Research Article

Statistics in Medicine

Received 18 September 2009, Accepted 31 January 2011 Published online 15 April 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4241

# Two-stage instrumental variable methods for estimating the causal odds ratio: Analysis of bias

**Bing Cai,[a,c]*† Dylan S. Small[b] and Thomas R. Ten Have[c]**

**We present closed-form expressions of asymptotic bias for the causal odds ratio from two estimation approaches of instrumental variable logistic regression: (i) the two-stage predictor substitution (2SPS) method and (ii) the two-stage residual inclusion (2SRI) approach. Under the 2SPS approach, the first stage model yields the predicted value of treatment as a function of an instrument and covariates, and in the second stage model for the outcome, this predicted value replaces the observed value of treatment as a covariate. Under the 2SRI approach, the first stage is the same, but the residual term of the first stage regression is included in the second stage regression, retaining the observed treatment as a covariate. Our bias assessment is for a different context from that of Terza (*J. Health Econ.* 2008; 27(3):531–543), who focused on the causal odds ratio conditional on the unmeasured confounder, whereas we focus on the causal odds ratio among compliers under the principal stratification framework. Our closed-form bias results show that the 2SPS logistic regression generates asymptotically biased estimates of this causal odds ratio when there is no unmeasured confounding and that this bias increases with increasing unmeasured confounding. The 2SRI logistic regression is asymptotically unbiased when there is no unmeasured confounding, but when there is unmeasured confounding, there is bias and it increases with increasing unmeasured confounding. The closed-form bias results provide guidance for using these IV logistic regression methods. Our simulation results are consistent with our closed-form analytic results under different combinations of parameter settings.** **Copyright © 2011 John Wiley & Sons, Ltd.**

**Keywords:** instrumental variable; two-stage residual inclusion; two-stage predictor substitution; logistic regression; bias

## 1. Introduction

Instrumental variable (IV) methods are used to estimate the effects of receiving treatment or exposure to risk factor on outcome when there is unmeasured confounding in medical research, such as in clinical trials under non-adherence to treatment [1] or observational studies [2, 3]. We present closed-form expressions of asymptotic bias for the causal odds ratio from two-stage logistic regressions, which is an extension of the conventional IV method for continuous outcomes to a binary outcome.

In the following discussion, we use 'treatment' to represent either treatment received or exposure to a risk factor. An IV has the following properties: (a) it is associated with treatment; (b) it has no direct causal effect on the outcome (exclusion restriction); and (c) it is independent of all (unmeasured) confounders of the treatment–outcome relationship [2, 4–6]. Note that in randomized trials, the randomized treatment assignment IV is independent of all confounders because it is randomized. In an

[a]*Merck Research Laboratories, UG1D-60, P.O. Box 1000, North Wales, PA 19454-1099, U.S.A.*
[b]*Department of Statistics, Wharton School, 464 JMHH/6340, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.*
[c]*Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Blockley Hall, 6th FLR 423 Guardian Dr., Philadelphia, PA 19104-6021, U.S.A.*
**Correspondence to: Bing Cai, Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Blockley Hall, Guardian Dr., Philadelphia, PA 19104-6021, U.S.A.*
†*E-mail: bcai@mail.med.upenn.edu*



Copyright © 2011 John Wiley & Sons, Ltd. 

observational study, the IV could be associated with measured confounders as long as it is independent of all unmeasured confounders of the treatment–outcome relationship conditional on the measured confounders, and the measured confounders are controlled for in the analysis [5]. Under these conditions, the IV analysis of the treatment-outcome relationship controls for measured and unmeasured confounding [4, 7–9].

In the context of randomized trials, the IV analysis has been used to adjust for all measured and unmeasured confounding due to treatment non-compliance when estimating the effect of actually receiving treatment. Such confounding factors impact the outcome while causing treatment non-compliance or switching from one treatment to another. While intent-to-treat (ITT) inference comparing randomized groups but ignoring treatment non-compliance is protected against such unmeasured confounding, this inference pertains to the effect of prescribing or assigning treatment in the population with the same rate and pattern of non-compliance in the particular trial. Using randomized treatment as an IV, IV inference for the effect of receiving treatment is not dependent on the rate of compliance in the trial except that lower compliance leads to higher variability [10]. This IV inference aims to estimate the effect of actually receiving treatment, which is useful for individual patient decisions and for predicting the effect of making the treatment available to populations in which the rate of compliance might differ from the trial [11, 12].

Besides clinical trials, IV methods are used in observational studies, such as data-based evaluations of the effect of medication on clinical or adverse outcomes. IVs, such as physician's prescribing preference [13–17], clinic or hospital [18], or geographic region [19–21], have been used to adjust for confounders of the intervention–outcome relationship.

For the additive effect of treatment, Angrist *et al.* [4] considered five assumptions for a setting with a proposed IV which are explained in detail in Section 2. Briefly, the key assumptions are that the proposed IV is associated with treatment, is independent of unmeasured confounders given the measured confounders, and that the IV only affects outcome through treatment received and there are no defiers (DFs). With these assumptions, they used principal stratification [22] to motivate interpretation of the IV estimand. Under the principal stratification framework, the population is divided into sub-classes based on the potential treatment receipt that would occur under each level of the IV. In the context of randomized trials with non-compliance, the principal strata are defined as compliers (Cs), who adhere to the assignment of treatment but do not take it when not assigned to it; always-takers (ATs) and never-takers (NTs), who respectively always or never take treatment regardless of assignment; and DFs, who only take treatment when not assigned to it. They proved that the probability limit of the two-stage least squares estimator, the usual IV estimator, is the average causal effect of receiving treatment among Cs, which is called the local average treatment effect (LATE) or the compiler average causal effect (CACE). Under certain no-interaction assumptions, this effect pertains to other sub-groups including anyone who takes the treatment or all patients. The estimands for other types of estimators based on structural mean models (SMMs) can be interpreted similarly [23, 24].

For binary outcomes, the IV approach has been extended in different ways for inference based on odds ratios under logistic models, where the odds ratio is interpreted as the effect of treatment on outcome in Cs. Those approaches include the Bayesian logistic model estimated with Markov-Chain Monte Carlo techniques [25], the SMM [26–28], and a multi-stage approach including an estimation step for the prediction of treatment as a function of the IV [29].

Terza *et al.* [30] extended the two-stage IV approach for non-linear models including the logistic regression model (two-stage predictor substitution (2SPS)), where the predictor of treatment as a function of the IV replaces observed treatment in the treatment–outcome model. This two-stage logistic regression IV approach was applied to observational studies and compared with other IV methods, such as the probit structural equation model and a generalized method of moment (GMM) IV approach [31]. Alternatively, Nagelkerke *et al.* [32] and Terza *et al.* [30] offered an approach where the treatment–outcome model includes a residual term from the treatment-IV model (two-stage residual inclusion (2SRI)). The 2SRI procedure is equivalent to the 2SPS approach under the linear model, but this is not the case under the logistic model. Terza *et al.* [30] showed analytical and simulation-based differences under a true model for the causal effect of treatment conditional on the unmeasured confounder.

Given the focus of much of the clinical trials literature on the causal effect of treatment in C, there is a need for assessment of the 2SPS and 2SRI two-stage logistic estimators with respect to this causal effect. We present analytical and simulation results for the bias of these two estimators under a causal logistic model expressed in terms of potential outcomes under the principal stratification framework,



Copyright © 2011 John Wiley & Sons, Ltd.

following the results of Angrist *et al.* [4] for the additive model. We also confirm our analytic result with simulations, and the simulations further reveal patterns of bias for different ranges of confoundings. Our bias evaluation is for a different context from that of Terza *et al.* [30], who focused on the causal odds ratio in the total population conditional on the unmeasured confounder, whereas we focus on the causal odds ratio among Cs.

## 2. Assumption and notation

We have the same five assumptions as Angrist *et al.* stated in their causal model [4]: (i) stable unit treatment value assumption (SUTVA) [33, 34], which means that the potential outcome for each person is unrelated to the treatment status of other individuals; this assumption also implies the consistency assumption, which means that the potential outcome of a certain treatment will be the same regardless of the treatment assignment mechanism [35]; (ii) random assignment assumption, which means that the IV is unrelated, as the randomized assignment, to all confounders in the randomized clinical trials, or it is unrelated to the unmeasured confounders (conditional on the measured confounders) of the treatment–outcome relationship in observational studies; (iii) exclusion restriction, which means that any effect of treatment assignment on outcomes must be via an effect of treatment assignment on treatment received; (iv) non-zero average causal effect of treatment assignment on treatment received, which means that the treatment assignment should be associated with treatment received; and (v) monotonicity, which means that there is no one who would do the opposite of his/her treatment assignment regardless of the actual assignment.

