The placement of instrumental variable methods in an epidemiologist's toolbox

Sonja A. Swanson
Department of Epidemiology, Erasmus MC
s.swanson@erasmusmc.nl

Danish Epidemiological Society 2019 Meeting

Disclaimer: why my emphases?

- Even if you never conduct an IV analysis, you will have:
  - Colleagues who use an IV analysis to study a topic in your area
  - Colleagues who propose conducting an IV analysis with your data
  - A need to weigh whether an IV analysis is appropriate for your research question and data setting
- Thus, you need to know for what questions IV methods work, when they work well, and how to weigh their strengths and limitations on a case-by-case basis

Outline

- Motivation for IV methods
- A gentle introduction to IV methods
- Three underappreciated practical issues
  - Sensitivity/bias analyses and falsification strategies

exist
- An IV analysis is not a cure-all
- Classical IV analyses are designed for point/time-fixed

exposures
- Summary

Outline

- Motivation for IV methods
- A gentle introduction to IV methods
- Three underappreciated practical issues
  - Sensitivity/bias analyses and falsification strategies

exist
- An IV analysis is not a cure-all
- Classical IV analyses are designed for point/time-fixed

exposures
- Summary

Why IVs?

- Most methods for estimating causal effects from observational studies rely on an assumption of no unmeasured confounding
- Dream with me: what if we can avoid this assumption?
  - IV methods to the rescue?

Use of IV methods

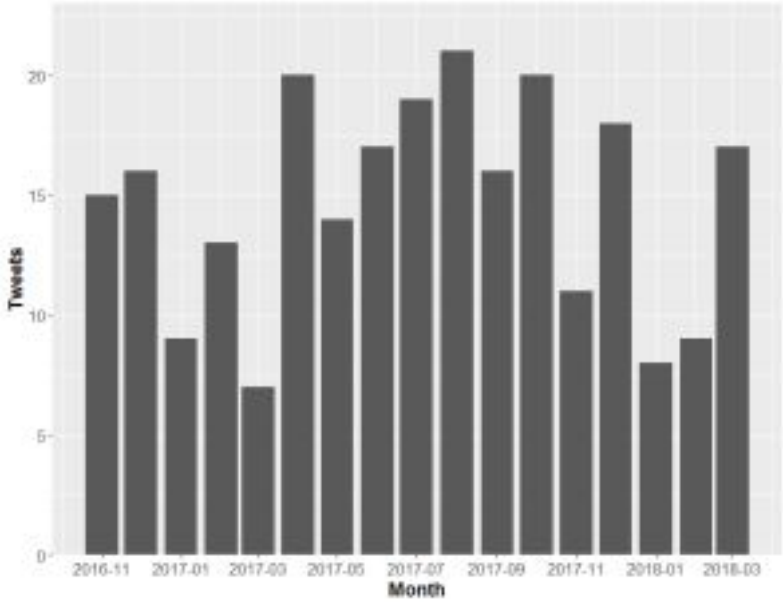
- IV analyses frequently conducted in pharmacoepidemiology and social epidemiology
- Mendelian randomization (MR), a special case of IV methods, is exploding with popularity

- Boef et al. 2015 *IJE* systematic review of MR studies

Exposure	Number of studies
C-reactive protein	29
BMI or fat mass	25
Alcohol consumption	12
Vitamin D	10
Other	103 (of 64 unique exposures)

Use of MR

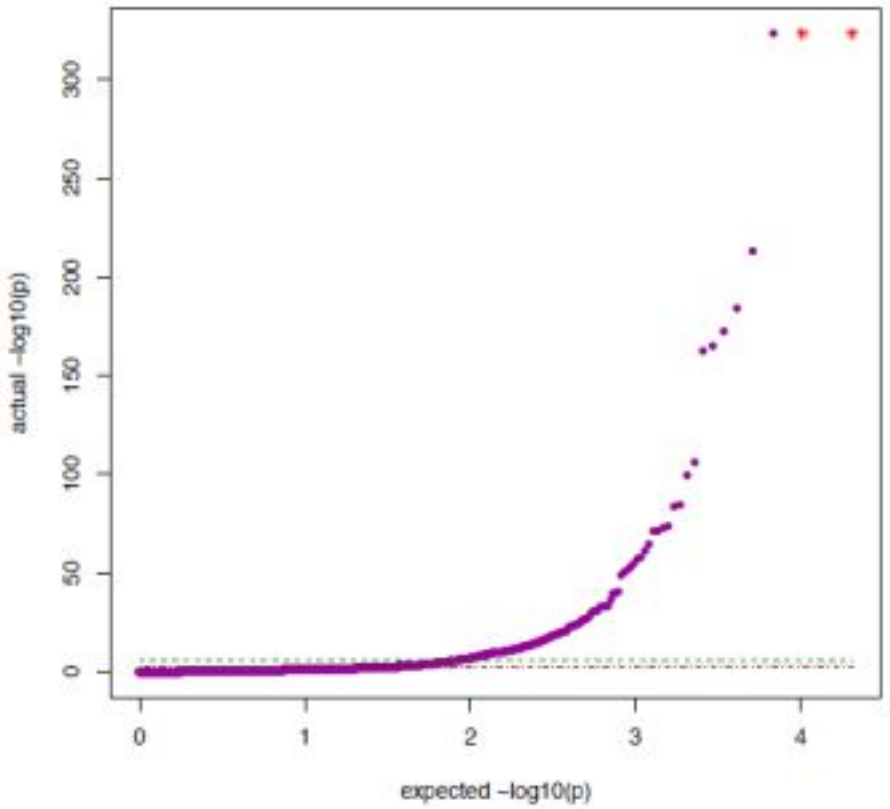
- Twitter MR publication alert
  - @Mendelian_lit
  - 22 tweets last month alone



Use of MR

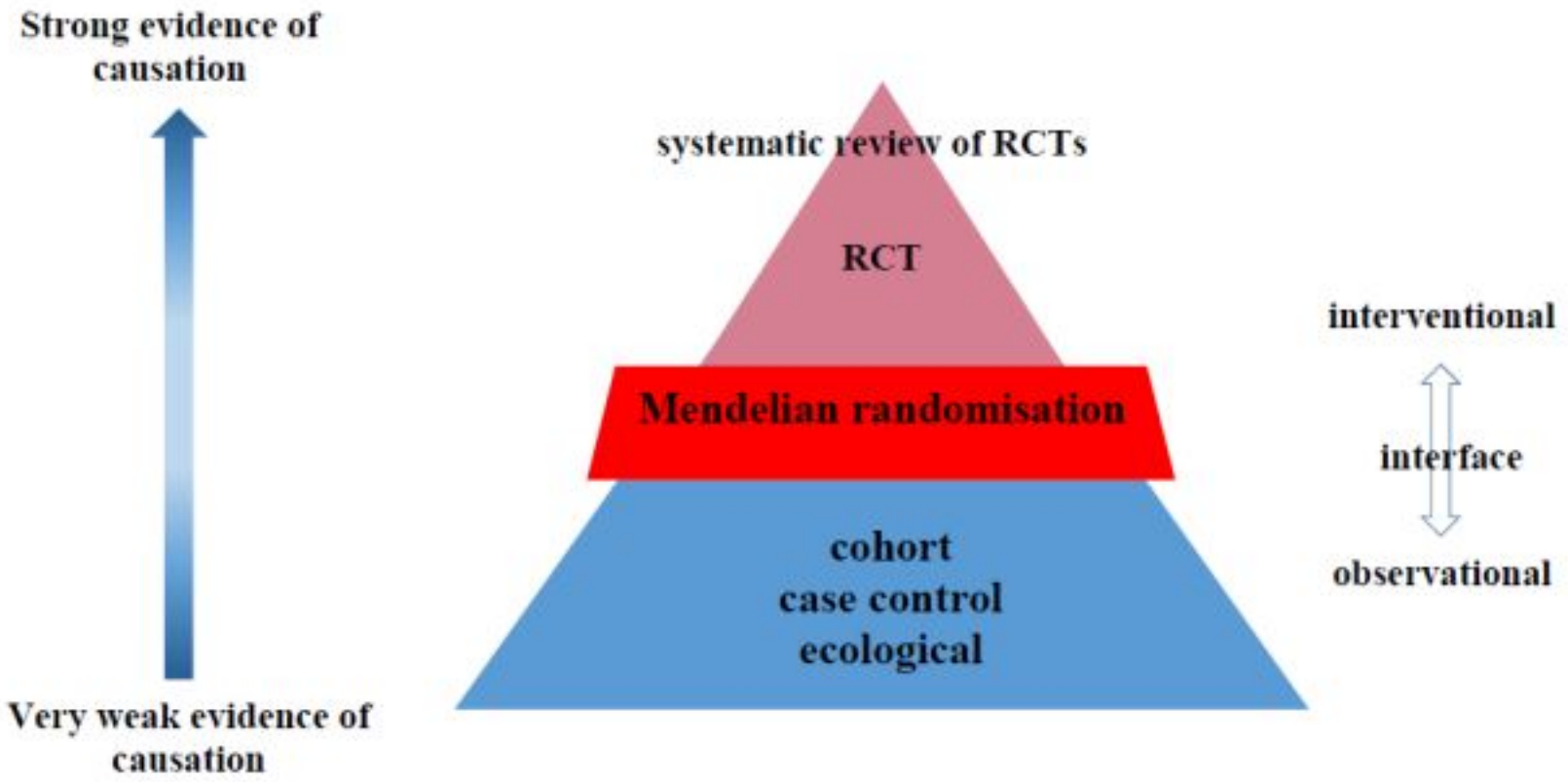
- Now estimating several causal effects within a single study
- E.g., effects of BMI on 20,461 unique outcomes

Figure 3: QQ plot of 20,461 MR-pheWAS results



Millard et al. 2017 *bioRxiv pre-print*

Description of MR (/IVs) in the literature



Zuccolo et al. 2017 *IJE*, Davies et al. 2018 *BMJ*

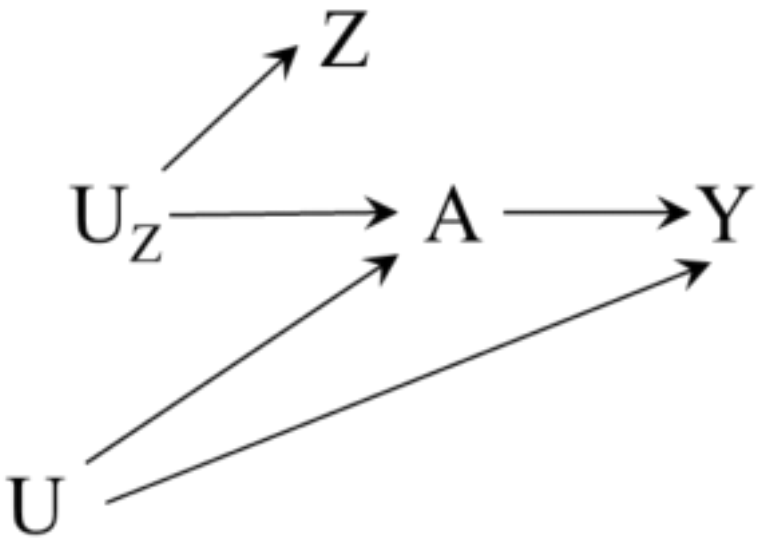
Outline

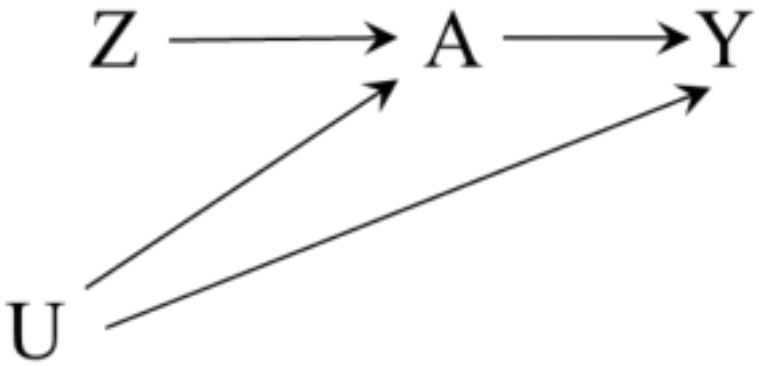
- Motivation for IV methods
- A gentle introduction to IV methods

