An Even Clearer Portrait of Bias in Observational Studies?

Neil M. Davies[a,b]

Instrumental variable analysis is an increasingly popular statistical method in epidemiologic research.[1] Epidemiologists' enthusiasm for this approach may be because it can potentially estimate causal effects in observational data in the presence of unmeasured confounding.[2] This overcomes a significant limitation of conventional epidemiologic methods, such as multivariable regression analysis: residual confounding. However, it is also possible for instrumental variable analyses to suffer from residual confounding. This can occur if the proposed instruments are associated with unmeasured confounding factors. Therefore, the key question for empirical researchers, regulators, and clinicians is: which is more biased—conventional multivariable adjusted regression or instrumental variable analysis?

In this issue of EPIDEMIOLOGY, Jackson and Swanson[3] elegantly describe a method for presenting and comparing the balance of potential confounders across values of the instrument and the actual treatment. This can allow researchers to assess the relative bias that could be caused by observed confounding factors. These methods may provide information about the relative bias of the unobserved confounders if they are correlated with the observed confounders. I will briefly discuss the methodologic improvements proposed by this article, its limitations, and finally a potential solution to these limitations.

A BRIEF DESCRIPTION OF THE MAIN RESULTS

The core of the paper is illustrated by a standard linear model:

$$Y(x) = \alpha_0 + \alpha_1 x + \alpha_2 U + \epsilon_x,$$

where Y, x, U, and z, respectively, represent a binary outcome, the treatment, a confounder of the outcome-treatment relationship, and the instrument. ϵ_x represents the error term, α_0 is a constant, α_1 is the effects of the treatment, and α_2 is the effects of the confounder on the outcome.

The simplest approach for assessing bias is to compare the difference in each confounder across values of the actual treatment, x, to the difference in the confounder across values of the instrument z. However, the same difference in a confounder across values of the instrument will result in much larger bias in the instrumental variable estimator than for the ordinary least squares (OLS) estimator.

To overcome this, Brookhart and Schneeweiss[4] proposed reporting the relative bias of the OLS and instrumental variable estimators of α_1 when the confounder U is omitted.

The OLS bias is

$$\text{Bias}(\hat{\alpha}_{1OLS}) = \alpha_2 (E[U \mid X = 1] - E[U \mid X = 0]).$$

The instrumental variable bias is

$$\text{Bias}(\hat{\alpha}_{1IV}) = \alpha_2 \frac{E[U \mid Z = 1] - E[U \mid Z = 0]}{E[X \mid Z = 1] - E[X \mid Z = 0]}.$$

These biases can be compared by computing the ratio of covariate imbalance across the actual treatment and instrument:

$$\frac{E[U \mid Z = 1] - E[U \mid Z = 0]}{E[U \mid X = 1] - E[U \mid X = 0]} < E[X \mid Z = 1] - E[X \mid Z = 0]$$

If the ratio on the left (known as the prevalence difference ratio) is larger than the term on the right (the strength of the instrument), then the instrumental variable bias is likely to be larger than the OLS bias.

Despite this approach's relative simplicity, it has been rarely used in the literature. Furthermore, Jackson and Swanson[3] note some limitations to prevalence difference ratios—they only provide information about relative bias and provide no information about the absolute bias. This means that if the instrumental variable and OLS biases are both very small a researcher could still find a very large prevalence difference ratio if the instrumental variable bias is small, but slightly larger than the OLS bias.

Jackson and Swanson[3] suggest presenting the bias components in graphical form. The effect of the confounder on the outcome, α_2, is the same for both the instrumental variable and OLS estimators, so any difference in bias between the two approaches must be due to the difference between the right hand terms in the bias equations. Therefore, it is sufficient to compare:

$$\frac{E[U \mid Z = 1] - E[U \mid Z = 0]}{E[X \mid Z = 1] - E[X \mid Z = 0]}$$

to

$$E[U \mid X = 1] - E[U \mid X = 0].$$

LIMITATIONS

The major limitation of this approach is that the authors do not propose any methods for estimating the standard errors of the instrumental variable bias and the OLS bias. As with instrumental variable estimates of the effects of the treatment, the scaled instrumental variable bias components will have much larger standard errors and confidence intervals than the OLS bias. This means that, solely by chance, across a set of potential confounding factors the scaled instrumental variable bias components are likely to be much larger for some

covariates. However, these differences may simply reflect statistical noise. To overcome this, for each covariate, we need to test the null hypothesis that there are no differences between instrumental variable and the OLS biases.

