CIMPOD 2017 – Day 1
Instrumental Variable (IV) Methods

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

Big picture overview

- Motivation for IV methods

- Key assumptions for identifying causal effects with IVs
  - Day 1: Per-protocol effects in trials with non-compliance
  - Day 2: Effects of initiating treatment in observational studies

- IV estimation and tools for understanding possible threats to validity
  - Day 1: Bounding, instrumental inequalities…
  - Day 2: Weak IVs, bias component plots…

- Extensions and further considerations

- Summary and Q&A

Some disclaimers

- My emphasis will be on addressing the following questions

  1. What are we hoping to estimate, and what can we actually estimate?

  2. Are the assumptions required to interpret our estimates as causal effects reasonable?

  3. Under plausible violations of these assumptions, how sensitive are our estimates?

- Provided R code will emphasize #2 and #3, as well as examples of how to implement IV estimation

- Ask questions!

Case study (Day 1): Swanson 2015 *Trials*

TRIALS

METHODOLOGY

Open Access

CrossMark

Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening

Sonja A. Swanson[1,2*], Øyvind Holme[3,4], Magnus Løberg[3,6], Mette Kalager[2,3,5], Michael Bretthauer[2,3,6], Geir Hoff[3,5,7], Eline Aas[3] and Miguel A. Hernán[2,8,9]

Case study (Day 2): Swanson 2015 *PDS*

ORIGINAL REPORT

Methodological considerations in assessing the effectiveness of antidepressant medication continuation during pregnancy using administrative data[†]

Sonja A. Swanson[1]*, Sonia Hernandez-Diaz[1], Kristin Palmsten[2], Helen Mogun[3], Mark Olfson[4] and Krista F. Huybrechts[3]

Overview

- Motivation for IV methods

- Key assumptions for identifying causal effects with IVs

- IV estimation and tools for understanding possible threats to validity

- Extensions and further considerations

- Summary and Q&A

Motivation for IV methods

- Most methods for causal inference rely on the assumption that there is no unmeasured confounding

  - Regression, propensity score methods, and other forms of stratification, restriction, or matching

  - G-methods (inverse probability weighting, parametric g-formula, usual form of g-estimation of structural nested models)

- HUGE assumption

- Dream with me: what if we could make causal inferences without this assumption?

- More specifically…
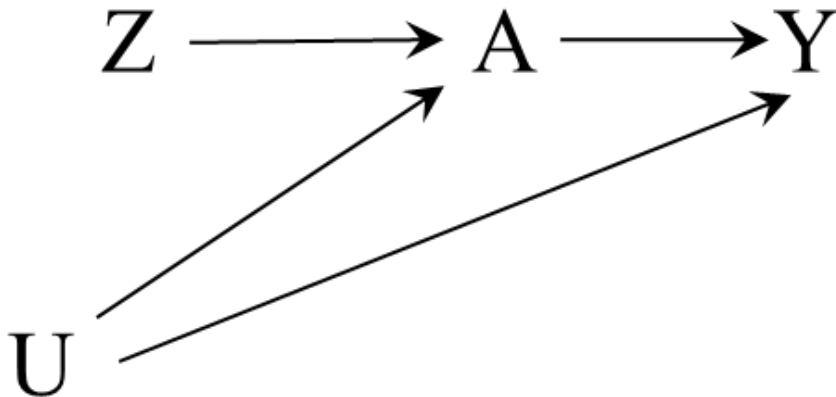
Problem #1: trials with non-compliance

- First, consider a hypothetical double-blind, placebo-controlled, single-dose randomized trial with complete follow-up
    - But with non-compliance
- We can readily estimate the intention-to-treat (ITT) effect
    - The effect of randomization
- But the ITT effect is hard to interpret because it critically depends on the degree of adherence

Problem #1: trials with non-compliance and estimating per-protocol effects

- We may be interested in a per-protocol effect
  - The effect of following the protocol (i.e., of actual treatment)
- How can we estimate a per-protocol effect?
  - This effect is confounded!
  - Usual strategies analyze the randomized trial data like an observational study, adjusting for measured confounders
  - IV methods offer an alternative strategy

$$Z \longrightarrow A \longrightarrow Y$$

with $U \rightarrow A$ and $U \rightarrow Y$

Problem #1: trials with non-compliance and our case study

- Consider the NORCCAP pragmatic trial of colorectal cancer screening vs. no screening
  - We may be interested in a per-protocol effect of screening versus no screening

- How can we estimate a per-protocol effect?
  - This effect is confounded!
  - Usual strategies analyze the randomized trial data like an observational study, adjusting for measured confounders
  - IV methods offer an alternative strategy

Swanson et al. 2015 *Trials*

Problem #2: observational studies with unmeasured confounding

- Often observational studies are our only hope for estimating treatment effects

- Treatment effects can be confounded (e.g., by indication)

  - Usual methods for analyzing treatment effects in observational studies rely on measuring and appropriate adjusting for confounders

  - IV methods offer an alternative strategy

Problem #2: observational studies with unmeasured confounding and our case study

- Suppose we want to estimate the risks and benefits of continuing antidepressant medication use during pregnancy among women with depression
  - Observational studies may be our best hope
- Treatment effects could be confounded by depression severity, healthy behaviors, etc.
  - Usual methods for analyzing treatment effects would require we measure (or come very close to approximating) these confounders
  - IV methods offer an alternative strategy

Overview

- Motivation for IV methods

- Key assumptions for identifying causal effects with IVs

- IV estimation and tools for understanding possible threats to validity

- Extensions and further considerations

- Summary and Q&A

Some notation

- Z: proposed instrument (defined on next slide)

- A: treatment

- Y: outcome

- U, L: unmeasured/measured relevant covariates

- Counterfactual notation: $E[Y^a]$ denotes the average counterfactual outcome Y had everybody in our study population been treated with $A=a$

IV conditions

1. Instrument and treatment are associated
2. Instrument causes the outcome only through treatment
3. Instrument and outcome share no causes
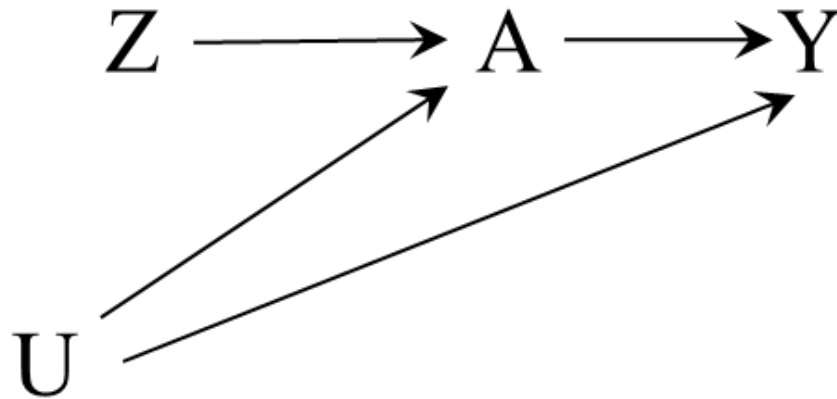
IV conditions

1. Instrument and treatment are associated
2. Instrument causes the outcome only through treatment
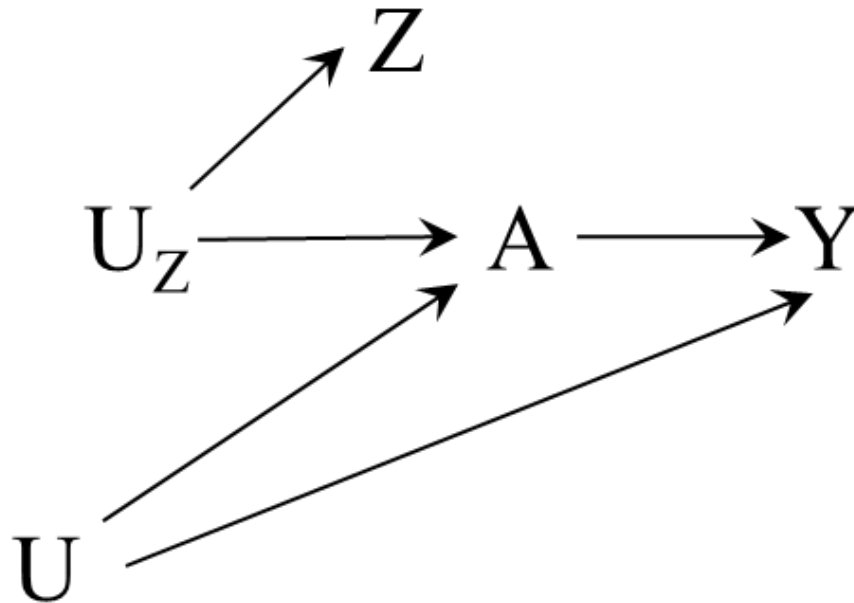3. Instrument and outcome share no causes

IV conditions

1. Instrument and treatment are associated
2. Instrument causes the outcome only through treatment
3. Instrument and outcome share no causes

Under these conditions, we can use the standard IV ratio or related methods to identify treatment effects

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[A|Z=1] - E[A|Z=0]}$$

IV methods in randomized trials

The randomization indicator as a proposed instrument to help estimate a per-protocol effect (focus of Day 1)