- Three underappreciated practical issues
  - Sensitivity/bias analyses and falsification strategies exist
  - An IV analysis is not a cure-all
  - Classical IV analyses are designed for point/time-fixed exposures
  - Summary

Instrumental conditions

i. Instrument *Z* and treatment *A* are associated
ii. Instrument *Z* causes outcome *Y* at most through treatment *A*
iii. Instrument *Z* and outcome *Y* share no causes





Z=proposed instrument; *A*=treatment;

Y=outcome; *U*=unmeasured confounders; U_Z=unmeasured causal instrument

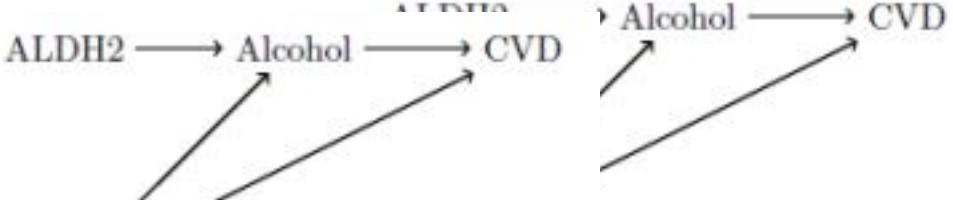
Example: physician's preference

i. Preference (measured by treatment prescribed to prior patients) and prescribed treatment are associated ii. Preference causes outcome at most through treatment iii. Preference and outcome share no causes

- Other commonly proposed IVs in pharmacoepi:
  - Distance
  - Calendar time
  - Geographic variation

Example: genetic variants (MR)

i. Genetic variant and exposure



are

associated ii. Genetic variant causes outcome only through exposure iii. Genetic variant and outcome share no causes

Considering condition (i)

i. Instrument *Z* and treatment *A* are associated ii. Instrument *Z* causes outcome *Y* at most through treatment *A*

iii. Instrument *Z* and outcome *Y* share no causes

- Condition (i) is verifiable
- Weak associations can be problematic
  - "Weak instrument bias"

Considering conditions (ii)-(iii)

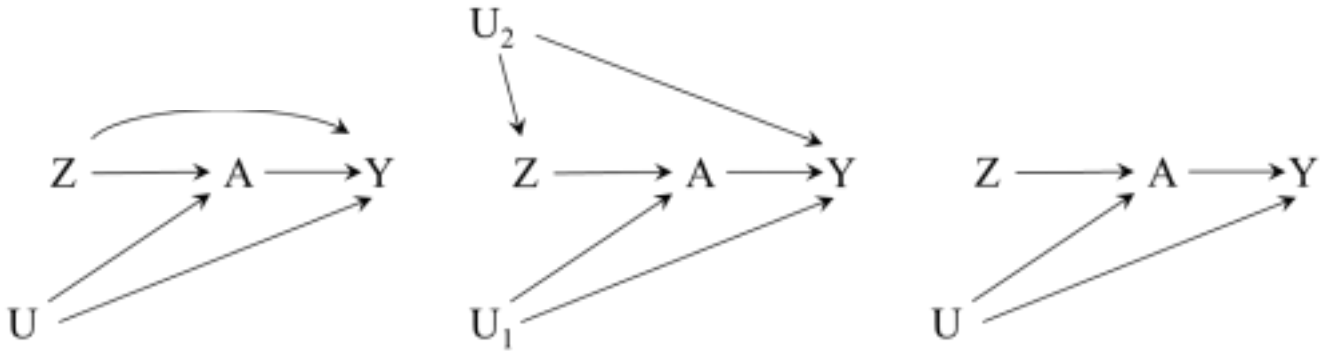
i. Instrument *Z* and treatment *A* are associated ii. Instrument *Z* causes outcome *Y* at most through treatment *A*

iii. Instrument *Z* and outcome *Y* share no causes

- Conditions (ii)-(iii) are not verifiable
  - Need to be justified with subject matter knowledge
  - Can sometimes be falsified
- Violations of conditions (ii)-(iii) can be subtle
- Biases due to violations of conditions (ii)-(iii) can be large and in counterintuitive directions

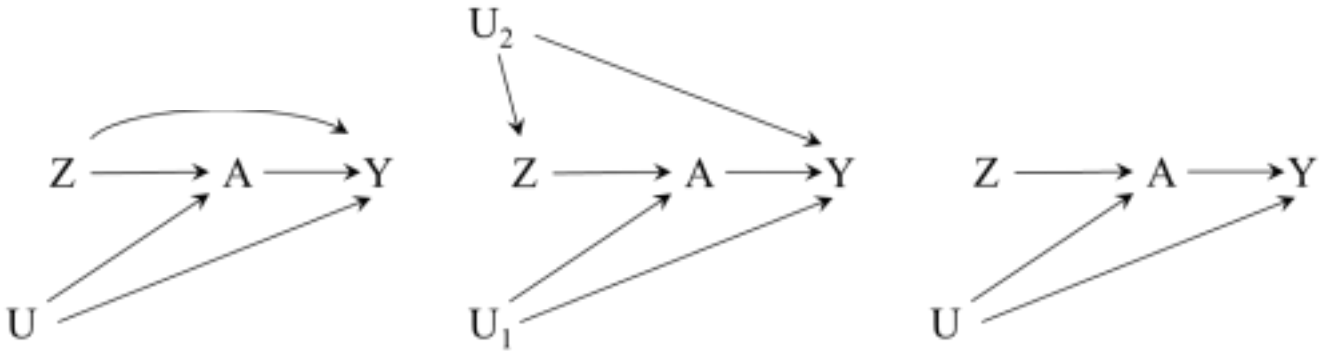
Violations of (ii) and (iii) examples

- Examples when preference may not be a valid IV
  - Physicians who prefer to prescribe the treatment also tend to prescribe another medication concomittantly
  - Patients with similar characteristics tend to see the same physicians (e.g., specialty services)
- Also, subtle violations
  - Selecting on treatment, measurement error in treatment, etc.



Violations of (ii) and (iii) examples

- Examples when genetic variants may not be a valid IV
  - A genetic variant may have pleiotropic effects on risk factors that affect the outcome, not via first affecting the exposure of interest
    - Ancestry may confound the relationship between the genetic variant and outcome
- Also, subtle violations
  - Selection biases etc.



But an instrument is not enough…

- IV analyses require a "fourth" assumption to obtain point estimates
- Identifying the effect in the full study population requires some form of homogeneity
- Identifying the effect in the "compliers" requires monotonicity

Stated effect of interest (N=81 IV studies)

LATE ATE Both Unclear

38%

49%

10% 3%

Swanson & Hernan *Epi* 2013

Outline

- Motivation for IV methods
- A gentle introduction to IV methods
- Three underappreciated practical issues
  - Sensitivity/bias analyses and falsification strategies exist
  - An IV analysis is not a cure-all
  - Classical IV analyses are designed for point/time-fixed exposures
  - Summary

Bias analyses and falsification tools exist

- Just as with our "usual" analyses, we can use some strategies to consider the plausible magnitude of bias
- Several tools already exist that are underutilized
- IV inequalities, negative controls, bias component plots, various specific

bias analyses, overidentification tests, relative confounding comparisons, preference surveys…

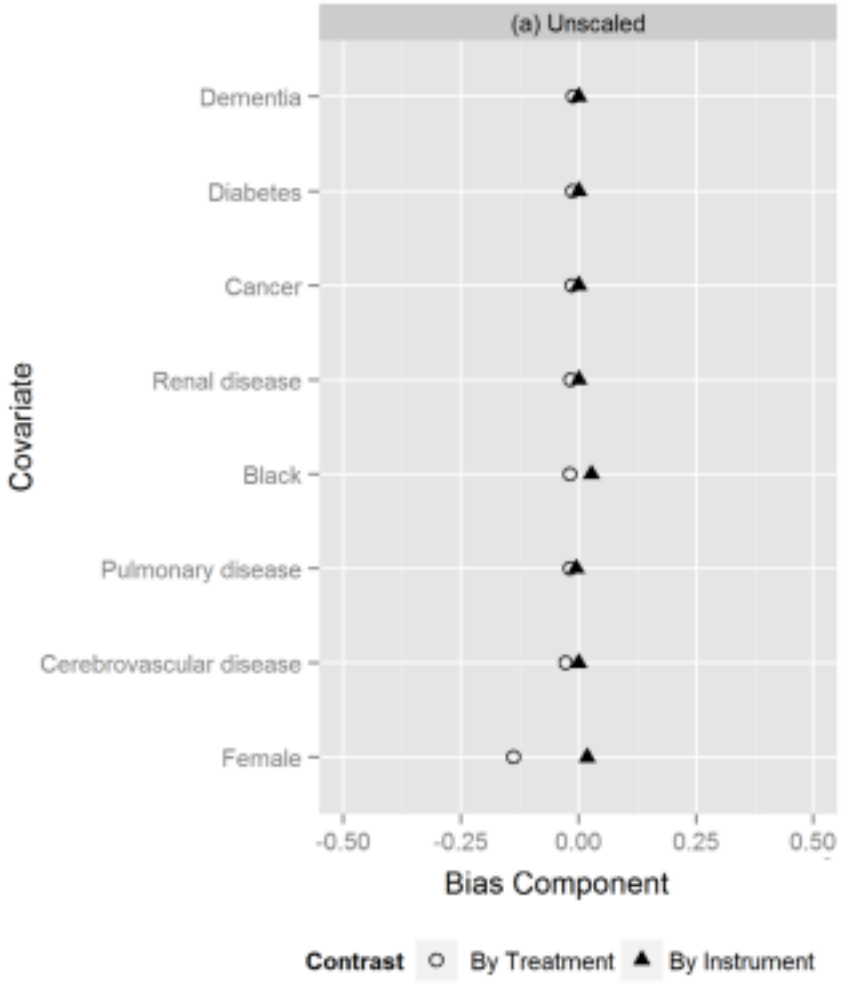
- See Labrecque & Swanson *Curr. Epi. Reports* 2018 for a review
- Some types of potential biases remain poorly understood
- Opportunities for methods development?
  - Need to help epidemiologists develop intuitions and heuristics?