With the above five assumptions, we first define $R$ and $Z$ as the treatment assignment and treatment received variables, respectively. First, $R=1$ denotes that a patient is assigned to the study treatment, and $R=0$ means a patient is assigned to the other treatment (or non-treatment), thus $R$ is the IV. Similarly, $Z=1$ means that a patient receives the study treatment, and $Z=0$ means that a patient receives the other treatment (or non-treatment). Additionally, $Y^{(1)}$ and $Y^{(0)}$ are the variables for potential outcomes. $Y^{(1)}$ indicates what the outcome for a patient would be if this patient were to take the study treatment, and $Y^{(0)}$ indicates what the outcome for this patient would be if he/she were to take the other treatment (or non-treatment). In contrast, $Y$ is the variable for the observed outcome. Similarly, $Z^{(1)}$ and $Z^{(0)}$ are the variables for potential treatment. $Z^{(1)}$ indicates what treatment a patient would take if this patient were assigned to the study treatment, and $Z^{(0)}$ indicates what treatment this patient would take if he/she were assigned to the other treatment (or non-treatment). Based on the principal stratification and potential outcome framework, patients are defined as ATs if $Z^{(1)}=1$ and $Z^{(0)}=1$; Cs if $Z^{(1)}=1$ and $Z^{(0)}=0$; NTs if $Z^{(1)}=0$ and $Z^{(0)}=0$; and DFs if $Z^{(1)}=0$ and $Z^{(0)}=1$.

Accordingly, we define the following parameters in the principal stratification framework:

$$\begin{aligned}
\omega_A^1 &= \Pr(Y^{(1)}=1|AT),\\
\omega_C^1 &= \Pr(Y^{(1)}=1|C),\\
\omega_N^1 &= \Pr(Y^{(1)}=1|NT),\\
\omega_A^0 &= \Pr(Y^{(0)}=1|AT),\\
\omega_C^0 &= \Pr(Y^{(0)}=1|C),\\
\omega_N^0 &= \Pr(Y^{(0)}=1|NT),\\
r &= \Pr(R=1),\\
\rho_A &= \Pr(AT),\\
\rho_C &= \Pr(C).
\end{aligned}$$

With our monotonicity assumption, there are no DFs [4], i.e. $\Pr(DF)=0$. Hence,

$$\Pr(NT)=\rho_N=1-\rho_A-\rho_C.$$



Copyright © 2011 John Wiley & Sons, Ltd.

The causal log odds ratio for C is parameterized as

$$\begin{aligned}\psi &= \text{logit}[\Pr(Y^{(1)}=1|C)] - \text{logit}[\Pr(Y^{(0)}=1|C)] \\ &= \text{logit}(\omega_C^1) - \text{logit}(\omega_C^0).\end{aligned}$$

The parameter $\psi$ is the log of the odds ratio that compares the probability of $Y=1$ if all Cs received the study treatment to the probability of $Y=1$ if all Cs received the other treatment (or no treatment).

## 3. Bias of 2SPS

In this section, we derive a closed-form expression for the probability limit of the 2SPS logistic regression estimator based on the principal stratification framework and assumptions. We can then obtain closed-form expressions for the bias, which is the difference between the expected value of the two-stage regression estimator and the causal log odds ratio.

### *3.1. Probability limit of the estimator*

The first stage regression is the treatment received on the treatment assignment $R$ as the IV. Let $D=E(Z|R)$ and $\hat{D}$ be an estimator of $D$ (e.g. maximum likelihood) such that $\hat{D}$ converges in probability to $D$, $\hat{D}=\hat{E}(Z|R)$. Two-stage logistic regression estimates the causal log odds ratio with the coefficient for $\hat{D}$ in the logistic regression of $Y$ on $\hat{D}$. Let $\hat{\xi}$ be an estimator (e.g. maximum likelihood) of the log odds ratio for $D$ in the logistic regression of $Y$ on $D$, and let $\hat{\xi}^*$ be the estimator of the log odds ratio for $\hat{D}$ in the logistic regression of $Y$ on $\hat{D}$ (i.e. the 2SPS estimator). As the sample size gets larger, $\hat{D} \longrightarrow D$ and $|\hat{\xi}^* - \hat{\xi}| \xrightarrow{p} 0$ [36, 37], i.e. $\hat{\xi}^*$ converges in probability to $\xi$ under the true model conditional on $D$, which is $P(Y=1|D)=\text{expit}(\eta+\xi D)$. We now find an expression for $\xi$ as a function of the log odds ratio for treatment received among C under the principal stratification framework.

When $R=0$, only ATs will receive the treatment; when $R=1$, both ATs and Cs will get the treatment. It follows that

$$d_0 = E(Z|R=0) = \rho_A \tag{1}$$

and

$$d_1 = E(Z|R=1) = \rho_A + \rho_C. \tag{2}$$

Then for the second stage logistic regression we have

$$\begin{aligned}&\text{logit}\Pr(Y=1|R=0) \\ &\quad = \text{logit}\Pr(Y=1|D=d_0) \\ &\quad = \eta + \xi d_0, \\ &\text{logit}\Pr(Y=1|R=1) \\ &\quad = \text{logit}\Pr(Y=1|D=d_1) \\ &\quad = \eta + \xi d_1.\end{aligned}$$

Solving the above two equations for $\xi$, we have

$$\xi = \frac{\text{logit}\Pr(Y|R=1) - \text{logit}\Pr(Y|R=0)}{d_1 - d_0}.$$

Under the five assumptions stated in Section 2 and the above parameter settings, the probability of observed $Y$ given $R$ can be expressed as the conditional probability of potential outcome $Y^{(0)}$ and $Y^{(1)}$. We can then calculate $\Pr(Y|R=1)$ and $\Pr(Y|R=0)$ as follows:

$$\begin{aligned}\text{logit}\Pr(Y|R=1) &= \text{logit}(\rho_A\omega_A^1 + \rho_C\omega_C^1 + \omega_N^0 - \rho_A\omega_N^0 - \rho_C\omega_N^0), \\ \text{logit}\Pr(Y|R=0) &= \text{logit}(\rho_A\omega_A^1 + \rho_C\omega_C^0 + \omega_N^0 - \rho_A\omega_N^0 - \rho_C\omega_N^0).\end{aligned}$$



Copyright © 2011 John Wiley & Sons, Ltd.

The full proof of these equations is in Appendix A1. From the above equation, we can calculate $\xi$ as follows:

$$\xi = \frac{\text{logit}\Pr(Y|R=1) - \text{logit}\Pr(Y|R=0)}{d_1 - d_0}$$
$$= \frac{\text{logit}(\rho_A\omega_A^1 + \rho_C\omega_C^1 + \omega_N^0 - \rho_A\omega_N^0 - \rho_C\omega_N^0) - \text{logit}(\rho_A\omega_A^1 + \rho_C\omega_C^0 + \omega_N^0 - \rho_A\omega_N^0 - \rho_C\omega_N^0)}{\rho_C}. \quad (3)$$

Since $\hat{\xi}$ converges in probability to $\xi$, equation (3) is a closed-form expression for the probability limit of the two-stage logistic regression estimator of $\hat{\xi}$.

### 3.2. Bias analysis

Having derived the closed-form expression of $\xi$, we can calculate the difference between $\psi$ and $\xi$, the asymptotic bias of the two-stage logistic regression.

$$B_{2\text{SPS}} = \xi - \psi$$
$$= \frac{1}{\rho_C}\begin{pmatrix} \text{logit}(\rho_A\omega_A^0 + \rho_C\omega_C^1 + \omega_N^0 - \rho_A\omega_N^0 - \rho_C\omega_N^0) \\ -\text{logit}(\rho_A\omega_A^0 + \rho_C\omega_C^0 + \omega_N^0 - \rho_A\omega_N^0 - \rho_C\omega_N^0) \end{pmatrix} - (\text{logit}(\omega_C^1) - \text{logit}(\omega_C^0))$$
$$= \frac{1}{\rho_C}\begin{pmatrix} \text{logit}(\rho_A\omega_A^0 + \rho_C\omega_C^1 + \text{expit}(\text{logit}(\omega_C^0) + \delta)\rho_N) \\ -\text{logit}(\rho_A\omega_A^0 + \rho_C\omega_C^0 + \text{expit}(\text{logit}(\omega_C^0) + \delta)\rho_N) \end{pmatrix} - (\text{logit}(\omega_C^1) - \text{logit}(\omega_C^0)). \quad (4)$$

In the above equation, we re-parameterize the $\omega_N^0$ and introduce a new parameter $\delta$ as follows:

$$\text{logit}(\omega_N^0) = \text{logit}(\omega_C^0) + \delta,$$

then

$$\omega_N^0 = \text{expit}(\text{logit}(\omega_C^0) + \delta) = \omega_C^0 \frac{e^\delta}{\omega_C^0 e^\delta - \omega_C^0 + 1}.$$

The parameter $\delta$ is the difference between $\omega_N^0$ and $\omega_C^0$ on the logit scale, so it is the log odds ratio of NTs over Cs regarding the outcome. Given that the differences between principal strata are due to unmeasured confounders related to outcome, $\delta$ in equation (4) can be interpreted as the magnitude of confounding, where $\delta = 0$ implies no confounding because $\omega_N^0 = \omega_C^0$.