1. Randomization indicator and treatment are associated
2. Randomization indicator causes the outcome only through treatment
3. Randomization indicator and outcome share no causes
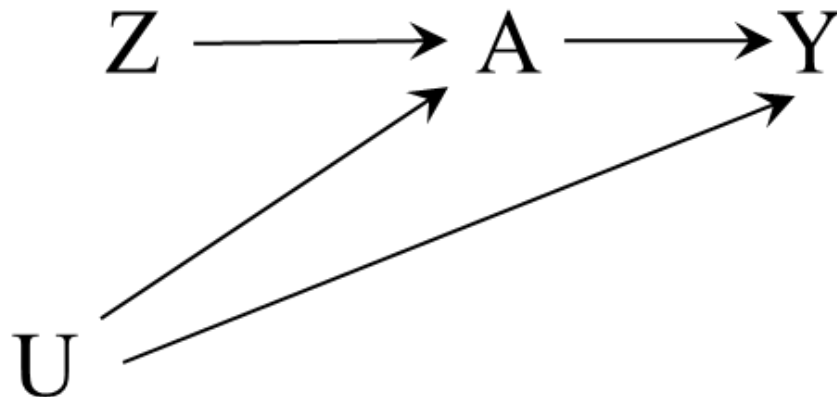
IV methods in observational studies

- Propose/find a "natural experiment" measured in your observational study that meets the IV conditions (focus of Day 2)

- Commonly proposed IVs in PCOR

  - Physician or facility preference

  - Calendar time

  - Geographic variation

Example of a proposed IV: preference

Propose physician/facility preference (e.g., as measured via prescriptions to prior patients) as an IV

1. Preference and patients' treatments are associated
2. Preference affects outcomes only through treatment
3. Preference and outcome share no causes

Example of a proposed IV: geographic variation

Propose geographic variation as an IV

1. Location and patients' treatments are associated

2. Location affects outcomes only through treatment

3. Location and outcome share no causes

Example of a proposed IV: calendar time

Propose pre- versus post-warning calendar period as an IV

1. Calendar period and patients' treatments are associated
2. Calendar period related to patient outcomes only through treatment
3. Calendar period and outcome share no causes

The ideal: calendar time as a proposed IV

Hypothetical Treatment Probability by Calendar Time
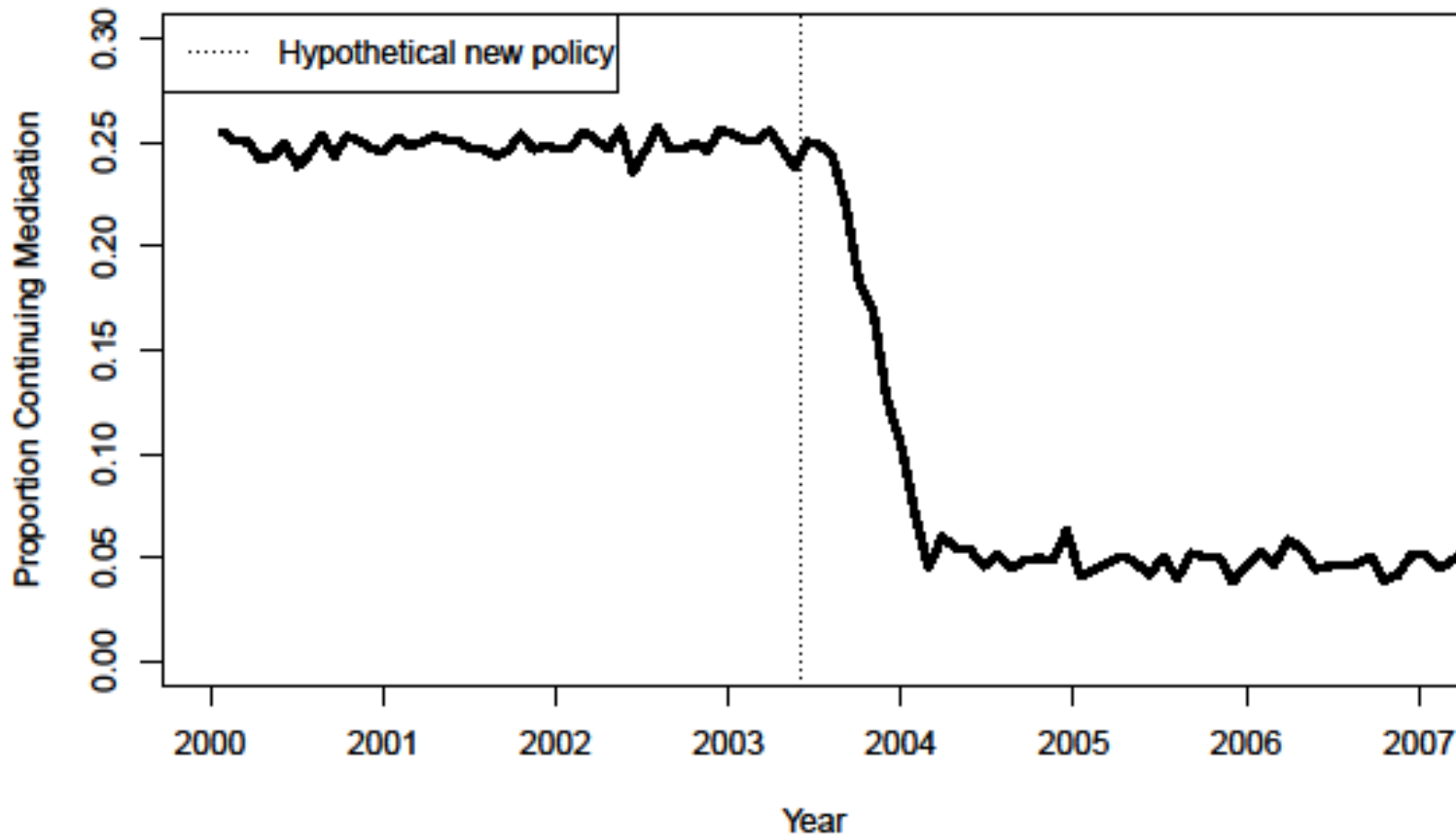
The reality: calendar time as a proposed IV

Treatment Continuation Probability by Calendar Time
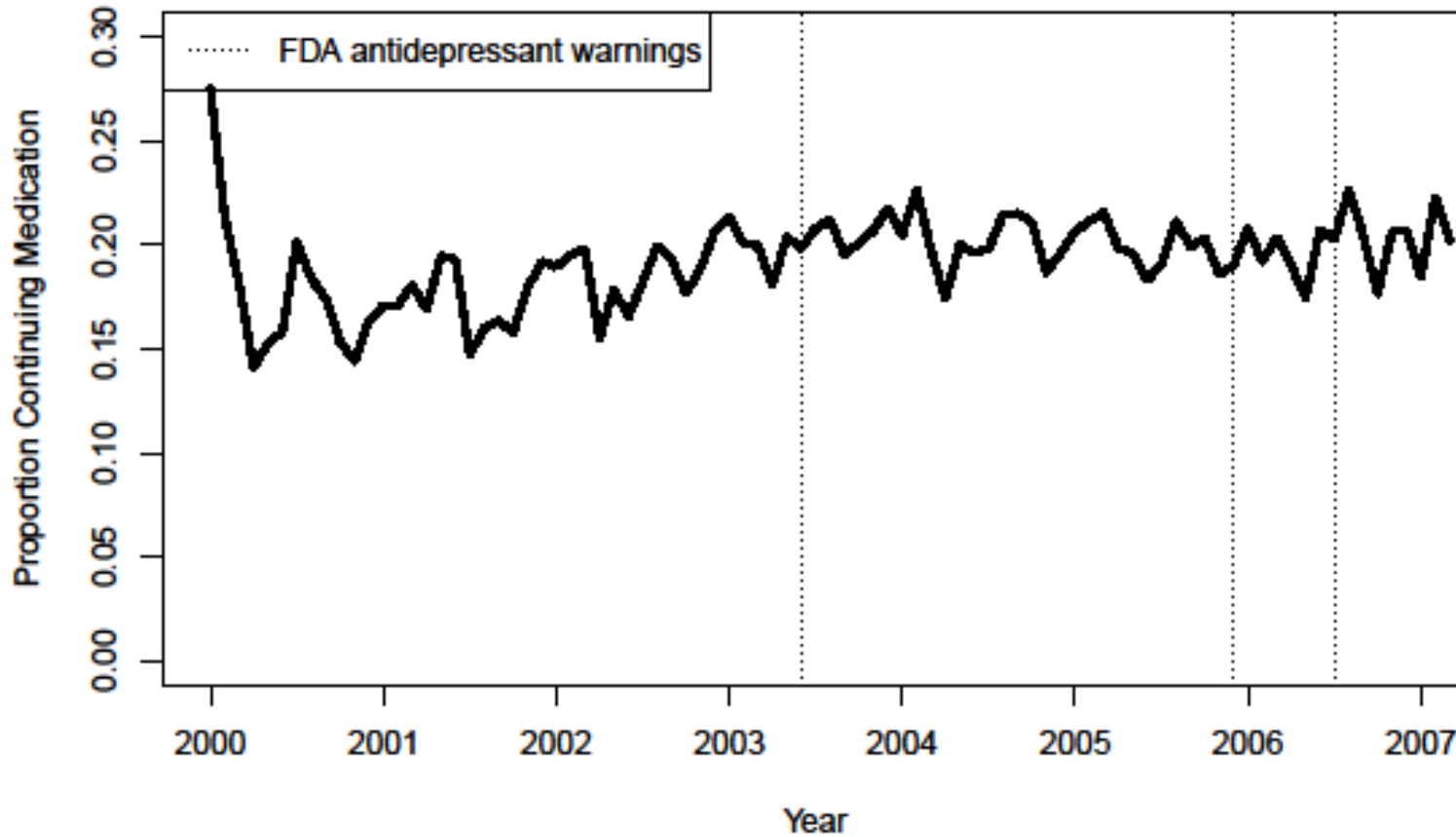
However, an IV not enough

