JAMA Guide to Statistics and Methods

Instrumental Variables and Heterogeneous Treatment Effects

Matthew L. Maciejewski, PhD; Bryan E. Dowd, PhD; Edward C. Norton, PhD

A randomized clinical trial (RCT) can be used to estimate the average treatment effect for a population. Some patients experience a treatment effect that is larger than the average, while others experience a smaller-than-average treatment effect. Subgroup analyses often are used to evaluate heterogeneity in the treatment effect.[1] When it is infeasible or unethical to randomize patients to a treatment, the average treatment effect may be a combination of the true treatment effect and the effects of confounders—factors that influence both the treatment selected and patient outcomes.[2] When confounding factors are unknown or unobserved, correcting for their effect in statistical analyses is challenging. Instrumental variable analysis is one approach to address unobserved confounding.

Instrumental variable analysis is designed to reduce or eliminate unobserved confounding in observational studies and thus allow unbiased estimation of treatment effects. Results from an instrumental variable analysis typically apply to a subgroup of patients in the study. In a 2019 publication in *JAMA Internal Medicine*, Werner and colleagues[3] reported the results of an instrumental variable analysis that compared postacute care outcomes between Medicare beneficiaries discharged from the hospital to home with home health care or discharged to a skilled nursing facility. The authors also described how their results applied not to all patients but instead to a distinct subgroup of patients.

Use of the Instrumental Variable Method

Why Use an Instrumental Variable Analysis in the Setting of Heterogeneity of the Treatment Effect?

An instrumental variable analysis is conducted to reduce bias from unmeasured confounding in the estimation of the effect of a treatment or exposure from an observational study.[2] Instrumental variable analysis begins by identifying an observed explanatory variable that, like randomization, influences assignment to the treatment, but has no direct effect on the outcome of interest, referred to as an "instrumental variable." Unlike randomization, however, the instrumental variable may not act like randomization for all patients but only for a subset of patients who were effectively quasi-randomized to treatment or no treatment based on the value of the instrumental variable. In this analytic approach, those patients are referred to as the "marginal patients." The overall goal of an instrumental variable analysis is to measure the treatment effect free of bias. A major trade-off for reducing bias (increased internal validity) is a loss of generalizability, because results apply only to marginal patients and it cannot be known with certainty what subset of the overall cohort are the marginal patients (although methods exist to be able to describe them).

Description of Instrumental Variable Analysis in the Setting of Heterogeneity of the Treatment Effect

In an instrumental variable analysis, treated patients can be classified as either "compliers with the instrumental variable" or as "always-takers," meaning they would have taken the treatment regardless of the value of the instrumental variable. Untreated patients can also be classified as "compliers" or as "never-takers" of the treatment, meaning they would not have taken the treatment regardless of the value of the instrumental variable. The decisions of compliers, the marginal patients, are strongly influenced by the value of the instrumental variable, whereas the decisions of the always-takers and never-takers are not influenced at all by the value of the instrumental variable.

For example, in a hypothetical study in a city that has 2 hospitals that offer different emergency (nonelective) treatments, the hospital on the east side of town specializes in laparoscopic colectomy and the hospital on the west side specializes in open colectomy. Some patients always will choose laparoscopic colectomy, and other patients always will choose open colectomy; both will therefore travel across town if necessary to receive their preferred treatment. A third group of patients will choose the hospital, and thus the treatment, based on the distance from their home to the hospital. Those in the third group are the marginal patients with respect to distance, and the relative distances can serve as an instrumental variable under the assumption that relative distance to the 2 hospitals affects treatment assignment but has no direct effect on the outcome of interest. There is no observable variable in the data set that explicitly differentiates always-takers from marginal patients in the treatment group or that distinguishes never-takers from marginal patients in the untreated group, but the relative distances to facilities offering laparoscopic vs open colectomy are observed for all patients.

A common instrumental variable approach to estimate the effect of a binary treatment on a continuous outcome is to implement a 2-stage regression model. The first-stage equation estimates the probability of receiving the treatment, primarily as a function of the instrumental variable while also adjusting for other factors. The first stage is used to estimate a predicted probability of receipt of the treatment for each patient. In the second stage, a linear regression model of the outcome is estimated as a function of the predicted probability of receiving the treatment while again adjusting for other factors. The coefficient on the predicted probability of receiving the treatment is interpreted as the local average treatment effect because it applies only "locally" to the subgroup of marginal patients. Instrumental variable analysis with binary outcomes requires a different estimation approach.[4]

If the treatment response is *exactly* the same for all patients (and if patients in a nonrandomized study mirror those in an RCT), then the results from an instrumental variable analysis will match those from the RCT. However, if the treatment response varies from one patient to another, referred to as heterogeneous treatment effect,[1,5] then the results may not match. The instrumental variable analysis measures the treatment effect only for the marginal patients whose treatment choice was directly affected by the instrumental variable, so the local average treatment effect may differ from the average treatment effect if there is treatment effect heterogeneity.

What Are the Limitations of Instrumental Variable Analysis in the Setting of Heterogeneity of Treatment Effect?

