Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships

Jeremy A. Rassen[a,b,c,*], **M. Alan Brookhart**[a,b], **Robert J. Glynn**[a,b,d], **Murray A. Mittleman**[b,c,e], and **Sebastian Schneeweiss**[a,b,c]

[a]Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham & Women's Hospital, Boston, MA, USA

[b]Harvard Medical School, Boston, MA, USA

[c]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

[d]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

[e]Cardiovascular Epidemiology Research Group, Beth Israel Deaconess Medical Center, Boston, MA,

Abstract

The gold standard of study design for treatment evaluation is widely acknowledged to be the randomized controlled trial (RCT). Trials allow for the estimation of causal effect by randomly assigning participants either to an intervention or comparison group; through the assumption of "exchangeability" between groups, comparing the outcomes will yield an estimate of causal effect. In the many cases where RCTs are impractical or unethical, instrumental variable (IV) analysis offers a nonexperimental alternative based on many of the same principles. IV analysis relies on finding a naturally varying phenomenon, related to treatment but not to outcome except through the effect of treatment itself, and then using this phenomenon as a proxy for the confounded treatment variable.

This article demonstrates how IV analysis arises from an analogous but potentially impossible RCT design, and outlines the assumptions necessary for valid estimation. It gives examples of instruments used in clinical epidemiology and concludes with an outline on estimation of effects.

Keywords

Pharmacoepidemiology; Instrumental variable; Confounding factor (epidemiology); Bias (epidemiology); Physician prescribing preference; Unmeasured confounding

1. Introduction

When questions of causality arise, epidemiologists widely acknowledge the randomized controlled trial (RCT) as the "gold standard" of research designs. In cases where RCTs are not possible–for financial, ethical, practi-cal, or other reasons–alternative methods must be used. These alternatives comprise the time-tested epidemiology toolbox: cohort studies, case-control studies, case-crossover studies, and their brethren. Although these approaches are appropriate in many instances, they do fundamentally lack an intervention; the classic experimental method

*Corresponding author. Division of Pharmacoepidemiology and Phare macoeconomics, Brigham & Women's Hospital, 1620 Tremont Street, Suite 3030, Boston, MA 02120, USA. jrassen@post.harvard.edu .

of establishing causality is to intervene in one group while leaving a second control group aside. Nonexperimental methods of causal inference must rely on an assumption of no unmeasured confounding [1,2], an assumption that is hard to justify in many cases, particularly in pharmacoe-pidemiologic studies based on health care claims and utilization data [3–5].

For decades, economists have been using instrumental variable (IV) analysis as a method of causal inference in cases where an RCT is not possible and when an assumption of no unmeasured confounding is unwarranted (Table 1). Although IV analysis is certainly no panacea for all that ails the non-randomized study, it does offer a tool for instances when the alternative methods do not work.

As an example, consider a question in cardiac care: does catheterization prevent death after myocardial infarction (MI)? This question has been addressed in several IV studies by McClellan and Newhouse [6,7].

In response, consider a group of patients who have experienced an MI. Divide these patients into two observed groups: those who were catheterized after their MI and those who were not; divide them again by who did and did not die. From the fabricated numbers in Table 2a, an odds ratio of 0.211 and a risk difference (RD) of 0.150 can be calculated, indicating that catheterization is strongly associated with reduced risk of death. Causality, however, is unknown: the treatment may be highly protective, or selection into the catheterization group may be indicative of overall health and reduced risk of death. In this setting, covariates typically available in health care utilization data (prior MI, age, and history of various comorbid conditions), or even covariates frequently available in prospective cohort studies (smoking, body mass index, or blood pressure) are unlikely to be sufficient to control for confounding. If the decision to catheterize depends on these variables, the assumption of no unmeasured confounding cannot be justified.

What is new?

- Instrumental variable (IV) analysis provides a method to obtain a potentially unbiased estimate of treatment effect, even in the presence of strong unmeasured confounding.

- This article outlines the analytical method and the assumptions required for IV analysis.

