Inference about ATE from Observational Studies with Continuous Outcome and Unmeasured Confounding

TAO LIU*, JOSEPH W. HOGAN

Department of Biostatistics, Center for Statistical Sciences,
Brown University School of Public Health,
Box GS-121-7, Providence, RI 02912, U.S.A.
*tliu@stat.brown.edu

March 26, 2013

Abstract

For settings with a binary treatment and a binary outcome, instrumental variables can be used to construct bounds on a causal treatment effect. With continuous outcomes, meaningful bounds are more difficult to obtain because the domain of the outcome is typically unrestricted. In this paper, we combine an instrumental variable and subjective assumptions in the context of an observational cohort study of HIV-infected women to construct meaningful bounds on the initial-stage causal effect of antiretroviral therapy on CD4 count. The subjective assumptions are encoded in terms of the potential outcomes that are identified by observed data as well as a sensitivity parameter that captures the impact of unmeasured confounding. Measured confounding is adjusted using the method of inverse probability weighting (IPW). With extra information from an IV, we quantify both the causal treatment effect and the degree of the unmeasured confounding. We demonstrate our method by analyzing data from the HIV Epidemiology Research Study.

1 Introduction

Observational studies offer an important alternative to randomized clinical trials when assigning a treatment to study subjects is unethical or practically impossible (Rosenbaum 2002). However analyzing data from such studies often confronts the difficulty that a direct comparison between the treated and untreated subjects does not necessarily show the causal effect of the treatment due to confounding. Informally, confounding is caused by factors that have causal effects on both treatment and outcome (see VanderWeele and Shpitser 2011, for more rigorous discussions and definitions of confounding and confounders). To adjust for the confounding effect, observational studies typically include collecting a set of covariates with the hope that most confounders if not all are measured. Statistical methods for controlling for the confounding effect under the assumption of no unmeasured confounding include multivariate adjustment via regression models, propensity score risk adjustment, propensity score matching, and inverse probability weighting (IPW) (Bang and Robins 2005; Kang and Schafer 2007; Robins et al. 1994; Rosenbaum and Rubin 1984; Hogan and Lancaster 2004; D'Agostino 1998; Robins et al. 2000, among others).

The assumption of no unmeasured confounding is untestable in general and very often implausible. In this case, making causal inference on the treatment effect as well as obtaining a quantitative measure about the degree of unmeasured confounding is imperative. The two objectives are generally not achievable in a single observational study, but possible given the existence of an instrumental variable (IV).

The IV methods can be traced back to 1920's (Stock and Trebbi 2003; Wright 1928), and have been extensively implemented in econometric and recent bio-medical research. Loosely speaking, an IV can be envisioned as a 'randomizer' which varies exogenously, has a causal effect on treatment received, but has no causal effect on the outcome except through treatment. By convention, these three conditions are referred to as the exogeneity, monotonicity, and exclusion restriction assumptions, respectively (Angrist et al. 1996). When a valid IV exists, it can be used to draw inference about the causal treatment effect, despite the existence of unmeasured confounding. However, the IV estimate of the treatment effect applies only to a specific non-identifiable subpopulation unless additional assumptions are made (Angrist et al. 1996; Imbens and Angrist 1994). In simple settings with a binary treatment and a binary outcome, IV methods also can be used to construct bounds on a population causal treatment effect (Robins 1989; Manski 1990; Joffe 2001; Cheng and Small 2006; Zhang and Rubin 2003), where the uncertainty of the impact of unmeasured confounding is accounted for by a bound (in contrast to a point) estimate. With continuous outcomes, meaningful bounds are not straightforward to obtain because the domain of the outcome is typically unrestricted.

In this paper, we consider the case of having an observational study with a continuous outcome and a valid IV. We propose to combine the IV and subjective assumptions, in the context of the HIV Epidemiology Research Study (HERS) (Smith et al. 1997), to construct meaningful bounds on (1) the population average treatment effect (ATE)

and (2) the degree of unmeasured confounding. The HERS was conducted when the highly active antiretroviral therapy (HAART) first became available to HIV-infected patients but was not randomly assigned. Of particular interest was the initial-stage causal effect of HAART on patient's CD4+ T lymphocytes (CD4) count, an important immunological marker for immune system function and disease stage. The study was conducted at two types of study site, community clinics and academic health centers, which we use as an IV. The study had collected an extensive set of covariates which can be used to adjust for part of the confounding effect, but unmeasured confounding may still exist and the magnitude of its impact is unclear. To have a sense of its impact, many HIV-positive individuals in the early HAART era were reluctant to initiate therapy due to fear of adverse side effects and toxicity. At the same time, physicians tended to prescribe HAART to patients with poor health condition, particularly with low CD4 count. These confounding factors were not fully measured and could possibly confound the HAART effect in a non-negligible way.

To account for the measured confounding effect, we implement the inverse probability weighting method proposed by Robins et al. (1994), assuming that each subject has a probability between 0 and 1 to receive HAART. In the ideal case when unmeasured confounding is absent, the IPW method can consistently estimate the ATE, and can be augmented to achieve double robustness (see Bang and Robins 2005). With unmeasured confounding, we adopt the method of Robins et al. (1999) and incorporate a sensitivity parameter into the IPW estimating equations to capture the effect of unmeasured confounding. The sensitivity parameter, defined as the systematic difference between the treated and untreated patients if hypothetically having these patients exposed to the same treatment condition, provides a measure of the magnitude of unmeasured confounding. Without external information, however, the sensitivity parameter is not identified by observed data. So in practice, this parameter is often used to conduct a sensitivity analyses to assess the robustness of the estimated causal treatment effect (Ko et al. 2003; Brumback et al. 2004) to unmeasured confounding.

In this paper, differential HAART prescription rates at the two types of study site (study site used as an IV) provide an extra piece of information that indeed allows us to infer the magnitude of the sensitivity parameter. In this paper, we impose a set of subjective assumptions in the context of the HERS. These subjective assumptions are encoded in terms of the potential outcomes as well as a sensitivity parameter for unmeasured confounding. Putting together the sensitivity parameter, the IPW estimating equations to account for measured confounding, and the IV estimating equation leads to a system of estimating equations, unified under a constraint imposed by the principal stratification (Frangakis and Rubin 2002). By solving the unified equations, we achieve our two objectives to (1) estimate the population average treatment effect of HAART at the initial treatment stage and (2) quantify the magnitude of unmeasured confounding.