From equation (4), we can easily see that

(a) When $\rho_C = 1$ (everyone is a C in a randomized controlled trial with perfect adherence), $B_{2\text{SPS}} = 0$. This is because when $\rho_C = 1$, both $\rho_A$ and $\rho_N$ are 0. In equation (4), if we replace $\rho_C$ by 1 and both $\rho_A$ and $\rho_N$ by 0, we have $B_{2\text{SPS}} = 0$.
(b) When $\omega_C^1 = \omega_C^0$ (there is no causal effect), $B_{2\text{SPS}} = 0$. If we replace $\omega_C^1$ by $\omega_C^0$ in equation (4), all terms are canceled out and we have $B_{2\text{SPS}} = 0$.
(c) The bias function does not include $R$, thus bias is not related to $\Pr(R=1)$.
(d) Bias can exist even when there is no confounding, that is, when $\rho_A = 0$ and $\omega_C^0 = \omega_N^0$. Replacing $\rho_A$ by 0 in equation (4), we have

$$B_{2\text{SPS}} = \frac{\begin{matrix}\text{logit}(\rho_A\omega_A^1 + \rho_C\omega_C^1 + \omega_N^0 - \rho_A\omega_N^0 - \rho_C\omega_N^0) \\ -\text{logit}(\rho_A\omega_A^1 + \rho_C\omega_C^0 + \omega_N^0 - \rho_A\omega_N^0 - \rho_C\omega_N^0)\end{matrix}}{\rho_C} - \text{logit}(\omega_C^1) + \text{logit}(\omega_C^0)$$
$$= \frac{\text{logit}(\rho_C\omega_C^1 + \omega_N^0 - \rho_C\omega_N^0) - \text{logit}(\omega_N^0)}{\rho_C} - \text{logit}(\omega_C^1) + \text{logit}(\omega_C^0).$$



Copyright © 2011 John Wiley & Sons, Ltd.

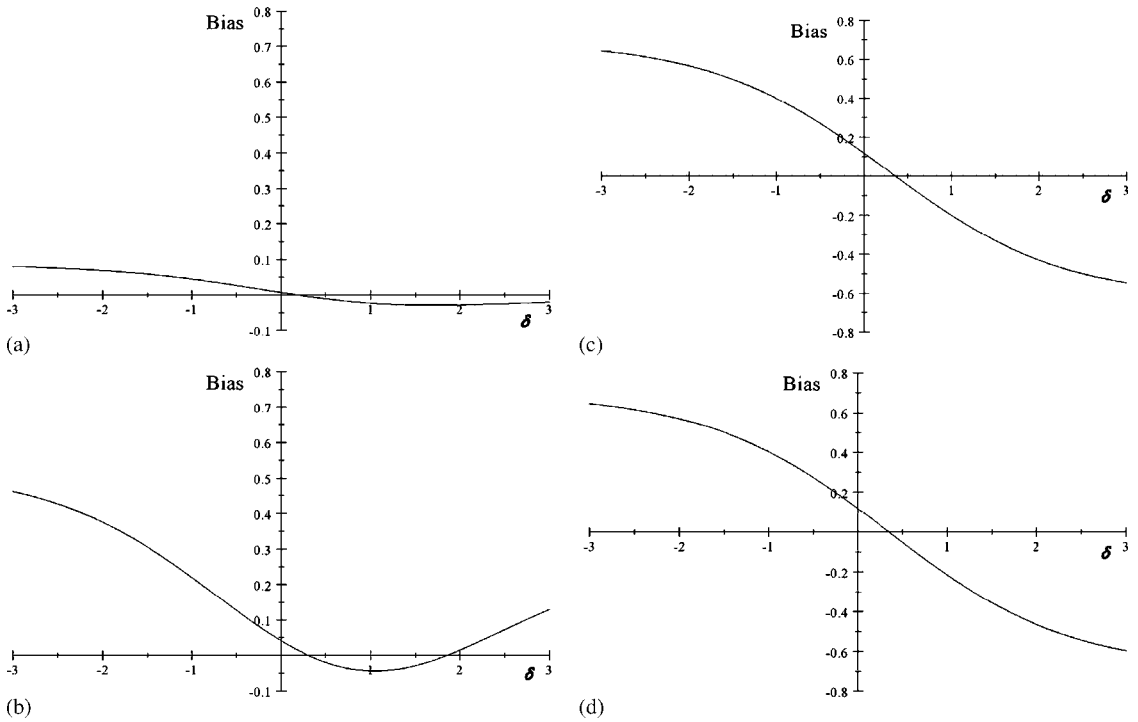


**Figure 1**. Plot of bias on magnitude of confounding $\delta$ with 2SPS approach: (a) $\rho_A=0$, $\rho_C=0.8$, $\omega_C^1=0.6$, $\omega_C^0=0.3$; (b) $\rho_A=0$, $\rho_C=0.5$, $\omega_C^1=0.6$, $\omega_C^0=0.3$; (c) $\rho_A=0$, $\rho_C=0.5$, $\omega_C^1=0.06$, $\omega_C^0=0.03$; and (d) $\rho_A=0$, $\rho_C=0.5$, $\omega_C^1=0.006$, $\omega_C^0=0.003$.

In this equation, $B_{2SPS}$ is generally not 0, because $\rho_C$ in the denominator cannot be canceled out with the $\rho_C$ in the logit function of the numerator. The no AT condition occurs when patients in a trial cannot possibly have access to the treatment without being assigned to that treatment. Since NTs cannot get the study treatment, confounding occurs only when there is a difference between the probability of outcome of Cs and that of NTs when they are not given the treatment. Thus, there is no confounding when $\rho_A=0$ and $\omega_C^0=\omega_N^0$.

With the closed-form expression (4), we can analyze the magnitude of bias under different parameter settings according to specific studies. To simplify the analysis and show the relationship between bias and confounding, we create four such scenarios when there are no ATs. We plot bias against $\delta$ while fixing all other parameters (Figure 1(a)–(d)).

All four plots show that the bias is not 0 when there is no confounding ($\delta=0$). When the compliance rate decreases from 0.8 to 0.5, the bias on the logit scale is about five times larger (compare plot 1a and plot 1b). Comparing plot 1b and plot 1c, we can see that when the event rate is lower, the bias range is larger, but when the event rate is decreased from 0.03 to 0.003, the absolute bias does not increase further.

## 4. Bias of 2SRI

In this section, we extend to the 2SRI estimator, the derivation in Section 3 of bias of the 2SPS under the principal stratification framework. In the first stage regression of treatment received on the treatment assignment $R$ as an IV, the residual is $E=Z-E(Z|R)$, and the second stage regression model is

$$\Pr(Y=1)=\text{expit}(\lambda_0+\lambda_1 Z+\lambda_2 E). \tag{5}$$

The estimator of $\lambda_1$ is an estimate of the causal log odds ratio for receiving treatment among Cs. We derive a closed-form expression for the probability limit of the estimator of $\lambda_1$. This enables us to derive a closed-form expression for the asymptotic difference between the probability limit of the estimator of $\lambda_1$ and the causal log odds ratio among Cs.



Copyright © 2011 John Wiley & Sons, Ltd.

### 4.1. Closed-form expression for the probability limit of the estimator

For the 2SRI approach, in general, equation (5) is not the true model for $\Pr(Y=1|Z,E)$, as the true model includes the interaction term between $Z$ and $E$; this makes it much more difficult to develop a closed-form expression for the probability limit of the estimator. However, if we assume that there are no ATs, so that $\Pr(Z=1,R=0)=0$, then the true model does not have the interaction term and the 2SRI model in equation (5) is the true model (see the details in Appendix A2). In this section, we develop a closed-form expression for the probability limit of the estimator of $\lambda_1$ only under the no AT assumption. The no AT assumption is true in clinical trials when patients in the placebo group cannot access the study drug. In contrast, the bias results for the 2SPS estimator depend on a true model conditional on just $Z$ (treatment-received) that does not require the absence of ATs.

The residual $E=Z-E(Z|R)$ is estimated from the first stage regression, and is included as a covariate in the second stage regression. Letting $\hat{E}=Z-\hat{E}(Z|R)$, we consider the second stage regression $\Pr(Y=1|Z,\hat{E})=\text{expit}(\lambda_0+\lambda_1 Z+\lambda_2\hat{E})$. The 2SRI approach estimates the causal log odds ratio with the estimated coefficient for $Z$ in the logistic regression of $Y$ on $Z$ and $\hat{E}$. Let $\hat{\lambda}_1$ denote the estimated coefficient for $Z$ in the logistic regression of $Y$ on $Z$ and $E$, and let $\hat{\lambda}_1^*$ denote the estimated coefficient for $Z$ in the logistic regression of $Y$ on $Z$ and $\hat{E}$. As the sample size gets larger, $\hat{E}\longrightarrow E$ and $|\hat{\lambda}_1^*-\hat{\lambda}_1|\xrightarrow{p}0$ [36, 37]. The estimator $\hat{\lambda}_1^*$ converges in probability to $\lambda_1$ under the model $\Pr(Y=1|Z,E)=\text{expit}(\lambda_0+\lambda_1 Z+\lambda_2 E)$ when there are no ATs. When there are ATs, the 2SRI model is misspecified. In this situation, $\hat{\lambda}_1^*$ estimated from the second stage logistic regression converges to the point that minimizes the Kullback–Leibler distance between the family of probability distributions being maximized over the true probability distribution [38].

