Introduction to Instrumental Variables

OCTOBER 29, 2019

SCOTT QUINLAN, PHD

ASSISTANT TEACHING PROFESSOR

DEPARTMENT OF EPIDEMIOLOGY AND BIOSTATISTICS

Outline

The challenges of observational, real-world research

The limits of other options available

What is an instrumental variable?

The key requirements of an instrumental variable

Some examples

Limitations and opportunities

Smoking and birth weight

We want to examine the impact of smoking on birth weight.

Randomized trial → not ethically feasible

Observational study → could be done, but likely problematic

- Women who smoke are likely quite different from those who do not on a number of factors, many of which we cannot measure.

Does the high dose flu vaccine provide better protection than the low dose?

High dose has been recommended in past seasons for the elderly to provide greater protection against flu.

Randomized trials showed that the high dose was more effective than low dose at preventing hospital admissions.

- Does this apply in the real-world setting?

But…

Observational studies have been very mixed, some showing no benefit.

Why might we be concerned about these studies?

A conundrum

Selective COX-2 inhibitors were created to reduce gastrointestinal complications associated with other commonly used NSAIDs.

They were shown to be successful in randomized trials.

But…

In real-world clinical settings and observational studies the risk of GI complications has sometimes been higher in people prescribed selective COX-2 inhibitors than those prescribed other NSAIDs.

The challenge

Randomized trials are the **gold standard** for comparing two different therapies, interventions, surgeries, etc.

- But, they may not be practical or feasible in all settings.
- Results from a randomized trial may not always be easily applicable to the real-world.

Observational studies are an alternative, but **exposure selection process** can lead to bias.

- Those exposed (i.e. treated) are sometimes quite different from those not exposed.

What can we do here?

Other options

Propensity score (or similar) matching

Stratification

Adjustment during analysis

But, all are limited to what?

Those characteristics we can **measure**.

Instrumental variables

Randomization allows for all relevant information (both measured and not measured) to be balanced between groups.

- Not always feasible.

We are often not able to identify and measure all clinically relevant information that leads to some being exposed and others not.

- Imbalances likely still exist.

The purpose of instrumental variables is to be able to **identify quasi-random treatment choices**.

- Think of this as trying to make use of a **natural experiment** to mimic what might have happened if we had been able to actually randomize people.

What is an instrumental variable?

An instrumental variable is something that is strongly related to actual treatment/exposure status.

Often considered as a system of two equations:

- $Outcome = \alpha + \beta Exposure + Covariates + E$ — Plug in results from below to estimate β
- $Exposure = \gamma + \delta IV + Covariates + Error$ — Start here to estimate exposure using IV

- IV = proposed instrumental variable.
- Two stage least squares methods are commonly used (SAS and R both have procedures for this)

The devil is in the details

To be a good instrumental variable, three important assumptions must be met:

1) There has to be some correlation between the proposed instrumental variable and the exposure of interest (stronger the better)

2) The relationship between the instrumental variable and exposure of interest is not confounded by other factors.

- The instrumental variable should not be related to patient characteristics

3) The instrumental variable does not have an direct or indirect impact on the outcome of interest, except through the exposure of interest.

- **Most important assumption**!

Smoking and birth weight

A randomized trial was conducted amongst women who were smoking during pregnancy.

Some women were randomly assigned to participate in an intervention program to help encourage them to stop smoking.

Other women were randomly assigned to not participate in the intervention program.

The original random assignment (encouragement or not) was the instrumental variable.

High dose versus standard dose flu vaccine

Researchers were concerned about using actual vaccine received due to potential channeling of sickest people to high dose vaccine.

Used data from the Veterans Health Administration (VHA).

Facilities often have autonomy over influenza vaccination policy, including what vaccines to administer and to whom.

Patient characteristics at different VHA facilities are pretty similar.

Used facility preference for high dose versus standard dose as an instrumental variable.

Vaccine 37 (2019) 1484–1490

Selective COX-2 inhibitors

Researchers were concerned that physicians may selectively prescribe COX2 inhibitors to patients at higher risk of GI complications.

- Called *confounding by indication*.

Used Medicare data for people 65 years and older who were first prescribed either a selective COX-2 inhibitor or other NSAID.

Looked at risk of GI complications during follow-up.

Instrumental variable was prescribing physician preference for either COX2 or other NSAID.

Epidemiology 2006;17: 268–275

Do we have a strong instrument?

