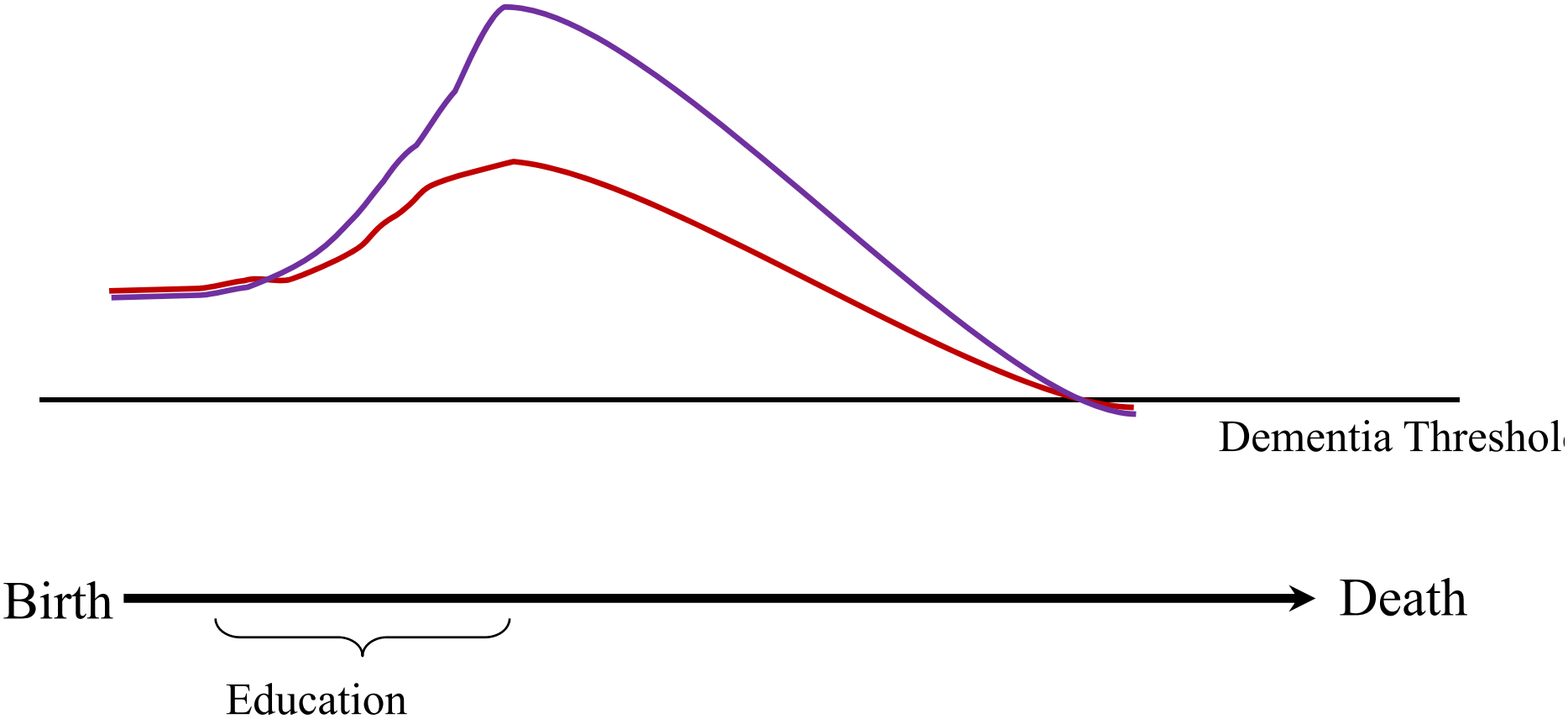
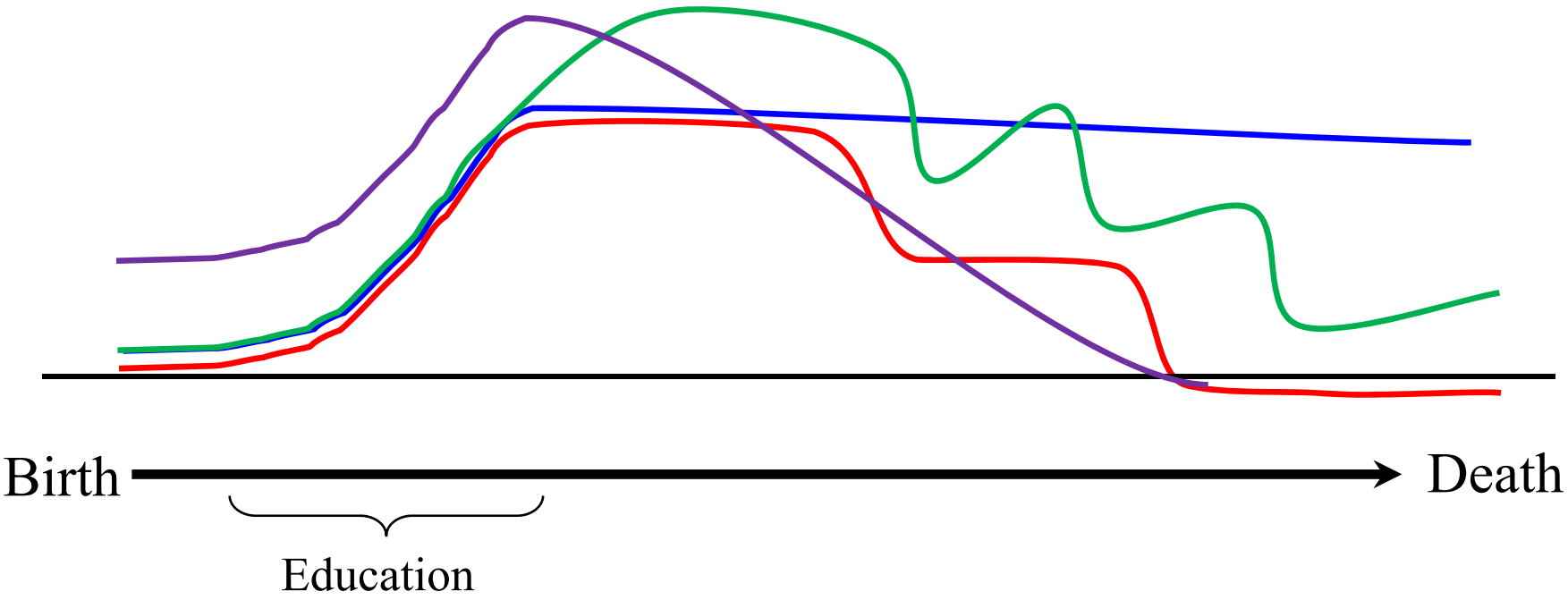
Choosing an analysis method

- We have focused on continuous outcomes: blood pressure, cognitive function, depressive symptoms
  - For repeated measures, use mixed models or a GEE framework focusing on rate of change
  - Key hypothesis test: does risk factors modify rate of change?
- Often the outcome is incidence of a binary condition: hypertension, dementia, depression
  - Usually, restrict to people without the condition at baseline and model time to event, e.g., with a Cox model, an Aalen additive hazards model, or a discrete time survival mode.
  - Key hypothesis test: does risk factor increase reference category incidence rate?
- Focusing on a binary outcome can introduce serious problems with identifying *current* risk factors
- Note some binary outcomes can occur more than once
- And there are categorical outcomes and variations on regression models can be applied to model transitions between states

Comparing possible cognitive trajectories: focusing on time to dementia is misleading



Some possible trajectories



Variability in cognitive test scores due to age-related decline

<<

Variability due to individual differences in functioning of people at the same age

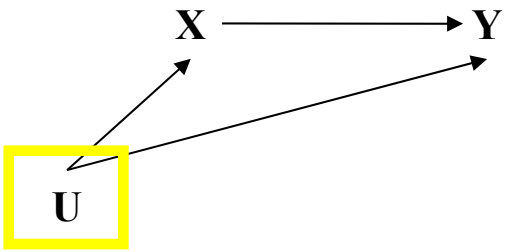
Negative ICC?