- With only these three conditions that define an IV, we cannot generally obtain a point estimate for a causal effect
  - Can estimate "bounds"

- What does the standard IV methods estimate then?
  - Depends on what further assumptions we are willing to make

"Fourth" assumptions: homogeneity

- Under strong homogeneity assumptions, IV methods estimate the average causal effect

$$E[Y^{a=1} - Y^{a=0}] = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[A|Z=1] - E[A|Z=0]}$$

- Most extreme type of homogeneity assumption: constant treatment effect

  - $Y^{a=1} - Y^{a=0}$ is the same for all individuals

- Less extreme (but still strong) version: no additive effect modification by the IV among the treated and untreated

  - $E[Y^{a=1} - Y^{a=0}|Z=1, A=1] = E[Y^{a=1} - Y^{a=0}|Z=0, A=1]$
  - $E[Y^{a=1} - Y^{a=0}|Z=1, A=0] = E[Y^{a=1} - Y^{a=0}|Z=0, A=0]$

"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 ($A^{z=1}=0$)
Random-ized to placebo arm ($Z=0$)	Treated ($A^{z=0}=1$)	**Always-taker** ($A^{z=0}=A^{z=1}=1$)	**Defier** ($A^{z=0}>A^{z=1}$)
	Not treated ($A^{z=0}=0$)	**Complier** ($A^{z=0}<A^{z=1}$)	**Never-taker** ($A^{z=0}=A^{z=1}=0$)

Compliance types: any causal IV Z

$Z=1$

	$A^{z=1}=1$	$A^{z=1}=0$
$A^{z=0}=1$	**Always-taker** $(A^{z=0}=A^{z=1}=1)$	**Defier** $(A^{z=0}>A^{z=1})$
$A^{z=0}=0$	**Complier** $(A^{z=0}<A^{z=1})$	**Never-taker** $(A^{z=0}=A^{z=1}=0)$

$Z=0$

Compliance types: preference

Prefers treatment ($Z=1$)

		Treated ($A^{z=1}=1$)	Not treated ($A^{z=1}=0$)
Prefers no treatment ($Z=0$)	Treated ($A^{z=0}=1$)	**Always-taker** ($A^{z=0}=A^{z=1}=1$)	**Defier** ($A^{z=0}>A^{z=1}$)
	Not treated ($A^{z=0}=0$)	**Complier** ($A^{z=0}<A^{z=1}$)	**Never-taker** ($A^{z=0}=A^{z=1}=0$)

Compliance types: geographic variation

Location with high treatment rate
($Z=1$)

		Treated ($A^{z=1}=1$)	Not treated ($A^{z=1}=0$)
Location with low treatment rate ($Z=0$)	Treated ($A^{z=0}=1$)	**Always-taker** ($A^{z=0}=A^{z=1}=1$)	**Defier** ($A^{z=0}>A^{z=1}$)
	Not treated ($A^{z=0}=0$)	**Complier** ($A^{z=0}<A^{z=1}$)	**Never-taker** ($A^{z=0}=A^{z=1}=0$)

Compliance types: calendar time

Post-warning period (Z=1)

		Treated ($A^{z=1}$=1)	Not treated ($A^{z=1}$=0)
Pre-warning period (Z=0)	Treated ($A^{z=0}$=1)	**Always-taker** ($A^{z=0}=A^{z=1}=1$)	**Defier** ($A^{z=0}>A^{z=1}$)
	Not treated ($A^{z=0}$=0)	**Complier** ($A^{z=0}<A^{z=1}$)	**Never-taker** ($A^{z=0}=A^{z=1}=0$)

Monotonicity and the LATE

- Under the IV conditions plus assuming there are no defiers (monotonicity), we can estimate the effect in the compliers
    - The local average treatment effect (LATE)
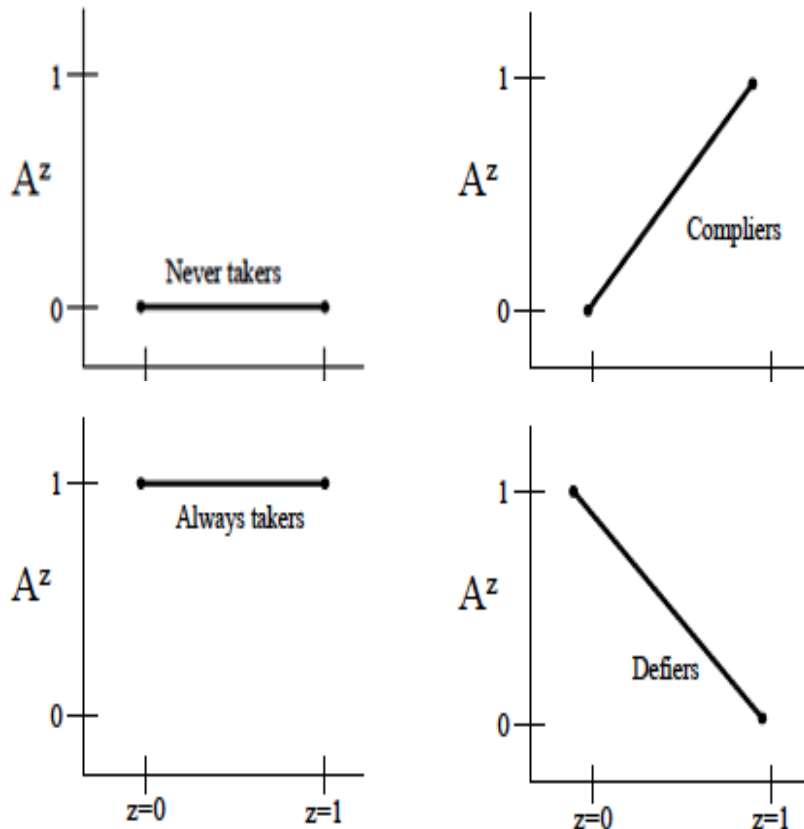
$$E[Y^{a=1} - Y^{a=0} | A^{z=0} < A^{z=1}]$$

$$= \frac{E[Y|Z=1] - E[Y|Z=0]}{E[A|Z=1] - E[A|Z=0]}$$

Identification of LATE: sketch of proof (1)

- The ITT effect is a weighted average of the ITT effects in our four compliance types

$$E[Y^{z=1}\text{-}Y^{z=0}] =$$

$$E[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}<A^{z=1}]\Pr[A^{z=0}<A^{z=1}] \qquad \text{(compliers)}$$

$$+ E[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}=A^{z=1}=1]\Pr[A^{z=0}=A^{z=1}=1] \qquad \text{(always-takers)}$$

$$+ E[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}=A^{z=1}=0]\Pr[A^{z=0}=A^{z=1}=0] \qquad \text{(never-takers)}$$

$$+ E[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}>A^{z=1}]\Pr[A^{z=0}>A^{z=1}] \qquad \text{(defiers)}$$

Identification of LATE: sketch of proof (2)

- Because an always-taker would always take treatment regardless of what she was randomized to, the effect of randomization in this subgroup is 0

$$\mathrm{E}[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}\text{=}A^{z=1}\text{=}1] = \mathrm{E}[Y^{a=1}\text{-}Y^{a=1}|A^{z=0}\text{=}A^{z=1}\text{=}1] = 0$$

- Similar logic applies to the never-takers

$$\mathrm{E}[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}\text{=}A^{z=1}\text{=}0] = \mathrm{E}[Y^{a=0}\text{-}Y^{a=0}|A^{z=0}\text{=}A^{z=1}\text{=}0] = 0$$

Identification of LATE: sketch of proof (3)

- Because a complier would take the treatment she was randomized to, the effect of randomization in this subgroup is exactly the average causal effect of the treatment in this subgroup

$$\mathrm{E}[Y^{z=1} - Y^{z=0}|A^{z=0} < A^{z=1}] = \mathrm{E}[Y^{a=1} - Y^{a=0}|A^{z=0} < A^{z=1}]$$

Identification of LATE: sketch of proof (4)

- Let's return to our ITT effect to see what happens if zero defiers

$$\mathrm{E}[Y^{z=1}\text{-}Y^{z=0}] =$$

$$\mathrm{E}[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}<A^{z=1}]\mathrm{Pr}[A^{z=0}<A^{z=1}] \qquad \text{(compliers)}$$

$$+\ \mathrm{E}[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}=A^{z=1}=1]\mathrm{Pr}[A^{z=0}=A^{z=1}=1] \qquad \text{(always-takers)}$$

$$+\ \mathrm{E}[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}=A^{z=1}=0]\mathrm{Pr}[A^{z=0}=A^{z=1}=0] \qquad \text{(never-takers)}$$

$$+\ \mathrm{E}[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}>A^{z=1}]\mathrm{Pr}[A^{z=0}>A^{z=1}] \qquad \text{(defiers)}$$

Identification of LATE: sketch of proof (4)