One of the first questions we need to ask is how well does our instrument predict our exposure?

A strong instrument is one that predicts exposure well.

How can we do this?

- F statistic and partial R^2 → A good rule of thumb is an F statistic of 10 or more is needed.
- Proportion compliant with instrumental variable.

$$Exposure = \gamma + \delta IV + Covariates + Error$$

The problem of weak instruments

Standard error for our estimates goes up.

Bias is likely, especially in small samples.

The impact of minor deviations of the other assumptions become magnified.

High dose versus standard dose flu vaccine

They used the F statistic for the prediction equation for actual vaccine received based on facility preference and other characteristics.

They found the **F statistic was >1,000**, indicating the facility preference was a useful predictor of actual vaccine receipt.

Test 1: Correlation

To test whether our instrument met the first requirement, we evaluated the F statistics of the first stage equations. The F statistic measures whether the instrument was sufficiently correlated with the endogenous variable (in this case, provision of HD). An F statistic greater than 10 is generally considered sufficient [22]. The F statistic for our sample was greater than 1,000, which easily satisfied this condition. We also performed a logistic regression with provision of HD as the outcome and HD proportion as the independent variable, in unit of 10%. We found that, for every increment of 10% of HD adoption, the odds ratio was 2 (2.06, 95% CI, 2.05-2.07, $p<0.0001$). In other words, every increment of 10% in a facility's HD proportion was associated with doubling the likelihood of a patient being provided with HD.

Selecting COX-2 inhibitors

Found that if a physician's last prescription was for a COX-2 inhibitor there was a **77% chance** their next prescription was for a COX-2 inhibitor.

Found that if a physician's last prescription was for a different NSAID, there was a **55% chance** the next prescription was for a COX-2 inhibitor.

Is our instrument related to patient characteristics?

The second key assumption is that our instrument is unrelated to patient characteristics.

We often test this by looking at patient characteristics across levels of the instrumental variable to see if we note any key differences.

It is also helpful to do the same for the actual treatment as a comparison.

Depending on sample size, p-values may not be as useful.

Example

TABLE 1. Characteristics of Patients Stratified on Type of NSAID Prescription Assigned at Index Date

Characteristics	COX-2 Inhibitor Users (n = 32,273) %	Nonselective NSAID Users (n = 17,646) %
Female	86	81
Age ≥75 at index date	75	65
Charlson comorbidity score ≥1	76	71
Hospitalized in prior year	31	26
Nursing home stay in prior year	8	6
History of warfarin use	13	7
History of oral glucocorticoids use	9	8
History of osteoarthritis	49	33
History of rheumatoid arthritis	5	3
History of peptic ulcer disease	4	2
History of gastrointestinal hemorrhage	2	1
History of hypertension	73	70
History of congestive heart failure	30	25
History of coronary artery disease	16	15
History of gastroprotective drug use	27	20
Five or more prescription drugs in prior year	75	67
Five or more doctor visits in prior year	72	64

TABLE 2. Associations Between Patient Risk Factors and Actual Treatment, the Instrumental Variable, and the Instrumental Variable in a Sample Restricted to Patients of Primary Care Physicians*

Characteristics	Actual Treatment (All Patients)†	Instrumental Variable (All Patients)‡	Instrumental Variable (PCPs Only)‡
Female	8.2	0.5	1.3
Age ≥75 at index date	11.0	1.3	0.9
Charlson comorbidity score ≥1	5.9	2.7	2.4
Hospitalized in prior year	5.0	1.4	0.8
Nursing home stay in prior year	9.1	2.9	1.8
History of warfarin use	15.8	4.0	3.8
History of oral glucocorticoids use	2.8	0.9	1.0
History of osteoarthritis	14.0	3.4	2.9
History of rheumatoid arthritis	13.1	5.1	5.2
History of peptic ulcer disease	9.4	1.4	0.3
History of gastrointestinal hemorrhage	9.3	1.0	−1.4
History of hypertension	3.0	1.3	1.6
History of congestive heart failure	6.5	1.4	0.8
History of coronary artery disease	2.9	1.1	1.2
History of gastroprotective drug use	8.5	0.4	0.2
Five or more prescription drugs in prior year	9.2	2.8	2.3
Five or more doctor visits in prior year	7.6	2.2	2.2

*The measures of association that we report are probability differences multiplied by 100.

†The probability of COX-2 inhibitor exposure among patients with risk factor minus the probability of this exposure among patients without the risk factor.