- From: http://faculty.umb.edu/peter_taylor/09b.pdf
- ICC refers to a number of quantities, but the simplest form is the usual linear (Pearson product-moment) correlation among a set of pairs of values when the order in each pair is arbitrary. That would be the case if we wanted to know the correlation of heights in same sex couples or the agreement of two independent ratings of an exam done by a set of students (where the two raters differed from one student to the next) -
- Harris (1913; as described in Haggard 1958) showed that it is equivalent to finding the ratio of the variance of the means of the classes to the variance of the entire set of values, which is easier to calculate
- As a ratio of variances, the ICC should be in the interval [0,1]. Yet, without comment, negative estimates are included in Field's (2005) encyclopedia entry.
- In fact, as best I can tell, the ICC calculated with the variance component formula is identical to the rho calculated as the correlation between individuals in a cluster *unless* the rho is negative, in which case the ICC goes to null

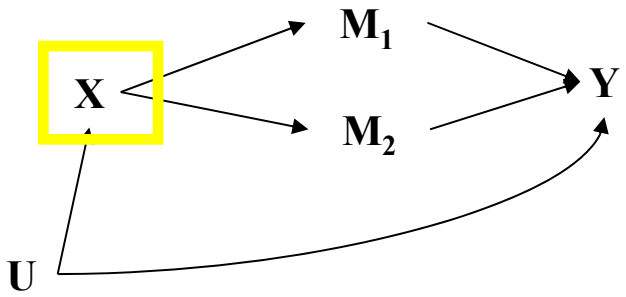
Instrumental Variables: Randomization andPseudoRandomization

Demonstrating Causation

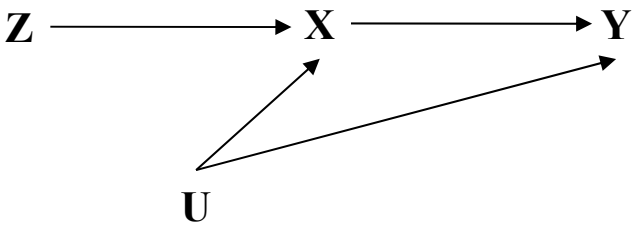
- Condition on all CPCs



- Measure all pathways



- Use an instrument



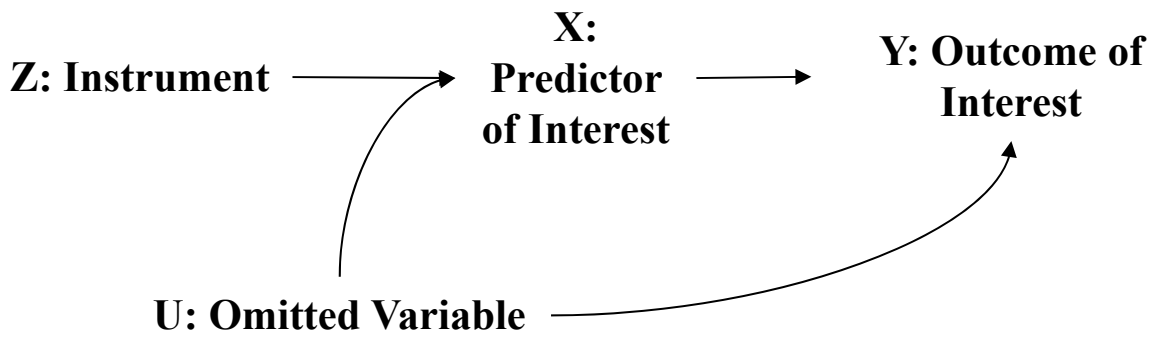
Motivations for IV Analyses

- Avoid confounding if a plausible natural experiment is available
- Estimate the effect of *receiving* treatment (IV) instead of being *assigned* treatment (ITT)
- Estimate the effect of treatment on the "marginal patient" or the compliers

Instrument Assumptions

Z is an instrument for the influence of X on Y if:

- Z is not an effect of X
- Z and X are associated
- every unblocked path between Z and Y has an arrow into X



Variations on the Definition

- Standard econ:
  - If you are interested in estimating $Y_i=\beta X_i+e_i$ then Z is an instrument if "z_i is correlated with x_i but not with ε_i" (Greene 2000, pg 370)
- Counterfactual:
  - Z is independent of Y_x (the counterfactual value of Y under any given value of X), and
  - Z is not independent of X. (Pearl, 2000, pg 248)
- AIR 1994 (developed originally for dichotomous Z, X):
  - SUTVA
  - Random assignment of Z
  - Exclusion: $Y(Z,X)=Y(Z',X)$ for all X, Z and Z'
  - Z has to have some effect on the probability of X
  - Montonicity: $X(Z=1)\geq X(Z=0)$ for everyone in the population
- DAGs: No paths not mediated by X

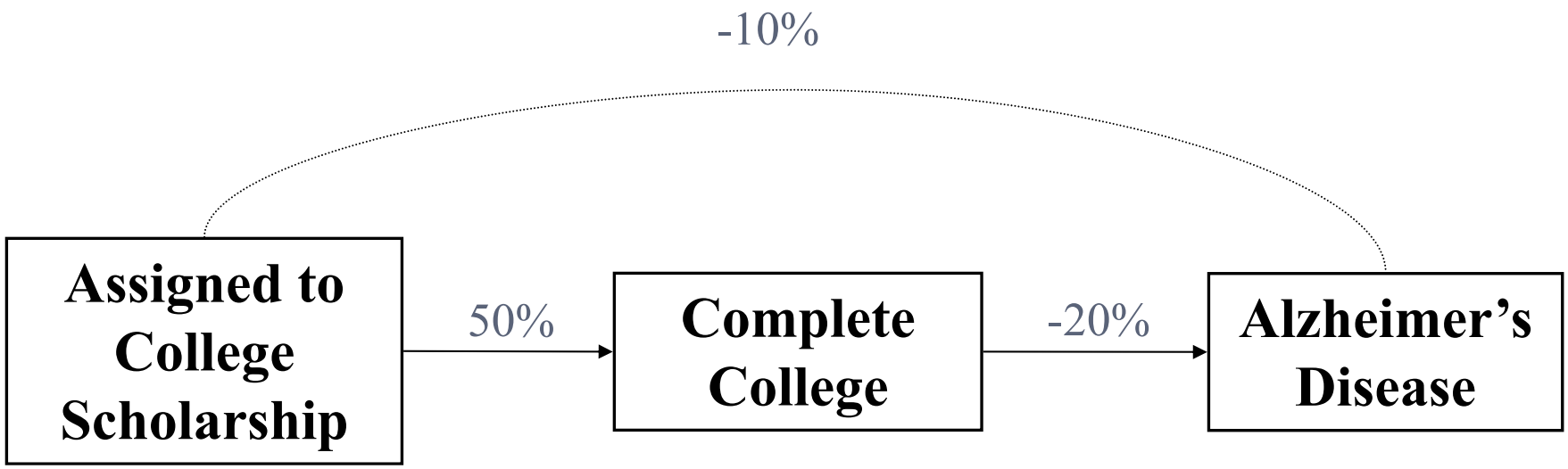
Familiar Instruments

- Instruments arise in the context of experiments
  - Randomized controlled trials
  - Natural experiments
  - Mendelian randomization
- Intent to treat analyses estimate the relation between treatment *assignment* and the outcome
- IV analyses estimate the relation between treatment and the outcome
  - Similar to correction for non-compliance
  - Estimates the effect of treatment on those who were treated because of their assignment
  - NOT a comparison of the treated to the untreated

Why are natural experiments important to epidemiology?

- Epidemiologists have long considered RCTs the "gold standard" for causal inference.
- Recent results call into question the reliability of causal inference from observational studies.
- Some exposures cannot be tested in RCTs.
- We must seek all approaches to bolster observational evidence.

IV Effect Estimates



Obtaining IV Estimates

Wald Estimates:

Effect = $\frac{\text{(Difference in Avg Outcome Score between Instr Groups)}}{\text{(Difference in Avg Exposure Level between Instr Groups)}}$

Two Stage Least Squares Estimates:

$\hat{e}xposure = \beta_0 + \beta_1 Instrument + \beta_k Other\ Covariates$

$Outcome = \beta_0 + \beta_1 \hat{e}xposure + \beta_k Other\ Covariates$

Generalized Method of Moments: Use assumption that residuals are independent of exposure

Obtaining IV Estimates

Wald Estimates:

Effect = $\frac{\text{(Difference in Avg Outcome Score between Instr Groups)}}{\text{(Difference in Avg Exposure Level between Instr Groups)}}$

Two Stage Least Squares Estimates:

$\hat{e}xposure = \beta_0 + \beta_1 \text{Instrument} + \beta_k \text{Other Covariates}$

$\text{Outcome} = \beta_0 + \beta_1 \hat{e}xposure + \beta_k \text{Other Covariates}$

Generalized Method of Moments: Use assumption that residuals are independent of exposure

Obtaining IV Estimates

Wald Estimates:

$$\text{Effect} = \frac{\text{(Difference in Avg Outcome Score between Instr Groups)}}{\text{(Difference in Avg Exposure Level between Instr Groups)}}$$

Note that the Wald Estimate is essentially the Intent to Treat effect estimate in the numerator Divided by the effect of randomization on adherence in the denominator

Both ITT and adherence can be estimated without bias in an RCT, so you can use these two unbiased estimates to derive the estimated effect of exposure on outcome.