A SIMPLE SOLUTION TO THESE LIMITATIONS

A simple way to overcome this limitation is to note that the two bias terms are equivalent to the instrumental variable estimate (a Wald estimator) of the effects of the treatment x on the confounder, and the OLS estimate of the association of the treatment and the confounder. Recapitulating these results using standard instrumental variable estimation methods allows researchers to estimate the bias terms using existing packages, such as reg and ivreg2 in Stata.[5,6] This allows us to estimate the confidence intervals of the bias terms. These confidence intervals can be added to the covariates balance plots. Furthermore, within this framework, we can test the null hypothesis of no differences between the OLS and instrumental variables biases using Hausman tests.[7]

EMPIRICAL ILLUSTRATION

To illustrate the benefits of this approach, I reanalyzed the results of my paper investigating the relative effects of paroxetine versus other selective serotonin reuptake inhibitors (SSRIs) on self-harm and suicide.[8] The instrumental variable is the patient's physician's preferences for paroxetine or another SSRI. This is unmeasured, so we used the physicians' previously prescribed prescriptions as a proxy for their preferences. Brookhart et al.[9] argued that physicians' preferences for medications were plausible instruments because they are related to the medications they issue and may not be related to patient-level confounding factors. Please see the full paper for details of the sample and methods. I previously reported that the prevalence difference ratios for six of the 12 covariates suggested that the instrumental variable bias was larger than the OLS bias (Table 4 of the referenced paper). In Table, I report (1) the estimates of the OLS bias of the actual treatment (equal to one if the patient was prescribed paroxetine zero otherwise) and each of the 12 covariates, (2) the estimates of instrumental variables bias, and (3) Hausman tests of the difference between the estimated biases.

I found evidence that patients prescribed paroxetine were different to those prescribed other SSRIs for eight of the 12 covariates. The instrumental variable biases were much less precise, but there was weak evidence of differences for four of the 12 covariates by values of the instrument. The differences between the OLS and instrumental variable biases, as indicated by the Hausman tests, were substantial. For six of the 12 covariates these tests suggested that the instrumental variable bias was either smaller or in the opposite direction to the OLS bias. The importance of presenting confidence intervals can clearly be seen in the Figure. If only the point estimates were presented, we might erroneously conclude that the instrumental variable bias is larger for six of 12 covariates. However, we

TABLE. Ordinary Least Squares and Instrumental Variable Bias When Comparing Paroxetine and Other SSRIs (N = 359,736)[a]

| Confounder | Ordinary Least Squares Bias | | | | Instrumental Variable Bias | | | | | | | | | |
| | | | | | One Prior Prescription | | | | | Seven Prior Prescriptions | | | | |
	Risk Difference *100	Confidence Interval Lower	Upper	P	Risk Difference *100	Confidence Interval Lower	Upper	P	Hausman Text P	Risk Difference *100	Confidence Interval Lower	Upper	P	Hausman Test P
BMI > 25 kg/m² (N = 268,696)	-1.01	-1.72	-0.30	0.005	1.73	-0.74	4.20	0.17	0.02	1.02	-0.79	2.83	0.27	0.01
Hospitalized in prior year	0.04	-0.20	0.27	0.77	0.61	-0.35	1.56	0.21	0.17	0.99	0.08	1.89	0.03	0.03
More than 13 consultations in prior year	-2.36	-3.10	-1.62	<0.001	-0.99	-3.52	1.54	0.44	0.23	-1.61	-3.92	0.70	0.17	0.40
Older than 40 at first prescriptions	0.30	-0.45	1.05	0.44	-0.10	-2.50	2.29	0.93	0.72	-0.61	-2.71	1.49	0.57	0.40
More than five prescriptions in prior year	-2.11	-2.80	-1.42	<0.001	-0.31	-2.68	2.06	0.80	0.10	-1.39	-3.38	0.59	0.17	0.35
Male	4.20	3.46	4.93	<0.001	2.15	-0.27	4.56	0.08	0.06	2.90	0.77	5.03	0.008	0.13
Ever smoked	0.45	-0.22	1.11	0.19	1.33	-1.02	3.67	0.27	0.41	2.02	0.07	3.97	0.04	0.08
Prior diagnosis depression	-5.36	-6.57	-4.15	<0.001	-1.77	-5.62	2.07	0.37	0.02	-0.96	-4.98	3.06	0.64	0.02
Prior diagnosis definite self-harm	0.29	0.03	0.54	0.03	0.78	-0.13	1.69	0.09	0.26	0.70	0.03	1.37	0.04	0.14
Prior hypnotic prescription	0.16	-0.25	0.57	0.44	-0.99	-2.37	0.40	0.16	0.07	-1.46	-2.65	-0.26	0.02	0.001
Prior antipsychotic prescription	0.71	0.49	0.92	<0.001	-0.01	-0.69	0.67	0.98	0.03	0.06	-0.45	0.57	0.82	0.008
Charlson Index not zero	-1.37	-2.03	-0.71	<0.001	1.96	-0.31	4.23	0.09	0.001	1.03	-0.89	2.96	0.29	0.002

