Technical Report 29

Instrumental Variables with VA Data

Todd H. Wagner, Elizabeth H. Cowgill

April 2014

Instrumental Variables with VA Data. Technical Report 28.
Health Economics Resource Center (HERC)
VA Palo Alto Healthcare System
795 Willow Road (152 MPD)
Menlo Park, CA 94025
650-617-2630
650-716-2639 (fax)
herc@va.gov

Suggested citation: Wagner TH, Cowgill EH. Instrumental Variables with VA Data. Technical Report 29. Menlo Park, CA. VA Palo Alto, Health Economics Resource Center; April 2014.

Acknowledgements: We thank Christine Chee for her helpful comments. This work was supported by HSR&D. The views, opinions, and content of this publication are those of the authors and do not necessarily reflect the views, opinions, or policies of the Department of Veterans Affairs.

Contents

Terms v

1. Introduction 1

2. Instrumental Variables 2

3. Good Instruments 4

3.1 Relevance 4

3.2 Strength 4

3.3 Common Instruments 4

4. Summary 8

References 9

Figures

Figure 1. Instrumental Variables Method 3

Terms

AA	Alcoholics Anonymous
AIDS	Autoimmune deficiency syndrome
CBOC	Community based outpatient clinic
HAART	Highly-active anti-retroviral treatment
HIV	Human immunodeficiency virus
IVs	Instrumental variables
VA	U.S. Department of Veterans Affairs
VA ACT	VA Assertive Community Treatment
VHA	Veterans Health Administration

1. Introduction

Understanding causal relationships is critical for researchers. Although data from randomized controlled trials is preferred for understanding causal relationships, randomization is not always ethical or feasible, due to the high monetary and time costs.[1] With greater access to observational data, more people are questioning how to use large existing data to inform clinical decisions and policy.[1] Unfortunately results from observational analyses are prone to bias, especially when the primary right-hand-side variable (i.e., the treatment) is correlated with other factors not included in the analysis; this is often referred to as endogeneity.[1,2]

Why is endogeneity a problem? Regression models assume that all right-hand-side variables are exogenous, hence the right-hand-side variables are often referred to as independent variables. When a variable is endogenous (correlated with unobserved variables), it violates an underlying assumption in the statistical model, resulting in a biased regression coefficient. Instrumental variables (IVs) is a statistical modeling technique to correct for endogeneity. This report describes the use of IVs in VA data. Section 2 provides background on IVs and how to use them, section 3 reviews common examples of IVs in VA data and their pitfalls, and the final section summarizes our discussion.

2. Instrumental Variables

Newspaper headlines often report results from observational studies in which a causal relationship is suggested. Coffee has seen its fair share of assertions: "coffee consumption can cut skin cancer risk"[3] or "coffee may make you lazy."[4] Given the proliferation of observational data, especially in the VA, it is only natural to ask whether all observational research is flawed or whether it can be used to understand causation. The key, it turns out, largely has to do with what is not measured. In econometrics, features that are not measured are referred to as *unobservables* and a correlation between the right hand side variable of interest and unobservables creates a problem known as *endogeneity*.

To stay concrete, let's use an example. Let's assume you wanted to examine whether outcomes following a heart attack (myocardial infarction) were better at hospitals with more specialized equipment. You obtain a sufficiently large database of people who had a myocardial infarction, and this database includes an accurate measure of the hospital's specialized cardiac equipment (*technology*). You regress mortality (the dependent variable) on technology and find a negative association (outcomes were better in hospitals with more technology). To control for possible confounding, you then include patient risk, age, gender and education in the regression model. After adding these variables, you continue to see a negative correlation between technology and mortality.

You show your results to your health economist friend, who points out that you aren't controlling for many other variables, making the technology variable endogenous. Your friend suggests that cardiologists choose their jobs after residency based on the hospital's cardiac program and its interest in cardiac technology. Thus, your friend warns that the correlation you see could be a function of cardiologist sorting, rather than the technology itself.

One way of mitigating the endogeneity problem is to include as many possible control variables in the regression model as possible. In the cardiology example, you could include as many control variables as possible. After all, omitting an important variable biases the results, whereas including an irrelevant variable is a much less serious issue. The unfortunate reality is that you usually do not observe control variables for all of the possible confounders. And even if you do, sometimes those variables are measured with error that is associated with unobservables (a problem known as errors in variables). Thus, while adding more control variables can help mitigate endogeneity, it does not solve the problem.