Variations on IV Estimators

Wald Estimate: = (Difference in Avg Outcome between IV values) / (Difference in Avg Treatment between IV values)

Two Stage Least Squares (2SLS):

- Multiple instruments
- Control for covariates

Separate Sample IV:

- 1st and 2nd stage of a 2SLS are from different data sets
- Often relevant in Mendelian Randomization

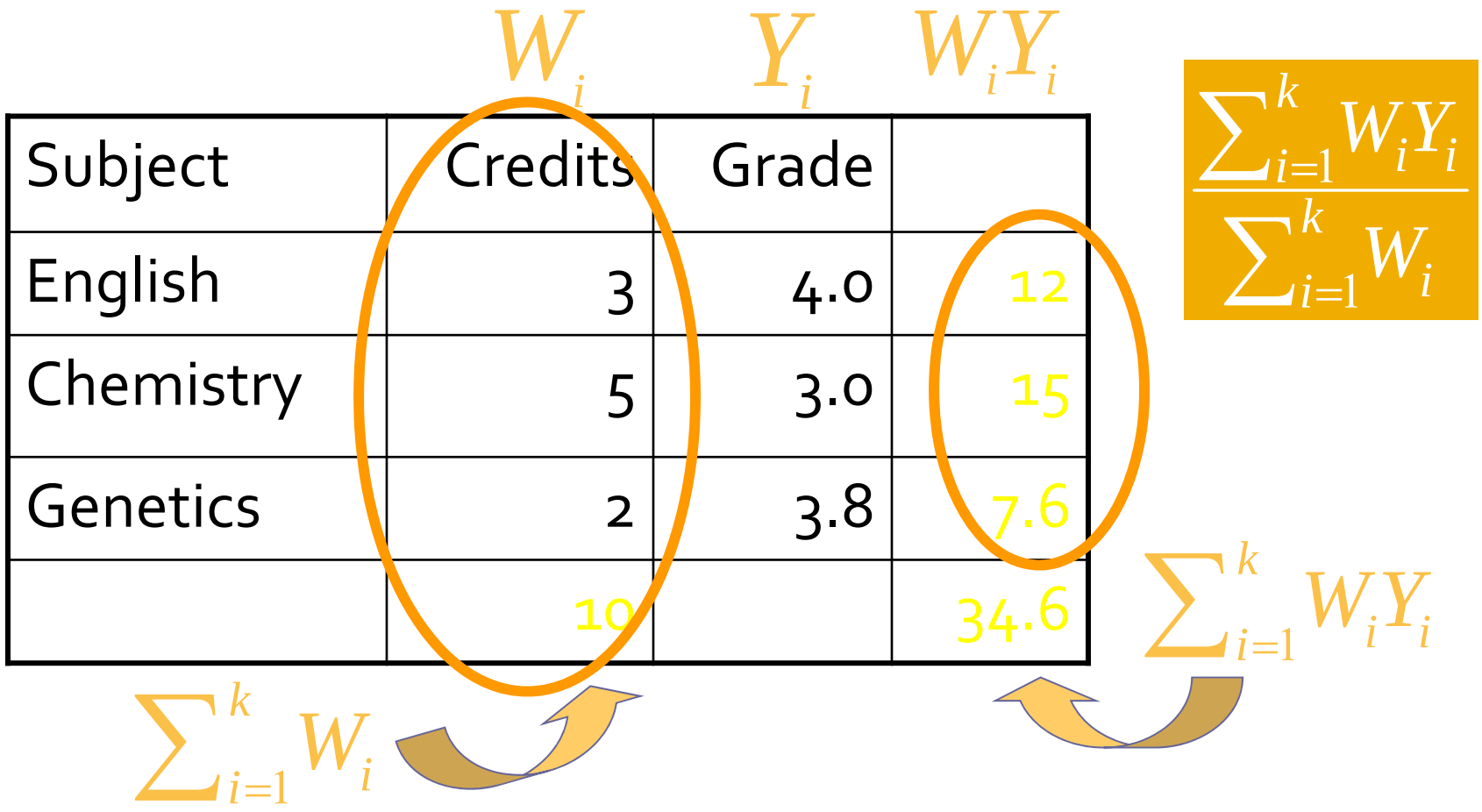
Generalized method of moments

Residual control approaches

From summarized data

$$\frac{\sum_{j=1}^{J} \hat{\gamma}_j^2 \sigma_{Yj}^{-2} \hat{\beta}_j}{\sum_{j=1}^{J} \hat{\gamma}_j^2 \sigma_{Yj}^{-2}}.$$