- Several examples are provided to illustrate the strengths and potential pitfalls of the IV approach.

This article and its companion, "Instrumental variable application: In 25 variations, the physician prescribing preference generally was strong and reduced covariate imbalance," together introduce the concept of IV analysis and examine some of the key assumptions underlying the technique. Taken together, the articles show how IVs arise in observational data and how IV analysis parallels randomized trial designs, and also examine the key notions of instrument strength and validity. Each of them describes instruments that have been used in clinical epidemiology and gives examples of IV analysis.

The problem of deducing causality is familiar to clinical epidemiologists; treatment-outcome relationships can be obscured by the combined effect of measured and unmea-sured confounders. Examples that may be affected include the use of hormone replacement therapy and incidence of coronary heart disease (CHD) [8] and vitamin E supple-ments and CHD [9].

In each of these cases, the nonrandom-ized results had been tantalizing, but their perceived unreliability [10,11] prompted randomized trials to confirm or refute their findings [12].

This article will introduce the use of IV analysis as a sup-plement to standard epidemiologic methods. It will explain IVs from a conceptual perspective by looking at how IV studies arise from their randomized trial analogs.

2. The interventional approach: a randomized trial

We began with the premise that the most reliable test of causality is the RCT. In that spirit, imagine that on an MI patient's entry into the emergency room, a coin is flipped; heads means that the patient will be slated to receive cath-eterization and tails indicates that he will not. Assume that all other hospital care would be equivalent whether or not this patient is catheterized, except to the extent that cathe-terization itself leads to changes in clinical care. This heads/tails assignment is an intervention in which the course of events will be dictated by the reading of the coin rather than by their natural course. It is important to note that the intervention in question is the assignment of treat-ment than the treatment actually received.

If this study were carried out today, it may not meet ethical standards for equipoise in post-MI care. Therefore, as an al-ternative, imagine examining data on those same MI patients 30 days after their hospital care. If it were possible to observe something about those patients, other than their health status, which could in retrospect serve to separate them into two ran-dom groups, then that random group assignment should serve the same function as a coin. In this sense, we are looking for a "natural experiment" in the data, a happenstance occur-rence whose randomness can be exploited to perform a retro-spective, nonexperimental "trial." A marker for this occurrence is called an instrument or IV. Like the coin in an RCT, it must influence treatment, but have no independent effect on the outcome.

3. The instrumental variable approach: distance as an instrument for catheterization

The challenge is to identify such an instrument. In this case, McClellan et al. observed that some hospitals provide catheterization, whereas others do not (or do so only infre-quently) [6,7]. They hypothesized that the patient's differen-tial distance from catheterization-providing hospital may be a determinant of receiving catheterization. Differential dif-ference was defined as the extra distance that an ambulance would have to travel to deliver the patient to a catheteriza-tion-providing hospital as opposed to a hospital without cath-eterization facilities. They hypothesized that the paramedic was more likely to go to the nearer hospital rather than select a farther one based on the availability of particular facilities, and that, all things equal, patients living short differential distances to catheterization-providing hospitals would be more likely to receive catheterization solely as a result of their proximity. As such, a short differential distance would be a predictor of receiving catheterization.

3.1. The three basic assumptions of instrumental variable analysis

If short differential distance is a valid proxy for a ran-domizing coin–a valid instrument–it must meet three fundamental criteria. As mentioned earlier (Table 1), (1) the instrument has to predict the actual treatment a patient received. The frequency that the instrument predicts the actual exposure is called the strength of the instrument.

Like the coin flipped at the beginning of an experiment, the instrument also cannot have any bearing on the outcome by either (2) direct associations or (3) associations as a result of common causes of the instrument and the out-come; it can only affect the outcome by the treatment itself. These assertions are termed the independence assumption and exclusion

restriction, respectively. In randomized ex-periments, independence and exclusion should be met by design. In nonexperimental designs using IV analysis, the exclusion restriction can be violated by the existence of common causes of both the instrument and the outcome, and is met only by assumption.

