Editorial

The Reliability of Instrumental Variables in Health Care Effectiveness Research: Less Is More

Observational studies of medical treatment effectiveness have increased substantially during the last several decades (Garber 2011) in part due to the growing realization that randomized controlled trials, the presumed gold standard of such research, are often not generalizable to the real world (Koppel 2013). Moreover, the economic stimulus of 2009 spurred the application of new medical effectiveness research methods using observational data, such as instrumental variable (IV) analysis (Garabedian et al. 2014).

In this issue of the journal, Sanwald and Schober (2016) analyzed the effects on survival of access to catheterization (cath) laboratories and invasive treatment of heart attack using an IV, distance to the hospital. We discuss the strengths and limitations of this study and demonstrate why this method more often produces untrustworthy estimates of the effects of medical treatments. IV analyses are statistical analyses, not research designs as articulated by their users. Many, weak research designs do not protect against bias even with heroic statistical adjustment to control for differences between the groups being studied (Soumerai, Starr, and Majumdar 2015). Unfortunately, most, but not all, IV studies use the weakest observational designs which do not demonstrate cause and effect. In the wise words of Light, Singer, and Willet (1990): “You can’t fix by analysis what you bungled by design.”

WHAT IS AN INSTRUMENTAL VARIABLE?

An instrumental variable (IV) is a variable, generally found in administrative data, that is assumed to randomize a treatment to estimate cause and effect

relationships, thus controlling for known and unknown patient characteristics affecting health outcomes. An important assumption is that the IV randomizes treatment but does not directly affect the patient outcome. In the study by Sanwald and Schober, whether a patient lived close or far from a hospital with a cath laboratory is the IV because it results in different levels of invasive treatment (McClellan, McNeil, and Newhouse 1994; McClellan 1996; Brooks, McClellan, and Wong 2000; Glickman and Normand 2000; Beck et al. 2003; Cutler 2007). In other words, the central and (we argue, dubious) assumption is that isolated people having a heart attack who live hours away from a hospital with a cath laboratory—and are therefore less likely to receive invasive procedures—are identical in their likelihood to survive as those lucky patients living very close to a cath hospital. Thus, we suggest that IVs, like distance to the hospital, do not randomize treatment. Instead, they likely further bias treatment effects.

Like any cross-sectional analysis, IV analysis relies on the absence of any unmeasured patient and health system confounders (e.g., socioeconomic status, health status, and other lifesaving treatments, such as medications) that may provide an alternative explanation for the relationship between the IV and the patients' survival. This assumption is the Achilles heel of IV studies. Most administrative data lack important variables correlated with survival (e.g., urban/rural status, or receipt of other lifesaving treatments), or they measure them poorly (e.g., race), representing a violation of the IV assumptions. Yet far too few IV studies investigate prior research to rule out such potential biases (Garabedian et al. 2014).

THE EFFECT OF CATH LABORATORIES ON SURVIVAL IN MYOCARDIAL INFARCTION USING THE IV METHOD

Sanwald and Schober used an IV approach and administrative data on 4,920 Austrian acute myocardial infarction (MI) patients to assess the effect of access (distance) to a cath laboratory on mortality and costs (Sanwald and Schober 2016). This IV study is stronger than most. For example, other RCTs and observational studies have shown some efficacy of invasive treatment on

Address correspondence to Stephen B. Soumerai, Sc.D., Harvard Medical School Department of Population Medicine and Harvard Pilgrim Health Care Institute, Landmark Center, 401 Park Drive, Suite 401 East, Boston, MA 02215; e-mail: ssoumerai@hms.harvard.edu. Ross Koppel, Ph.D., University of Pennsylvania, Sociology Department, Philadelphia, PA.

survival of acute MI patients. The authors forthrightly outline many limitations of their research. For example, they attempt to account for many confounders, such as socioeconomic status and several comorbidities. In addition, the authors conduct numerous types of sensitivity analyses to test the robustness of their findings (e.g., the effects are similar in smaller groups of urban and rural patients). They emphasize that any of hundreds of unmeasured, simultaneous, and effective treatments in a hospital with a cath laboratory might be responsible for reducing mortality, not the cath laboratory (and invasive procedures) by themselves. However, if policy makers and hospital administrators do not have an explanation for improved survival and cannot pinpoint the responsible factors affecting mortality, how can they hope to improve it?

