

Assessing The Performance of Physician's Prescribing Preference As An Instrumental Variable in Comparative Effectiveness Research With Moderate and Small Sample Sizes: A Simulation Study

Lisong Zhang (✉ l.zhang.3@research.gla.ac.uk)

University of Glasgow

Jim Lewsey

University of Glasgow

David McAllister

University of Glasgow

Research Article

Keywords: simulation study, comparative effectiveness research, instrumental variables, Physician's prescribing preference

Posted Date: December 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1138222/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Instrumental variable (IV) analyses are used to account for unmeasured confounding in Comparative Effectiveness Research (CER) in pharmacoepidemiology. To date, simulation studies assessing the performance of IV analyses have been based on large samples. However, in many settings, sample sizes are not large.

Objective

In this simulation study, we assess the utility of Physician's Prescribing Preference (PPP) as an IV for moderate and smaller sample sizes.

Methods

We designed a simulation study in a CER setting with moderate (around 2500) and small (around 600) sample sizes. The outcome and treatment variables were binary and three variables were used to represent confounding (a binary and a continuous variable representing measured confounding, and a further continuous variable representing unmeasured confounding). We compare the performance of IV and non-IV approaches using two-stage least squares (2SLS) and ordinary least squares (OLS) methods, respectively. Further, we test the performance of different forms of proxies for PPP as an IV.

Results

The PPP IV approach results in a percent bias of approximately 20%, while the percent bias of OLS is close to 60%. The sample size is not associated with the level of bias for the PPP IV approach. However, smaller sample sizes led to lower statistical power for the PPP IV. Using proxies for PPP based on longer prescription histories result in stronger IVs, partly offsetting the effect on power of smaller sample sizes.

Conclusion

Irrespective of sample size, the PPP IV approach leads to less biased estimates of treatment effectiveness than conventional multivariable regression adjusting for known confounding only. Particularly for smaller sample sizes, we recommend constructing PPP from long prescribing histories to improve statistical power.

Background

As a source of natural variation, physician's prescribing preference (PPP) has been increasingly used as an instrumental variable (IV) in Comparative Effectiveness Research (CER)(1). Multiple simulation and applied studies have discussed the use of PPP in comparing the effectiveness of two drug classes. In many recent applied papers about PPP IV, they have large sample sizes of around 30,000 (2–9). However,

in many contexts the sample size will be smaller, for example, Nelson and colleagues conducted a PPP study of HIV using a sample size of less than 2000 (10). Smaller sample sizes are likely to occur in studies of rare outcomes or where drugs have only recently become available (e.g. in a single administrative area).

Boef and colleagues argued that the sample size put limits on the performance of IVs (11). Further, they concluded that the bias in IV estimates relative to conventional approaches (e.g. ordinary least squares (OLS)) is determined both by the strength of the IV as well as the strength of unmeasured confounders. With an aim to widen the applicability of PPP IV, we test the performance of the method in moderate and small sample sizes using a simulation study.

Method

In order to be comparable with OLS, we use two-stage least squares (2SLS) as the main statistical method to generate the IV estimates of treatment effectiveness. Despite the fact that 2SLS may cause model misspecification for binary outcomes and treatment, the 2SLS is the most common method and a common starting point for the IV method (12). In addition, in many settings, when the outcome is not rare, the 2SLS estimates generates similar estimates to non-linear two stage regression (prevalence between 1.5–50%) (13, 14).

A summary of how performance was assessed is shown below (See Table 1). We use percent bias to assess the performance of PPP IVs for different levels of unmeasured confounding. The strength of IV is calculated as the F-statistic of the first stage. We use the coverage rate to compare the stability of OLS and 2SLS at the different levels of unmeasured confounding.

Table 1
measurement of performance.

Measurement	Calculation
Percent bias	$\frac{trueRiskDifference - estimatedRiskDifference}{trueRiskDifference} * 100\%$
Coverage rate	% of iterations when 95% CI includes the true risk difference across 1000 simulations
F-statistics of the first stage regression	$F\ statistics = \frac{SumofSquaresforModel / DegreesofFreedomForModel}{\frac{SumofSquaresforError / DegreesofFreedomforError}{MeanofSquaresforErrors}} =$

Simulation design:

Study population

For the moderate sample size study, we set the number of physicians to 80, the lower bound of the number of patients/physicians was 10 and the upper bound of patients/physician is 50. These numbers are The sample size is 2452 in this case. For small sample size study, the number of physicians is set to 20. The smaller sample size is 620. The prevalence of outcomes varies between 20–60%.