The rest of the paper is organized as follows: More details about the HERS are provided in Section 2. Notations and models are given in Section 3. In Section 4, we review the IV method and the IPW method, and introduce a unified system of

estimating equations based on them. In Section 5, we present a set of contextually plausible constraints and assumptions and develop bounds on the average treatment effect of HAART on CD4 count and 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 Motivating Example: HERS

The HERS was conducted from 1993-2001 to investigate the natural history of HIV progression in women. Details of the HERS have been reported in Smith et al. (1997). In the study, a total of 871 HIV-infected women were enrolled at four study sites: Baltimore, New York City, Detroit, and Providence. The first two sites were community clinics while the other two were academic medical centers. Clinical outcomes such as CD4 count were recorded about every six months since enrollment. Starting from 1996, HAART became the recommended treatment regimen for HIV infected people, especially for those with low CD4 counts (Carpenter et al. 2000). Our analyses use data extracted from 201 women at their 8th visit, who 6 months previously were HAART-naive and had low CD4 counts (< 350 cells/mm^3). Using the HERS data, we want to estimate the initial-stage causal effect of HAART on CD4 count among this population. The study had collected a rich set of covariates, but unmeasured confounding might still exist.

The characteristics of the 201 women are summarized in Table 1. Among them, 46 (23%) have initiated the HAART. Those receiving HAART have a higher CD4 count on average than those not on HAART, but this "as-received" treatment effect (Ten Have et al. 2008) is not statistically significant (standard normal z statistic = 0.58). Ko et al. (2003) analyzed the data from the HERS and screened out several candidate confounders, which we list in the upper panel of Table 1. In brief, patients receiving HAART are more likely to be aware their HIV status and on HAART at enrollment and at the previous visit; less likely to show any HIV symptom and be a drug user; have higher viral loads (HIV-RNA) at enrollment and at the previous visit; and consist of relatively more white and less black.

In this paper, the type of study site is used as an IV assuming that conventional IV assumptions (outlined in Section 3.1) are satisfied. The validity of making these assumptions will be discussed in Section 7. In the lower panel of Table 1, we summarize the patient's CD4 count and HAART receipt rates stratified by the type of study site. Notably, patients at academic centers are more likely to be prescribed HAART (28 versus 18%) than those at the community clinics, and their average CD4 count is slightly higher. With the exogeneity assumption, this difference in CD4 count is the causal effect of study site. Further with the exclusion restriction, it is the causal effect of the differential HAART assignment between the two types of study sites. In the following, we will explore using this extra piece of information in conjunction with other assumptions to infer the causal effect of HAART and unmeasured confounding.

Table 1: Summary of patient demographic characteristics by HAART receipt status and study site. The numbers inside parentheses are standard errors. z stands for a standard normal test for comparing two sample means, and χ^2_2 for a chi-squared statistic for Pearson's chi-squared test.

	Received HAART?		Comparison
	Yes	No	statistic
Number of patients, n	46	155	-
Average CD4 counts, cell/mm^3	229 (19)	216 (11)	z = 0.58
Candidate confounders			
Race:			
black; white; others	46; 28; 26%	61; 15; 24%	$\chi^2_2 = 5.1$
ART 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 symptom, %	26% (6.5)	37% (3.9)	z = 1.2
HIV RNA, $\log_{10}$ copy/mm^3			
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^3	230 (14)	210 (12)	z = 1.0

3 Notations and Definitions

3.1 Notations

We use Z to denote the IV (in the HERS, $Z = 1$ if the study site is an academic medical center and $= 0$ otherwise), and A_z the potential treatment status that an individual would receive should Z be set to z. Hence, each individual has a pair of potential treatments (A_1, A_0) that she would potentially receive at the two types of study site. The *actual* treatment received is $A = A_Z = A_1 Z + A_0(1 - Z)$, where $A = 1$ means that the individual receives HAART, and 0 otherwise. With the Stable Unit Treatment Value Assumption (Rubin 1974), 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. Thus, the *actual* outcome observed is $Y = Y_Z(A) = Y_Z(A_Z)$. Further, we denote all confounders by a vector X, and the measured confounders by V, a subvector of X. The observed data consist of n identically and independently distributed copies of $\{X_i, Z_i, A_i, Y_i\}$, $i = 1, 2, \ldots, n$.

We assume that the conventional IV assumptions – the *exogeneity*, *exclusion restriction*, and *monotonicity* assumptions (Angrist et al. 1996; Imbens and Angrist 1994) – are satisfied. The exclusion restriction assumes $Y(a) \equiv Y_1(a) = Y_0(a)$, i.e. the IV has no direct effect on the outcome beyond its impact on individual's treatment. The monotonicity assumption requires $\Pr(A_1 \geq A_0) = 1$, i.e. an individual who would not receive HAART at an academic medical center would not do so at a community clinic either. 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

The causal effect of HAART treatment can be defined at different levels. The average treatment effect (ATE), $\mathrm{E}\{Y(1) - Y(0)\}$, is defined over the entire population. The ATE is of broad interest in public health and epidemiology, and is the parameter of interest in this paper. With an IV, the local average treatment effect (LATE; Imbens and Angrist 1994), $\mathrm{E}\{Y(1) - Y(0) \mid A_0 = 0, A_1 = 1\}$, is defined as the average treatment effect among a subpopulation who would receive the treatment only when $Z = 1$. The LATE can be estimated given a valid IV, not subject to the presence of unmeasured confounding. However, the facts that this subpopulation is not fully identifiable and the interpretation of the LATE depends on the choice of IV pose a significant limitation for generalizing results to a broader population and to other settings.

