Advanced Statistical Methods for Observational Studies

LECTURE 05

class management

- Questions?

practical issue

venturing out of the ivory tower.

implementation: sensitivity analysis

- In practice, it's common to just report the value of Γ which nullifies your qualitative conclusions (i.e., goes from significant to insignificant), and to help the reader in interpreting the meaning of Γ.
- For example, $\Gamma = 2$ means that within a given pair – even though the two matched individuals looked identical in the data set – the actual odds of assignment was up to twice as likely for one member in the pair than the other. Likely this difference is due to the unobserved covariates.
- The question then becomes: Is what's left lingering out there, outside of your data set, enough to cause that level of confounding?

sensitivity analysis: understanding gamma

- When you report Gamma, you report the maximal Gamma that still returns a qualitatively similar conclusion (e.g., if you found the treatment positive and significant then you report the value of Gamma that just barely makes the test not-significant).
- Note: This says nothing about the case at hand. You are looking at the strength of the argument (i.e., how many in treated need to be switched to control before null). But we haven't measured the level of unobserved (it's still unobserved!!).
- This is like saying a building can withstand a magnitude 8 earthquake. There is no statement in there about what magnitude earthquake the building will experience.

sensitivity analysis: understanding gamma

- Explaining Gamma sensitivity is hard because the concept of confounding is tough.
- Confounding (usually) arises from sorting into treatment/control (propensity - Fisher) on variables important in determining the outcome (prognostic – Mill).
- Gamma sensitivity is kind of weird because it only focusses on the propensity, and then assumes the worst case for prognostic.

missing covariates

- Missing covariates

obs	b_weight	bw_mis	gest_age	ga_mis	dose	death
1	2412	0	36	0	1	0
2	NA	1	29	0	1	1
3	2569	0	36	0	1	0
4	2443	0	34	0	1	0
5	2569	0	36	0	0	0
6	2436	0	NA	1	0	0
7	2461	0	34	0	0	0
8	2759	0	32	0	0	0
9	2324	0	27	0	0	1
10	2667	0	34	0	0	0

missing covariates

- Missing covariates

obs	b_weight	bw_mis	gest_age	ga_mis	dose	death
1	2412	0	36	0	1	0
2	**2515**	1	29	0	1	1
3	2569	0	36	0	1	0
4	2443	0	34	0	1	0
5	2569	0	36	0	0	0
6	2436	0	**33**	1	0	0
7	2461	0	34	0	0	0
8	2759	0	32	0	0	0
9	2324	0	27	0	0	1
10	2667	0	34	0	0	0

(i) Build pscores using the imputed value and the missing indicators.
(ii) Use imputed values and missing indicators in calculating the Mahalanobis distance.

a small but important point

obs	b_weight	bw_mis	gest_age	ga_mis	dose	death
1	2412	0	36	0	1	
2	2515	1	29	0	1	
3	2569	0	36	0	1	
4	2443	0	34	0	1	
5	2569	0	36	0	0	
6	2436	0	33	1	0	
7	2461	0	34	0	0	
8	2759	0	32	0	0	
9	2324	0	27	0	0	
10	2667	0	34	0	0	

an example

matching on more than one metric

- Intuition: matching on just propensity scores is like uniform randomization, whereas a Mahalanobis & pscores is more like a matched pairs randomization.

example

- Example: House examines the patient and wants to treat for sarcoidosis. He is always considering treating for sarcoidosis… but in a way that is unrelated to how sick the patient is.
- To make this example easier to follow, let's consider two data generating functions

1. Treatment:

B

A	0	1
0	0.5	0.1
1	0.5	0.1

2. Outcome:

B

A	0	1
0	0.1	0.1
1	0.5	0.5

propensity score vs. prognostic score

- This departure arises when the variables predictive of treatment differs from the prognostically relevant variables
- This insight led to an interesting paper:
  - Bhattacharya & Vogt "Do Instrumental Variables Belong in Propensity Scores?"
- Prognostic score is one way to address this:
  - Ben Hansen "The prognostic analog of the propensity score"