- Let's return to our ITT effect to see what happens if zero defiers

$$\mathrm{E}[Y^{z=1}\text{-}Y^{z=0}] =$$

$$\mathrm{E}[Y^{z=1}\text{-}Y^{z=0}|A^{z=0}{<}A^{z=1}]\mathrm{Pr}[A^{z=0}{<}A^{z=1}] \qquad \text{(compliers)}$$

$$+\ 0 \qquad \text{(always-takers)}$$

$$+\ 0 \qquad \text{(never-takers)}$$

$$+\ 0 \qquad \text{(defiers)}$$

Identification of LATE: sketch of proof (5)

- By randomization and monotonicity, we have:

$$\mathrm{E}[Y^{z=1}\text{-}Y^{z=0}] = \mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]$$

$$\mathrm{Pr}[A^{z=1}<A^{z=0}] = \mathrm{E}[A|Z=1] - \mathrm{E}[A|Z=0]$$

- Thus, we have:

$$\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0]$$

$$= \mathrm{E}[Y^{a=1}\text{-}Y^{a=0}|A^{z=0}<A^{z=1}](\mathrm{E}[A|Z=1] - \mathrm{E}[A|Z=0])$$

- Rearranging terms, we have identified the LATE:

$$\mathrm{E}[Y^{a=1}\text{-}Y^{a=0}|A^{z=0}<A^{z=1}]$$

$$= (\mathrm{E}[Y|Z=1] - \mathrm{E}[Y|Z=0])/(\mathrm{E}[A|Z=1] - \mathrm{E}[A|Z=0])$$

Overview

- Motivation for IV methods

- Key assumptions for identifying causal effects with IVs

- IV estimation and tools for understanding possible threats to validity

- Extensions and further considerations

- Summary and Q&A

Concept of bounding

- The effect will be within a certain range

    - Causal risk difference: $-1 \leq RD \leq 1$

    - Not very informative

- Often, we combine data with assumptions to estimate the effect along that range, but that may require too strong of assumptions

- What if we could use weaker assumptions to identify a range of possible values?

    - Less information but lower risk of being wrong

Heuristics of bounding the average causal effect

Reporting two point estimates

Causal risk difference

1

-1
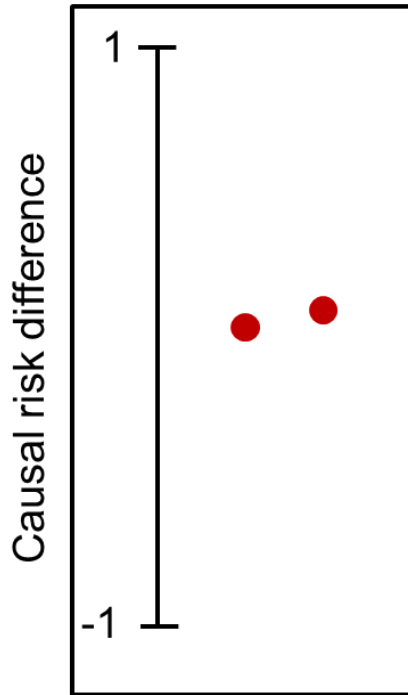
Reporting bounds and point estimates
under multiple sets of assumptions

Causal risk difference

1

-1

Increasingly strong assumptions
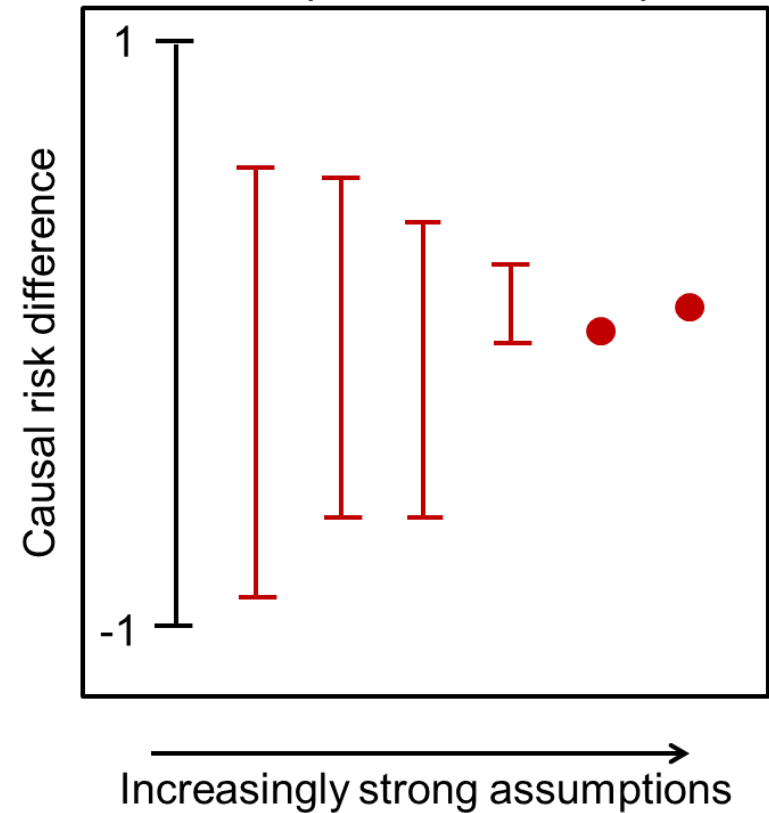
Partial and point identification

- Imagine we had infinite data

- A causal effect is (point-) identified if the data combined with our assumptions results in a single number: the point estimate

- A causal effect is partially identified if the data combined with our assumptions results in a range of numbers defined by lower and upper bounds
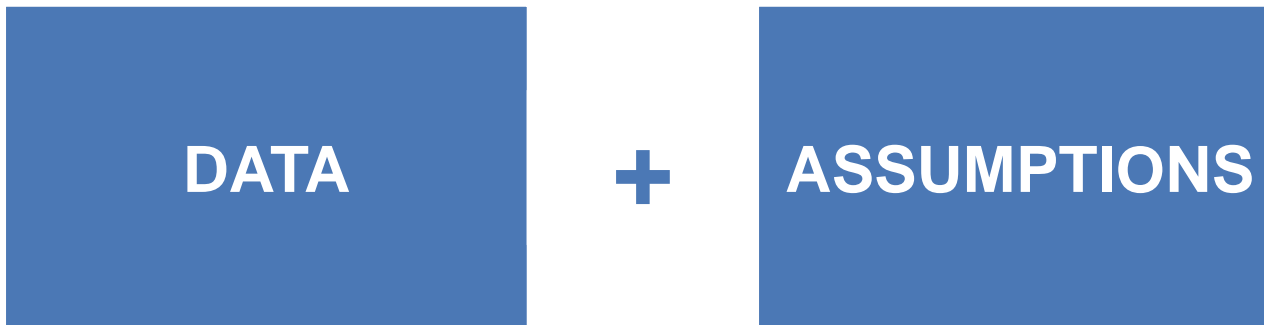
DATA **+** **ASSUMPTIONS**

Bounds with no data, no assumptions

- Before we look at our dataset or make any assumptions, our counterfactual risks and causal effects are naturally bounded:

  - $0 \leq \Pr[Y^{a=0}=1] \leq 1$

  - $0 \leq \Pr[Y^{a=1}=1] \leq 1$

  - Causal risk difference: $-1 \leq RD \leq 1$

  - Causal risk ratio: $0 \leq RR \leq \infty$

Bounds with data but no assumptions

- **With the dataset on this slide (but no assumptions), let's compute bounds:**
  - Pr[$Y^{a=0}=1$]
  - Pr[$Y^{a=1}=1$]
  - Causal risk difference

ID	A	Y	$Y^{a=0}$	$Y^{a=1}$
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
5	0	1	1	
6	1	0		0
7	1	0		0
8	1	0		0
9	1	1		1
10	1	1		1

Bounds with data but no assumptions

$$\underline{\hspace{2cm}} \leq \text{Pr}[\, Y^{a=0}=1] \leq \underline{\hspace{2cm}}$$

ID	*A*	Y	$Y^{a=0}$	$Y^{a=1}$
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
5	0	1	1	
6	1	0		0
7	1	0		0
8	1	0		0
9	1	1		1
10	1	1		1

Bounds with data but no assumptions

$$0.1 \leq \Pr[Y^{a=0}=1] \leq 0.6$$

$$\underline{\hspace{2cm}} \leq \Pr[Y^{a=1}=1] \leq \underline{\hspace{2cm}}$$

ID	A	Y	$Y^{a=0}$	$Y^{a=1}$
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
5	0	1	1	
6	1	0		0
7	1	0		0
8	1	0		0
9	1	1		1
10	1	1		1

Bounds with data but no assumptions

$0.1 \leq \Pr[Y^{a=0}=1] \leq 0.6$

$0.2 \leq \Pr[Y^{a=1}=1] \leq 0.7$

$\underline{\hspace{2cm}} \leq RD \leq \underline{\hspace{2cm}}$

ID	A	Y	$Y^{a=0}$	$Y^{a=1}$
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
5	0	1	1	
6	1	0		0
7	1	0		0
8	1	0		0
9	1	1		1
10	1	1		1

Bounds with data but no assumptions

$$0.1 \leq \Pr[Y^{a=0}=1] \leq 0.6$$

$$0.2 \leq \Pr[Y^{a=1}=1] \leq 0.7$$

$$-0.4 \leq RD \leq 0.6$$

ID	A	Y	$Y^{a=0}$	$Y^{a=1}$
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
5	0	1	1	
6	1	0		0
7	1	0		0
8	1	0		0
9	1	1		1
10	1	1		1