There are several limitations with instrumental variable analysis. First, the treatment effect rarely is homogeneous and often differs across patients, perhaps by age or comorbidity. When treatment effects are heterogeneous, then the local average treatment effect might not equal the average treatment effect. Instrumental variable analysis can miss subsets of patients for whom the treatment is effective or may find significant effects that apply only to a small subset of patients. That limitation can be addressed by attempting to identify the marginal patients, as Werner and colleagues did.[3]

Second, different instrumental variables potentially identify different subgroups of marginal patients. As a result, analyses using 2 different instrumental variables could yield 2 different results, both correct but generalizing to different (but potentially overlapping) subgroups of marginal patients.[6]

How Did Werner et al Use Instrumental Variable Analysis and Address Heterogeneity of the Treatment Effect?

In their instrumental variable analysis of postacute care outcomes between Medicare beneficiaries discharged from the hospital to home with home health care or discharged to a skilled nursing facility, Werner et al[3] used differential distance—from the patient's home to the nearest home health care location vs the nearest skilled nursing facility—as the instrumental variable. Differential distance strongly predicted the discharge setting for patients whose better health or ability to recover with home support (instead of skilled nursing facility) made them candidates for either setting.[7] The authors also suggested that differential distance should not affect subsequent health outcomes directly but only through the choice of setting. Hence, the marginal patients in this analysis were those whose choice of treatment setting was influenced by differential distance.

The authors characterized the marginal patients using a method suggested by Baiocchi et al[8] in which the effect of the instrumental variable on treatment assignment for subsets of the population, defined by observed covariates, is compared with the average effect of the instrumental variable on treatment assignment for the entire analytic sample. Marginal patients are those who have larger-than-average effect of the instrumental variable on treatment assignment, ie, the compliers.

How Should the Results of the Instrumental Variable Analysis in Werner et al Be Interpreted?

Werner et al[3] found that marginal patients discharged to home with home health care had higher rates of hospital readmission but lower Medicare costs and had similar mortality and functional outcomes than patients discharged to skilled nursing facilities. They also found that marginal patients were younger and more likely to be Black or Hispanic, dually enrolled in Medicare and Medicaid, and to be enrolled in a Medicare Advantage plan. The results from this instrumental variable analysis generalize only to those marginal patients whose site of postdischarge care was likely influenced by distance to each type of care. This result minimizes bias from unobserved confounding for the marginal patients, but in doing so may yield a result that differs from the average treatment effect for all patients that would be estimated from an RCT. The greater internal validity in instrumental variable analyses has a trade-off in external validity.

ARTICLE INFORMATION

Author Affiliations: Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Medical Center, Durham, North Carolina (Maciejewski); Department of Population Health Sciences, Duke University School of Medicine, Durham, North Carolina (Maciejewski); Division of General Internal Medicine, Department of Medicine, Duke University School of Medicine, Durham, North Carolina (Maciejewski); Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis (Dowd); Department of Health Management and Policy, University of Michigan, Ann Arbor (Norton); Department of Economics, University of Michigan, Ann Arbor (Norton); National Bureau of Economic Research, Cambridge, Massachusetts (Norton).

Corresponding Author: Matthew L. Maciejewski, PhD, Department of Population Health Sciences, Duke University Medical Center, 508 Fulton St, Ste 600, Durham, NC 27705 (matthew.maciejewski@va.gov).

Section Editors: Robert Golub, MD, Deputy Editor; and Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA.

Conflict of Interest Disclosures: Dr Maciejewski reported receiving grants from US Department of Veterans Affairs Health Services Research & Development (RCS 10-391) and that he owns Amgen stock through his spouse's employment. No other disclosures were reported.

REFERENCES

1. Angus DC, Chang CH. Heterogeneity of treatment effect: estimating how the effects of interventions vary across individuals. *JAMA*. 2021;326(22):2312-2313. doi:10.1001/jama.2021.20552

2. Maciejewski ML, Brookhart MA. Using instrumental variables to address bias from unobserved confounders. *JAMA*. 2019;321(21):2124-2125. doi:10.1001/jama.2019.5646

3. Werner RM, Coe NB, Qi M, Konetzka RT. Patient outcomes after hospital discharge to home with home health care vs to a skilled nursing facility. *JAMA Intern Med*. 2019;179(5):617-623. doi:10.1001/jamainternmed.2018.7998

4. Bhattacharya J, Goldman D, McCaffrey D. Estimating probit models with self-selected treatments. *Stat Med*. 2006;25(3):389-413. doi:10.1002/sim.2226

5. Kent DM, Shah ND. Risk models and patient-centered evidence: should physicians expect one right answer? *JAMA*. 2012;307(15):1585-1586. doi:10.1001/jama.2012.469

6. Mogstad M, Torgovitsky A, Walters C. The causal interpretation of two-stage least squares with multiple instrumental variables. *Am Econ Rev*. 2021;111(11):3663-3698. doi:10.1257/aer.20190221

7. Lee DS, Moreira MJ, McCrary J, Porter J. Valid t-ratio inference for IV. *arXiv*. Preprint posted online October 10, 2020. https://arxiv.org/abs/2010.05058

8. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med*. 2014;33(13):2297-2340. doi:10.1002/sim.6128