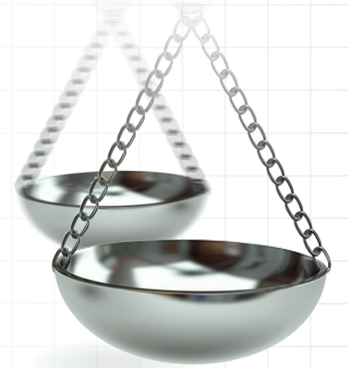
Research Article

For reprint orders, please contact: info@becaris.com



Journal of Comparative Effectiveness Research

Comparing the performance of two-stage residual inclusion methods when using physician's prescribing preference as an instrumental variable: unmeasured confounding and noncollapsibility

Lisong Zhang*,[1] & Jim Lewsey[2]

[1]Department of Population Health Sciences, University of Leicester, Leicester, LE1 7RH, UK

[2]School of Health and Well-Being, University of Glasgow, Glasgow, G12 8TB, UK

*Author for correspondence: lz236@leicester.ac.uk

Aim: The first objective is to compare the performance of two-stage residual inclusion (2SRI), two-stage least square (2SLS) with the multivariable generalized linear model (GLM) in terms of the reducing unmeasured confounding bias. The second objective is to demonstrate the ability of 2SRI and 2SPS in alleviating unmeasured confounding when noncollapsibility exists. **Materials & methods:** This study comprises a simulation study and an empirical example from a real-world UK population health dataset (Clinical Practice Research Datalink). The instrumental variable (IV) used is based on physicians' prescribing preferences (defined by prescribing history). **Results:** The percent bias of 2SRI in terms of treatment effect estimates to be lower than GLM and 2SPS and was less than 15% in most scenarios. Further, 2SRI was found to be robust to mild noncollapsibility with the percent bias less than 50%. As the level of unmeasured confounding increased, the ability to alleviate the noncollapsibility decreased. Strong IVs tended to be more robust to noncollapsibility than weak IVs. **Conclusion:** 2SRI tends to be less biased than GLM and 2SPS in terms of estimating treatment effect. It can be robust to noncollapsibility in the case of the mild unmeasured confounding effect.

First draft submitted: 26 May 2023; Accepted for publication: 18 March 2024; Published online: 3 April 2024

Keywords: instrumental variables • noncollapsibility • physician's prescribing preferences • two-stage prediction substitution • two-stage residual inclusion • unmeasured confounding

In order to address unmeasured confounding bias concerns in observational CERs, the IV approach is widely used. In this approach, the most commonly used estimation method is the two-stage least square (2SLS) which consists of two stage linear regression. The 2SLS estimator is normally consistent when the outcome measure is represented as a numerical variable [1]. However, if one requires to estimate the treatment effect using an odds ratio (OR) for a binary outcome, the method needs to be adapted to the nonlinear setting. One such approach is two-stage predictor substitution (2SPS). The first stage regression of 2SPS is treatment regressed upon the covariates; the second stage is the outcome regressed upon predicted results from the first stage together with covariates.

Another nonlinear method, two-stage residual inclusion (2SRI), has the same first stage regression as 2SPS, but use the residuals from the first stage as an additional covariate in the second stage. It was first introduced by Hausman [2] in order to test endogeneity in the linear context. Currently, there are simulation studies [3,4,5] as well as real-world studies that provide evidence for the 2SRI being generally less biased than 2SPS when estimating a treatment effect in the presence of unmeasured confounding. However, unlike the risk difference, odds ratio, is not collapsible which means that it cannot always be expressed as the weighted average of stratum-specific OR. This characteristic is also referred to noncollapsibility [4,5,6,7,8]. For example, if one adjusts for covariates that are not associated with both outcome and treatment in a logistic regression model (i.e., not a true confounder), the

Table 1. Measurement of performance.

Measurement	Calculation
Percent bias (in GLM, 2SRI, 2SPS)	$\frac{\textit{true odds ratio} - \textit{estimated adds ratio(from GLM, 2SRI and 2SPS)}}{\textit{true odds ratio}} \times 100\%$
Coverage rate	Iterations when 95% CI includes the true OR across 1000 simulations (%)
F-statistics of the first stage regression	$\text{F-statistics} = \frac{\textit{Sum of squares for Model/Degrees of Freedom For Model}}{\textit{Sum of Squares for Error/Degrees of Freedom for Error}}$ $= \frac{\textit{Mean of Squares for Model}}{\textit{Mean of Squares for Errors}}$

2SLS: Two-stage least square; 2SRI: Two-stage residual inclusion; CI: Confidence interval; GLM: Generalized linear model; OR: Odds ratio.

adjusted OR may differ from the unadjusted OR. Therefore, in such contexts the difference between adjusted and unadjusted logistic regression consists of is made of two parts: confounding effect and noncollapsibility effect.

