Statistics
in Medicine

Research Article

Received 13 December 2013, Accepted 8 December 2014 Published online 29 December 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6404

On a preference-based instrumental variable approach in reducing unmeasured confounding-by-indication

Yun Li,[a,b*†] Yoonseok Lee,[c] Robert A. Wolfe,[b] Hal Morgenstern,[b,d] Jinyao Zhang,[b] Friedrich K. Port[b] and Bruce M. Robinson[b,e]

Treatment preferences of groups (e.g., clinical centers) have often been proposed as instruments to control for unmeasured confounding-by-indication in instrumental variable (IV) analyses. However, formal evaluations of these group-preference-based instruments are lacking. Unique challenges include the following: (i) correlations between outcomes within groups; (ii) the multi-value nature of the instruments; (iii) unmeasured confounding occurring between and within groups. We introduce the framework of between-group and within-group confounding to assess assumptions required for the group-preference-based IV analyses. Our work illustrates that, when unmeasured confounding effects exist only within groups but not between groups, preference-based IVs can satisfy assumptions required for valid instruments. We then derive a closed-form expression of asymptotic bias of the two-stage generalized ordinary least squares estimator when the IVs are valid. Simulations demonstrate that the asymptotic bias formula approximates bias in finite samples quite well, particularly when the number of groups is moderate to large. The bias formula shows that when the cluster size is finite, the IV estimator is asymptotically biased; only when both the number of groups and cluster size go to infinity, the bias disappears. However, the IV estimator remains advantageous in reducing bias from confounding-by-indication. The bias assessment provides practical guidance for preference-based IV analyses. To increase their performance, one should adjust for as many measured confounders as possible, consider groups that have the most random variation in treatment assignment and increase cluster size. To minimize the likelihood for these IVs to be invalid, one should minimize unmeasured between-group confounding.

Keywords: bias formula; causal inference; instrumental variables; observational study; unmeasured confounders

1. Introduction

As the health care system becomes more digitalized, large administrative databases become increasingly available. This provides valuable opportunities to conduct observational studies to evaluate the effectiveness and quality of care in actual practice and on a large scale. However, the validity of these studies is often threatened by confounding-by-indication, a source of bias which is common but particularly difficult to remedy [1]. Confounding-by-indication arises when doctors assign different treatment plans to patients based on perceived patient risk or prognosis [2]. In order to handle this, statistical methods, such as regression analyses and propensity score adjustments [3], have been widely used to obtain treatment effect estimates. However, such methods generally result in bias if an important confounder is not controlled for. This can frequently happen because, usually, not all the information on the confounders is

[a]*Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[b]*Arbor Research Collaborative for Health, Ann Arbor, MI 48104, U.S.A.*
[c]*Department of Economics and Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Syracuse University, Syracuse, NY 13244, U.S.A.*
[d]*Departments of Epidemiology, Environmental Health Sciences, and Urology, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[e]*Department of Medicine, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[*]*Correspondence to: Yun Li, Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109 and Arbor Research Collaborative for Health, Ann Arbor, MI 48104, U.S.A.*
[†]*E-mail: yunlisph@umich.edu*

available in observational studies. For example, patient disease status may be available, but not the disease severity or its complications, which leads to confounding-by-indication and subsequent bias. Additional examples of covariate omissions can occur in the context of simultaneity and measurement errors. Simultaneity arises when the exposure of interest is also partly determined by the outcome of interest. It often occurs in studying chronic diseased patients who receive treatments that are ongoing and change over time. For example, in a study of examining the relationship between erythropoiesis-stimulating agents (ESA) use and hemoglobin (Hgb) level for managing anemia patients, ESA dose is the exposure of interest, and Hgb is the outcome. However, ESA dose is also partly determined by the Hgb level. It can be challenging to include all the variables and their correct functional forms in order to capture the temporal relationship between ESA dose and Hgb accurately.

These challenges motivate us to examine an instrumental variable (IV) analysis that can lead to more robust effect estimates, even in the absence of complete covariate adjustments. In particular, we focus on a preference-based IV because treatment preferences across different groups, such as clinical centers, physician practices, or service areas, have often been proposed as instruments to control for confounding-by-indication due to unmeasured factors [4, 5]. In an IV analysis, an instrument is used to mimic an experiment mechanism that can randomly 'assign' patients to different treatment plans, and it can work virtually as well as randomization of treatment assignment. An instrument can only affect the outcome through the treatment plans which it assigns patients to. IV analyses have been considered to work like a randomized encouragement design [6]. In a randomized encouragement design [7, 8], patients are randomly assigned to either treatment or control groups. Patients are then 'encouraged' to receive either treatment or control. However, patients may or may not comply, and non-compliance can lead to selection bias. Randomization can operate as a perfect IV to overcome the bias [9–11] with higher compliance implying a stronger instrument. A preference-based IV analysis works like these randomized encouragement trials adapted for observational studies in which patients are 'assigned' to different groups through 'randomization' after sufficiently adjusting for measured confounders. Different groups then 'encourage' patients to receive differential treatment dosage levels even after sufficient covariate adjustments, but patients are free to take dosage levels different from what they are 'encouraged' to and can choose to 'comply' to various degrees. These differential dosage levels 'encouraged' at different groups are not usually completely 'random'. However, they are often partially 'random' because of differential group policies, mix of insurance coverage, and patient/physician's preferred drug use or medical knowledge level. And it may be reasonable to assume that these factors could be generally independent of unmeasured confounders, such as patients' disease severity or complications after adjusting for measured confounders, or at least much more independent than the treatment plan a patient actually receives. See related argument by Baiocchi *et al* [12]. The preference-based IV analysis utilizes this 'random' component of the variation in the treatment assignment across groups to obtain valid treatment effect estimates.

The preference-based IV analysis has been used in a wide range of medical research. It is closely related, but not equivalent to the group-treatment approach [13], in which the proportion treated in each group is substituted for each patient's actual treatment status in a conventional regression analysis. See related applications of preference-based IV analyses and their variations [14–17]. It is also related to multiple-instruments-based analyses, which have been used in studies such as Mendelian randomization [18]. However, preference-based IVs differ from multiple instruments in several ways: (i) intra-cluster correlations among outcomes often exist, which may arise from unadjusted group-level characteristics such as other group treatment practices beyond the treatment of interest; (ii) unmeasured confounding occurring within and/or between clusters; (iii) the preference-based IVs have multiple values and function as categorical variables, and the number of different values (or categories) of a preference-based IV often increases as the number of groups increases; (iv) unlike Mendelian randomization studies where a patient can have several genetic variants as multiple instruments, a patient in a preference-based IV analysis cannot belong to multiple groups, and his group membership is unique. These unique features raise unique challenges in examining the properties of preference-based IV estimators. Note that we use 'group' and 'cluster' interchangeably.

In this paper, we consider continuous outcomes with continuous treatment. We propose a two-stage generalized least squares (2SGLS) estimator and its variance estimator to accommodate the correlations of outcomes within groups and measured confounders when the outcome model is a linear mixed

model (LM) with known variance–covariance structures. We introduce the concept of between-cluster and within-cluster confounding to examine the assumptions required for preference-based IV analyses. We formalize the assumptions required for the IVs themselves and for the corresponding IV models to be valid. We raise the concern over the validity of preference-based IV estimators when unmeasured between-cluster confounding effects exist. Further, we derive a closed-form expression of asymptotic bias for the 2SGLS IV estimator when the IV is valid. While two-stage least squares (2SLS) IV estimators are usually consistent, that is, unbiased when the number of independent units goes to infinity, their evaluations have been conducted within simpler settings without clustering or measured confounders or multi-valued instruments [19]. Our bias calculation accommodates all these complexities. Additional IV methods that can incorporate covariates have been described elsewhere in [20–26]. With valid preference-based IVs, when the number of independent units (i.e., clusters) goes to infinity, we demonstrate that the 2SGLS preference-based IV estimator is usually biased unless the number of patients per cluster (i.e., cluster size) also goes to infinity. We conduct simulations to confirm that our asymptotic bias formula approximates bias quite well in finite samples, even when the number of clusters is not very large. We also examine bias of the preference-based IV analysis through simulations in the presence of unmeasured between-cluster confounders when the IVs are invalid. In all simulations, we compare the IV estimator with a commonly used LM estimator. We then provide practical guidance in reducing bias and the possibility of violating assumptions required for preference-based IV analyses.

Our motivating example comes from the Dialysis Outcomes and Practice Patterns Study (DOPPS), an international prospective cohort study of patients receiving hemodialysis for end-stage kidney disease. Patients entered the study at various times after dialysis initiation. The objective is to estimate the effect of ESA use on Hgb level for managing anemia patients. Physicians prescribe ESA dosage primarily based on patients' Hgb levels, health status, and ESA responsiveness. However, many observational databases lack information that sufficiently captures ESA responsiveness. We illustrate the use of the IV analysis without controlling for ESA responsiveness.

2. Between-cluster and within-cluster confounding

In this section, we define the types of confounders in a clustered data setting. We are interested in the causal relationship between the treatment T and outcome Y. We denote any confounder of the $T-Y$ association in a clustered observational study as U_{ij} with $i = 1, ..., m$ indexing clusters and $j = 1, ..., n_i$ for subjects within clusters. Without losing generality, we assume $n_i = n$ for any i. Neuhaus [27] introduced cluster-level and *designed* within-cluster covariates to assess their respective effects on longitudinal outcomes. He separated a *designed* within-cluster covariate into a between-cluster component and a within-cluster component. Usually, the effects of between- and within-cluster components of the same covariate on outcomes are assumed to be the same in regression models; however, Neuhaus and Kalbfleisch [28] present some examples to illustrate that effects of these two components on outcomes can be very different. We adopt this framework to study confounding and required assumptions in the use of preference-based IVs. We assume there are two types of confounders: cluster-level confounders and *designed* within-cluster confounders. A cluster-level confounder has identical values for all the subjects with the same cluster, that is, $U_{ij} = U_i$ for all j with U_i denoting the cluster mean of U_{ij}. Examples include cluster-level characteristics, such as free-standing or hospital-based dialysis facilities, teaching-based or non-teaching-based facilities, or facility nurse/patient ratios in the DOPPS data example, and other characteristics that do not change across subjects within the same cluster, such as county membership. A *designed* within-cluster confounder generally has different values for subjects within the same cluster, although these values in the same cluster come from an identical distribution. On the other hand, the variances and cluster means of a *designed* within-cluster confounder generally vary across different clusters. Examples include patients' age, Hgb level, income, and body mass index. We assume all confounding factors of the $T-Y$ relationship can be summarized by two components: (i) within-cluster components of confounders, which include the deviations from cluster means of any *designed* within-cluster confounders, $(U_{ij}-U_i)$; and (ii) between-cluster components of confounders, which include the cluster-level confounders and the cluster means of any *designed* within-cluster confounders (U_i). We can assess how each component affects the treatment/outcome and define the effect of (i) on the treatment/outcome as the within-cluster confounding and the effect of (ii) as the between-cluster confounding. We further assume

that some confounders are observed and adjusted for in the models, which we denote as C_{ij} for simplicity, a K_c-dimensional vector consisting of both within-cluster and between-cluster components of observed confounders. That is, for some k, $C_{kij} = C_{ki}$ for any j with C_{ki} representing the cluster mean of C_{kij}. For unobserved or unadjusted confounders, we use separate matrix notations for the within-cluster components (P_{ij}) and the between-cluster components (G_i). We assume that P_{ij} is a K_p dimensional vector consisting of all within-cluster components of any *designed* within-cluster unobserved confounders, and G_i is a K_g dimensional vector consisting of all between-cluster components of any *designed* within-cluster unobserved confounders and any cluster-level unobserved confounders. The separation between unmeasured between-cluster confounding and within-cluster confounding helps assess the assumptions required for the preference-based IV approach in the next section.