In non-experimental settings, and even in randomized trials, the independence assumption and exclusion restric-tion are fundamentally unverifiable. Indeed, many of the problems with RCTs, such as poor randomization leading to treatment group imbalance, are empirical violations of independence or exclusion. We will return to the topic of assumptions later.

3.2. Compliance and the effect on the marginal subject

In a randomized trial, there are three categories of partic-ipants: those who follow the coin flip (compliers); those who note the advice of the coin but do what they were going to do anyway (noncompliers; in drug studies, these are always-takers or never-takers); and those who will always do the opposite of what the coin flip tells them to do (defiers). In an RCT, blinding removes the possibility of defiance.

In an RCT, the participants who do follow the coin flip, the compliers, will provide the statistical information that will determine the effect measure of the study, because by ran-domization, the noncompliers should be equally distributed in the two treatment groups. In the usual intention-to-treat (ITT) analysis, the random distribution of noncompliance will yield a bias toward the null.

In the IV setting, some patients will also "comply," that is, their treatment status will be determined by the status of the instrument (living close to a catheterization-providing facility) and not by another choice process (severity of the cardiac condition). The compliers are often the people who could have benefited equally from either treatment, and therefore, the natural randomness incorporated in the instrument became the factor that tipped them toward one therapy or the other. As in an RCT, those who comply–termed mar-ginal subjects in the IV arena–will provide information about the effect of treatment, as they are the ones whose exposure was directly affected by the instrument.

For this reason, IV analysis provides an estimate of the ef-fect of treatment among the marginal subjects (compliers). This estimate is then scaled to a figure that reflects the effect of treatment had everyone in the population been marginal [13]. If it seems plausible that the treatment has the same ef-fect on everyone in the population, then the scaled estimate can be interpreted as an estimate of the population average treatment effect. If not, the parameter should be interpreted carefully, but can be especially meaningful in clinical circumstances where the effect among patients whose treatment choice is not clearcut is of substantive interest.

3.3. Treatment-effect heterogeneity

The basic IV analytical framework presented here makes a fundamental assumption that the effect of treatment is constant among the population under study [14]. If a young person has a 20% benefit from treatment, then an old person in the population should receive the same 20%. This treatment-effect homogeneity assumption is parallel to the assumption of no effect modification made in many clinical epidemiology studies, and is considered a reasonable place to begin an analysis [15].

Like Mantel-Haenszel (M-H) and other traditional epide-miology methods used to summarize effect estimates over strata, an overall IV-based estimate is a weighted average of a number of stratum-specific estimates. If there is a differ-ential effect of treatment in any of the strata, the overall average may not be wrong *per se*, but will need to be interpreted in an appropriate light. The M-H method up-weights strata with small variances, whereas an IV analysis up-weights

strata where the IV strongly affected the treatment, that is, the strata with the most marginal patients [16].

In instances where there is treatment-effect heterogeneity, it is possible to estimate an average effect of treatment in a specific subgroup of marginal patients. This estimate is called a local average treatment effect (LATE). However, to do this, one must assume that there are no defiers in the study, an assumption termed monotonicity [17]. As defiance is indistinguishable from noncompliance in the data, and because the treatment effect is unequal from stratum to stratum, the presence of defiance could introduce bias. Monotonicity is reasonable in most RCT examples but has to be carefully evaluated in many IV applications.

4. Instrumental variables in epidemiology

With much of the theory in place, the challenge remains to find strong, valid instruments. Cole et al. used insurance benefit status as a proxy for how adherent a patient was likely to be [18]. They hypothesized that as insurance companies modify reimbursement policies over time, patients may react by altering adherence rates (amount of medication consumed). This notion would hold if all patients remain in the plan before and after the change, but if certain patients self-select into different plans as a result of the modification, the IV assumptions would be violated.