Bounds with data but no assumptions

$$0.1 \leq \Pr[Y^{a=0}=1] \leq 0.6$$

$$0.2 \leq \Pr[Y^{a=1}=1] \leq 0.7$$

$$-0.4 \leq RD \leq 0.6$$

- Without assumptions, we do not get very far
  - Assumption-free bounds for the average causal effect will always cover the null
  - Remember the data do not speak for themselves!

| ID	A	Y	$Y^{a=0}$	$Y^{a=1}$
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
5	0	1	1	
6	1	0		0
7	1	0		0
8	1	0		0
9	1	1		1
10	1	1		1

Introducing our data setting

- Suppose our (simulated) dataset came from a study that is similar to the NORCCAP trial in our case study paper

- Specifically, suppose our data come from a pragmatic randomized trial

  - Randomized to colorectal cancer screening versus no screening

  - Treatment is unavailable to the control arm

  - Outcome of interest is 10-year cancer risk

  - Complete follow-up (for illustrative purposes)

- See R code for data

Our dataset in R
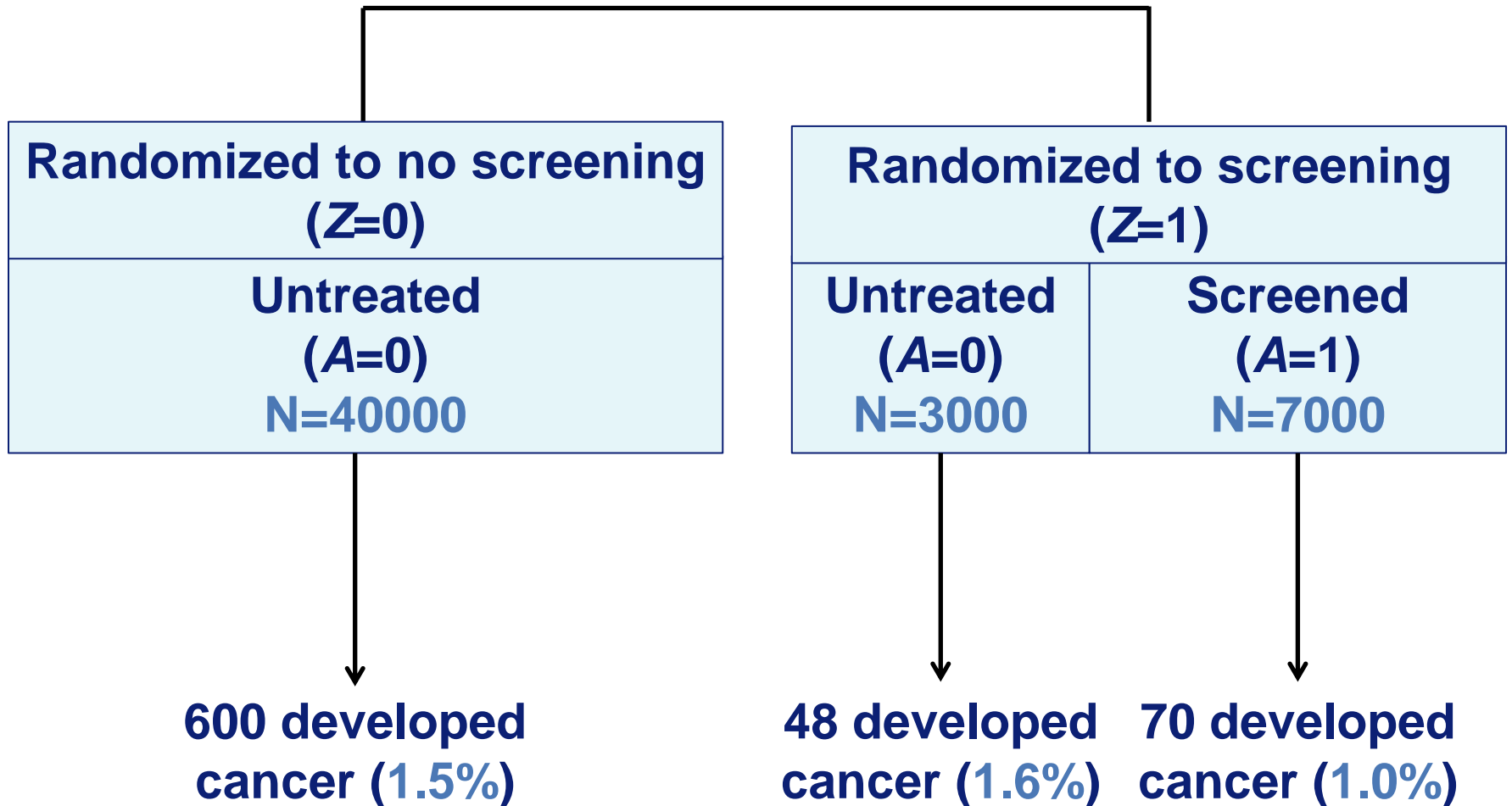
Recall the IV conditions

1. The randomization indicator and treatment are associated

    - Pr[A=1|Z=1] − Pr[A=1|Z=0] ≠ 0

2. The randomization indicator only affects the outcome through encouraging treatment

    - To discuss: when would this be a reasonable assumption?

3. The randomization indicator and outcome do not share causes

    - Expected by design

Compliance types in the context of a trial

Randomized to treatment arm ($Z=1$)

		Treated ($A^{z=1}=1$)	Not treated ($A^{z=1}=0$)
Random-ized to placebo arm ($Z=0$)	Treated ($A^{z=0}=1$)	**Always-taker** ($A^{z=0}=A^{z=1}=1$)	**Defier** ($A^{z=0}>A^{z=1}$)
	Not treated ($A^{z=0}=0$)	**Complier** ($A^{z=0}<A^{z=1}$)	**Never-taker** ($A^{z=0}=A^{z=1}=0$)

Compliance types for one-sided non-compliance

Randomized to treatment arm ($Z=1$)

		Treated ($A^{z=1}=1$)	Not treated ($A^{z=1}=0$)
Random-ized to placebo arm ($Z=0$)	Treated ($A^{z=0}=1$)	**Always-taker** ($A^{z=0}=A^{z=1}=1$)	**Defier** ($A^{z=0}>A^{z=1}$)
	Not treated ($A^{z=0}=0$)	**Complier** ($A^{z=0}<A^{z=1}$)	**Never-taker** ($A^{z=0}=A^{z=1}=0$)

Identifying compliance types in our trial

- In treatment arm, we know all subjects' compliance types

  - Z=1 and A=0 implies she is a never-taker

  - Z=1 and A=1 implies she is a complier

- In the control arm, we do not know who is a complier versus never-taker

  - Z=0 and A=0 is either a never-taker or a complier

- In the control arm, we know the distribution is the same as the treatment arm

  - By randomization

What we will show we can bound or identify…

Population	Risk under no treatment	Risk under treatment	Causal effect (RD or RR)
Compliers	Point identification	Point identification	Point identification
Never-takers	Point identification	No information (between 0 and 1)	Bounded
Full study population	Point identification	Bounded	Bounded

$Y^{a=1}$ in the never-takers

- By definition, we have no information on what would happen to the never-takers had we somehow forced them to be treated

- $0 \leq \Pr[Y^{a=1}=1|A^{z=0}=A^{z=1}=0] \leq 1$

$Y^{a=0}$ in the never-takers

- Under the IV conditions, we can (point-) identify this!

- Recall we know who the never-takers are in treatment arm

  - By condition (3), they are exchangeable with the placebo arm never-takers

  - By condition (2) and consistency, their counterfactual risk under no treatment is their observed risk

$$\Pr[Y^{a=0}=1|A^{z=0}=A^{z=1}=0] = \Pr[Y=1|Z=1,A=0]$$

$Y^{a=1}$ in the compliers

- Under the IV conditions, we can (point-) identify this!

- Recall we know who the compliers are in the treatment arm

  - By condition (3), they are exchangeable with the placebo arm compliers

  - By condition (2) and consistency, their counterfactual risk under treatment is their observed risk

$$\Pr[Y^{a=1}=1|A^{z=0}<A^{z=1}] = \Pr[Y=1|Z=1,A=1]$$

$Y^{a=0}$ in the compliers

- Under the IV conditions, we can (point-) identify this!