The relationship between the ATE and LATE can be expressed using the principal stratification (Frangakis and Rubin 2002). 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 our case, the potential treatments have the following four possible combinations, $(A_0, A_1) \in \{(0,0), (0,1), (1,0), (1,1)\}$, where $\{(A_0, A_1) = (0,0)\}$

indicates the subpopulation who would never receive the HAART; $\{(A_0, A_1) = (0, 1)\}$ is the subpopulation who would receive HAART only at academic medical centers; and so forth. The monotonicity assumption, $\Pr(A_1 \geq A_0) = 1$, implies that the subpopulation $\{(A_0, A_1) = (1, 0)\}$ is an empty set. Henceforth, we denote the remaining three subpopulations by $\mathcal{P}_{jk} = \{A_0 = j, A_1 = k\}$, $j \leq k$.

We denote the ATE and LATE by β^{ATE} and β^{LATE}, respectively. Then their relationship can be expressed by

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

where $\pi_{jk} = \Pr(\mathcal{P}_{jk})$ and $\mu_{jk}(a) = \mathrm{E}\{Y(a) \mid \mathcal{P}_{jk}\}$. In this paper, this relationship is used to unify the IPW and IV estimation methods, and based on that, a system of estimating equations is developed to draw inference on the ATE of HAART as well as the magnitude of unmeasured confounding.

4 Review of Estimation Methods

We review the IPW and IV methods in this section.

4.1 The IPW Method

Putting aside the covariates for the moment, the potential outcomes can be expressed by a marginal structural mean model (Gange et al. 2007; Robins 1999; Robins et al. 2000)

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

Assuming that $Y(a) \perp A|V$, i.e. unmeasured confounding is absent, we can estimate β^{ATE} by the solution $\hat{\beta}_{\text{IPW}}$ to the IPW estimating equations

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

where $W_{1i} = A_i/e(V_i;\gamma) + (1 - A_i)/\{1 - e(V_i;\gamma)\}$, and $e(V;\gamma) = \Pr(A = 1|V)$ is the propensity score (Rosenbaum and Rubin 1983) with a l-dimension parameter γ. We assume that $0 < e(V;\gamma) < 1$.

The IPW method has several properties that are worth mentioning. The efficiency of the resulting estimator can be improved by using stabilized weights to replace W_{1i} (Miguel et al. 2001). The estimator can be augmented to achieve double robustness if we further specify an outcome regression model on Y (Bang and Robins 2005). Moreover, if γ is unknown, $\hat{\beta}_{\text{IPW}}$ remains consistent when γ is replaced by a consistent estimator $\hat{\gamma}$ that solves

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

where W_{2i} is an appropriate weight function; e.g. $W_{2i} = \partial e(V_i;\gamma)/\partial\gamma$ in logistic regressions.

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 (1999) proposed to introduce a sensitivity parameter τ and then estimate β^{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 $Y_i^* = Y_i - \tau\{A_i - e(V_i; \gamma)\}$ is the "outcome" corrected for the selection bias due to unmeasured confounding. For a binary treatment as in this paper, the sensitivity parameter can be defined as the contrast of the potential outcomes between the treated and untreated conditional on V.

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

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

Without additional information from data, the parameter τ is not identified. Hence, the resulting estimator $\hat{\beta}_{\text{MIPW}}(\tau)$ is typically used to conduct a sensitivity analysis. That is, estimate β^{ATE} using $\hat{\beta}_{\text{MIPW}}(\tau)$ as if τ is known, and then examine the sensitivity of $\hat{\beta}_{\text{MIPW}}(\tau)$ by varying the value of τ over its plausible range (Ko et al. 2003; Brumback et al. 2004).

With an IV and information extracted by IV, it becomes possible to draw inference about the ATE as well as τ which quantifies the degree of unmeasured confounding.

4.2 The IV Method

The IV methods have been widely used in econometric research (c.f. Wooldrige 2002). In our just-identified case with a single binary IV and a binary treatment, the standard IV estimating equations are

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

Under the IV assumptions and $\text{Cov}(Z, A) \neq 0$, the solution

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

is consistent for β^{LATE} (Imbens and Angrist 1994; Angrist et al. 1996; Hernan and Robins 2006). The bars in (3) calculate sample averages, e.g. $\overline{YZ} = \sum_i Y_i Z_i/n$. One important property of the IV method is that $\hat{\beta}_{\text{IV}}$ remains consistent despite of the existence of unmeasured confounding.

Under the framework of the generalized method of moments, the IV estimating equations are solved using the two-stage least squares method (Angrist and Imbens 1995), and can further incorporate a weight matrix to allow for heteroskedastic or correlated residuals. The IV methods also can be generalized to deal with multiple IVs and non-continuous outcomes (c.f. Wooldrige 2002).

5 A Unified System of Estimating Equations

With a set of covariates and an IV in the HERS, we propose to combine the IV and IPW methods, and develop a unified system of estimation equations as follows,

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

with a constraint of (1). Note that in (1), parameters $\mu_{11}(0)$ and $\mu_{00}(1)$ are the averages of unobserved potential outcomes and not identified. All other parameters are identified because $\pi_{11} = \text{E}(A = 1|Z = 0)$, $\pi_{00} = \text{E}(A = 0|Z = 1)$, $\pi_{01} = 1 - \pi_{00} - \pi_{11}$, $\mu_{11}(1) = \text{E}(Y|A = 1, Z = 0)$, and $\mu_{00}(0) = \text{E}(Y|A = 0, Z = 1)$ (Angrist et al. 1996, and Rejoinder). When natural limits exist on $\mu_{11}(0)$ and $\mu_{00}(1)$ e.g. with binary outcomes, both the ATE and τ are partially identified to bounds. When no natural limits exist as is our case with continuous outcomes, additional prior information is needed to implement the constrained estimating equations system, which we will discuss next.

We present three sets of assumptions in the context of the HERS. Each allows us to identify bounds on the ATE and unmeasured confounding parameter τ. In Sections 5.1 - 5.3, we assume that the sample size n is sufficiently large such that the sampling variations of the estimating equations (4) is ignored. In Section 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 both $\mu_{00}(1)$ 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 ξ_0 and ξ_1.