Smith et al. suggested a genetic marker as an IV in a technique called Mendelian randomization [19]. To study the cardioprotective effects of alcohol, confounded by factors including the potentially negative health behaviors associated with those who are heavy drinkers, they proposed using the aldehyde dehydrogenase gene as an instrument. Lack of this gene, which inhibits the ability to efficiently metabolize alcohol, makes alcohol consumption unpleasant, and is therefore a predictor of lower alcohol use. They theorized that the gene is not associated with cardiovascular disease, but if there were a direct association between the gene and outcome, or an indirect one through other genes working in combination, then the IV assumptions would not hold.

Stukel et al. used differences in regional catheterization rates as a proxy for whether a patient at a particular hospital would receive catheterization after MI [20]. Treatment predicted exposure: an MI patient living in a particular region was more likely than not to get that region's standard of care. A violation of the exclusion restriction could come from a link between the regional rate and outcome: if the regional rate were low and that low rate were in turn associated with generally worse health state, perhaps because of access-to-care issues, then assumption (3) would not hold. In the distance example, if great differential distance to a catheterization facility were also a proxy for poor access to other health care services, then exclusion would be violated.

Like the regional rate of catheterization, many of the IVs that have been used in clinical epidemiology fall into the category of preference-based instruments, where a behavior pattern at the regional, facility, or physician level is used to predict treatment for a particular patient [21–24]. Brookhart et al. considered physician-level preference: they used the physician's preference for prescribing one treat-ment over another as an IV [25,26]. This example will be considered in greater detail in the following section.

4.1. Making use of natural variation in treatment choice: physician prescribing preference

The use of physician prescribing preference (PPP) as an IV is based on the observation that in some instances, prescribing varies more among physicians than it does within a particular physician's practice [27,28]. It is posited that this diminished within-physician variation is a result of doctors' simple preference for one drug over another. The preference could have any

sort of basis: drug A might have worked well in a previous patient, or drug B might have been marketed heavily. Whatever the motivation, when pre-sented with a patient who could benefit equally from either treatment, the hypothesis says that underlying preference will govern the doctor's choice [27].

To motivate PPP, consider a simple interventional study of Cox-2 inhibitors (coxibs) vs. nonselective nonsteroidal anti-inflammatory drugs (NSAIDs) for pain control and protection against gastrointestinal bleeds. A patient for whom either treatment is appropriate presents himself to a study panel; a coin is flipped and the patient is random-ized to a treatment arm. The coin will therefore predict treatment.

With this hypothetical intervention in place, we now seek to replace the coin with the PPP instrument using the following logic. If preference shows natural variation, and if patients choose their doctors without knowledge or sense of that preference (or factors associated with preference, such as quality of care), then PPP can be substituted for the randomizing coin. In short, physician preference lets patients be "quasi-randomized" to coxib vs. nonselective NSAID treatment.

5. Evaluating the instrumental variable assumptions

For PPP to work as an IV, it must meet assumptions (1) through (3) as stated earlier. Assumption (1) says that pref-erence is related to treatment choice. With an appropriate measure of a physician's preference, we can test whether assumption (1) holds: the strength of the association can be quantified and the assumption verified by means of goodness-of-fit measures, such as the F statistic, often cited by economists, or the partial r^2 value [29,30].

Assumption (2) states that there is no direct relationship from PPP to outcome, except through the treatment pre-scribed. For this assumption to be met, preference cannot be associated with the physician's overall outcomes or quality of care: consider that coxibs are a new, beneficial treatment and nonselective NSAIDs are existing standard of care. If a doctor is using nonselective NSAIDs not because he thinks they are better but rather because he is not aware of newer treatment alternatives, then the nonse-lective NSAID-preferring physician might have poorer overall outcomes in his patients, and his preference for NSAIDs would be correlated with worse outcome [31].