Example: violations of condition (iii)

- As condition (iii) is essentially shifting our no unmeasured confounding assumption from *A*-*Y* to *Z*-*Y*, can we repurpose tools to the IV setting?
  - Yes and no: bias will be amplified by the *Z*-*A* relationship
  - Tools like covariate balance checks used as diagnostics for non-IV methods may be misleading when applied to IVs

Brookhart et al. 2007 *IJB*; Jackson & Swanson 2015 *Epidemiology*

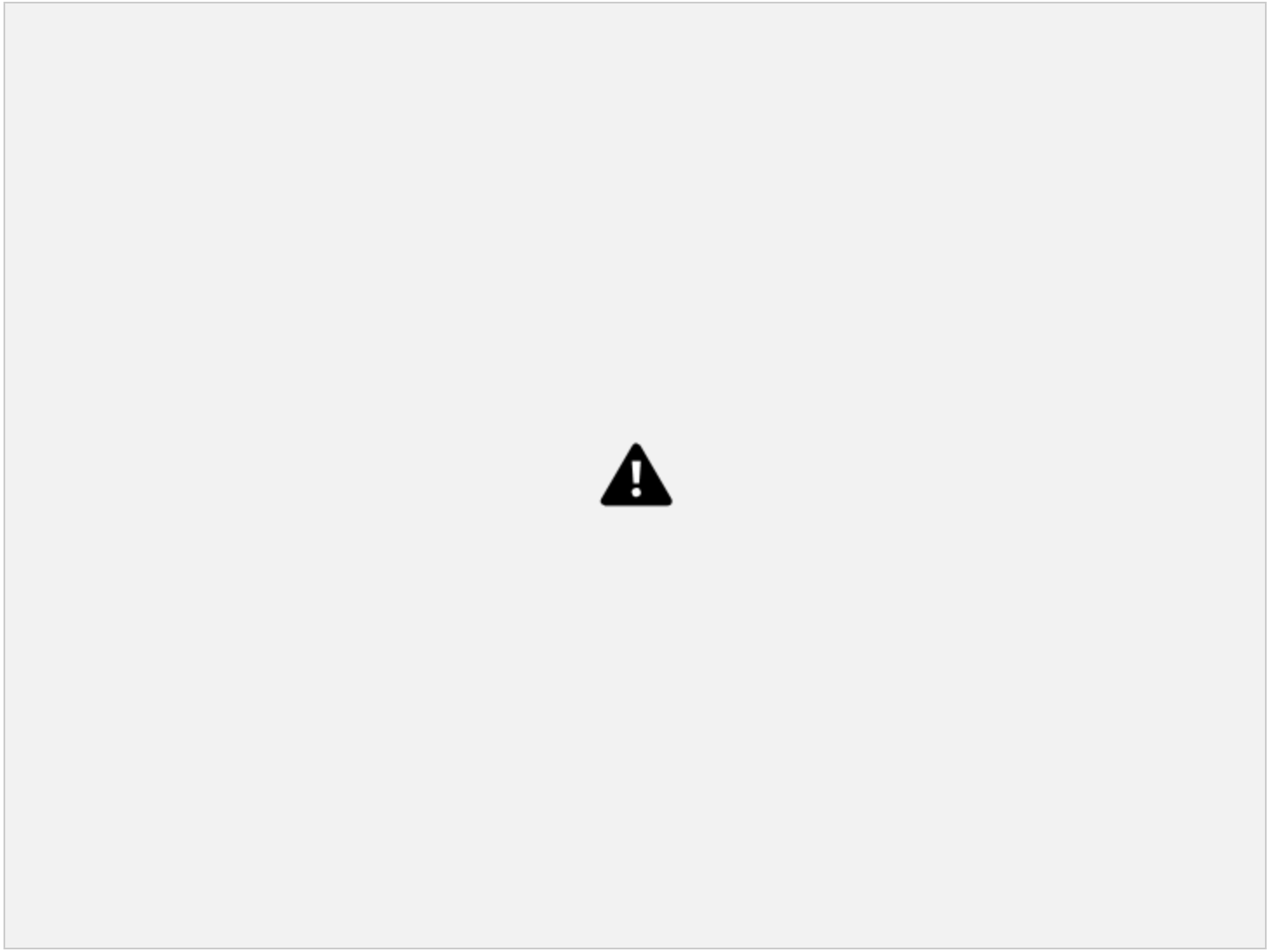
Covariate balance example



McClellan et al. 1994 *JAMA*; Jackson & Swanson 2015 *Epidemiology*

Covariate balance scaled

appropriately



McClellan et al. 1994 *JAMA*; Jackson & Swanson 2015 *Epidemiology*

Outline

- Motivation for IV methods
- A gentle introduction to IV methods
- Three underappreciated practical issues
  - Sensitivity/bias analyses and falsification strategies exist
  - An IV analysis is not a cure-all
  - Classical IV analyses are designed for point/time-fixed exposures
  - Summary

IVs are not a cure-all

- IV analyses may address unmeasured (baseline) confounding, but that is not the only issue we may encounter with observational data
  - Selection bias, information bias, ill-defined interventions, time

dependent confounding…

- Structure of selection (e.g.) needs to be considered to know whether and how possible bias can be addressed
  - Loss to follow-up
  - Selecting on treatment
  - Misalignment of “time zero”
  - …etc.

Time zero

- In a trial, three events occur at “time

zero” Treatment adherence and

outcome

1. Treatment assigned (randomization)
2. Eligibility criteria applied 3. Outcome recording begins

occurrence assessed during a specified period

Time zero

End of study period

Swanson et al. 2017 *Epi*

Time zero

- In MR (and other IVs), these three events may not co-occur 3. Outcome recording begins

Exposure level and outcome

1. Genetic variant inherited (conception)

Time
occurrence assessed during an unspecified period

End of zero

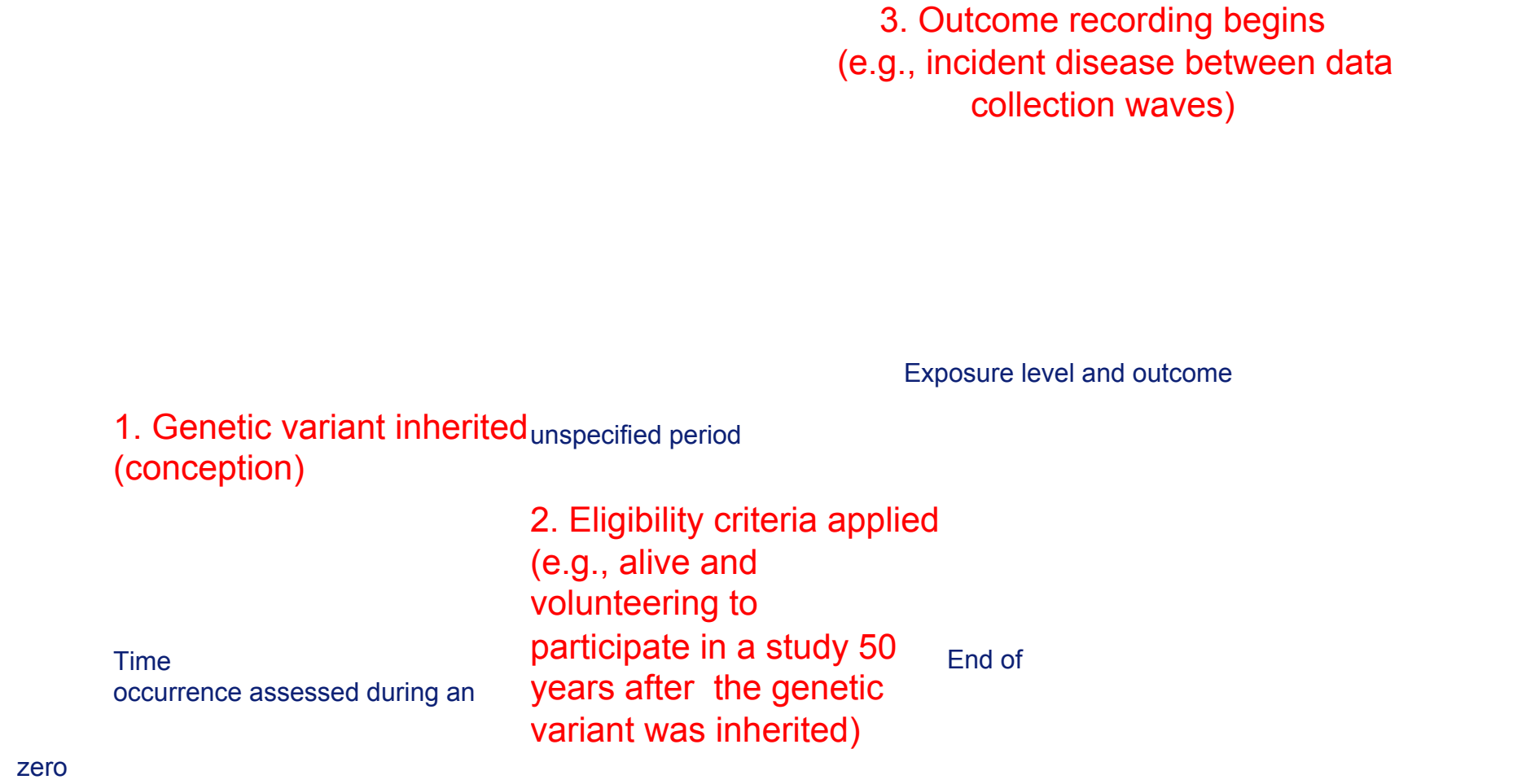
2. Eligibility criteria applied

Swanson et al. 2017 *Epi*

study period

Time zero

- In MR (and other IVs), these three events may not co-occur



study period

Swanson et al. 2017 *Epi*

Time zero

- Unclear time zero could create similar selection biases as would occur if we conditioned on post-randomization events in a trial