3. Preference-based instrumental variable analysis

We first describe preference-based IV analysis models. Then, we assess the validity of the IV analysis in the presence of unmeasured between-cluster or within-cluster confounding. Finally, we examine the assumptions for the preference-based IV analysis models and for preference-based IVs to be valid instruments.

3.1. The models and assumptions

The models using the preference-based IVs we are interested in are written as simultaneous equations:

$$T_{ij} = \gamma_i + C_{ij}^{'}\alpha_{Ic} + e_{ij}^{t}, \tag{1}$$

$$Y_{ij} = \beta_I T_{ij} + C_{ij}^{'}\beta_{Ic} + v_i + e_{ij}^{y}, \tag{2}$$

where β_I, α_{Ic} and β_{Ic} are respective fixed effects; β_I is the parameter of interest estimating the $T-Y$ relationship; e_{ij}^{t} and e_{ij}^{y} are within-cluster errors; γ_i and v_i are between-cluster errors which accommodate the intra-cluster correlation arisen from the fact that subjects within the same cluster tend to be more alike than those from different clusters; C_{ij} contains the measured confounders and $C_{1ij} = 1$ for the intercept. All error terms e_{ij}^{t}, e_{ij}^{y}, γ_i and v_i have mean zeros.

Alternatively, the models (1) can be expressed as

$$T_{ij} = Z_{ij}^{'}\theta + C_{ij}^{'*}\alpha_{Ic}^{*} + e_{ij}^{t}\,, \tag{3}$$

where Z_{ij} is an $m \times 1$ indicator vector with its elements being $\mathrm{I}(\ell = i)$ for $\ell = 1, \ldots, m$ (i.e., if $\ell = i$, $\mathrm{I}(\ell = i) = 1$; otherwise, $\mathrm{I}(\ell = i) = 0$, with ℓ representing any potential group memberships), θ is an $m \times 1$ vector of parameters such that $\theta = \left(\gamma_1 + \alpha_{1Ic}, \ldots, \gamma_m + \alpha_{1Ic}\right)^{'}$ where α_{1Ic} is a constant and corresponds to the intercept term, $C_{ij}^{'*} = \left(C_{2ij}, \ldots, C_{K_c ij}\right)$ and $\alpha_{Ic}^{*} = \left(\alpha_{2Ic}, \ldots, \alpha_{K_c Ic}\right)$. In this formulation, we regard Z_{ij} as random variables and θ as unknown parameters. Note that $Z_{ij}^{'}\theta = \gamma_i + \alpha_{1Ic}$, which represents treatment preference levels across m clusters. We will consider models (3) and (2) in the estimation of the IV models. Let $\xi_i = v_i J_n + e_i^{y}$ where $e_i^{y} = \left(e_{i1}^{y}, \cdots, e_{in}^{y}\right)^{'}$. We assume that $v_i \sim N\left(0, \sigma_v^2\right)$ and $e_{ij}^{y} \sim N\left(0, \sigma_{ey}^2\right)$; hence, $\xi_i \sim N(0, \Omega)$ with $\Omega = \sigma_v^2 I_n + \sigma_{ey}^2 J_n J_n^{'}$ where I_n is an identity matrix with rank n, and J_n is a $n \times 1$ vector of ones. We further assume that $\gamma_i \sim N\left(0, \sigma_r^2\right)$ and $e_{ij}^{t} \sim N\left(0, \sigma_{et}^2\right)$. Although it is not necessary, for simplicity, we make the normality assumptions for e_{ij}^{t}, e_{ij}^{y}, γ_i and v_i here. Additional assumptions for the IV model equations include $\left(e_{ij}^{t}, e_{ij}^{y}, v_i\right) \perp C_{ij}$, $e_{ij}^{y} \perp T_{ij}$ and $(\gamma_i, v_i) \perp \left(e_{ij}^{t}, e_{ij}^{y}\right)$. These assumptions are usually standard in linear mixed models.

We now explore the relationship between e_{ij}^{t} and e_{ij}^{y}, and that between γ_i and v_i in the presence of unmeasured between-cluster and within-cluster confounders. The IV analysis aims to obtain valid effect estimates by utilizing the random component of the treatment assignment, which may arise from differential group policies or preferences and is independent of unmeasured confounders (conditional on measured confounders). The random component is captured by γ_i and needs to be independent of v_i, e_{ij}^{t}, and e_{ij}^{y}. In the presence of unmeasured within-cluster confounders P, but not unmeasured between-cluster confounders G, P is absorbed by e_{ij}^{t} and e_{ij}^{y}. Let $e_{ij}^{t} = P_{ij}^{'}\alpha_p + \varepsilon_{ij}^{t}$ and $e_{ij}^{y} = P_{ij}^{'}\beta_p + \varepsilon_{ij}^{y}$. Even if $\varepsilon_{ij}^{t} \perp \varepsilon_{ij}^{y}$ and $P_{ij} \perp \left(\varepsilon_{ij}^{t}, \varepsilon_{ij}^{y}\right)$, we have $\mathrm{Cov}(e_{ij}^{t}, e_{ij}^{y}) = \alpha_p^{'} V(P)\beta_p \neq 0$ where $V(P)$ represents the variance of P. This

non-zero correlation between the within-cluster error terms usually serves as the motivation to use an IV analysis approach because fitting a single model equation (2) can typically result in biased estimates of β_I. Here, when P exists, $\gamma_i \perp (e^t_{ij}, e^y_{ij}, v_i)$ remains true after adjusting for C_{ij}. In order to make these assumptions remain plausible, we further assume that P_{ij} represents unadjusted residual within-cluster confounding after controlling for C_{ij}. On the other hand, in the presence of unmeasured between-cluster confounding from G, G is absorbed by γ_i and v_i. Let $\gamma_i = G_i'\alpha_g + r_{0i}$ and $v_i = G_i'\beta_g + u_{0i}$. Even when $r_{0i} \perp u_{0i}$, we have $\mathrm{Cov}(\gamma_i, v_i) \neq 0$, and γ_i is not independent of v_i anymore (conditional on C). Hence, the IV analysis cannot utilize γ_i to obtain valid effect estimation when G exists.

Examples of the between-cluster confounding factors may include physicians' training levels, group preferences of other treatments, environmental factors, and average social economic status of patients at group levels. If these factors are not adjusted for, it can result in $\mathrm{Cov}(\gamma_i, v_i) \neq 0$ and invalid estimates of the causal effect. In practice, some medical practices at group levels are correlated with each other while others are not. For example, during anemia management of patients receiving hemodialysis for end-stage kidney disease, ESA and iron prescriptions are often considered simultaneously in raising Hgb. Higher iron dosage often accompanies lower ESA dosage prescriptions, and the cost of these two treatments is now reimbursed as a bundle. If the group preferences of iron prescriptions are not adjusted for in the model examining the ESA-Hgb relationship, it will likely induce a correlation between γ_i and v_i and a direct effect between IV and the outcome. On the other hand, group preferences of vascular access type prescriptions or blood flow managements are medical practices that are likely independent of the ESA prescription. Even though they may influence the Hgb levels, it is unlikely that they will induce the correlation between γ_i and v_i. Therefore, when unadjusted group treatment/practice preferences are not correlated with group ESA prescription preferences, group preferences can still serve as instruments to represent the variation of ESA dosage across groups that is captured by γ_i and remains random. These unadjusted group practices contribute to the between-cluster error term v_i and partially explain the correlations among Hgb levels within the same group. Our IV models are designed to accommodate these correlations.

Prior explorations have focused on assumptions related to the IV models, and here we attempt to formalize the assumptions required for preference-based IVs to be valid in a counterfactual framework. We consider the subscripts ij are purely labels, and subjects are randomly assigned to the ith group and then the jth member within the group within levels of C; all information about subjects is described by the underlying true models, including all confounders. For simplicity, we assume that γ_i is ordered such that $\gamma_i > \gamma_{i'}$ for any $i > i'$. That is, we order the groups from the lowest group treatment dosage preference levels to the highest. In literature, the m indicator variables in Z_{ij} are sometimes referred to as the IVs [15,29]. In our setting, the presence of random effect v_i in the IV model Equation (2) allows for additional unadjusted between-cluster factors to exist as long as they are uncorrelated with γ_i and are not $T - Y$ confounders. This may give the misconception of a violation of valid IV assumptions because a direct effect from Z_{ij} to the outcome may seem to have come from these additional unadjusted between-cluster factors. In actuality, the purpose of using Z_{ij} is to extract the various treatment preference levels across m clusters represented by γ_i, which is achieved by fitting Equation (3); hence, γ_i is the essential player whose properties are what matter here. Additionally, there is a one-to-one relationship between Z_{ij} and γ_i, such that $Z_{ij}'\theta = \gamma_i + \alpha_{1Ic}$ and a higher order of $\ell = i$ (i.e., the element of Z_{ij}), corresponds to a higher value of γ_i. Therefore, it is more accurate to explore the IV assumptions using γ_i directly instead of Z_{ij} in our setting. That is, it is essential for γ_i not to have a direct effect on Y, which implies that γ_i should not be part of the Equation (2) and that γ_i needs to be independent of the unmeasured factors such that $\gamma_i \perp v_i$. Because the subscripts are purely labels, for simplicity of notations, we drop the subscripts in examining the IV assumptions. To formalize the assumptions, we first hypothesize that, in a counterfactual framework, for any subject, $Y^{\gamma,t}(v, C)$ is the potential outcome that would have been observed under the treatment preference level of γ, and the treatment dosage t actually received by any subject, conditional on v and C, and $T^{\gamma}(C)$ is the potential treatment a subject would have received under the group treatment preference level of γ given C. In reality, we can observe only one of the potential outcomes of Y for any subject because the subject can only be assigned to one group with the corresponding group-specific treatment preference level and one treatment dosage, and similarly for T.

We adopt the IV assumptions specified in [9, 30] for a preference-based IV:

(1) γ is positively associated with the treatment received: $E(T^{\gamma}|C) > E(T^{\gamma'}|C)$ for any $\gamma > \gamma'$. Here, we assume that the level γ of the IV means that the subject is encouraged to take level γ of the treatment.
(2) γ must be independent of unmeasured confounders, conditional on measured covariates: $\gamma \perp [T^{\gamma}(C), Y^{\gamma,t}(v, C)]$. This assumption is implied by random assignment of the preference-based instruments, conditional on measured covariates. When between-cluster confounders G are unadjusted for, this assumption is violated because γ is not independent of G, conditional on measured covariates.
(3) There must not be a direct effect between γ and Y, conditional on v and C. The group treatment preference level γ must affect outcome only through its effect on the treatment received: $Y^{\gamma,t}(v, C) = Y^{\gamma',t}(v, C)$. This is the exclusion restriction (ER). Under the ER, $Y^{\gamma,t}(v, C) \equiv Y^{\gamma',t}(v, C)$.