Instrumental variables (IVs) were developed to address the underlying endogeneity problem. But before we get into the nuts and bolts of IVs, it is worthwhile to mention propensity scores. Researchers have recently started using propensity scores in great abundance. The challenge with multivariate techniques, including propensity scores, is that they only focus on observables, not the unobservables. Unfortunately, the bigger problem is the potential of unobserved confounding, and there is no way of knowing, *a priori*, how much this could bias the estimates.[1,2] IVs, unlike propensity scores, are designed to tackle the endogeneity.

An IV is directly linked to the independent (i.e., treatment) variable, but it is not directly linked to the outcome variable.[1,2,5,6] The IV must be irrefutably linked to the treatment, and this relationship must be strong. The instrument must also not be linked to the outcome (or alternatively associated with the error term). These pathways are shown below.

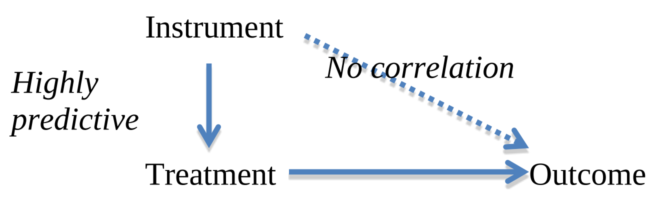

Figure 1. Instrumental Variables Method

Let's return to our example on mortality following a myocardial infarction. In this situation, let's assume you realize that when a person has a myocardial infarction and calls an ambulance, the ambulance usually takes the patient to the closest hospital with an emergency room. The ambulance driver does not consider the hospital's cardiac equipment. In this case, travel distance (or travel time) is the instrumental variable. Choice of hospital is determined by where the person was at the time they had the heart attack, which seems plausibly unrelated to the outcome. This example is based on a well-known 1994 *JAMA* paper by McClellan, McNeil, and Newhouse.[7]

Many econometric textbooks describe the math underlying IVs and the alternative estimators (e.g., two stage least squares, Wald). But it is easy to get lost in the discussion of Y, X, and Z and how Z acts a projection matrix, especially for people unfamiliar with econometrics. The goal here is not to replicate the math presented in detail elsewhere, but to reinforce the underlying intuition. Readers interested in the math are encouraged to read Bill Greene's book entitled, Econometric Analysis.

3. Good Instruments

There is a large economic literature on IVs. Two necessary features of IVs are relevance and strength.

3.1 Relevance

Relevance is the lack of association between the instrument and the outcome (dependent) variable.[8] In less technical terms, the IV must be directly related to the treatment, causing significant variation in the treatment, but only influences the outcome through the IV's influence on the treatment.[1,5,9] IVs are irrelevant if they are correlated with the outcome. They can also be irrelevant if the IV is measured with error, which is, in turn, correlated with the outcome. There are empirical tests for exogeneity, but they are of limited power and most people use common sense.[1,10] If the instrument could be plausibly endogenous (correlated with the outcome), then it probably is not a good instrument. We will discuss this more in section 3.3 with regard to common instruments.

3.2 Strength

An IV's *strength* (i.e., its predictive power) is measured by computing a partial F-statistic. A partial F-statistic greater than 10 is often used as the measure of a minimally adequate IV. An instrument with a partial F-statistic less than 10 is often considered weak; the easiest way to calculate the partial F-test is to square the t-statistic on the IV. Using a weak instrument can create serious bias and estimation problems.[11] Weak instruments are more likely in small samples given that precision decreases with sample size (SE=SD/sqrt(n)).

3.3 Common Instruments

Finding instruments that meet both the relevance and strength requirements is difficult, and this is one of the biggest limitations with this method. Published studies provide examples of IVs. Below we highlight some common instruments and discuss their strengths and weaknesses.

<u>3.3.1 Randomization</u>

Randomized trials generally use an intent-to-treat analysis to examine safety and efficacy. In behavioral trials, poor compliance and use of the intervention by control group members (i.e., unintended crossover) can impede the intent to treat analysis. For example, consider a trial in which patients with alcohol abuse problems were randomized to usual care or usual care plus a referral to Alcoholics Anonymous (AA). Many people in the AA group may not go to AA and many people in the usual care group may go to AA. The intent-to-treat analysis provides information on the effectiveness of referral, but it may not provide information on the effectiveness of AA attendance.