In terms of using PPP as an IV and applying the 2SRI method, Koladjo *et al.* concluded that 2SRI is less biased than IV-based generalized method of moments (GMM) [5] in estimation of treatment effect. It is widely acknowledged that 2SPS is not superior to 2SRI in terms of dealing with endogeneity in health research [3,6]. However, there are also studies indicating that 2SRI produces biased estimates of average treatment effect (ATE) and local average treatment effect (LATE), compared with 2SLS [7]. In this study, we focused on the nonlinear settings. For the conventional approaches which do not account for the unmeasured confounding issue, we chose the generalized linear model (GLM) as it is a one of the most intuitive approaches in nonlinear settings. There are two objectives in this study: compare 2SRI and 2SPS with the generalized linear models (GLMs), which can only adjust for measured confounders, in a drug comparison simulation study using physician's prescribing preference as instrumental variable in the presence of unmeasured confounding bias; test the robustness of 2SRI to noncollapsibility, using simulated data and real-life data from a real-world UK population health dataset (CPRD).

Method

Data generating process

In Box 1 in the Appendix, we present the details of the data generating process. In order to construct an observational CER, we set the total number of physicians as 80. The patients per physicians is in range from 10 to 50. The simulated data consists of 2442 records (n = 2442). X_1 and X_2 are the measured confounders. 'un' is the unmeasured confounder. The R code for constructing the treatment (X) and the outcome (Y) is listed in the Appendix.

X_3 in research objective 2 is the variable that induces the noncollapsibility effect which is based on a scenario that the variable is associated with the outcome but not associated with the treatment [4]. X_3 is formed with a mean value of 10 and 1 as standard deviation to ensure an adequate noncollapsibility effect. We used the γ_2 (ranges from 0 to 1.9) to control the unmeasured confounding level and γ_3 (ranges from 0 to 0.95) to control the level of noncollapsibility effect. The PPP IV is formed by the prior n prescription of drug A and divided by n prescribed by the same physician. The strength of IV is tested using the F-statistics. All assumptions of a valid IV are assumed to be met in this simulated dataset.

Study design

According to recent studies, 2SRI is sensitive to the choice of residuals [7]. In this study, we selected Pearson residuals to be used in the second stage of regression after initial analysis using raw residuals (results from raw residuals are extremely biased and not shown in this thesis). Percent bias and coverage rate are used to measure the performance of the estimation methods (Table 1). In order to obtain more precise estimates, each simulation is run for 1000-times. All simulations and statistical analyses are conducted using R version 4.1.1.

Design of the empirical illustration

An empirical example is presented in this section to demonstrate the ability performance of 2SRI in dealing with the accounting for a noncollapsibility effect. It is a CER which compares the effectiveness of acamprosate and disulfiram in reducing the risk of alcohol use disorder (AUD) hospitalizations in England using data from CPRD. Unlike adjusted logistic regression, the inverse propensity score weighting (IPSW) method using stabilized weights is considered to estimate the marginal treatment effect (MTE) and is free from the impact from noncollapsibility.

Table 2. The measurement of confounding bias and noncollapsibility effect.

Forms of bias	Ways of presentation using estimate of treatment effect from univariable, multivariable logistic regression and IPSW
Confounding bias	$\beta' - \beta^{IPSW}$
Noncollapsibility effect	$\beta^{IPSW} - \beta$

IPSW: Inverse propensity score weighting.

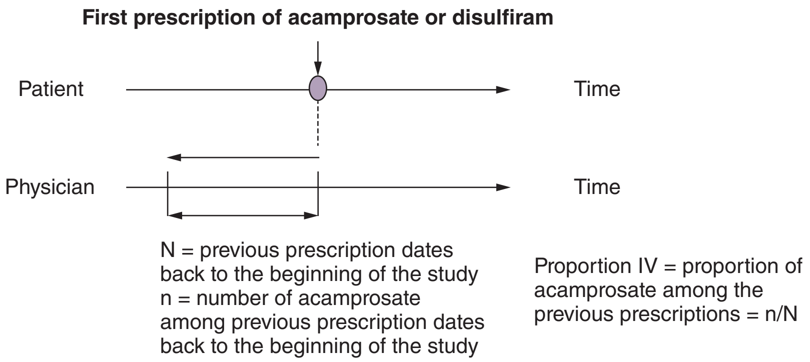

First prescription of acamprosate or disulfiram
Patient
Time
Physician
Time
N = previous prescription dates back to the beginning of the study
n = number of acamprosate among previous prescription dates back to the beginning of the study
Proportion IV = proportion of acamprosate among the previous prescriptions = n/N

Figure 1. Constructing instrumental variable in the empirical illustration.
IV: Instrumental variable.