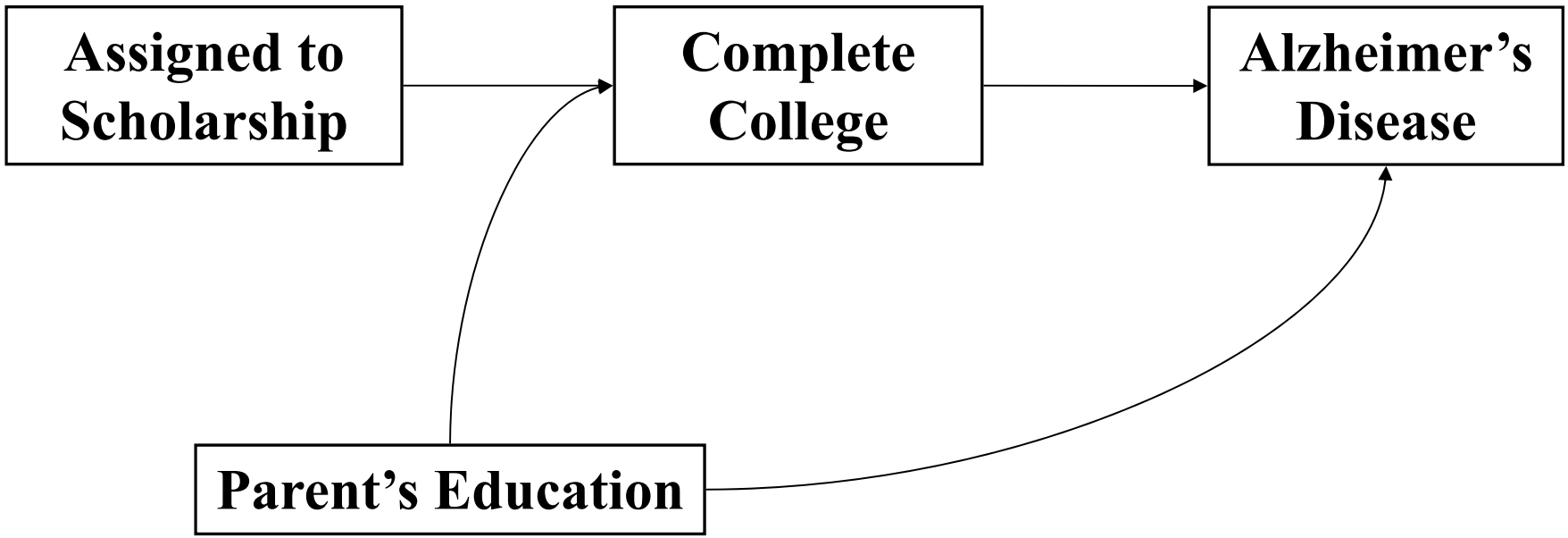
Weighted Averages



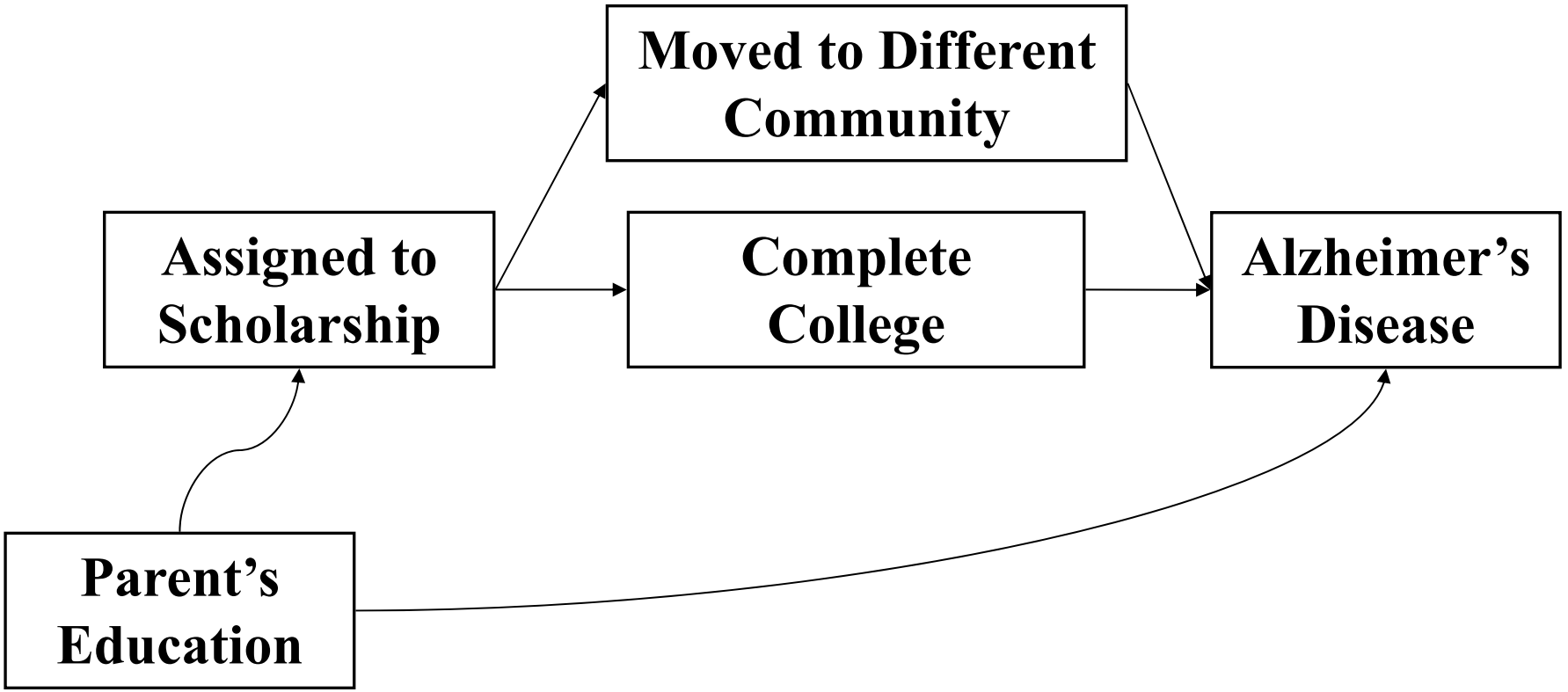
–Weighted Average =

–Slide thanks to Braithwaite, Allen, Krishnaswami, and Fan. Thanks!

IV can solve unmeasured confounding



But only under strong structural assumptions



What IV Analyses Estimate

- Most commonly, assume monotonicity: the instrument does not have *opposite* effects on individuals within the population.
  - "no contrarians"
  - This is violated if increasing mandatory school increases Joe's educational attainment, but *decreases* Earl's educational attainment.
- IV analyses estimate the effect of the treatment on people induced to receive the treatment because of the instrument
- This may or may not equal the population average effect.

What IV Analyses Estimate

- The 3 structural assumptions are enough to use IV to test the sharp null of no effect of treatment on exposure for anyone in the population
- But if you want to interpret the IV parameter as a particular causal effect, you need to adopt another assumption
- Different assumptions allow you to interpret this estimate as different causal estimands
- Consider the parameters you might be interested in:
  - Population average treatment effect
  - Effect of treatment on those who were treated
  - Effect of treatment on those who were treated *because of the value of the IV*, called the local average treatment effect
- The most common interpretation of the IV estimate is the last, the LATE, and for this we use the monotonicity assumption

Whose Causal Effect?

Response if assigned to treatment A:	*Response if assigned to treatment B:* Take A	Take B
Take A	Never-Takers	Compliers
Take B	Contrarians/ Defiers	Always Takers

Montonicity assumption: there are no contrarians

Under monotonicity, the IV estimate is the effect for compliers (LATE)

Variations on the Definition

- Standard econ:
  - If you are interested in estimating $Y_i=\beta X_i+e_i$ then Z is an instrument if "Z_i is correlated with X_i but not with ε_i" (Greene 2000, pg 370)
- Counterfactual:
  - Z is independent of Y_x (the counterfactual value of Y under any given value of X), and
  - Z is not independent of X. (Pearl, 2000, pg 248)
- AIR 1994 (developed originally for dichotomous Z, X):
  - SUTVA
  - Random assignment of Z
  - Exclusion: $Y(Z,X)=Y(Z',X)$ for all X, Z and Z'
  - Z has to have some effect on the probability of X
  - Montonicity: $X(Z=1)\geq X(Z=0)$ for everyone in the population

Economists' Account of IV

This [instrument] is a new independent variable which must have two characteristics. First, it must be contemporaneously uncorrelated with the error; and second, it must be correlated (preferably highly so) with the regressor for which it is to serve as an instrument.

Economists' Account of Confounding

$$\text{Outcome}_i=\beta_0+\beta_1 \text{ exposure}_i + \beta_k \text{ Other Covariates}_i+\varepsilon_i$$

Consider estimating the above as a structural (causal) equation. If the errors are correlated with exposure, the estimate of β_1 is biased.

Why would ε_i be correlated with exposure?

Because of a common cause of exposure and the outcome

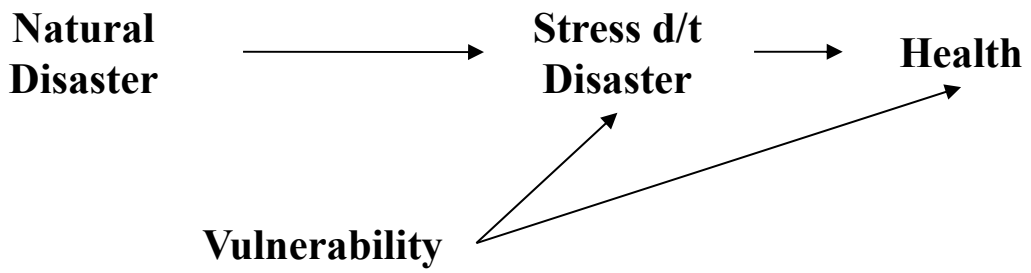
Or in econ parlance, because of a variable that should be included in the structural equation but has been omitted. Omitted variable bias=Confounding.

Separate Samples

- Sometimes, the exposure and the outcome are not available in the same data set
- If there are two data sets,
  - One with instrument and exposure
  - One with instrument and outcome
- You can estimate the IV effect

Common Confusions

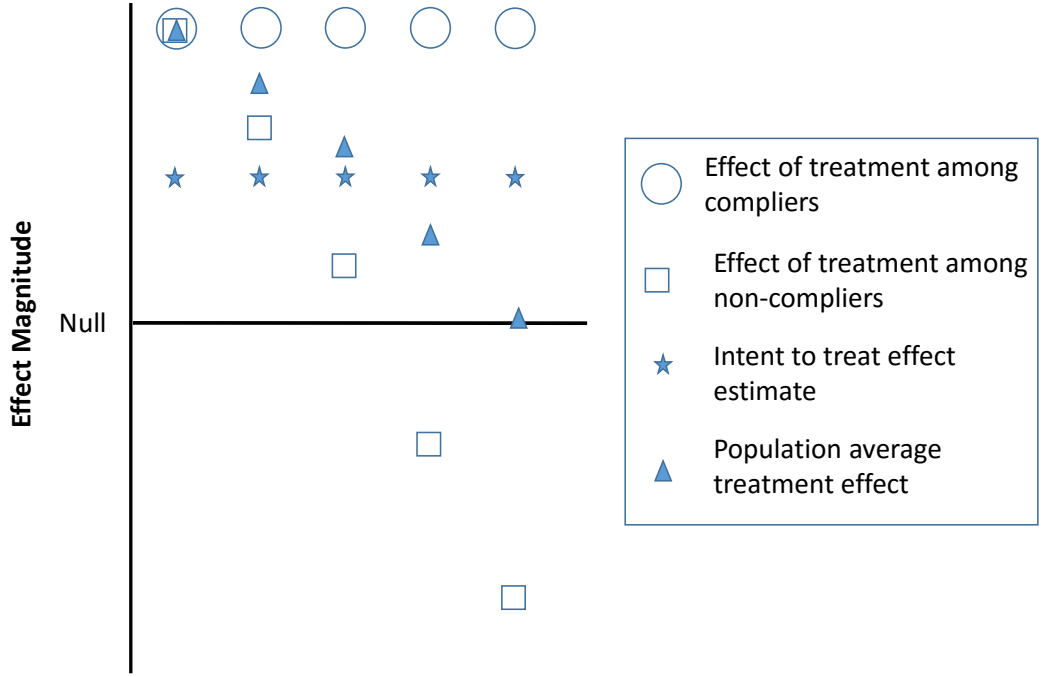
- Is an instrument a confounder?
- Should I adjust for an IV in a conventional model
- Is 2sls like a propensity score model?
- If variation was induced by a natural experiment… do I compare the exposed to the unexposed?



ITT vs PATE vs IV, under partial non-compliance

- If no individualized treatment selection, the effect among compliers is the same as among non-compliers. ITT<PATE, and PATE = IV
- Imagine that the effect of treatment among compliers remains the same, but the effect among non-compliers is increasingly smaller, so PATE gets smaller and ITT estimate stays constant in all scenarios.

- When the effect among non-compliers is less than the null (i.e., has a different sign than the effect among compliers), the ITT estimate is farther from the null than is the PATE, so the ITT estimate is no longer "conservative" compared to the PATE.