Under the no AT assumption, we can find an expression for $\lambda_1$ as follows. From equations (1) and (2), we have

$$E(Z|R)=\rho_A+\rho_C R,$$

so

$$E=Z-E(Z|R)=Z-\rho_A-\rho_C R.$$

Note that $Z,E$ and $Z,R$ contain the same information; i.e. knowing $Z,E$ tells us $Z,R$ and vice versa, so that $\Pr(Y=1|Z,E)=Pr(Y=1|Z,R)$. For the second stage regression, we have

$$\begin{aligned}\text{logit}\Pr(Y=1|Z,E)&=\lambda_0+\lambda_1 Z+\lambda_2 E\\&=\lambda_0+\lambda_1 Z+\lambda_2(Z-\rho_A-\rho_C R)\\&=\lambda_0-\lambda_2\rho_A+(\lambda_1+\lambda_2)Z-\lambda_2\rho_C R\\&=\text{logit}\Pr(Y=1|Z,R).\end{aligned}\tag{6}$$

Then we have three equations based on the possible values of $Z$ and $R$ ($(Z=1,R=0)$ is not possible because there are no ATs):

$$\begin{aligned}\text{logit}\Pr(Y=1|Z=1,R=1)&=\text{logit}\Pr(Y^{(1)}=1|Z=1,R=1)\\&=\text{logit}\left(\frac{\rho_A}{\rho_A+\rho_C}\omega_A^1+\frac{\rho_C}{\rho_A+\rho_C}\omega_C^1\right)\\&=\lambda_0-\lambda_2\rho_A+(\lambda_1+\lambda_2)-\lambda_2\rho_C,\end{aligned}\tag{7}$$

$$\begin{aligned}\text{logit}\Pr(Y=1|Z=0,R=1)&=\text{logit}\Pr(Y^{(0)}=1|Z=0,R=1)\\&=\text{logit}\Pr(Y^{(0)}=1|NT)\\&=\text{logit}(\omega_N^0)\\&=\lambda_0-\lambda_2\rho_A-\lambda_2\rho_C,\end{aligned}\tag{8}$$



Copyright © 2011 John Wiley & Sons, Ltd.

$$\begin{aligned}\operatorname{logit}\Pr(Y=1|Z=0,R=0) &= \operatorname{logit}\Pr(Y^{(0)}=1|Z=0,R=0)\\ &= \operatorname{logit}\left(\frac{1-\rho_A-\rho_C}{1-\rho_A}\omega_N^0+\frac{\rho_C}{1-\rho_A}\omega_C^0\right)\\ &= \lambda_0-\lambda_2\rho_A. \end{aligned} \tag{9}$$

Solving equations (7), (8), and (9) for $\lambda_1$ yields the closed-form expression for $\lambda_1$ as

$$\begin{aligned}\lambda_1 = &\operatorname{logit}\left(\frac{\rho_A}{\rho_A+\rho_C}\omega_A^0+\frac{\rho_C}{\rho_A+\rho_C}\omega_C^1\right)-\operatorname{logit}(\omega_N^0)\\ &-\frac{1}{\rho_C}\operatorname{logit}\left(\frac{1-\rho_A-\rho_C}{1-\rho_A}\omega_N^0+\frac{\rho_C}{1-\rho_A}\omega_C^0\right)+\frac{1}{\rho_C}\operatorname{logit}(\omega_N^0). \end{aligned} \tag{10}$$

### 4.2. *Bias analysis*

With the closed-form expression for the probability limit of $\hat{\lambda}_1$, we can calculate $B_{2\text{SRI}}$, the bias defined as the difference between the log odds ratio for treatment-received among Cs and the estimated log odds ratio with the 2SRI approach.

$$\begin{aligned}B_{2\text{SRI}} &= \lambda_1-\psi\\ &= \operatorname{logit}\left(\frac{\rho_A}{\rho_A+\rho_C}\omega_A^1+\frac{\rho_C}{\rho_A+\rho_C}\omega_C^1\right)-\operatorname{logit}(\omega_N^0)\\ &\quad -\frac{1}{\rho_C}\operatorname{logit}\left(\frac{1-\rho_A-\rho_C}{1-\rho_A}\omega_N^0+\frac{\rho_C}{1-\rho_A}\omega_C^0\right)+\frac{1}{\rho_C}\operatorname{logit}(\omega_N^0)-\operatorname{logit}(\omega_C^1)+\operatorname{logit}(\omega_C^0)\\ &= \operatorname{logit}\left(\frac{\rho_A}{\rho_A+\rho_C}\omega_A^1+\frac{\rho_C}{\rho_A+\rho_C}\omega_C^1\right)-\operatorname{logit}(\operatorname{expit}(\operatorname{logit}(\omega_C^0)+\delta))\\ &\quad -\frac{1}{\rho_C}\operatorname{logit}\left(\frac{1-\rho_A-\rho_C}{1-\rho_A}(\operatorname{expit}(\operatorname{logit}(\omega_C^0)+\delta))+\frac{\rho_C}{1-\rho_A}\omega_C^0\right)+\frac{1}{\rho_C}\operatorname{logit}(\operatorname{expit}(\operatorname{logit}(\omega_C^0)+\delta))\\ &\quad -\operatorname{logit}(\omega_C^1)+\operatorname{logit}(\omega_C^0). \end{aligned} \tag{11}$$

$\delta$ is the same parameter as in equation (4). The following conclusions follow from equation (11):

(a) When $\rho_C=1$ (everyone is a C), $B_{2\text{SRI}}=0$. If $\rho_C=1$, both $\rho_A$ and $\rho_N$ equal 0. Plugging in these values of $\rho_C$, $\rho_A$, and $\rho_N$ to equation (11), $B_{2\text{SRI}}=0$. $\rho_C=1$ can only occur in a randomized control trial with perfect adherence.
(b) When $\omega_C^0=\omega_N^0$ and $\omega_A^1=\omega_C^1$ (there is no confounding), we replace $\omega_N^0$ with $\omega_C^0$ and $\omega_A^1$ with $\omega_C^1$ in equation (11), yielding $B_{2\text{SRI}}=0$. That is, when there is no confounding, the 2SRI approach is unbiased.

As in Section 3 with the 2SPS estimator, we use equation (11) to analyze the magnitude of bias of the 2SRI estimator under different scenarios as follows.

All four plots (Figures 2(a)–(d)) show that when there is no confounding ($\delta=0$), the bias of the 2SRI estimator is zero. The first scenario shows that when the compliance rate is high (0.8), the bias is small for a wide range of confounding. The second scenario shows that if the outcome is not rare, the bias is very small unless $\delta$ is smaller than $-1$ or greater than 2, which means that the odds ratio comparing Cs to NTs with respect to the potential outcomes is smaller than 0.37 or greater than 7.4. These scenarios correspond to very strong confounding. Figure 2(c) shows the scenario when the outcome is rare, with $\omega_C^1$ and $\omega_C^0$ one-tenth of those in scenario 1, The bias for this scenario is larger than that of scenario 1, but the bias is still moderate if the confounding is not very severe. In scenario 4, we make the outcome even rarer. The magnitude of bias does not change much compared to the bias under scenario 3. Therefore, we can conclude that for the 2SRI model, there is bias when there is confounding, but the bias is small to moderate if the confounding is not severe.



Copyright © 2011 John Wiley & Sons, Ltd.