Assumption (A) leads to a simplified version of the Robins-Manski type bound on the ATE (Robins 1989; Manski 1990; Zhang and Rubin 2003). It is straightforward to show that the ATE falls within the interval

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

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

Then the bound on τ can be inferred by finding the values of τ such that the corresponding solutions to $\hat{\beta}^{\text{MIPW}}(\tau)$ are consistent with the above bound on ATE. For a given β^{ATE}, the solution to $U_3(\beta_{\text{MIPW}}, \tau) = 0$ for τ is

$$\hat{\tau}_n(\beta^{\text{ATE}}, \gamma) = \frac{\overline{W_1 A} * \overline{W_1(Y - \beta^{\text{ATE}} A)} - \bar{W}_1 * \overline{W_1 A(Y - \beta^{\text{ATE}} A)}}{\overline{W_1 A} * \overline{W_1(A - e(V;\gamma))} - \bar{W}_1 * \overline{W_1 A(A - e(V;\gamma))}}.$$

It is straightforward to verify that $\hat{\tau}_n(\beta^{\text{ATE}}, \gamma)$ is a non-increasing function of β^{ATE}, so the unmeasured confounding parameter τ is bounded by

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

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

5.2 Constraint on Relationships between $\mu_{11}(0)$ and $\mu_{00}(1)$ and Identified Quantities

Assumption (A) alone is sufficient for identifying the bounds on ATE and τ, but in practice the two upper limits need to sufficiently large and the resulting bounds can be wide. In the following, we consider making assumptions on the relative magnitude between the unidentifiable and identifiable quantities.

Assumption (B): We assume that

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

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

   A plausible choice for δ_{11} is zero, that is, we assume that on average, $\mathcal{P}_{11}$ who would *always* receive HAART can *on average* benefit from HAART. We make this assumption because although suffering from confounding bias, the efficacy of HAART on the treated patients has been demonstrated by several contemporary studies. Further, we impose a known lower bound on the average treatment effect among $\mathcal{P}_{00}$ that

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

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

2. The difference on $\mathrm{E}\{Y(0)\}$ between $\mathcal{P}_{11}$ and $\mathcal{P}_{00}$ is bounded above,

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

   We can set $\delta_{y0} = 0$ by our intuition that in the untreated condition, people who would *always* receive HAART have higher degree of HIV progression (lower CD4 on average, compared to those who would *never* receive HAART).

3. The difference of treatment effects between those who would *always* receive HAART and those who would *never* receive HAART is bounded below,

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

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

Under this set of assumptions, we show that the ATE is bounded by

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

and τ by

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

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

For the HERS, it may be more realistic to assume that Assumption (B) holds conditional on clinically important covariates V. So we propose our third set of assumptions as

Assumption (B'):

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

$$\begin{aligned}\beta = \pi * \beta^{\mathrm{LATE}} &+ \int_{\mathcal{V}} \mathrm{E}\{Y(1) - Y(0)|\mathcal{P}_{11}, V\} P(\mathcal{P}_{11}|V) \mathrm{d}F(V) \\ &+ \int_{\mathcal{V}} \mathrm{E}\{Y(1) - Y(0)|\mathcal{P}_{00}, V\} P(\mathcal{P}_{00}|V) \mathrm{d}F(V), \end{aligned} \tag{5}$$

   where $\mathcal{V}$ is the support of V with a distribution $F(V)$. We write the product $\pi_{01} * \beta^{\mathrm{LATE}}$ as before because both the LATE and π_{01} are identified by the data. Again, no observed data are available for $\mathrm{E}\{Y(0)|\mathcal{P}_{11}, V\}$ and $\mathrm{E}\{Y(1)|\mathcal{P}_{00}, V\}$, which are denoted by $\mu_{11}(0, V)$ and $\mu_{00}(1, V)$, respectively.

Under (B') and (5), we obtain a bound on ATE

$$[\pi_{01} \times \mathrm{LATE} + \int_{\mathcal{V}} b_V(c_0(V), c_2(V)) \mathrm{d}F,\ \pi_{01} \times \mathrm{LATE} + \int_{\mathcal{V}} b_V(0, c_1(V)) \mathrm{d}F]$$

and a bound on τ

$$[\tau(\pi_{01} \times \mathrm{LATE} + \int_{\mathcal{V}} b_V(0, c_1(V)) \mathrm{d}F, \gamma),\ \tau(\pi_{01} \times \mathrm{LATE} + \int_{\mathcal{V}} b_V(c_0(V), c_2(V)) \mathrm{d}F, \gamma)],$$

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

5.4 Inference from Finite Samples

To implement (1), we can assume two regression models $E(A|Z) = \text{logit}^{-1}(\eta(Z;\theta_1))$ and $E(Y|A,Z) = \kappa(A,Z;\theta_2)$ for some known functions $\eta(Z;\theta_1)$ and $\kappa(A,Z;\theta_2)$. For binary Z and A, we 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), since one can show that (2) suggests that $\text{E}(Y|A,Z)$ is linear in A, Z, and AZ.

Given two regression models, we can obtain consistent estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ 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 $W_{4i} = (1,Z_i,A_i,A_iZ_i)^\top$. Then, we estimate $\hat{\pi}_{11} = \text{logit}^{-1}\{\eta(0;\hat{\theta}_1)\}$, $\hat{\pi}_{00} = 1 - \text{logit}^{-1}\{\eta(1;\hat{\theta}_1)\}$, $\hat{\pi}_{01} = \text{logit}^{-1}\{\eta(1;\hat{\theta}_1)\} - \text{logit}^{-1}\{\eta(0;\hat{\theta}_1)\}$, $\hat{\mu}_{11}(1) = \kappa(1,0;\hat{\theta}_2)$, and $\hat{\mu}_{00}(0) = \kappa(0,1;\hat{\theta}_2)$.

For Assumptions (A) and (B), the function $b()$ can be estimated by replacing those identifiable quantities with their consistent estimators, i.e. $\hat{b}(\mu_{11}(0),\mu_{00}(1)) = \hat{\pi}_{01}\hat{\beta}_{\text{IV}} + \hat{\pi}_{11}\{\hat{\mu}_{11}(1) - \mu_{11}(0)\} + \hat{\pi}_{00}\{\mu_{00}(1) - \hat{\mu}_{00}(0)\}$. Further, we estimate c_0 by $\hat{c}_0(\delta_{11},\delta_{y0}) = \min\{(\hat{\mu}_{11}(1)-\delta_{11}),(\hat{\mu}_{00}(0)+\delta_{y0})\}$ and c_1 by $\hat{c}_1(\delta_{trt}) = \hat{\mu}_{11}(1)+\hat{\mu}_{00}(0)-\delta_{trt}$. By substituting these estimates for their estimands in the bounds, we obtain bound estimates on the ATE and τ. The results are summarized in Table 2.