Treatment and outcome

In this paper, we focus on scenarios where the treatment and outcome are both binary.

The formula for the probability of being prescribed a certain treatment ($X = 1$) and the probability of the outcome of interest ($Y = 1$) are listed below:

$$Prob(X = 1) = \alpha_0 + \alpha_z PPP + \alpha_1 X1 + \gamma_x X2 + \alpha_3 X3$$

$$Prob(Y = 1) = \beta_0 + \beta_x * Prob(X = 1) + \beta_1 X1 + \gamma_y X2 + \beta_3 X3$$

PPP stands for instrumental variable. We set PPP 70% of chance equals to 1, 30% of chance equals to 0. This imbalance reflects a common situation that treatment providers tend to prefer one type of treatment than another (perhaps based on following clinical guidelines). $X1$ is a binary covariate, and $X2$, $X3$ are continuous covariates. We assume $X1$ and $X3$ are measured covariates and $X2$ is an unmeasured covariate. In the Data generation process, $X1$ follows binominal distribution, $X2$ and $X3$ follows the normal distribution. These are implemented using R functions `rbinom` and `rnorm` (please see code in supplementary material for full details). α_z controls the strength of association between the instrumental variable and exposure. The PPP is the 'true' prescribing preference that in practice is a latent variable and is a binary variable. The parameter values for the data generation process are listed in equation (1) and (2).

The focus of this study is to investigate the impact of unmeasured confounding. Therefore, we keep $\alpha_z = 0.4$ to ensure the IV strength is fixed.

The parameter value for treatment in equation (2) is 0.1 and this represents the 'true' risk difference between the two treatments. β_x is the observed estimate of this risk difference.

$$[Prob(X = 1)] = \alpha_z PPP + 0.053X1 + 0.1X2 + 0.02X3(1)$$

$$[Prob(Y = 1)] = 0.10 * treatment + 0.04X1 + \gamma_y X2 + 0.01X3 - 0.01(2)$$

Draw from the existing literature(1, 2, 4, 6), we constructed the proxies for PPP mainly based on the prescription history. The prior 1 to prior 4 prescriptions are investigated in this study. The prior 1 prescription is the most recent prescription made by the same physician. Likewise, the prior 2 prescription is prior 2 prescriptions from the same physician and the same for prior 3 and prior 4 prescriptions. For example, possible values for prior 4 prescriptions are 0,1,2,3,4. The proportion PPP is the number of certain treatment ($X=1$) divided by the number of all prescriptions made by this physician (See below).

$$\text{Proportion PPP} = \frac{\text{Number of drug made by one physician}}{\text{Number of all prescriptions made by the same physician}}$$

All analysis is done in R studio with R version 3.6.1. The R code that generates the simulated datasets and the regression models can be seen in the supplementary material.

Results

Percent bias:

Figure 1 presents the percent bias of the 2SLS and OLS in moderate and small sample sizes.

OLS is subject to unmeasured confounding bias (See Figure 1). In the case of a lower unmeasured confounding level, the 2SLS is more biased than OLS. The advantage of 2SLS appears after the level of the unmeasured confounding increases. The sample size does not influence the percent bias in general.

The coverage rate shows that the 2SLS covers nominal 95% while the coverage rate of OLS drop dramatically in both sample sizes (see Figure 2).