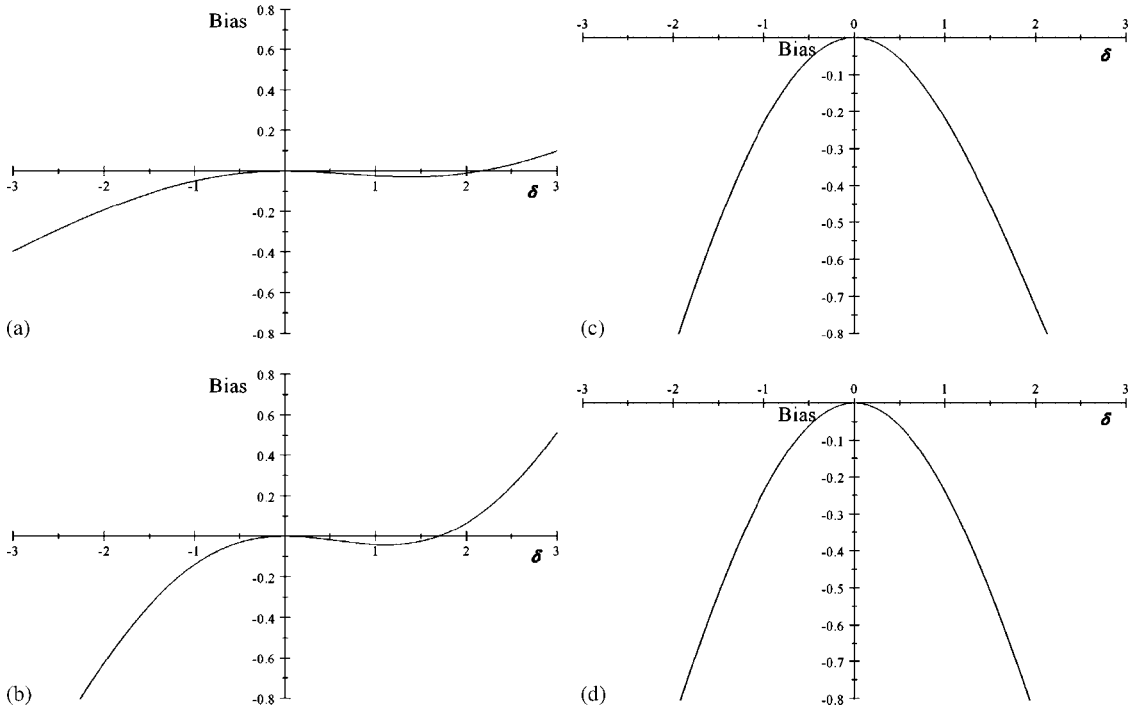


**Figure 2**. Plot of bias on magnitude of confounding $\delta$ with 2SRI approach: (a) $\rho_A=0$, $\rho_C=0.8$, $\omega_C^1=0.6$, $\omega_C^0=0.3$; (b) $\rho_A=0$, $\rho_C=0.5$, $\omega_C^1=0.6$, $\omega_C^0=0.3$; (c) $\rho_A=0$, $\rho_C=0.5$, $\omega_C^1=0.06$, $\omega_C^0=0.03$; and (d) $\rho_A=0$, $\rho_C=0.5$, $\omega_C^1=0.006$, $\omega_C^0=0.003$.

## 5. Simulation

### *5.1. Simulation algorithm*

We simulated the data sets according to the following algorithm:

*Step 1*: Generate a data set with a total number of $N$ subjects. Among these subjects, ATs, Cs, and NTs are generated from a multinomial distribution with probability of $\rho_A$ for ATs, probability of $\rho_C$ for Cs and probability of $\rho_N$ for NTs. With the statistical programming package R, this step can be implemented by $W=t(\text{rmultinom}(n,1,c(\rho_A,\rho_C,\rho_N)))$.

*Step 2*: With the probability of $\Pr(R=1)=r$, randomly assign about $rN$ of the subjects to $R=1$ and the rest of $(1-r)N$ subject to $R=0$. This step can be implemented by $R=t(\text{rmultinom}(n,1,c(r,1-r)))$ in the package R.

*Step 3*: Simulate $Y^{(0)}$ and $Y^{(1)}$ based on the value of AT, C, or NT, and the parameter $\omega_A^1$, $\omega_C^1$, $\omega_N^1$, $\omega_A^0$, $\omega_C^0$, and $\omega_N^0$. For instance, if a subject is AT, then $Pr(Y^{(0)}=1)=\omega_A^0$ and $Pr(Y^{(1)}=1)=\omega_A^1$. With these probabilities, we can create $Y^{(1)}$ and $Y^{(0)}$ with the binomial distribution. We implemented this step in the package R with the following program:

```
prY0=W[, 1]*ω_A^0+W[, 2]*ω_C^0+W[, 3]*ω_N^0
dim(prY0)=c(n, 1)
prY1=W[, 1]*ω_A^1+W[, 2]*ω_C^1+W[, 3]*ω_N^1
dim(prY1)=c(n, 1)
Y0=apply(prY0, 1, function (x) rbinom(1,1,x))
Y1=apply(prY1, 1, function (x) rbinom(1,1,x))
```

*Step 4*: Based on AT, C, or NT, and R, determine $Z$. For instance, if an observation is in either the AT or C group, and the treatment assignment $R=1$, then $Z=1$.

*Step 5*: Based on $Z$ , $Y^{(0)}$, and $Y^{(1)}$, determine $Y$

$Y=Y^{(1)}Z+Y^{(0)}(1-Z)$.



Copyright © 2011 John Wiley & Sons, Ltd.

**Table I**. Comparison of simulation result and analytic result when there are no always-takers.

| | | | | 2SPS | | | | 2SRI | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\omega_C^0$ | $\omega_C^1$ | True LogOR | $\delta$ | Log OR by regression | Observed bias | Analytic result of bias | MSE | LogOR by regression | Observed bias | Analytic result of bias | MSE |
| 0.3 | 0.60 | 1.2528 | −2.0 | 1.6295 | 0.3768 | 0.3754 | 0.1500 | 0.6256 | −0.6272 | −0.6266 | 0.4095 |
| | | | −1.5 | 1.5601 | 0.3073 | 0.3061 | 0.1024 | 0.9112 | −0.3416 | −0.3415 | 0.1295 |
| | | | −1.0 | 1.4740 | 0.2213 | 0.2200 | 0.0567 | 1.1127 | −0.1400 | −0.1410 | 0.0301 |
| | | | −0.5 | 1.3813 | 0.1286 | 0.1263 | 0.0238 | 1.2244 | −0.0284 | −0.0309 | 0.0095 |
| | | | 0.0 | 1.2961 | 0.0433 | 0.0405 | 0.0088 | 1.2559 | 0.0031 | 0.0000 | 0.0075 |
| | | | 0.5 | 1.2362 | −0.0166 | −0.0200 | 0.0069 | 1.2383 | −0.0145 | −0.0179 | 0.0071 |
| | | | 1.0 | 1.2079 | −0.0449 | −0.0435 | 0.0090 | 1.2103 | −0.0425 | −0.0413 | 0.0088 |
| | | | 1.5 | 1.2228 | −0.0300 | −0.0289 | 0.0081 | 1.2268 | −0.0259 | −0.0250 | 0.0079 |
| | | | 2.0 | 1.2666 | 0.0138 | 0.0145 | 0.0080 | 1.3172 | 0.0644 | 0.0651 | 0.0123 |
| 0.03 | 0.0600 | 0.7246 | −2.0 | 1.2894 | 0.5648 | 0.5666 | 0.3901 | −0.1732 | −0.8978 | −0.8474 | 0.9745 |
| | | | −1.5 | 1.2215 | 0.4969 | 0.4973 | 0.3131 | 0.2011 | −0.5235 | −0.5015 | 0.3865 |
| | | | −1.0 | 1.1225 | 0.3980 | 0.3994 | 0.2181 | 0.4788 | −0.2458 | −0.2314 | 0.1432 |
| | | | −0.5 | 0.9900 | 0.2654 | 0.2709 | 0.1232 | 0.6522 | −0.0724 | −0.0589 | 0.0666 |
| | | | 0.0 | 0.8374 | 0.1128 | 0.1175 | 0.0585 | 0.7161 | −0.0084 | 0.0000 | 0.0485 |
| | | | 0.5 | 0.6770 | −0.0475 | −0.0459 | 0.0387 | 0.6630 | −0.0616 | −0.0571 | 0.0406 |
| | | | 1.0 | 0.5198 | −0.2048 | −0.2005 | 0.0705 | 0.5002 | −0.2243 | −0.2169 | 0.0790 |
| | | | 1.5 | 0.3911 | −0.3334 | −0.3310 | 0.1335 | 0.2658 | −0.4587 | −0.4525 | 0.2339 |
| | | | 2.0 | 0.2932 | −0.4314 | −0.4306 | 0.2026 | −0.0107 | −0.7352 | −0.7297 | 0.5593 |

*Note*: The probability of always-takers $\rho_A = 0$, the probability of compliers $\rho_C = 0.5$, and the probability of never-takers $\rho_N = 0.5$.

### 5.2. *Simulation results*

For each setting, we ran the simulation 2000 times, with the sample size of $n = 10000$. For both 2SPS and 2SRI approaches, we simulated data with different selection of parameters. As examples, Table I shows the results with the parameter settings without ATs: $\rho_A = 0$; $\rho_C = 0.5$ (thus $\rho_N = 0.5$); $\omega_C^0 = 0.3$ or $\omega_C^0 = 0.03$; $\omega_C^1 = 0.6$ or $\omega_C^1 = 0.06$; $\delta$ varies among 2, 1.5, 1, 0.5, 0, −0.5, −1, −1.5, or −2. For these simulations, the bias is calculated as the difference between the mean of estimated log odds ratio ($\hat{\xi}$ for 2SPS and $\hat{\lambda}_1$ for 2SRI) and the log odds ratio among Cs $\psi$. The mean square of error (MSE) is calculated as the mean square of the difference between the estimated log odds ratio and the log odds ratio among Cs.