Therefore, the noncollapsibility is usually quantified as the difference between IPSW-adjusted results and multivariable logistic regression, while the confounding bias is quantified using the difference between the univariable logistic regression and the IPSW-adjusted results [4,8] (Table 2).

The IV used in this case is the proportion of acamprosate among the last year prescriptions (see Figure 1 for the details of constructing the IV).

Results

Research objective 1: assessing the ability of 2SRI & 2SPS of alleviating the unmeasured confounding bias

Figure 2 shows that 2SPS is consistently more biased than 2SRI (γ_2 more than 0.75). When γ_2 is less than 0.75, the 2SRI estimates are not always less biased than 2SPS, but the percent bias is consistently at a low level (below 12.5%). The percent bias from 2SRI does not always inflate as the unmeasured confounding level rises; in the case of prior 1 and prior 2 as IV, we observed a rather low percent bias (less than 12.5%) for 2SRI throughout the range of γ_2 values from 0 to 1.9. The estimates from GLM are only at a low level when the unmeasured confounding level is small.

Despite the point estimate deviating from the 'true' OR, the coverage rates of 2SPS are around 95% most of cases. When the IV strength increases (F-statistics from 105 to 250), the coverage rates from 2SRI are more likely to achieve 95% (Figure 3).

Research objective 2: assessing the ability of 2SPS & 2SRI of alleviating noncollapsibility

This simulation tested whether the 2SRI or 2SPS estimates are able to reduce unmeasured confounding bias. However, the percent bias from the research objective 1 is free from the noncollapsibility effect. Research objective 2 is to assess the performance of 2SRI and 2SPS with the existence of unmeasured confounding as well as the noncollapsibility effect. We minimized the unmeasured confounding effect by selecting two scenarios where γ_2 equals 1.0 and 1.5 where 2SRI is generally unbiased (percent bias less than 10%) against the unmeasured confounding according to the results from the research objective 1. The results are shown in Figure 4.

It can be seen from Figure 4 that the 2SRI is generally less biased than 2SPS and GLM when the noncollapsibility effect is not severe. When the γ_2 equals 1.0, the percent bias of 2SRI is at a low level at the beginning but rise dramatically when the non-collapsibility effect increases. Same trend is found in GLM and 2SPS. When the γ_2 equals 1.5 (the unmeasured confounding effect at higher level), the percent bias of 2SRI fluctuated below 50% and hits a high level as γ_3 increases. The threshold of γ_3 where the percent bias (b) of 2SRI becomes extremely biased is