To estimate the bounds on ATE and τ under Assumption (B'), we proceed as follows:

Step 1. We assume two observed-data models conditional on V, $\text{E}(A|Z,V) = \text{logit}^{-1}(\eta(Z,V;\theta_3))$, and $\text{E}(Y|Z,A,V) = \kappa(Z,A,V;\theta_4)$ for known $\eta(Z,V;\theta_3)$ and $\kappa(Z,A,V;\theta_4)$. For example, we can assume two linear models without interaction that $\eta(Z,V;\theta_3) = \theta_{30} + \theta_{31}Z + \theta_{32}V$ with $\theta_3 = (\theta_{30},\theta_{31},\theta_{32}^\top)^\top$ and $\kappa(Z,A,V;\theta_4) = \theta_{40} + \theta_{41}Z + \theta_{42}A + \theta_{43}^\top V$ with $\theta_4 = (\theta_{40},\theta_{41},\theta_{42},\theta_{43}^\top)^\top$. Let $\hat{\theta}_4$ and $\hat{\theta}_3$ denote the estimators of θ_4 and θ_3 by solving the corresponding estimating equations.

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

Step 3. We estimate the distribution $F(V)$ by the empirical cumulative density function of V and integrals by empirical sums, e.g. estimate $\int_{\mathcal{V}} b_V(c_0(V),c_2(V))\text{d}F$ by $\sum_{i=1}^n \hat{b}_V(\hat{c}_0(V_i),\hat{c}_2(V_i))/n$. Then, we substitute the parameters in (5) by their estimates.

The resulting bound estimates on the ATE and τ are summarized in Table 2.

5.5 Uncertainty Region for Estimated Bounds

An interval that provides $(1-\theta)100\%$ coverage probability on a bound estimate is often called the $(1-\theta)100\%$ *Uncertainty Region* (UR) to distinguish it from a confidence interval. A UR takes into account both the sampling variability and partial identifiability. Two types of URs are considered in this paper, *point-wise* and *strong* $(1-\theta)100\%$ coverage URs.

A point-wise UR $(\hat{L}, \hat{U})$ contains any particular value $\varrho \in (L, U)$ with a probability of at least $(1-\theta)$, where (L, U) denotes the true bound and ϱ is the parameter generating the data. If the $\hat{L}$ and $\hat{U}$ are consistent estimates and asymptotically normally distributed (CAN), a $(1-\theta)100\%$ point-wise UR is given by

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

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 $\Phi^{-1}(1-\theta)$ where Φ is the normal cumulative density function (Vansteelandt et al. 2006).

A strong UR is defined as an interval that contains the entire set (L, U) with a probability of at least $(1-\theta)$ (Horowitz and Manski 2000; Vansteelandt et al. 2006). If both $\hat{L}$ and $\hat{U}$ are CAN, a strong $(1-\theta)100\%$ UR is

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

with $\text{c} = \Phi^{-1}(1-\theta/2) = 1.96$. Without assuming $\hat{L}$ and $\hat{U}$ to be CAN, a strong $(1-\theta)$ UR can be obtained using the bootstrap method. Specifically, let $(\tilde{L}^*, \tilde{U}^*)$ denote the estimated bound from a bootstrapped sample. A bootstrap strong 95% UR is the interval (L^*, U^*) satisfying $\Pr^*(L^* \leq \tilde{L}^*, \tilde{U}^* \leq U^*) = 1-\theta$ and $\Pr^*(\tilde{L}^* < L^*) = \Pr^*(\tilde{U}^* > U^*)$, where $\Pr^*$ is the probability measure induced by the bootstrapped resamples (Bickel and Freeman 1981), and so can be obtained by finding the shortest interval $\text{UR}_{\text{S-BTS}} = (L^*, U^*)$ that satisfies $\frac{\#(L^* \leq \tilde{L}_k^* < \tilde{U}_k^* \leq U^*)}{K} \geq 1-\theta$, and $\frac{\#(L^* \leq \tilde{L}_k^*)}{K} \simeq \frac{\#(U^* \geq \tilde{U}_k^*)}{K}$, where $\#()$ counts the number of statements that hold for $k = 1, 2, \ldots, K$.

6 HERS Data: Treatment Effect Estimation and Confounding Assessment

6.1 Preliminary Analyses

The upper panel of Table 3 summarizes the IPW estimate of the ATE and IV estimate of the LATE of HAART on CD4 count. The IPW uses the variables listed in Table 1 as the measured confounders of V and assumes that $e(V;\gamma) = \text{logit}^{-1}(\gamma^\top V)$. The IPW estimate of the ATE suggests that HAART can boost patient's CD4 count by 27 cells/mm^3 on average with a 95% confidence interval of $(-16, 70)$. The IV estimate of the LATE suggests that for those who would receive HAART at academic medical centers but not at community clinics, HAART can increase CD4 count by 207 on

average with a 95% CI of (−250, 664). In Table 3, we also list the "as-treated" (AT) treatment effect, which is estimated by the contrast of the average CD4 counts between those actually receiving HAART and those not. The difference between the IPW and AT estimates can be regarded as the bias of AT estimate that is attributable to the *measured* confounders.

