

# Unifying instrumental variable and inverse probability weighting approaches for inference of causal treatment effect and unmeasured confounding in observational studies

Statistical Methods in Medical Research

2021, Vol. 30(3) 671–686

© The Author(s) 2020

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0962280220971835

[journals.sagepub.com/home/smm](https://journals.sagepub.com/home/smm)Tao Liu  and Joseph W Hogan 

## Abstract

Confounding is a major concern when using data from observational studies to infer the causal effect of a treatment. Instrumental variables, when available, have been used to construct bound estimates on population average treatment effects when outcomes are binary and unmeasured confounding exists. With continuous outcomes, meaningful bounds are more challenging to obtain because the domain of the outcome is unrestricted. In this paper, we propose to unify the instrumental variable and inverse probability weighting methods, together with suitable assumptions in the context of an observational study, to construct meaningful bounds on causal treatment effects. The contextual assumptions are imposed in terms of the potential outcomes that are partially identified by data. The inverse probability weighting component incorporates a sensitivity parameter to encode the effect of unmeasured confounding. The instrumental variable and inverse probability weighting methods are unified using the principal stratification. By solving the resulting system of estimating equations, we are able to quantify both the causal treatment effect and the sensitivity parameter (i.e. the degree of the unmeasured confounding). We demonstrate our method by analyzing data from the HIV Epidemiology Research Study.

## Keywords

Causal treatment effect, identification bounds, instrumental variable, inverse probability weighting, unmeasured confounding

## 1 Introduction

Observational studies offer an important alternative to randomized clinical trials when randomly assigning treatments to study subjects is unethical or practically impossible.<sup>1</sup> Analyzing data from such studies, however, confronts a difficulty that a direct comparison between the treated and untreated subjects does not necessarily reflect the causal effect of the treatment due to confounding.<sup>2</sup>

To control for the confounding effect, study investigators typically collect a rich set of covariates with the hope that these covariates capture all background differences between treated and untreated subjects. If all background differences between the comparison groups have been correctly measured, the confounding bias can be corrected by adjustments such as multivariable regressions, stratified analyses, propensity score matching, and inverse

---

Department of Biostatistics, Center for Statistical Sciences, Brown University, Providence, RI, USA

### Corresponding author:

Tao Liu, Department of Biostatistics, Center for Statistical Sciences, Brown University School of Public Health Providence, RI 02912, USA.

Email: [tliu@stat.brown.edu](mailto:tliu@stat.brown.edu)

probability weighting (IPW).<sup>3–11</sup> Absence of unmeasured confounding (i.e. ignorability) is a strong and untestable assumption, and often implausible in observational studies.

When ignorability cannot be assumed, estimating the causal effect of a treatment and sometimes quantifying the degree of unmeasured confounding are imperative. The former is often the primary research objective, while the latter becomes important when the robustness of analyses assuming ignorability needs to be objectively assessed, or when the impact of unmeasured confounding needs to be evaluated, e.g., for planning future studies in similar settings. These two objectives are typically not achievable in a single observational study, but possible given the existence of an instrumental variable (IV).

Instrumental variable methods can be traced back to 1920s,<sup>12,13</sup> and have been extensively implemented in econometric and recently in biomedical research. Loosely speaking, an IV can be envisioned as a “randomizer” which (1) varies independent of confounders, (2) has a causal effect on treatment received, but (3) has no direct effect on the outcome of interest. These three conditions are conventionally referred to as the exogeneity, monotonicity, and exclusion restriction assumptions, respectively.<sup>14</sup> An IV allows for drawing causal inference about treatment effect despite the existence of unmeasured confounding. However, without additional assumptions, the IV estimate of the treatment effect applies only to a non-identifiable subpopulation of those whose treatment can be changed by the IV.<sup>15–17</sup>

The population average treatment effect (ATE) is generally of broad interest in public health and epidemiology. To infer the ATE, IVs (if available) have been used to construct bound estimates in simple settings (e.g. when both treatment and outcome are binary).<sup>18–23</sup> In those settings, the uncertainty of unmeasured confounding effect is accounted for by a bound estimate instead of a point estimate. With continuous outcomes, obtaining bounds on ATE becomes a challenge because the domain of the outcome is typically unrestricted. In this case, additional properties of data may be implemented to construct contextually proper constraints on the observed and counterfactual data so as to identify meaningful bounds on the ATE. In this paper, we use the HIV Epidemiology Research Study (HERS)<sup>24,25</sup> as an example to explore such constraints. Our interest is to estimate the ATE of highly active antiretroviral therapy (HAART) on patients’ CD4+ T lymphocytes (CD4) count and to quantify the degree of unmeasured confounding in the HERS.

The HIV Epidemiology Research Study was conducted when the HAART first became available to HIV-infected patients. It was a cohort study and prescription of HAART to study participants was not random. One of investigators’ interests was the initial-stage causal effect of HAART on patients’ CD4 count, an immunological marker for immune system function and disease stage. The study had collected an extensive set of covariates, but like many observational studies unmeasured confounding might still exist<sup>26,27</sup> and its impact was unclear. To have a sense of unmeasured confounding, many HIV-positive individuals in the early HAART era were reluctant to initiate therapy due to the fear of adverse side effects and toxicity, and physicians at the time tended to prescribe HAART to patients with poor health condition, particularly those with low CD4 count. These prognostic factors were not fully measured and possibly confounded the HAART effect in a non-negligible way by affecting both treatment decisions and outcomes.<sup>25</sup>

In this paper, we describe an approach that unifies the IV and IPW methods to simultaneously quantify (1) the ATE and (2) the degree of unmeasured confounding. To account for *measured* confounding, we use the IPW method<sup>4</sup> to “restore” the balance on measured covariates between treated and untreated subjects. To capture the *unmeasured* confounding, a sensitivity parameter is incorporated into the IPW estimating equations using the approach of Robins et al.<sup>28</sup> The sensitivity parameter is defined as the systematic difference between the treated and untreated patients if hypothetically having these patients exposed to the same treatment condition, after the measured confounding has been balanced out. This sensitivity parameter has been previously used to conduct sensitivity analyses to assess the robustness of estimated causal treatment effects to unmeasured confounding.<sup>25,29</sup> In this paper, we assume that an IV is available. The HERS was conducted at two types of study sites: academic medical centers and community health clinics. This motivates us to consider using study site of the HERS as an instrument variable. Instead of conducting a sensitivity analysis, we take advantage of additional information provided by IV to estimate the sensitivity parameter. We propose to unify the IV and IPW estimating equations with a constraint imposed by the principal stratification<sup>30</sup> and contextually suitable assumptions. By solving the resulting system of estimating equations, we obtain causal bound estimates on both the ATE and the sensitivity parameter for unmeasured confounding.

The rest of the paper is organized as follows: More details about the HERS and motivations of using HERS site as an IV are provided in Section 2. Notations and models are elaborated in Section 3. In Section 4, we review the IV and IPW methods, and then introduce a unified system of estimating equations derived from them. In Section 5, we present three sets of constraints and assumptions in the context of HERS, and develop bounds on

the initial-stage ATE of HAART on CD4 count and bounds on the degree of unmeasured confounding. In Section 6, we analyze the HERS data, and finally in Section 7, we offer some points for discussion.

## 2 The HIV Epidemiology Research Study (HERS)

### 2.1 Study overview

The HERS was conducted from 1993 to 2001 to investigate the natural history of HIV progression in women. Details of the study have been reported previously.<sup>24</sup> The study enrolled a total of 871 HIV-infected women at four study sites: Detroit, Providence, Baltimore, and New York City. Clinical outcomes (e.g. CD4 count) of each participant were recorded about every six months since enrollment. Starting around 1996, HAART became the recommended treatment regimen for HIV infected people, especially for those with low CD4 counts.<sup>31</sup> In this paper, we used data extracted for 201 HERS participants who completed both their seventh and eighth visits. They also met the following two conditions: (a) They were HAART naive before their seventh visit, and (b) had a low CD4 count of less than 350 cells/mm<sup>3</sup> before their eighth visit which indicated having a deteriorating immune system. Some of them were prescribed HAART after the seventh visit. Their CD4 counts at the eight visit were used as the outcome. The study had collected a rich set of covariates, but unmeasured confounding might still exist.

Table 1 summarizes the key demographic and clinical characteristics of the 201 women. In brief, 46 women (23%) had initiated the HAART. Those receiving HAART had a higher CD4 count on average than those not on HAART, but this “as-received” treatment effect<sup>32</sup> was not statistically significant (standard normal  $z$  statistic = 0.58) and was certainly a biased estimate of HAART causal effect.