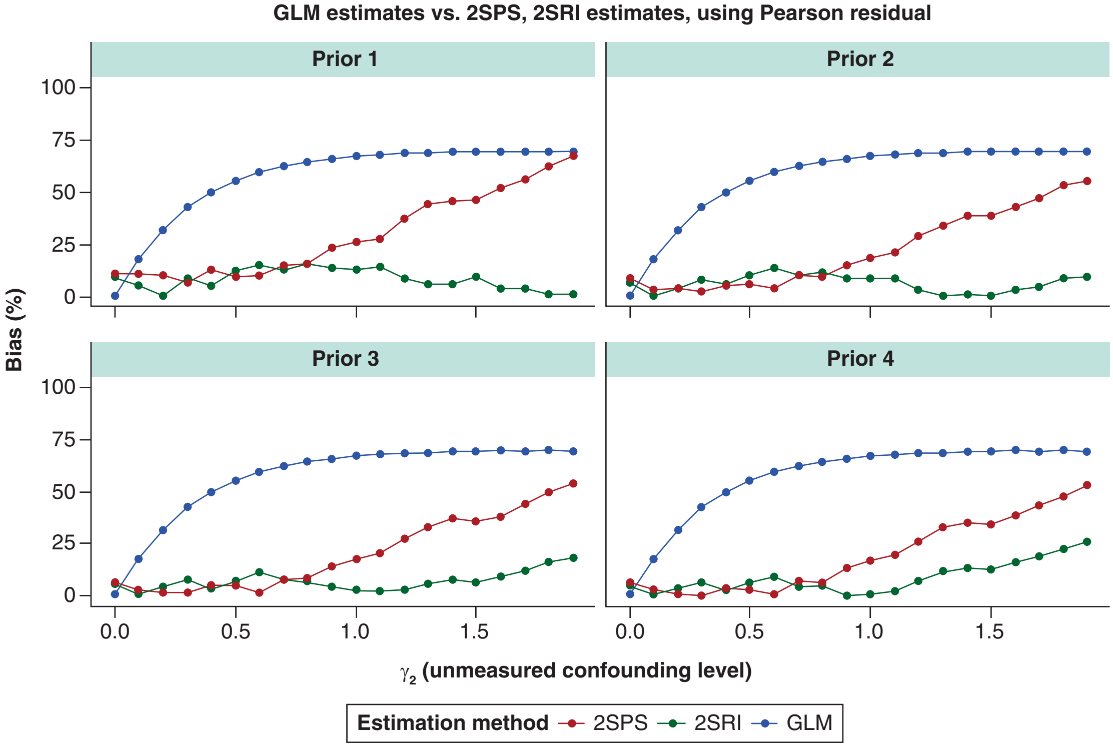


Figure 2. Estimates from generalized linear model, two-stage residual inclusion and two-stage least square.
GLM: Generalized linear model; 2SLS: Two-stage least square; 2SRI: Two-stage residual inclusion.

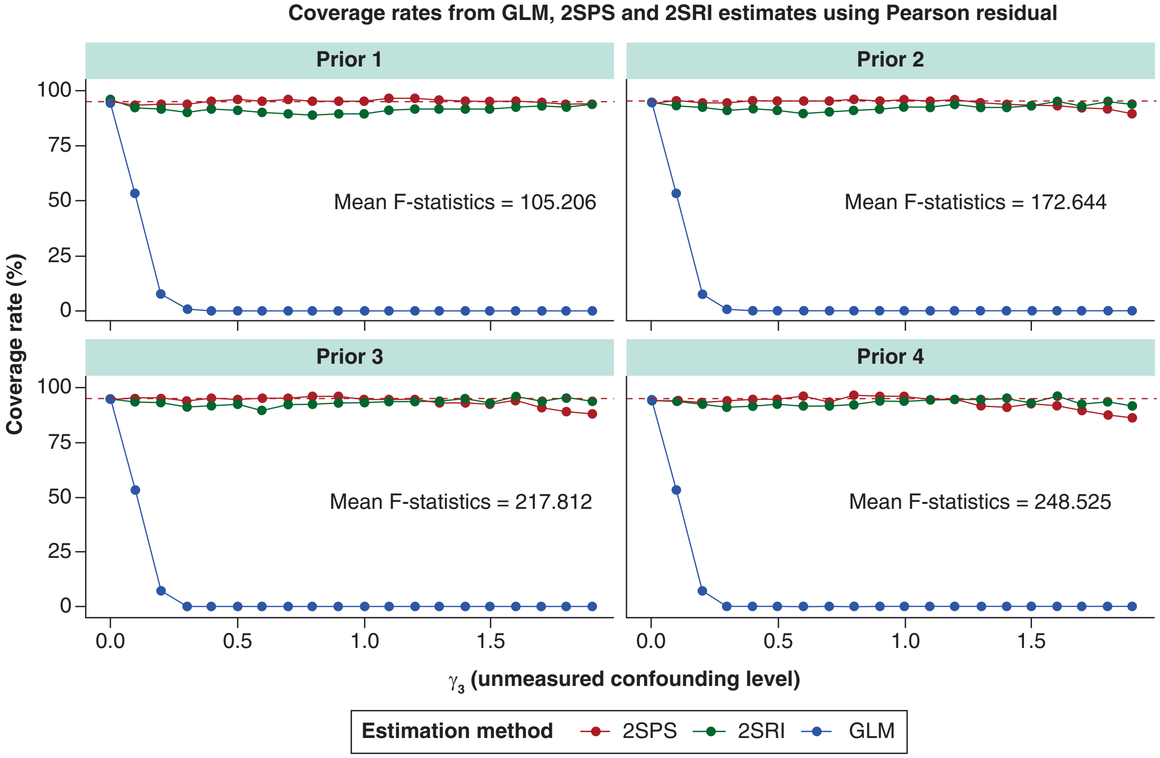


Figure 3. Coverage rates from generalized linear model, two-stage residual inclusion and two-stage least square.
Red dash line represents the 95% nominal.
GLM: Generalized linear model; 2SLS: Two-stage least square; 2SRI: Two-stage residual inclusion.