The strength of the IV increases, and the p-value of the 2SLS treatment effect decreases, as the number of previous prescriptions used in the PPP construction increases (See Figure 3). The level of unmeasured confounding does not influence these results. However, the strength of IV decreases noticeably when the sample size decreases. The ρ_{ZX} does not change much in these two cases (around 0.14 to 0.15) indicating that the strength of the association between the exposure and instruments does change. Rather, it is the sample size that decreases the F-statistic values and makes the IV weaker (see the equation below). The p-values of 2SLS in N=620 sample are consistently larger than that of N=2452 which means the statistical power of 2SLS is limited by the sample size.

$$\text{var}(\hat{\beta}_n^{IV}) = \frac{\sigma_{Y,X}^2}{n\sigma_X^2\rho_{X,Z}^2}$$

$$\text{var}(\hat{\beta}_n^{OLS}) = \frac{\sigma_{Y,X}^2}{n\sigma_X^2}$$

$\sigma_{Y,X}^2$: The residual variance of the outcome after adjusting the treatment (X);

σ_X^2 : The variance of the treatment.

$\rho_{X,Z}^2$: The correlation between the treatment(X) and the instrumental variable (Z).

(15)

$$Fstatistics = \frac{\rho_{ZX}^2(n - 2)}{1 - \rho_{ZX}^2}$$

As an IV, the true preference (PPP in the equation (1)) also shows a strong ability to reduce the unmeasured confounding bias. The F-statistics of true preference reaches 500 which is much higher than all proxies mentioned above (See Figure 9 in the Appendix) which align with the finding from Ionescu-Ittu (14) that the true preference has the smallest variance. The p-values for 2SLS estimates are close to conventional statistical significance (p value <0.05). The bias-variance trade-off for IV methods also exist for the 'true preference' but not as critical as for the proxy PPP indicating stronger instrument reduces the variance of instrumental variable estimates (16). For the reason that time cannot be simulated, we test the time-fixed proxy for PPP (proportion preference). It turns out that the proportion preference is the strongest IV among these proxies. It is associated with the smallest p-value which leads to the 2SLS estimates having small p-values.

As summarised in Figure 4, the 95% confidence intervals of IV estimates narrow as the strength of instruments increases (from prior 1 prescription to the proportional preference). As discussed above, the confidence intervals of OLS estimates are narrower than for 2SLS, and this is also shown in Figure 4. It can also be seen that the OLS estimates are severely biased when the unmeasured confounder covariate parameter (γ_2) is set at a high level. Although the IV estimates are generally less precise, it is feasible when the IV is strong enough that an IV estimate can achieve statistical significance while at the same time reducing the influence of unmeasured confounding bias.

Discussion

The sample size limits the performance of IVs (11, 15). A straightforward explanation for this is that smaller sample sizes make it harder for the IV to meet the relevance assumption. In real life CERs, sample sizes are often large enough which can avoid such pitfalls, but when the outcome of interest is rare or a drug has only recently become available the corresponding CERs will have smaller sample sizes. This simulation aimed to test the performance PPP IV on the different level of unmeasured confounding level and generate supporting evidence that the PPP IV can perform well in reducing bias in studies of moderate or small sample sizes.

Our results show that the 2SLS does reduce the unmeasured confounding to a considerable extent compared to conventional analyses even in a small sample size. As the same time, the standard deviation of 2SLS estimates is generally many times larger than OLS and the confidence interval wide and crossing the null hypothesis. However, if the instrumental variable is strong enough, 2SLS estimates could be statistically significant. We also conducted the same experiment on a larger sample size (N=5869) (See Figure 8 in the Appendix) and even smaller sample size (N=255) (See Figure 5 in the Appendix). In terms of reducing bias, the sample size is not the determinant as it does not impact percent

bias. The figures show that the 2SLS percent bias line is lower than OLS in most cases in N=255 sample. Nevertheless, the sample size does limit the statistical power. In a smaller sample size, the difference of percent bias from 2SLS and OLS can still be an indicator to see if an unmeasured confounding is a major problem although the weak statistical power makes the 2SLS estimate less useful.

The results of this simulation study show that using PPP as an IV is effective at minimising bias caused by unmeasured confounding relative to only adjusting for measured confounding in CER. The PPP is a latent variable that cannot be measured directly using routinely collected data (1). Our results show that increasing the number of previous prescriptions used in constructing the PPP leads to power gains which could be particularly important for studies with small or moderate sample sizes. It is worth noting that using PPP with only one previous prescription is a popular strategy in the applied literature. According to our results, prior 2, prior 3, prior 4 and the proportion IV performs better than prior one since the IV strength increases as we account for longer prescription history. It is worth pointing out that by using a larger history in calculating PPP, this implicitly assumes that a physician's preference does not change over time. This can be empirically tested using study data.