- The observed risk in the placebo arm is the counterfactual risk under no treatment in everybody

  - And just showed that we can identify the counterfactual risk under no treatment in the never-takers

  - Using the known distribution of compliance types, can solve

$Y^{a=0}$ and $Y^{a=1}$ in the full study population

- Under the IV conditions, we can (point-) identify the counterfactual risk under no treatment
  - Exactly the observed risk in the placebo arm
- Under the IV conditions, we can only bound the counterfactual risk under treatment
  - Because we have no information on the never-takers
- What about causal effects?
  - Identified in the compliers
  - Bounded in the never-takers
  - Bounded in the fully study population

What we can bound or identify…

Population	Risk under no treatment	Risk under treatment	Causal effect (RD or RR)
Compliers	Point identification	Point identification	Point identification
Never-takers	Point identification	No information (between 0 and 1)	Bounded
Full study population	Point identification	Bounded	Bounded

Computed in R under the IV conditions

Population	Risk under no treatment	Risk under treatment	Causal effect (RD)
Compliers	1.5%	1.0%	-0.5%
Never-takers	1.6%	[0.0%, 100.0%]	[-1.6%, 98.4%]
Full study population	1.5%	[0.7%, 30.7%]	[-0.8%, 29.2%]

Computed in R under IV + additive homogeneity

Population	Risk under no treatment	Risk under treatment	Causal effect (RD)
Compliers	1.5%	1.0%	-0.5%
Never-takers	1.6%		Assume: -0.5%
Full study population	1.5%		

Computed in R under IV + other restrictions

Population	Risk under no treatment	Risk under treatment	Causal effect (RD)
Compliers	1.5%	1.0%	-0.5%
Never-takers	1.6%	Assume: [0.0%, x%]	
Full study population	1.5%		

Bounds in trials with two-sided non-compliance

- Bounds for the per-protocol effect in trials with non-compliance in both arms can be achieved
  - See R code and equations on page 5 of our case study
- Note the possibility of all four compliance types
  - Bounds within compliance types can be achieved for a specified feasible proportion of defiers
  - One-sided non-compliance makes it easier to see the intuitions of how these bounds work

A brief history on bounds

- Robins 1989 and Manski 1990 derived the "natural bounds"

- Balke & Pearl 1997 derived bounds that can be sometimes narrower

  - The difference between the natural and Balke-Pearl bounds is in how the IV conditions are formalized (specifically, exchangeability)

  - In practice, the bounds are often equivalent

- Richardson & Robins 2010 described the Balke-Pearl bounds in relation to compliance types

  - A substantially more general approach than what we covered in our special case of a trial with one-sided non-compliance

- Lots of literature on combining IV conditions with additional assumptions

Bounds and the IV inequalities

- For dichotomous *Z, A, Y*, the IV conditions imply certain constraints on the observed data
  - IV inequalities also for some non-binary settings
  - Mathematically related to how bound expressions are derived
- Can be used to detect extreme violations of the IV conditions
  - That is, can falsify but not verify that the conditions hold
- See R code

Balke & Pearl 1997 *JASA*; Bonet 2001 *PUAI*; Glymour et al. 2012 *AJE*

Why bound? Three reasons

1. Bounds remind us to remain humble about our point estimates.

2. The exercise of bounding can sometimes illuminate subgroups we have more (or less) information on.

3. Bounding the causal effect under several sets of assumptions shifts the scientific debate to what assumptions are most reasonable and therefore what effect sizes are most plausible.

Reason #1: a reminder to remain humble about our point estimates

"Some argue against reporting bounds for nonidentifiable parameters, because bounds are often so wide as to be useless in making public health decisions. But we view the latter problem as a reason for reporting bounds in conjunction with other analyses: wide bounds make clear the degree to which public health decisions are dependent on merging the data with strong prior beliefs."

- Robins & Greenland 1996 *JASA*

Why bound? Three reasons.

1. Bounds remind us to remain humble about our point estimates.

2. The exercise of bounding can sometimes illuminate subgroups we have more (or less) information on.

3. Bounding the causal effect under several sets of assumptions shifts the scientific debate to what assumptions are most reasonable and therefore what effect sizes are most plausible.

Reason #2: bounds sometimes illuminate subgroups we have more/less information on

- In the special case of trials with one-sided non-compliance:
  - We learned a lot about the compliers
  - We learned less about the never-takers
  - We could have described compliance types in the treatment arm (but are not generally identifiable)

- In other settings, bounds for the average causal effect and effects within subgroups may provide similar clarity

Why bound? Three reasons.

1. Bounds remind us to remain humble about our point estimates.

2. The exercise of bounding can sometimes illuminate subgroups we have more (or less) information on.

3. Bounding the causal effect under several sets of assumptions shifts the scientific debate to what assumptions are most reasonable and therefore what effect sizes are most plausible.

Reason #3: shifts scientific debate to reasonableness of assumptions

- We should always be transparent about the assumptions underlying any effect estimate

- Bounding causal effects under several sets of assumptions serves as a reminder that the scientific debate should be about the reasonableness of the assumptions

  - If we agree the assumptions are reasonable, then we agree what range of effect sizes are plausible

Consider this scenario

- Three investigators conduct analyses on the same dataset to compute the causal risk difference (ignore sampling variability)

Investigator	Assumptions	Bounds for RD
	Assumption-free	-0.3 to 0.7
Investigator #1	A	-0.1 to 0.4
Investigator #2	A, B	0.1 to 0.4
Investigator #3	A, B, C	0.3

What should we conclude/do?

- …if we had a consensus on what assumptions are reasonable?

- …if we did not have a consensus?

Investigator	Assumptions	Bounds for RD
	Assumption-free	-0.3 to 0.7
Investigator #1	A	-0.1 to 0.4
Investigator #2	A, B	0.1 to 0.4
Investigator #3	A, B, C	0.3

Considerations for bounding

Three reasons for bounding and estimating treatment effects under several sets of assumptions:

1. Bounds remind us to remain humble about our point estimates.

2. The exercise of bounding can sometimes illuminate subgroups we have more (or less) information on.

3. Bounding the causal effect under several sets of assumptions shifts the scientific debate to what assumptions are most reasonable and therefore what effect sizes are most plausible.

IV (point) estimation

- Previously discussed the standard IV ratio and estimating a per-protocol effect

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[A|Z=1] - E[A|Z=0]}$$

- Two other modeling procedures are highlighted in the provided R code

  - Two-stage least squares estimation
  - G-estimation of an additive structural mean model

Two-stage least squares estimation

- Stage 1: Fit a linear model for treatment

    - $E[A|Z] = \alpha_0 + \alpha_1 Z$

    - Generate the predicted values $\hat{E}[A|Z]$ for each individual

- Stage 2: Fit a linear model for the outcome

    - $E[Y|Z] = \beta_0 + \beta_1 \hat{E}[A|Z]$

- The parameter estimate of β_1 is the IV estimate

IV estimation

- The two-stage estimator is frequently used, while IV g-estimators of structural mean models are less common approaches

- Some benefits/extensions of these modeling approaches:
  - Introduce covariates
  - Handle continuous treatments
  - Consider multiple instruments simultaneously (of interest in observational studies)
- See R code for examples

Overview

- Motivation for IV methods

- Key assumptions for identifying causal effects with IVs

- IV estimation and tools for understanding possible threats to validity

- Extensions and further considerations

- Summary and Q&A

Further points for consideration when computing bounds in trials

- Confidence intervals

- Relaxations of the IV conditions

- Conditional randomization and bounds
  - See standardization approach in our case study

- Possible collider stratification biases
  - Could combine with inverse probability weighting

- Three or more trial arms

- Non-binary treatment options

- Non-binary outcomes
  - E.g., continuous or failure-time

Tamer 2010 *ARE*; Robins 1989 & 1994

Bounding with continuous *Y*

- Recall how we computed assumption-free bounds for a dichotomous *Y*

- How could we extend this to a continuous *Y*?

ID	*A*	*Y*	$Y^{a=0}$	$Y^{a=1}$
1	0	0	0	
2	0	0	0	
3	0	0	0	
4	0	0	0	
5	0	1	1	
6	1	0		0
7	1	0		0
8	1	0		0
9	1	1		1
10	1	1		1

Bounding with continuous *Y*

- **Recall how we computed assumption-free bounds for a dichotomous *Y***

- **How could we extend this to a continuous *Y*?**

ID	*A*	*Y*	$Y^{a=0}$	$Y^{a=1}$
1	0	200	200	
2	0	220	220	
3	0	250	250	
4	0	300	300	
5	0	180	180	
6	1	300		300
7	1	320		320
8	1	480		480
9	1	250		250
10	1	250		250

Further points for consideration when proposing IV estimation in trials
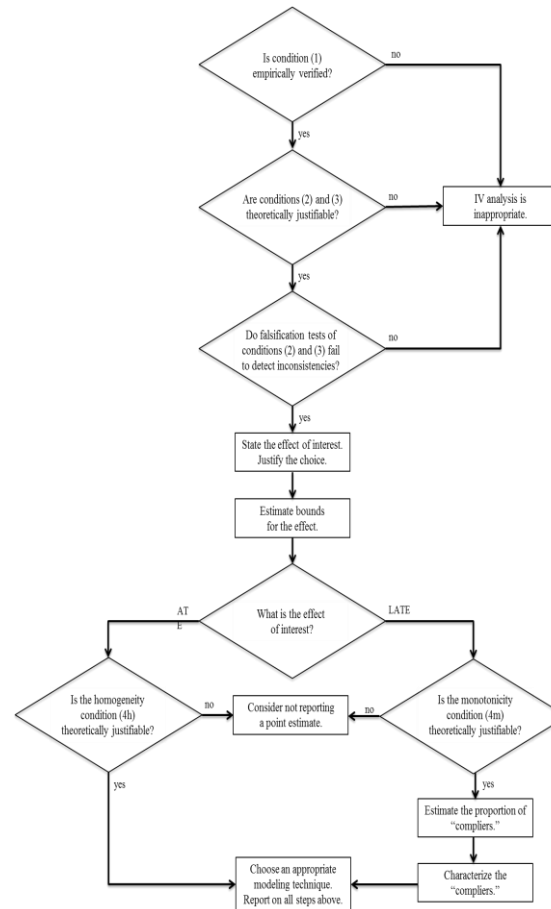
- Loss to follow-up and other collider stratification biases

- Conditional and/or sequential randomization

- Non-binary treatment strategies

- Active treatment comparisons

- Non-binary outcomes

  - E.g., continuous or failure-time

See Robins 1989 & Robins 1994 for further discussion of g-estimation.