Under all parameter settings without ATs, the bias resulting from simulations is consistent with the analytic results, and when there is no confounding, the bias is not zero for 2SPS but is zero for 2SRI (Table I). The simulation results of MSE follow the same pattern as the results for absolute bias with these large sample simulations. We are currently doing further research on the MSE properties of the different estimators.

We also performed simulations including ATs with the parameter settings: $\rho_A = 0.2$; $\rho_C = 0.5$ (thus $\rho_N = 0.3$); $\omega_C^0 = 0.3$ or $\omega_C^0 = 0.03$; $\omega_C^1 = 0.6$ or $\omega_C^1 = 0.06$; $\delta$ varies among 2, 1.5, 1, 0.5, 0, −0.5, −1, −1.5, or −2. Under these parameter settings, the analytic results are available for the 2SPS procedure, but are not possible for the 2SRI approach as discussed in Section 4. As shown in Table II, the bias from simulated data is consistent with the analytic results for the 2SPS approach when there are ATs. For 2SRI, the results show that the bias is smaller than for 2SPS, and is close to 0 when $\delta$ is 0, but for some parameter settings with strong confounding, the bias is larger than for 2SPS.

## 6. Discussion

The IV approach has been applied to logistic regression to control for unmeasured confounding in estimating treatment effects under non-adherence in randomized trials and under actual medical care in observational studies. However, there has been little if no evaluation of the bias of this use of IV in the context of estimating the effect of treatment among those who are Cs or take the treatment. Accordingly, we have developed closed-form expressions for the asymptotic bias of the 2SRI and 2SPS approaches to two-stage logistic regression, and we have shown that these analytic results are



Copyright © 2011 John Wiley & Sons, Ltd.

**Table II**. Comparison of simulation result and analytic result when there are always-takers.

| | | | | 2SPS | | | | 2SRI | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\omega_C^0$ | $\omega_C^1$ | True LogOR | $\delta$ | Log OR by regression | Observed bias | Analytic result of bias | MSE | LogOR by regression | Observed bias | Analytic result of bias | MSE |
| 0.3 | 0.60 | 1.2528 | −2.0 | 1.3159 | 0.0631 | 0.0615 | 0.0098 | 1.2554 | 0.0026 | NA | 0.0090 |
| | | | −1.5 | 1.3007 | 0.0480 | 0.0461 | 0.0081 | 1.2624 | 0.0096 | NA | 0.0085 |
| | | | −1.0 | 1.2809 | 0.0281 | 0.0257 | 0.0065 | 1.2677 | 0.0149 | NA | 0.0079 |
| | | | −0.5 | 1.2574 | 0.0046 | 0.0016 | 0.0057 | 1.2668 | 0.0140 | NA | 0.0074 |
| | | | 0.0 | 1.2338 | −0.0190 | −0.0220 | 0.0061 | 1.2559 | 0.0031 | NA | 0.0066 |
| | | | 0.5 | 1.2167 | −0.0361 | −0.0389 | 0.0073 | 1.2380 | −0.0148 | NA | 0.0067 |
| | | | 1.0 | 1.2112 | −0.0416 | −0.0434 | 0.0083 | 1.2221 | −0.0306 | NA | 0.0077 |
| | | | 1.5 | 1.2201 | −0.0327 | −0.0346 | 0.0077 | 1.2216 | −0.0311 | NA | 0.0076 |
| | | | 2.0 | 1.2393 | −0.0135 | −0.0162 | 0.0071 | 1.2410 | −0.0118 | NA | 0.0071 |
| 0.03 | 0.0600 | 0.7246 | −2.0 | 0.8826 | 0.1580 | 0.1583 | 0.0753 | 0.9577 | 0.2331 | NA | 0.1092 |
| | | | −1.5 | 0.8623 | 0.1378 | 0.1390 | 0.0677 | 0.9177 | 0.1931 | NA | 0.0895 |
| | | | −1.0 | 0.8312 | 0.1067 | 0.1093 | 0.0578 | 0.8633 | 0.1387 | NA | 0.0677 |
| | | | −0.5 | 0.7880 | 0.0634 | 0.0652 | 0.0483 | 0.7983 | 0.0737 | NA | 0.0507 |
| | | | 0.0 | 0.7276 | 0.0030 | 0.0034 | 0.0410 | 0.7250 | 0.0005 | NA | 0.0413 |
| | | | 0.5 | 0.6471 | −0.0774 | −0.0766 | 0.0421 | 0.6443 | −0.0803 | NA | 0.0427 |
| | | | 1.0 | 0.5549 | −0.1696 | −0.1704 | 0.0598 | 0.5541 | −0.1705 | NA | 0.0600 |
| | | | 1.5 | 0.4575 | −0.2671 | −0.2683 | 0.0971 | 0.4389 | −0.2857 | NA | 0.1073 |
| | | | 2.0 | 0.3686 | −0.3560 | −0.3586 | 0.1472 | 0.2962 | −0.4284 | NA | 0.2042 |

*Note*: The probability of always-takers $\rho_A = 0.2$, the probability of compliers $\rho_C = 0.5$, and the probability of never-takers $\rho_N = 0.3$.

consistent with the simulation results under different parameter settings. Terza *et al.* [30] showed that the 2SRI approach is unbiased when the true model is conditional on the unmeasured confounder. For the treatment effect conditional on compliance or receiving treatment, Nagelkerke *et al.* [32] and Ten Have *et al.* [28] presented simulations showing that the bias of 2SRI approach increases as the magnitude of confounding increases. Our analytical and simulation results confirm such bias for the 2SRI as well as for the 2SPS approach. We further show that unlike the 2SRI approach, the 2SPS procedure is biased even when there is no unmeasured confounding.

An important contribution of this research is the expression of the conditional distribution of observed outcomes $Y$ given treatment assignment R as a function of the probability of compliance and the conditional distribution of potential outcomes $Y^{(0)}$ and $Y^{(0)}$, given compliance status. With this contribution, we can analytically present probability limits and therefore the bias of the estimators of the causal effects of treatment given compliance and treatment status. Further, we provide analytic estimates of bias for a variety of situations. These analytic estimates of bias can help researchers evaluate if the bias is small under specific conditions (e.g. high compliance and moderate confounding). Hence, our results can be used as a guide for deciding whether the 2SRI or 2SPS strategy is appropriate. This method can be potentially applied to the bias analysis of causal inference with other non-linear two-stage regressions, such as regressions of probit models and log linear models.

When the 2SRI or 2SPS is appropriately used, these approaches have the advantage that they are very easy to implement with any software package that can do logistic regression (e.g. SAS, R, or STATA). Logistic regression is used for both the first and second stages of either the 2SRI or 2SPS procedures. The predicted or residual values from the first stage logistic regression of treatment on the IV are used as covariates in the second stage logistic regression: the predicted value of treatment replaces the observed treatment for 2SPS, whereas the residual from the first stage regression is added as a covariate along with observed treatment for 2SRI.

The bias for both the 2SPS and 2SRI approaches occurs even when all the IV assumptions are met. Additional research is needed in resolving such bias, and also in assessing departures from the IV assumptions under the logistic IV model. To resolve the bias of the 2SRI and 2SPS approaches, the logistic structural nested mean model of Vansteelandt and Goetghebeur [39] in the randomized trial context when controls do not have access to the treatment can be extended to the observational data context when all subjects have access to treatment. Additionally, such a modeling approach may be modified to assess departures from the exclusion restriction using a similar weighted estimating



Copyright © 2011 John Wiley & Sons, Ltd.

equations approach as in Ten Have *et al.* [40]. Our bias analysis for the two-stage logistic regression can help researchers decide in which situations the bias of two stage logistic regression is small, in which case the two stage logistic regression maybe a reasonable method to use in contrast to more complicated methods.

