CLEAR Center-PM&R Journal Methods Webinar Series

Methods for Dealing with Confounding in Observational Studies: Instrumental Variables

Kristin Sainani, PhD

April 26, 2019

Introduction to Instrumental Variables

- Key advantage: Unlike stratification, matching, regression, and propensity scores, instrumental variable analysis addresses *unmeasured and residual confounding*.
- Instrumental variable analysis exploits "natural experiments."

Example: Alcohol and CVD

- We've all heard that "moderate drinking is good for the heart."
- But is the association causal or spurious?

Residual and unmeasured confounding

- Residual confounding arises because SES is often crudely measured.
- Unmeasured confounding arises because we haven't accounted for everything (e.g., what about the ability to practice moderation?)

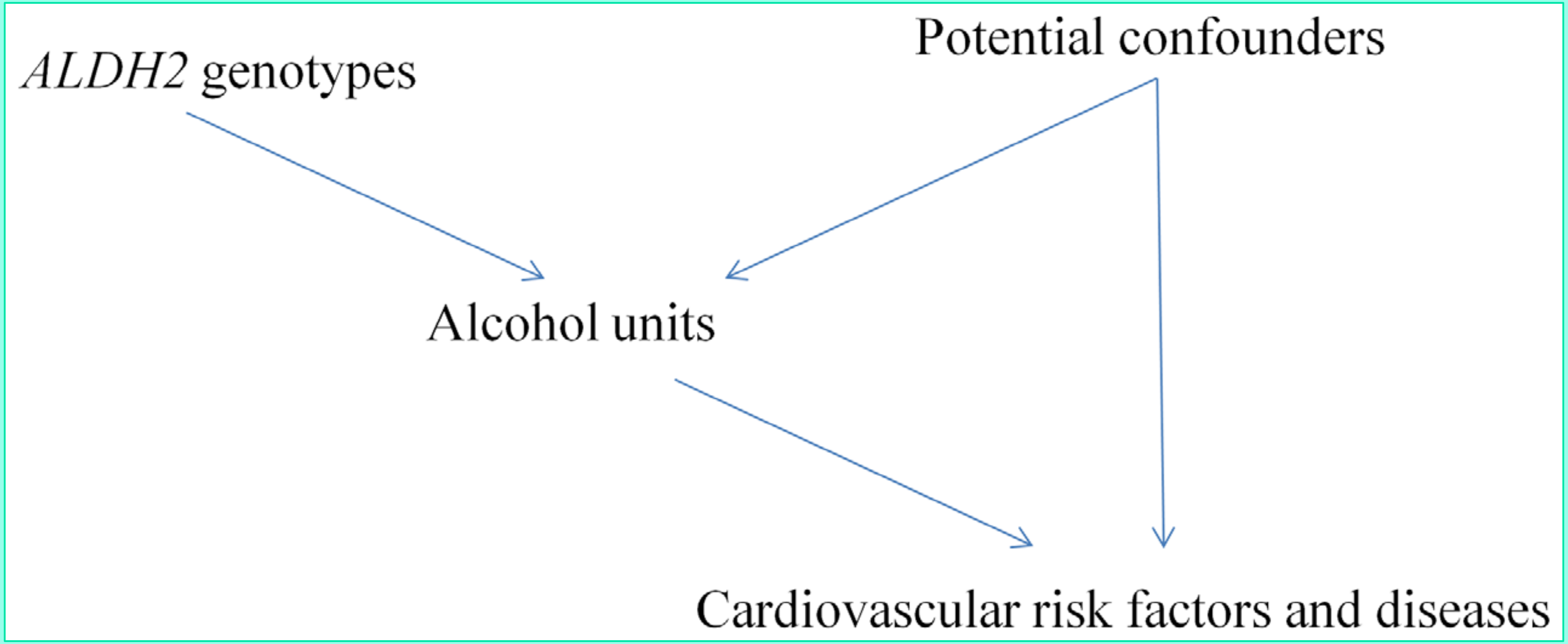
Solution: Instrumental Variable

- An instrumental variable is a naturally occurring phenomenon that imperfectly randomizes people to an exposure or treatment.
- For example, some people carry a gene that makes alcohol consumption unpleasant. Carrier status is randomly assigned at birth and partially determines one's alcohol exposure.
- Instrumental variable analysis focuses solely on the variation in alcohol exposure that is determined by this gene to estimate an unconfounded effect of alcohol on cardiovascular disease risk.