Ko et al.<sup>25</sup> analyzed the same data set and screened out several candidate confounders, which are listed in the upper panel of Table 1.]Notably, we found that patients receiving HAART were (a) more likely to be aware their HIV status and on antiretroviral medicines at their enrollment and at the previous visit; (b) presenting less HIV symptoms and less likely to be a drug user; (c) having higher viral loads (HIV-RNA) at their enrollment and at the

**Table 1.** Summary of patient demographic characteristics by HAART receipt status and study site.

	Received HAART?		Comparison Statistic
	Yes	No	
Number of patients ( $n$ )	46	155	–
Average CD4 counts (cell/mm <sup>3</sup> )	229 (19)	216 (11)	$Z = 0.58$
<i>Candidate confounders</i>			
Race:			
Black; white; others	46; 28; 26%	61; 15; 24%	$\chi^2_2 = 5.1$
ARV med. receipt rate			
At enrollment (%)	50% (7.4)	39% (3.9)	$Z = 1.2$
At previous visit (%)	74% (6.5)	57% (4.0)	$Z = 1.9$
Presence of HIV symptoms (%)	26% (6.5)	37% (3.9)	$Z = 1.2$
HIV RNA (log <sub>10</sub> copy/mm <sup>3</sup> )			
At enrollment (average)	3.2 (.15)	3.1 (.07)	$Z = .78$
At previous visit (average)	3.7 (.15)	3.4 (.09)	$Z = 1.5$
Intravenous drug use			
Recent (%)	22% (6.1)	.25 (.035)	$Z = .19$
Lifetime (%)	61% (7.2)	.63 (.039)	$Z = .04$
Aware of HIV status (%)	83% (5.6)	.81 (.032)	$Z = .08$
The HERS study site			
	Academic centers	Community clinics	
Number of patients ( $n$ )	93	108	–
HAART received, $n$ (%)	26; 28%	20; 18%	$Z = 1.4$
Average CD4 (cell/mm <sup>3</sup> )	230 (14)	210 (12)	$Z = 1.0$

Note: The numbers inside parentheses are standard errors.  $z$  stands for a standard normal test for comparing two sample means, and  $\chi^2_2$  for a chi-squared statistic for Pearson's chi-squared test. HAART: highly active antiretroviral therapy.

previous visit; and (d) consisting of relatively more white and less black. As pointed out earlier, other confounders could likely exist and were not fully captured by the study.

## 2.2 HERS study site as IV

The HERS was a multi-center study and designed to recruit participants from two types of study sites for increased study generalizability. The study sites in Detroit and Providence were academic medical centers, while the other two study sites in Baltimore and New York City were community health clinics. The two types of study sites differed in many aspects. For example, the HERS investigators have noted that the academic sites had higher referral rates to the HERS by physicians and study clinic nurses and had higher HAART uptake rates among their participants.<sup>24</sup> Generally speaking, compared with community health clinics, academic medical centers tended to involve more actively in research on cutting-edge therapies and innovative HIV treatments besides routine patient care. As a result, physicians at the academic medical centers were more likely to be aware of the latest breakthroughs on HIV treatment, and hence when HAART first became available, they were more likely to prescribe it to HIV patients.

These differences motivate us to consider using the type of HERS study site as an IV. Using different characteristics of hospitals or physicians as IVs has been explored in other studies.<sup>27,33,34</sup> As noted by authors of these studies, differences in health care facilities/giver can be a reasonable but not a perfect IV. In the following section, we formalize the assumptions that are needed to use HERS study site as an IV. Potential violations of these assumptions are pointed out and their impacts are discussed later in Section 7.

## 3 Notations and definitions

### 3.1 Notation

We use  $Z$  to denote an IV (in the HERS,  $Z=1$  if the study site is an academic medical center and  $=0$  a community health clinic) and  $A_z$  the potential treatment status that an individual would receive should  $Z$  be set to  $z$ . This notation implies that each individual has a pair of potential treatments  $(A_1, A_0)$ <sup>35,36</sup> that she would potentially receive if seen by doctors at the two types of study site, and the *actual* treatment received is  $A = A_Z = A_1Z + A_0(1 - Z)$ , where  $A=1$  means that the individual receives HAART, and 0 otherwise. With the Stable Unit Treatment Value Assumption,<sup>36</sup> we use  $Y_z(a)$  to denote the potential outcome for an individual should we hypothetically set the IV to  $z$  and her treatment to  $a$ . The *actual* outcome observed is therefore  $Y = Y_Z(A) = Y_Z(A_Z)$ . In the following, we denote all confounders by a vector  $\mathbf{X}$  and the measured confounders by  $\mathbf{V} \subseteq \mathbf{X}$ . The observed data consist of  $n$  identically and independently distributed copies of  $\{\mathbf{V}_i, Z_i, A_i, Y_i\}$ ,  $i = 1, 2, \dots, n$ .

We assume that the conventional IV assumptions – the *exogeneity*, *exclusion restriction*, and *monotonicity* assumptions<sup>15,16</sup> – are satisfied. The exclusion restriction implies that  $Y(a) \equiv Y_1(a) = Y_0(a)$ , i.e. the IV has no direct effect on the outcome beyond its impact on individual's treatment. The monotonicity assumption assumes that  $\Pr(A_1 \geq A_0) = 0$ , i.e. an individual who was not prescribed HAART at an academic medical center would not either at a community health clinic. The exogeneity assumes that  $Z$  is jointly independent of the potential outcomes and treatments,  $Z \perp (A_0, A_1, Y(0), Y(1))$ .

### 3.2 Definitions of causal treatment effect

Using potential outcomes, the causal effect of a treatment can be defined at different levels. The ATE =  $E\{Y(1) - Y(0)\}$  is the treatment effect defined over the entire population, which is the interest of this paper. With an IV, the local average treatment effect (LATE)<sup>15</sup> is defined as the average treatment effect among the subpopulation whose treatment assignment can be changed by the IV, i.e.  $\text{LATE} = E\{Y(1) - Y(0) | A_0 = 0, A_1 = 1\}$ . For a given IV, the LATE can be estimated consistently even with the presence of unmeasured confounding. However, a limitation of LATE estimates is that the subpopulation is not fully identified and the interpretation of the LATE depends on the choice of IV, which poses a significant drawback for generalizing results to the general population and to other settings.

The relationship between the ATE and LATE can be expressed using the principal stratification.<sup>30</sup> For a binary instrument and a binary treatment, the principal stratification suggests that the population can be partitioned into four mutually exclusive subpopulations based on the potential treatments each individual would have: In HERS,  $\mathcal{P}_{00} = \{(A_0, A_1) = (0, 0)\}$  is the subpopulation who would never receive HAART;  $\mathcal{P}_{01} = \{(A_0, A_1) = (0, 1)\}$  is the

subpopulation who would receive HAART only at academic medical centers;  $\mathcal{P}_{10} = \{(A_0, A_1) = (1, 0)\}$  is the subpopulation who would receive HAART only at community health clinics; and  $\mathcal{P}_{11} = \{(A_0, A_1) = (1, 1)\}$  is the subpopulation who would always receive HAART. The monotonicity assumption implies that the subpopulation  $\mathcal{P}_{10}$  is an empty set.

Let us denote the estimands of ATE and LATE by  $\beta^{\text{ATE}}$  and  $\beta^{\text{LATE}}$ , respectively. Then the relationship between ATE and LATE is expressed as

$$\beta^{\text{ATE}} = \pi_{01}\beta^{\text{LATE}} + \pi_{00}\{\mu_{00}(1) - \mu_{00}(0)\} + \pi_{11}\{\mu_{11}(1) - \mu_{11}(0)\} \quad (1)$$

where  $\pi_{jk} = \Pr(\mathcal{P}_{jk})$  and  $\mu_{jk}(a) = E\{Y(a)|\mathcal{P}_{jk}\}$ . We will use this relationship to unify the IPW and IV estimation methods.

## 4 Review of estimation methods

### 4.1 The IPW method

Putting aside the covariates for the moment, the potential outcomes can be rewritten using a marginal structural mean model<sup>6,37,38</sup>

$$E[Y(a)] = \beta_0 + \beta^{\text{ATE}}a, \quad a = 0, 1 \quad (2)$$

When unmeasured confounding is absent (i.e.  $Y(a) \perp A|\mathbf{V}$ ),  $\beta^{\text{ATE}}$  can be consistently estimated by the solution  $\hat{\beta}_{\text{IPW}}$  to the following IPW estimating equations<sup>4</sup>

$$U_1(\beta_{\text{IPW}}) := \sum_{i=1}^n (1, A_i)^\top W_{1i} (Y_i - \beta_1 - A_i \beta_{\text{IPW}}) = 0,$$

where  $W_{1i} = A_i/e(\mathbf{V}_i; \gamma) + (1 - A_i)/\{1 - e(\mathbf{V}_i; \gamma)\}$ , and  $e(\mathbf{V}; \gamma) = \Pr(A = 1|\mathbf{V})$  is the propensity score<sup>39</sup> with a  $l$ -dimension parameter  $\gamma$ . We assume that  $0 < e(\mathbf{V}; \gamma) < 1$  so that the estimating equations are well defined. This condition is referred to as the positivity assumption in the causal inference literature. It assumes that all individuals have positive probabilities of receiving the treatment and positive probabilities of not receiving the treatment. In other words, for any subpopulation on the support of  $\mathbf{V}$ , we have information available about the distributions of both  $Y(0)$  and  $Y(1)$ .