2. Eligibility criteria applied

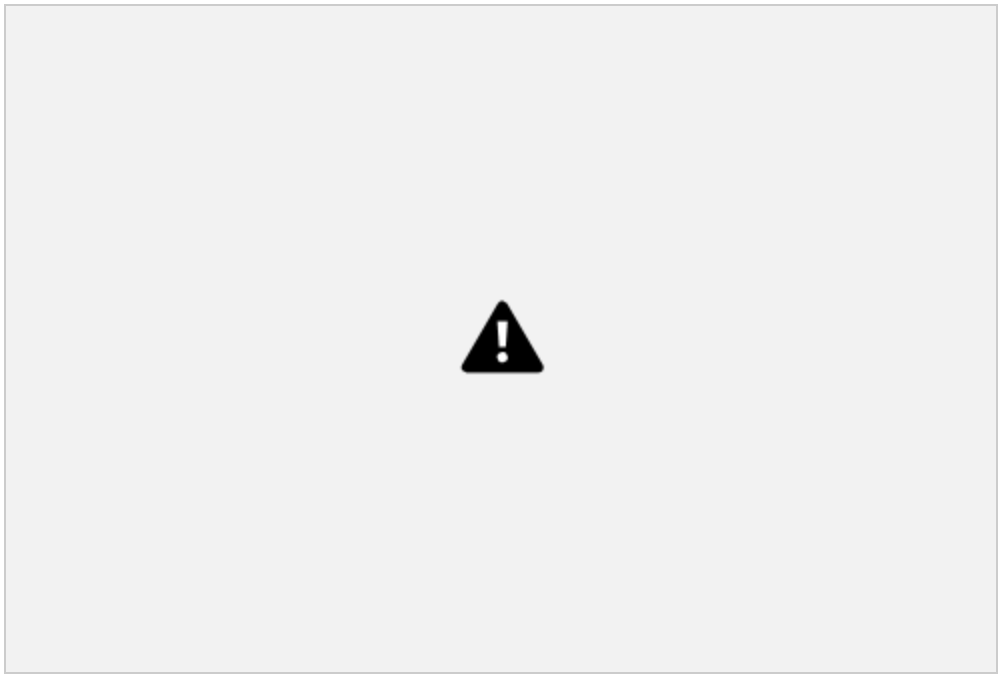
3. Outcome recording begins

1. Genetic variant inherited (conception)

Exposure level and outcome occurrence assessed during an unspecified period

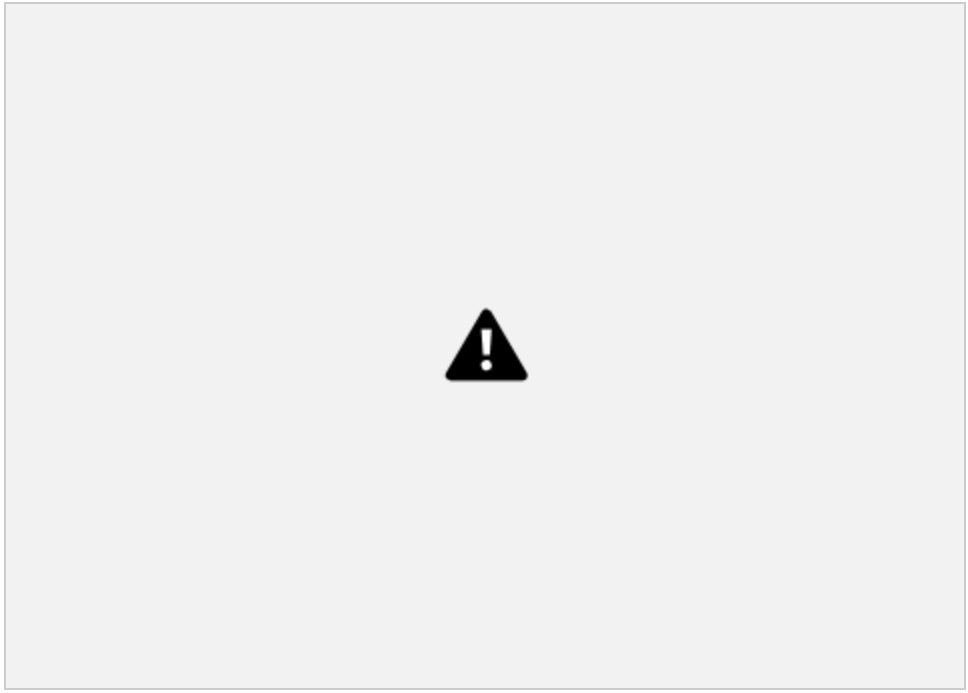
Swanson et al. 2017 *Epi*; see also: Robins 1998 *Stat Med*, Hernán et al. 2013 *AIM*, Swanson et al. 2015 *AJE,* Boef et al. 2015 *Epi*, Hernán et al. 2016 *JCE*, Hernán & Robins 2016 *AJE*

Selection bias in IV analyses



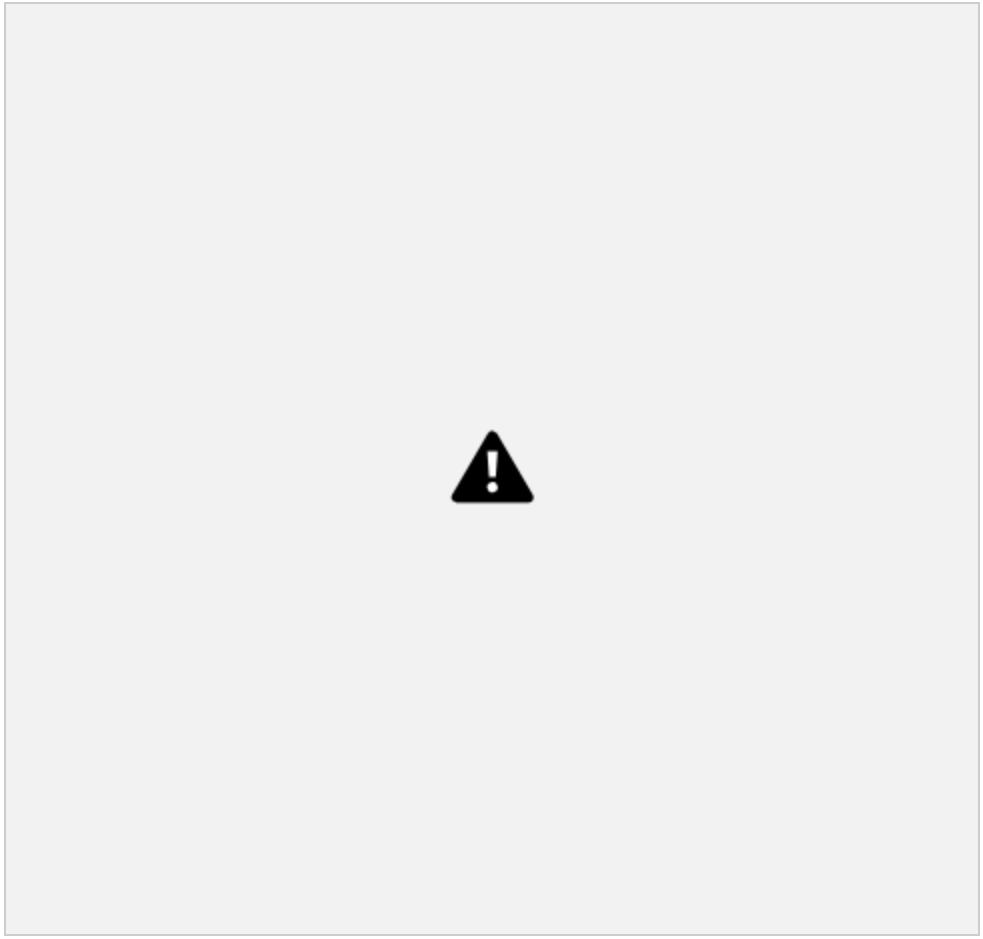
Swanson *Epi* 2019

Selection bias in IV analyses



Swanson *Epi* 2019

Selection bias in IV analyses



Swanson *Epi* 2019

Outline

- Motivation for IV methods
- A gentle introduction to IV methods
- Three underappreciated practical issues
  - Sensitivity/bias analyses and falsification strategies exist
  - An IV analysis is not a cure-all
  - Classical IV analyses are designed for point/time-fixed exposures
  - Summary

Classical IV analyses and time

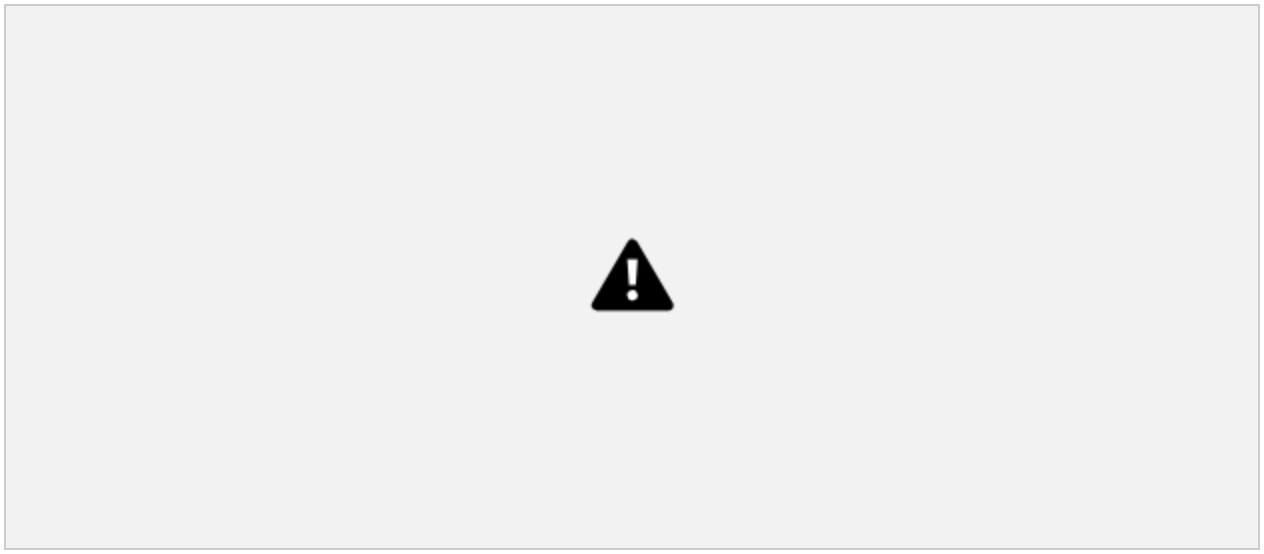
- Classical IV analyses are developed for a time-fixed or point exposure
- Can we use these classical IV analyses to study

sustained treatment strategies?

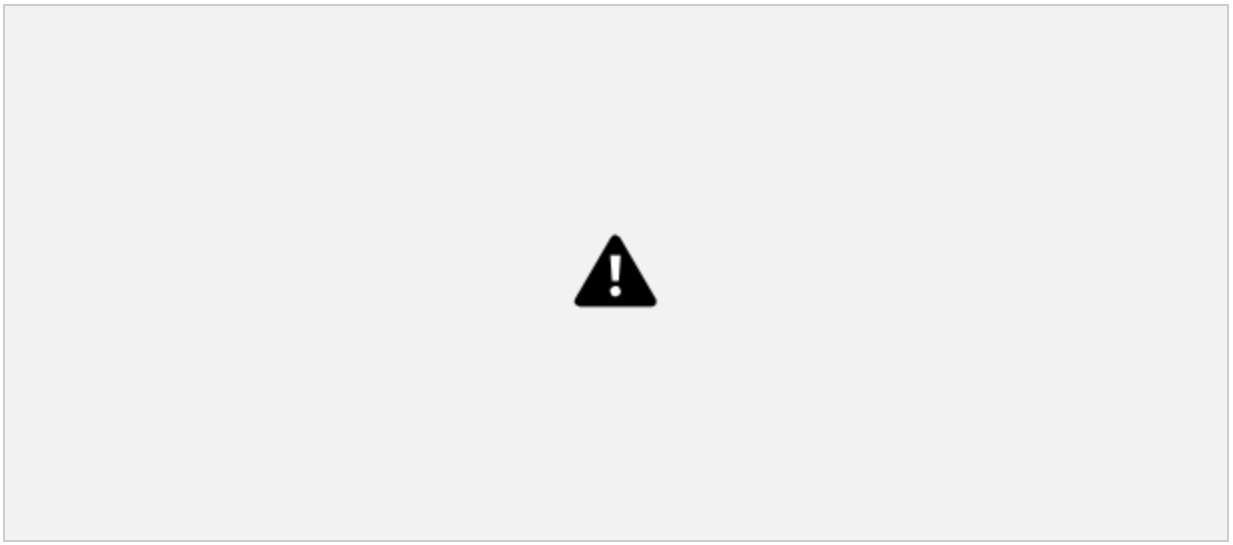
- Analogous to estimating the per-protocol effect of sustained continuous treatment vs. usual care by adjusting for non-adherence measured at a single moment in a long follow-up period

Swanson et al. *Epi* 2017, Labrecque & Swanson *AJE* 2018

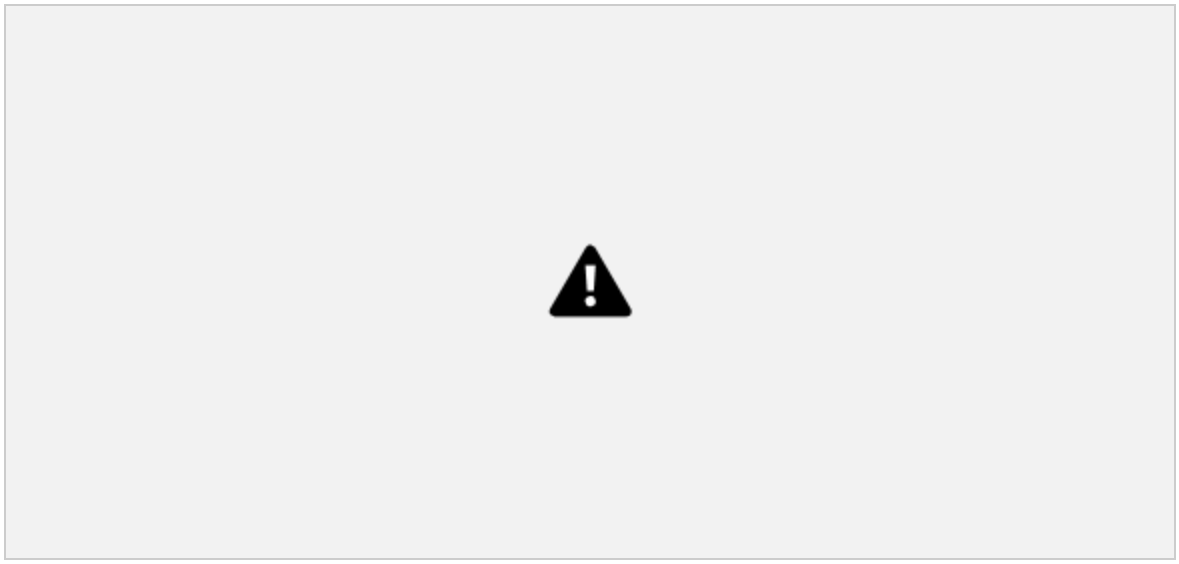
The problem simplified to two time

points 

The problem simplified to two time

points 

The problem simplified to two time

points 

Are these all “lifetime” effect estimates?