6.2 Bounds on HAART Treatment Effect and Unmeasured Confounding

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 much lower than 350 and at the eighth visit, the average CD4 count was 229 for those treated and 216 for those untreated (refer to Tables 1). For Assumptions (B) and (B'), we let $\delta_{11} = \delta_{00} = \delta_{y0} = \delta_{trt} = 0$, and further for (B') let V be the variables listed in Table 1. The lower panel of Table 3 summarizes the bounds estimates of the ATE and τ. The bound estimate of the ATE $(-196, 256)$ under (A) is not informative and much wider than those under (B) (20, 231) and (B') (18, 218). Assumption (B') is not necessarily stronger than (B), but by imposing observed data models and having the estimated bounds smoothed over covariates, tighter bounds are resulted in. To obtain uncertainty regions on these bound estimates, we draw $K = 1,000$ bootstrap samples, fixing the number of patients at the two types of study sites (academic medial centers versus community clinics). Because the bounds estimates contain min() operation which complicates the derivation of their standard errors, we use the K bootstrapped samples to calculate the standard errors of the two ends of bound estimates. Table 3 summarizes the point-wise, strong, and bootstrap strong 95% coverage URs. The difference between the 95% CI of the IPW estimate and 95% URs under (B) and (B') can be regarded as the bias of the IPW estimate due to *unmeasured* confounding. The results suggest that unmeasured confounding tends to cause a downward bias and the true ATE is likely to higher than what the IPW 95% CI suggests.

The bound estimates on τ under (B) and (B') are $(-204, 7.5)$ and $(-191, 9.1)$, which are much tighter and more informative than the bound estimate under (A) $(-229, 223)$. The 95% URs on τ are listed in Table 3. A possibly negative value of τ implies that unmeasured factors resulted in preferential prescriptions of HAART to those with fewer CD4 count in the HERS, and those on HAART might have up to 200 fewer CD4 count (if left untreated) on average compared with those not on HAART.

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 ξ vary from 300 to 500. The bound estimates on the ATE and τ along with bootstrap strong 95% URs are shown in Figure 1 (First row). For the considered range, ξ has more influence on the upper (lower) bound estimate on ATE

Table 2: Estimated bounds on ATE and τ under Assumptions (A), (B) and (B').

	Parameter	Estimated bound
(A)	ATE	$[\hat{b}(\xi_0, 0),\ \hat{b}(0, \xi_1)]$
	τ	$[\hat{\tau}_n(\hat{b}(0, \xi_1), \hat{\gamma}),\ \hat{\tau}_n(\hat{b}(\xi_0, 0), \hat{\gamma})]$
(B)	ATE	$[\hat{b}(\hat{c}_0, \hat{\mu}_{00}(0) + \delta_{00}),\ \hat{b}(0, \hat{c}_1)]$
	τ	$[\hat{\tau}_n(\hat{b}(0, \hat{c}_1), \hat{\gamma}),\ \hat{\tau}_n(\hat{b}(\hat{c}_0, \hat{\mu}_{00}(0) + \delta_{00}), \hat{\gamma})]$
(B')	ATE	$[\hat{\pi}_{01}\hat{\beta}_{\text{IV}} + \frac{\sum_i \hat{b}_v(\hat{c}_0(V_i), \hat{c}_2(V_i))}{n}, \hat{\pi}_{01}\hat{\beta}_{\text{IV}} + \frac{\sum_i u\hat{b}_v(0, \hat{c}_1(V_i))}{n}]$
	τ	$[\hat{\tau}_n\{\hat{\pi}_{01}\hat{\beta}_{\text{IV}} + \frac{\sum_i \hat{b}_v(0, \hat{c}_1(V_i))}{n}\}, \hat{\tau}_n\{\hat{\pi}_{01}\hat{\beta}_{\text{IV}} + \frac{\sum_i \hat{b}_v(\hat{c}_0(V_i), \hat{c}_2(V_i))}{n}\}]$

Table 3: Estimates of HAART treatment effect on CD4 and τ. The 95% confidence intervals (CI) for point estimates and 95% uncertainty regions for bound estimates ($\text{UR}_{\text{P-CAN}}$, $\text{UR}_{\text{S-CAN}}$, and $\text{UR}_{\text{S-BOOT}}$; see Section 5.5) are shown in **bold** font. We assume that $\xi_0 = \xi_1 = 500$ for (A); and that $\delta_{00} = \delta_{11} = \delta_{y0} = \delta_{trt} = 0$ for (B) and (B').

		ATE	τ
AT	Point estimate	13	-
	95% CI	**(−30, 56)**	
IPW	Point estimate	27	-
	95% CI	**(−16, 70)**	
IV	Point estimate	207	-
	95% CI	**(−250, 664)**	
Assumption:			
(A):	Bound estimate	(−196, 256)	(−229, 223)
	95% $\text{UR}_{\text{P-CAN}}$	**(−229, 289)**	**(−269, 266)**
	95% $\text{UR}_{\text{S-CAN}}$	**(−235, 295)**	**(−277, 274)**
	95% $\text{UR}_{\text{S-BTS}}$	**(−233, 294)**	**(−274, 273)**
(B):	Bound estimate	(20, 231)	(−204, 7.5)
	95% $\text{UR}_{\text{P-CAN}}$	**(−9, 280)**	**(−260, 49)**
	95% $\text{UR}_{\text{S-CAN}}$	**(−15, 289)**	**(−271, 57)**
	95% $\text{UR}_{\text{S-BTS}}$	**(−14, 285)**	**(−270, 57)**
(B'):	Bound estimate	(18, 218)	(−191, 9.1)
	95% $\text{UR}_{\text{P-CAN}}$	**(−10, 261)**	**(−234, 48)**
	95% $\text{UR}_{\text{S-CAN}}$	**(−16, 270)**	**(−243, 56)**
	95% $\text{UR}_{\text{S-BTS}}$	**(−14, 270)**	**(−243, 56)**

(on τ), and the resulting bound estimates remain wide and non-informative.

For (B) and (B'), we let δ_{00} (the lower limit of treatment effect among $\mathcal{P}_{00}$) rang from -60 to 20 and fix $\delta_{11} = \delta_{y0} = \delta_{trt} = 0$. (More sophisticated sensitivity analyses that jointly evaluate δ_{11}, δ_{00}, δ_{y0} and δ_{trt} are possible.) We choose this range for δ_{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 δ_{00} only affects the lower (upper) bound estimates of the ATE (of τ). The ATE can be as high as over 200 cell/mm^3, and the lower bound of ATE varies around zero depending on the value of δ_{00}. Again, a possible negative value of τ suggests that unmeasured confounding likely causes HAART to be preferentially prescribed to those with poorer health.