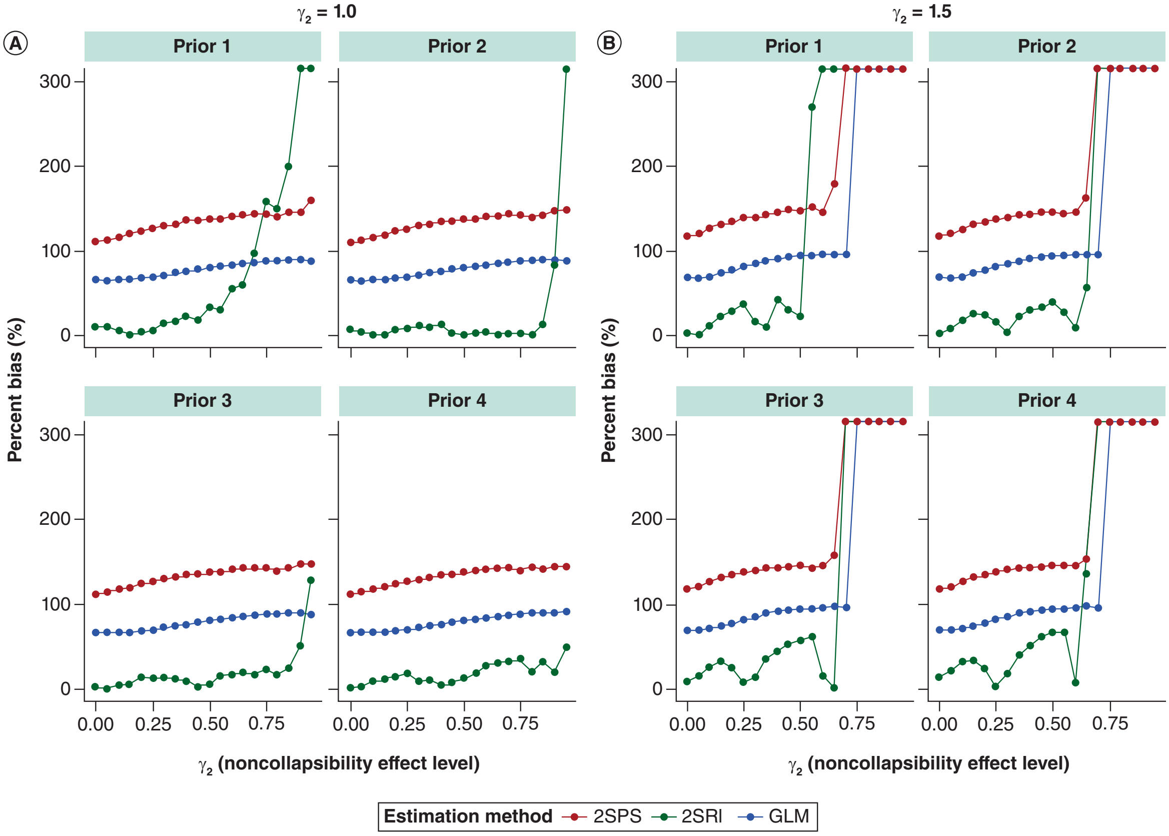


Figure 4. Percent bias of generalized linear model, two-stage residual inclusion and two-stage least square when there is noncollapsibility. (A) Percent bias represents percent bias of the estimate when γ_2 equals 1.0. **(B)** Percent bias represents percent bias of the estimate when γ_2 equals 1.5.
GLM: Generalized linear model; 2SLS: Two-stage least square; 2SRI: Two-stage residual inclusion.

smaller than percent bias (a) of 2SRI. The percent bias of 2SPS and GLM exceeds 100% when γ_3 is at low level in both scenarios indicating they are less robust to noncollapsibility. Note that for 2SPS and GLM, the percent bias is high even when γ_3 equals 0. This indicates that 2SPS and GLM are not robust to a random variable which is not a true confounder adjusted in the covariates. The coverage rates are presented in Figure 5. The coverage rates of 2SRI are around 95% when the γ_3 equals 1.0 and drop when the noncollapsibility effect increases. Note that, the coverage rate (b) increases after a certain point because the confident interval of 2SRI estimate is extremely wide, where the estimates of 2SRI become biased.