These are the three main assumptions for valid preference-based IVs. Several additional assumptions for IVs can often be found in literature. One is the stable unit treatment value assumption (SUTVA) [9], which is commonly assumed: (a) If $\gamma = \gamma'$, then $T^{\gamma}(C) = T^{\gamma'}(C)$; (b) If $\gamma = \gamma'$ and $t = t'$, then $Y^{\gamma,t}(v, C) = Y^{\gamma',t'}(v, C)$. The assumptions 1−3 and SUTVA do not point identify a treatment effect [31]. In order to point identify the treatment effect, we will need to assume either monotonicity or homogeneous effect. Monotonicity assumes that there are no subjects who are 'defiers', that is, $T^{\gamma}(C) > T^{\gamma'}(C)$ for any $\gamma > \gamma'$. With the assumptions $1 - 3$, SUTVA and monotonicity, the IV estimate is interpreted as the causal treatment effect for the subpopulation of compliers, that is, complier average causal effect (CACE, [30]), but the IV estimate may not generalize to the whole population. A homogeneous effect assumes that the treatment has the same effect for compliers and defiers. With the assumptions $1 - 3$, SUTVA and the homogeneous effect, the IV estimate is a consistent estimate of the CACE, which is also the average causal effect for the whole population. Note that in this manuscript, except for the data analysis, we assume that the treatment effect is homogeneous. In the data analysis, we do not really know whether the true treatment effect is homogeneous or heterogeneous or whether the monotonicity assumption holds.

3.2. *Two-stage generalized least squares estimator*

Assume $\beta_{Ic} = \left(\beta_{1Ic}, \ldots, \beta_{K_cIc}\right)'$ in the model (2) corresponds to the effects of $\left(C_{1ij}, \ldots, C_{K_cij}\right)'$. Let $\eta_I = \left(\beta_I, \beta_{1Ic}, \ldots, \beta_{K_cIc}\right)'$. With preference-based IVs, the 2SGLS estimator of η_I is given by:

$$\hat{\eta}_I = \left(\sum_{i=1}^{m} \hat{O}_i' \hat{\Omega}^{-1} \hat{O}_i\right)^{-1} \left(\sum_{i=1}^{m} \hat{O}_i' \hat{\Omega}^{-1} Y_i\right), \tag{4}$$

where $\hat{\Omega}$ is an estimate of Ω and $\hat{O}_i = \left(\hat{O}_{i1}, \ldots, \hat{O}_{in}\right)'$ with $\hat{O}_{ij} = (\hat{T}_{ij}, C_{1ij}, \cdots, C_{K_cij})'$. And $\hat{T}_{ij}$ is the predicted T obtained from the equation (3) using ordinary least squares (OLS) estimation by regressing T_{ij} on C_{ij}' and Z_{ij}' (namely a fixed effect estimation instead of a random effect estimation in economics). Let $X_{ij}' = \left(Z_{ij}', C_{ij}'^{*}\right) = \left(Z_{1ij}, \cdots, Z_{mij}, C_{2ij}, \cdots, C_{K_cij}\right)$ and $X_i = (X_{11}, \cdots, X_{in})$. Hence, Equation (3) can be rewritten as $T_i = X_i'\zeta + e_i'$, where $\zeta' = \left(\theta', \alpha_{Ic}'^{*}\right)$ and $e_i = \left(e_{i1}, e_{i2}, \ldots, e_{in}\right)'$. Subsequently, we have $\hat{\zeta} = \left[\sum_{i=1}^{m}(X_i X_i')\right]^{-1}\left[\sum_{i=1}^{m} X_i T_i\right]$ and $\hat{T}_{ij} = X_{ij}'\hat{\zeta}$. The 2SGLS estimator of β_I is given by $\hat{\beta}_I = (1, 0, \ldots, 0) \times \hat{\eta}_I$. The variance of $\hat{\eta}_I$ is estimated by:

$$\mathrm{Var}(\hat{\eta}_I) = \left(\sum_{i=1}^{m} \hat{O}_i' \hat{\Omega}^{-1} \hat{O}_i\right)^{-1}.$$

Hence, $\mathrm{Var}(\hat{\beta}_I) = (1, 0, \ldots, 0)\mathrm{Var}(\hat{\eta}_I)(1, 0, \ldots, 0)'$. Note that we propose this variance estimator to accommodate the known variance-covariance structure. The estimator is different from the Hubert-White variance estimator, which is robust when variance-covariance structure is unknown [32] and often used in software packages.

In practice, the components of Ω $\left(\text{i.e., } \sigma_v^2 \text{ and } \sigma_{ey}^2\right)$ are replaced by their estimates, which we obtain by taking the following steps:

(1) Run the initial pooled 2SLS regression without any random effect, ignoring the variance-covariance structure: $\tilde{\eta}_I = \left(\sum_{i=1}^m \widehat{O}_i' \widehat{O}_i\right)^{-1} \sum_{i=1}^m \widehat{O}_i' Y_i$.
(2) Calculate the 2SLS residual $\tilde{e}_{ij}$ such that $\tilde{e}_{ij} = Y_{ij} - \tilde{\beta}_I T_{ij} - C_{ij}' \tilde{beta}_{Ic}$ where $\tilde{\beta}_I$ and $\tilde{\beta}_c$ are components of $\tilde{\eta}_I$.
(3) Obtain the estimates of σ_v^2 and σ_{ey}^2 as:

$$\hat{\sigma}_v^2 = \frac{1}{nm(n-1)/2 - K} \sum_{i=1}^{m} \sum_{j=1}^{n-1} \sum_{h=j+1}^{n} \tilde{e}_{ij} \tilde{e}_{ih},$$

$$\hat{\sigma}_{ey}^2 = \frac{1}{nm - K} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{e}_{ij}^2 \right) - \hat{\sigma}_v^2.$$

(4) Estimate $\hat{\eta}_I$ using equation (4).
(5) Iterate steps 2-4 using a new 2SGLS regression residual $\hat{e}_{ij} = Y_{ij} - \hat{\beta}_I T_{ij} - C_{ij}' \hat{\beta}_{Ic}$ to replace $\tilde{e}_{ij}$ until convergence.

The method proposed in step 3 to obtain $\hat{\sigma}_v^2$ is based on the fact that $E\left(e_{ij}^y e_{ih}^y\right) = \sigma_v^2$ for all $j \neq h$ and $\hat{\sigma}_{ey}^2$ is based on the standard OLS error-variance estimator $\hat{\sigma}_e^2 = \hat{\sigma}_v^2 + \hat{\sigma}_{ey}^2 = [1/(nm-K)] \sum_{i=1}^m \sum_{j=1}^n \tilde{e}_{ij}^2$ [33]. Note that K is used for degrees of freedom correction in estimating σ_v^2 and σ_{ey}^2 in finite samples and $K = \text{rank E}\left(\widehat{O}_i' \widehat{\Omega}^{-1} \widehat{O}_i\right)$.

Note that the residuals in steps 2 and 5 are not the same as the residuals from $Y_{ij} - \tilde{\beta}_I \widehat{T}_{ij} - C_{ij}' \tilde{\beta}_{Ic}$ or $Y_{ij} - \hat{\beta}_I \widehat{T}_{ij} - C_{ij}' \hat{\beta}_{Ic}$, which do not give correct estimates of σ_v^2, σ_{ey}^2 or β_I.

4. Bias in the presence of within-cluster unmeasured confounding

In this section, we first introduce two commonly used true models for T and Y. We then derive the expression of asymptotic bias of the 2SGLS IV estimator based on the true models when group treatment preferences are valid instruments and when unmeasured within-cluster confounding, but not between-clustering confounding, may exist.

4.1. True models

We consider two commonly used LMs as true models for T and Y in clustered data settings, with C_{ij} representing the measured confounders and P_{ij} for the unmeasured within-cluster confounders. The distributions of T and Y are specified as follows:

$$T_{ij} = a_{0i} + C_{ij}' \alpha_c + P_{ij}' \alpha_p + \epsilon_{ij}^t, \tag{5}$$

$$Y_{ij} = b_{0i} + \beta T_{ij} + C_{ij}' \beta_c + P_{ij}' \beta_p + \epsilon_{ij}^y, \tag{6}$$

where α_c, β, β_c, and β_p are fixed effects; a_{0i}, and b_{0i} are between-cluster random errors; ϵ_{ij}^t and ϵ_{ij}^y are within-cluster random errors. We assume $a_{0i} \sim N\left(0, \sigma_a^2\right)$, $b_{0i} \sim N\left(0, \sigma_b^2\right)$, $\epsilon_{ij}^t \sim N\left(0, \sigma_{\epsilon t}^2\right)$, and $\epsilon_{ij}^y \sim N\left(0, \sigma_{\epsilon y}^2\right)$. The parameter of interest is β, the causal effect of T on Y. The key assumptions include $a_{0i} \perp (\epsilon_{ij}^t, \epsilon_{ij}^y, P_{kij})$, $b_{0i} \perp (a_{0i}, C_{kij}, T_{ij}, \epsilon_{ij}^t, \epsilon_{ij}^y, P_{kij})$, $\epsilon_{ij}^t \perp (C_{kij}, P_{kij})$ and $\epsilon_{ij}^y \perp (\epsilon_{ij}^t, C_{kij}, T_{ij}, P_{kij})$ for any k. Of note, the between-cluster errors a_{0i} and b_{0i} are also random effects which capture the random variation across clusters beyond C, P for a_{0i}, and beyond C, P, T for b_{0i}.

Let $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in})'$, $T_i = (T_{i1}, T_{i2}, \ldots, T_{in})'$, $P_i = (P_{i1}, P_{i2}, \ldots, P_{in})'$, $C_i = (C_{i1}, C_{i2}, \ldots, C_{in})'$, $\epsilon_i^t = \left(\epsilon_{i1}^t, \epsilon_{i2}^t, \ldots, \epsilon_{in}^t\right)'$, and $\epsilon_i^y = \left(\epsilon_{i1}^y, \epsilon_{i2}^y, \ldots, \epsilon_{in}^y\right)'$. Here, Y_i, T_i, ϵ_i^t, and ϵ_i^y are $n \times 1$ vectors, P_i and C_i are

$n \times K_p$, and $n \times K_c$ matrices respectively. The true models in (5) and (6) can be expressed in a matrix form as follows:

$$T_i = a_{0i} J_n + C_i \alpha_c + P_i \alpha_p + \epsilon_i^t,$$

$$Y_i = b_{0i} J_n + \beta T_i + C_i \beta_c + P_i \beta_p + \epsilon_i^y.$$

A more compact matrix form for the true models can be written as below and will be used in the next section:

$$T = A_0 + C\alpha_c + P\alpha_p + \epsilon^t, \tag{7}$$

$$Y = B_0 + \beta T + C\beta_c + P\beta_p + \epsilon^y, \tag{8}$$

where Y, T, ϵ^t, ϵ^y, A_0, and B_0 consist of stacked elements of Y_i, T_i, ϵ_i^t, ϵ_i^y, $a_{0i}J_n$, and $b_{0i}J_n$ respectively. Note that Y, T, ϵ^t, ϵ^y, A_0, and B_0 are $mn \times 1$ vectors; and P and C are $mn \times K_p$ and $mn \times K_c$ matrices respectively.

4.2. Asymptotic bias

Here, we will derive the asymptotic bias under these true models and examine the consistency property of the 2SGLS IV estimator. To simplify the derivations, we make the assumptions that the means of T, C, and P are zeros because these means only impact the estimator of the intercept but not the estimator of β.