- MR studies often described as measuring “lifetime” effects
- Suppose we wanted to estimate the “lifetime” effect of BMI on systolic blood pressure proposing *FTO* as an IV
- Often

done with cross-sectional data but suppose we happened to have a cohort study with repeated assessments over several ages

- Using BMI measured at different ages in the Rotterdam Study, we get different effect estimates with MR:
  - BMI measured at age 55: 0.6 mmHg/kg/m^2
  - BMI measured at age 70: 4.7 mmHg/kg/m^2
  - Average BMI measurement: 2.2 mmHg/kg/m^2
  - How could these all be measuring "lifetime" effects?

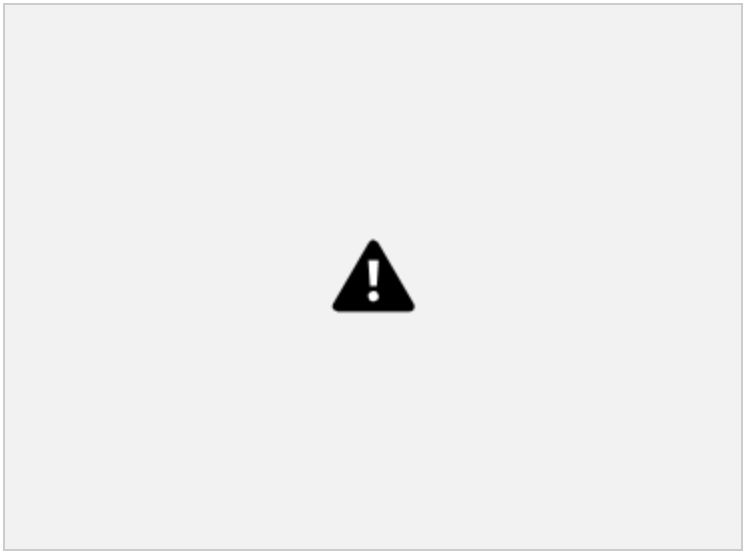
Labrecque & Swanson *AJE* 2018

Valid estimation of a lifetime effect?

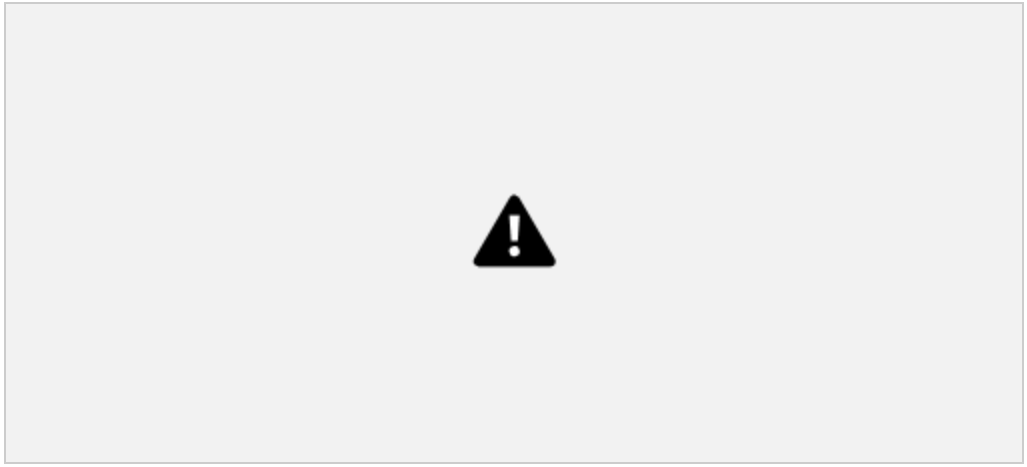
- First, need a clear definition of a lifetime effect
  - E.g., $\mathrm{E}[Y^{\bar{a}=\bar{1}}] - \mathrm{E}[Y^{\bar{a}=\bar{0}}]$
- With a clear definition, can begin to study possible

bias

- Simulations
- Bias formulas
- Identify bias-free special cases (e.g., *Z*-*A* effect constant by time) Labrecque & Swanson *AJE* 2018



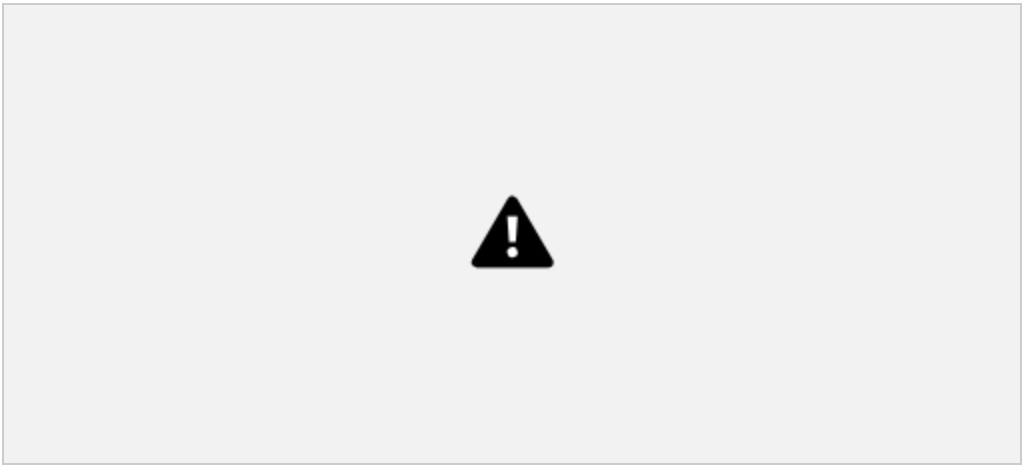
Valid testing of a lifetime effect?



- Even if valid estimation is not possible, could simply test
  - E.g., if Z and Y_1 are associated, then the joint sharp causal null hypothesis does not hold
  - But, if Z and Y_1 are associated, does not inform other causal null hypotheses (or effect sizes) without further assumptions

Swanson et al. 2018 *EJE*

Valid testing of a lifetime effect?



- Consider the canonical MR analyses of alcohol and CVD
  - Assuming valid IVs: for at least one person, changing alcohol consumption levels by some (unspecified) amount at some (unspecified) point in time would affect CVD risk
  - Estimating size of effect or testing more specific hypotheses requires more assumptions

Swanson et al. 2018 *EJE,* Chen et al. 2008 *PLoS Med*

Outline

- Motivation for IV methods
- A gentle introduction to IV methods
- Three underappreciated practical issues
  - Sensitivity/bias analyses and falsification strategies exist
  - An IV analysis is not a cure-all
  - Classical IV analyses are designed for point/time-fixed exposures
- Summary

Summary

- IV analyses have a place in the epidemiologists' toolbox, but, as with any tool, we need to:

- Collectively understand when the tool works well
- Continuously evaluate the tool's strengths and limitations
- Be aware of our own blindspots
- Interpret results appropriately

Select references

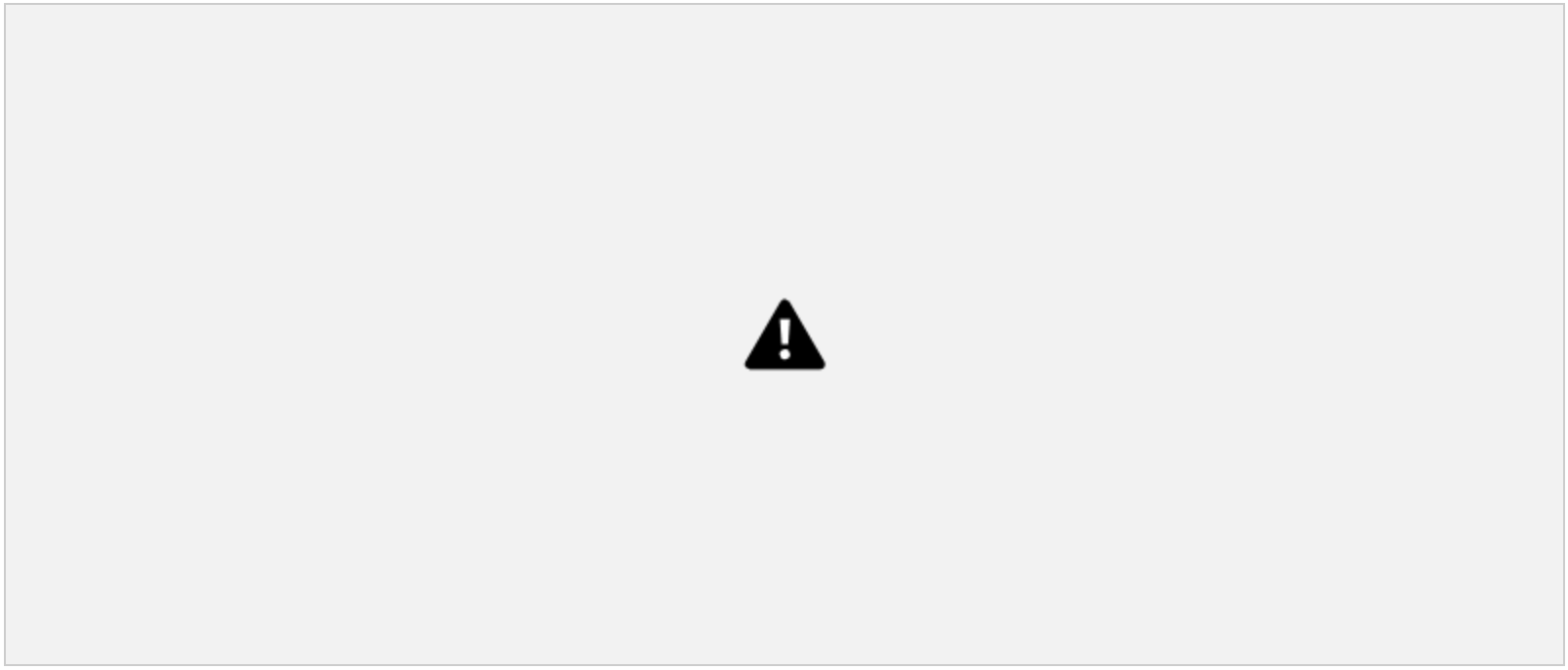
- **Questions? Email: s.swanson@erasmusmc.nl**

- Swanson SA, Hernán MA. How to report instrumental variable analyses (suggestions welcome). *Epidemiology* 2013;24(3):370-4.
- Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a trialist? Deconstructing the analogy between Mendelian randomization and randomized trials. *Epidemiology* 2017;28(5):653-9.
- Swanson SA. Instrumental variable analyses in pharmacoepidemiology: What target trials do we emulate? *Current Epidemiology Reports* 2017;4(4):281-7.
- Labrecque JA, Swanson SA. Understanding the assumptions underlying instrumental variable analyses: A brief review of falsification strategies and related tools. *Current Epidemiology Reports* 2018;5(3):214-20.
- Swanson SA, Labrecque JA, Hernán MA. Causal null hypotheses of sustained treatment strategies: What can be tested with an instrumental variable? *European Journal of Epidemiology* 2018;33(8):723-8.
- Labrecque JA, Swanson SA. Interpretation and potential biases of Mendelian randomization estimates with time-varying exposures. *American Journal of Epidemiology* 2018;188(1):231-8.
- Swanson SA, Hernán MA, Miller M, Robins JM, Richardson TS. Partial identification of the average treatment effect using instrumental variables. *Journal of the American Statistical Association* 2018;113(522):933-47.
- Swanson SA. A practical guide to selection bias in instrumental variable analyses. *Epidemiology* 2019 in press.