## Appendix A

A1. Prove that the probability of observed $Y$ given R can be expressed by the following equations:

$$\Pr(Y|R=1)=\rho_A\omega_A^0+\rho_C\omega_C^1+\omega_N^0-\rho_A\omega_N^0-\rho_C\omega_N^0$$

and

$$\Pr(Y|R=0)=\rho_A\omega_A^0+\rho_C\omega_C^0+\omega_N^0-\rho_A\omega_N^0-\rho_C\omega_N^0.$$

In these equations, AT means always-taker, C means compiler, and NT means never-taker, and

$$\begin{aligned}
\omega_A^1 &= \Pr(Y^{(1)}=1|AT),\\
\omega_C^1 &= \Pr(Y^{(1)}=1|C),\\
\omega_N^1 &= \Pr(Y^{(1)}=1|NT),\\
\omega_A^0 &= \Pr(Y^{(0)}=1|AT),\\
\omega_C^0 &= \Pr(Y^{(0)}=1|C),\\
\omega_N^0 &= \Pr(Y^{(0)}=1|NT),\\
r &= \Pr(R=1),\\
\rho_A &= \Pr(AT),\\
\rho_C &= \Pr(C),\\
\rho_N &= \Pr(NT).
\end{aligned}$$

*Proof*

$$\begin{aligned}
\Pr(Y^{(1)}=1|Z=1,R=1) &= \Pr(Y^{(1)}=1,Z=1,R=1)/\Pr(Z=1,R=1)\\
&= \frac{\Pr(Y^{(1)}=1,AT,R=1)+\Pr(Y^{(1)}=1,C,R=1)}{\Pr(R=1,AT)+\Pr(R=1,C)}\\
&= \frac{\Pr(Y^{(1)}=1,AT)\Pr(R=1)+\Pr(Y^{(1)}=1,C)\Pr(R=1)}{\Pr(R=1)\Pr(AT)+\Pr(R=1)\Pr(C)}\\
&= \frac{\Pr(Y^{(1)}=1|AT)\Pr(AT)+\Pr(Y^{(1)}=1|C)\Pr(C)}{\Pr(R=1)\Pr(AT)+\Pr(R=1)\Pr(C)}\\
&= \frac{\Pr(AT)}{\Pr(AT)+\Pr(C)}\Pr(Y^{(1)}=1|AT)+\frac{\Pr(C)}{\Pr(AT)+\Pr(C)}\Pr(Y^{(1)}=1|C)\\
&= \frac{\rho_A}{\rho_A+\rho_C}\omega_A^1+\frac{\rho_C}{\rho_A+\rho_C}\omega_C^1.
\end{aligned}$$

Note: According to the assumptions of the IV, $R$ is independent of $Y^{(1)}$ and the principal stratum, thus in the above equation, $Pr(Y^{(1)}=1,AT,R=1)=Pr(Y^{(1)}=1,AT)Pr(R=1)$ and $Pr(Y^{(1)}=1,C,R=1)=Pr(Y^{(1)}=1,C)Pr(R=1)$.



Copyright © 2011 John Wiley & Sons, Ltd.

$$\begin{aligned}
&\Pr(Y^{(0)}=1|Z=0,R=0)\\
&=\frac{\Pr(NT)}{\Pr(NT)+\Pr(C)}\Pr(Y^{(0)}=1|NT)+\frac{\Pr(C)}{\Pr(NT)+\Pr(C)}\Pr(Y^{(0)}=1|C)\\
&=\frac{1-\rho_A-\rho_C}{1-\rho_A}\omega_N^0+\frac{\rho_C}{1-\rho_A}\omega_C^0,\\
&\Pr(Y=1|R=1)\\
&=\Pr(Y^{(1)}=1,Z=1|R=1)+\Pr(Y^{(0)}=1,Z=0|R=1)\\
&=\Pr(Y^{(1)}=1|Z=1,R=1)\Pr(Z=1|R=1)+\Pr(Y^{(0)}=1|Z=0,R=1)\Pr(Z=0|R=1)\\
&=\left(\frac{\rho_A}{\rho_A+\rho_C}\omega_A^0+\frac{\rho_C}{\rho_A+\rho_C}\omega_C^1\right)(\rho_A+\rho_C)+\omega_N^0(1-\rho_A-\rho_C)\\
&=\rho_A\omega_A^0+\rho_C\omega_C^1+\omega_N^0-\rho_A\omega_N^0-\rho_C\omega_N^0,\\
&\Pr(Y=1|R=0)\\
&=\Pr(Y^{(1)}=1,Z=1|R=0)+\Pr(Y^{(0)}=1,Z=0|R=0)\\
&=\Pr(Y^{(1)}=1|Z=1,R=0)\Pr(Z=1|R=0)+\Pr(Y^{(0)}=1|Z=0,R=0)\Pr(Z=0|R=0)\\
&=\omega_A^0\rho_A+\left(\frac{1-\rho_A-\rho_C}{1-\rho_A}\omega_N^0+\frac{\rho_C}{1-\rho_A}\omega_C^0\right)(1-\rho_A)\\
&=\rho_A\omega_A^0+\rho_C\omega_C^0+\omega_N^0-\rho_A\omega_N^0-\rho_C\omega_N^0.
\end{aligned}$$

A2. Prove: $\Pr(Y=1|Z,E)=\text{expit}(\lambda_0+\lambda_1 Z+\lambda_2 E)$ is not the true model and the true model should include the interaction between $Z$ and $E$, or the interaction between $Z$ and $R$. When there are no ATs, the true model does not include the interaction.

*Proof*
The true model is

$$\begin{aligned}
\Pr(Y=1|Z,E)&=\Pr(Y=1|Z,R)\\
&=E(Y|Z,R)\\
&=I_{(Z=0,R=0)}E(Y|Z=0,R=0)+I_{(Z=1,R=0)}E(Y|Z=1,R=0)\\
&\quad+I_{(Z=0,R=1)}E(Y|Z=0,R=1)+I_{(Z=1,R=1)}E(Y|Z=1,R=1)\\
&=E(Y|Z=0,R=0)+Z[E(Y|Z=1,R=0)-E(Y|Z=0,R=0)]\\
&\quad+R[E(Y|Z=0,R=1)-E(Y|Z=0,R=0)]\\
&\quad+ZR\begin{bmatrix}E(Y|Z=1,R=1)-E(Y|Z=1,R=0)\\-E(Y|Z=0,R=1)+E(Y|Z=0,R=0)\end{bmatrix}\\
&=\lambda_0+\lambda_1 Z+\lambda_2 R+\lambda_3 ZR.
\end{aligned}$$



Copyright © 2011 John Wiley & Sons, Ltd.

In the above equations,

$$\begin{aligned}
\lambda_0 &= E(Y|Z=0, R=0),\\
\lambda_1 &= [E(Y|Z=1, R=0) - E(Y|Z=0, R=0)],\\
\lambda_2 &= [E(Y|Z=0, R=1) - E(Y|Z=0, R=0)],\\
\lambda_3 &= E(Y|Z=1, R=1) - E(Y|Z=1, R=0)\\
&\quad - E(Y|Z=0, R=1) + E(Y|Z=0, R=0)\\
&= E(Y|Z=1, R=1) - (\lambda_0 + \lambda_1 + \lambda_2).
\end{aligned}$$

Thus the true model includes the interaction between $Z$ and $R$.

When there are no ATs, we have $I_{(Z=1,R=0)} \equiv 0$, then the true model becomes

$$\begin{aligned}
\Pr(Y=1|Z,E) &= \Pr(Y=1|Z,R)\\
&= E(Y|Z,R)\\
&= I_{(Z=0,R=0)} E(Y|Z=0, R=0)\\
&\quad + I_{(Z=0,R=1)} E(Y|Z=0, R=1) + I_{(Z=1,R=1)} E(Y|Z=1, R=1)\\
&= E(Y|Z=0, R=0) + R[E(Y|Z=0, R=1) - E(Y|Z=0, R=0)]\\
&\quad + Z[E(Y|Z=1, R=1) - E(Y|Z=0, R=1)]\\
&= \lambda_0 + \lambda_1 R + \lambda_2 Z.
\end{aligned}$$

In the above equations,

$$\begin{aligned}
\lambda_0 &= E(Y|Z=0, R=0),\\
\lambda_1 &= [E(Y|Z=0, R=1) - E(Y|Z=0, R=0)],\\
\lambda_2 &= [E(Y|Z=1, R=1) - E(Y|Z=0, R=1)].
\end{aligned}$$

The true model does not include the interaction term.

A3. Some details about the bias analysis.

(a) When there is no confounding, the treatment effect estimated with 2SPS can be biased. The bias of 2SPS estimator is

$$B_{\text{2SPS}} = \frac{\begin{array}{c}\text{logit}(\rho_A \omega_A^1 + \rho_C \omega_C^1 + \omega_N^0 - \rho_A \omega_N^0 - \rho_C \omega_N^0)\\ -\text{logit}(\rho_A \omega_A^1 + \rho_C \omega_C^0 + \omega_N^0 - \rho_A \omega_N^0 - \rho_C \omega_N^0)\end{array}}{\rho_C} - \text{logit}(\omega_C^1) + \text{logit}(\omega_C^0).$$

One no-confounding scenario is that there are no ATs, and Cs and NTs have the same probability of potential outcome, e.g. $\rho_A = 0$ and $\omega_C^0 = \omega_N^0$. Plugging in these values to the above equation, we have

$$\begin{aligned}
B_{\text{2SPS}} &= \frac{\begin{array}{c}\text{logit}(0\omega_A^1 + \rho_C \omega_C^1 + \omega_C^0 - 0\omega_C^0 - \rho_C \omega_C^0)\\ -\text{logit}(0\omega_A^1 + \rho_C \omega_C^0 + \omega_C^0 - 0\omega_C^0 - \rho_C \omega_C^0)\end{array}}{\rho_C} - \text{logit}(\omega_C^1) + \text{logit}(\omega_C^0)\\
&= \frac{\text{logit}(\rho_C \omega_C^1 + \omega_C^0 - \rho_C \omega_C^0) - \text{logit}(\omega_C^0)}{\rho_C} - \text{logit}(\omega_C^1) + \text{logit}(\omega_C^0).
\end{aligned}$$

This equation is generally not 0. We can easily see that it is 0 if on linear scale instead of on a logit scale.