We first examine the impact of measured confounders C (including the intercept term) on the bias derivation. We let $M_C = I_{mn} - C(C'C)^{-1}C'$ and transform the data by pre-multiplying M_C to the regression equations in (7) and (8) to obtain

$$T^* = A_0 + P\alpha_p + \epsilon^t,$$

$$Y^* = B_0 + \beta T^* + P\beta_p + \epsilon^y,$$

where $T^* = M_C T$ and $Y^* = M_c Y$ which are the projection errors of Y and T on the space spanned by C respectively. Because $C_{ij} \perp \left(\epsilon_{ij}^t, \epsilon_{ij}^y, P_{ij}, a_{0i}, b_{0i}\right)$, we have $M_c\epsilon^t = \epsilon^t$, $M_c\epsilon^y = \epsilon^y$, $M_cP = P$, and so on. As we shall find, the transformation by multiplying M_C does not impact the derivation of any element in the bias formula because the formula (9) will remain the same with transformed data. Therefore, to further simplify the derivation process, we will assume that there are no measured confounders C. Theoretically, the bias formula should be the same with or without C, which we will also confirm through simulations.

In the absence of C, the 2SGLS estimator of β obtained by fitting models (3) and (2) can be simplified to

$$\begin{aligned}\hat{\beta}_I &= \left(\sum_{i=1}^{m} \widehat{T}_i' \widehat{\Omega}^{-1} \widehat{T}_i\right)^{-1} \left(\sum_{i=1}^{m} \widehat{T}_i' \widehat{\Omega}^{-1} Y_i\right) \\ &= \left[\sum_{i=1}^{m} \widehat{T}_i' \left(I_n - \hat{\pi} J_n J_n'\right) \widehat{T}_i\right]^{-1} \sum_{i=1}^{m} \left[\widehat{T}_i' \left(I_n - \hat{\pi} J_n J_n'\right) \left(\beta T_i + P_i \beta_p + b_{0i} J_n + \epsilon_i^y\right)\right],\end{aligned}$$

where $\widehat{T}_i = QT_i = J_n J_n' T_i / n$, with Q being the orthogonal projection matrix and $\hat{\pi} = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_{ey}^2 + n\hat{\sigma}_v^2}$. Of note, since $\Omega = \sigma_v^2 I_n + \sigma_{ey}^2 J_n J_n'$, we have $\Omega^{-1} = \frac{1}{\sigma_{ey}^2}\{I_n - \frac{\sigma_v^2 J_n J_n'}{\sigma_{ey}^2 + n\sigma_v^2}\} = \frac{1}{\sigma_{ey}^2}\left\{I_n - \pi J_n J_n'\right\}$ with $\pi = \frac{\sigma_v^2}{\sigma_{ey}^2 + n\sigma_v^2}$; the expression of $\widehat{\Omega}^{-1}$ follows. Suppose $\hat{\pi} \to_p w$. Consistent estimates of σ_v^2 and σ_{ey}^2 guarantee that $w = \pi$, which usually requires a consistent estimate of β. Otherwise, an inconsistent estimate of β can lead to an inconsistent estimate of π (i.e., $w \neq \pi$).

The asymptotic bias of $\hat{\beta}_I$ can be calculated as follows:

$$\begin{aligned}\hat{\beta}_I - \beta &= \frac{m^{-1}\sum_{i=1}^{m}\left[\widehat{T}_i'\left(I_n - \hat{\pi}J_nJ_n'\right)\left(b_{0i}J_n + P_i\beta_p + \epsilon_i^y\right)\right]}{m^{-1}\sum_{i=1}^{m}\left[\widehat{T}_i'\left(I_n - \hat{\pi}J_nJ_n'\right)\widehat{T}_i\right]} \\ &\to_p \frac{lim_{m\to\infty}m^{-1}\sum_{i=1}^{m}\left[\left(\alpha_{0i}J_n + P_i\alpha_p + \epsilon_i^t\right)'Q\left(I_n - \hat{\pi}J_nJ_n'\right)\left(b_{0i}J_n + P_i\beta_p + \epsilon_i^y\right)\right]}{lim_{m\to\infty}m^{-1}\sum_{i=1}^{m}[\left(\alpha_{0i}J_n + P_i\alpha_p + \epsilon_i^t\right)'Q\left(I_n - \hat{\pi}J_nJ_n'\right)Q\left(\alpha_{0i}J_n + P_i\alpha_p + \epsilon_i^t\right)]} \\ &\to_p \frac{\alpha_p'V_p\beta_p(1-nw)}{n\sigma_a^2(1-nw) + \left(\alpha_p'V_p\alpha_p + \sigma_{\epsilon t}^2\right)(1-nw)} \\ &\to_p \frac{\alpha_p'V_p\beta_p/n}{\sigma_a^2 + \left(\alpha_p'V_p\alpha_p + \sigma_{\epsilon t}^2\right)/n}.\end{aligned} \tag{9}$$

Note that w is conveniently canceled out from the formula. Because $E(P_i) = 0$, $V(P_i) = E(P_i'P_i)$, which is denoted as V_p in the formula. See Appendix for more details of the derivations. When n_is are not the same, n in (9) should be replaced by the mean sample size of the clusters (by applying the Law of Large Numbers [LLN]).

In a preference-based IV analysis, when the number of clusters increases, the number of categories or values of the IV variable also increases. Based on the bias formula, when there is no unmeasured confounding (i.e., $\alpha_p = 0$ or $\beta_p = 0$), the 2SGLS estimator of β is consistent. When $m \to \infty$ but the cluster size n is finite, the estimator is inconsistent when unmeasured confounders exist. The magnitude of the asymptotic bias is impacted by the magnitudes of n, α_p, β_p, V_p, σ_a^2 and $\sigma_{\epsilon t}^2$. Subsequently, we cannot consistently estimate σ_v^2 or σ_{ey}^2 or Var$\left(\hat{\beta}_I\right)$. Only when both $n \to \infty$ and $m \to \infty$, the asymptotic bias of $\hat{\beta}_I$ disappears, and $\hat{\beta}_I$ becomes a consistent estimator. This consistency property of the 2SGLS IV estimator is unique and different from several other estimators. For example, for OLS, 2SLS, or LM estimators, their asymptotic bias usually disappears when the number of independent units goes to infinity.

Regular regression estimators such as OLS and LM estimators are unbiased not only in large samples but also in small samples when the assumptions for the corresponding models are met. In contrast, neither 2SLS nor 2SGLS estimators are unbiased in small samples, even when there are no violations of assumptions. The weaker the instruments, the more biased these estimators are in small samples, particularly when there are multiple instruments [34, 35] or multiple clusters (in our setting). The strength of the instruments evaluates the correlation between the instruments and endogenous variables and can be measured by the first-stage partial F statistic [35]. In the absence of unmeasured between-cluster confounding, with the true model (5), we have E(F) $= \frac{\left(n-1-\frac{1}{m}\right)\sigma_a^2}{\alpha_p'V_p\alpha_p+\sigma_{\epsilon t}^2}$. Similar to the 2SLS estimator, we expect that the relative bias of the 2SGLS estimator to the OLS estimator in finite samples is approximately $\frac{1}{E(F)+1}$ [36]. Similar to 2SLS, obtaining F statistics can give us some insight into the relative bias in practice; the larger the F value, the stronger the IVs, and the smaller the relative bias of the 2SGLS estimator. Based on the expression of F, we note that large σ_a^2, large n, small β_p, and small V_p lead to large F. See Angrist and Pischke [19] for related discussions on 2SLS. In this manuscript, we focus on the closed-form expression of asymptotic bias we derived, which can provide guidance to study design, data analysis, and sensitivity analysis.

5. Simulation

We conduct simulations with several objectives in mind: (i) to examine how well the asymptotic bias formula approximates bias in finite samples when IVs are valid; (ii) to further reveal patterns of bias and examine factors that influence bias and the coverage rate (CR) of confidence intervals (CI) of the IV estimates when IVs are valid; (iii) to investigate the bias patterns when IVs are invalid due to the presence of unmeasured between-cluster confounders; and (iv) to compare IV analyses with LMs commonly used

to analyze clustered data. We simulate T_{ij} and Y_{ij} using the true models (5) and (6) which are specified as follows:

$$T_{ij} = a_{0i} + \alpha_{1c} + \alpha_{2c}C_{2ij} + \alpha_{3c}C_{3i} + \alpha_{1p}P_{1ij} + \alpha_{1g}G_{1i} + \epsilon^t_{ij},$$
$$Y_{ij} = b_{0i} + \beta_{1c} + \beta T_{ij} + \beta_{2c}C_{2ij} + \beta_{3c}C_{3ij} + \beta_{1p}P_{1ij} + \beta_{1g}G_{1i} + \epsilon^y_{ij}.$$

We first conduct simulations without the presence of unmeasured between-cluster confounding, that is, $\alpha_{1g} = 0$ and $\beta_{1g} = 0$. Because the simulations all demonstrate similar patterns with a wide range of combinations of parameter values, we only present the results when we vary one parameter while holding other parameters constant with the following default parameter specifications: $m = 200, n = 20, \alpha_{1c} = 18, \alpha_{2c} = -1, \alpha_{3c} = -1, \alpha_{1p} = 0.6, \alpha_{1g} = 0, \beta_{1c} = 3, \beta = 0.7, \beta_{2c} = 1, \beta_{3c} = 1, \beta_{1p} = 0.6, \beta_{1g} = 0$. We specify that $C_{2ij} \sim N(0,1)$, $C_{3i} \sim N(11,1)$, $P_{1ij} \sim N(1,1)$, $G_{1i} \sim N(1,1)$, $\epsilon^t_{ij}, \epsilon^y_{ij} \sim N(0,1)$, $a_{0i} \sim N(0, 0.3^2)$ and $b_{0i} \sim N(0,1)$. Note that previously, in order to simplify the bias derivation, we assumed mean zeros for P and no presence of C; in these simulations, we allow the presence of C and non-zero means for T, C, and P. We then conduct simulations allowing the presence of unmeasured between-cluster confounding with corresponding changes to the previous default parameter specifications, that is, $\alpha_{1p} = 0, \alpha_{1g} = 0.6, \beta_{1p} = 0, \beta_{1g} = 0.6$. For each set of parameter specifications, we simulate 1000 data sets. For each data set, we estimate β using both IV and LM methods. For each parameter specification, we report bias of the estimates from each method averaged over 1000 simulations and the bias calculated directly from the bias formula for the 2SGLS estimator. Additionally, we report the CR of CIs of the IV and LM estimates.

Figures 1, 2, and 3 show results when there is no unmeasured between-cluster confounding, specifically, the bias and CR when the number of clusters (m) and the cluster size (n) vary, when the unmeasured within-cluster confounder effect on T (α_{1p}) and on Y (β_{1p}) vary, and when the variance of the random effect for T $\left(\sigma^2_a\right)$ and the variance of the within-cluster unmeasured confounder $\left(\sigma^2_p\right)$ vary, respectively. Figure 1 shows that, when m is small ($m = 10$), the asymptotic bias formula for the IV estimator approximates bias in finite samples not very well, but not badly either. As m increases to $m = 40$ and then $m = 200$, the bias formula approximates bias in finite samples quite well across a range of n (Figure 1), α_{1p}, β_{1p} (Figure 2), σ^2_p, and σ^2_a (Figure 3). Regardless of m, as n decreases, the IV bias increases approaching the bias of the LM estimates, and the CR also decreases; as n increases, the IV bias decreases toward zero, and the CR increases toward 95%. Figure 2 shows that, when α_{1p} increases, the bias of IV estimates increases before decreasing, reflecting the nonlinear relationship between bias and α_{1p} captured in the bias formula. On the other hand, the increase of β_{1p} linearly increases bias of IV estimates. The CR monotonically decreases as α_{1p} and β_{1p} increase. Figure 3 shows that, when σ^2_a increases, that is, the amount of randomness of the treatment assignment captured by the preference-based instruments increases, the IV bias disappears quickly, and the CR increases to 95%. As σ^2_p increases, the IV bias increases, and the CR decreases. In almost all scenarios, when an unmeasured within-cluster confounder P_{1ij} exists, IV estimates are much less biased and have higher CRs than LM estimates.