More important, while the authors recognize this quandary, they fail to moderate their enthusiasm for causal effects of invasive heart attack treatments on survival—effects that may well be unsupported by their results. For example, in their abstract, they state: “place of residence affects the access of patients to invasive heart attack treatment and therefore their chance of survival.” Of even greater concern, the introduction states that such a cross-sectional analysis can “estimate the causal effects of an initial admission to a hospital equipped with a cath laboratory on mortality and follow up costs.” And the discussion states: “we conclude that providing more heart attack patients immediate treatment at PCI hospitals should be beneficial.” In contrast, we emphasize that the long time from the heart attack to 3-year mortality makes it difficult to exclude the possibility that other illnesses or treatments might explain differences in long-term mortality. It is indeed likely that invasive procedures save lives, but these correlational data cannot prove it. As in other IV studies, it would be prudent to include specialists (e.g., cardiologists) as coinvestigators because of their knowledge, especially for identifying possible confounders in the care of heart attack patients.

A SYSTEMATIC REVIEW OF THE EVIDENCE ON THE VALIDITY OF COMMON IVS

The weakness of IV studies is not a theoretical concern. In one of the only large systematic reviews of potential bias in research comparing the effectiveness of medical treatments using IVs, Garabedian et al. (2014) identified 65 studies using one of the four most common IVs (including distance, Sanwald and Schober’s IV). The authors identified major confounders likely to affect

survival in all of the 65 studies. The degree of bias is difficult to calculate, but in some cases, the confounders may have completely reversed the direction of reported effects, including studies of cath laboratories and invasive procedures among heart attack patients. Thus, although certain observational research methods that use longitudinal data and control groups are trustworthy, the majority of IVs are not a reliable way to control for bias in medical effectiveness research.

Although most IVs are cross-sectional and therefore do not meet the minimum standards for inclusion in Cochrane systematic evidence reviews (Effective Practice Organisation of Care [EPOC] 2015), several IVs do exist that reduce the likelihood of confounding and bias. A number of public policies have been “randomized” to subjects using lotteries, an IV that is almost equivalent to the validity of a randomized controlled trial. For example, lotteries using birth date have approximately randomized men to the draft during the Vietnam War (Angrist, Chen, and Frandsen 2010). This landmark study rigorously demonstrated the war’s profound and long-lasting negative effects on the health and well-being of drafted soldiers. The IV worked because only a small proportion of randomly selected draftees were able to avoid service due to illness, disability, and so on.

A longitudinal rather than a cross-sectional design, such as an interrupted time series analysis, that observed more hospitalizations immediately after reduced medication access because of a drug benefit limit (Soumerai et al. 1994) might represent an acceptable IV, because all of the confounding variables discussed above are generally constant over time (e.g., income location, ethnicity). Unfortunately, however, the IV method generally does not control for prior secular trends, and thus it is unable to control for what would have happened in the absence of the intervention. This lack of control produces questionable results. A formal interrupted time series design is more resistant to bias (Shadish, Cook, and Campbell 2002).

CONCLUSION

As noted, IV analyses are statistical analyses, not research designs (Sanghavi et al. 2015). IV studies can utilize powerful quasi-randomized public lotteries and interruptions in trends associated with interventions. Or, as is common, they can be weak correlational studies at one point in time that are inadequate to distinguish causes from effects in medical effectiveness research (Garabedian et al. 2014; Effective Practice Organisation of Care [EPOC] 2015).

Investigators seldom *search* for unmeasured characteristics that may offer alternative causal explanations. This could explain why Sanwald and Schober measured an effect on survival that was much greater than the most cited IV study of cath laboratories among heart attack patients (McClellan, McNeil, and Newhouse 1994). More important, they state in the paper that the survival advantage could have resulted from many non-cath-related treatments in the hospital.

Yet hundreds of such weak studies, spurred on by the U.S. economic stimulus package that promoted observational comparative effectiveness research (Soumerai 2009), confuse the public and provide counterproductive advice to clinicians and policy makers. As we wrote on a different subject in the Health Affairs Blog last year (Soumerai and Koppel 2015), “We spend a lot of money on medical research for our well-being. While no research is flawless, everyone should understand the strengths and weaknesses of the studies on which we base our policies, our economy, and our lives.” Weak observational and randomized research designs, tossed out from systematic reviews conducted by the Cochrane Collaboration, produce misleading research on medical interventions that can harm patients and inflate medical expenditures. We can do better. It is time to restore public confidence in what so often appears to be our flip-flopping field of contradictory evidence. Given that over 50 percent of health care studies use the weakest research designs (Soumerai, Starr, and Majumdar 2015), we can substantially increase the trustworthiness of published research by simply adhering to the Cochrane Collaboration’s design standards. Both we and the research upon which we base our health care and health policies deserve nothing less.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors report no relevant disclosure for this commentary. We are grateful for the advice from Dr. Laura Garabedian on the use of instrumental variables in comparative effectiveness research. We are also indebted to Caitlin Lupton for expert research assistance and reference management.