takeaway

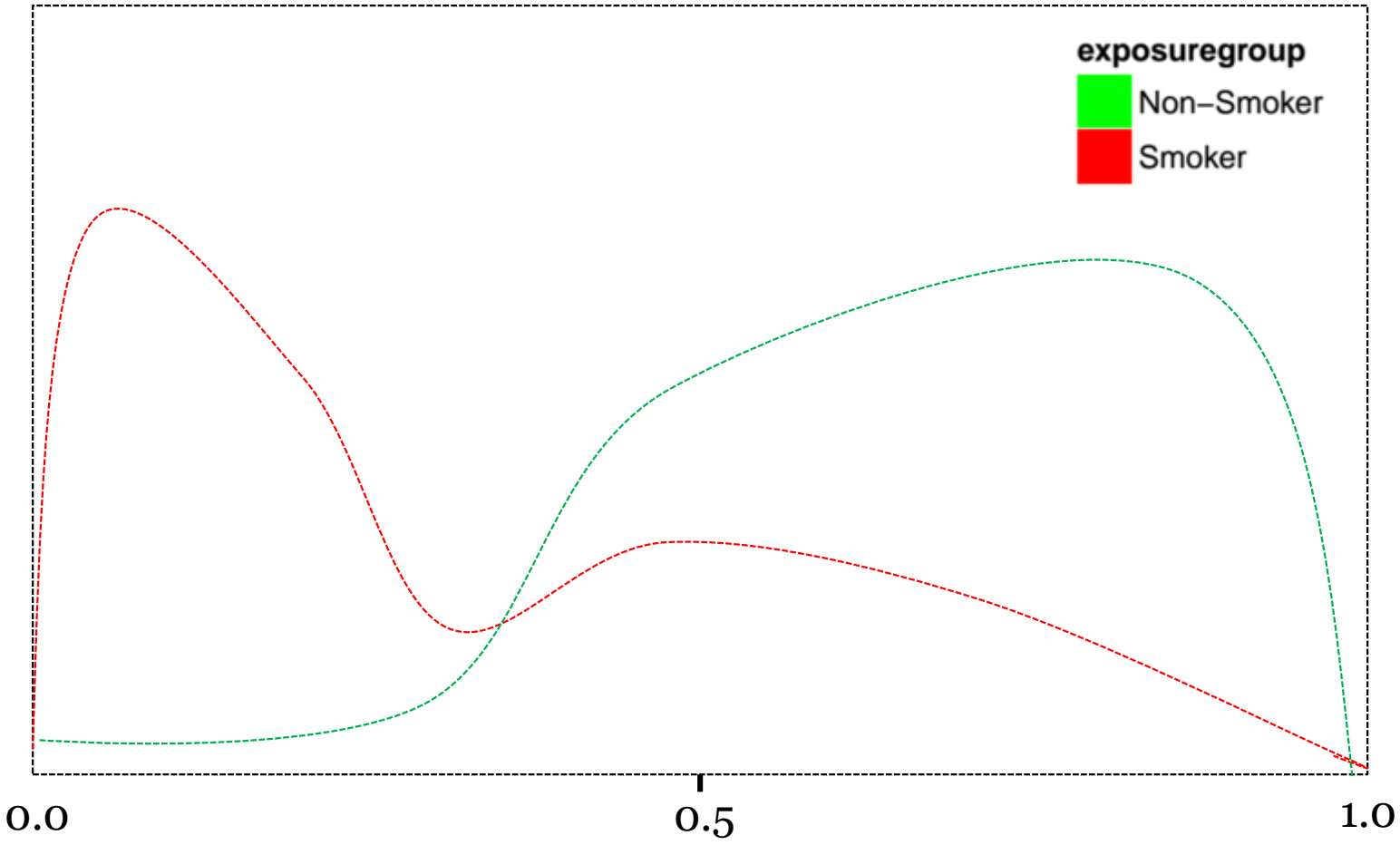
- This toy example highlights that the propensity score focuses on treatment, which may be unrelated to outcomes.
- This is OK – the theory of inference is predicated on randomization, not identical units going into the groups (Fisher)
- But it is better to start wish similar groups (Mill)
- Next is an example of matching using both