‡The probability that a patient's physician most recently prescribed a COX-2 inhibitor among patients with the risk factor minus the same probability among patients without the risk factor.

PCP indicates primary care physician.

Epidemiology 2006;17: 268–275

High dose versus standard dose flu vaccine

Table 2
Patient characteristics during study period by vaccine type.

Actual vaccine receipt

	HD	SD
Comorbidity		
Any malignancy	24,188 (15%)	441,466 (13%)
Metastatic solid tumor	1165 (1%)	19,448 (1%)
Congestive heart failure	13,538 (9%)	233,461 (7%)
Chronic pulmonary disease	30,026 (19%)	561,231 (16%)
Cerebrovascular disease	12,591 (8%)	225,186 (6%)
Dementia	3393 (2%)	50,679 (1%)
Diabetes with complications	12,460 (8%)	235,526 (7%)
Diabetes without complications	68,075 (43%)	1,393,512 (40%)
HIV/AIDS	945 (0.6%)	9394 (0.3%)
Mild liver disease	3442 (2.2%)	46,183 (1.3%)
Moderate/severe liver disease	375 (0.2%)	5607 (0.2%)
Myocardial infarction	2102 (1.3%)	43,722 (1.3%)
Hemiplegia/paraplegia	1131 (0.7%)	19,748 (0.6%)
Peptic ulcer disease	1114 (0.7%)	20,006 (0.6%)
Peripheral vascular disease	12,277 (8%)	233,269 (7%)
Rheumatoid disease	2815 (2%)	53,870 (2%)
Renal disease	16,510 (10%)	316,305 (9%)

Using facility preference

	Below Median (%)	Median+ (%)	SMD*
Patient Characteristics			
Male	98	98	0
Married	61	60	2
White	76	75	2
Admitted to a Nursing Home	2	1	8
Vaccinated in the Previous Season	80	77	7
Patient Conditions			
Chronic cardiac disease	28	27	2
Chronic pulmonary	13	13	0
Neurological/musculoskeletal	5	5	0
Other metabolic and immunity disorders	1	1	0
Diabetes mellitus	34	33	2
Liver diseases	1	1	0
Malignancies	14	14	0
Immunosuppressive disorders	4	5	-5
Chronic renal disease	6	6	0
Hemoglobinopathies	0.3	0.3	0
At least 3 risk factors	9	9	0

Is our instrument related to outcome?

Remember that a valid instrument can only be linked to our outcome through the exposure of interest.

How can we test this assumption?

Can only really be decided theoretically, but since this is the most important assumption we should really think carefully about it.

Are facilities different?

Facility-Level Comparison			
Complexity (1 most complex, 3 least)			P-value
1	73	80	
2	15	14	
3	12	7	0.213
Quality of Care (1 lowest, 5 highest)			
1	5	2	
2	16	16	
3	41	38	
4	24	32	
5	14	11	0.311
Region			
Midwest	30	32	
Northeast	20	19	
South	32	31	
West	17	18	0.96
Rurality			
Rural	27	29	
Urban	73	71	0.541

Smoking and birth weight

Remember, our instrument was the randomized assignment (either encouragement intervention or not).

Is randomized assignment related to smoking status?

- Yes (hopefully)

Is randomized assignment related to other confounders?

- Hopefully not, if randomization was done correctly.

Could randomized assignment impact our outcome (low birth weight)?

- Maybe

Common instruments and their limits

Distance to specialty care

Application: Impact of cardiac catheterization on survival post acute myocardial infarction (AMI). They wanted to see if catheterization improves outcomes compared to no catheterization.

The theory was that when someone has an AMI they would be taken to the nearest hospital and some hospitals are much more likely to perform a catheterization than others.

- Therefore distance was a strong predictor of whether someone would be catheterized or not

The concern: Distance could also be related to patient characteristics.

Common instruments and their limits

Provider preference

Application: Assumes that providers or groups have different preferences regarding treatment regiments, medications, or procedures.

Instead of actual treatment assignment, use provider preference as an instrumental variable.

How do we define preference?

How do we handle changes with time?

The concern: Patient characteristics cannot vary between physicians. Other differences in provider preference may also impact the outcome.

Common instruments and their limits

Day of the week

Application: For certain acute injuries that require prompt surgery (hip fracture) there was interest in seeing if waiting time between injury and surgery impacted outcomes.