Further points for consideration when proposing IV methods in observational studies
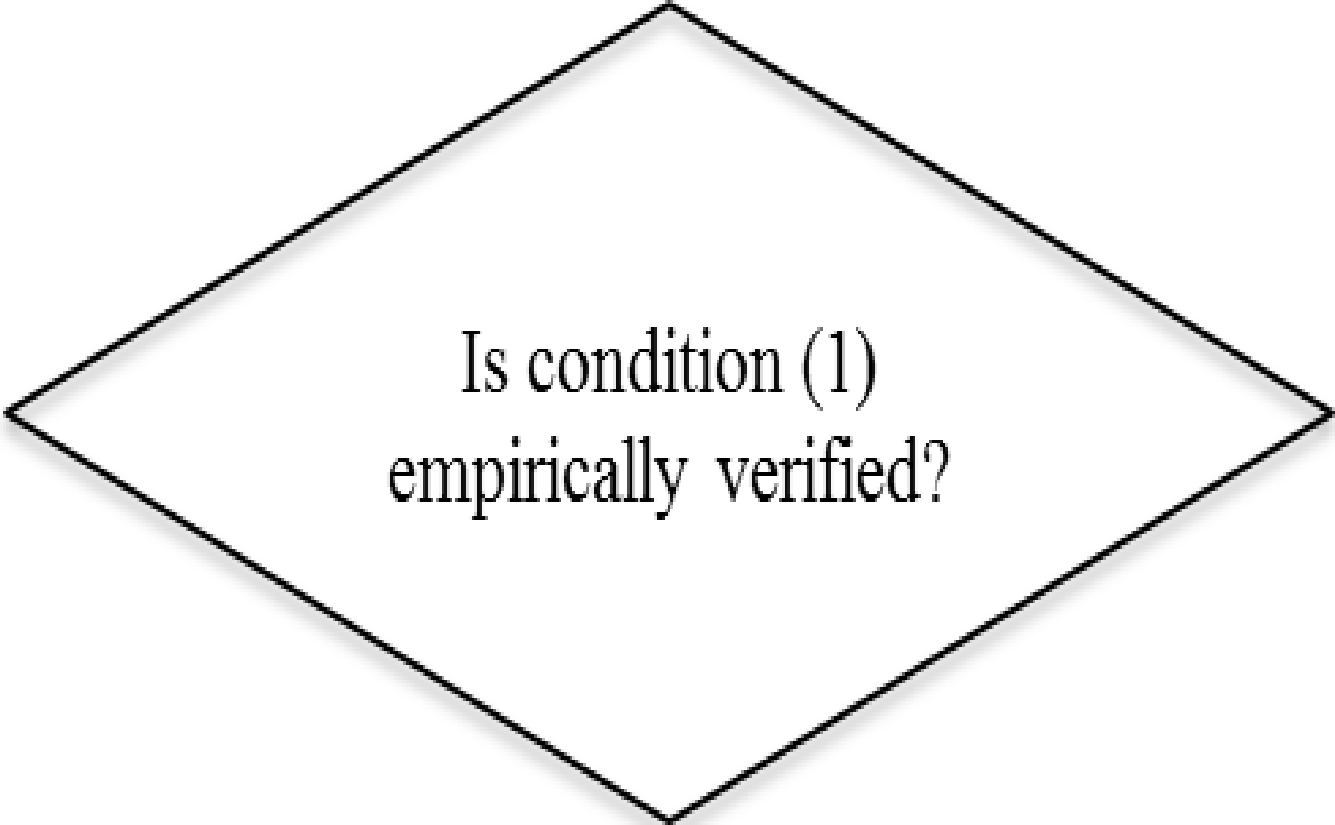
- Same structure of methods and threats to validity apply

- However, no IV conditions are expected to hold by design!

  - Need to be vigilant in our attempts to support/falsify each condition

  - May be more interested in the trade-offs of relative bias in an IV versus a non-IV approach

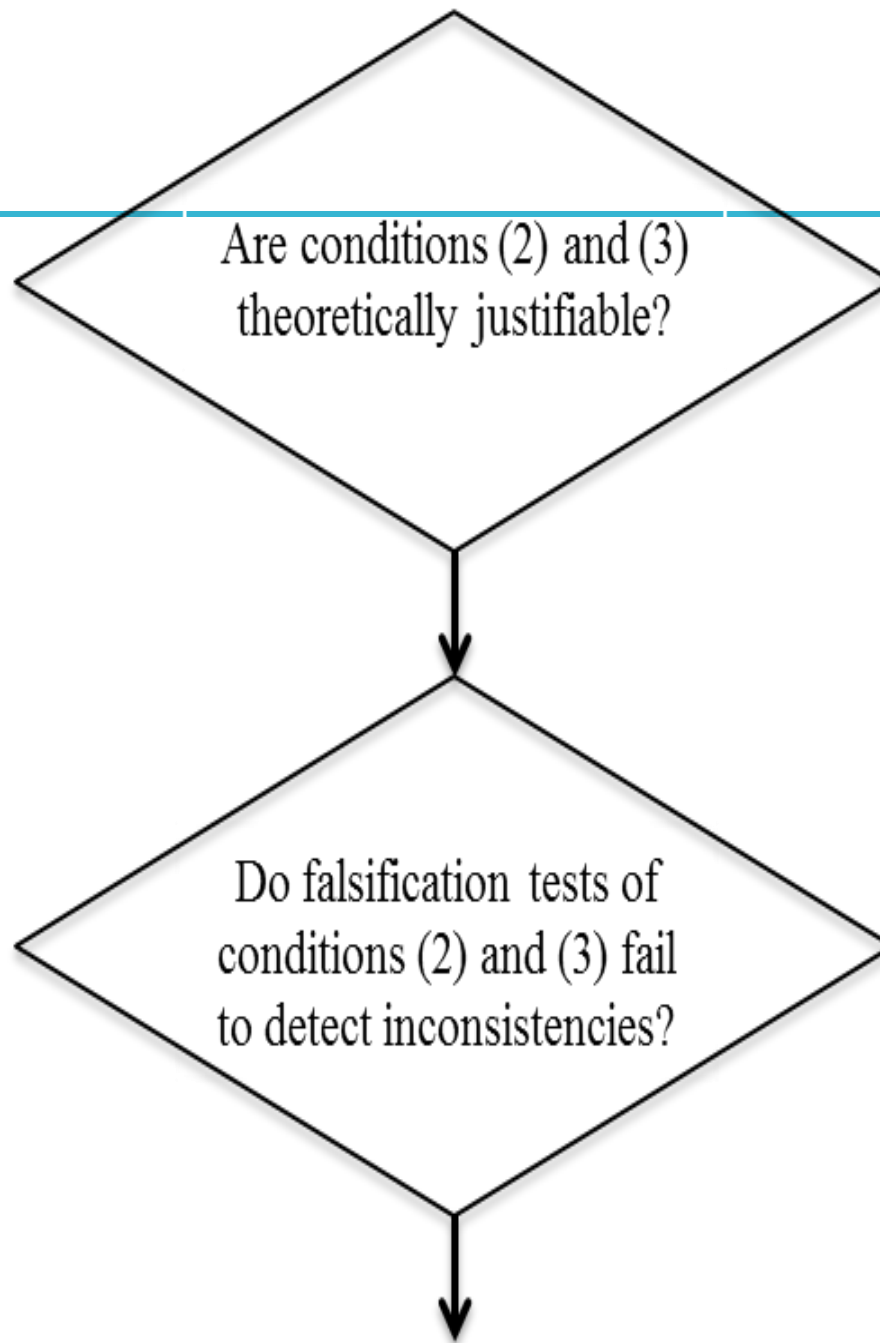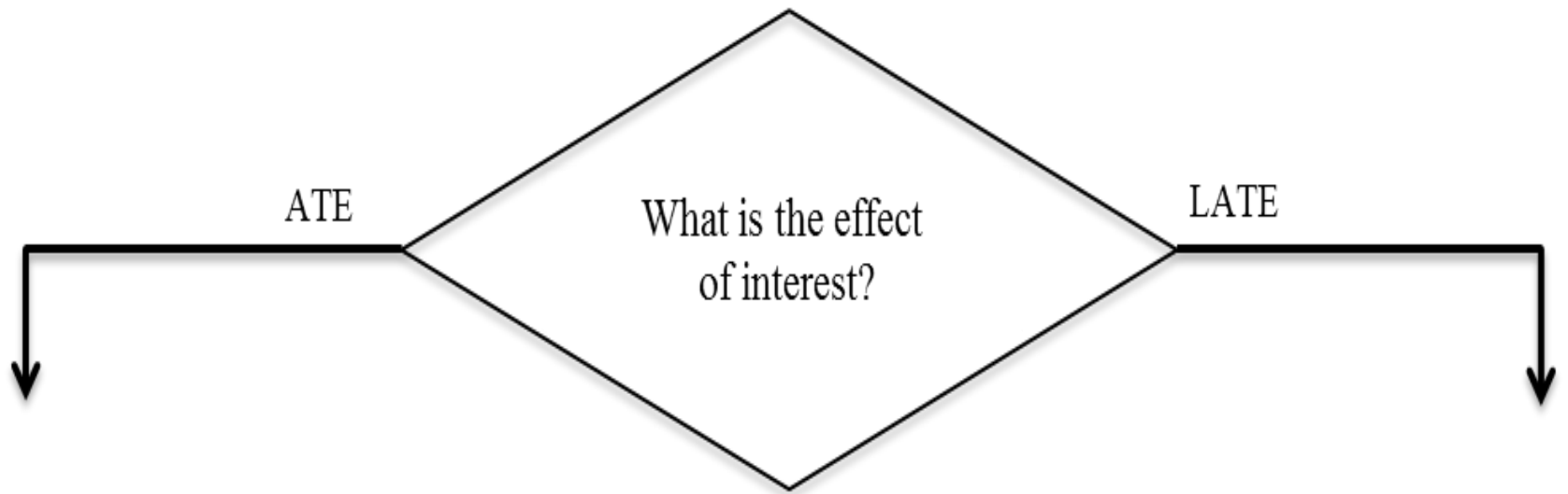- Interpretation considerations

Reporting guidelines

Swanson & Hernan 2013 *Epi*

State the effect of interest.
Justify the choice.