MR: Genetic variants as proposed IVs

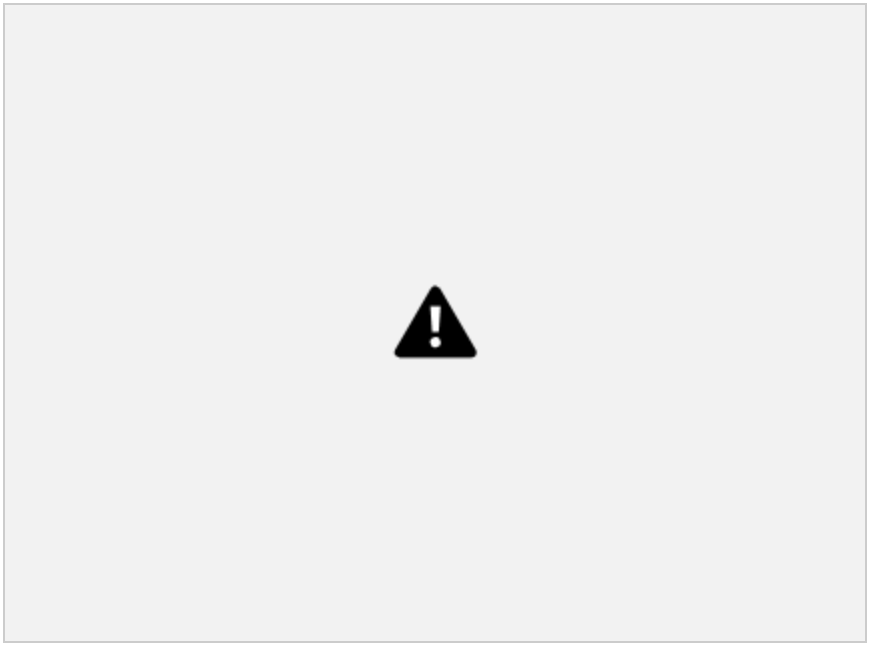
- Instrumental conditions:

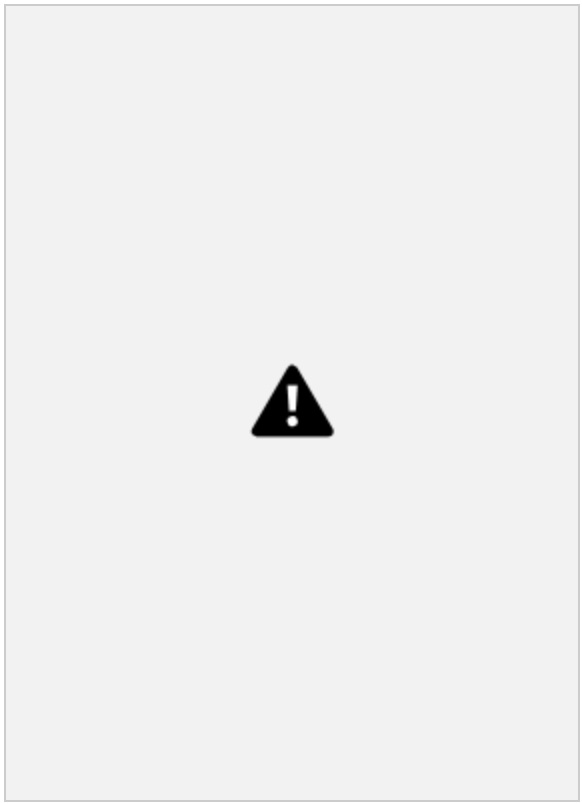
1. Genetic variant and exposure are associated
2. Genetic variant causes outcome only through exposure 3. Genetic variant and outcome share no causes



Z = instrument; *A* = treatment; *Y* = outcome; *U* = unmeasured covariates

Effects of BMI-related variants by age

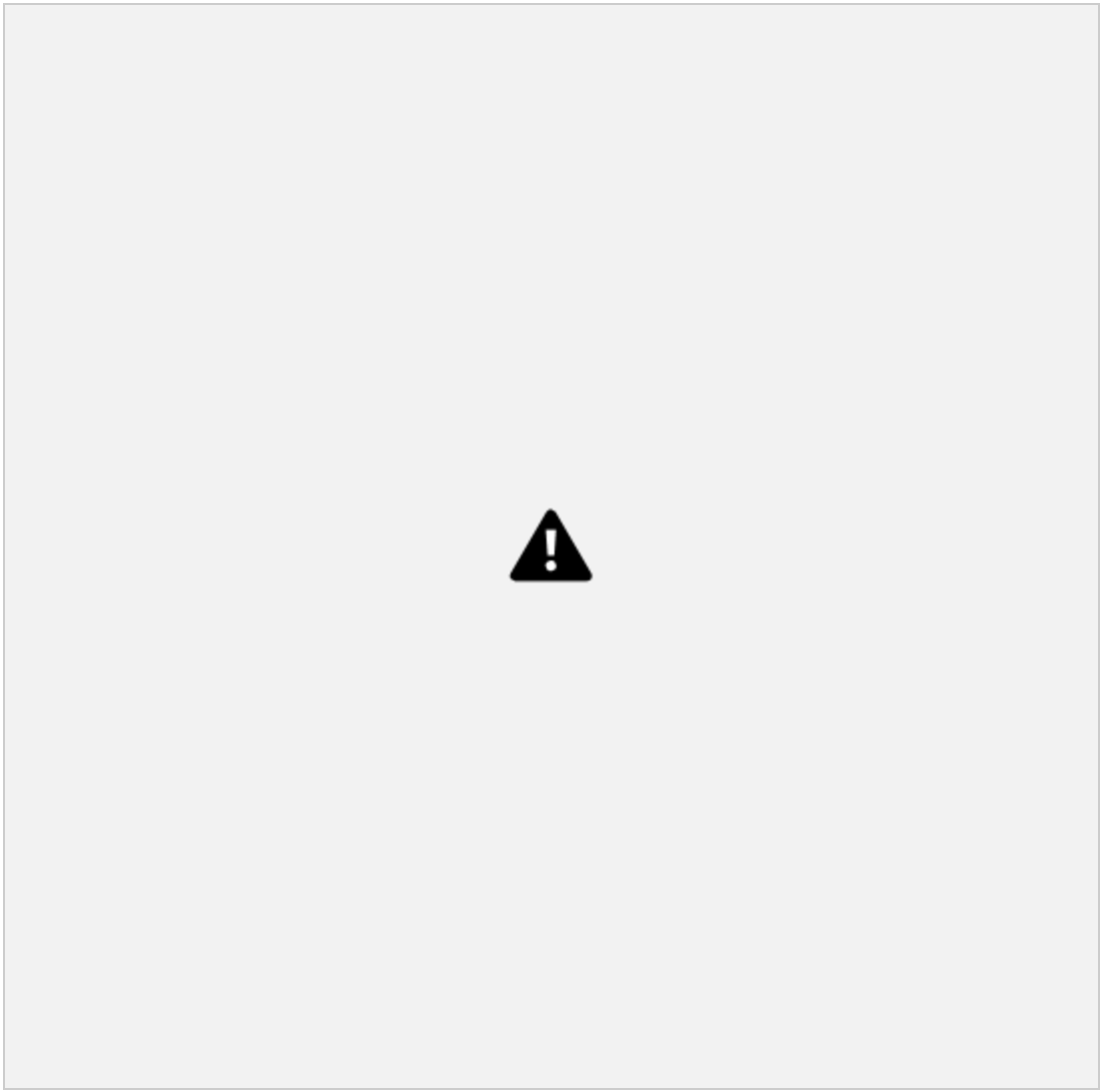




- Simulations based on *FTO*-BMI-SBP relationship in Rotterdam Study suggest relative bias of >50%

Labrecque & Swanson 2018 *AJE*, Winkler et al. 2015 *PLOS Genetics*

Effects of alcohol-related variants by age



"Fourth" assumptions: monotonicity

- Under a monotonicity assumption, IV methods estimate a causal effect in only a subgroup of the study population
  - Local average treatment effect (LATE)
    - Complier average causal effect (CACE)
    - Angrist, Imbens, & Rubin 1996 *JASA*

Compliance types in the context of a trial

Randomized to treatment arm (Z=1)

Treated ($A^{z=1}$=1) Not treated ($A^{z=1}$=0)

taker ($A^{z=0}>A^{z=1}$)

($A^{z=0}=A^{z=1}=1$)

Randomi zed to placebo arm (Z=0)

Treated ($A^{z=0}$=1)

Not treated ($A^{z=0}$=0)

Always

Complier ($A^{z=0}<A^{z=1}$)

Defier

Never-taker ($A^{z=0}=A^{z=1}$=0)

Compliance types: any causal IV *Z*

Z=1

$A^{z=1}$=1 $A^{z=1}$=0

$(A^{z=0}=A^{z=1}=1)$

$Z=0$

$A^{z=0}=1$ **Always taker**

Defier
$(A^{z=0}>A^{z=1})$

Never-taker
$(A^{z=0}=A^{z=1}=0)$

$A^{z=0}=0$ **Complier**
$(A^{z=0}<A^{z=1})$

Example of compliance types: MR Has

the genetic variant ($Z=1$)

		$(A^{z=1}=1)$	$(A^{z=1}=0)$
Does not have the genetic variant ($Z=0$) Obese ($A^{z=0}=1$)	Not obese ($A^{z=0}=0$) Obese	**Always taker** ($A^{z=0}=A^{z=1}=1$)	**Defier** ($A^{z=0}>A^{z=1}$)
		Complier ($A^{z=0}<A^{z=1}$) Not obese	**Never-taker** ($A^{z=0}=A^{z=1}=0$)

Problems with LATE approaches

- For many commonly proposed IVs, monotonicity may not be reasonable
- For non-causal IVs, or the meaning of this effect is ambiguous at best

- Even in the ideal situation of a causal well-defined IV with all assumptions holding perfectly, the effect in the "compliers" is not directly of policy or clinical interest
  - We don't even know who they are!
    - Definition of group is study-specific and instrument-dependent

Prenatal MR trial analogue

- Prenatal MR has unclear time zero

2. Eligibility criteria applied (e.g., mother's pregnancy)

Exposure level assessed during pregnancy

1. Genetic variant inherited (mother's conception)

(postnatal)