A violation of assumption (3) is also easily conceived. A clustering of high-risk patients might arise around specialist physicians, or patients at higher risk may "doctor shop," seeking out physicians likely to prescribe a particular med-ication. This clustering could create a pool of severity: all the patients at this doctor's office have risk factors for the outcome, and they have chosen the doctor based on her known or perceived preference for a particular treatment. This self-assignment of patients to doctors who prefer a particular drug will create a violation of assumption (3). Differences in case mix are an important potential violation which can be reduced by focusing the analysis on a fairly homogenous group of physicians [32].

It may be apparent that assumptions (2) and (3) can be examined but are fundamentally unverifiable. As was stated earlier, IVs are not a panacea for the problems of non-randomized studies, but rather, IV analyses trade one set of unverifiable assumptions (no unmeasured confounding) for another (unconfounded instruments). A belief that as-sumptions (2) and (3) do indeed hold can come from empirical evaluation, subject matter expertise reading or reasoning, but not from any statistical test [17].

6. Analyzing the data: causal effect on the marginal subject

Going back to the example of distance as a proxy for catheterization, if the data from Table 2a (crude RD 5 0.150) are reanalyzed by using "short differential distance" in place of "received catheterization" and "long differential distance" in place of "didn't receive catheterization" (Table 2b; RD 5 −0.100), then the confounding effect of selection for catheterization and death should be removed by the quasi-randomized treatment arising from the natural variation in the place where patients live. In this case, moving from the treatment-based estimate to the IV-based estimate switches the direction of the effect estimate. This estimate of differential distance on catheterization may be muted because there might be a significant number of nonmarginal patients, patients for whom distance was not the factor that determined their treatment (Table 2c;RD 5 0.494). To assess the full effect, the association observed in Table 2b must be rescaled by the degree to which short differential distance was truly associated with catheterization, by dividing the estimate by the instrument strength (Table 2c), a number between −1 and 1.

The simple calculation of the IV estimate on the RD scale is as follows:

$$RD = \frac{\text{Instrument} \rightarrow \text{Outcome association}}{\text{Instrument strength}}$$
$$= \frac{\text{Distance} \rightarrow \text{Death association}}{\text{Distance} \rightarrow \text{Catheterization association}}$$
$$= \frac{-0.100}{0.494} = -0.202$$

The numerator in the fraction is the IV-to-outcome relationship, and will also range from −1 to 1; in a randomized study, the numerator is simply the ITT estimate. The denominator is the scaling factor that accounts for compliance. A strong instrument will yield a rescaling factor toward ±1, whereas a weak instrument will be closer to zero. Importantly, if any of the assumptions have been violated, scaling may magnify any bias from residual unmeasured confounding that is factored into the numerator [33–35].

This fraction, the so-called Wald estimator, is useful for only the most basic IV estimates. As in most epidemiological studies, models are almost always used in place of simple 2×2 tables. In the IV case, the most common estimation technique is known as two-stage least squares (2SLS).

2SLS applies two ordinary least square (OLS) models sequentially to create an estimate of effect [33,34]. The first stage predicts the expected value of treatment for patient i, $E[X_i]$ based on the instrument Z_i, that is, it uses the instrument Z_i and any covariate C_i to predict what the treatment "should" have been based on the data. The second stage then predicts the outcome $E[Y_i]$ as a function of the predicted treatment, which is fed in from the first stage along with the same covariates. The basic notion is that if we replace the confounded treatment with a prediction of treatment that by the IV assumptions we can say is unconfounded, then we get an unconfounded estimate of the causal RD. Note that by using covariates C_i, we can relax assumption (3) by asserting that the IV is not indirectly related to the outcome after adjusting for measured confounders. Because of the two-stage construction of the model and the imperfect prediction of treatment by the IV, IV analyses are generally less efficient than similar conventionally adjusted studies and have wider confidence intervals.

7. Conclusion

The IV analysis of nonrandomized data in clinical epidemiology can be a gift that comes with certain strings attached. If a valid instrument can be identified, it has the potential for unbiased estimation of treatment effects, but at the same time, it is impossible to be certain that all

necessary assumptions for instrument validity have been ful filled. With that major caveat in mind, we believe that with proper design and due caution, IVs are a sensible addition to the toolbox of clinical epidemiology. Several successful examples for IV analyses in clinical epidemiology have already demonstrated the promise of this method.

