Instrumental variables analysis and mendelian randomization for causal inference

Moodie, E.E.M.; Cessie, S. le

The Journal of Infectious Diseases
VIEWPOINT

Instrumental Variables Analysis and Mendelian Randomization for Causal Inference

Erica E. M. Moodie[1] **and Saskia le Cessie**[2]

[1]Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada; and [2]Department of Clinical Epidemiology and Department of Biomedical Data Sciences, Leiden University, The Netherlands

Keywords. causal inference; instrumental variable; Mendelian randomization; unmeasured confounding.

Causal inference is a branch of statistics that attempts to quantify the impact of a treatment or intervention on an outcome, recognizing that, when experimentation is impossible, the treatment allocation may not be independent of other factors that predict the outcome. Rather, there may be common factors that affect both the treatment and the outcome; for instance, in a nonrandomized study, sicker individuals—who may be less likely to experience a full recovery—may receive a more aggressive treatment; untangling differences in recovery that are due to the treatment versus other factors such as the baseline level of health requires statistical adjustment. Frequently, such adjustment is direct—for example, via choosing pairs of individuals, each one having received one of 2 competing treatments, where the individuals are matched with respect to initial health status, or by a regression analysis where the health status measure is included as a covariate in the regression model. However, for such adjustment strategies to be successful in removing any confounding bias (ie, the bias that arises when a variable that predicts both the treatment and the outcome and thus distorts the treatment effect), *all* confounding variables must be recorded and available to the analyst. Unfortunately, this cannot always be ensured. In such circumstances, an alternative causal approach can be employed: an instrumental variables analysis, as was recently performed by Reilly et al in *The Journal of Infectious Diseases* [1].

As outlined by Goetghebeur et al [2] and previously highlighted by Moodie in *JID* [3], learning about causal relationships requires the explicit formalization of all definitions (eg, of the population of interest, the exposure, the outcome, the timeframe, and so on), the target causal effect, and the method of estimation with all the associated assumptions on data availability. Causal graphs, also called directed acyclic graphs, are useful tools for encoding beliefs on proposed relationships between variables and to center the analyses on a single treatment or exposure. In Figure 1, we display an assumed data-generating mechanism that lends itself to an instrumental variables analysis. This analysis relies on the existence of an instrument or instrumental variable that acts as a substitute for randomization to a treatment group, in a setting where individuals may not comply with the treatment assignment or randomization group. For example, different centers may have different treatment policies for certain diseases. If the different centers treat similar patients, patients are more or less randomly allocated to the different treatment policies, based on the center the patients happened to visit. The center can be used in this situation, under certain assumptions, as an instrumental variable. In general, an instrumental variable must satisfy 3 conditions. In particular, (*i*) it must be associated with the exposure, (*ii*) there are no common causes of the instrumental variable and the outcome, and (*iii*) the instrumental variable affects the outcome variable *only* through its effect on treatment but does not have a direct influence on the outcome. Condition (*ii*) states there is no confounding for the effect of the instrument on the outcome; (*iii*) is called the exclusion restriction. In Figure 1, condition (*ii*) is seen by the lack of any direct arrow from the instrument to the outcome, while condition (*iii*) is seen by the lack of any variables with a direct effect on *both* the outcome and the instrument. Note, however, that the treatment is confounded, due to the presence of a common cause of both treatment and the outcome.

Instruments may, in general, be difficult to identify; however, there are some special circumstances in which they arise. The first is in the setting of a randomized controlled trial where there is imperfect compliance. In such a setting, the randomly assigned treatment is the instrument, and the treatment actually taken is the exposure. Clearly, when the instrument is a randomization process, there is no confounding of the randomization and the decision to comply with assigned treatment (condition [*ii*]), nor is there an effect of the randomization on the

Published online 6 August 2024
Correspondence: Erica E. M. Moodie, PhD, Department of Epidemiology and Biostatistics, McGill University, 2001 McGill College Ave, Montreal, QC H3A 1G1, Canada (erica.moodie@mcgill.ca).