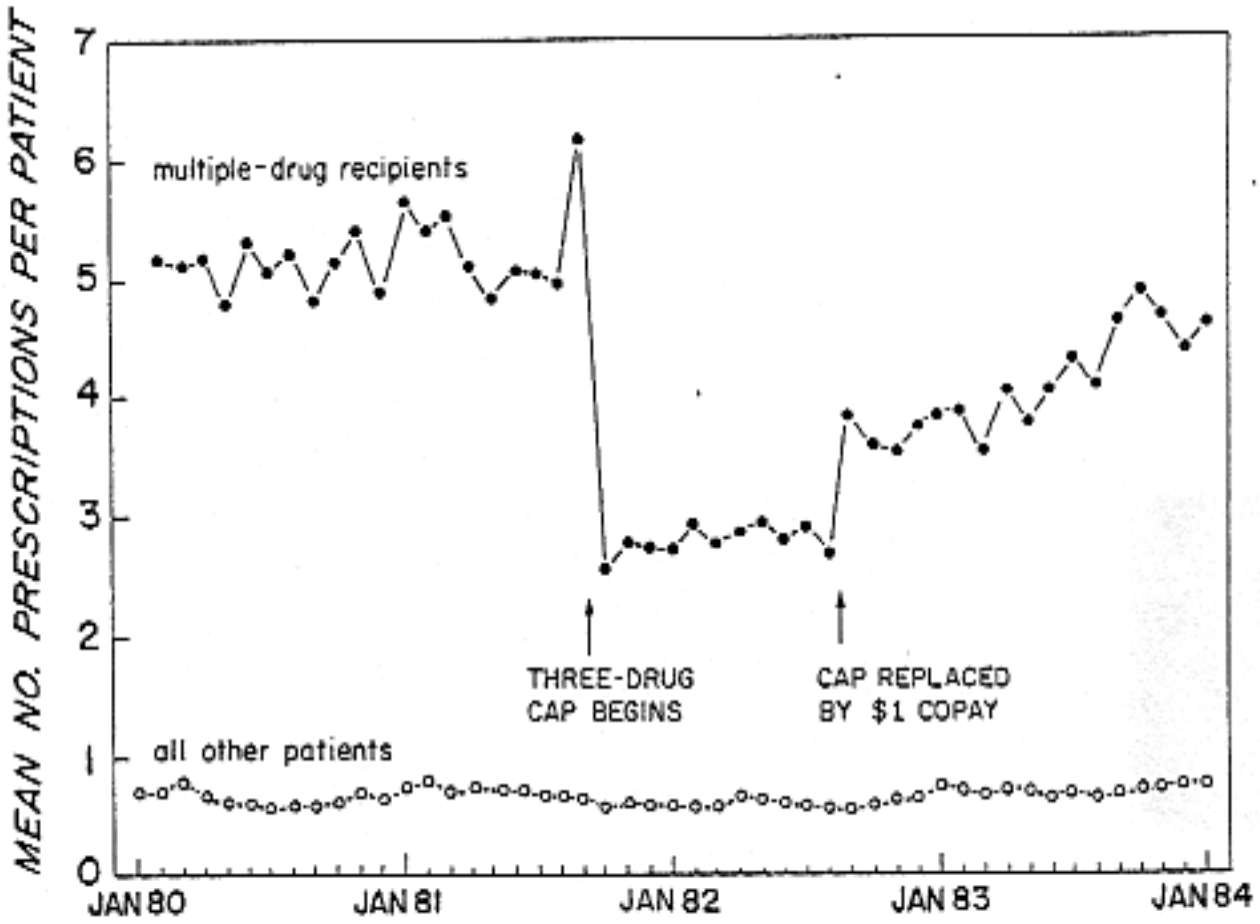


IV and related methods

- IV Analyses are a tool to combine information on Z-Y and Z-X to estimate X-Y.
- Sometimes used informally when people didn't measure X and possible only to test the null that X has no effect on Y.
- Often the null or the ITT is of most interest.
  - Policy makers: the instrument is the policy
  - For an RCT: not everyone will take up the drug, so what's the effect of availability?

IV and related methods

- Conceptualize a group of methods that are about moving a step back from a direct correlation (or regression) of exposure and outcome, using something that influences exposure to test the null
- Regression discontinuity: "in a highly rule based world, some rules are arbitrary and therefore provide good experiments"
- RDs often strengthened with a comparison group, otherwise risk being pre-post comparisons (remember threats to validity, e.g., history, maturation, regression)
- Can distinguish between "sharp" and "fuzzy" RD



STEPHEN B. SOUMERAI, SC.D., JERRY AVORN, M.D.,
DENNIS ROSS-DEGNAN, SC.D., AND STEVEN GORTMAKER, PH.D.
New England Journal of Medicine 317:550-556 (August 27), 1987

IV and related methods

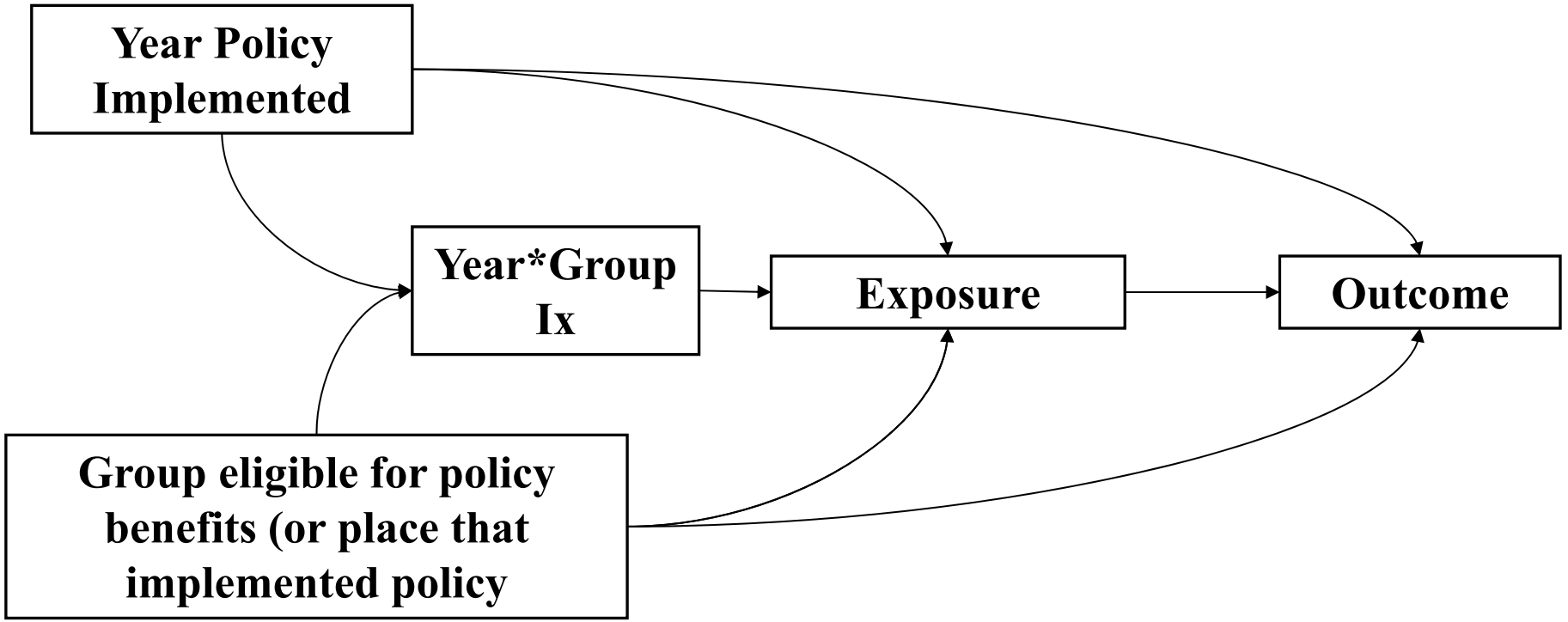
- "sharp" RD
- Exposure is a deterministic function of an underlying continuous characteristic
  - Test score (college eligibility)
  - Date (policy implementation)
  - Class size (additional class resources)
  - Weeks of gestation (abortion access)
  - Income (Medicaid eligibility)
  - BP (? Receive medications)
- X_i=1 if z_i>=cutoff
- X_i=0 if z_i<cutoff
- Outcome$_i$=β_0+β_1 (Z_i>=cutoff) +β_2 Zi+ β_k Other Covariates+ε_i
- β_1 now corresponds with the effect of X.

IV and related methods

- "fuzzy" RD
- Exposure is influenced by an underlying continuous characteristic
  - Mandatory school dropout age or work permit age
  - Amount of $ your MD received in marketing incentives
  - Distance from specialized care facility (probability receiving specialized care)
  - Blood type (transplant waiting time)
- These are almost equivalent to IVs as we have discussed above.
- Economists used to think of these as separate – conceptualization merging (Angrist and Pischke say: fuzzy RD=IV)

Regression Discontinuity/Fixed Effects

- Caveat controversy over whether this is a valid DAG, so consider it a heuristic.