Acknowledgments

References

1. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000;11:561–70. [PubMed: 10955409]

2. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. Commun Statist Theory Meth 1994;23:2379–412.

3. Walker AM. Confounding by indication. Epidemiology 1994;7:335–6. [PubMed: 8793355]

4. Schneweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin epidemiol 2005;58:323–37. [PubMed: 15862718]

5. Glynn RJ, Schneeweiss S, Wang PS, Levin R, Avorn J. Selective prescribing led to overestimation of the benefits of lipid-lowering drugs. J Clin epidemiol 2006;59:819–28. [PubMed: 16828675]

6. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. JAMA 1994;272:859–66. [PubMed: 8078163]

7. Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health 1998;19:17–34. [PubMed: 9611610]

8. Hsia J, Criqui MH, Herrington DM, Manson JE, Wu L, Heckbert SR, et al. Conjugated equine estrogens and peripheral arterial disease risk: the Women's Health Initiative. Am Heart J 2006;152:170–6. [PubMed: 16824852]

9. Rimm EB, Stampfer MJ, Ascherio A, Giovannucci E, Colditz GA, Willett WC. Vitamin E consumption and the risk of coronary heart disease in men. N engl J Med 1993;328:1450–6. [PubMed: 8479464]

10. Taubes G. Do we really know what makes us healthy? New York Times Sunday Magazine. September 16;2007

11. Avorn J. In defense of pharmacoepidemiology–embracing the yin and yang of drug research. N engl J Med 2007;357:2219–21. [PubMed: 18046025]

12. Lee IM, Cook NR, Gaziano JM, Gordon D, Ridker PM, Manson JE. Vitamin E in the primary prevention of cardiovascular disease and cancer: the Women's Health Study: a randomized controlled trial. JAMA 2005;294:56–65. [PubMed: 15998891]

13. Harris KM, Remler DK. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. Health Serv Res 1998;33(5 Pt 1):1337–60. [PubMed: 9865223]

14. Brooks JM, Chrischilles EA. Heterogeneity and the interpretation of treatment effect estimates from risk adjustment models and instrumental variable methods. Med Care 2007;45(10 Suppl 2):S123–30. [PubMed: 17909370]

15. Angrist JD. Estimations of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. J Business Econ Stat 2001;19:2–16.

16. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. Int J Biost 2007;3 (article 14).

17. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology 2006;17:360–72. [PubMed: 16755261]

18. Cole JA, Norman H, Weatherby LB, Walker AM. Drug copayment and adherence in chronic heart failure: effect on cost and outcomes. Pharmacotherapy 2006;26:1157–64. [PubMed: 16863491]

19. Davey, Smith G.; Ebrahim, S. What can Mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ 2005;330:1076–9. [PubMed: 15879400]

20. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA 2007;297:278–85. [PubMed: 17227979]

21. Wang PS, Schneeweiss S, Avorn J, Fischer MA, Mogun H, Solomon DH, et al. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. N engl J Med 2005;353:2335–41. [PubMed: 16319382]

22. Johnston SC. Combining ecological and individual variables to reduce confounding by indication: case study–subarachnoid hemorrhage treatment. J Clin epidemiol 2000;53:1236–41. [PubMed: 11146270]

23. Brooks JM, Chrischilles EA, Scott SD, Chen-Hardee SS. Was breast conserving surgery underutilized for early stage breast cancer? Ine strumental variables evidence for stage II patients from Iowa. Health Serv Res 2003;38(6 Pt 1):1385–402. [PubMed: 14727779]

24. Wen SW, Kramer MS. Uses of ecologic studies in the assessment of intended treatment effects. J Clin epidemiol 1999;52:7–12. [PubMed: 9973068]

25. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating shorteterm drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology 2006;17:268–75. [PubMed: 16617275]

26. Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis. Arthritis Rheum 2006;54:3390–8. [PubMed: 17075817]

27. Schneeweiss S, Glynn RJ, Avorn J, Solomon DH. A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of firsttime pree scriptions for COX-2 inhibitors. J Clin epidemiol 2005;58:98–102. [PubMed: 15649677]

28. Solomon DH, Schneeweiss S, Glynn RJ, Levin R, Avorn J. Determinants of selective cyclooxygenase-2 inhibitor prescribing: are patient or physician characteristics more important? Am J Med 2003;115:715–20. [PubMed: 14693324]

29. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J Am Stat Assoc 1995;90:443–50.

30. Staiger D, Stock JH. Instrumental variable regression with weak instruments. Econometrica 1997;65:557–86.

31. Brookhart MA, Rassen J, Wang PS, Dormuth CA, Mogun H, Schneeweiss S. Evaluating the validity of an instrumental variable study of neuroleptics: can between-physician differences in prescribing patterns be used to estimate treatment effects? Med Care 2007;45(10 Suppl 2):S116–22. [PubMed: 17909369]

32. Rassen JA, Brookhart MA, Mittleman MA, Glynn RJ, Schneeweiss S. Instrumental variables II: in 25 variations, the physician prescribing preference generally was strong and reduced imbalance. J Clin epidemiol 2009;62:1235–43.

33. Wooldridge, JM. Introductory econometrics: a modern approach. 3rd edition. Thomson/South-Western; Mason, OH: 2006.

34. Greene, WH. Econometric analysis. 5th edition. Prentice Hall; Upper Saddle River, NJ: 2003.

35. Greenland S. An introduction to instrumental variables for epidemiologists. Int J epidemiol 2000;29:722–9. [PubMed: 10922351]

Table 1

Key points about IV analysis

An instrumental variable is a variable in nonexperimental data that can be thought to mimic the coin toss in a randomized trial.

If an appropriate and valid instrument is found, then the effects of measured and unmeasured confounding can be mitigated.

An IV analysis always has an experimental analog, however absurd the experiment sounds. The IV analysis is therefore based on "a natural experiment."

Assumption (1): The IV must predict treatment but that prediction does not have to be perfect. An IV that does a poor job of prediction is said to be weak.

Assumption (2): A valid IV will not be directly related to outcome, except through the effect of the treatment.

Assumption (3): A valid IV will also not be related to outcome through either measured or unmeasured paths.

In a randomized trial, the assumptions are met by design in the act of randomization. In an IV analysis, these assumptions must be empirically checked to the extent possible or assumed based on context and subject matter knowledge.

In cases of treatment-effect homogeneity, IV studies estimate the effect on the marginal subject, the average treatment effect for patients whose treatment was determined by the instrument [14].

Abbreviation: IV, instrumental variable.

Table 2a

Association between catheterization (*X*) and death (*D*) (crude exposure to outcome association: RD = 0.150)

	Catheterization (*X*+)	No catheterization (*X*−)	Total
Died (*D*+)	100	25	125
Did not die (*D*−)	400	475	875
Total	500	500	1,000

Abbreviation: RD, risk difference.

Table 2b

Association between closeness to catheterization facility (Z) and death (D) (instrument to outcome association assuming quasi-randomization: RD = −0.100)

	Small Diff. Dist. to Cath. facility (Z+)	Large Diff. Dist. to Cath. facility (Z−)	
Died ($D+$)	6	119	125
Did not die ($D-$)	144	731	875
Total	150	850	1,000

Abbreviation: Diff. Dist.: differential distance; Cath., catheterization.

Table 2c

Association between closeness to catheterization facility (Z) and catheterization (X) (strength of instrument and amount of compliance: RD = 0.494)

	Small Diff. Dist. to Cath. facility ($Z+$)	Large Diff. Dist. to Cath. facility ($Z-$)	
Catheterization ($X+$)	138	362	500
No catheterization ($X-$)	12	488	500
Total	150	850	1,000