Assumptions of IVs

1. The IV must be related to the exposure or treatment.
2. The IV must be unrelated to confounders (at least after adjusting for measured confounders).
3. The IV must have no direct effect on the outcome except through its effect on exposure/treatment.

Figure 1. Directed acyclic graph showing the framework of Mendelian randomization analyses in this study.



Au Yeung SL, Jiang C, Cheng KK, Cowling BJ, Liu B, et al. (2013) Moderate Alcohol Use and Cardiovascular Disease from Mendelian Randomization. PLOS ONE 8(7): e68054. https://doi.org/10.1371/journal.pone.0068054
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0068054

Table 1. Alcohol consumption and socio-demographic characteristics by ALDH2 genotype among men from the Guangzhou Biobank Cohort Study (2003–8).

		ALDH2 genotype (from rs671)			
		Two inactive alleles (AA)	**One inactive allele (AG/GA)**	**No inactive alleles (GG)**	**§P value**
Alcohol units (10g	n	416	2,023	2,428	
ethanol) per day	mean (SD)	0.09 (0.79)	0.24 (1.22)	0.90 (2.52)	$<$0.001
Age group (%) years	n, years	417	2,053	2,457	
	50–54	11.0	10.0	9.2	0.41
	55–59	20.9	20.9	21.2	
	60–64	25.9	23.9	26.3	
	65–69	19.7	23.8	23.4	
	70–74	16.3	15.7	14.6	
	75–79	5.5	4.2	3.7	
	80+	0.7	1.5	1.4	
Education (%)	n	417	2,051	2,455	
	Less than primary	2.6	2.3	2.3	0.63
	Primary	24.7	27.3	26.2	
	Junior middle	29.0	30.3	31.1	
	Senior middle	27.1	25.1	23.5	
	Junior college	10.3	8.5	9.2	
	College	6.2	6.5	7.7	
Smoking status (%)	n	416	2,045	2,444	
	Never	41.1	40.4	40.1	0.88
	Former	29.3	27.7	27.8	
	Current	29.6	31.9	32.2	
Physical activity	n	417	2,053	2,457	
(IPAQ) (%)	Inactive	9.1	8.5	8.1	0.23
	Minimally active	36.9	38.8	41.6	
	†HEPA active	54.0	52.7	50.3	
Antihypertensive	n	416	2,045	2,451	
drugs (%)	Current user	19.5	18.7	20.2	0.49
Lipid modifying	n	417	2,052	2,453	
drugs (%)	Current user	5.5	5.4	6.3	0.44
drugs (%)	Current user	6.2	6.2	6.6	0.87
Systolic blood	n	416	2,046	2,449	
pressure (mmHg)	mean (SD)	131.2 (19.3)	132.7 (21.1)	133.0 (21.7)	0.31
Diastolic blood	n	415	2,046	2,446	
pressure (mmHg)	mean (SD)	75.3 (10.4)	75.8 (10.9)	76.5 (11.4)	0.05
Body Mass	n	416	2,048	2,448	
Index (kg/m^2)	mean (SD)	23.5 (3.0)	23.5 (3.1)	23.5 (3.2)	0.84

[§]P-value from ANOVA for continuous variables and from a χ^2 test for categorical variables, 2 sided.
[†]HEPA: health-enhancing physical activity (i.e., vigorous activity at least 3 days a week achieving at least 1,500 metabolic equivalent (MET) minutes per week or activity on 7 days of the week, achieving at least 3,000 MET minutes per week; IPAQ: International Physical Activity Questionnaire.
doi:10.1371/journal.pone.0068054.t001

Au Yeung SL, Jiang C, Cheng KK, Cowling BJ, Liu B, et al. (2013) Moderate Alcohol Use and Cardiovascular Disease from Mendelian Randomization. PLOS ONE 8(7): e68054. https://doi.org/10.1371/journal.pone.0068054
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0068054

Table 1. Alcohol consumption and socio-demographic characteristics by ALDH2 genotype among men from the Guangzhou Biobank Cohort Study (2003–8).

		ALDH2 genotype (from rs671)			
		Two inactive alleles (AA)	One inactive allele (AG/GA)	No inactive alleles (GG)	§P value
Alcohol units (10g	n	416	2,023	2,428	
ethanol) per day	mean (SD)	0.09 (0.79)	0.24 (1.22)	0.90 (2.52)	<0.001

.09 versus .24 versus 0.90 alcohol units per day—big difference!