The IPW method has several properties that are worth mentioning. First, the efficiency of the resulting estimator can be improved by using stabilized weights to replace  $W_{1i}$ .<sup>40</sup> Second, the estimating equations can be augmented to achieve double robustness if we further specify an outcome regression model of  $Y$  on  $A$  and  $\mathbf{V}$ .<sup>8</sup> Finally, if  $\gamma$  is unknown,  $\hat{\beta}_{\text{IPW}}$  remains consistent when  $\gamma$  is replaced by a consistent estimator  $\hat{\gamma}$  that solves

$$U_2(\gamma) := \sum_{i=1}^n W_{2i} \{A_i - e(\mathbf{V}_i; \gamma)\} = 0$$

where  $W_{2i}$  is an appropriate weight function; for example,  $W_{2i} = \partial e(\mathbf{V}_i; \gamma)/\partial \gamma$  if a logistic model can describe the relationship between  $A$  and  $\mathbf{V}$ .

When unmeasured confounding exists,  $U_1(\beta_{\text{IPW}})$  is biased (i.e.  $E\{U_1(\beta_{\text{IPW}})\} \neq 0$ ). In this case, Robins et al.<sup>38</sup> proposed to introduce a sensitivity parameter  $\tau$  and estimate  $\beta^{\text{ATE}}$  by the solution  $\hat{\beta}_{\text{MIPW}}(\tau)$  to the following modified IPW estimating equations

$$U_3(\beta_{\text{MIPW}}, \tau) := \sum_{i=1}^n (1, A_i)^\top W_{1i} \{Y_i^* - \beta_2 - A_i \beta_{\text{MIPW}}\} = 0$$



where  $Y_i^* = Y_i - \tau\{A_i - e(\mathbf{V}_i; \gamma)\}$  is the “outcome” corrected for the selection bias due to unmeasured confounding. For binary treatments, the sensitivity parameter can be simplified as the contrast of the potential outcomes between the treated and untreated conditional on  $\mathbf{V}$ <sup>25</sup>

$$\tau = (a - a') [E\{Y(a)|A = a, \mathbf{V}\} - E\{Y(a)|A = a', \mathbf{V}\}]$$

with  $a = 1 - a'$ . In the context of the HERS,  $\tau > 0$  means that the HAART might be preferentially given to those with higher CD4 counterfactuals  $Y(a)$ ; while  $\tau < 0$  means the opposite; and when  $\tau = 0$ , no unmeasured confounding is implied and the resulting estimator  $\hat{\beta}_{\text{MIPW}}(0) = \hat{\beta}_{\text{IPW}}$ .

Without additional assumptions, the parameter  $\tau$  is not identified by the data. The resulting estimator  $\hat{\beta}_{\text{MIPW}}(\tau)$  is typically used to conduct a sensitivity analysis—estimate  $\beta^{\text{ATE}}$  using  $\hat{\beta}_{\text{MIPW}}(\tau)$  as if  $\tau$  is known and examine the sensitivity of  $\hat{\beta}_{\text{MIPW}}(\tau)$  to unmeasured confounding by varying the value of  $\tau$  over its plausible range.<sup>25,29</sup> In this paper, we assume that an IV is available. Instead of conducting a sensitivity analysis, we propose to use the extra information extracted by IV to quantify  $\tau$  (i.e. the degree of unmeasured confounding).

## 4.2 The IV method

The IV methods have been widely used in econometric research.<sup>41</sup> In the just-identified case with a single binary IV and a binary treatment, the standard IV estimating equations are

$$U_4(\beta_{\text{IV}}) := \sum_{i=1}^n (1, Z_i)^\top (Y_i - \beta_3 - \beta_{\text{IV}} A_i) = 0$$

When the IV assumptions given in Section 3.1 are satisfied, the solution

$$\hat{\beta}_{\text{IV}} = \frac{\overline{YZ}/\overline{Z} - \overline{Y(1-Z)}/\overline{(1-Z)}}{\overline{AZ}/\overline{Z} - \overline{A(1-Z)}/\overline{(1-Z)}} \quad (3)$$

is consistent for  $\beta^{\text{LATE}}$ .<sup>15,16,42</sup> An important property of the IV estimand is that  $\hat{\beta}_{\text{IV}}$  remains consistent despite the existence of unmeasured confounding.

Under the framework of the generalized method of moments, the IV estimating equations can be readily solved using the two-stage least squares method.<sup>14,43,44</sup> The IV estimating equations also can incorporate a weight matrix to allow for heteroskedastic or correlated residuals, and be generalized to deal with multiple IVs and non-continuous outcomes.<sup>41,45</sup>

## 5 A unified system of estimating equations

We propose to use principal stratification and the resulting constraint (1) to unify the IV and IPW methods. Specifically, we propose to jointly solve the following system of constrained estimation equations

$$(U_2(\gamma), U_3(\beta_{\text{MIPW}}, \tau), U_4(\beta_{\text{IV}}))^\top = 0 \quad (4)$$

and use the solution  $\hat{\beta}_{\text{MIPW}}$  to estimate the ATE and  $\hat{\tau}$  to estimate the degree of unmeasured confounding.

One problem of using the constraint (1) is that  $\mu_{11}(0)$  and  $\mu_{00}(1)$  in the equation are the averages of unobserved potential outcomes and hence not identified, while all other parameters are identified because  $\pi_{11} = E(A = 1|Z = 0)$ ,  $\pi_{00} = E(A = 0|Z = 1)$ ,  $\pi_{01} = 1 - \pi_{00} - \pi_{11}$ ,  $\mu_{11}(1) = E(Y|A = 1, Z = 0)$ , and  $\mu_{00}(0) = E(Y|A = 0, Z = 1)$ .<sup>16</sup> So to implement the above constrained estimating equations system, additional prior information of  $\mu_{11}(0)$  and  $\mu_{00}(1)$  is needed.

In this following, we explore and present three sets of assumptions in the context of the HERS. Each allows us to identify bounds on the ATE and unmeasured confounding parameter  $\tau$ . In Sections 5.1 to 5.3, we first assume that the sample size  $n$  is sufficiently large such that the sampling variation of the estimating equations (4) is ignored. Then in Sections 5.4 and 5.5, we discuss inferences on the sampling uncertainty of bound estimates for a finite  $n$ .

### 5.1 Assumption on the upper limits of $\mu_{11}(0)$ and $\mu_{00}(1)$

The outcome variable of our interest is CD4 count, so  $\mu_{00}(1)$  and  $\mu_{11}(0)$  must be greater than zero. In our first set of assumptions, we make a simple assumption that there exist two upper bounds that

**Assumption (A):**  $0 \leq \mu_{00}(1) \leq \xi_1$ ,  $0 \leq \mu_{11}(0) \leq \xi_0$ , with known  $\xi_0$  and  $\xi_1$ .

Assumption (A) leads to a simplified version of the Robins-Manski type bound on the ATE.<sup>18,19,23</sup> It is straightforward to show that with known  $\xi_0$  and  $\xi_1$ , the ATE falls within the interval

$$[b(\xi_0, 0), b(0, \xi_1)]$$

where to emphasize the unidentifiable parameters in equation (1), we define  $b(\mu_{11}(0), \mu_{00}(1)) = \pi_{01} \times \text{LATE} + \pi_{11}\{\mu_{11}(1) - \mu_{11}(0)\} + \pi_{00}\{\mu_{00}(1) - \mu_{00}(0)\}$

The bound on  $\tau$  can be inferred by finding the values of  $\tau$  such that the corresponding solutions to equation (4) are consistent with the above bound on ATE. It is straightforward to verify that for a given  $\beta^{\text{ATE}}$ , the solution of  $\tau$  is

$$\hat{\tau}_n(\beta^{\text{ATE}}, \gamma) = \frac{\overline{W_1 A} * \overline{W_1 (Y - \beta^{\text{ATE}} A)} - \overline{W_1} * \overline{W_1 A (Y - \beta^{\text{ATE}} A)}}{\overline{W_1 A} * \overline{W_1 (A - e(\mathbf{V}; \gamma))} - \overline{W_1} * \overline{W_1 A (A - e(\mathbf{V}; \gamma))}}$$

which is a non-increasing function of  $\beta^{\text{ATE}}$ . So the unmeasured confounding parameter  $\tau$  is bounded by

$$[\tau(b(0, \xi_1), \gamma), \tau(b(\xi_0, 0), \gamma)]$$

where  $\tau(\beta^{\text{ATE}}, \gamma) \equiv \hat{\tau}_\infty(\beta^{\text{ATE}}, \gamma)$ .

Assumption (A) alone is sufficient to identify the bounds on ATE and  $\tau$ , but the two upper limits  $\xi_0$  and  $\xi_1$  need to be sufficiently large, making the two bounds too wide to be of practical value. In practices, contextually plausible constraints, such as that the average treatment effect among  $\mathcal{P}_{11}$  (those always receiving treatment) is higher than other subpopulations, are often made to tighten the bounds.<sup>22,46-48</sup> That motivates us to consider making assumptions on the relative magnitudes between the unidentified and identified quantities.