The Journal of Infectious Diseases® 2025;231:556–8

https://doi.org/10.1093/infdis/jiae357

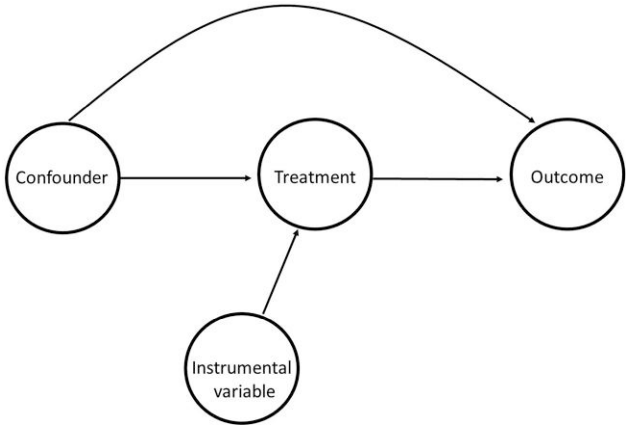

Figure 1. Example of a causal diagram showing a simple setting with a treatment or exposure whose effect on the outcome is *confounded* (eg, because of age, sex at birth, underlying health status, diet, socioeconomic status, and other variables that are predictive of both the exposure and the outcome. The *instrumental variable* is one or more variables that is associated with the exposure level but has no direct effect on the outcome. In these diagrams, the direction of an arrow is indicative of a causal relationship, where changing the level or value of the variable at the origin of the arrow will result in changes in the variable to which it points.

outcome (condition [*iii*]); furthermore, it is reasonable to suppose that most participants in a trial will comply with treatment, such that randomization causally affects the treatment taken (condition [*i*]). The second setting in which an instrument may be plausibly identified is one in which genetic variants predict the exposure but are unrelated to the outcome. In this case, the instrumental variables analysis is often referred to as a Mendelian randomization (MR) analysis; the genetic variant acts as a random allocation mechanism, with the randomness inherent to assortment of parental genes during meiosis. MR analyses are thus relevant to genetic variants that modify a particular exposure or risk factor of interest, such as obesity or ability to metabolize a particular compound.

The simplest instrumental variable analysis is a 2-stage regression approach, which can be used within a single dataset, combining statistics derived from the estimated association between the instrument and the outcome (model 1; note that this effect is due only to the path from the instrument to the outcome via the treatment of interest) and the association between the instrument and the exposure itself (model 2). Alternatively, a 2-sample MR analysis (see, eg, [4, 5]) can be employed in which the estimates in models 1 and 2 are estimated in different datasets. This could be advantageous if it is difficult to identify a single dataset that measures the instrumental variable, the exposure, and the outcome; it may be easier, for example, to study the association between the instrumental variable and the exposure in one dataset such as a large biobank or genetic dataset, and the association between the instrumental variable and the outcome in another dataset. As many such datasets are publicly available, this may facilitate instrumental variables analyses and rapid production of analytic results; however, the ready availability of such data may also encourage researchers to employ the method, potentially making bold causal claims, with insufficient knowledge of the subject matter under study.

Whether one-stage or 2-stage, the validity of any findings from an MR analysis will rely crucially on the 3 assumptions needed for instrumental variables analyses. In the MR context, the third assumption requires that there is no horizontal pleiotropy, meaning that the genetic variant(s) have no effect on disease except through the effect on the exposure. In MR studies, the association between genetic variants and exposures is often very weak. To increase power, often multiple genetic variants are used simultaneously in an MR analysis. New MR methods exist that do not require horizontal pleiotropy for *each* genetic variant. For example, the weighted median method only requires that at least 50% of the genetic variants are valid instruments.

Because of the relative ease with which researchers can now access large genetic datasets, instrumental variables analyses may become more common. While they are in practice easy to carry out, great care is still needed to ensure the analyses are validly conducted. Making explicit use of the Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) checklist [6] or an alternative tool such as the critical appraisal checklist of Davies et al [7] is an important but not sufficient element for documenting the analysis. Finally, as with any nonexperimental study, caution is warranted in interpreting the findings, which should be viewed as contributing to the body of evidence on a particular causal relationship rather than a definitive answer.

Notes

Financial support. E. E. M. M. acknowledges funding from a Discovery Grant from the Natural Sciences and Engineering Research Council of

Canada; is supported by a Chercheur de mérite career award from the Fonds de recherche du Québec, Santé; and holds a Canada Research Chair in Statistical Methods for Precision Medicine from the Canadian Institutes of Health Research.

Potential conflicts of interest. The authors: No reported conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Reilly CS, Borges ÁH, Bake JV, et al. Investigation of causal effects of protein biomarkers on cardiovascular disease in persons with HIV. J Infect Dis **2023**; 227:951–60.
2. Goetghebeur E, le Cessie S, De Stavola B, Moodie EEM, Waernbaum I. Tutorial: formulating causal questions and principled statistical answers. Stat Med **2020**; 39:4922–48.
3. Moodie EEM. Causal inference and confounding: a primer for interpreting and conducting infectious disease research. J Infect Dis **2023**; 228: 365–7.
4. Lawlor DA. Commentary: two-sample Mendelian randomization: opportunities and challenges. Int J Epidemiol **2016**; 45:908–15.
5. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Int J Epidemiol **2016**; 45:1717–26.
6. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization: the STROBE-MR statement. J Am Med Assoc **2021**; 326:1614–21.
7. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ **2018**; 362:k601.