Au Yeung SL, Jiang C, Cheng KK, Cowling BJ, Liu B, et al. (2013) Moderate Alcohol Use and Cardiovascular Disease from Mendelian Randomization. PLOS ONE 8(7): e68054. https://doi.org/10.1371/journal.pone.0068054
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0068054

PLOS | ONE
TENTH ANNIVERSARY

		ALDH2 genotype (from rs671)			
		Two inactive alleles (AA)	One inactive allele (AG/GA)	No inactive alleles (GG)	§P value
Education (%)	n	417	2,051	2,455	
	Less than primary	2.6	2.3	2.3	0.63
	Primary	24.7	27.3	26.2	
	Junior middle	29.0	30.3	31.1	
	Senior middle	27.1	25.1	23.5	
	Junior college	10.3	8.5	9.2	
	College	6.2	6.5	7.7	
Smoking status (%)	n	416	2,045	2,444	
	Never	41.1	40.4	40.1	0.88
	Former	29.3	27.7	27.8	
	Current	29.6	31.9	32.2	
Physical activity	n	417	2,053	2,457	
(IPAQ) (%)	Inactive	9.1	8.5	8.1	0.23
	Minimally active	36.9	38.8	41.6	
	†HEPA active	54.0	52.7	50.3	
Body Mass	n	416	2,048	2,448	
Index (kg/m^2)	mean (SD)	23.5 (3.0)	23.5 (3.1)	23.5 (3.2)	0.84

Au Yeung SL, Jiang C, Cheng KK, Cowling BJ, Liu B, et al. (2013) Moderate Alcohol Use and Cardiovascular Disease from Mendelian Randomization. PLOS ONE 8(7): e68054. https://doi.org/10.1371/journal.pone.0068054
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0068054

PLOS ONE
TENTH ANNIVERSARY

Results of numerous IV studies on this topic:

- “These data show that individuals of European descent with a genetic predisposition to consume less alcohol had a reduced risk of coronary heart disease and ischaemic stroke, and lower levels of several established and emerging risk factors for cardiovascular disease.”
- “Our results therefore challenge the concept of a cardioprotective effect associated with light to moderate alcohol consumption reported in observational studies and suggest that this effect may have been due to residual confounding or selection bias.”

Michael V Holmes et al. *BMJ* 2014;349:bmj.g4164

Commonly used IVs

- Genotype (“Mendelian randomization”)
- Differential distance from specialty care
- Policy change
- Physician or institution preference
- Prescribing trends over time
- Treatment assignment in an RCT with non-compliance

Example: Policy change

Smoking and pregnancy:

- **Exposure:** smoking during pregnancy
- **Outcome:** low birth weight
- **IV:** large cigarette tax hikes that occurred in four states. Rates of smoking in pregnant women dropped after the tax hikes in these states.

Lien DS, Evans WN. Estimating the impact of large cigarette tax hikes. *The Journal of Human Resources*. 2005;15:373–92.

Example: Prescribing change

Hormone therapy and stroke/heart attack risk:

- **Exposure:** hormone replacement therapy
- **Outcome:** stroke and heart attack
- **IV:** calendar time. Use of hormone replacement therapy in postmenopausal women was widespread before 2002, but dropped sharply in 2002 due to the results of the Women's Health Initiative randomized trial.

Shetty KD, Vogt WB, Bhattacharya J. Hormone replacement therapy and cardiovascular health in the United States. *Med Care* 2009; 47(5): 600–606. *Med Care*. 2009;47:600-6.

Example: Proximity to specialized care

Admissions to a dedicated stroke center and stroke mortality:

- **Exposure:** admission to a dedicated stroke center
- **Outcome:** stroke mortality
- **IV:** differential distance to a stroke center: the distance from a patient's residence to the nearest stroke center minus the distance from a patient's residence to the nearest hospital of any kind.

Xian Y, Holloway RG, Chan PS, et al. Association Between Stroke Center Hospitalization for Acute Ischemic Stroke and Mortality. *JAMA.* 2011;305(4):373–380. doi:10.1001/jama.2011.22.

Example: RCT with noncompliance

RCT of integrated care (the Children's Treatment Network) versus usual care for children with special needs:

- **Intervention**: integrated care delivered through the Children's Treatment Network versus usual care
- **Outcome**: psychosocial quality of life score
- **IV**: randomization assignment in the trial. Only 48% of those assigned to integrated care were compliant. (All control patients received usual care.)

Chenglin Y, Gina B, Joseph B, Lehana T.A sensitivity analysis of the Children's Treatment Network trial: a randomized controlled trial of integrated services versus usual care for children with special health care needs. *Clin Epidemiol* 2013; 5: 373–385.