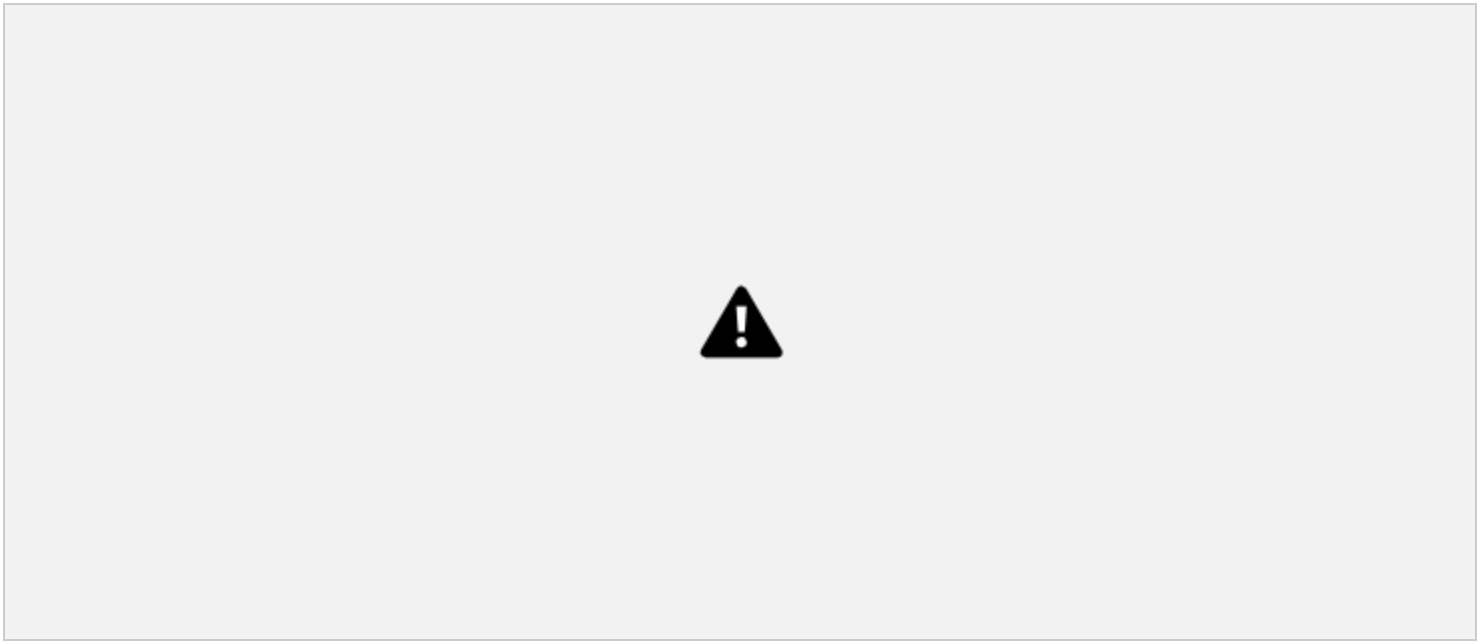
Time zero

3. Outcome recording

End of study period

Two unique concerns with prenatal MR

- Selection bias related to “time zero” misalignment
- “Pleiotropic” effect of postnatal exposure

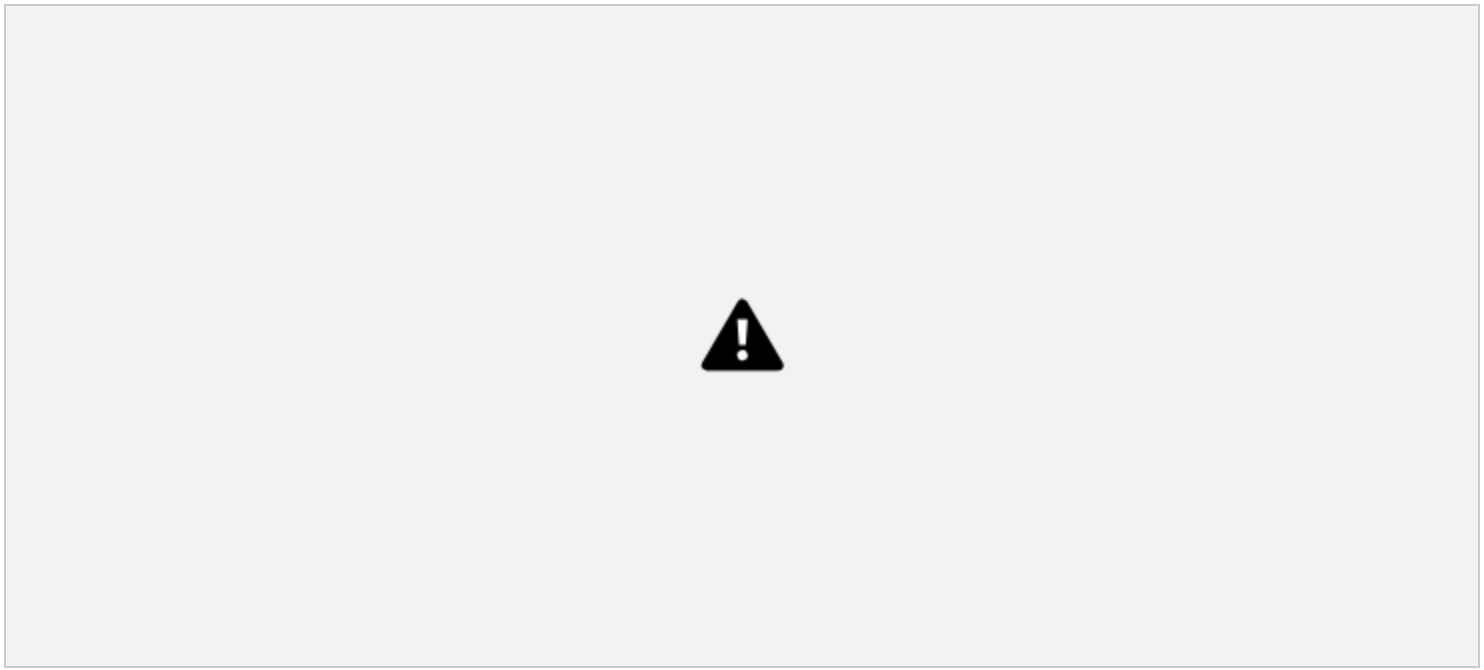


Two unique concerns with prenatal MR

- Selection bias related to “time zero” misalignment

E.g., if pre-pregnancy exposure affects fertility

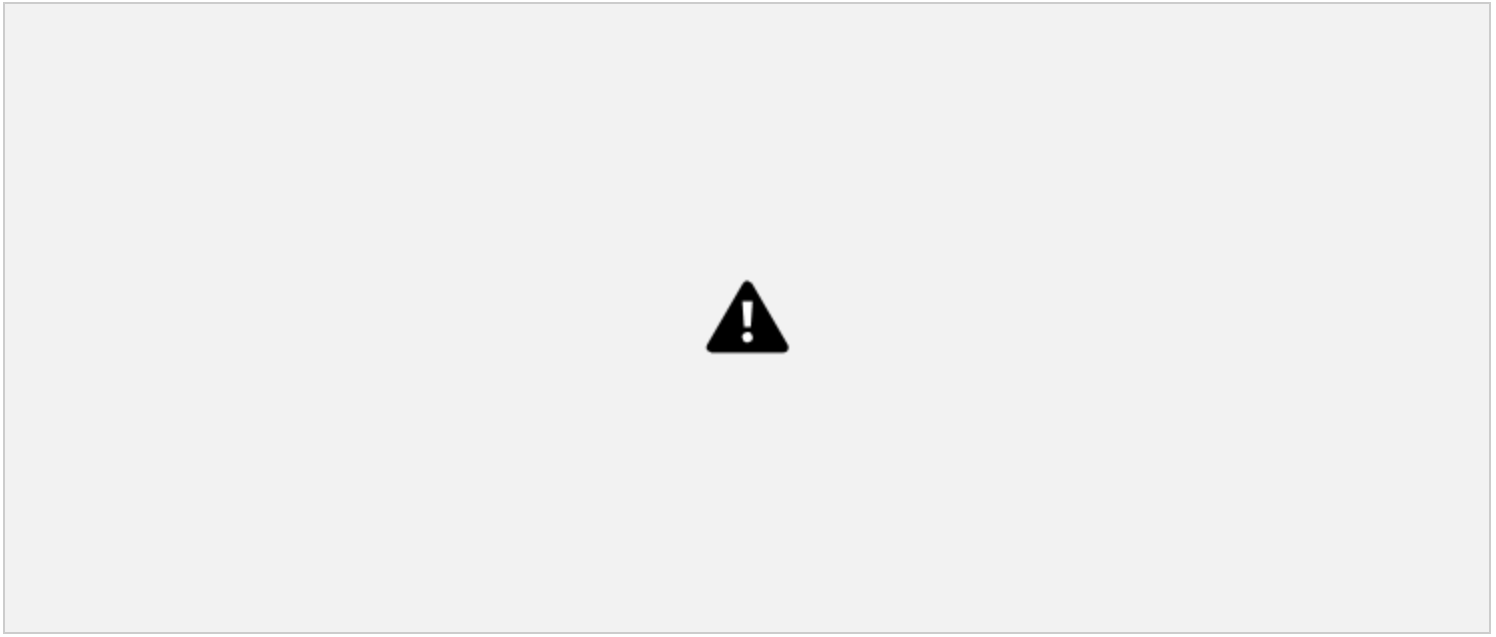
- Challenging bias to address because prenatal MR studies are usually conducted in birth cohort studies



Two unique concerns with prenatal MR

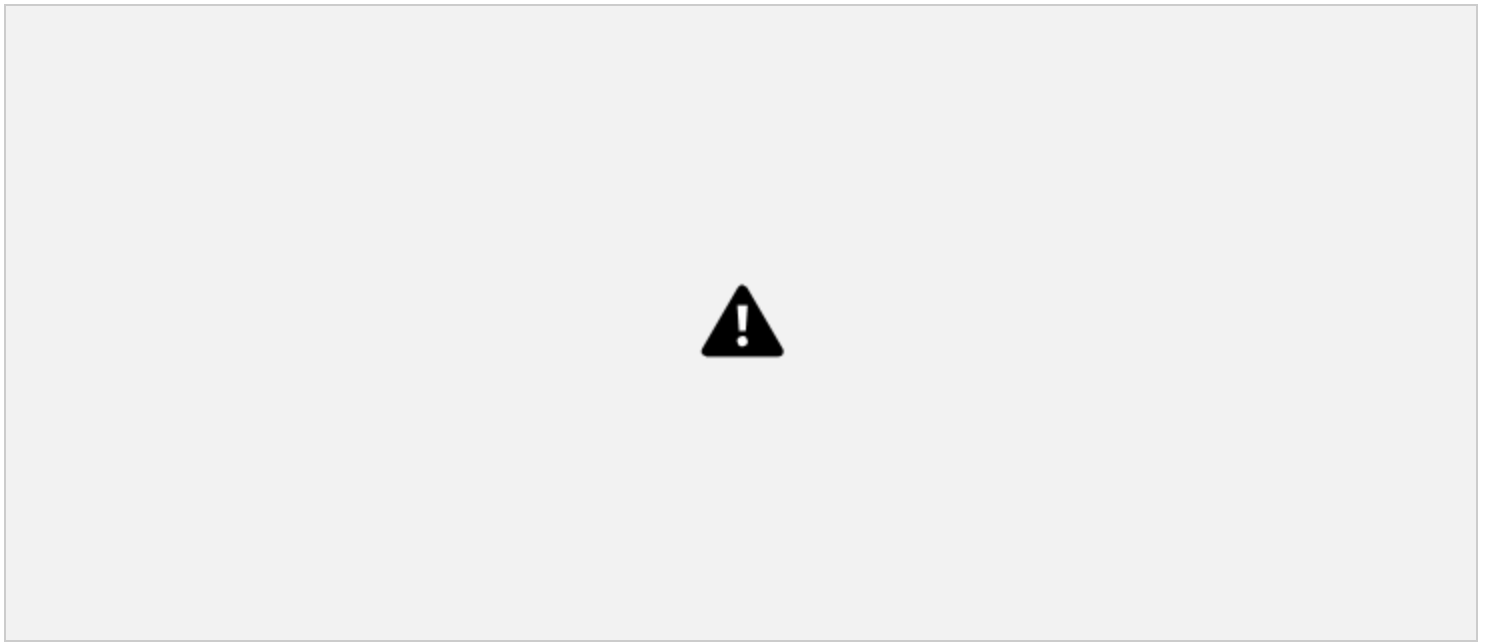
- “Pleiotropic” effect of postnatal exposure
  - E.g., if mother’s behavior post-pregnancy affects outcome

