3] Rubin DB. Randomization analysis of experimental data: The Fisher randomization test comment. J Amer Stat Assoc. 1980;75(371):591–3.

[4] Cox DR. Planning of experiments. New York: Wiley; 1958.

[5] Zigler C, Kim C, Choirat C, Hansen J, Wang Y, Hund L, et al. Causal inference methods for estimating long-term health effects of air quality regulations. Res Report (Health Effects Institute). 2016;187:5–49.

[6] Hyland A, Ambrose BK, Conway KP, Borek N, Lambert E, Carusi C, et al. Design and methods of the population assessment of Tobacco and health (PATH) study. Tobacco Control. 2017;26(4):371–8.

[7] Tchetgen Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. Stat. Methods Med Res. 2012;21:55–75.

[8] Hudgens MG, Halloran ME. Toward causal inference with interference. J Amer Stat Assoc. 2008;103:832–42.

[9] Liu L, Hudgens MG, Saul B, Clemens JD, Ali M, Emch ME. Doubly robust estimation in observational studies with partial interference. Stat. 2019;8(1):e214.

[10] Papadogeorgou G, Mealli F, Zigler CM. Causal inference with interfering units for cluster and population level treatment allocation programs. Biometrics. 2019;75(3):778–87.

[11] Sun Z, Liu L. Semiparametric inference of causal effect with nonignorable missing confounders. Stat Sinica. 2021;31:1–20.

[12] Perez-Heydrich C, Hudgens MG, Halloran ME, Clemens JD, Ali M, Emch ME. Assessing effects of cholera vaccination in the presence of interference. Biometrics. 2014;70(3):731–41.

[13] Ogburn EL, Sofrygin O, Diaz I, Van Der Laan MJ. Causal inference for social network data. 2017. arXiv: http://arXiv.org/abs/arXiv:170508527.

[14] Ding P, Geng Z. Identifiability of subgroup causal effects in randomized experiments with nonignorable missing covariates. Stat Med. 2014;33:1121–33.

[15] Miao W, Tchetgen Tchetgen EJ. On varieties of doubly robust estimators under missingness not at random with a shadow variable. Biometrika. 2016;103:475–82.

[16] Chen YH. A semiparametric odds ratio model for measuring association. Biometrics. 2007;63(2):413–21.

[17] van der Vaart A. Asymptotic statistics. Cambridge: Cambridge University Press; 1998.

[18] Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Software. 2015;67(1):1–48.

[19] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. https://www.R-project.org/.

[20] Kang JDY, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. Stat Sci. 2007;22:523–39.