### 5.2 Constraints on relationships between $\mu_{11}(0)$ and $\mu_{00}(1)$ and identified quantities

**Assumption (B):** We assume that

1. The average treatment effect among  $\mathcal{P}_{11}$  is no less than a known  $\delta_{11}$

$$E\{Y(1) - Y(0) | \mathcal{P}_{11}\} = \mu_{11}(1) - \mu_{11}(0) \geq \delta_{11}$$

A plausible choice for  $\delta_{11}$  is zero; that is, we assume that *on average*, the subpopulation  $\mathcal{P}_{11}$  (who would *always* receive HAART) would benefit from HAART. We make this assumption because during the HERS, HAART was becoming a recommended therapy particularly for HIV patients with low CD4 count. Further, we impose a known lower bound on the average treatment effect among  $\mathcal{P}_{00}$  that

$$E\{Y(1) - Y(0) | \mathcal{P}_{00}\} = \mu_{00}(1) - \mu_{00}(0) \geq \delta_{00}$$

A negative value of  $\delta_{00}$  implies that HAART can potentially be harmful for those who would never receive HAART at either site, while setting  $\delta_{00} = 0$  implies that HAART is also beneficial for them on average.

3. The difference on  $E\{Y(0)\}$  between  $\mathcal{P}_{11}$  and  $\mathcal{P}_{00}$  is bounded above by  $\delta_{y0}$

$$E\{Y(0) | \mathcal{P}_{11}\} - E\{Y(0) | \mathcal{P}_{00}\} = \mu_{11}(0) - \mu_{00}(0) \leq \delta_{y0}$$

In the HERS, it is sensible to set  $\delta_{y0} = 0$ . Intuitively, it means that in the untreated condition, people who would *always* receive HAART had higher degree of HIV progression (lower CD4 on average, compared to those who would *never* receive HAART).

1. The difference in average treatment effects between those who would *always* receive HAART and those who would *never* receive HAART is bounded below

$$E\{Y(1) - Y(0)|\mathcal{P}_{11}\} - E\{Y(1) - Y(0)|\mathcal{P}_{00}\} = \{\mu_{11}(1) - \mu_{11}(0)\} - \{\mu_{00}(1) - \mu_{00}(0)\} \geq \delta_{trt}$$

For example, letting  $\delta_{trt} = 0$  implies that the average treatment effect on those who would always receive HAART is greater than those would never do so.

Under this set of assumptions, it can be shown that the ATE is bounded by

$$[b(c_0, \mu_{00}(0) + \delta_{y0}), b(0, c_1)]$$

and  $\tau$  by

$$[\tau(b(0, c_1), \gamma), \tau(b(c_0, \mu_{00}(0) + \delta_{y0}), \gamma)]$$

where  $c_0 = \min\{\mu_{11}(1) - \delta_{11}, \mu_{00}(0) + \delta_{y0}\}$  and  $c_1 = \mu_{11}(1) + \mu_{00}(0) - \delta_{trt}$ .

### 5.3 Constraint conditional on measured covariates

Given the HERS data, it may be more realistic to assume that Assumption (B) holds conditional on the measured covariates  $\mathbf{V}$ . So we propose our third set of assumptions as

**Assumption (B')**: We assume that for known  $\delta_{11}$ ,  $\delta_{00}$ ,  $\delta_{y0}$  and  $\delta_{trt}$ .

1.  $E\{Y(1) - Y(0)|\mathcal{P}_{11}, \mathbf{V}\} \geq \delta_{11}$ ;  $E\{Y(1) - Y(0)|\mathcal{P}_{00}, \mathbf{V}\} \geq \delta_{00}$ .
2.  $E\{Y(0)|\mathcal{P}_{11}, \mathbf{V}\} - E\{Y(0)|\mathcal{P}_{00}, \mathbf{V}\} \leq \delta_{y0}$ .
3.  $E\{Y(1) - Y(0)|\mathbf{V}\} - E\{Y(1) - Y(0)|\mathcal{P}_{00}, \mathbf{V}\} \geq \delta_{trt}$ .
4. Further, we assume that the monotonicity and exclusion restriction assumptions hold conditional on  $\mathbf{V}$ . So the constraint equation (1) becomes

$$\begin{aligned} \beta^{ATE} = & \pi_{01}\beta^{LATE} + \int_{\mathbb{V}} E\{Y(1) - Y(0)|\mathcal{P}_{11}, \mathbf{V}\}P(\mathcal{P}_{11}|\mathbf{V})dF(\mathbf{V}) \\ & + \int_{\mathbb{V}} E\{Y(1) - Y(0)|\mathcal{P}_{00}, \mathbf{V}\}P(\mathcal{P}_{00}|\mathbf{V})dF(\mathbf{V}) \end{aligned} \quad (5)$$

where  $\mathbb{V}$  is the support of  $\mathbf{V}$  with a distribution  $F(\mathbf{V})$ . We write the product  $\pi_{01}\beta^{LATE}$  as before because both the LATE and  $\pi_{01}$  are identified by the data. Again, no observed data are available for  $E\{Y(0)|\mathcal{P}_{11}, \mathbf{V}\}$  and  $E\{Y(1)|\mathcal{P}_{00}, \mathbf{V}\}$  and therefore the two quantities are not identified. Henceforth, we denote them by  $\mu_{11}(0, \mathbf{V})$  and  $\mu_{00}(1, \mathbf{V})$ , respectively.

Under (B') and equation (5), it can be shown that a bound on ATE is

$$\left[ \pi_{01} \times \text{LATE} + \int_{\mathbb{V}} \tilde{b}\{c_0(\mathbf{V}), c_2(\mathbf{V})\}dF, \pi_{01} \times \text{LATE} + \int_{\mathbb{V}} \tilde{b}\{0, c_1(\mathbf{V})\}dF \right]$$

and a bound on  $\tau$  is

$$\left[ \tau(\pi_{01} \times \text{LATE} + \int_{\mathbb{V}} \tilde{b}\{0, c_1(\mathbf{V})\}dF, \gamma), \tau(\pi_{01} \times \text{LATE} + \int_{\mathbb{V}} \tilde{b}\{c_0(\mathbf{V}), c_2(\mathbf{V})\}dF, \gamma) \right]$$



where  $\tilde{b}(\mu_{11}(0, \mathbf{V}), \mu_{00}(1, \mathbf{V})) = [\mathbb{E}\{Y(1)|\mathcal{P}_{11}, \mathbf{V}\} - \mu_{11}(0, \mathbf{V})]\Pr(\mathcal{P}_{11}|\mathbf{V}) + [\mu_{00}(1, \mathbf{V}) - \mathbb{E}\{Y(0)|\mathcal{P}_{00}, \mathbf{V}\}]\Pr(\mathcal{P}_{00}|\mathbf{V})$ ,  $c_0(\mathbf{V}) = \min(\mathbb{E}\{Y(1)|\mathcal{P}_{11}, \mathbf{V}\} - \delta_{11}, \mathbb{E}\{Y(0)|\mathcal{P}_{00}, \mathbf{V}\} + \delta_{y0})$ ,  $c_1(\mathbf{V}) = \mathbb{E}\{Y(1)|\mathcal{P}_{11}, \mathbf{V}\} + \mathbb{E}\{Y(0)|\mathcal{P}_{00}, \mathbf{V}\} - \delta_{trt}$ , and  $c_2(\mathbf{V}) = \mathbb{E}\{Y(0)|\mathcal{P}_{00}, \mathbf{V}\} + \delta_{00}$ .