what are we estimating

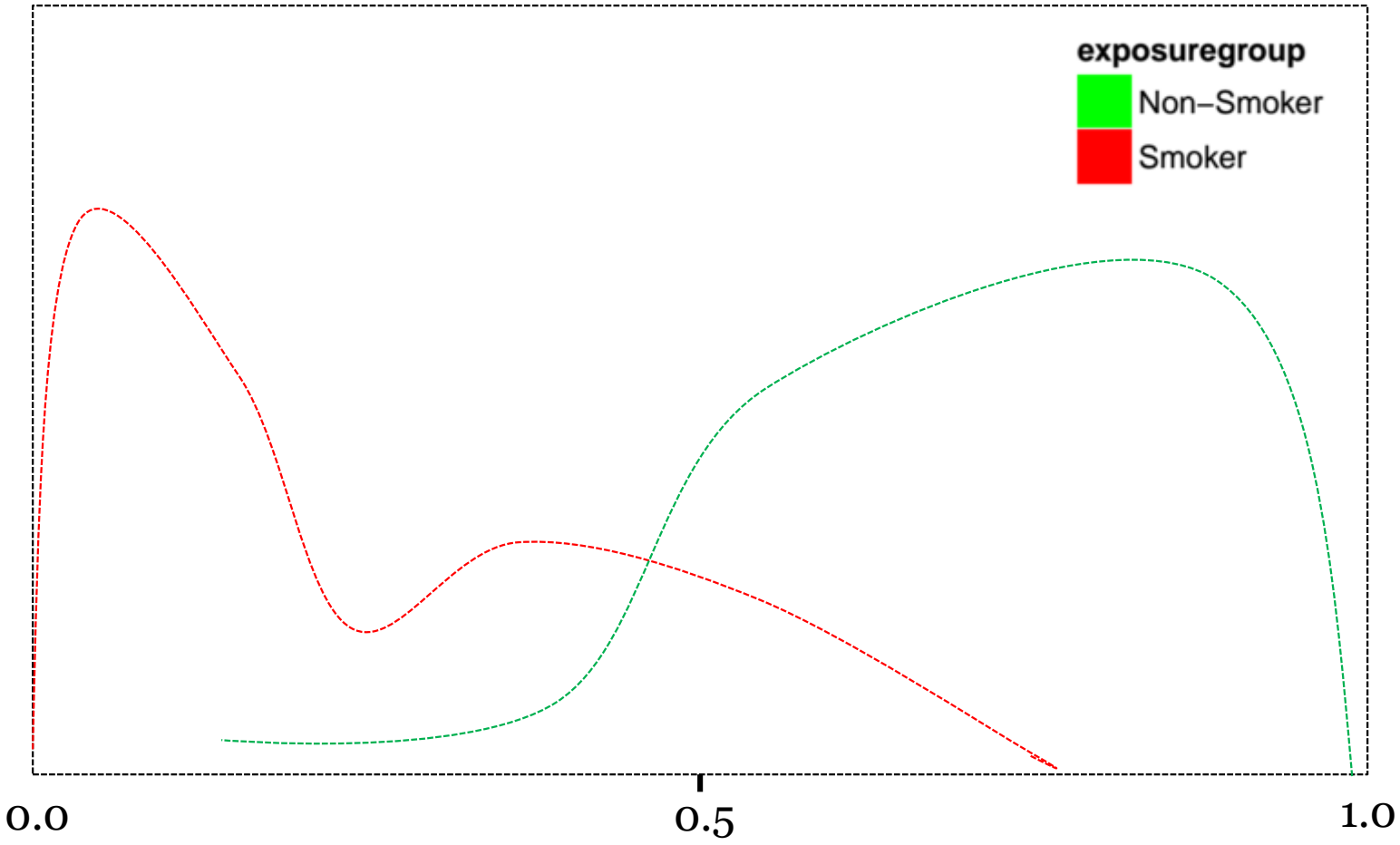
TREATMENT EFFECT

treatment effect: smoking example

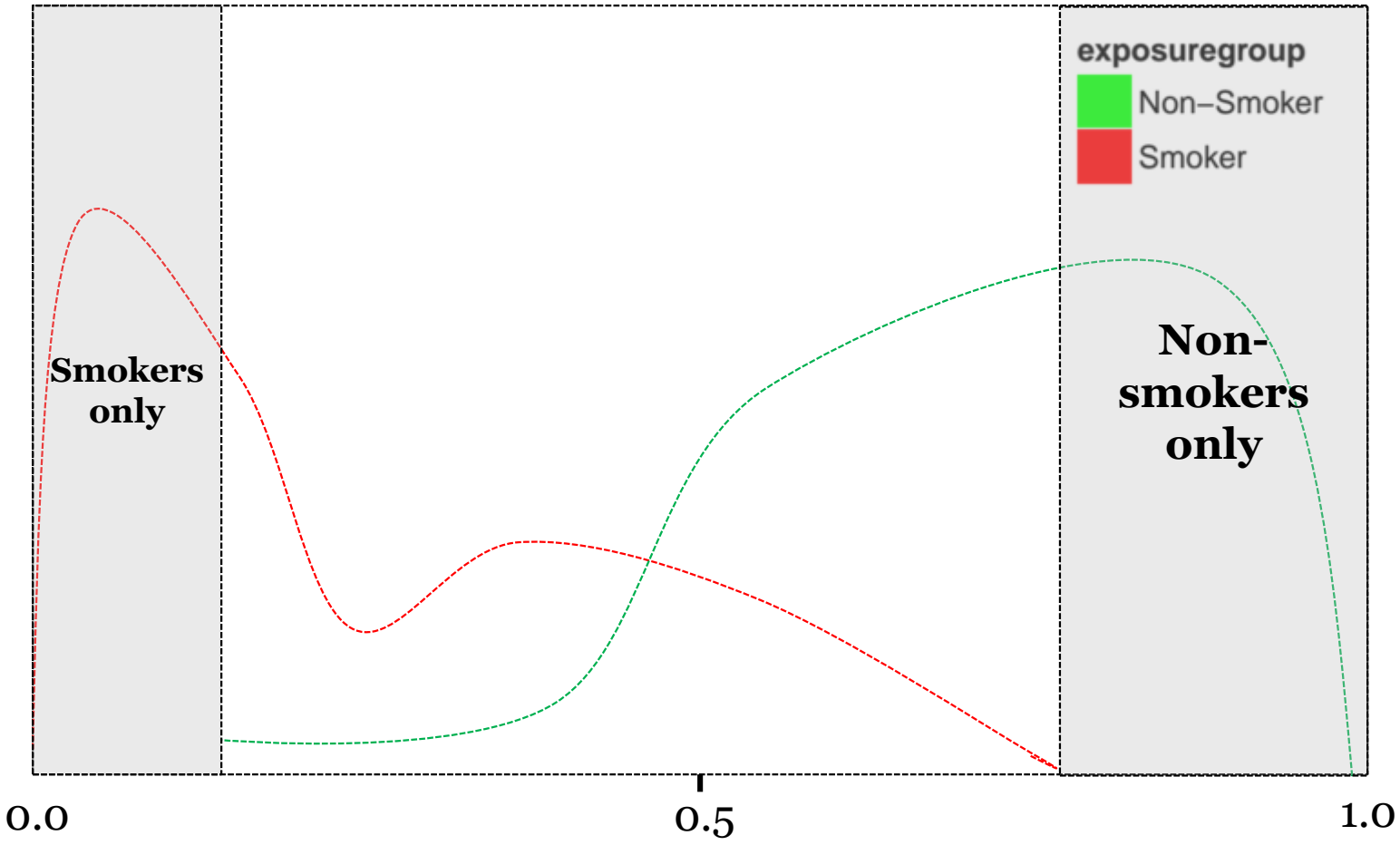
- Example: We collected data on people who reported for job training in the Bay Area. Roughly half smoked. We collected 20ish variables at baseline. We then looked at employment at 12 months.
- Let's consider matching one treated to one control.



P(non-S|**X**) = Probability of being a non-smoker, given covariates



P(non-S|**X**) = Probability of being a non-smoker, given covariates



P(non-S|**X**) = Probability of being a non-smoker, given covariates

treatment effect: smoking example

- Non-overlap
  - Look at a histogram
  - Upper and lower
  - Violation of strongly ignorable treatment assignment
  - Careful, need to consider what effect you're estimating
    - What's actually estimable and what isn't

treatment effect: smoking example

- Non-overlap
  - Look at a histogram
  - Upper and lower
  - Violation of strongly ignorable treatment assignment
  - Careful, need to consider what effect you're estimating
    - What's actually estimable and what isn't
  - Focus on the 50% range because that's actually where the debate is happening
  - Trim at the edges because that's where you're pretty sure the violation of SITA is going to happen
  - More detail here: *Crump et al*

treatment effect: smoking example

- Consider how to remove the observations that you can't/don't want to include in your study.
- This is roughly equivalent to the inclusion/exclusion criteria of a randomized controlled trial.
- Examine the pscore fitted model and see what parts of the covariate space are in the non-overlap
- Use a regression tree (or some other classifier) to make it intelligible. Citation: Traskin & Small (2011)