Another analytical approach would be to compare people who followed the protocol as intended; these are often referred to as a per protocol analysis. However, the decision to attend AA meetings is endogenous (the decision to attend could be correlated with many unobserved factors) and randomization (randomly referring people to AA) could be an excellent instrument. Randomization meets the relevance criterion and one can test for the instrument's strength.

There are other examples where randomization could act as an instrument. The Veterans Victory over Tobacco study[12] was a VA randomized controlled trial of all Veterans who are current smokers, as identified by their electronic medical record. The study randomized Veterans to a proactive intervention or usual care, using the randomization process to minimize bias. Using randomization in this way has attracted limited attention. Sussman and colleagues refer to this technique as contamination-adjusted clinical trial,[13] suggesting that it is a good way to handle unintended bias in trials, as was encountered by McKellar et al.[14] when studying follow-up for substance use treatment.

Unfortunately, a major limitation to using randomization as an instrument is sample size since many clinical trials are small. A partial F-statistic greater than 10 is often the litmus test for a strong instrument, and F-statistic is inversely related to the sample size, all else being equal. In these cases, investigators might have to pool data across multiple trials to get a sufficiently large enough sample size.

3.3.2 Travel Distance and Travel Time

Travel distance is one of the most common IVs,[15-19] especially in VA studies.[20-23] Travel time is the primary nonmonetary price for obtaining care,[24] and distance is often strongly associated with health care utilization.

Patient zip codes are listed in the PTF and SE datasets, making it relatively straightforward to calculate distances from the patient's zip code to the medical center. The VA Planning Systems Support Group has a dataset called VAST in which these distances are already calculated.

Using travel distance or travel time assumes that travel is associated with receiving care.[2] Distance can simply measure the straight line distance from the subject's home to the nearest hospital (Euclidian distance)[25] or can measure the difference of distances between the subject's home and two different locations (differential distance). Most research has found an inverse relationship between travel time and use of care. For example, a study[20] examining the impact of newly opened VA community based outpatient clinics (CBOCs) on Veteran use of primary care services, used the change in (Euclidean) travel distance as an instrument. They found that decreased travel distance to a CBOC predicted a significant increase in primary care encounters.[20] The strength of the instrument remains an empirical question that needs to be tested in each study.

In a study examining the cost savings from the VA Assertive Community Treatment (VA ACT) program for patients with severe mental illness, the unmeasured confounders were controlled for using the distance from the address of the patient's residence to nearest VA ACT team.[26] The IV estimate found that use of the VA ACT program yielded fewer inpatient mental health bed days but did not significantly lower probability of an inpatient mental health admission nor did it significantly lower total mental health costs.[26]

Is distance exogenous? Like many things, the answer depends on the situation. In the most famous study using distance as an instrument, McClellan and colleagues[7] studied the effect of hospital technology on treatment for heart attacks (myocardial infarction). They argued that distance was exogenous because the heart attack was not expected and because of the emergent nature of the attack, the person was taken to the nearest emergency department. In this case, the instrument was both a strong predictor and reasonably exogenous.

Other studies have attempted to use distance without success. An unpublished study attempted to assess the effect of using drug treatment on health outcomes. In this case, distance proved problematic for two reasons. First, people who have to make many trips to receive treatment often relocate themselves to make it easier. This raises serious concerns about the validity (exogeneity) of the instrument. Second, drug treatment facilities are not randomly distributed across geographical areas. They are disproportionately located in urban areas and many affluent communities fight the decision to open a treatment facility in their neighborhood. This implies that the instrument is not strong for all groups (distance is not a strong determinant for people living in urban areas, many of whom are minorities).

Caution is warranted when using distance. Both strength and exogeneity can fail for various reasons.

3.3.3 Changes in Co-Payments

Demand for health care is price sensitive and this sensitivity (often referred to as elasticity) is greater for ambulatory care than emergency or specialty care. Calculating change in co-payments is simple in the VA for those Veterans who are required to pay. Dual eligibility for VHA and non-VHA care (Medicare, Medicaid, private) is common and complicates the picture because it creates questions about non-VHA care and cross-price elasticities (does an increase in VA copayments increase non-VA use).

A non-VA study[27] of the impact of patient adherence to ACE inhibitors or β-blockers on health outcomes in patients with chronic heart failure used changes in prescription drug co-payments as an IV. They found that as prescription drug costs increased, patients were less likely to adhere to drug regimens, yielding worse health outcomes.

