VIEWPOINT

Can Sophisticated Study Designs With Regression Analyses of Observational Data Provide Causal Inferences?

Tyler J. VanderWeele, PhD
Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts.

This Viewpoint presents considerations for assessing evidence for causal inference when using sophisticated study designs with regression analyses of longitudinal observational data. A view is sometimes expressed that regressions with observational data can never give causal conclusions. I argue this position is too extreme. While observational data rarely conclusively demonstrate causality, some study designs may provide evidence, and sometimes that evidence can be strong. However, the extent of evidence depends on a number of considerations. These considerations are narrower than those discussed decades ago by Hill,[1] which covered evidence from numerous sources, not just that from observational studies. I will begin with considerations concerning regression analysis using a single observational study and then return to broader considerations on the synthesis of evidence across studies.

Considerations for a Single Study

First, it is critical whether the observational data come from a cross-sectional or longitudinal study. Too often in psychological or psychiatric research, if a randomized trial cannot be used, then it is presumed all observational data are similarly inferior. However, among observational studies, the distinction between cross-sectional and longitudinal designs is critical. Typical cross-sectional studies cannot provide evidence for causal relationships. Cross-sectional associations reflect potential relationships in both directions; there is no way to separate them. For example, an association between antidepressants and depressive symptoms may reflect both how antidepressant use alters depressive symptoms, but also the fact that those with more severe symptoms may be more likely to use antidepressants. With cross-sectional data, unless exposure and confounder data are assessed retrospectively, we simply cannot distinguish these. With rigorous longitudinal study designs there is some hope by examining how each predicts the other over time.

However, a second important consideration is whether control has been made for the outcome at baseline. This may help rule out reverse association.[2] For example, if we use sophisticated and rigorous study methods that control for depressive symptoms at baseline, and other potential confounding factors, and examine how use of antidepressants may be associated with subsequent symptoms, we may help control for the reverse association that those with depressive symptoms may be more likely to use antidepressants.

Third, it is critical to control for other potential confounding factors at baseline, including social, demographic, genetic, economic, health, and related psychological variables. In doing so, we hope to minimize potential bias associated with these confounding factors, but there of course still always remains the possibility of bias from unknown confounders.

Fourth, if 3 waves of data are available and properly analyzed, it is also possible to control for prior levels of exposure in wave 1 and examine how wave 2 exposure is associated with wave 3 outcomes, controlling also for wave 1 confounding factors.[2] This can help further address reverse association and also may further minimize bias due to unmeasured confounding because the unmeasured confounders (eg, personality variables, like conscientiousness or neuroticism) would have to be statistically associated with wave 2 exposure beyond their association with wave 1 exposure to generate substantial bias.[2]

Fifth, high-quality measurements of the exposures, outcome, and covariates help address measurement-error biases; and minimizing attrition and missing data helps address potential selection bias. As discussed below, these potential biases can also be evaluated quantitatively.

Sixth, a large sample size with a precise estimate and narrow confidence interval may provide more evidence than when considerable statistical uncertainty remains.

Seventh, careful flexible statistical modeling of the associations between the outcome, exposure, and covariates; examining robustness to modeling decisions; and possibly using ensemble and/or doubly robust methods[2,3] may help ensure the observed associations are not simply artifacts of poor modeling.

Eighth, evidence can be strengthened further by examining the robustness of associations to biases, such as unmeasured confounding, measurement error, and selection using straightforward sensitivity analyses.[2,4] The E-value measure[4] reports how strong an unmeasured confounder would need to be associated with both the exposure and outcome to explain away the observed association, and sensitivity analysis for measurement error and selection bias can be similarly implemented in a straightforward manner.[2] Considerable robustness to potential biases can strengthen evidence substantially or, alternatively, make clear that caution is needed. Large effect size estimates will generally be more robust than modest effect sizes.[1,4]

A single observational study is unlikely to be definitive. Nevertheless, a large longitudinal study with good measurements; with control for baseline outcome, prior levels of exposure, and a rich set of covariates; with careful statistical modeling; and with sensitivity analyses suggesting robustness to potential biases may provide substantial evidence for causal inference. However, even then, effect sizes may be exaggerated.

Corresponding Author: Tyler J. VanderWeele, PhD, Department of Epidemiology, Harvard T. H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115 (tvanderw@hsph.harvard.edu).

Broader Considerations

The evidence from regressions using observational data will typically be less than that from large randomized trials with little loss to follow-up. But this is relative to the quality of the trial. A trial with considerable attrition may result in larger biases than those in a well-designed observational study. Nevertheless, randomized trials have the advantage of generally ruling out unmeasured confounding, and thus often provide stronger evidence.