Copyright © 2011 John Wiley & Sons, Ltd.

(b) When there is no confounding, the treatment effect estimated with 2SRI is unbiased.

The bias of the 2SRI estimator with no ATs is

$$\begin{aligned}B_{2\text{SRI}} &= \lambda_1 - \psi \\ &= \text{logit}\left(\frac{\rho_A}{\rho_A+\rho_C}\omega_A^1 + \frac{\rho_C}{\rho_A+\rho_C}\omega_C^1\right) - \text{logit}(\omega_N^0) - \frac{1}{\rho_C}\text{logit}\left(\frac{1-\rho_A-\rho_C}{1-\rho_A}\omega_N^0 + \frac{\rho_C}{1-\rho_A}\omega_C^0\right) \\ &\quad + \frac{1}{\rho_C}\text{logit}(\omega_N^0) - \text{logit}(\omega_C^1) + \text{logit}(\omega_C^0).\end{aligned}$$

Plugging in $\rho_A = 0$ and $\omega_C^0 = \omega_N^0$ to this equation, we have

$$\begin{aligned}B_{2\text{SRI}} &= \lambda_1 - \psi \\ &= \text{logit}\left(\frac{0}{0+\rho_C}\omega_A^1 + \frac{\rho_C}{0+\rho_C}\omega_C^1\right) - + \text{logit}(\omega_C^0) - \frac{1}{\rho_C}\text{logit}\left(\frac{1-0-\rho_C}{1-0}\omega_C^0 + \frac{\rho_C}{1-0}\omega_C^0\right) \\ &\quad + \frac{1}{\rho_C}\text{logit}(\omega_C^0) - \text{logit}(\omega_C^1) + \text{logit}(\omega_C^0) \\ &= \text{logit}(\omega_C^1) - \text{logit}(\omega_C^0) - \frac{1}{\rho_C}\text{logit}(\omega_C^0) + \frac{1}{\rho_C}\text{logit}(\omega_C^0) - \text{logit}(\omega_C^1) + \text{logit}(\omega_C^0) \\ &= 0.\end{aligned}$$

## References

1. Bellamy S, Lin J, Ten Have T. An introduction to causal modelling in clinical trials. *Clinical Trials* 2007; **4**(1):58–73.
2. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; **29**(4):722–729.
3. Hernan M, Robins J. Instruments for causal inference—An epidemiologist's dream?. *Epidemiology* 2006; **17**(4):360–372.
4. Angrist J, Imbens G, Rubin DB. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**(434):444–455.
5. Abadie A. Semiparametric instrumental variable estimation of treatment response models. *Journal of the American Econometrics* 2003; **113**:231–263.
6. Didelez V, Meng S, Sheehan NA. Assumptions of IV methods for observational epidemiology. *Statistical Science* 2010; **25**(1):22–40.
7. Sommer A, Zeger S. On estimating efficacy from clinical-trials. *Statistics in Medicine* 1991; **10**(1):45–52.
8. Frangakis C, Rubin D. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999; **86**(2):365–379.
9. Tan Z. Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association* 2006; **101**(476):1607–1618.
10. Small D, Rosenbaum P. War and wages: The strength of instrumental variables and their sensitivity to unobserved biases. *Statistics in Medicine* 2006; **25**(12):1981–2007.
11. Small D, Ten Have T, Joffe M, Cheng J. Random effects logistic models for analyzing efficacy of a longitudinal randomized treatment with non-adherence. *Journal of the American Statistical Association* 2008; **103**(483):924–933.
12. Sheiner L, Rubin D. Intention-to-treat analysis and the goal of clinical trials. *Clinical Pharmacology and Therapeutics* 1995; **56**(1):6–10.
13. Korn E, Teeter D, Baumrind S. Using explicit clinician preferences in nonrandomized study designs. *Journal of Statistical Planning and Inference* 2001; **96**(1):67–82.
14. Korn E, Rosenbaum P, Fienberg S, Rubin D. Causal inference through potential outcomes and principal stratification: Application to studies with 'censoring' due to death—comments and rejoiners. *Statistical Science* 2006; **21**(3):310–321.
15. Brookhart M, Wang P, Solomon D, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; **17**(3):268–275.
16. Wang P, Schneeweiss S, Avorn J, Fischer M, Mogun H, Solomon D, Brookhart M. Risk of death in elderly users of conventional vs atypical antipsychotic medications. *New England Journal of Medicine* 2005; **353**(22):2335–2341.
17. Hennessy S, Leonard C, Palumbo C, Shi X, Ten Have T. Instantaneous preference was a stronger instrumental variable than 3-and 6-month prescribing preference for nsaids. *Journal of Clinical Epidemiology* 2008; **61**(12):1285–1288.
18. Johnston S. Combining ecological and individual variables to reduce confounding by indication: case study—subarachnoid hemorrhage treatment. *Journal of Clinical Epidemiology* 2000; **53**(12):1236–1241.
19. Wen S, Kramer M. Uses of ecologic studies in the assessment of intended treatment effects. *Journal of Clinical Epidemiology* 1999; **52**(1):7–12.



Copyright © 2011 John Wiley & Sons, Ltd.

20. Brooks J, Chrischilles E, Scott S, Chen-Hardee S. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage ii patients from Iowa. *Health Services Research* 2003; **38**(6):1385–1402.
21. Stukel T, Fisher E, Wennberg D, Alter D, Gottlieb D, Vermeulen M. Analysis of observational studies in the presence of treatment selection bias—effects of invasive cardiac management on ami survival using propensity score and instrumental variable methods. *Journal of the American Medical Association* 2007; **297**(3):278–285.
22. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; **58**(1):21–29.
23. Joffe M, Brensinger C. Weighting in instrumental variables and g-estimation. *Statistics in Medicine* 2003; **22**(8):1285–1303.
24. Hogan J, Lancaster T. Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. *Statistical Methods in Medical Research* 2004; **13**(1):17–48.
25. Hirano K, Imbens W, Rubin B, Zhou X. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 2000; **1**(1):69–88.
26. Goetghebeur E, Molenberghs G. Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *Journal of the American Statistical Association* 1996; **91**(435):928–934.
27. Vansteelandt S, Goetghebeur E. Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society Series B—Statistical Methodology* 2003; **65**(4):817–835.
28. Ten Have T, Joffe M, Cary M. Causal logistic models for non-compliance under randomized treatment with univariate binary response. *Statistics in Medicine* 2003; **22**(8):1255–1283.
29. Robins J, Rotnitzky A. Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika* 2004; **91**(4):763–783.
30. Terza J, Basu A, Rathouz P. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics* 2008; **27**(3):531–543.
31. Rassen J, Schneeweiss S, Glynn R, Mittleman M, Brookhart M. Instrumental variable analysis for estimation of treatment effects with dichotomous outcomes. *American Journal of Epidemiology* 2009; **169**(3):273–284.
32. Nagelkerke N. Estimating treatment effects in randomized clinical trials in the presence of non-compliance (vol. 19, p. 1849, 2000). *Statistics in Medicine* 2001; **20**(6):982.
33. Rubin DB. Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics* 1978; **6**:34–58.
34. Rubin DB. Statistics and causal inference—Which ifs have causal answers. *Journal of the American Statistical Association* 1989; **81**(396):961–962.
35. Lin JY, Ten Have T, Elliott MR. Longitudinal nested compliance class model in the presence of time-varying noncompliance. *Journal of the American Statistical Association* 2008; **103**:462–473.
36. Newey W, Mcfadden D. *Large Sample Estimation and Hypothesis Tesing*, Chapter 36. Elsevier: Amsterdam, North-Holland, 1994; 2111–2245.
37. Wooldridge J. *M-Estimation*, Chapter 12. The MIT Press: Cambridge, MA/London/England, 2002; 341–384.
38. Nishii R. Maximum likelihood principle and model selection when the true model is unspecified. *Journal of Multivarite Analysis* 1988; **27**:392–403.
39. Vansteelandt S, Goetghebeur E. Using potential outcomes as predictors of treatment activity via strong structural mean models. *Statistica Sinica* 2004; **14**(3):907–925.
40. Ten Have T, Joffe M, Lynch K, Brown G, Maisto S, Beck A. Causal mediation analyses with rank preserving models. *Biometrics* 2007; **63**:926–924.



Copyright © 2011 John Wiley & Sons, Ltd.