One challenge with using copayments as an instrument is that changes in co-payments tend to be very small, as are the resulting effect on demand. Thus, the price change may not be big enough, relative to a person's willingness to pay, to act as a strong instrument.

3.3.4 Calendar Time

New health technologies are continually rolling out. One question is whether calendar time can act as an instrument. In this case, calendar time is capturing the implicit change in the supply of care (e.g., a pre and post variable). For example, recommended HIV/AIDS treatments have varied considerably over the past 20 years. One study[28] examined the effect of HAART (highly-active anti-retroviral treatment) in preventing the progression of HIV to AIDS through the use of calendar period (pre-HAART versus HAART) as an IV.[28] People have also used the placement of drugs on formularies as an instrument (again pre and post).

Calendar time needs to be checked to determine if it is a strong instrument. More importantly, however, great care must be used in determining whether it is relevant. Frequently clinicians (and sometimes patients) are aware of pending changes and change their behaviors accordingly. Also other related treatments change over time and care must be given to control for these changes.

3.3.5 Provider Preference or Organizational Capability

Brookhart, Rassen, and Schneeweiss[2] define a category of IVs as "preference-based instruments" based on the assumption that providers, provider groups, or patients have different preferences. The dominant example in this category is physician preference or choice.[29,30] Physician preference refers to a physician's inclination to prescribe one treatment over another. For example a study[30] examining the VA's use of palliative care consultation in relationship to hospital costs compared the costs of palliative care patients versus non-palliative care patients with advanced disease. Using the attending physician's preference for offering palliative care consultation as an IV, the estimate found that costs of palliative care patients were significantly less than those of non-palliative care patients with advanced disease.[30]

These are highly suspicious IVs by definition. We know that physicians have differential preference and they are attracted to where they work by their preferences. These preferences are not fully observed and may be correlated with the key variables of interest. Great caution is warranted in using these variables as instruments.

3.3.6 Genomics

One can expect that advances in genomics will provide many opportunities for using them as instruments. At face value, many may be strong predictors and seem plausibly exogenous. However, caution is warranted given that we do not fully understand how genetics affect systems and behaviors.

4. Summary

Large health care databases offer many opportunities to study the organization and delivery of health care. The growth of information, however, must be tempered with the realization that more information does not yield more wisdom. What is needed is the way to discern causal relationships from mere associations. To gain insights into causal relationships, one needs to adjust for unobserved confounders, something that multivariate regression and propensity scores cannot do. IV regression provides an empirical tool for understanding causal relationships with observational data. This report highlighted the intuition behind IVs and commonly used instruments in health care. Great caution is warranted in using instruments as none are applicable in all circumstances. Analysts must consider the strength of the instrument and whether the instrument is plausible exogenous. Many proposed instruments have failed on both grounds.

References

1. Sinnott P. Instrumental Variables Models [PPT]. Menlo Park: Health Economics Resource Center (HERC); 2011.
2. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental Variable Methods in Comparative Safety and Effectiveness Research. Effective Health Care Research Report No. 22. (Prepared by Brigham and Women's Hospital DEcIDE Center Under Contract No. 290-2005-0016I T03). Rockville, MD: Agency for Healthcare Research and Quality (AHRQ); April 2010.
3. Roizen M, Oz M. Coffee consumption can cut skin cancer risk. *TheRecord.com.* 2011. http://www.therecord.com/living/healthfitness/article/626550--coffee-consumption-can-cut-skin-cancer-risk. Accessed April 23, 2012.
4. Tarkan L. Study: Coffee may make you lazy. *Fox News* 2012 Mar 28; http://www.foxnews.com/health/2012/03/28/study-coffee-may-make-lazy/.
5. Newhouse JP, McClellan M. Econometrics in Outcomes Research: The Use of Instrumental Variables. *Annual Review of Public Health.* 1998;19(1):17.
6. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *Journal of Clinical Epidemiology.* 2009;62(12):1226-1232.
7. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables [see comments]. *JAMA.* 1994;272(11):859-866.
8. Hall AR, Rudebusch GD, Wilcox DW. Judging Instrument Relevance in Instrumental Variables Estimation. *International Economic Review.* 1996;37(2):283-298.
9. Barnett P, Gage M, Finney J. Comparison of Cost and Outcomes of Inpatient and Outpatient Treatment for Substance Use Disorders [PPT]. Menlo Park, CA: Health Economics Resource Center (HERC); 2008.
10. University of Utah. Instrumental-variables Estimators. *[Econometrics 2009 Course Notes]*: University of Utah; 2009:1-13.
11. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *JASA.* 1995;90(430):443-450.
12. Fu S, van Ryn M, Sherman S, et al. Population-based tobacco treatment: study design of a randomized controlled trial. *BMC Public Health.* . 2012;12(159).
13. Sussman JB, Hayward RA. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. *BMJ.*340:1756-1833.
14. McKellar J, Wagner T, Harris A, Oehlert M, Buckley S, Moos R. One-year outcomes of telephone case monitoring for patients with substance use disorder. *Addict Behav.* Oct;37(10):1069-1074.
15. Brooks JM, Irwin CP, Hunsicker LG, Flanigan MJ, Chrischilles EA, Pendergast JF. Effect of Dialysis Center Profit-Status on Patient Survival: A Comparison of Risk-Adjustment and Instrumental Variable Approaches. *Health Services Research.* 2006;41(6):2267-2289.
16. Gore JL, Litwin MS, Lai J, et al. Use of Radical Cystectomy for Patients With Invasive Bladder Cancer. *Journal of the National Cancer Institute.* June 2, 2010 2010;102(11):802-811.
17. Hadley J, Cunningham P. Availability of Safety Net Providers and Access to Care of