How do we estimate effects?

- Simple estimate (binary IV, no confounders)
- Two-stage regression

Complier class

- **Complier:** someone whose treatment/exposure level depends on the instrument. E.g., would take the treatment if assigned to the treatment group but would take the control if assigned to the control group
- **Noncomplier:** someone whose treatment/exposure level does not depend on the instrument. E.g., someone who would always take the treatment, even if assigned to control ("always taker") or someone who would never take the treatment ("never taker").
- **Defier:** someone whose treatment exposure level is affected by the instrument, but in the opposite direction than expected. E.e., someone who would take the treatment when assigned to the control group and would take the control when assigned to the treatment group. **We are going to assume that defiers don't exist." (the "no defiers" assumption).**

Complier class is often unobservable. If a control patient takes control, is this because they are a complier or a never taker?

Complier class example

Take recent legalization of marijuana in California

- **Complier:** someone who would smoke if it was legal, but not if it was illegal (someone who changes their behavior due to the legislation)
- **Noncomplier:** someone who would always smoke even if it was illegal (“always takers”) or someone who would never smoke even if was legal (“never taker”)
- **Defier:** someone who would choose to smoke only if it was illegal

Simple IV estimate

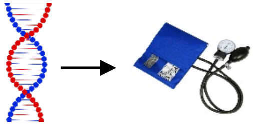
Effect of the exposure on the outcome=

Unconfounded effect

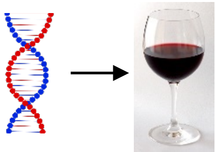
$$\frac{\text{Effect of the instrument on the outco}}{\text{Effect of the instrument on the exposu}}$$

Rescaling to units of the exposure rather than units of the instrument (accounts for the amount of variation in the exposure that is due to the instrument)

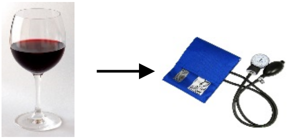
Simple IV estimate:



Effect of genotype on BP = -1 mmHg



Effect of genotype on alcohol consumption = -0.81 standard drinks/day



☐Effect of alcohol consumption on BP=

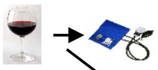
$$\frac{\text{-1 mmHg}}{-0.81\ \text{standard drinks/day}}$$ =1.2 mmHg per standard drink/day

Derivation of IV estimate:

- For simplicity, we will assume that the population has just compliers and non-compliers (binary).
- For the ALDH2 genotype study:
  - **Complier** would drink alcohol if they had the normal genotype but would avoid alcohol if they had the mutant genotype.
  - **Noncomplier** would drink exactly the same amount of alcohol whether they had the mutant genotype or not.
- IV analysis assumes no defiers.

Derivation of IV estimate:

 Effect of genotype on blood pressure = average BP in ALDH2-/-group – average BP in non-carriers



average BP in ALDH2-/-group – average BP in non-carriers =

(Effect of alcohol on blood pressure in **compliers**) x (difference in average alcohol intake between ALDH2-/- compliers and non-carrier compliers) x (proportion who are compliers)

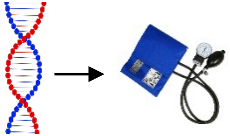
+

~~(Effect of alcohol on blood pressure in **non-compliers**) x (difference in average alcohol intake between ALDH2-/- non-compliers and non-carrier non-compliers) x (proportion who are non-compliers)~~

+

~~Average effect of confounders on blood pressure~~

Derivation of IV estimate:

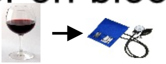


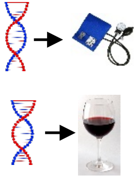
Effect of genotype on blood pressure =

(Effect of alcohol on blood pressure in **compliers**) x (difference in average alcohol intake between ALDH2-/- compliers and non-carrier compliers) x (proportion who are compliers)

= Effect of alcohol on blood pressure in compliers x
(difference in average alcohol intake in the ALDH2-/- group and the non-carrier group)

$$\therefore \text{Effect of alcohol on blood pressure in compliers} = \frac{\text{Effect of genotype on blood pressure}}{\text{Effect of genotype on alcohol intake}}$$





Since estimate is based on “compliers” only:

1. Estimate may not be generalizable to noncompliers.
2. The effective sample size is reduced.
   - Precision is decreased!

Table 2. Mendelian randomization estimates, obtained from instrumental variable analysis using 2SLS and probit regression, and multivariable linear and probit regression estimates of the association of alcohol use (1 unit) with CVD risk factors and morbidity.