Instead of actual waiting time (which could be impacted by injury severity), used day of the week (weekend or not) with the assumption that weekends would lead to longer wait times than week days.

The concern: Are patients different who are admitted on weekends than on weekdays? Other differences in hospital care between weekdays and weekends?

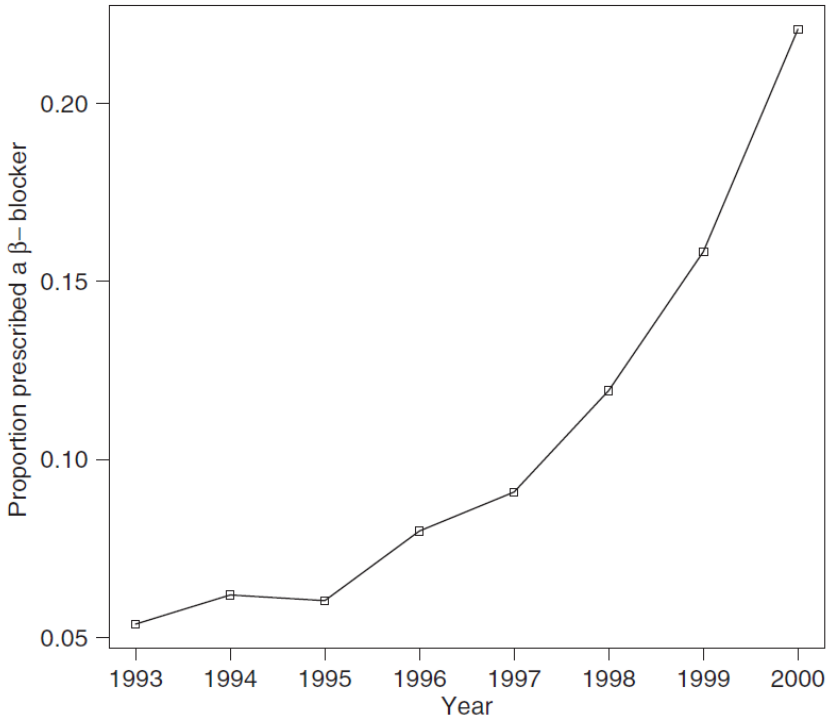
Common instruments and their limits

Calendar year

Application: Secular trends in medication use. This can result from changes in guidelines, formularies, preferences, safety information, etc.

Beta-blocker use after hospitalization for heart failure and impact on all-cause mortality.

The concern: Rarely used since other changes over time may also impact outcomes. Best used when there is a dramatic change in practice in a short period of time.



Statist. Med. 2008; **27**:1539–1556

What does our instrument actually measure?

Instrumental variable analyses are usually conducted as part of a 2 stage ordinary least squares process.

We want to know what impact our exposure has on our outcome, adjusting for the instrumental variable:

$$\hat{\beta}_{IV} = \frac{\widehat{\beta_{ols(IV \to Y)}}}{\widehat{\beta_{ols(IV \to X)}}}$$ Stage 2 / Stage 1

Can be simplified to:

$$\hat{\beta}_{IV} = \frac{P(Y = 1|Z = 1) - P(Y = 1|Z = 0)}{P(X = 1|Z = 1) - P(X = 1|Z = 0)}$$

"Intent to treat" estimate

Reflects the strength of our instrument.

Perfect instrument = 1

As weakens, approaches 0

Smoking and birth weight

$$\hat{\beta}_{IV} = \frac{Mean\ birthweight\ in\ encouragement\ group - Mean\ birthweight\ in\ control\ group}{\widehat{p_{enc}} - \widehat{p_{control}}}$$

Researchers found:

Mean birthweight in those randomized to encouragement group was 98 grams higher than those randomized to control group.

In the encouragement group, 57% still smoking

In the control group, 80% still smoking

$$\hat{\beta}_{IV} = \frac{98\ grams}{0.80 - 0.57} = \frac{98\ grams}{0.23} = 430\ grams$$

High dose versus standard dose flu vaccine

Table 3
Relative vaccine effectiveness of HD versus SD influenza vaccines using instrumental variable analysis.