Estimate bounds
for the effect.

ATE

What is the effect
of interest?

LATE

ATE

Is the homogeneity
condition (4h)
theoretically justifiable?

LATE

Is the monotonicity condition (4m) theoretically justifiable?

Estimate the proportion of "compliers."

Characterize the "compliers."
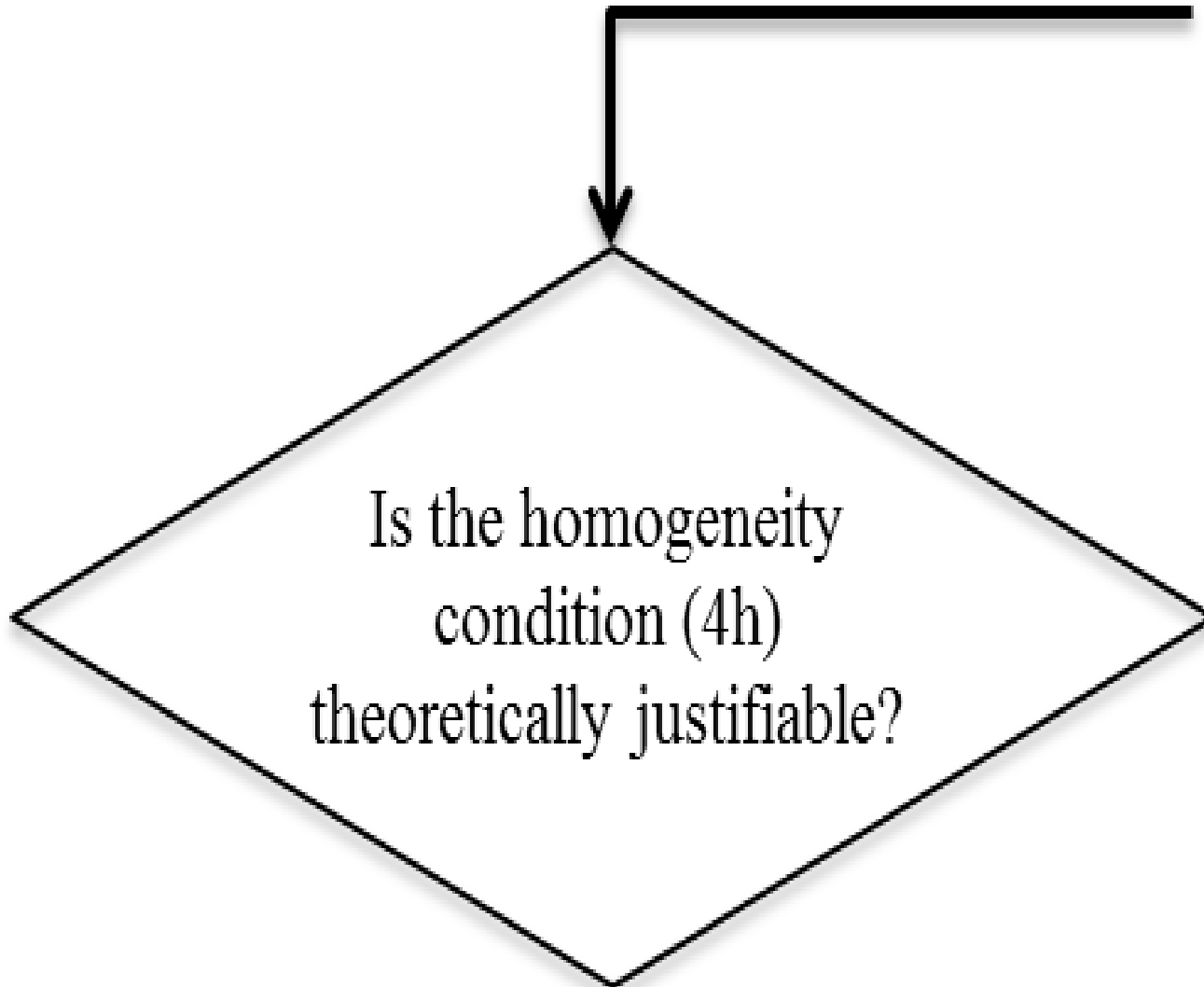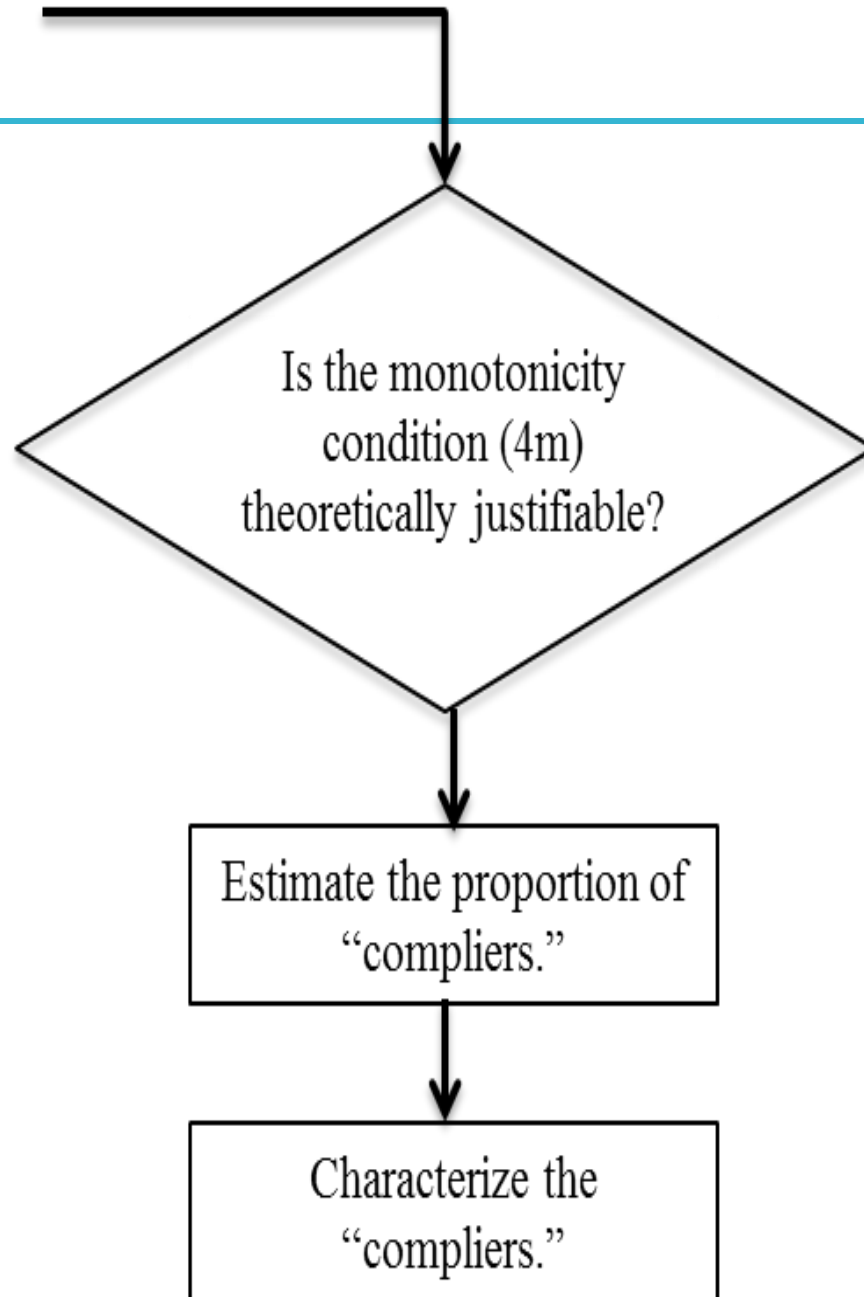
Choose an appropriate modeling technique.
Report on all steps above.

Reporting guidelines

Swanson & Hernan 2013 *Epi*

Overview

- Motivation for IV methods

- Key assumptions for identifying causal effects with IVs

- IV estimation and tools for understanding possible threats to validity

- Extensions and further considerations

- Summary and Q&A

Summary: key conditions

- IV methods require strong, untestable assumptions
  - Three IV conditions for bounding
  - Three IV conditions plus additional conditions for point estimation
- Applying IV methods requires concerted efforts to attempt to falsify assumptions and quantify possible biases
- Under these key conditions, IV methods offer opportunities for estimating:
  - Per-protocol effects in randomized trials
  - Treatment effects in observational studies

Summary: transparent reporting

- Transparent reporting is a key component of PCOR

- Major themes in reporting guidelines apply to both IV and non-IV studies

  - Should always clearly state and discuss assumptions
  - Should always state the effect we are estimating

- IV reporting also needs to address unique challenges

  - Requires applying different subject matter expertise
  - Seemingly minor violations of assumptions can result in large or counterintuitive biases
  - Interpreting "local" effects requires special care

Q&A

- Email: s.swanson@erasmusmc.nl