Empirical illustration using CPRD data

As it has been mentioned in the method section, the empirical illustration of research objective is based on a CER that compares the effectiveness of acamprosate and disulfiram in preventing AUD hospitalization. The confounders adjusted for in the models are Charlson Comorbidity Index (CCI), and 'prescription year'. CCI is associated with the outcome, but not associated with the treatment. 'Prescription year' is associated with both treatment and outcome (Figure 6). The presence of confounding bias and/or noncollapsibility is defined by adjusting for or not adjusting for these confounders.

In order to illustrate how 2SRI perform in different practical scenarios, we built three models. Model 1 is a logistic regression with or without adjusting for CCI. Model 2 is a logistic regression with or without adjusting for 'prescription year'. Model 3 is a logistic regression adjusts for 'prescription year', with or without adjusting for CCI.

The results of the case study using CPRD data are shown in Table 3. Since the CCI is associated with outcome but not associated with the treatment, the difference between the coefficient from β′ and β in Model 1 should be totally due to the noncollapsibility. In Model 2, adjusting for 'prescription year' bring in both noncollapisibility and confounding bias, where the noncollapsibility is less significant than the confounding effect. In Model 3, adjusting

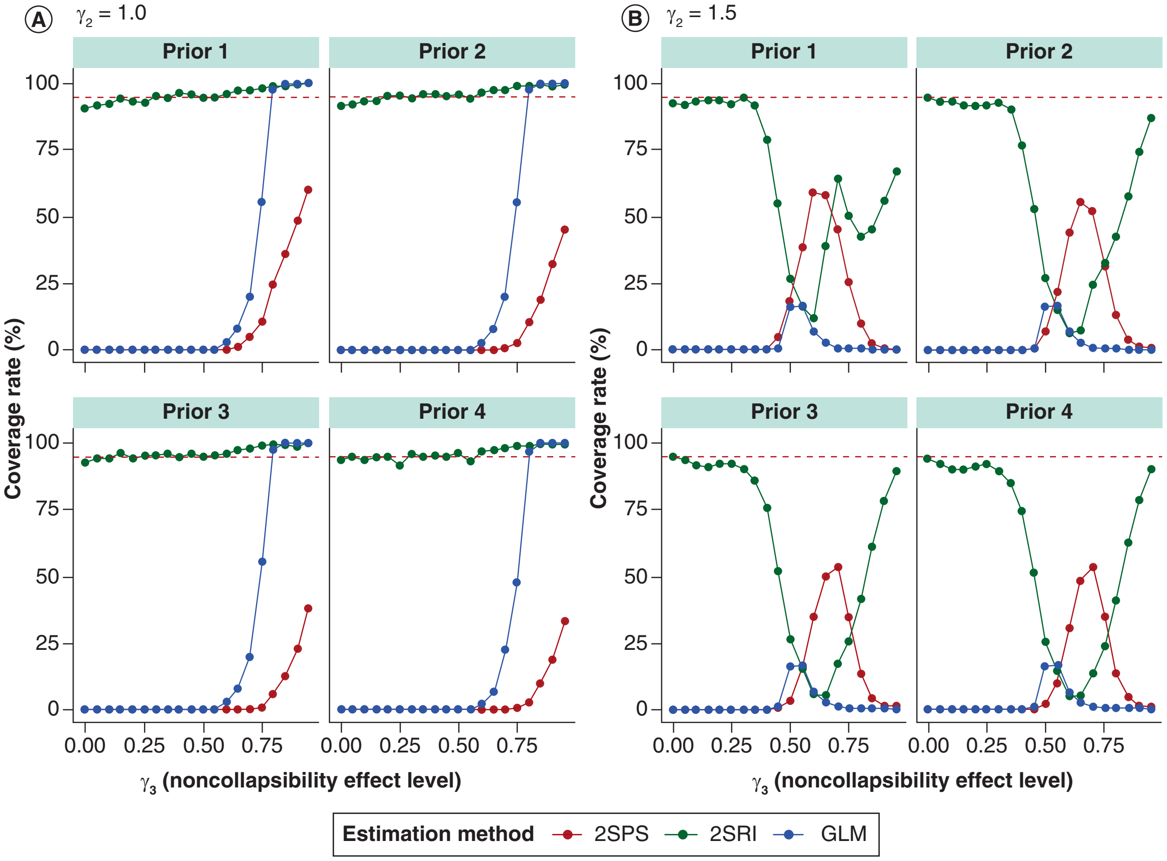


Figure 5. Coverage rate of generalized linear model, two-stage residual inclusion and two-stage least square. (A) Coverage raterepresents coverage rate of the estimate when γ_2 equals 1.0. **(B)** Coverage rate represents coverage rate of the estimate when γ_2 equals 1.5.
GLM: Generalized linear model; 2SLS: Two-stage least square; 2SRI: Two-stage residual inclusion.