Hospitalizations (Primary Diagnoses)	All-cause	Cardiorespiratory	Influenza/Pneumonia	Urinary Tract Infection
2010–2011	0.96 (0.91–1.01)	0.83 (0.76–0.91)	0.89 (0.78–1.02)	1.05 (0.85–1.30)
2011–2012	0.94 (0.90–0.98)	0.83 (0.77–0.90)	0.84 (0.67–1.05)	1.03 (0.85–1.25)
2012–2013	0.89 (0.85–0.94)	0.81 (0.74–0.90)	0.90 (0.79–1.03)	1.08 (0.86–1.36)
2013–2014	0.90 (0.85–0.95)	0.81 (0.74–0.89)	0.86 (0.66–1.13)	1.06 (0.86–1.32)
2014–2015	0.90 (0.87–0.93)	0.82 (0.77–0.88)	0.82 (0.70–0.96)	0.99 (0.84–1.17)
Longitudinal 5-season analysis	0.90 (0.88–0.92)	0.82 (0.79–0.85)	0.86 (0.78–0.94)	1.05 (0.82–1.34)

Selective COX-2 inhibitors

TABLE 4. Instrumental Variable and Conventional Multivariable Regression Estimates of the Risk Differences of Gastrointestinal Toxicity per 100 Patients Treated With COX-2 Inhibitors Compared With Nonselective NSAIDs

	Estimated Risk Difference per 100 Patients (95% CI)*			
	Conventional Unadjusted	Conventional Adjusted†	Instrumental Variable Unadjusted	Instrumental Variable Adjusted†
GI event within 60 d				
All patients	0.03 (−0.12 to 0.18)	−0.04 (−0.20 to 0.10)	−0.92 (−1.74 to −0.10)	−1.02 (−1.88 to −0.16)
Patients treated by PCPs	0.11 (−0.05 to 0.28)	0.03 (−0.14 to 0.20)	−0.75 (−1.73 to 0.23)	−0.81 (−1.84 to 0.22)
Patients with OA or RA	0.10 (−0.13 to 0.33)	0.07 (−0.17 to 0.30)	−1.80 (−3.31 to −0.29)	−1.81 (−3.34 to −0.28)
GI event within 120 d				
All patients	0.09 (−0.10 to 0.29)	−0.06 (−0.26 to 0.14)	−1.15 (−2.20 to −0.09)	−1.31 (−2.42 to −0.20)
Patients treated by PCPs	0.03 (−0.20 to 0.26)	−0.13 (−0.37 to 0.11)	−0.93 (−2.24 to 0.39)	−1.04 (−2.41 to 0.34)
Patients with OA or RA	0.14 (−0.17 to 0.45)	0.03 (−0.28 to 0.35)	−2.06 (−3.99 to −0.13)	−2.05 (−4.00 to −0.09)
GI event within 180 d				
All patients	0.19 (−0.02 to 0.41)	−0.03 (−0.26 to 0.19)	−0.94 (−2.14 to 0.25)	−1.21 (−2.46 to 0.04)
Patients treated by PCPs	0.09 (−0.17 to 0.35)	−0.15 (−0.42 to 0.12)	−0.61 (−2.12 to 0.89)	−0.82 (−2.40 to 0.75)
Patients with OA or RA	0.24 (−0.12 to 0.60)	0.07 (−0.30 to 0.43)	−1.45 (−3.65 to 0.75)	−1.52 (−3.74 to 0.71)

*All confidence limits were estimated robustly to account for within-physician correlation of outcomes.

†Adjusted for age, sex, Charlson comorbidity score, calendar year, hospitalization in previous year, number of doctor visits within previous year, history in the last year of: warfarin use, glucocorticoid use, gastroprotective drug use, congestive heart failure, osteoarthritis, rheumatoid arthritis, coronary artery disease, hypertension, GI hemorrhage, and peptic ulcer disease.

Sample sizes for the instrumental variable estimates are smaller because the instrument is undefined for the first NSAID prescription written by each physician during study period.

Conclusions

Instrumental variables offer a unique approach to controlling potential confounders both known and unknown.

The concept is relatively simple to implement and has shown success in some areas.

Finding a strong instrumental variable is tough and relies on many assumptions that may or may not be testable.

Can be considered as a sensitivity analysis or secondary analysis to more traditional observational methods.

Learn more

Brookhart MA, et al. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiology and Drug Safety* 2010; 19: 537-554

Ertefaie A, et al. A tutorial on the use of instrumental variables in pharmacoepidemiology. *Pharmacoepidemiology and Drug Safety* 2017;

Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology* 2000; 29: 722-729

Thank you!

SQUINLAN@GWU.EDU