## 5.4 Inference from finite samples

With a finite sample size  $n$ , we can estimate the bounds on ATE and  $\tau$ , based on the results in Sections 5.1 to 5.3. We proceed by first estimating the identifiable parameters in the constraint (1). Specifically we assume two regression models

$$E(A|Z) = \text{logit}^{-1}(\eta(Z; \theta_1)) \text{ and } E(Y|A, Z) = \kappa(A, Z; \theta_2)$$

for some known functionals  $\eta(Z; \theta_1)$  and  $\kappa(A, Z; \theta_2)$ . For binary  $Z$  and  $A$ , we can use two saturated models and specify that  $\eta(Z; \theta_1) = \theta_{10} + \theta_{11}Z$  and  $\kappa(A, Z; \theta_2) = \theta_{20} + \theta_{21}Z + \theta_{23}A + \theta_{23}AZ$  with  $\theta_1 = (\theta_{10}, \theta_{11})^\top$  and  $\theta_2 = (\theta_{20}, \theta_{21}, \theta_{22}, \theta_{23})^\top$ . Here, a saturated additive model for  $\kappa(A, Z; \theta_2)$  is compatible with the structural model (2), which suggests that  $E(Y|A, Z)$  is linear in  $A$ ,  $Z$ , and  $AZ$ . The estimated regression parameter  $\hat{\theta}_1$  and  $\hat{\theta}_2$  can be obtained by solving

$$U_5(\theta_1, \theta_2) := \begin{pmatrix} \sum_{i=1}^n W_{3i}[A_i - \text{logit}^{-1}\{\eta(Z_i; \theta_1)\}] \\ \sum_{i=1}^n W_{4i}\{Y_i - \kappa(A, Z; \theta_2)\} \end{pmatrix} = 0$$

where  $W_{3i} = \partial \text{logit}^{-1}\{\eta(Z_i; \theta_1)\} / \partial \theta_1$  and  $W_{4i} = (1, Z_i, A_i, A_i Z_i)^\top$ . Based on the above regression models, we have  $\hat{\pi}_{11} = \text{logit}^{-1}\{\eta(0; \hat{\theta}_1)\}$ ,  $\hat{\pi}_{00} = 1 - \text{logit}^{-1}\{\eta(1; \hat{\theta}_1)\}$ ,  $\hat{\pi}_{01} = \text{logit}^{-1}\{\eta(1; \hat{\theta}_1)\} - \text{logit}^{-1}\{\eta(0; \hat{\theta}_1)\}$ ,  $\hat{\mu}_{11}(1) = \kappa(1, 0; \hat{\theta}_2)$ , and  $\hat{\mu}_{00}(0) = \kappa(0, 1; \hat{\theta}_2)$ .

For Assumptions (A) and (B), we then estimate the function  $b(\cdot)$  by  $\hat{b}(\mu_{11}(0), \mu_{00}(1)) = \hat{\pi}_{01}\hat{\beta}_{IV} + \hat{\pi}_{11}\{\hat{\mu}_{11}(1) - \mu_{11}(0)\} + \hat{\pi}_{00}\{\mu_{00}(1) - \hat{\mu}_{00}(0)\}$ . Further, estimate  $c_0(\cdot)$  by  $\hat{c}_0(\delta_{11}, \delta_{y0}) = \min\{(\hat{\mu}_{11}(1) - \delta_{11}), (\hat{\mu}_{00}(0) + \delta_{y0})\}$  and  $c_1(\cdot)$  by  $\hat{c}_1(\delta_{trt}) = \hat{\mu}_{11}(1) + \hat{\mu}_{00}(0) - \delta_{trt}$ . By substituting these estimates for their estimands, we obtain the bound estimates on the ATE and  $\tau$ , as shown in Table 2.

For Assumption (B'), we estimate the bounds on ATE and  $\tau$  using the following steps

Step 1. We assume two observed-data models conditional on  $\mathbf{V}$ ,  $E(A|Z, \mathbf{V}) = \text{logit}^{-1}\eta(Z, \mathbf{V}; \theta_3)$ , and  $E(Y|Z, A, \mathbf{V}) = \kappa(Z, A, \mathbf{V}; \theta_4)$  for known  $\eta(Z, \mathbf{V}; \theta_3)$  and  $\kappa(Z, A, \mathbf{V}; \theta_4)$ . For example, we can assume two linear models without interactions that  $\eta(Z, \mathbf{V}; \theta_3) = \theta_{30} + \theta_{31}Z + \theta_{32}\mathbf{V}$  with  $\theta_3 = (\theta_{30}, \theta_{31}, \theta_{32}^\top)^\top$  and  $\kappa(Z, A, \mathbf{V}; \theta_4) =$

**Table 2.** Bounds estimates on ATE and  $\tau$  under assumptions (A), (B) and (B').

	Parameter	Estimated bound
(A)	ATE	$[\hat{b}(\hat{\xi}_0, \mathbf{0}), \hat{b}(\mathbf{0}, \hat{\xi}_1)]$
	$\tau$	$[\hat{\tau}_n(\hat{b}(\mathbf{0}, \hat{\xi}_1), \hat{\gamma}), \hat{\tau}_n(\hat{b}(\hat{\xi}_0, \mathbf{0}), \hat{\gamma})]$
(B)	ATE	$[\hat{b}(\hat{c}_0, \hat{\mu}_{00}(\mathbf{0}) + \delta_{00}), \hat{b}(\mathbf{0}, \hat{c}_1)]$
	$\tau$	$[\hat{\tau}_n(\hat{b}(\mathbf{0}, \hat{c}_1), \hat{\gamma}), \hat{\tau}_n(\hat{b}(\hat{c}_0, \hat{\mu}_{00}(\mathbf{0}) + \delta_{00}), \hat{\gamma})]$
(B')	ATE	$\left[ \hat{\pi}_{01}\hat{\beta}_{IV} + \frac{\sum_i \hat{b}(\hat{c}_0(\mathbf{V}_i), \hat{c}_2(\mathbf{V}_i))}{n}, \hat{\pi}_{01}\hat{\beta}_{IV} + \frac{\sum_i \hat{b}(\mathbf{0}, \hat{c}_1(\mathbf{V}_i))}{n} \right]$
	$\tau$	$\left[ \hat{\tau}_n \left\{ \hat{\pi}_{01}\hat{\beta}_{IV} + \frac{\sum_i \hat{b}(\mathbf{0}, \hat{c}_1(\mathbf{V}_i))}{n} \right\}, \hat{\tau}_n \left\{ \hat{\pi}_{01}\hat{\beta}_{IV} + \frac{\sum_i \hat{b}(\hat{c}_0(\mathbf{V}_i), \hat{c}_2(\mathbf{V}_i))}{n} \right\} \right]$

ATE: average treatment effect.

$\theta_{40} + \theta_{41}Z + \theta_{42}A + \theta_{43}^\top \mathbf{V}$  with  $\theta_4 = (\theta_{40}, \theta_{41}, \theta_{42}, \theta_{43}^\top)^\top$ . Let  $\hat{\theta}_3$  and  $\hat{\theta}_4$  denote the corresponding estimated parameters.

Step 2. With the monotonicity assumption and exclusion restriction, we estimate that  $\hat{E}\{Y(0)|\mathcal{P}_{00}, \mathbf{V}\} = \hat{E}\{Y|Z=1, A=0, \mathbf{V}\} = \mu(1, 0, \mathbf{V}; \hat{\theta}_4)$ ,  $\hat{E}\{Y(1)|\mathcal{P}_{11}, \mathbf{V}\} = \mu(0, 1, \mathbf{V}; \hat{\theta}_4)$ ,  $\hat{\Pr}(\mathcal{P}_{11}|\mathbf{V}) = \pi(0, \mathbf{V}; \hat{\theta}_3)$ , and  $\hat{\Pr}(\mathcal{P}_{00}|\mathbf{V}) = 1 - \pi(1, \mathbf{V}; \hat{\theta}_3)$ . Then we estimate  $\hat{b}(\cdot)$ ,  $\hat{c}_0(\cdot)$ ,  $\hat{c}_1(\cdot)$ , and  $\hat{c}_2(\cdot)$  by bringing in these estimates.

Step 3. We estimate the distribution of  $F(\mathbf{V})$  in (5) by the empirical cumulative density function and integrals by empirical sums, e.g. estimate  $\int_{\mathbf{V}} b(c_0(\mathbf{V}), c_2(\mathbf{V}))dF$  by  $\sum_{i=1}^n \hat{b}(\hat{c}_0(\mathbf{V}_i), \hat{c}_2(\mathbf{V}_i))/n$

The resulting bound estimates on the ATE and  $\tau$  for Assumption (B') are shown in Table 2.

## 5.5 Uncertainty region for estimated bounds

With a finite sample size, we quantify the sampling uncertainty of estimated bounds using *uncertainty regions* (URs), which are defined as intervals that provide a  $(1 - \alpha)$  100% coverage probability on the true bounds. Unlike usual confidence intervals (CIs), a UR takes into account both the sampling variability and partial identifiability. We considered two types of URs in this paper: *point-wise* and *strong*  $(1 - \alpha)$  100% coverage URs.

Let  $(L, U)$  denote the true bound. A point-wise UR is defined as an interval  $(\hat{L}, \hat{U})$  that contains any particular value  $\varrho \in (L, U)$  with a probability of at least  $(1 - \alpha)$ , where  $\varrho$  is the parameter generating the data. If the  $\hat{L}$  and  $\hat{U}$  are consistent estimates and asymptotically normal, a large-sample  $(1 - \alpha)$ 100% point-wise UR is given as<sup>49</sup>

$$\text{UR}_{\text{P-CAN}} = [\hat{L} - c^* \text{se}(\hat{L}), \hat{U} + c^* \text{se}(\hat{U})]$$

where  $\text{se}(\cdot)$  is the standard error and  $c^*$  is a critical value. When  $(U - L)$  is large compared to  $\text{se}(\hat{L})$  and  $\text{se}(\hat{U})$ ,  $c^*$  can be approximated by  $\Phi^{-1}(1 - \alpha)$  with  $\Phi$  being the normal distribution function.