Baiocchi and colleagues suggest that researchers should consider the necessity for using IV method (17). If the unmeasured confounder is not a major problem statistically, IV methods may not be necessary. We support this conclusion with our result. According to the Figures, it is quite noticeable that there a threshold where the per cent bias of conventional methods become larger than that of IV methods. If we conduct IV methods at those points, the IV estimates may not be reliable, especially when we use 2SLS (see the 2SLS vs. OLS figure when the γ^2 equals 0) to compare with the OLS.

The limitation of this study mainly rests on the simplicity of the design. By moderate sample size, we used approximately 2500 which is derived from a research study the authors led on investigating prescribing for alcohol dependence in Scotland. Also, we synthesised the sample sizes that in the current CERs papers that focus on the PPP IV and found that most of them are above 10,000. We did not consider survival analysis including censored outcomes (18) or non-linear two-stage approaches, like two-stage predictor substitution and two-stage residual inclusion, in the simulation design. Finally, we need to emphasise that an essential limitation of studying the time-based proxies for IVs is that the time cannot be truly simulated as all data generated at the same time. The prior 1/2/3/4 prescription proxy estimates are based on real time in applied studies. Strictly speaking, this simulation demonstrates valid proxies for PPP IV, rather than the "true" prior 1,2,3,4 prescriptions as proxies.

Conclusion

Using PPP as an IV for CER is less biased than conventional approaches and can achieve adequate statistical power in smaller sample sizes if the IV strength is high enough. If it can be assumed that a physician's preference does not change over time, we recommend constructing PPP using entire prescribing history to gain power.

Abbreviations

2SLS:two-stage least squares; OLS: ordinary least squares; IV: Instrumental variable; PPP: Physician's Prescribing Preference; CER: Comparative Effectiveness Research

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and materials: The R code that generated the simulated data and regression models can be found in the supplementary material.

Competing interests: No interest declared

Funding: Not applicable

Authors' contributions: ZL simulated the data, analysed the results and drafted the manuscript. LJ and MD are two supervisors of ZL. They revised the draft. And all authors approved the final manuscript.

Acknowledgements: Abstract of this work has been presented at the 13th Asian Conference on Pharmacoepidemiology (ACPE 2021). Record can be found on the websites.

References

1. Brookhart MA, 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).
2. Davies NM, Taylor AE, Taylor GM, Itani T, Jones T, Martin RM, et al. Varenicline versus nicotine replacement therapy for long-term smoking cessation: an observational study using the Clinical Practice Research Datalink. *Health Technology Assessment (Winchester, England)*. 2020;24(9):1.
3. Thomas KH, Martin RM, Davies NM, Metcalfe C, Windmeijer F, Gunnell D. Smoking cessation treatment and risk of depression, suicide, and self harm in the Clinical Practice Research Datalink: prospective cohort study. *Bmj*. 2013;347:f5704.
4. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *Journal of clinical epidemiology*. 2013;66(12):1386–96.
5. Davies NM, Taylor GM, Taylor AE, Jones T, Martin RM, Munafò MR, et al. The effects of prescribing varenicline on two-year health outcomes: an observational cohort study using electronic medical records. *Addiction*. 2018;113(6):1105–16.