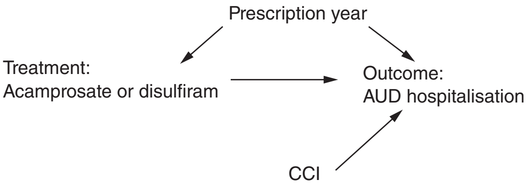


Figure 6. Assumed relations between the prescription year, Charlson Comorbidity Index and treatment and outcome.
AUD: Alcohol use disorder; CCI: Charlson Comorbidity Index.

for CCI brings more noncollapsibility compared with in Model 2. In Model 1 and Model 2, 2SRI does not show much ability to remove noncollapsibility from the adjusted models. However, results from the Model 3 echoes the simulation in the research objective 2 that the 2SRI can alleviate noncollapsibility when true confounders are adjusted in the model. Note the percent difference in Table 3 may be partly due to the residual unmeasured confounding in the observational studies.

Discussion

Results from the simulation study show that the percent bias of 2SRI is less than 15% in most scenarios while the percent bias of 2SPS reaches 50%. This echoes the findings from Cai *et al.* [6], who found that two-stage logistic regression 2SRI is asymptotically unbiased when the unmeasured confounding effect is not severe. However, our results are inconsistent with their conclusion that percent bias of 2SRI tends to rise as the unmeasured confounding level increases. Our findings indicate that the percent bias of 2SRI fluctuates when the unmeasured confounding is moderate but does not increase monotonically following the increase of an unmeasured confounding effect. According to the results, the IV strength does not affect the consistency of 2SRI estimates.

Table 3. Results from empirical case study.

Model		β'	β	β^{IPSW}	β^{2SRI}	Confounding bias = $\beta' - \beta^{IPSW}$	Noncollapsibility = $\beta^{IPSW} - \beta$	The estimate of treatment that excludes noncollapsibility: $\beta^* =$ $\beta - (\beta - \beta^{IPSW})$	Percent difference = $\frac{\lvert\beta^{2SRI} - \beta^*\rvert}{\beta^*}$ * 100%
1	Unadjusted logistic regression	0.168							
	Adjusting for CCI		0.211	0.184	0.152	0.168–0.184 = -0.016	0.184–0.211 = -0.027	0.238	33.9%
2	Unadjusted logistic regression	0.168							
	Adjusting for 'prescription year'		0.439	0.420	0.326	0.168–0.420 = -0.252	0.420–0.439 = -0.019	0.458	28.8%
3	Adjusting for 'prescription year' but not 'CCI'	0.439							
	Adjust for 'prescription year' and 'CCI'		0.452	0.379	0.608	0.439–0.379 = 0.06	0.379–0.452 = -0.073	0.525	13.6%

CCI: Charlson Comorbidity Index; IPSW: Inverse propensity score weighting.

We discussed the potential bias from the noncollapsibility on the 2SRI estimate using a simulated scenario where a covariate is associated with outcome but not associated with the treatment. The results indicate that percent bias of 2SRI has potential to be robust with minor or moderate noncollapsibility effect, and robustness is associated with the level of unmeasured confounding effect. We can see from Figure 4 that robustness is shown to be more resilient when the unmeasured confounding is smaller in magnitude (γ_2 equals 1.0). It is also reflected in the case study that 2SRI provides a close estimate to the unadjusted logistic regression when a variable that should not be adjusted appears in a model and brings noncollapsibility. However, for 2SPS and 2SRI, the noncollapsibility leads to major distortions on the estimates.

In this study, 2SRI is shown to be superior to 2SPS. There are studies that argue the consistency of 2SRI estimate is associated with the collapsibility of the model. Normally, the 2SRI estimator is consistent when the model is collapsible, for example, in the addictive hazards models [9]. However, our results show that the 2SRI estimate is not certainly biased with noncollapsibility effect. Despite our results supporting the preference of 2SRI, the nonlinear extension of 2SRI and 2SPS in the binary exposure are not studied adequately. Wan *et al.* proved the consistency of 2SRI estimates as the same time pointed out that the original framework proposed by Terza is used for the continuous treatment variable [10]. Further theoretical and methodological research is needed for 2SRI used for binary treatment.