A strong UR is an interval that contains the entire true bound  $(L, U)$  with a probability of at least  $(1 - \alpha)$ .<sup>49,50</sup> If both  $\hat{L}$  and  $\hat{U}$  are consistent and approximately normal, a strong  $(1 - \alpha)$ 100% UR is

$$\text{UR}_{\text{S-CAN}} = [\hat{L} - c \text{se}(\hat{L}), \hat{U} + c \text{se}(\hat{U})]$$

with  $c = \Phi^{-1}(1 - \alpha/2)$ .

Without assuming  $\hat{L}$  and  $\hat{U}$  to be normally distributed, we also can obtain a strong  $(1 - \alpha)$  UR using the bootstrap method.<sup>51</sup> Specifically, let  $(\tilde{L}^*, \tilde{U}^*)$  denote the estimated bound from a bootstrapped sample. A bootstrap strong 95% UR can be defined as the interval  $(L^*, U^*)$  that has  $\Pr^*(L^* \leq \tilde{L}^*, \tilde{U}^* \leq U^*) = 1 - \alpha$  with  $\Pr^*(\tilde{L}^* < L^*) = \Pr^*(\tilde{U}^* > U^*)$ , where  $\Pr^*$  is the empirical probability function induced by bootstrapped resamples. So a bootstrap strong UR, henceforth denoted by  $\text{UR}_{\text{S-BTS}} = (L^*, U^*)$ , can be obtained by finding the shortest interval that satisfies  $\frac{\#\{L^* \leq \tilde{L}^* < \tilde{U}^* \leq U^*\}}{K} \geq 1 - \alpha$ , and  $\frac{\#\{L^* \leq \tilde{L}^*\}}{K} \simeq \frac{\#\{U^* \geq \tilde{U}^*\}}{K}$ , where  $\#(\cdot)$  counts the number of statements that hold and  $K$  is the number of bootstrap resamples.

## 6 Application to the HERS data

### 6.1 Preliminary analyses

The upper panel of Table 3 shows the “as-treated” (AT) effect of initial-stage HAART on CD4 count, the IPW estimate of the ATE, and IV estimate of the LATE. The AT effect is estimated by the contrast of the average CD4 counts between those actually receiving HAART and those not. The IPW uses the variables listed in Table 1 as the measured confounders  $\mathbf{V}$  and assumes that  $e(\mathbf{V}; \gamma) = \text{logit}^{-1}(\gamma^\top \mathbf{V})$ .

The IPW estimate suggests that at initial stage, HAART could boost patient's CD4 count by an average of 27 cells/mm<sup>3</sup> among all patients with a 95% CI = (−16, 70) cells/mm<sup>3</sup>. The IPW estimate is higher than the AT estimate but the difference is not statistically significant. The difference between the IPW and AT estimates can be regarded as the bias of AT estimate that is attributable to the *measured* confounders. The IV estimate of the LATE suggests that HAART could increase CD4 count by 207 cells/mm<sup>3</sup> on average with a 95% CI of (−250, 664) cells/mm<sup>3</sup> among those who would receive HAART at academic medical centers but not at community health clinics.

**Table 3.** Estimates of HAART treatment effect on CD4 and  $\tau$ .

		ATE	$\tau$
AT	Point estimate 95% CI	13 (−30, 56)	—
IPW	Point estimate 95% CI	27 (−16, 70)	—
IV	Point estimate 95% CI	207 (−250, 664)	—
Assumption: (A):	Bound estimate	(−195, 256)	(−229, 223)
	95% UR <sub>P-CAN</sub>	<b>(−229, 289)</b>	<b>(−269, 266)</b>
	95% UR <sub>S-CAN</sub>	<b>(−235, 295)</b>	<b>(−277, 274)</b>
	95% UR <sub>S-BTS</sub>	<b>(−233, 294)</b>	<b>(−274, 273)</b>
Assumption: (B):	Bound estimate	(20, 231)	(−204, 7.5)
	95% UR <sub>P-CAN</sub>	<b>(−9, 280)</b>	<b>(−260, 49)</b>
	95% UR <sub>S-CAN</sub>	<b>(−15, 289)</b>	<b>(−271, 57)</b>
	95% UR <sub>S-BTS</sub>	<b>(−14, 285)</b>	<b>(−270, 57)</b>
Assumption: (B′):	Bound estimate	(18, 218)	(−191, 9.1)
	95% UR <sub>P-CAN</sub>	<b>(−10, 261)</b>	<b>(−234, 48)</b>
	95% UR <sub>S-CAN</sub>	<b>(−16, 270)</b>	<b>(−243, 56)</b>
	95% UR <sub>S-BTS</sub>	<b>(−14, 270)</b>	<b>(−243, 56)</b>

Note: We assume that  $\xi_0 = \xi_1 = 500$  for Assumption (A); and that  $\delta_{00} = \delta_{11} = \delta_{y0} = \delta_{irt} = 0$  for Assumptions (B) and (B′). Uncertainty regions are highlighted using bold font. HAART: highly active antiretroviral therapy; ATE: average treatment effect.

The IV estimate is not subject to the impact of unmeasured confounding but applies only to an unidentified subpopulation.

## 6.2 Bounds on HAART treatment effect and unmeasured confounding

We then estimate the initial-stage HAART treatment effect and the degree of unmeasured confounding using our proposed method. For Assumption (A), we let the upper limits  $\xi_0 = \xi_1 = 500$ . (Recall that the two limits are on the expected values of  $Y(0)$  among  $\mathcal{P}_{11}$  and  $Y(1)$  among  $\mathcal{P}_{00}$ ). We choose the two limits based on the facts that the average CD4 count at the previous visit was lower than 350 cells/mm<sup>3</sup> and at “current” (the eighth) visit, the average CD4 count was 229 cells/mm<sup>3</sup> for those treated and 216 cells/mm<sup>3</sup> for those untreated (refer to Tables 1). For Assumptions (B) and (B′), we let  $\delta_{11} = \delta_{00} = \delta_{y0} = \delta_{irt} = 0$ . The implications of choosing these values have been discussed in Section 5.2.

The lower panel of Table 3 summarizes the bound estimates of the ATE and  $\tau$ . The bound estimate of the ATE under Assumption (A) is (−196, 256) which is not informative compared with those under Assumption (B) (20, 231) and (B′) (18, 218). Assumption (B′) is not necessarily a stronger assumption than (B), but by imposing observed data models and having the estimated bounds smoothed over covariates, (B′) leads to slightly tighter bounds than (B) but their difference is negligible. The difference between the IPW (point) estimate and the bound estimates under (B) and (B′) is likely due to the bias of unmeasured confounding. Noticeably, the IPW estimate is close to the lower bound estimates under Assumptions (B) and (B′), suggesting that accounting only for measured confounders may not be adequate.

To obtain uncertainty regions on these bound estimates, we draw  $K = 1000$  bootstrap resamples, fixing the number of patients at the two types of study sites. Because the bounds estimates contain  $\min(\cdot)$  operation which complicates the derivation of their standard errors, we use the  $K$  bootstrapped resamples to calculate the standard errors of the two ends of bounds.