Disclosures: None.

Disclaimers: None.

Stephen B. Soumerai
Ross Koppel

REFERENCES

- Angrist, J. D., S. H. Chen, and B. R. Frandsen. 2010. "Did Vietnam Veterans Get Sicker in the 1990s? The Complicated Effects of Military Service on Self-Reported Health." *Journal of Public Economics* 94 (11–12): 824–37.
- Beck, C. A., J. Penrod, T. W. Gyorkos, S. Shapiro, and L. Pilote. 2003. "Does Aggressive Care Following Acute Myocardial Infarction Reduce Mortality? Analysis with Instrumental Variables to Compare Effectiveness in Canadian and United States Patient Populations." *Health Services Research* 38 (6 Pt 1): 1423–40.
- Brooks, J. M., M. McClellan, and H. S. Wong. 2000. "The Marginal Benefits of Invasive Treatments for Acute Myocardial Infarction: Does Insurance Coverage Matter?" *Inquiry* 37 (1): 75–90.
- Cutler, D. M. 2007. "The Lifetime Costs and Benefits of Medical Technology." *Journal of Health Economics* 26 (6): 1081–100.
- Effective Practice Organisation of Care [EPOC]. 2015. "EPOC-specific resources for review authors" [accessed on April 27, 2015]. Available at <http://epoc.cochrane.org/epoc-specific-resources-review-authors>
- Garabedian, L. F., P. Chu, S. Toh, A. M. Zaslavsky, and S. B. Soumerai. 2014. "Potential Bias of Instrumental Variable Analyses for Observational Comparative Effectiveness Research." *Annals of Internal Medicine* 161 (2): 131–8.
- Garber, A. M. 2011. "How the Patient-Centered Outcomes Research Institute Can Best Influence Real-World Health Care Decision Making." *Health Affairs (Millwood)* 30 (12): 2243–51.
- Glickman, M. E., and S. T. Normand. 2000. "The Derivation of a Latent Threshold Instrumental Variables Model." *Statistica Sinica* 10: 517–44.
- Koppel, R. 2013. *2013 Keynote Chapter: Is HIT Evidenced-Based?*. International Medical Informatics Yearbook. Stuttgart, Germany: Schattauer Publishers.
- Light, R. J., J. D. Singer, and J. B. Willet. 1990. *By Design Planning Research on Higher Education*. Cambridge, MA: Harvard University Press.
- McClellan, M. 1996. "Are the Returns to Technological Change in Health Care Declining?" *Proceedings of the National Academy of Sciences of the United States of America* 93 (23): 12701–8.
- McClellan, M., B. J. McNeil, and J. P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *Journal of the American Medical Association* 272 (11): 859–66.
- Sanghavi, P., A. B. Jena, J. P. Newhouse, and A. M. Zaslavsky. 2015. "Outcomes of Basic versus Advanced Life Support for Out-of-Hospital Medical Emergencies, Author's Response." *Annals of Internal Medicine* 163 (9): 681–90.
- Sanwald, A., and T. Schober. 2016. "Follow Your Heart: Survival Chances and Costs after Heart Attacks-An Instrumental Variable Approach." *Health Services Research* 52 (1 Pt 1): 16–34.
- Shadish, W. R., T. D. Cook, and D. T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, WA: Wadsworth Cengage Learning.

- Soumerai, S. B. 2009. *Breakthrough Science Can't Be Rushed*. Boston: The Boston Globe.
- Soumerai, S. B., and R. Koppel. 2015. "Avoiding Expensive and Consequential Health Care Decisions Based on Weak Research Designs." *Health Affairs Blog*.
- Soumerai, S. B., D. Starr, and S. R. Majumdar. 2015. "How Do You Know Which Health Care Effectiveness Research You Can Trust? A Guide to Study Design for the Perplexed." *Preventing Chronic Disease* 12: E101.
- Soumerai, S. B., T. J. McLaughlin, D. Ross-Degnan, C. S. Casteris, and P. Bollini. 1994. "Effects of a Limit on Medicaid Drug-Reimbursement Benefits on the Use of Psychotropic Agents and Acute Mental Health Services by Patients with Schizophrenia." *New England Journal of Medicine* 331 (10): 650–5.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.