Evidence for causal inference can certainly come from other study designs as well, including instrumental-variable designs and mendelian randomization analyses that use genetic variants as instruments, difference-in-difference methods and interrupted time-series designs, regression-discontinuity designs, and genetic co-relative designs. Ohlsson and Kendler[5] recently reviewed these designs. Each has advantages and disadvantages relative to regressions using observational data as to the assumptions made. A view is sometimes expressed that so-called natural experiments, as ostensibly used in mendelian randomization analyses, provide stronger forms of evidence for causal inference than regressions using observational data.[6] This is an oversimplification. Mendelian randomization analyses make numerous assumptions about the absence of direct effects of the variant on the outcome not through the exposure, and the absence of various selection biases.[7,8] These assumptions are different from those in regressions with observational data but just as susceptible to violations. As noted above, the quality of evidence from regressions using observational data may vary substantially. Similar variation in quality of evidence pertains to mendelian randomization and other so-called quasi-experimental designs.

Evidence accumulates over multiple studies. Regression analyses of different observational studies with consistent results, especially when combined meta-analytically with indications of robustness to potential biases, further strengthens evidence.[9] However, multiple studies with similar designs may be subject to similar biases. Thus, even better is the accumulation of evidence from different designs (each subject to different biases), an approach sometimes called *triangulation*.[5] As noted by Hill,[1] when the sources of evidence are convergent, this can be compelling; however, when there are differences, this does not necessarily constitute evidence against causation. Indeed, different designs and methods often estimate effect sizes for different subpopulations. Instrumental-variable or mendelian randomization analyses typically provide estimates for the subpopulation for whom the instrument changes the exposure rather than the total sample.[5-7] Sometimes, underlying true effect sizes for these 2 populations could even plausibly be in different directions. Evidence must thus be weighted carefully.

The editorial policy of the JAMA Network journals is that causal language should be used only for randomized trials. This is a conservative position, though one that beneficially keeps readers from overconfidence in clinical study results. However, in the weighing of evidence as a scientific community, other designs may contribute, which should be acknowledged.[10] Study interpretation should be nuanced and account for stringent assessments of the study quality, such as the considerations above for analyses of longitudinal observational data, or analogues for other designs to make determinations as to when language concerning evidence for causation is appropriate. As argued above, evidence from sophisticated regression analyses of longitudinal observational data can sometimes be strong and thereby assist us in drawing causal inferences.

ARTICLE INFORMATION

Published Online: September 9, 2020. doi:10.1001/jamapsychiatry.2020.2588

Conflict of Interest Disclosures: None reported.

Funding/Support: The writing of this Viewpoint was supported in part by grant CA222147 from the National Institutes of Health.

Role of the Funder/Sponsor: The National Institutes of Health had no role in the preparation, review, or approval of the manuscript or the decision to submit the manuscript for publication.

REFERENCES

1. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(5):295-300. doi:10.1177/003591576505800503

2. VanderWeele TJ, Mathur MB, Chen Y. Outcome-wide longitudinal designs for causal inference: a new template for empirical studies. *Statistical Science*. Accessed June 16, 2020. https://www.e-publications.org/ims/submission/STS/user/submissionFile/38456?confirm=800c6919

3. Schuler MS, Rose S. Targeted maximum likelihood estimation for causal inference in observational studies. *Am J Epidemiol*. 2017;185(1):65-73. doi:10.1093/aje/kww165

4. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-value to assess the potential effect of unmeasured confounding in observational studies. *JAMA*. 2019;321(6):602-603. doi:10.1001/jama.2018.21554

5. Ohlsson H, Kendler KS. Applying causal inference methods in psychiatric epidemiology: a review. *JAMA Psychiatry*. 2020;77(6):637-644. doi:10.1001/jamapsychiatry.2019.3758

6. Davies NM, Holmes MV, Davey Smith G. Reading mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018;362:k601. doi:10.1136/bmj.k601

7. VanderWeele TJ, Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology*. 2014;25(3):427-435. doi:10.1097/EDE.0000000000000081

8. Swanson SA. A practical guide to selection bias in instrumental variable analyses. *Epidemiology*. 2019;30(3):345-349. doi:10.1097/EDE.0000000000000973

9. Mathur M, VanderWeele TJ. Sensitivity analysis for unmeasured confounding in meta-analyses. *J Am Stat Assoc*. 2020;115(529):163-172. doi:10.1080/01621459.2018.1529598

10. Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am J Public Health*. 2018;108(5):616-619. doi:10.2105/AJPH.2018.304337