[a]Confidence intervals account for clustering between physicians. All estimates adjusted for year of prescription. Data taken from Davies et al. (2013). Restricted to patients whose physicians had issued at least seven previous SSRI prescriptions.
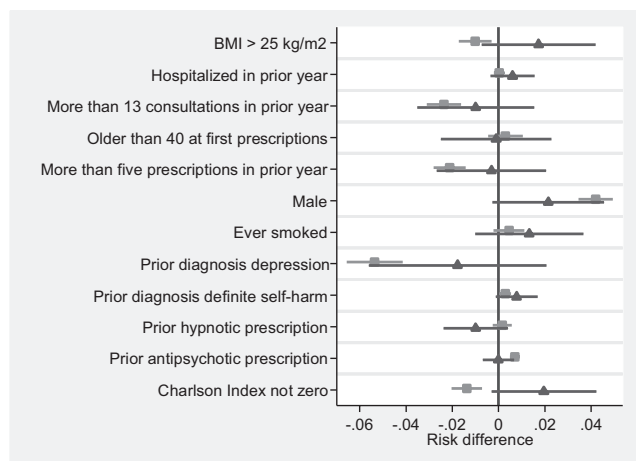
BMI indicates body mass index.

FIGURE. Covariate balance by levels of treatment (*squares*) and levels of the proposed instrument (*triangles*) using individual level data published by Davies et al.[7] (N = 359,736). Notes: Covariates binary variables, robust standard errors clustered by physician.

can only reject the Hausman test for two of differences (body mass index and Charlson index), and for these covariates the instrumental variable and OLS biases are opposite directions.

Two further advantages of using a standard instrumental variable framework to estimate the bias terms is that it is generalizable to multiple instrument settings, and we can test for differences in the biases between different sets of instruments using Hansen tests.[10]

CONCLUSIONS, ONGOING WORK, AND SUGGESTIONS FOR FUTURE RESEARCH

Jackson and Swanson[3] make an important contribution to the literature. When comparing OLS and instrumental variable biases, researchers should account for the instrument strength. However, these graphical specification tests are even clearer if researchers present confidence intervals and test hypotheses. These tests are simple to compute and should be reported by studies using instrumental variable analysis in pharmacoepidemiology and genetic epidemiology. These techniques are a welcome addition to a rapidly growing literature describing and critiquing the use of instrumental variable methods.[11–15] I fully expect that these and other developments will overcome some of the known limitations of instrumental variable methods. And perhaps more importantly, researchers

will precisely describe situations when instrumental variable analysis is likely to be more biased. Only if this is accomplished will epidemiologists be in a position to credibly advise regulators, clinicians, and patients about whether instrumental variable analysis or conventional multivariable regression gives the least biased indication of causal treatment effects.

ABOUT THE AUTHOR

NEIL DAVIES is a Research Associate at the MRC Integrative Epidemiology Unit at the University of Bristol. He has published a number of studies that have used instrumental variable methods in genetic epidemiology and pharmacoepidemiology.

REFERENCES

1. Davies NM, Davey Smith G, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology*. 2013;24:363–369.
2. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006;17:360–372.
3. Jackson JW, Swanson SA. Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology*. 2015;26:498–504.
4. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *Int J Biostat*. 2007;3:14.
5. Baum CF, Schaffer ME, Stillman S. ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML and k-class; 2010.
6. StataCorp. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP; 2013.
7. Hausman JA. Specification tests in econometrics. *Econometrica*. 1978;46:1251–1271.
8. 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. *J Clin Epidemiol*. 2013;66:1386–1396.
9. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Instrumental variable analysis of secondary pharmacoepidemiologic data. *Epidemiology*. 2006;17:373–374.
10. Hansen LP. Large sample properties of generalized method of moments estimators. *Econom J Econom Soc*. 1982;50:1029–1054.
11. Tchetgen Tchetgen EJ, Walter S, Vansteelandt S, Martinussen T, Glymour M. Instrumental variable estimation in a survival context. *Epidemiology*. 2015;26:402–410.
12. Swanson SA, Robins JM, Miller M, Hernán MA. Selecting on treatment: a pervasive form of bias in instrumental variable analyses. *Am J Epidemiol*. 2015;181:191–197.
13. Bulik-Sullivan B, Loh P-R, Finucane HK et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–295.
14. Franklin JM, Schneeweiss S, Huybrechts KF, Glynn RJ. Evaluating possible confounding by prescriber in comparative effectiveness research. *Epidemiology*. 2015;26:238–241.
15. Swanson SA, Miller M, Robins JM, Hernán MA. Definition and evaluation of the monotonicity condition for preference-based instruments. *Epidemiology*. 2015;26:414–420.