7 Discussions

In the study of HERS, we propose to use an IV and sets of contextually plausible assumptions to quantify the causal effect of a treatment as well as the degree of unmeasured confounding. We consider three sets of assumptions. 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, and lead to bounds that can be much tighter than (A). The 95% uncertainty regions for the ATE under (B) and (B') are more informative than the 95% CI of the IV estimate, and have less concern of having unmeasured confounding bias compared to the 95% CI of the IPW estimate. The bound estimates on the ATE and τ 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 (e.g. Gange et al. 2007) have been conducted contemporarily as the HERS, and could suffer from similar amount of unmeasured confounding. In those studies when unmeasured confounding is of concern, analyses should be complemented with a sensitivity analyses as described in Section 6.3. A plausible range for τ 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. Such an observation suggests that the deterministic monotonicity $\Pr(A_1 \geq A_0) = 1$ is plausible, but this assumption cannot be 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, the bias due to the violation of the monotonicity assumption is probably negligible. Alternatively, one can assume that $\mathcal{P}_{00}$ is absent,

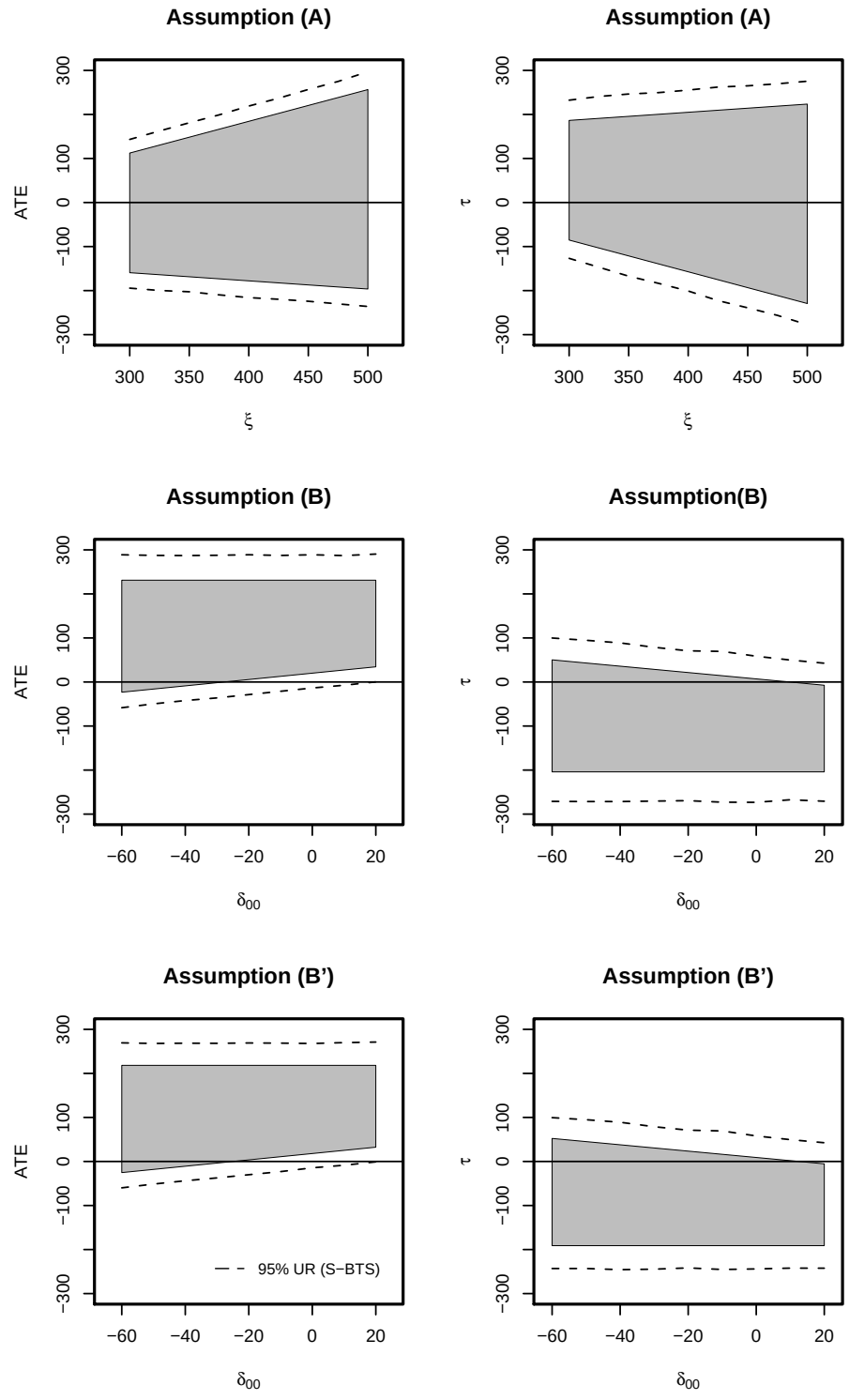

Figure 1: Sensitivities of bound estimates to $\xi = \xi_0 = \xi_1$ under Assumption (A); to δ_{00} under Assumptions (B) and (B'). The gray zones show the bound estimates as a function of ξ or δ_{00}. The bootstrap strong 95% $\text{UR}_{\text{S-BTS}}$'s are shown as dashed lines.

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 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: $\mathrm{E}\{Y(0)|\mathcal{P}_{01}\} = \mathrm{E}(Y|A = 0, Z = 0)$, $\mathrm{E}\{Y(0)|\mathcal{P}_{10}\} = E(Y|A = 0, Z = 1)$, $\mathrm{E}\{Y(1)|\mathcal{P}_{11}\mathrm{or}\mathcal{P}_{01}\} = \mathrm{E}(Y|A = 1, Z = 1)$, and $\mathrm{E}\{Y(1)|\mathcal{P}_{11}\mathrm{or}\mathcal{P}_{10}\} = \mathrm{E}(Y|A = 1, Z = 0)$. 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 (1). This may be worth further investigation.

Moreover, replacing the deterministic monotonicity with a stochastic monotonicity assumption deserves explorations in the future. Roy et al. (2008) assumed $\Pr(A_1 = 1|A_0 = 1, V) \geq \Pr(A_1 = 1|A_0 = 0, V)$, and proposed to use auxiliary covariates to estimate the memberships of principal strata. Small and Tan (2008) 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 $\mathcal{P}_{10}$ and may be more realistic in the HERS than the deterministic monotonicity.