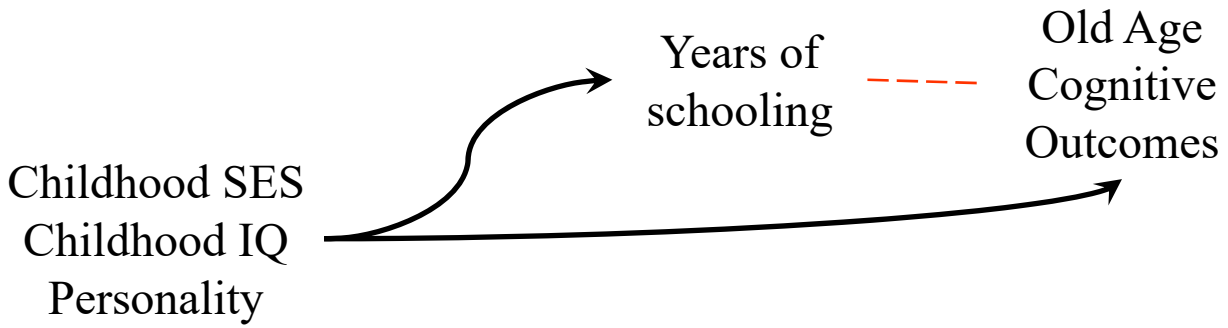


An example: CSLs and education

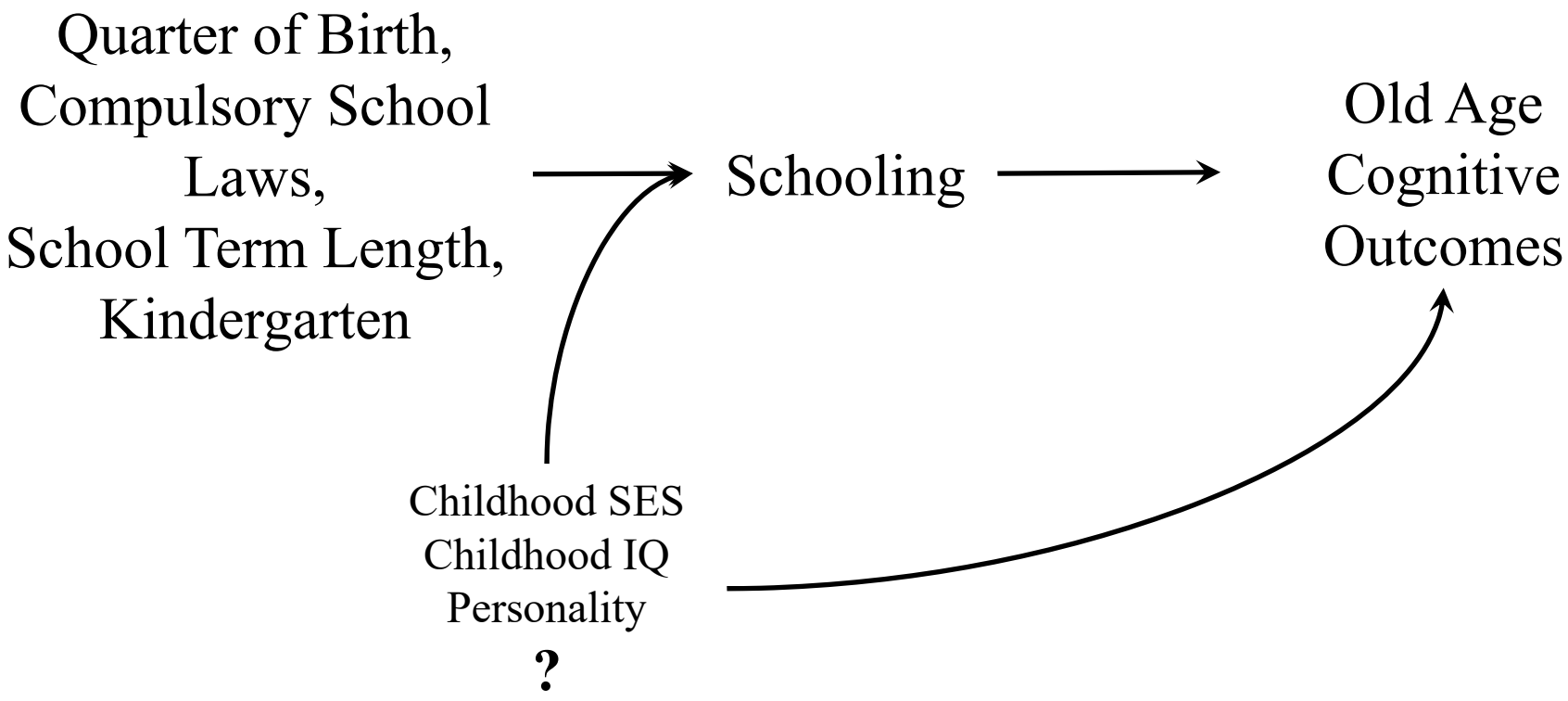
Substantive Question

Multiple studies show that years of education predicts old age cognitive function, cognitive change, and dementia.

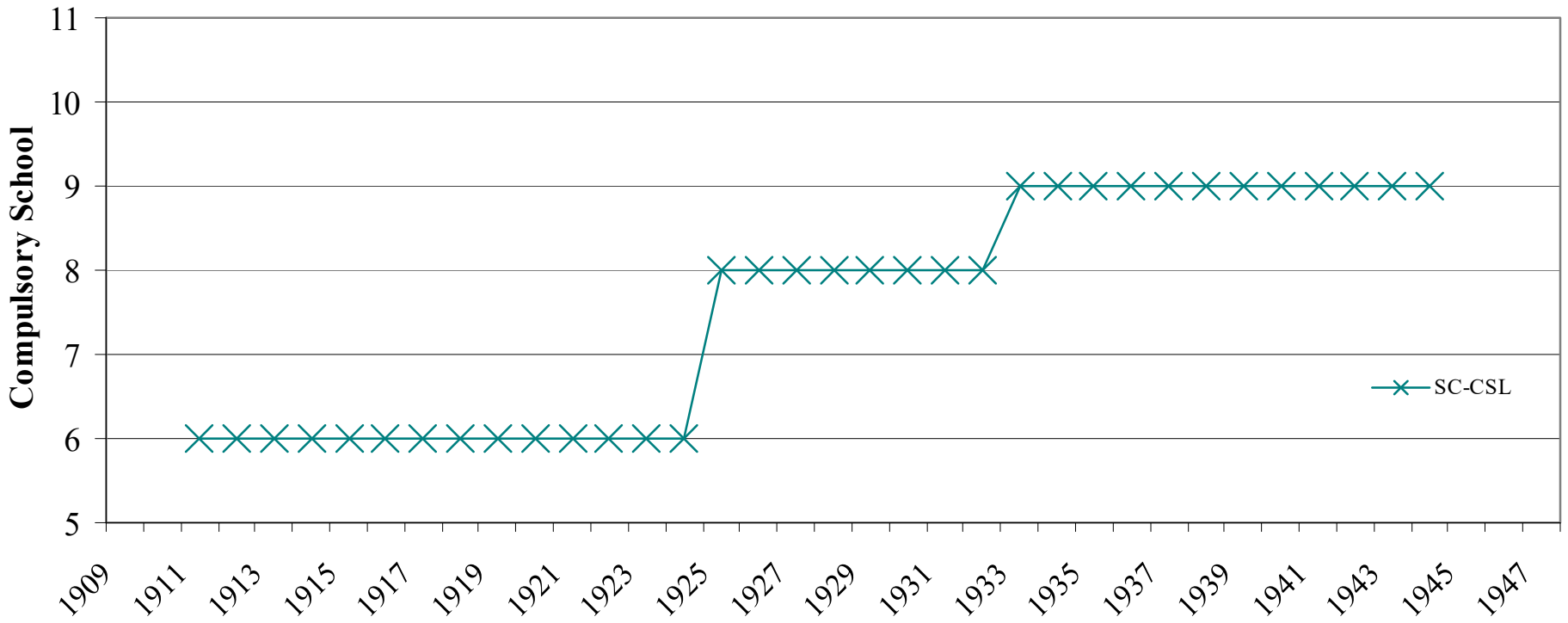
Causality questionable.



Natural Experiments for Education



Early 20th Century CSL Changes



Hypotheses

Variation in schooling is induced by CSLs, CSL-Ws

&

these instruments also predict old age cognitive test scores.