6. Taylor GM, Taylor AE, Thomas KH, Jones T, Martin RM, Munafo MR, et al. The effectiveness of varenicline versus nicotine replacement therapy on long-term smoking cessation in primary care: a prospective cohort study of electronic medical records. *International journal of epidemiology*. 2017;46(6):1948–57.
7. Kollhorst B, Abrahamowicz M, Pigeot I. The proportion of all previous patients was a potential instrument for patients' actual prescriptions of nonsteroidal anti-inflammatory drugs. *J Clin Epidemiol*. 2016;69:96–106.
8. Boef AG, le Cessie S, Dekkers OM, Frey P, Kearney PM, Kerse N, et al. Physician's Prescribing Preference as an Instrumental Variable: Exploring Assumptions Using Survey Data. *Epidemiology*. 2016;27(2):276–83.
9. Kuo YF, Montie JE, Shahinian VB. Reducing bias in the assessment of treatment effectiveness: androgen deprivation therapy for prostate cancer. *Med Care*. 2012;50(5):374–80.
10. Nelson RE, Nebeker JR, Hayden C, Reimer L, Kone K, LaFleur J. Comparing adherence to two different HIV antiretroviral regimens: an instrumental variable analysis. *AIDS and Behavior*. 2013;17(1):160–7.
11. Boef AG, Dekkers OM, Vandenbroucke JP, le Cessie S. Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. *Journal of clinical epidemiology*. 2014;67(11):1258–64.
12. Zhang Z, Uddin MJ, Cheng J, Huang T. Instrumental variable analysis in the presence of unmeasured confounding. *Annals of translational medicine*. 2018;6(10).
13. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application—in 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance. *Journal of clinical epidemiology*. 2009;62(12):1233–41.
14. Ionescu-Iltu R, Delaney JA, Abrahamowicz M. Bias–variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiology and drug safety*. 2009;18(7):562–71.
15. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. *Epidemiology*. 2006:260–7.
16. Ionescu-Iltu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *Journal of clinical epidemiology*. 2012;65(2):155–62.
17. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Statistics in medicine*. 2014;33(13):2297–340.
18. Tchetgen EJT, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context. *Epidemiology (Cambridge, Mass)*. 2015;26(3):402.