Challenging bias to address because postnatal exposure may be confounded by the same confounders that motivated use of MR



Approaching these biases

- Need subject matter expertise to consider if these (or other!) biases are plausible in prenatal MR
- Can consider using available IV falsification strategies to identify and maybe quantify the bias if suspected

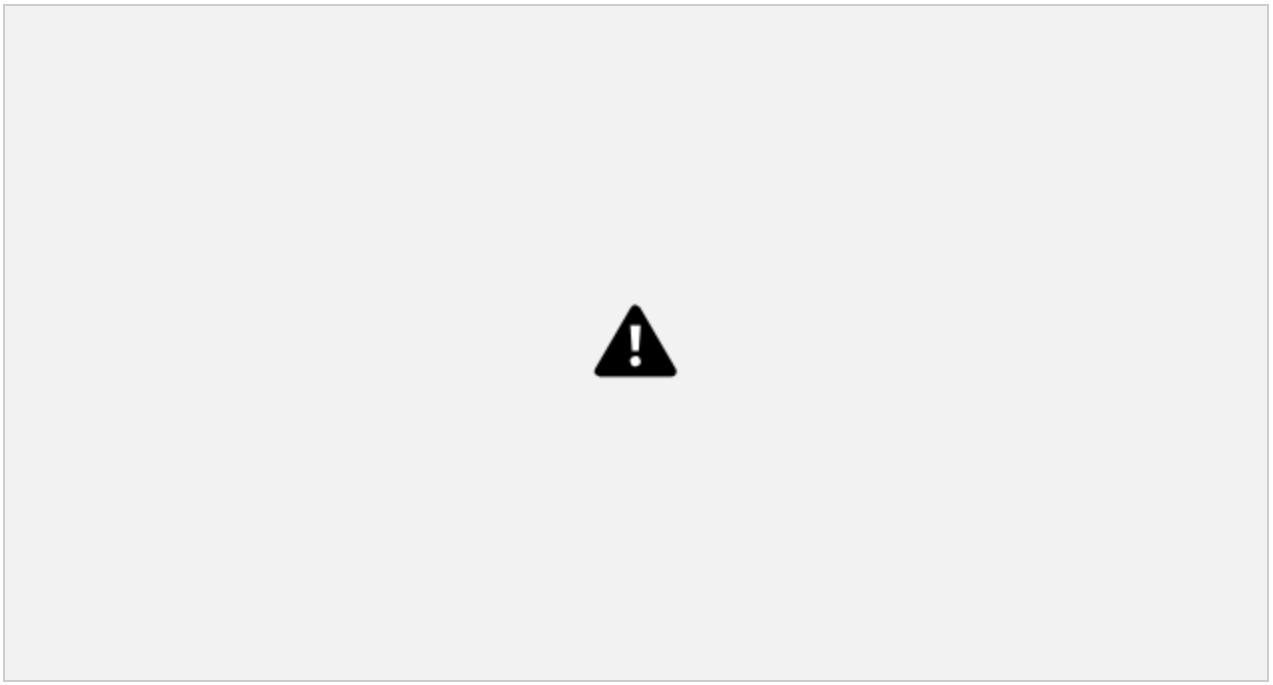


Balke & Pearl 1997 *JASA*, Glymour et al. 2012 *AJE*, Jackson & Swanson 2015 *Epi* , Swanson 2017 *Epi*, Labrecque & Swanson 2018 *CER*

MR as a test of causal null hypotheses

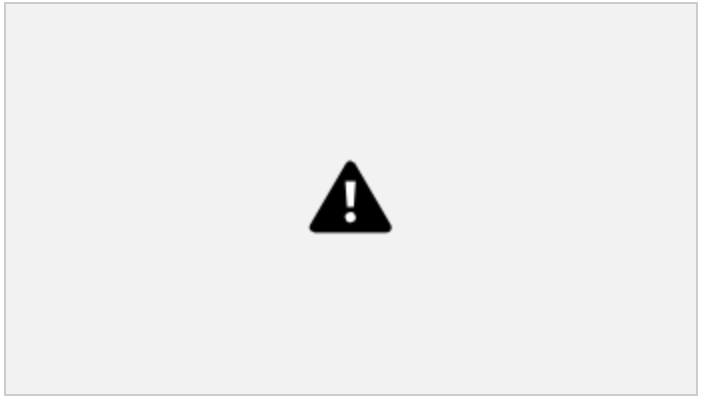
- Some investigators propose MR's goal is to test "the" causal null hypothesis
  - But which causal null hypotheses are we testing?
    - Analogous to trial ITT analyses being a test of treatment

effects?



Burgess et al. 2017 *Epi*

Sharp causal null (time-fixed



exposure)

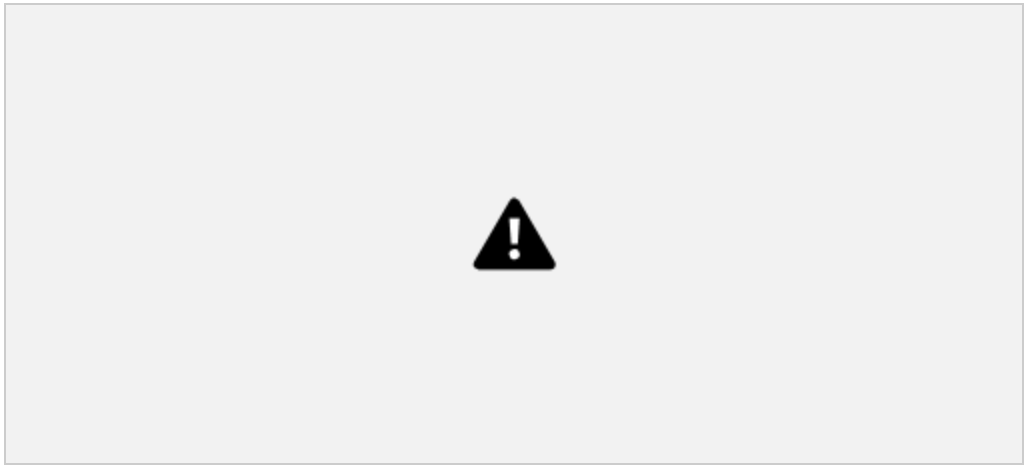
- Sharp causal null hypothesis of *A* on *Y* holds
  - $Y_i^{A=1} = Y_i^{A=0}$ for all individuals *i*
- *Z* and *Y* are d-separated
- Thus, if *Z* and *Y* are associated, then one or more of the following must be true:
  - *Z* is not an instrument
  - The sharp causal null hypothesis does not hold

Average causal null (time-fixed

exposure)

- Average causal null hypothesis of *A* on *Y* holds
  - $E\left[Y^{a=1}\right] = E\left[Y^{a=0}\right]$
- If *Z* and *Y* are associated, then we generally cannot conclude whether the average effect of *A* on *Y* is non-null
  - Need an additional assumption, e.g., monotonic treatment effect

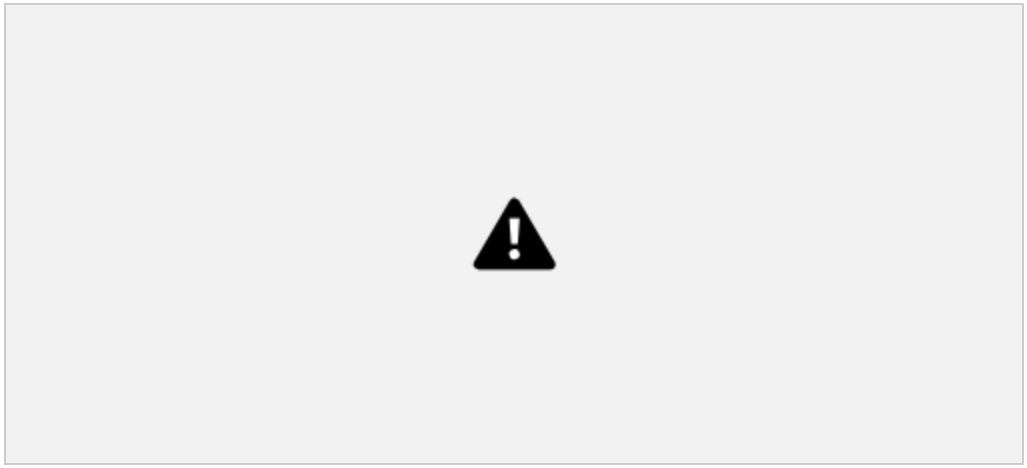
Joint sharp causal null



- Joint sharp causal null hypothesis of (A_0, A_1) on Y_1 holds
  - Z and Y_1 are d-separated
- Thus, if Z and Y_1 are associated, then one or more of the following must be true:
  - Z is not an instrument
  - The joint sharp causal null hypothesis does not hold

Swanson et al. 2018 *EJE*

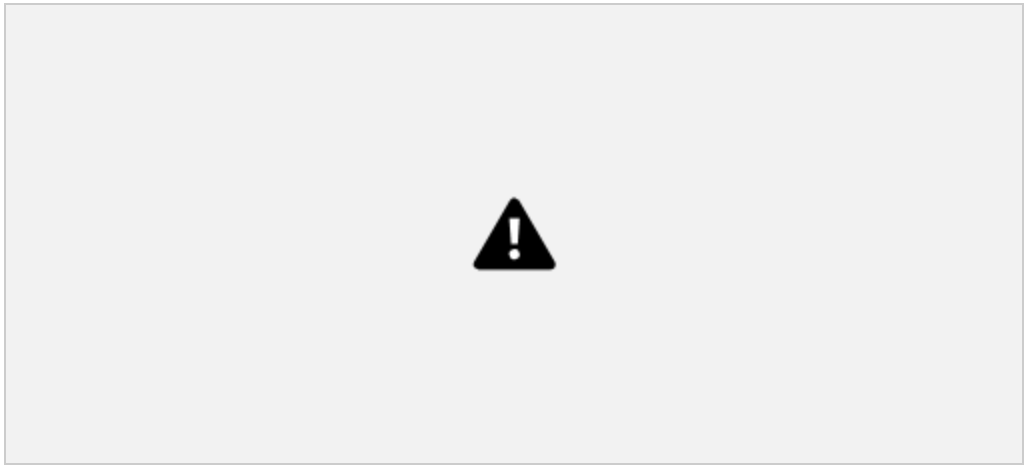
Non-joint sharp causal nulls



- If Z and Y_1 are associated, then we generally cannot conclude whether A_0 or A_1 or both have an effect
- Z is an instrument for the joint, not each separately

Swanson et al. 2018 *EJE*

Joint average causal null



- If Z and Y_1 are associated, then we generally cannot conclude whether the average effect of (A_0, A_1) on Y_1 is non-null
  - Need an additional assumption, e.g., monotonic treatment

effect Swanson et al. 2018 *EJE*