IV Analyses

- State schooling policies
  - Compulsory school to drop out (CSL) or receive a work-permit (CSL-W)
  - Based on policy in state of birth when school-age
  - 2-Sample least squares analysis
- Exposure (endogenous) variable:
  - Years of education (self-report)

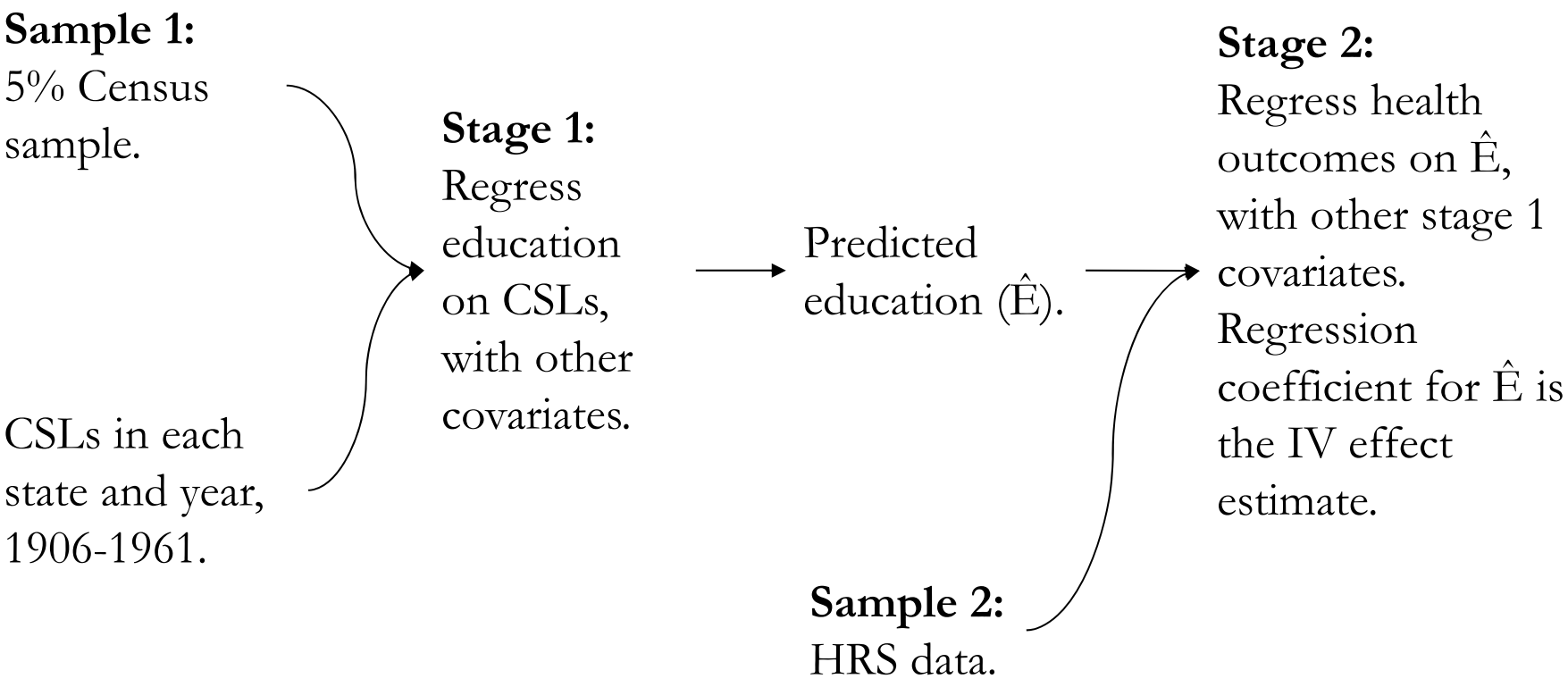
Data Set: 1st Stage

- IPUMS (Census) 5% 1980 sample,
- Birth years 1900-1947
- Years of education linked to CSLs and CSL-Ws based on state of birth
- Link predictions from 1st stage regression model to individual data in the 2nd stage based on state of birth and all covariates.

Data Set: 2nd stage

- Health & Retirement Study, 1992-2000: panel enrollment by birth cohort (whites only due to evidence on enforcement)
- Cognitive assessments and state of birth on 21,041 individuals born 1900-1947
- CSLs and CSL-Ws

Two-Sample Least Squares



Outcomes

- Immediate and delayed word list recall
  - Z-scores averaged over all years of interviews (1-5)
- Combined cognitive score
  - Modified telephone interview for cognitive status + serial-7 subtractions
  - Z-scores averaged over all years of interviews (1-4)

Covariates

1
- Unadjusted

2
- Sex
- Birthyear (indicators for every year)

3
- State of birth indicators

4
- State characteristics: age 6 % black, % urban, and % foreign born; age 14 manufacturing jobs per capita and wages per manufacturing job

HRS Sample

	n	%
n	11,439	100
Male	4,862	43
Birth Year		
<1914	1,278	11
1914-1921	2,078	18
1922-1930	2,656	23
1931-1941	4,241	37
1942-1947	1,186	10
Years of schooling		
<6	308	3
6-8	1,706	15
9-11	2,677	23
12	6,719	59

Distribution of Instruments

CSLs

	n	%
Not Restricted	115	1
< 7 years	323	2
7 years	1,481	7
8 or 9 years	12,928	60
10 years	1,461	7
11 years	76	0
12 or more years	5,056	24

CSL-Ws

	n	%
Not Restricted	174	1
< 6 years	406	2
6 years	2,645	12
7 years	5,562	26
8 or more years	12,653	59

Social and Economic Characteristics of States

	CSL		CSL-W	
	rho	**p**	**rho**	**p**
% Urban at age 6	0.15	<0.01	0.11	<0.01
% Foreign-born at age 6	0.01	0.65	-0.10	<0.01
% Black at age 6	-0.09	<0.01	0.06	0.01
% Manufacturing jobs at age 14	0.05	0.02	0.07	<0.01
Manufacturing wages per job	0.38	<0.01	0.36	<0.01
CSL-W	0.49	<0.01	-	-

Do the Instruments Predict Education?

First stage regression results (from IPUMS 5% sample)

	1. Unadjusted Model	2. Birthyear and sex	3. Model 2 + state of birth	4. Model 3 + state condns
CSLs	0.238 (0.236, 0.240)	0.110 (0.108, 0.112)	0.062 (0.059, 0.064)	0.037 (0.034, 0.040)
CSL-Ws	0.143 (0.146, 0.141)	-0.032 (-0.034, -0.029)	0.063 (0.060, 0.066)	0.044 (0.040, 0.048)
CSL-Ws UNR	-1.397 (-1.429, -1.365)	-0.282 (-0.315, -0.249)	-0.204 (-0.238, -0.17)	0.034 (0.000, 0.069)

How Strong is the 1st Stage?

	1. Unadjusted Model		2. Birthyear* and sex.		3. Model 2 + state of birth indicators		4. Model 3 + state characteristics#	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Model r^2 without instrumental variables	0.0000		0.1080		0.1599		0.1626	
Model r^2 including instrumental variables	0.0465		0.1127		0.1613		0.1631	
Variance explained by instrumental variables	0.0465		0.0047		0.0014		0.0005	

Not technically "weak" instruments, but clear that a small violation of the IV assumptions could introduce a large amount of bias.

IV Estimates for Education: CSLs

Estimated effect of 1 year ed'n on cognitive test scores.

	Memory		Cognition	
Model covariates	**β_{IV}**	**95% CI^**	**β_{IV}**	**95% CI^**
1. Unadjusted	0.33	(0.27, 0.39)	0.19	(0.12, 0.26)
2. Birthyear, and sex	0.30	(0.14, 0.46)	0.34	(0.05, 0.63)
3. Model 2 + birth state	0.18	(0.02, 0.33)	0.03	(-0.22, 0.27)
4. Model 3 + state condns	0.34	(0.11, 0.57)	-0.06	(-0.37, 0.26)
5. OLS estimates	0.09	(0.08, 0.10)	0.15	(0.14, 0.16)