Figures

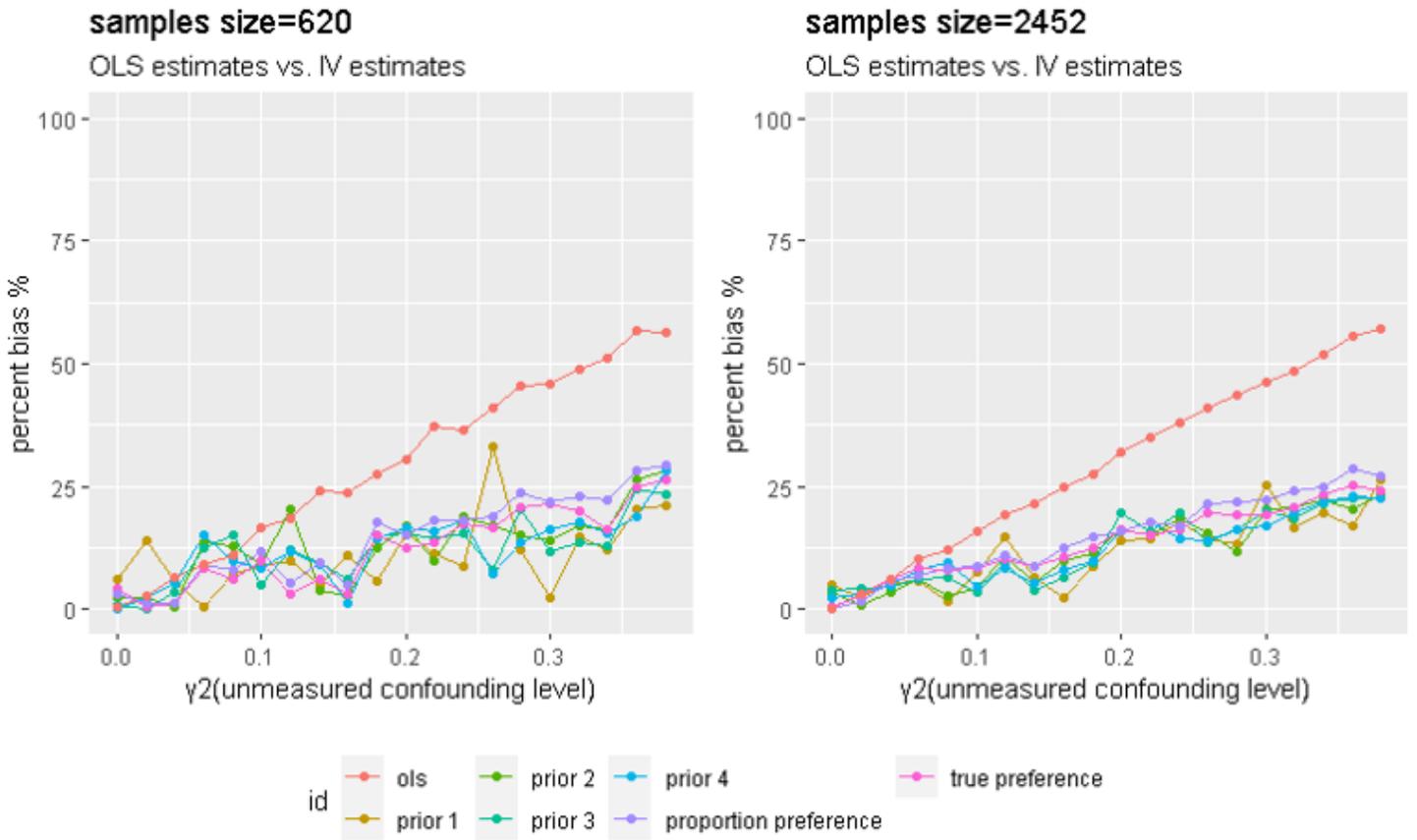


Figure 1

Percent bias of 2SLS and OLS

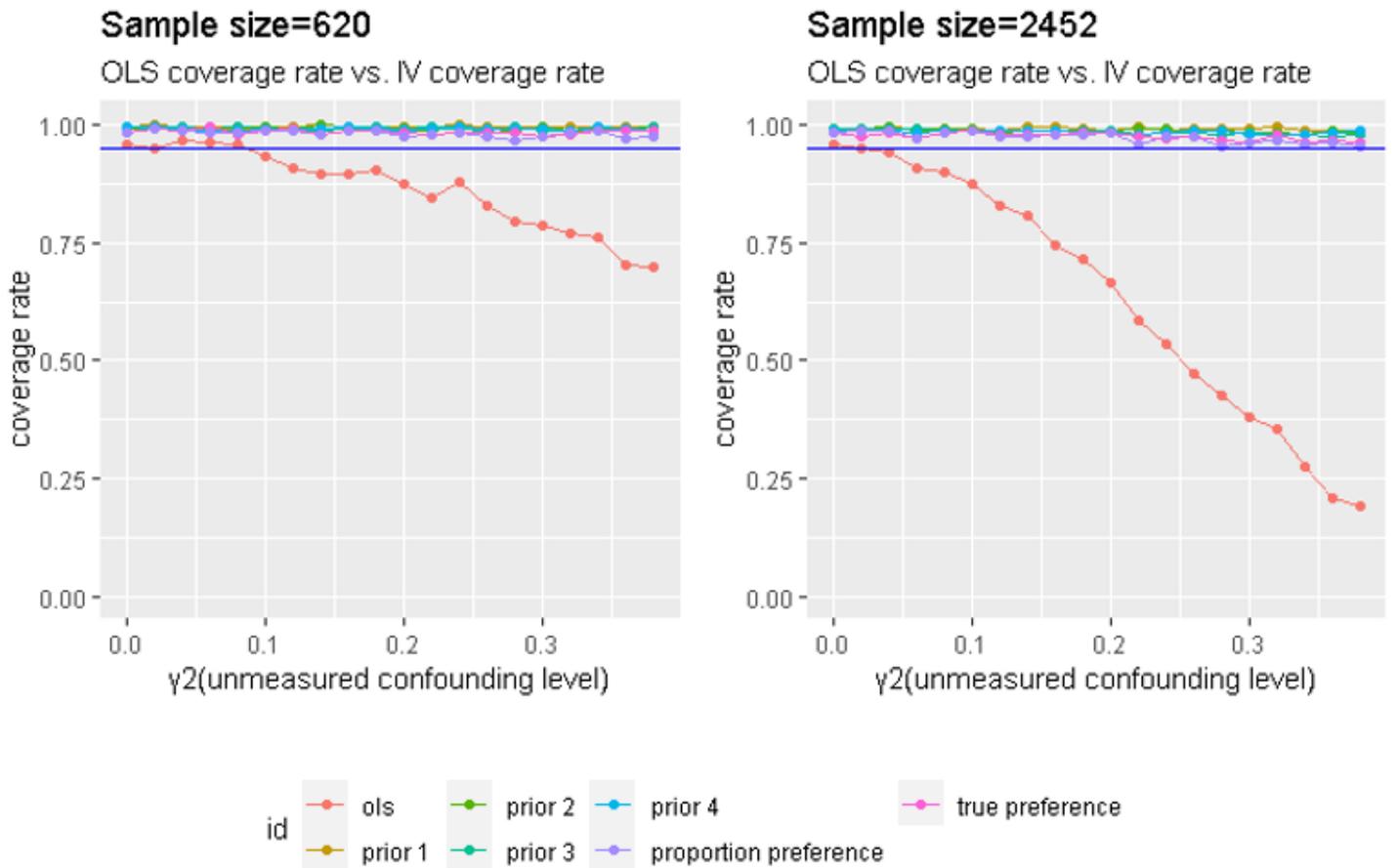


Figure 2

Coverage rate across 1000 times simulation. The blue intercept line represents the nominal 95%