Figure 4 presents the bias patterns when the unmeasured between-cluster confounding effect exists. When α_{1g} increases, the bias of IV estimates increases before decreasing. Meanwhile, the increase of β_{1g} linearly increases bias of IV estimates. Compared with LM estimates, IV estimates are much more biased and have lower CRs.

6. A case study

We illustrate the use of IV analysis using the DOPPS Phase 3 data (2005–2008) with 1434 dialysis patients in 67 facilities to estimate the effect of ESA on Hgb levels. Although it is well-known that ESA increases Hgb levels [37], several cross-sectional studies have found that patients who received the higher ESA dosage tended to have lower Hgb levels [38]. These studies may have not fully adjusted for all the major confounders or correct functional forms of the ESA prescription history in order to accurately capture the temporal relationships; thus, they failed to estimate the causal effect correctly. For illustration purposes, we use Hgb at the 14th month as the outcome (denoted by Hgb_0) and the ESA dosage one month prior to Hgb_0 ($\text{ESA}_{(-1)}$) as the main exposure since it typically takes about 4 weeks for ESA to show most of its effect on Hgb [37].

Our investigation of the DOPPS data shows that one of the major confounding factors omitted by previous work may have been the ESA responsiveness. We find that the ESA dosage at two months prior

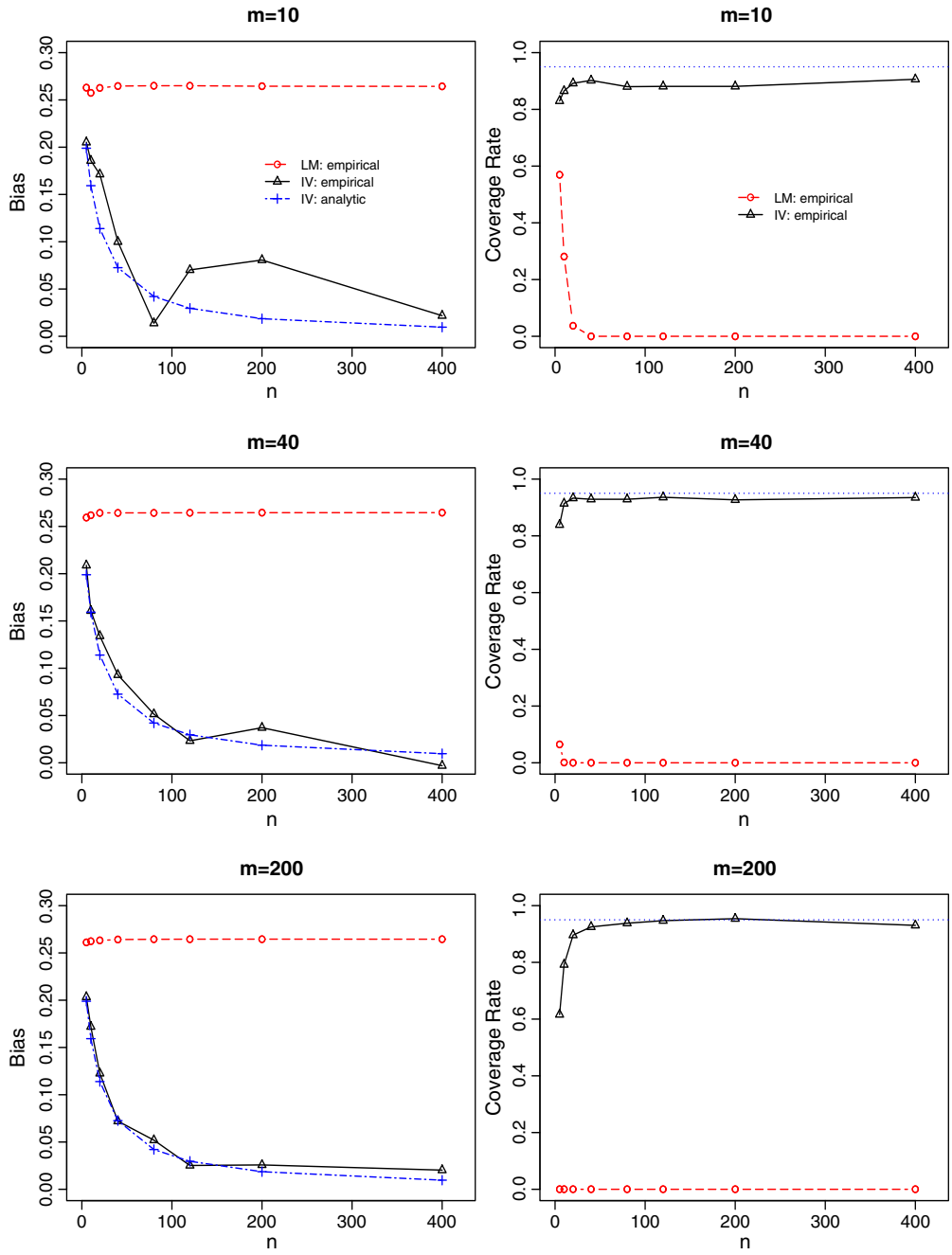

Figure 1. Bias and coverage rate of the IV and LM estimates. The true effect $\beta = 0.7$. m: the number of clusters, n: the cluster size, empirical: averaged over 1000 simulations, analytic: based on the asymptotic bias formula.

to Hgb_0 (denoted by $ESA_{(-2)}$) and Hgb at one month prior ($Hgb_{(-1)}$) strongly influence $ESA_{(-1)}$ and Hgb_0 and have been able to effectively capture ESA responsiveness. The rationale is that, if $ESA_{(-2)}$ is high and $Hgb_{(-1)}$ is still low, the patient may not be very responsive, and his physician will likely increase the $ESA_{(-1)}$ dosage. In contrast, the physician will likely keep $ESA_{(-1)}$ dosage low in a patient with low $ESA_{(-2)}$ and high $Hgb_{(-1)}$. As data illustration, we estimate the effect of $ESA_{(-1)}$ on Hgb_0 using both IV and LM methods without adjusting for the two confounders, $ESA_{(-2)}$ and $Hgb_{(-1)}$. The adjusted covariates include patients characteristics (i.e., age, sex, race, years on dialysis, history of coronary artery disease, congestive heart failure, cancer, cerebrovascular disease, diabetes, gastrointestinal bleeding, peripheral vascular disease, hypertension, intravenous iron, psychiatric disorder, intravenous iron use) as well as facility quality indicators (i.e., percentages of patients having albumin level < 3.5 g/dL, catheter use, phosphorus level > 5.5 mg/dL, and single pool kt/V < 1.2).

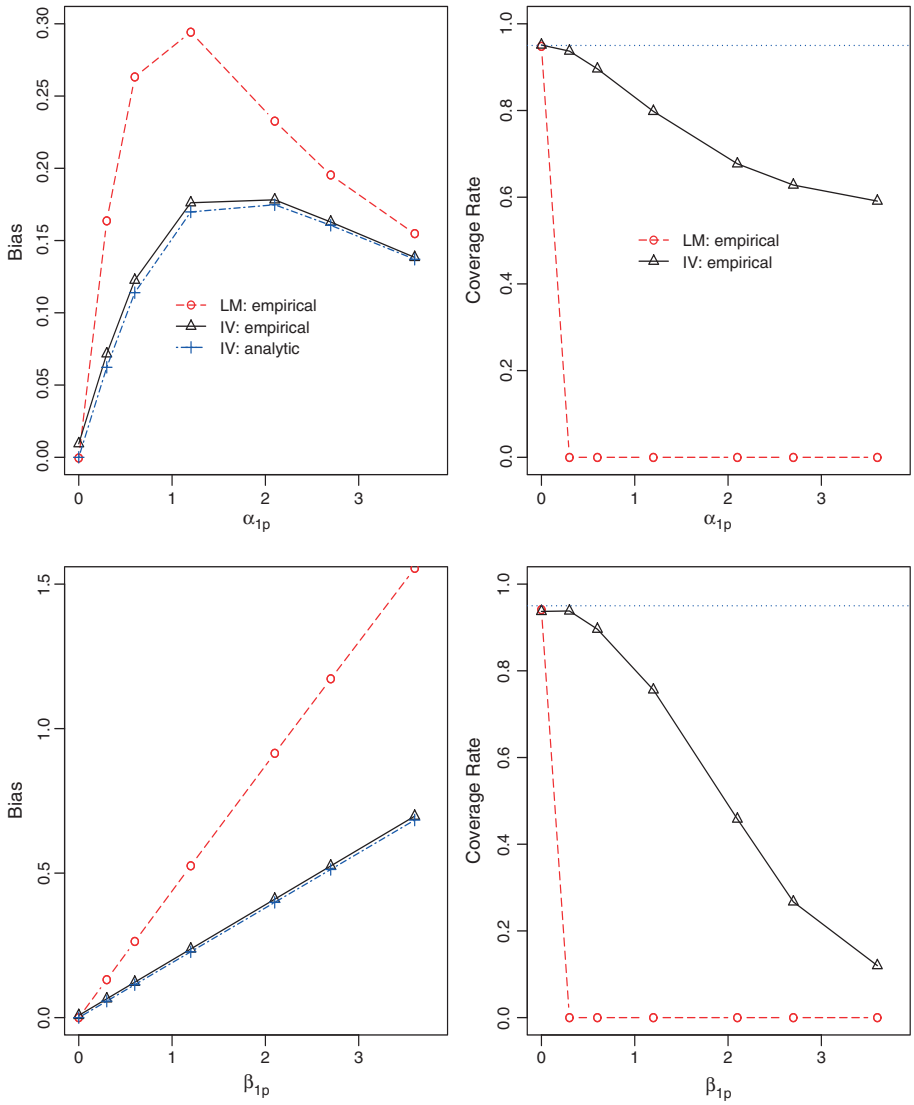

Figure 2. Bias and coverage rate of the IV and LM estimates. The true effect $\beta = 0.7$. α_{1p}: the effect of unmeasured within-cluster confounding on the treatment, β_{1p}: the effect of unmeasured within-cluster confounding on the outcome, empirical: averaged over 1000 simulations, analytic: based on the asymptotic bias formula.

We first examine whether the effects of the two unadjusted confounders that capture ESA responsiveness on outcome are largely within clusters or between clusters. We decompose these two variables into within-cluster components (denoted by $\text{ESA}_{(-2p)}$ and $\text{Hgb}_{(-1p)}$ respectively) and between-cluster components (denoted by $\text{ESA}_{(-2g)}$ and $\text{Hgb}_{(-1g)}$ respectively), adjusting for covariates. Then, we regress Hgb_0 on both between-cluster and within-cluster components adjusting for other covariates. The partial F statistics (p-values) are $15.12(p = 0.0001)$ for $\text{ESA}_{(-2p)}$, $0.03(p = 0.871)$ for $\text{ESA}_{(-2g)}$, $1455.74(p < .0001)$ for $\text{Hgb}_{(-1p)}$, and $320.97(p < 0.0001)$ for $\text{Hgb}_{(-1g)}$. Based on these F statistics, although there is evidence of between-cluster confounding effect of $\text{Hgb}_{(-1)}$, the within-cluster confounding effects appear to be much larger for both variables.