Sensitivity Analyses

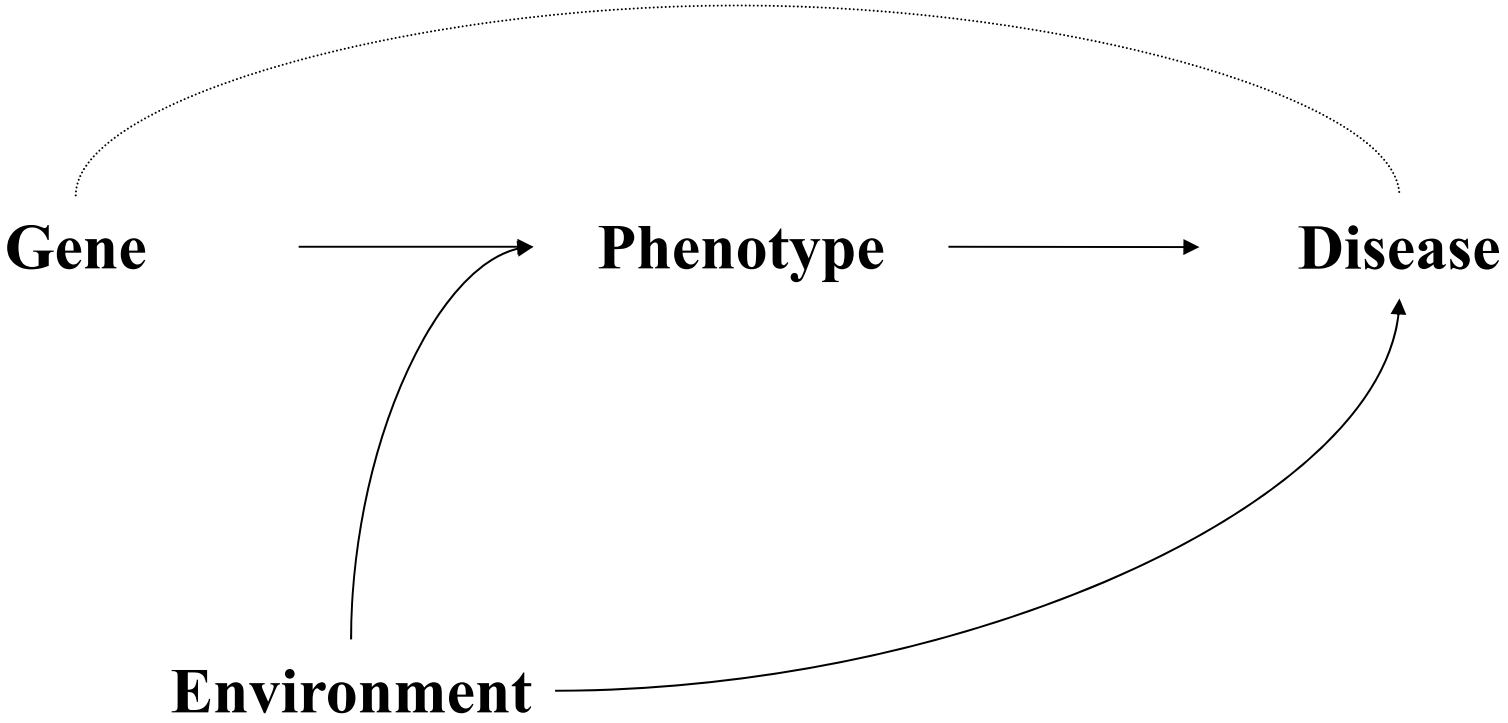
- Including education >13 years
  - β_{IV} (memory, model 3): 0.15 (-0.01, 0.31)
- Restricting to education > 13 years
  - Instruments do not predict education or memory for individuals with >13 years of school
  - β_{IV} (memory, model 3): -1.04 (-3.70, 1.62)
- Inverse probability weighted for missing Memory (parental SES, self-report chronic condns at baseline)
  - β_{IV} (memory, model 3): 0.19 (0.03, 0.36)

Conclusions

- CSLs and CSL-Ws applicable in state of birth predicted years of education completed.
- IV estimates of the effect of education on Memory score were significant and larger than OLS estimates.
- IV estimates of the effect of education on cognition were uninformative: CIs included OLS estimates and the null.
- Other state characteristics are plausible confounders, but must change in tandem with CSLs and affect only those with <13 years of education.

Mendelian Randomization

Mendelian Randomization & IV



Invalid Instruments in MR

- Canalization
- Pleiotropy

Direct pathways from the instrument to the outcome

- Population stratification
- Linkage disequilibrium

Common prior causes of the instrument and the outcome

Evaluating the assumptions

1. Constraints implied by theory
2. Over-identification tests
3. Stratification-based tests (similar to over-identification)
4. IV inequality constraints
5. Negative controls
6. Independent from known confounders
7. Egger tests

1 Evaluating MR Studies: leverage prior assumptions about confounding

- Often, field is only interested in knowing if a conventional effect estimate is biased up
- Other direction of bias not of interest, or not considered plausible
- Compare IV effect estimate to conventional effect estimate: if conventional effect estimate is positively confounded, IV<conventional
- 4 *equivalent* versions of this test: no more convincing to show all 4.
- Relies on assumption regarding the direction of confounding: if you don't know the direction, the test is not informative
- Not guaranteed to be consistent in non-linear causal structures (but doesn't completely rely on linearity).

2 Evaluating MR Studies: Over-identification tests

- Use multiple instrumental variables to conduct over-identification tests.
- Other genes or even polymorphisms of the same gene might provide additional instruments.
- Cannot detect violations of the IV assumptions if all instruments have identical biasing pathways.
- May also reject even when all instruments are valid if the model is incorrectly assumed to be linear or the phenotype is composite. These tests generally have low statistical power.

3 Evaluating MR Studies: instrumental inequality tests

- These tests are applicable only when the causal phenotype is known to be categorical.
- Certain inequalities are impossible given the IV assumptions: if you see them, IV must not be right

$$\max_i [\Pr(X=0, Y=1|G=i)] \leq \min_i [1 - \Pr(Y=0, X=0|G=i)]$$

$$\max_i [\Pr(X=0, Y=1|Z=i) + \Pr(X=1, Y=1|Z=i)] + \max_i [\Pr(X=0, Y=1|Z=i) + \Pr(X=1, Y=0|Z=i)] + \max_i [\Pr(X=0, Y=0|Z=i)] \leq 2$$

$$\max_i [\Pr(X=1, Y=1|Z=i)] \leq \min_i [1 - \Pr(Y=0, X=1|Z=i)]$$

- Detects extreme violations of the assumptions
- With more instrumental variables, more opportunities for violations of tests.

4 Evaluating MR Studies: Modifying factors

- Identify factors that modify the genotype-phenotype association.
- Compare the IV effect estimate across groups in which the population association between the instrument and the phenotype is either silenced or reversed.
- This test could identify a biased instrument if the biasing pathway is active in both subgroups.

5 Evaluating MR Studies: Negative controls (theoretically based)

- Identify situations (types of people, particular locations) in which the phenotype should not affect the outcome
- Identify outcomes that would be associated with similar pleiotropic pathways but not causal pathways
- Identify non-linearities (?? BMI only relevant above "obesity" threshold)

7 Egger Regression: treat each IV estimate as an element in a meta-analysis

- Under the InSIDE (instrument strength independent of direct effects) assumption, bias converges to zero.
- Regress the Z-Y associations on the Z-X associations.

The intercept is an estimate of average pleiotropy and the slope is an estimate of the true causal effect under InSIDE.

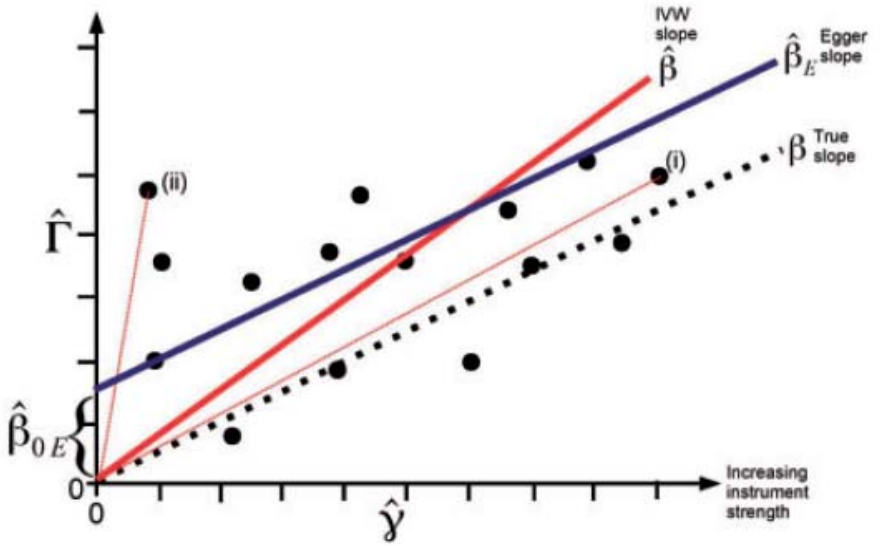


Figure 2. Plot of the gene–outcome ($\hat{\Gamma}$) vs gene–exposure ($\hat{\gamma}$) regression coefficients for a fictional Mendelian randomization analysis with 15 genetic variants. The true slope is shown by a dotted line, the inverse-variance weighted (IVW) estimate by a red line, and the MR-Egger regression estimate by a blue line. Refer to text for explanation of points

Bowden, Davey Smith, and Burgess, Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression, IJE 2015

Conclusions

- Honest epidemiologists tend to be worried epidemiologists
- Abandon the difficult questions?
  - Or learn what we can from fraught methods?
- IV adds:
  - A way forward with observational data
  - Sometimes a parameter estimate of special interest
  - Pushes us to identify interventions that change exposures
- A set of related methods use pseudo-randomization, strongly preferred in econ, less common in epi