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^*, i.e. $\{Y(1), Y(0)\} \perp Z|V^*$. In this case, the methods by Tan (2006) can be implemented for identifying the LATE, and our method for bound estimation on ATE and τ 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 (Bang and Robins 2005) 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 and Schafer (2007) suggest that IPW is relatively robust to the impact of unmeasured confounding in term of estimation bias. Because the focus issue of this paper is unmeasured confounding, we use the IPW for estimating ATE.

References

Joshua D. Angrist and Guido W. Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90(430):431–442, 1995.

Joshua D. Angrist, Guido W. Imbens, and Donald B. Rubin. Identification of causal

effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455, 1996. ISSN 0162-1459.

Heejung Bang and James M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61:962–972, 2005.

P. J. Bickel and D. A. Freeman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9:1196–1217, 1981.

Bagette A. Brumback, Miguel A. Hernán, Sebastien J. P. A. Haneuse, and James M. Robins. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Statistics in Medicine*, 23(5):749–767, 2004.

C.C.J. Carpenter, D.A. Cooper, J.M. Fischl, J.M. Gatell, B.G. Gazzard, S.M. Hammer, M.S. Hirsch, D.M. Jacobsen, D.A. Katzenstein, J.S.G. Montaner, D.D. Richman, M.S. Saag, M. Schechter, M. Schooley, M. A. Thompson, S. Vella, P.G. Yeni, and P. A. Volberding. Antiretroviral therapy in adults:updated recommendations of the international aids society-usa panel. *Journal American Medical Association*, 283: 381–391, 2000.

Jing Cheng and Dylan S. Small. Bounds on causal effects in three-arm trials with noncompliance. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 68(5):815–836, 2006.

R.B. D'Agostino. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17(19):2265–2281, 1998.

Constantine E. Frangakis and Donald B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002.

Stephen J Gange, Mari M Kitahata, Michael S Saag, David R Bangsberg, Ronald J Bosch, John T Brooks, Liviana Calzavara, Steven G Deeks, Joseph J Eron, Kelly A Gebo, M John Gill, David W Haas, Robert S Hogg, Michael A Horberg, Lisa P Jacobson, Amy C Justice, Gregory D Kirk, Marina B Klein, Jeffrey N Martin, Rosemary G McKaig, Benigno Rodriguez, Sean B Rourke, Timothy R Sterling, Aimee M Freeman, and Richard D Moore. Cohort profile: The North American AIDS cohort collaboration on research and design (NA-ACCORD). *International Journal of Epidemiology*, 36:294–301, 2007.

Miguel Angel Hernan and J. M. Robins. Instruments for causal inference: An epidemiologist's dream? *Epidemiology*, 17:360–372, 2006.

Joseph W. Hogan and Tony Lancaster. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. *Statistical Methods in Medical Research*, 13:17–48, 2004.

Joel L. Horowitz and Charles F. Manski. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, 95(449):77–84, 2000.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994. ISSN 0012-9682.

Marshall M. Joffe. Using information on realized effects to determine prospective causal effects. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 63 (4):759–774, 2001. ISSN 1369-7412.

J. D. Y. Kang and J. L. Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22:523–539, 2007.

Hyejin Ko, Joseph W. Hogan, and Kenneth H. Mayer. Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models. *Biometrics*, 59:152–162, 2003.

C.F. Manski. Nonparametric bounds on treatment effects. *American Economic Reviews, Papers and Proceedings*, 80:319–323, 1990.

Hernan A Miguel, Brumback Babette, and J. M. Robins. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96:440–448, 2001.

J. M. Robins. *The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies*, chapter Health Service Research Methodology: A Focus on AIDS, pages 213–290. Washington D.C.: U.S. Public Health Service, 1989.

James M. Robins. Association, causation, and marginal structural models. *Synthese*, 121:151–179, 1999.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. ISSN 0162-1459.

James M. Robins, Miguel Ángel Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560, 2000.

J.M. Robins, A. Rotnitzky, and D.O. Scharfstein. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. *Statistical Models in Epidemiology: The Environment and Clinical Trials*, 116:1–92, 1999.

Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

P.R. Rosenbaum. *Observational studies*. Springer, 2002.

P.R. Rosenbaum and D.B. Rubin. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524, 1984.

Jason Roy, Joseph W. Hogan, and Bess H. Marcus. Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics*, 9:277–289, 2008.

D. B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:668–701, 1974.

Dylan S. Small and Zhiqiang Tan. A stochastic monotonicity assumption for the instrumental variables method. Technical report, University of Pennsylvnia, 2008.

D.K. Smith, D. Warren, P. Vlahov, P. Schuman, M.D. Stein, B.L. Greenberg, and S.D. Holmberg. Design and baseline participant characteristics of the human immunodeficiency virus epidemiology research (HER) study: a prosepective cohort study of human immunodeficiency virus infection in us women. *American Journal of Epidemiology*, 146:459–469, 1997.

J.H. Stock and F. Trebbi. Retrospectives: Who invented instrumental variable regression? *The Journal of Economic Perspectives*, 17(3):177–194, 2003.

Zhiqiang Tan. Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association*, 101(476):1607–1618, 2006.

T.R. Ten Have, S.L.T. Normand, S.M. Marcus, C.H. Brown, P. Lavori, and N. Duan. Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psychiatric annals*, 38(12):772, 2008.

T.J. VanderWeele and I. Shpitser. On the definition of a confounder. Technical report, bepress, 2011.

Stijn Vansteelandt, Els Goetghebeurand, Michael G. Kenward, and Geert Molenberghs. Ignorance and uncerntainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16:953–979, 2006.

Jefferey M. Wooldrige. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press., 2002.

S Wright. *Appendix to the Tariff on Animal and Vegetable Oils, by P. G. Wright,*, volume 26. New York: MacMillan., 1928.

J. L. Zhang and D. B. Rubin. Estimation of causal effects via principal stratification when some outcomes are truncated by death. *Journal of Educational and Behavioral Statistics*, 28:353–368, 2003.