We then investigate the properties of the ESA preference of dialysis facilities as a potential instrument. We examine the strength of the IV, and its corresponding first-stage partial F statistic is calculated as 6.2, which is relatively small and indicates stronger instruments may be explored/preferred (see related discussions on weak instruments [18]). We subsequently examine the associations between the instrument and the unadjusted variables (to put it more accurately, their independent contributions beyond the adjusted covariates). These additional contributions are represented by the residuals from regressing the unadjusted variables on adjusted covariates. The facility ESA preference is estimated as the covariate-adjusted facility average $\text{ESA}_{(-1)}$ by fitting Equation (3) with the adjusted covariates set to their respective

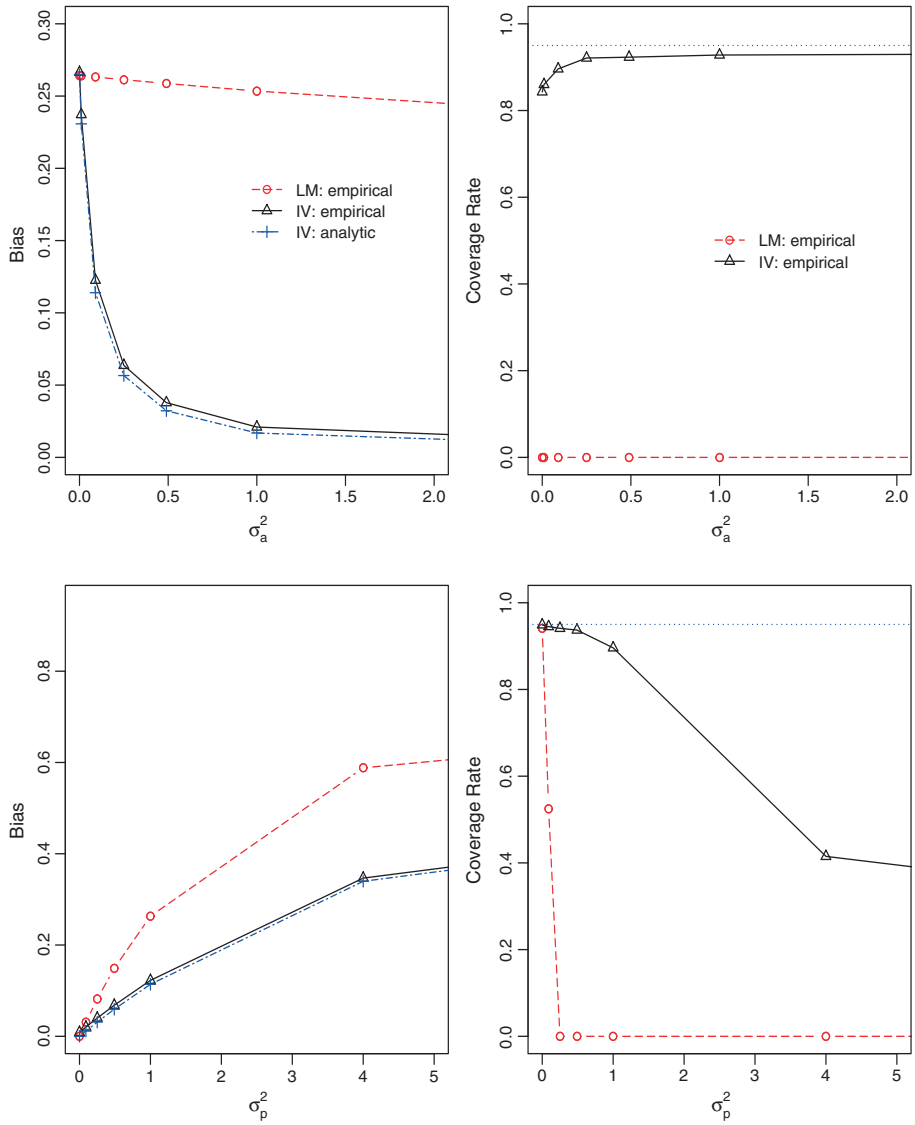

Figure 3. Bias and coverage rate of the IV and LM estimates. The true effect $\beta = 0.7$. σ_a^2: the variance of random treatment assignment across groups, σ_p^2: the variance of the unmeasure within-group confounder, empirical: averaged over 1000 simulations, analytic: based on the asymptotic bias formula.

means. We calculate the Pearson correlation coefficient (CC) between the residuals of $\text{ESA}_{(-2)}$ and the facility preference as $0.214 (p < 0.0001)$; although the CC is statistically significant, it is much smaller than the CC between the residuals of $\text{ESA}_{(-2)}$ and $\text{ESA}_{(-1)}$ (i.e., $0.625 (p < 0.0001)$). Similarly, the CC between the residuals of $\text{Hgb}_{(-1)}$ and the facility preference is $0.061 (p = 0.022)$, which is smaller than the CC between the residuals of $\text{Hgb}_{(-1)}$ and $\text{ESA}_{(-1)}$ (i.e., $-0.110 (p < 0.0001)$).

Finally, we conduct data analysis without adjusting for $\text{ESA}_{(-2)}$ and $\text{Hgb}_{(-1)}$. The IV analysis gives a positive association between $\text{ESA}_{(-1)}$ and Hgb_0, with an estimate of 0.047 and 95% CI of $(0.016, 0.078)$; and the estimate is qualitatively consistent with the established effect of ESA on Hgb. In contrast, LM yields an estimate of -0.010 with 95% CI of $(-0.016, -0.003)$, indicating an inverse association. We fit another LM model by adding $\text{Hgb}_{(-1)}$ and $\text{ESA}_{(-2)}$ to the model, and it results in a positive, albeit small, association of 0.017, with 95% CI of $(0.010, 0.023)$. Hence, without fully adjusting for important confounders, the IV method gives results qualitatively consistent with the truth. Even though the previous analyses in investigating the properties of the IV show that the facility preference is not a perfect instrument, the IV and LM analysis comparison demonstrates the potential advantage of the preference-based IV approach over LM in reducing the impact of unmeasured within-cluster confounding. The IV analysis

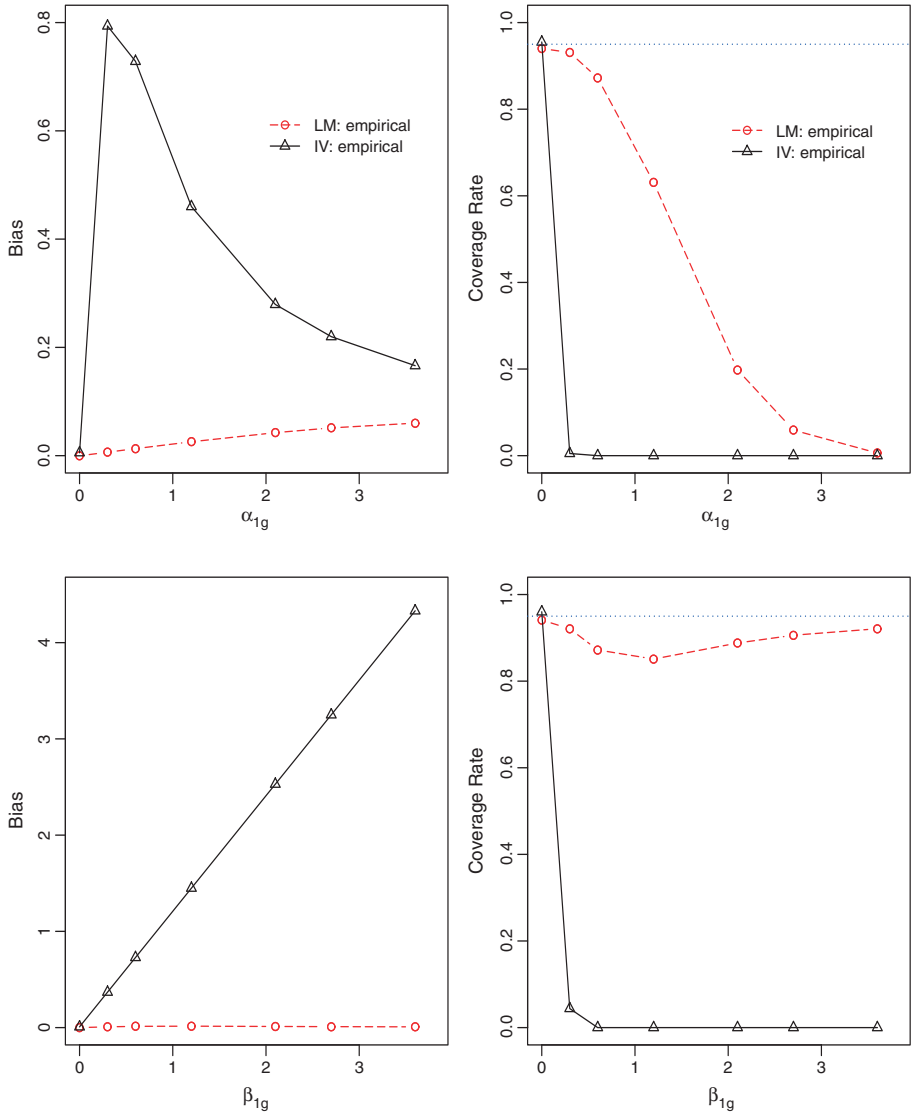

Figure 4. Bias and coverage rate of the IV and LM estimates. The true effect $\beta = 0.7$. α_{1g}: the effect of unmeasured between-cluster confounding on the treatment, β_{1g}: the effect of unmeasured between-cluster confounding on the outcome, empirical: averaged over 1000 simulations.

in this case study may outweigh its limitations, be potentially more beneficial than the LM analysis, and provide helpful information about the treatment effect, particularly considering that it is unlikely to ever find a perfectly valid instrument in practice [4].

7. Discussion

Preference-based instruments have often been used in IV analyses to reduce bias due to confounding-by-indication. However, to our knowledge, few formal evaluations of their properties exist in the literature; instead, findings in settings without clustering or multi-valued IVs or measured confounders have often been borrowed and applied. Our research attempts to fill this void. Several unique features of the data structure with preference-based instrument variables include the following: (i) correlations of the outcomes within groups, possibly due to unadjusted between-cluster factors; (ii) increasing number of groups implying increasing number of IV values or categories; (iii) unmeasured confounding effects occurring between-groups and within-groups. We investigated the assumptions required for preference-based IV analyses. We derived a closed-form expression of asymptotic bias of a 2SGLS IV estimator and assessed its consistency property when IVs are valid. We conducted simulations to evaluate how well the bias formula approximates bias in finite samples and to reveal bias patterns. Our bias assessment provides

researchers with guidance in the use of preference-based IVs. We also examined the magnitude of bias through simulations when IVs are not valid in the presence of unmeasured between-cluster confounders.