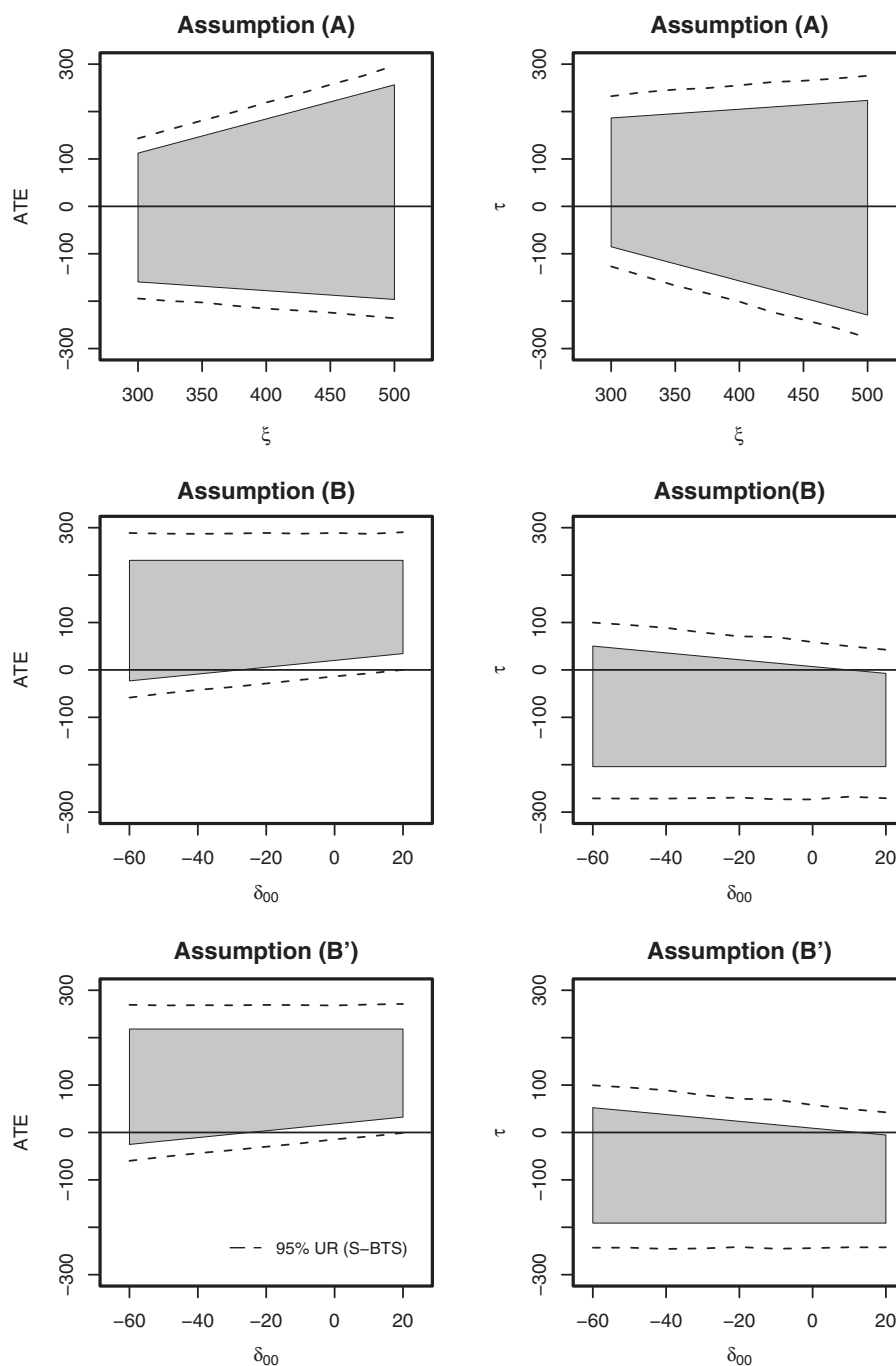
Table 3 summarizes the point-wise, strong, and bootstrap strong 95% coverage URs. The URs under Assumptions (B) and (B′) have comparable lower limits to that of the IPW 95% CI, while the difference in their upper limits indicates that unmeasured confounding might cause a downward bias and the true ATE is likely to be higher than what the IPW 95% CI suggests.

The bound estimates on  $\tau$  under Assumptions (B) and (B′) are (−204, 7.5) and (−191, 9.1), which are tighter and more informative than the bound estimate under Assumption (A) (−229, 223). A possibly negative value of  $\tau$ , as suggested by the 95% URs, implies that unmeasured factors might cause a selection bias in a way that resulted in preferential prescriptions of HAART to those with fewer CD4 count. The degree of unmeasured confounding,

defined as the adjusted difference in CD4 between those on HAART (if left untreated) and those not, is estimated to be roughly between  $-250$  and  $50$ .

### 6.3 Sensitivity to unknown parameters

In this section, we conduct a simple sensitivity analysis for the unknown parameters used in the three sets of assumptions. We impose a common upper limit  $\xi = \xi_0 = \xi_1$  for Assumption (A) and let  $\xi$  vary from 300 to 500. The bound estimates on the ATE and  $\tau$  along with bootstrap strong 95% URs are shown in Figure 1 (first row).



**Figure 1.** Sensitivities of bound estimates to  $\xi = \xi_0 = \xi_1$  under assumption (A); to  $\delta_{00}$  under assumptions (B) and (B'). The gray zones show the bound estimates as a function of  $\xi$  or  $\delta_{00}$ . The bootstrap strong 95% UR<sub>S-BTS</sub>'s are shown as dashed lines.

As suggested by the figure,  $\xi$  has more influence on the upper (lower) bound estimate on ATE (on  $\tau$ ) for the considered range of  $\xi$ , and the resulting bound estimates remain wide and non-informative.

For (B) and (B'), we let  $\delta_{00}$  (the lower limit of treatment effect among  $\mathcal{P}_{00}$ ) range from  $-60$  to  $20$  and fix  $\delta_{11} = \delta_{y0} = \delta_{trt} = 0$ . (More sophisticated sensitivity analyses that jointly evaluate  $\delta_{11}$ ,  $\delta_{00}$ ,  $\delta_{y0}$  and  $\delta_{trt}$  are possible.) We choose this range for  $\delta_{00}$  based on the magnitude of the AT and IPW estimates, and have it tilt toward the negative side for the possibility that HAART could be harmful for those never receiving HAART. Figure 1 (second and third rows) shows that  $\delta_{00}$  only affects the lower (upper) bound estimates of the ATE (of  $\tau$ ). The estimated ATE can be as high as over  $200 \text{ cell/mm}^3$ , and the lower bound of ATE varies around zero depending on the value of  $\delta_{00}$ . Again, a possible negative value of  $\tau$  suggests that unmeasured confounding likely caused HAART to be preferentially prescribed to those with poorer health.

## 7 Discussions

We propose to use an IV and sets of contextually plausible assumptions to quantify the population causal effect of a treatment as well as the degree of unmeasured confounding. We describe three sets of assumptions that are suitable in an observational study (the HERS). Assumption (A) specifies the limits of the expected unobservable potential outcomes, which leads to a simplified version of the Robins-Manski bounds on ATE. Assumptions (B) and (B') specify the relative magnitudes between identified and unidentified potential outcome averages. The bound estimates of ATE obtained under (B) and (B') are tighter than that under (A); have less concern of having unmeasured confounding bias compared with the IPW estimate; and are more informative than the IV estimate (if we look at the CI of the IV estimate). The bound estimates on the ATE and  $\tau$  reveal that unmeasured confounding could cause a downward bias on the ATE because of HAART being preferentially prescribed to those with poorer health condition.

Quantifying the degree of unmeasured confounding can be valuable for analysis of studies conducted in similar settings but having no IV. Several HIV observational studies<sup>37</sup> have been conducted contemporarily as the HERS, and could suffer from unmeasured confounding as well. In those studies when unmeasured confounding is of concern, analyses should be complemented with a sensitivity analyses as described in Section 6.3, where a plausible range for  $\tau$  can be informed from our study.

In this paper, we use the type of study site as an instrument variable, assuming that two crucial IV assumptions (monotonicity and exclusion restriction) are satisfied. The observed HAART assignment rate at academic centers is higher than that at community clinics, an observation suggesting that the deterministic monotonicity  $\Pr(A_1 \geq A_0) = 1$  is plausible, but the assumption is not verified. As one limitation of our study, this assumption will be violated if some individuals would receive HAART at community clinics but not at academic medical centers. If the proportion of these individuals ( $\mathcal{P}_{10}$ ) is small, it is reasonable to believe that the bias due to the violation of the monotonicity assumption is probably negligible. Alternatively, one can assume that  $\mathcal{P}_{00}$  is absent, so that  $\mathcal{P}_{11}$ ,  $\mathcal{P}_{01}$ , and  $\mathcal{P}_{10}$  form a partition of the population. This assumption allows everyone to have some chance of receiving HIV therapy, which is also sensible for the HERS because these patients' CD4 counts are less than  $350 \text{ cells/mm}^3$  six months before, and allows for the possibility that some people would potentially be treated at a community clinic but not an academic medical center. With this assumption, the proportions of  $\mathcal{P}_{11}$ ,  $\mathcal{P}_{01}$  and  $\mathcal{P}_{10}$  are identified because  $\pi_{01} = \Pr(A = 0|Z = 0)$ ,  $\Pr(\mathcal{P}_{10}) = \Pr(A = 0|Z = 1)$ , and  $\Pr(\mathcal{P}_{11}) = 1 - \pi_{01} - \Pr(\mathcal{P}_{10})$ . The following estimands are also identified:  $E\{Y(0)|\mathcal{P}_{01}\} = E(Y|A = 0, Z = 0)$ ,  $E\{Y(0)|\mathcal{P}_{10}\} = E(Y|A = 0, Z = 1)$ ,  $E\{Y(1)|\mathcal{P}_{11} \text{ or } \mathcal{P}_{01}\} = E(Y|A = 1, Z = 1)$ , and  $E\{Y(1)|\mathcal{P}_{11} \text{ or } \mathcal{P}_{10}\} = E(Y|A = 1, Z = 0)$ . A challenge here is how to incorporate the IV estimator, which now has an estimand as a “weighted” contrast of the average treatment effects between  $\mathcal{P}_{01}$  and  $\mathcal{P}_{10}$ , to construct constraint similar to equation (1). This is worth further investigation.

Moreover, replacing the deterministic monotonicity with a stochastic monotonicity<sup>52,53</sup> assumption deserves some explorations. Roy et al.<sup>54</sup> assumed  $\Pr(A_1 = 1|A_0 = 1, \mathbf{V}) \geq \Pr(A_1 = 1|A_0 = 0, \mathbf{V})$ , and proposed to use auxiliary covariates to estimate the memberships of principal strata. Small et al.<sup>55</sup> assumed  $\Pr(A_1 = 1|U) \geq \Pr(A_0 = 1|U)$  with  $U$  being a latent variable satisfying certain conditions. These stochastic monotonicity assumptions allow the possible presence of “defiers” and are generally a more plausible condition than the deterministic monotonicity. In the HERS study, given that the physicians at the academic centers were more likely to prescribe HAART to patients when it first became available, we assumed that the fraction of “defiers” was small and the deterministic monotonicity was a plausible condition in the context.