Uninsured Persons. *Health Services Research.* 2004;39(5):1527-1546.
18. Lu M. The productivity of mental health care: an instrumental variable approach. *J Ment Health Policy Econ.* 1999 Jun 1999;2(2):59-71.
19. Xian Y, Holloway RG, Chan PS, et al. Association Between Stroke Center Hospitalization for Acute Ischemic Stroke and Mortality. *JAMA: The Journal of the American Medical Association.* January 26, 2011 2011;305(4):373-380.
20. Fortney JC, Steffick DE, Burgess JF, Maciejewski ML, Petersen LA. Are Primary Care Services a Substitute or Complement for Specialty and Inpatient Services? *Health Services Research.* 2005;40(5p1):1422-1442.
21. Kim HM, Eisenberg D, Ganoczy D, et al. Examining the Relationship between Clinical Monitoring and Suicide Risk among Patients with Depression: Matched Case–Control Study and Instrumental Variable Approaches. *Health Services Research.* 2010;45(5p1):1205-1226.
22. Pracht EE, Bass E. Exploring the Link between Ambulatory Care and Avoidable Hospitalizations at the Veteran Health Administration. *Journal for Healthcare Quality.* 2011;33(2):47-56.
23. Simeonova E. Race, Quality of Care and Patient Outcomes: What Can We Learn from the Department of Veterans Affairs? *Atlantic Economic Journal.* 2009;37(3):279-298.
24. Acton JP. Nonmonetary factors in the demand for medical services - some empirical evidence. *JPE.* 1975;83(3):595-614.
25. Phibbs CS, Luft HS. Correlation of travel time on roads versus straight line distance. *Med. Care Res. Rev.* 1995;52(4):532-542.
26. Slade EP, McCarthy JF, Valenstein M, Visnic S, Dixon LB. Cost Savings from Assertive Community Treatment Services in an Era of Declining Psychiatric Inpatient Use. *Health Services Research.* 2012.
27. Cole JA, Norman H, Weatherby LB, Walker AM. Drug Copayment and Adherence in Chronic Heart Failure: Effect on Cost and Outcomes. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy.* 2006;26(8):1157-1164.
28. Cain LE, Cole SR, Greenland S, et al. Effect of Highly Active Antiretroviral Therapy on Incident AIDS Using Calendar Period as an Instrumental Variable. *Am J Epidemiol.* 2009 May 1 2009;69 (9):1124-1132.
29. Brooks JM, Chrischilles EA, Scott SD, Chen-Hardee SS. Was Breast Conserving Surgery Underutilized for Early Stage Breast Cancer? Instrumental Variables Evidence for Stage II Patients from Iowa. *Health Services Research.* 2003;38(6p1):1385-1402.
30. Penrod JD, Deb P, Dellenbaugh C, et al. Hospital-based palliative care consultation: effects on hospital cost. *J Palliat Med.* 2010 Aug 2010;13(8):973-979.