Strengths & Limitations

To my knowledge, this study is the first simulation study to discuss the 2SRI's robustness to the noncollapsibility effect. In addition to the simulated data, we demonstrated an empirical example from CPRD to illustrate the noncollapsibility effect is common in logistic regression and can cause misleading results in causal inferential studies. One limitation of this study is the simplicity of the design. We only considered the simplest scenarios, without out taking account for more covariates, or more than one IV, or another forms of residuals that used in 2SRI. In the empirical illustration, we only consider three models in which the noncollapisbility can be considered mild. We did not consider the settings with more different levels of unmeasured confounding effect. Without the rigorous mathematical reasoning, these simplicity put a caveat on generality of results. As there is always possibility that the results turn to an opposite way when we apply the same approach to more scenarios.

Conclusion

The findings of this simulation study show that 2SRI performs unbiasedly in nonlinear models when conducting comparative effectiveness research. Further, the results show that 2SRI is more likely to alleviate noncollapsibility when unmeasured confounding effects are at lower levels.

10.57264/cer-2023-0085

Summary points

- Two-stage residual inclusion (2SRI) can reduce unmeasured confounding in the nonlinear settings.
- 2SRI tend to be less biased than two-stage least square (2SLS) and generalized linear model (GLM) when there is unmeasured confounder.
- The percent bias of 2SRI in terms of treatment effect estimates to be lower than GLM and 2SPS and was less than 15% in most scenarios.
- 2SRI is more robust to the noncollapsibility than 2SPS.
- 2SRI can alleviate moderate noncollaspibility.
- The percent bias of 2SRI in terms of treatment effect are around 20% when the unmeasured confounding is moderate.
- The percent bias of 2SRI in terms of treatment effect are less than 50% when the unmeasured confounding is more noticeable.
- Strong IVs tended to be more robust to noncollapsibility than weak IVs.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: https://bpl-prod.literatumonline.com/doi/10.57264/cer-2023-0085

Author contributions

L Zhang was responsible for the study conception and design, writing of the manuscript and data analysis. J Lewsey reviewed the manuscript and commented on it.

Financial disclosure

The authors have no financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with anyorganization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Writing disclosure

No writing assistance was utilized in the production of this manuscript.

Data sharing statement

This manuscript reports secondary analysis of Clinical Practice Research Datalink (CPRD). CPRD data were analysed under licence and are not available for public sharing. Access to CPRD is available to all qualified researchers via procedures detailed on the CPRD website (https://www.cprd.com/home/). Applications for data access are submitted to isac@cprd.com.

Open access

References

1. Palmer TM, Holmes MV, Keating BJ, Sheehan NA. Correcting the standard errors of 2-stage residual inclusion estimators for mendelian randomization studies. *Am. J. Epidemiol.* 186(9), 1104–1114 (2017).
2. Hausman JA. Specification tests in econometrics. *Econometrica* 46(6), 1251–1271 (1978).
3. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *J. Health Econ.* 27(3), 531–543 (2008).
4. Schuster NA, Twisk JWR, Ter Riet G, Heymans MW, Rijnhart JJM. Noncollapsibility and its role in quantifying confounding bias in logistic regression. *BMC Med. Res. Methodol.* 21(1), 136 (2021).
5. Koladjo BF, Escolano S, Tubert-Bitter P. Instrumental variable analysis in the context of dichotomous outcome and exposure with a numerical experiment in pharmacoepidemiology. *BMC Med. Res. Methodol.* 18(1), 61 (2018).

6. Cai B, Small DS, Have TR. Two-stage instrumental variable methods for estimating the causal odds ratio: analysis of bias. *Stat. Med.* 30(15), 1809–1824 (2011).

7. Basu A, Coe NB, Chapman CG. 2SLS versus 2SRI: appropriate methods for rare outcomes and/or rare exposures. *Health Econ.* 27(6), 937–955 (2018).

8. Pang M, Kaufman JS, Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat. Methods Med. Res.* 25(5), 1925–1937 (2016).

9. Wang A, Nianogo RA, Arah OA. G-computation of average treatment effects on the treated and the untreated. *BMC Med. Res. Methodol.* 17(1), 3 (2017).

10. Wan F, Small D, Mitra N. A general approach to evaluating the bias of 2-stage instrumental variable estimators. *Stat. Med.* 37(12), 1997–2015 (2018).