The exclusion restriction could also be violated if the type of study site  $Z$  remains associated with the outcome  $Y$  after accounting for the effect of  $Z$  on HAART receipt. A weaker exclusion restriction assumption can be made,

if the association between the instrument and the outcome can be removed after conditioning on some measured covariate  $V^* \subseteq V$ , i.e.  $\{Y(1), Y(0)\} \perp Z | V^*$ . In this case, various methods<sup>14,22,56–58</sup> can be implemented for an IV analysis conditional on  $V^*$ , and our method for bound estimation on ATE and  $\tau$  still applies.

There are several ways to account for the measured confounding. We use the method of inverse probability weighting by specifying a propensity score model. Alternatively, we can specify both an outcome regression model and a propensity score model and use the doubly robust (DR) estimator<sup>8</sup> to estimate the ATE. We do not implement the DR estimator in this paper because when unmeasured confounding exists, the DR estimator is no longer guaranteed to be consistent for ATE and could suffer more bias than other estimators. The simulations of Kang et al.<sup>10</sup> suggest that IPW is relatively robust to the impact of unmeasured confounding in terms of estimation bias. Because the focus issue of this paper is unmeasured confounding, we use the IPW for estimating ATE.

Finally, it should be pointed out that bound estimates may not be normally distributed asymptotically, especially when a bound occurs at the boundary of the parameter space or when the likelihood is not smooth around their true values. So practically, data analysts should check these regularity conditions in a similar way as they do when conducting statistical inference with other methods.

### Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was facilitated by the Providence/Boston Center for AIDS Research (P30AI042853).

### ORCID iDs

Tao Liu  <https://orcid.org/0000-0002-5274-4445>

Joseph W Hogan  <https://orcid.org/0000-0001-7959-7361>

### References

1. Rosenbaum PR. *Observational studies*. New York, NY: Springer, 2002.
2. VanderWeele TJ and Shpitser I. On the definition of a confounder. *Ann Stat* 2013; **41**: 196–220.
3. Rosenbaum PR and Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984; **79**: 516–524.
4. Robins JM, Rotnitzky A and Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc* 1994; **89**: 846–866.
5. D’Agostino RB. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; **17**: 2265–2281.
6. Robins JM, Hernán MA and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**: 550–560.
7. Hogan JW and Lancaster T. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. *Stat Meth Med Res* 2004; **13**: 17–48.
8. Bang H and Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics* 2005; **61**: 962–972.
9. Cole SR, Hernán MA, Margolick JB, et al. Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count. *Am J Epidemiol* 2005; **162**: 471–478.
10. Kang JDY and Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci* 2007; **22**: 523–539.
11. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci* 2010; **25**: 1–21.
12. Wright S. *Appendix to the tariff on animal and vegetable oils*. vol. **26**. New York, NY: Macmillan; 1928.
13. Stock JH and Trebbi F. Retrospectives: who invented instrumental variable regression? *J Econ Perspect* 2003; **17**: 177–194.
14. Angrist JD and Imbens GW. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *J Am Stat Assoc* 1995; **90**: 431–442.
15. Imbens GW and Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994; **62**: 467–475.



16. Angrist JD, Imbens GW and Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996; **91**: 444–455.
17. Robins JM and Greenland S. Identification of causal effects using instrumental variables: comment. *J Am Stat Assoc* 1996; **91**: 456–458.
18. Robins JM. The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies. In Sechrest L, Freeman H and Bailey A (eds) *Health Service Research Methodology: A Focus on AIDS*. Washington, DC: U. S. Public Health Service, 1898, pp. 113–159.
19. Manski CF. Nonparametric bounds on treatment effects. *Am Econ Rev* 1990; **80**: 319–323.
20. Balke A and Pearl J. Bounds on treatment effects from studies with imperfect compliance. *J Am Stat Assoc* 1997; **92**: 1171–1176.
21. Joffe MM. Using information on realized effects to determine prospective causal effects. *J R Stat Soc Ser B* 2001; **63**: 759–774.
22. Cheng J and Small DS. Bounds on causal effects in three-arm trials with non-compliance. *J R Stat Soc Ser B* 2006; **68**: 815–836.
23. Zhang JL and Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by death. *J Edu Behav Stat* 2003; **28**: 353–368.
24. Smith DK, Warren D, Vlahov P, et al. Design and baseline participant characteristics of the Human Immunodeficiency Virus Epidemiology Research (HER) Study: a prospective cohort study of human immunodeficiency virus infection in US women. *Am J Epidemiol* 1997; **146**: 459–469.
25. Ko H, Hogan JW and Mayer KH. Estimating causal treatment effects from longitudinal {HIV} natural history studies using marginal structural models. *Biometrics* 2003; **59**: 152–162.
26. Walker AM. Confounding by indication. *Epidemiology* 1996; **7**: 335–336.
27. Brookhart MA and Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *Int J Biostat* 2007; **3**: 14.
28. Robins JM, Rotnitzky A and Scharfstein DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. *Stat Model Epidemiol: Environ Clin Trials* 1999; **116**: 1–92.
29. Brumback BA, Hernán MA, Haneuse SJPA, et al. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med* 2004; **23**: 749–767.
30. Frangakis CE and Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**: 21–29.
31. Carpenter CCJ, Cooper DA, Fischl JM, et al. Antiretroviral therapy in adults: updated recommendations of the International AIDS Society-USA Panel. *J Am Med Assoc* 2000; **283**: 381–391.
32. Ten Have TR, Normand SLT, Marcus SM, et al. Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psych Ann* 2008; **38**: 772.
33. Johnston SC. Combining ecological and individual variables to reduce confounding by indication. *J Clin Epidemiol* 2000; **53**: 1236–1241.
34. Brookhart MA, Wang PS, Solomon DH, et al. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; **17**: 268–275.
35. Neyman J. On the application of probability theory to agricultural experiments: essay on principles. *Stat Sci* 1923; **1990**: 465–472.
36. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974; **66**: 668–701.
37. Gange SJ, Kitahata MM, Saag MS, et al. Cohort profile: The North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD). *Int J Epidemiol* 2007; **36**: 294–301.
38. Robins JM. Association, causation, and marginal structural models. *Synthese* 1999; **121**: 151–179.
39. Rosenbaum PR and Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**: 41–55.
40. Miguel HA, Babette B and Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J Am Stat Assoc* 2001; **96**: 440–448.
41. Wooldridge JM. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press, 2002.
42. Hernan MA and Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; **17**: 360–372.
43. Davidson R and MacKinnon J. *Estimation and inference in econometrics*. Oxford, UK: Oxford University Press, 1993.
44. Stock JH. *Instrumental variables in statistics and econometrics*. Amsterdam: Elsevier, 2001.
45. Freedman D. *Statistical models: theory and practice*. Cambridge, UK: Cambridge University Press, 2009.
46. Bhattacharya J, Shaikh AM and Vytlacil E. Treatment effect bounds under monotonicity assumptions: an application to Swan-Ganz Catheterization. *Am Econ Review* 2008; **98**: 351–356.
47. Vansteelandt S, Bowden J, Babanezhad M, et al. On instrumental variables estimation of causal odds ratios. *Stat Sci* 2011; **26**: 403–422.

48. Siddique Z. Partially identified treatment effects under imperfect compliance: the case of domestic violence. *J Am Stat Assoc* 2014; 108: 504–513.
49. Vansteelandt S, Goetghebeurand E, Kenward MG, et al. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Stat Sinica* 2006; 16: 953–979.
50. Horowitz JL and Manski CF. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J Am Stat Assoc* 2000; 95: 77–84.
51. Bickel PJ and Freeman DA. Some asymptotic theory for the bootstrap. *Ann Stat* 1981; 9: 1196–1217.
52. DiNardo J and Lee DS. Program evaluation and research designs. In: Ashenfelter O and Card D (eds) *Handbook of labor economics*. vol. 4. Amsterdam, the Netherlands: Elsevier, 2011, pp.463–536.
53. de Chaisemartin C. Tolerating defiance? Local average treatment effects without monotonicity. *Quantitative Economics* 2017; 8: 367–396.
54. Roy J, Hogan JW and Marcus BH. Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* 2008; 9: 277–289.
55. Small DS, Tan Z, Ramsahai RR, et al. Instrumental variable estimation with a stochastic monotonicity assumption. *Stat Sci* 2017; 32: 561–579.
56. Hirano K, Imbens GW, Rubin DB, et al. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics (Oxford, England)* 2000; 1: 69–88.
57. Abadie A. Semiparametric instrumental variable estimation of treatment response models. *J Econom* 2003; 113: 231–263.
58. Tan Z. Regression and weighting methods for causal inference using instrumental variables. *J Am Stat Assoc* 2006; 101: 1607–1618.