	Mendelian randomization Instrumental variable analysis				†Observational Multivariable regression			
	n	§β	95% CI	p value	n	§β	95% CI	p value
Systolic blood pressure (mmHg)	4,853	1.00	−0.74 to 2.74	0.26	4,847	0.84	0.55 to 1.12	<0.001
Diastolic blood pressure (mmHg)	4,849	1.15	0.23 to 2.07	0.01	4,843	0.49	0.34 to 0.65	<0.001

95% CI
−0.74 to 2.74
0.23 to 2.07

IV Analysis

95% CI
0.55 to 1.12
0.34 to 0.65

Regression

Au Yeung SL, Jiang C, Cheng KK, Cowling BJ, Liu B, et al. (2013) Moderate Alcohol Use and Cardiovascular Disease from Mendelian Randomization. PLOS ONE 8(7): e68054. https://doi.org/10.1371/journal.pone.0068054
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0068054

PLOS ONE
TENTH ANNIVERSARY

Simple IV estimate, integrated care example:

Effect of treatment assignment on psychosocial quality of life (= intention-to-treat estimate from the RCT!) = + 1.5 points

Effect of treatment assignment on receipt of treatment (increase in proportion receiving treatment) = 48% - 0% = 48%

$\therefore$ Effect of treatment on psychosocial quality of life $= \dfrac{1.5}{0.48} = +3.1$

ITT vs. IV estimates

- ITT estimate: +1.5 (-1.5 to 4.5), p=.32
- IV estimate: +3.1 (-3.1 to 9.3), p=.33

Relative increase in effect size = 3.1/1.5=2.07

Relative increase in standard error = 3.1/1.5 =2.07

This means p-values will be nearly identical

- In limited simulations I've run, I've found this to be generally true→ the increase in effect size is very close to the increase in standard error, leading to little difference in the statistical inference.

Two-stage regression

Model 1: Exposure/treatment = instrument + confounders

Model 2: Outcome = predicted exposure/treatment (from model 1) + confounders

Two-stage regression

Model 1: alcohol units = genotype

Model 2: BP = predicted alcohol units (from model 1)

Regress BP on genotype: BP = $\alpha + \beta*(0 \text{ if ALDH2-/-}, 1 \text{ if non-carrier}) \rightarrow \beta=+1.0$mmHg
Regress BP on predicted alcohol units: BP = $\alpha + \beta*(0.09 \text{ if ALDH2-/-}, 0.90 \text{ if non-carrier}) \rightarrow$
$\beta=1.0$mmHg/0.81 =1.2 mmHg

Two-stage regression

Model 1: Exposure/treatment = instrument + **confounders**

Model 2: Outcome = predicted exposure/treatment (from model 1) + **confounders**

**Must use two-stage regression software to do this, or you will get the wrong standard errors!

Assumptions to check

- 1. Do I have a strong enough instrument?
  - Weak instruments are imprecise (huge standard errors)
  - Weak instruments are highly sensitive to violations of assumptions

1. Do I have a strong enough instrument?

- Commonly used criterion: F-statistic > 10 from regression of instrument on exposure (model 1 of the two-stage regression)
- Example of a strong instrument:
  - ALDH2 gene
  - 80% of stroke patients who lived closer to a stroke center than any other hospital went to a stroke center versus 25% of stroke patients who lived farther from a stroke center

Assumptions to check

- 2. Is the instrument truly a good randomizer?**
  - Could it be related to unmeasured confounders?
  - Could it be related directly to the outcome?

2. Is the instrument truly a good randomizer?

- Researchers can check for balance in measured confounders empirically, but can only argue for balance in unmeasured confounders on theoretical grounds.
- To check for balance in measured confounders, look for standardized differences <10%
  - For the stroke study, the standardized differences in ages and comorbidities between patients with differential distance = 0 and patients with differential distance >0 were all less than 10%. The researchers did find differences in race and rural status, which they adjusted for in their analyses.
- Perform sensitivity analyses to gauge the potential impact of unmeasured confounders on the results

2. Is the instrument truly a good randomizer?

- Researchers also need to think carefully about whether the instrument could have a direct impact on the outcome, and should perform sensitivity analyses to address this.

Assumptions to check

- 3. Could there be “defiers”?

Advantages of IV analysis

- Addresses unmeasured and residual confounding
- Exploits natural experiments
- Gives an alternative to ITT estimates for randomized trials with noncompliance

Disadvantages of IV analysis

- A good instrument doesn’t always exist
- Relies on assumptions that may be hard to test