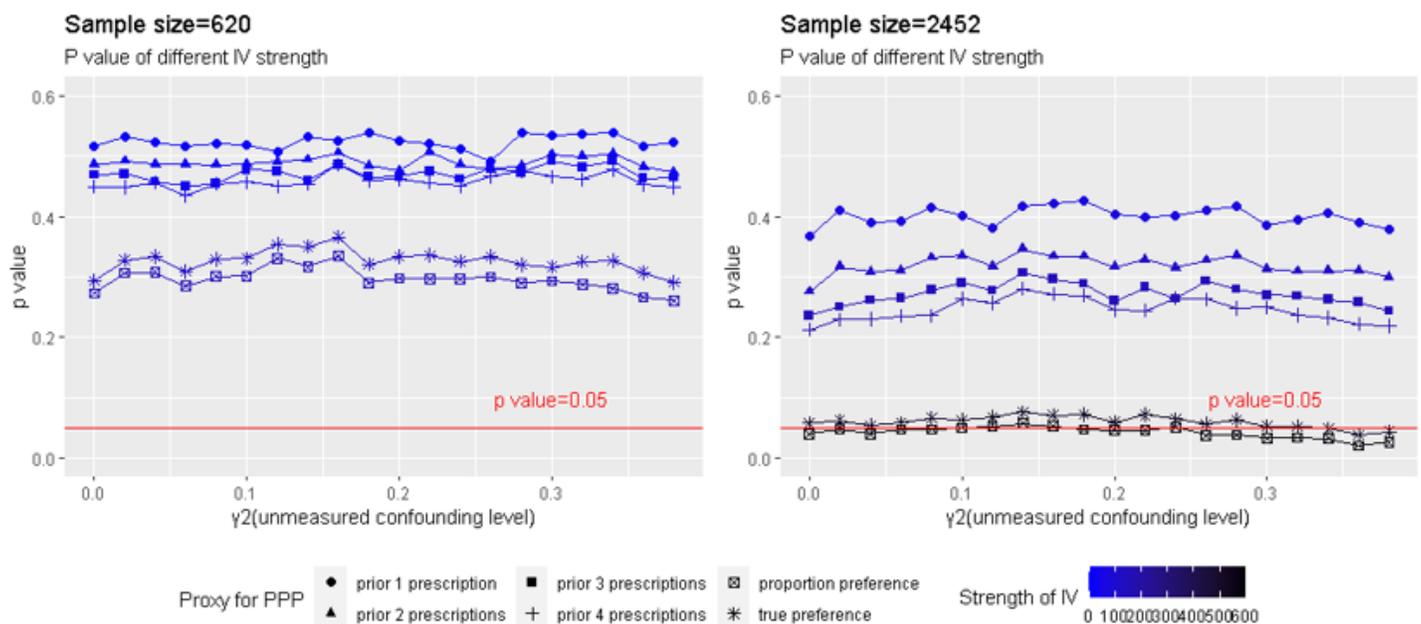


Figure 3

p-value of OLS and 2SLS estimates

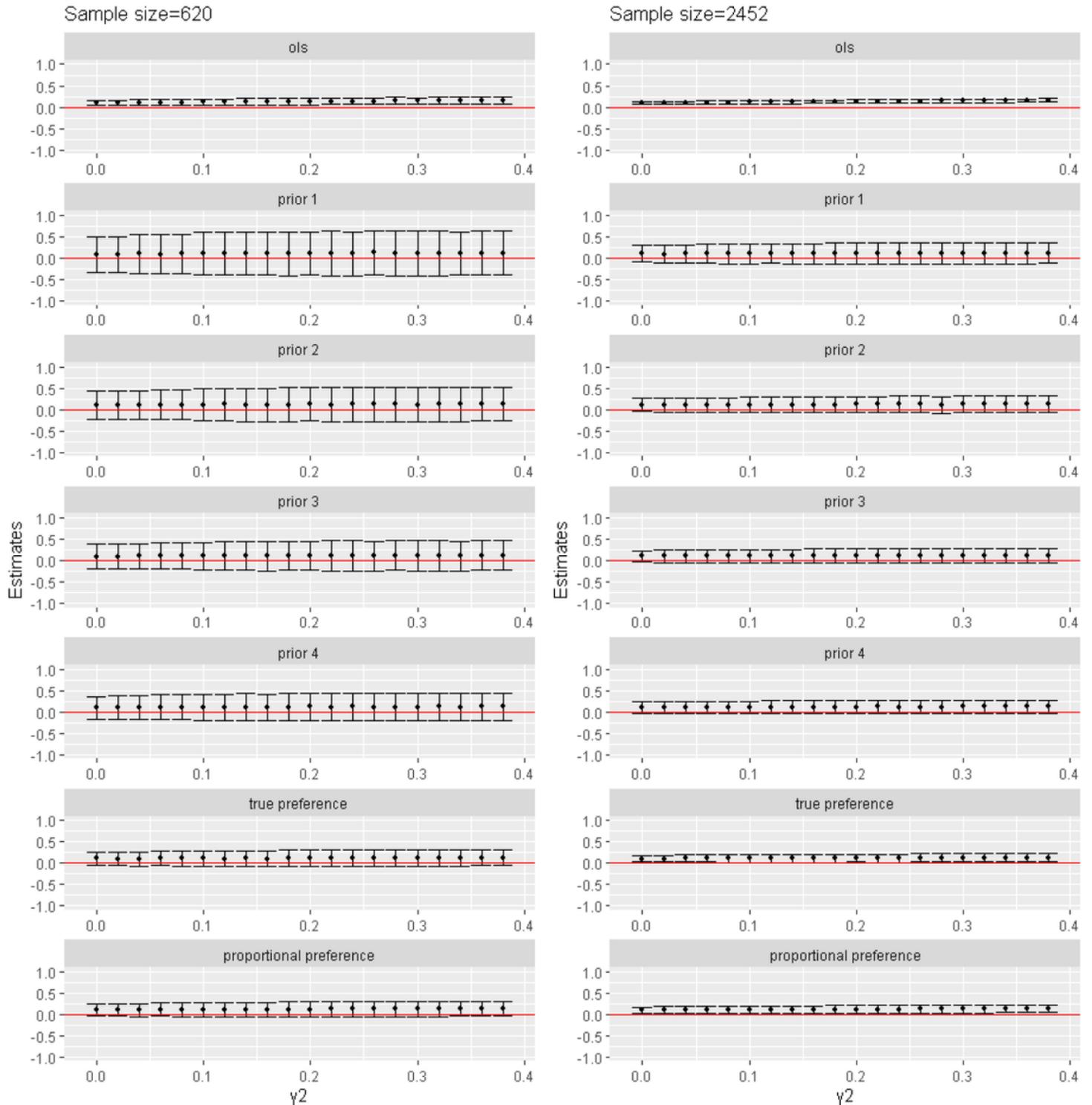


Figure 4

95% Confidence intervals of OLS and 2SLS estimates (the confidence intervals are calculated using cluster robust standard errors). The red line represents the null hypothesis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [appendix.docx](#)
- [simulationcode.txt](#)
- [simulationcoded1.txt](#)