We have introduced a framework of between-cluster and within-cluster confounding in the context of studying the preference-based IV analyses. For preference-based IVs to utilize the random component of the treatment assignment across groups free of unmeasured confounding, unmeasured within-group confounding may exist, but unmeasured between-group confounding must not. In practice, disease severity is a major cause of confounding-by-indication and often inadequately measured in databases. It may be reasonable to assume that the disease severity (or other indicators of treatment) is usually confounding mostly at the within-group level. In such scenarios, IV estimators may be preferred to LM or OLS estimators. However, if certain groups systematically admit patients with more severe diseases, unmeasured confounding occurs at both within-group and between-group levels. The treatment preferences across clusters will no longer be valid IVs even after covariate adjustments because they are contaminated by unmeasured between-cluster confounding from disease severity. In addition to patient characteristics aggregated at the group level, other sources of unmeasured between-group confounding come from group-level characteristics and practice patterns. For example, hospitals that more frequently prescribe the treatment of interest may also be more (or less) likely to adhere to practice guidelines and provide better (or worse) care that may confound the treatment effect on outcome. In such scenarios, IV estimators may not be advantageous to LM or OLS estimators if the unadjusted between-cluster confounding is relatively large compared with the within-cluster confounding. However, if additional unadjusted practice and treatment patterns at these hospitals are not $T - Y$ confounders and are uncorrelated with the prescription pattern of the treatment of interest at these hospitals, IV estimators can remain advantageous. Hence, it is important to measure and control for as many between-group confounding effects as we can to reduce the likelihood of violating IV assumptions. With our model specifications, the previous statements are consistent with the assumptions that γ_i needs to only capture the variation of treatment preference across groups and that $\gamma_i \perp v_i$ (conditional on adjusted covariates). Certain statistical tests can help examine whether group-preferences satisfy the IV assumptions. For example, the Sargan tests can provide information on whether any of the instruments are associated with unadjusted confounders [39, 40]. The Durbin-Wu-Hausman test can be used to examine the endogeneity of treatment [41–43]. Although these tests can help guide IV analyses, to our knowledge, there is no empirical method for knowing if these assumptions are satisfied, and no statistical tests are available to test $\gamma_i \perp v_i$ in our model specification when unmeasured confounders exist. Consequently, IV analysts need to apply substantive knowledge of their subject matter and study design to assess the plausibility of these assumptions [29, 44]. Be cautioned that, when an instrument is invalid such that unmeasured between-cluster confounding exists, a small violation of IV assumptions can lead to a large bias, as demonstrated through our simulations. More targeted work is currently being conducted to investigate the properties of the preference-based IV estimators when these key assumptions are not satisfied.

We have derived a closed-form expression for asymptotic bias of the 2SGLS IV estimator when group treatment preferences serve as valid instruments and correlations exist among outcomes within the same clusters, a setting commonly occurring in biomedical research. In finite samples, the estimated variance is one component of the IV estimator, which makes the assessment of the finite-sample bias intractable. However, our asymptotic bias proves to be a very good alternative and approximates the finite-sample bias very well except when the number of clusters is very small. Even when IV is valid, an important finding in our bias analysis is that the 2SGLS IV estimator is usually asymptotically biased when the number of clusters goes to infinity but the cluster size is finite. When the cluster size increases, asymptotic bias decreases. When the cluster size also goes to infinity, the IV estimator becomes asymptotically unbiased, that is, consistent. This property differs from the 2SLS estimator. In future work, it will be interesting to make correction to the finite-sample bias and extend the work to other types of models and endpoints [45–47].

Our simulations and bias formula provide some practical guidance in the use of preference-based instruments during study design and data analysis. When investigators suspect there may be relatively large confounding-by-indication due to unmeasured variables whose distributions do not differ much across groups and whose confounding effects are mostly within clusters, the preference-based IV method may be preferred over LM. Investigators should also try to incorporate as many measured confounders at both between-group and within-group levels as possible, particularly the ones that have strong effects on the outcome. Incorporating these measured confounders can effectively reduce the effects of either unmeasured confounders on the outcome (β_p) or the variance of an unmeasured confounder (V_p), as both are positively associated with larger bias of IV estimates and lower CRs of their CIs. Further, whenever

possible, investigators should select groups that maximize the proportion of random variation of treatment assignments across groups as instruments. Additionally, increasing the cluster size can reduce bias. Investigators should also account for correlations among outcomes within groups in the preference-based IV analysis because it helps explain other sources of variation in outcomes not explained by measured variables, but contributing to the correlation within groups. The first-stage partial F-statistics provide useful guidance for the strength of instruments and the extent of bias of the IV estimator relative to the OLS estimator. Additionally, the closed-form bias formula provides investigators with a method to conduct sensitivity analyses [48] in order to measure how robust the results of preference-based IV analyses are toward hypothetical unmeasured confounders. Additional sensitivity analysis methods have been developed for different IV methods to examine how sensitive the conclusions from an IV analysis are toward plausible violations of key assumptions [5, 6, 9, 49]. Note that we obtained OLS estimators using a fixed effect regression for the first-stage model (3), alternatively, a random effects estimator can be explored [50]. Here, we focus on the asymptotic distribution of the 2SGLS estimator; however, it will be worthwhile to investigate how to obtain the alternative approximation to the distribution of the IV estimator [51] in our setting, which may improve the asymptotic approximation. While we consider the use of preference-based instruments in the form presented in this paper, many variants of the instruments have been used in literature, such as percentages of patients with treatment usage (or adjusted mean dosage) at group levels in current or prior times [29, 52]. They perform generally similar to preference-based instruments in the current form, but with some important differences. This topic has been part of our ongoing investigation effort and warrants a separate in-depth discussion.

In conclusion, when an unmeasured confounding effect exists solely within clusters, group preferences can satisfy the assumptions for valid instruments. When group-preferences are valid instruments, the 2SGLS IV estimator remains advantageous in reducing bias from unmeasured confounding. To reduce the biases of IV estimates in finite samples, investigators should adjust for as many measured confounders as possible, consider groups that capture the most random variation in treatment assignments, and increase cluster size. To reduce the likelihood of violating IV assumptions, investigators should control for confounders that pose between-group contextual confounding effects.

Appendix *(Bias of a Preference-Based Instrumental Variable Estimator in the Presence of Unmeasured Within-Cluster Confounding)*

The bias of $\widehat{\beta}_I$ can be calculated as follows:

$$\begin{aligned}\hat{\beta}_I - \beta &= \frac{m^{-1}\sum_{i=1}^{m}\left[\widehat{T}_i'\left(I_n - \hat{\pi}J_nJ_n'\right)\left(P_i\beta_p + b_{0i}J_n + \epsilon_i^y\right)\right]}{m^{-1}\sum_{i=1}^{m}\left[\widehat{T}_i'\left(I_n - \hat{\pi}J_nJ_n'\right)\widehat{T}_i\right]^{-1}} \\ &\equiv B_I^{-1}A_I\end{aligned}$$

For the numerator A_I:

$$\begin{aligned}A_I &= m^{-1}\sum_{i=1}^{m}\left[\widehat{T}_i'\left(I_n - \hat{\pi}J_nJ_n'\right)\left(P_i\beta_p + b_{0i}J_n + \epsilon_i^y\right)\right] \\ &= m^{-1}\sum_{i=1}^{m}\left[\left(\alpha_{0i}J_n + P_i\alpha_p + \epsilon_i^t\right)'Q\left(I_n - \hat{\pi}J_nJ_n'\right)\left(P_i\beta_p + b_{0i}J_n + \epsilon_i^y\right)\right] \\ &= m^{-1}\sum_{i=1}^{m}\Big[\alpha_{0i}J_n'QP_i\beta_p + \alpha_p'P_i'QP_i\beta_p + \left(\epsilon_i^t\right)'QP_i\beta_p \\ &\quad - \alpha_{0i}J_n'Q\hat{\pi}J_nJ_n'P_i\beta_p - \alpha_p'P_i'Q\hat{\pi}J_nJ_n'P_i\beta_p - \left(\epsilon_i^t\right)'Q\hat{\pi}J_nJ_n'P_i\beta_p \\ &\quad + \alpha_{0i}J_n'Qb_{0i}J_n + \alpha_p'P_i'Qb_{0i}J_n + \left(\epsilon_i^t\right)'Qb_{0i}J_n \\ &\quad - \alpha_{0i}J_n'Q\hat{\pi}J_nJ_n'b_{0i}J_n - \alpha_p'P_i'Q\hat{\pi}J_nJ_n'b_{0i}J_n - \left(\epsilon_i^t\right)'Q\hat{\pi}J_nJ_n'b_{0i}J_n \\ &\quad + \alpha_{0i}J_n'Q\epsilon_i^y + \alpha_p'P_i'Q\epsilon_i^y + \left(\epsilon_i^t\right)'Q\epsilon_i^y \\ &\quad - \alpha_{0i}J_n'Q\hat{\pi}J_nJ_n'\epsilon_i^y - \alpha_p'P_i'Q\hat{\pi}J_nJ_n'\epsilon_i^y - \left(\epsilon_i^t\right)'Q\hat{\pi}J_nJ_n'\epsilon_i^y\Big]\end{aligned}$$

As $m \to \infty$, we apply the Law of Large Numbers (LLN) for i.i.d. sequences, and we have

$$\begin{aligned}
A_I \to_p & \; lim_{m\to\infty} \Big\{ E\left(\alpha_{0i} J_n' Q P_i \beta_p\right) + \alpha_p' E\left(P_i' Q P_i\right) \beta_p + E\left[\left(\epsilon_i^t\right)' Q P_i\right] \beta_p \\
& - E\left(\alpha_{0i} J_n' Q \hat{\pi} J_n J_n' P_i\right) \beta_p - E\left(\alpha_p' P_i' Q \hat{\pi} J_n J_n' P_i \beta_p\right) - E\left[\left(\epsilon_i^t\right)' Q \hat{\pi} J_n J_n' P_i \beta_p\right] \\
& + E\left(\alpha_{0i} J_n' Q b_{0i} J_n\right) + E\left(\alpha_p' P_i' Q b_{0i}\right) J_n + E\left[\left(\epsilon_i^t\right)' Q b_{0i} J_n\right] \\
& - E\left(\alpha_{0i} J_n' Q \hat{\pi} J_n J_n' b_{0i}\right) J_n - E\left(\alpha_p' P_i' Q \hat{\pi} J_n J_n' b_{0i} J_n\right) - E\left[\left(\epsilon_i^t\right)' Q \hat{\pi} J_n J_n' b_{0i}\right] J_n \\
& + E\left(\alpha_{0i} J_n' Q \epsilon_i^y\right) + E\left(\alpha_p' P_i' Q \epsilon_i^y\right) + E\left[\left(\epsilon_i^t\right)' Q \epsilon_i^y\right] \\
& - E\left(\alpha_{0i} J_n' Q \hat{\pi} J_n J_n' \epsilon_i^y\right) - E\left(\alpha_p' P_i' Q \hat{\pi} J_n J_n' \epsilon_i^y\right) - E\left[\left(\epsilon_i^t\right)' Q \hat{\pi} J_n J_n' \epsilon_i^y\right] \Big\} \\
= & \; \alpha_p' E(P_i' Q P_i) \beta_p - E(\alpha_p' P_i' Q \hat{\pi} J_n J_n' P_i \beta_p) \\
= & \; \alpha_p' V_p \beta_p - nw \alpha_p' V_p \beta_p \\
= & \; \alpha_p' V_p \beta_p (1 - nw)
\end{aligned}$$

For the denominator B_I:

$$\begin{aligned}
B_I &= m^{-1} \sum_{i=1}^{m} \left[\hat{T}_i' \left(I_n - \hat{\pi} J_n J_n'\right) \hat{T}_i\right] \\
&= m^{-1} \sum_{i=1}^{m} \left[\left(\alpha_{0i} J_n + P_i \alpha_p + \epsilon_i^t\right)' Q \left(I_n - \hat{\pi} J_n J_n'\right) Q \left(\alpha_{0i} J_n + P_i \alpha_p + \epsilon_i^t\right)\right] \\
&= m^{-1} \sum_{i=1}^{m} \Big[\alpha_{0i}^2 J_n' Q J_n + \alpha_{0i} \alpha_p' P_i' Q J_n + \alpha_{0i} \left(\epsilon_i^t\right)' Q J_n \\
&\quad - \alpha_{0i}^2 J_n' Q \hat{\pi} J_n J_n' J_n - \alpha_{0i} \alpha_p' P_i' \hat{\pi} J_n J_n' Q J_n - \alpha_{0i} \hat{\pi} \left(\epsilon_i^t\right)' J_n J_n' Q J_n \\
&\quad + \alpha_{0i} J_n' Q P_i \alpha_p + \alpha_p' P_i' Q P_i \alpha_p + \left(\epsilon_i^t\right)' Q P_i \alpha_p \\
&\quad - \alpha_{0i} \hat{\pi} J_n' Q J_n J_n' P_i \alpha_p - \hat{\pi} \alpha_p' P_i' Q J_n J_n' P_i \alpha_p - \hat{\pi} \left(\epsilon_i^t\right)' Q J_n J_n' P_i \alpha_p \\
&\quad + \alpha_{0i} J_n' Q \epsilon_i^t + \alpha_p' P_i' Q \epsilon_i^t + \left(\epsilon_i^t\right)' Q \epsilon_i^t \\
&\quad - \alpha_{0i} Q \hat{\pi} J_n' J_n J_n' \epsilon_i^t - \hat{\pi} \alpha_p' P_i' Q J_n J_n' \epsilon_i^t - \hat{\pi} \left(\epsilon_i^t\right)' Q J_n J_n' \epsilon_i^t\Big]
\end{aligned}$$

As $m \to \infty$, we apply the LLN for i.i.d. sequences, and we have

$$\begin{aligned}
B_I \to_p & \; lim_{m\to\infty} \Big\{ E\left(\alpha_{0i}^2\right) J_n' Q J_n + E(\alpha_{0i}) \alpha_p' E\left(P_i'\right) Q J_n + E(a_{0i}) E\left(\epsilon_i^t\right)' Q J_n \\
& - E\left(\alpha_{0i}^2\right) J_n' E(\hat{\pi}) J_n J_n' Q J_n - E(\alpha_{0i}) \alpha_p' E\left(P_i'\right) E(\hat{\pi}) J_n J_n' Q J_n - E(\alpha_{0i}) E(\hat{\pi}) E\left(\epsilon_i^t\right)' J_n J_n' Q J_n \\
& + E(\alpha_{0i}) J_n' Q E(P_i) \alpha_p + \alpha_p' E\left(P_i' Q P_i\right) \alpha_p + E\left(\epsilon_i^t\right)' Q E(P_i) \alpha_p \\
& - E(\alpha_{0i}) E(\hat{\pi}) J_n' Q J_n J_n' E(P_i) \alpha_p - E(\hat{\pi}) \alpha_p' E\left(P_i'\right) Q J_n J_n' E(P_i) \alpha_p - E(\hat{\pi}) E\left(\epsilon_i^t\right)' Q J_n J_n' E(P_i) \alpha_p \\
& + E(\alpha_{0i}) J_n' Q E\left(\epsilon_i^t\right) + \alpha_p' E\left(P_i' Q \epsilon_i^t\right) + E\left[\left(\epsilon_i^t\right)' Q \epsilon_i^t\right] \\
& - E(\alpha_{0i} \hat{\pi}) J_n' Q J_n J_n' E\left(\epsilon_i^t\right) - E(\hat{\pi}) \alpha_p' E\left(P_i'\right) Q J_n J_n' E\left(\epsilon_i^t\right) - E(\hat{\pi}) E\left[\left(\epsilon_i^t\right)' Q J_n J_n' \epsilon_i^t\right] \Big\} \\
= & \; E\left(\alpha_{0i}^2\right) J_n' Q J_n - E\left(\alpha_{0i}^2\right) J_n' E(\hat{\pi}) J_n J_n' Q J_n + \alpha_p' E\left(P_i' Q P_i\right) \alpha_p \\
& - E(\hat{\pi}) \alpha_p' E\left(P_i'\right) Q J_n J_n' E(P_i) \alpha_p E\left[\left(\epsilon_i^t\right)' Q \epsilon_i^t\right] - E(\hat{\pi}) E\left[\left(\epsilon_i^t\right)' Q J_n J_n' \epsilon_i^t\right] \\
= & \; n \sigma_a^2 - n^2 w \sigma_a^2 + \alpha_p' V_p \alpha_p - nw \alpha_p' V_p \alpha_p + \sigma_{\epsilon t}^2 - nw \sigma_\epsilon^2 \\
= & \; n \sigma_a^2 (1 - nw) + \left(\alpha_p' V_p \alpha_p + \sigma_{\epsilon t}^2\right) (1 - nw)
\end{aligned}$$

Thus, the bias of $\widehat{\beta_I}$ can be approximated as

$$\hat{\beta}_I - \beta \approx \frac{\alpha_p' V_p \beta_p / n}{\sigma_a^2 + \left(\alpha_p' V_p \alpha_p + \sigma_{\epsilon t}^2\right)/n}.$$

Acknowledgements

The authors thank Drs. Dylan Small, Min Zhang, Lu Wang, Jack Kalbfleisch, Brenda Gillespie, and Ronald Pisoni for helpful discussions and Heather Van Doren and Brett Griffiths for editorial assistance on this manuscript. The DOPPS is administered by Arbor Research Collaborative for Health and is supported by scientific research grants from Amgen (since 1996), Kyowa Hakko Kirin (since 1999, in Japan), Sanofi Renal (since 2009), Abbott (since 2009), Baxter (since 2011), and Vifor Fresenius Renal Pharma (since 2011), without restrictions on publications.

References

1. Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, Yood MU, Haque R, Wei F, Lash TL. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *Journal of Clinical Epidemiology* 2010; **63**(1):64–74.
2. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *American Journal of Epidemiology* 1999; **149**(11):981–983.
3. D'Agostino RB Jr. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine* 1998; **17**:2265–2281.
4. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Statistics in Medicine* 2014; **33**: 2297–2340.
5. Brookhart M, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *The International Journal of Biostatistics* 2007; **3**(1):14.
6. Small D, Rosenbaum PR. War and wages: the strength of instrumental variables and their sensitivity to unobserved biases. *Journal of the American Statistical Association* 2008; **103**:924–933.
7. Holland PW. Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology* 1988; **18**:449–484.
8. West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, Szapocznik J, Fishbein M, Rapkin B, Clatts M, Mullen PD. Alternatives to the randomized controlled trial. *American Journal of Public Health* 2008; **98**(8):1359–1366.
9. Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**:444–469.
10. Heckman JT. Randomization as an instrumental variable. *The Review of Economics and Statistics* 1996; **78**:336–341.
11. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; **29**(4):722–729.
12. Baiocchi M, Small DS, Yang L, Polsky D, Groeneveld PW. Near/far matching: a study design approach to instrumental variables. *Health Services and Outcomes Research Methodology* 2012; **12**:237–253.
13. Johnston SC, Henneman T, McCulloch CE, van der Laan M. Modeling treatment effects on binary outcomes with grouped-treatment variables and individual covariates. *American Journal of Epidemiology* 2002; **156**(8):753–760.
14. Brookhart MA, Schneeweiss S, Avorn J, Bradbury BD, Liu J, Winkelmayer WC. Comparative mortality risk of anemia management practices in incident hemodialysis patients. *The Journal of American Medical Association* 2010; **303**(9): 857–864.
15. Korn ED, Baumrind S. Clinician preferences and the estimation of causal treatment differences. *Statistical Science* 1998; **13**(3):209–235.
16. Chen Y, Briesacher BA. Use of instrumental variable in prescription drug research with observational data: a systematic review. *Journal of Clinical Epidemiology* 2011; **64**(6):687–700.
17. Davies NM, Smith GD, Windmeijer F, Martina RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology* 2013; **24**:363–369.
18. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *International Journal of Epidemiololgy* 2011; **40**(3):755–764.
19. Angrist JD, Pischke JS. *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press, 2008.
20. Little R, Yau L. Statistical techniques for analyzing data from prevention trials: treatment of no-shows using rubins causal model. *Psychological Methods* 1998; **3**:147–159.
21. Hirano K, Imbens G, Rubin D, Zhou X. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 2000; **1**(1):69–88.
22. Angrist J, Imbens G. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 1995; **90**:430–442.
23. Abadie A. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* 2003; **113**:231–263.
24. Tan Z. Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association* 2006; **101**:1607–1618.

25. O'Malley A, Frank R, Normand S. Estimating cost-offsets of new medications: use of new antipsychotics and mental health costs for schizophrenia. *Statistics in Medicine* 2011; **30**:1971–1988.
26. Okui R, Small D, Tan Z, Robins J. Doubly robust instrumental variables regression. *Statistica Sinica* 2012; **22**:173–205.
27. Neuhaus JM. Assessing change with longitudinal and clustered binary data. *Annual Review Public Health* 2001; **22**: 115–128.
28. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998; **54**:638–645.
29. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology* 2006; **17**(3):268–275.
30. Imbens G, Angrist J. Identification and estimation of local average treatment effects. *Econometrica* 1994; **62**(2):467–475.
31. Balke A, Pearl J. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* 1997; **92**:1171–1176.
32. White H. Instrumental variables regression with independent observations. *Econometrica* 1982; **50**:483–499.
33. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. MIT press, 2002; 260–262.
34. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association* 1995; **90**(430):443–450.
35. Stock J, Wright J, Yogo M. A survey of weak instruments and weak identification in generalized method of moments. *Journal of the American Statistical Association* 2002; **20**(4):518–529.
36. Staiger D, Stock J. Instrumental variables regression with weak instruments. *Econometrica* 1997; **65**(3):557–586.
37. Eschbach JW, Egrie JC, Downing MR, Browne JK, Adamson JW. Correction of the anemia of end-stage renal disease with recombinant human erythropoietin. Results of a combined phase I and II clinical trial. *New England Journal of Medicine* 1987; **316**(2):73–78.
38. Madore F, Lowrie EG, Brugnara C, Lew NL, Lazarus JM, Bridges K, Owen WF. Anemia in hemodialysis patients: variables affecting this outcome predictor. *Journal of American Society of Nephrololgy* 1997; **8**(12):1921–1929.
39. Sargan JD. The estimation of economic relationships using instrumental variables. *Econometrica* 1958; **26**(3):393–415.
40. Lee Y, Okui R. Hahn-Hausman Test as a Specification Test. *Journal of Econometrics* 2012; **167**:133–139.
41. Durbin J. Errors in variables. *Review of the International Statistical Institute* 1954; **22**:23–32.
42. Wu DM. Alternative tests of independence between stochastic regressors and disturbances. *Econometrica* 1973; **41**(4): 733–750.
43. Hausman J. Specification tests in econometrics. *Econometrica* 1978; **46**(3):1251–1271.
44. Glymour MM, Tchetgen EJ, Robins JM. Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *American Journal of Epidemiology* 2012; **175**(4):332–339.
45. Lee Y. Bias in dynamic panel models under time series misspecification. *Journal of Econometrics* 2012; **169**:54–60.
46. Lee Y. Nonparametric estimation of dynamic panel models with fixed effects econometric theory 2014; **30**:1315–1347.
47. Lee Y, Phillips PCB. Model selection in the presence of incidental parameters. *Journal of Econometrics*. In press.
48. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 1998; **54**:948–963.
49. Small D. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Association* 2007; **102**:1049–1058.
50. Chamberlain G, Imbens G. Random effects estimators with many instrumental variables. *Econometrica* 2004; **72**(1): 295–306.
51. Bekker PA. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 1994; **62**(3):657–681.
52. Newman TB, Vittinghoff E, McCulloch CE. Efficacy of phototherapy for newborns with hyperbilirubinemia: a cautionary example of an instrumental variable analysis. *Medical Decision Making* 2012; **32**(1):83–92.