CONSORT
Flow Diagram

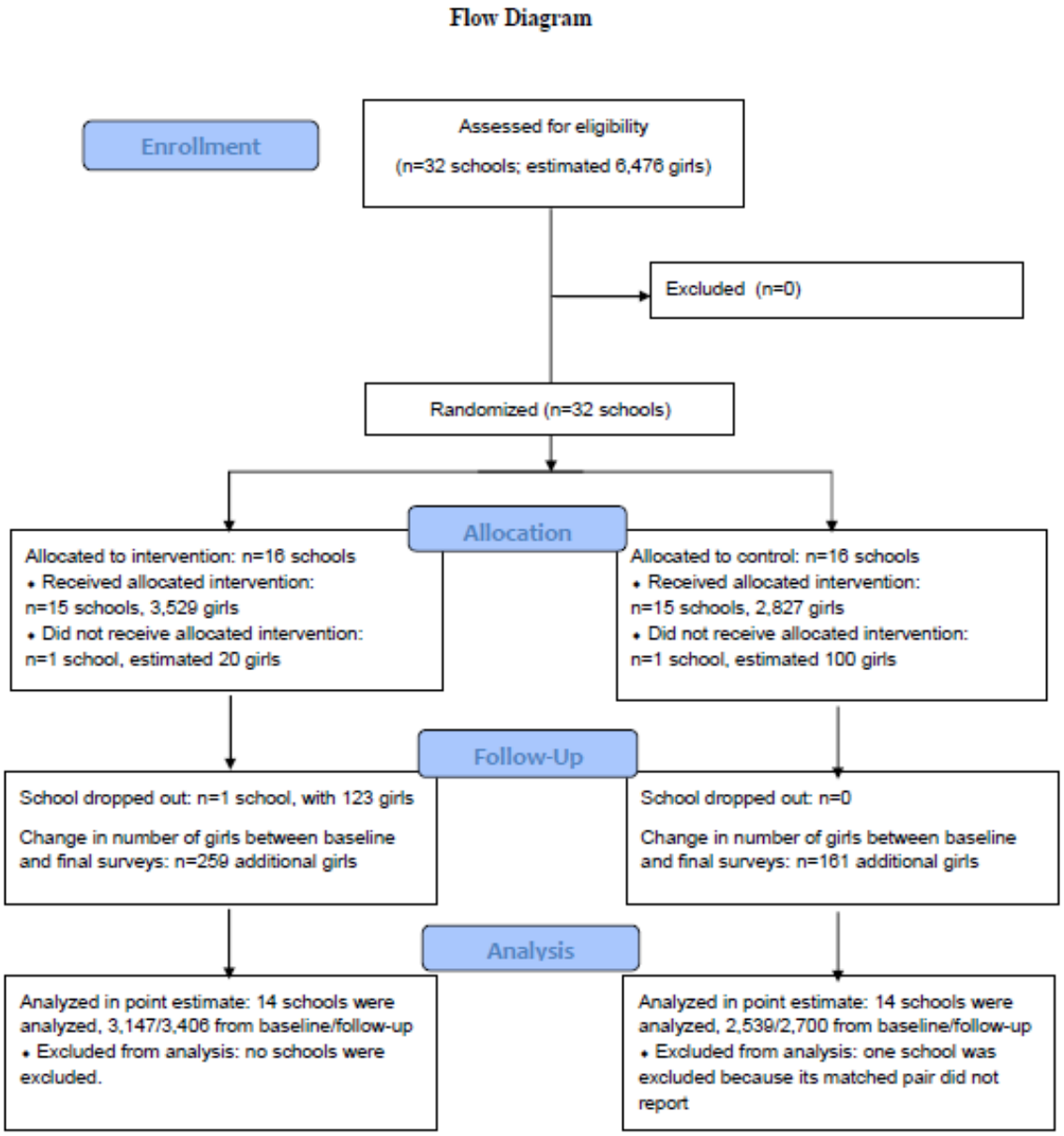
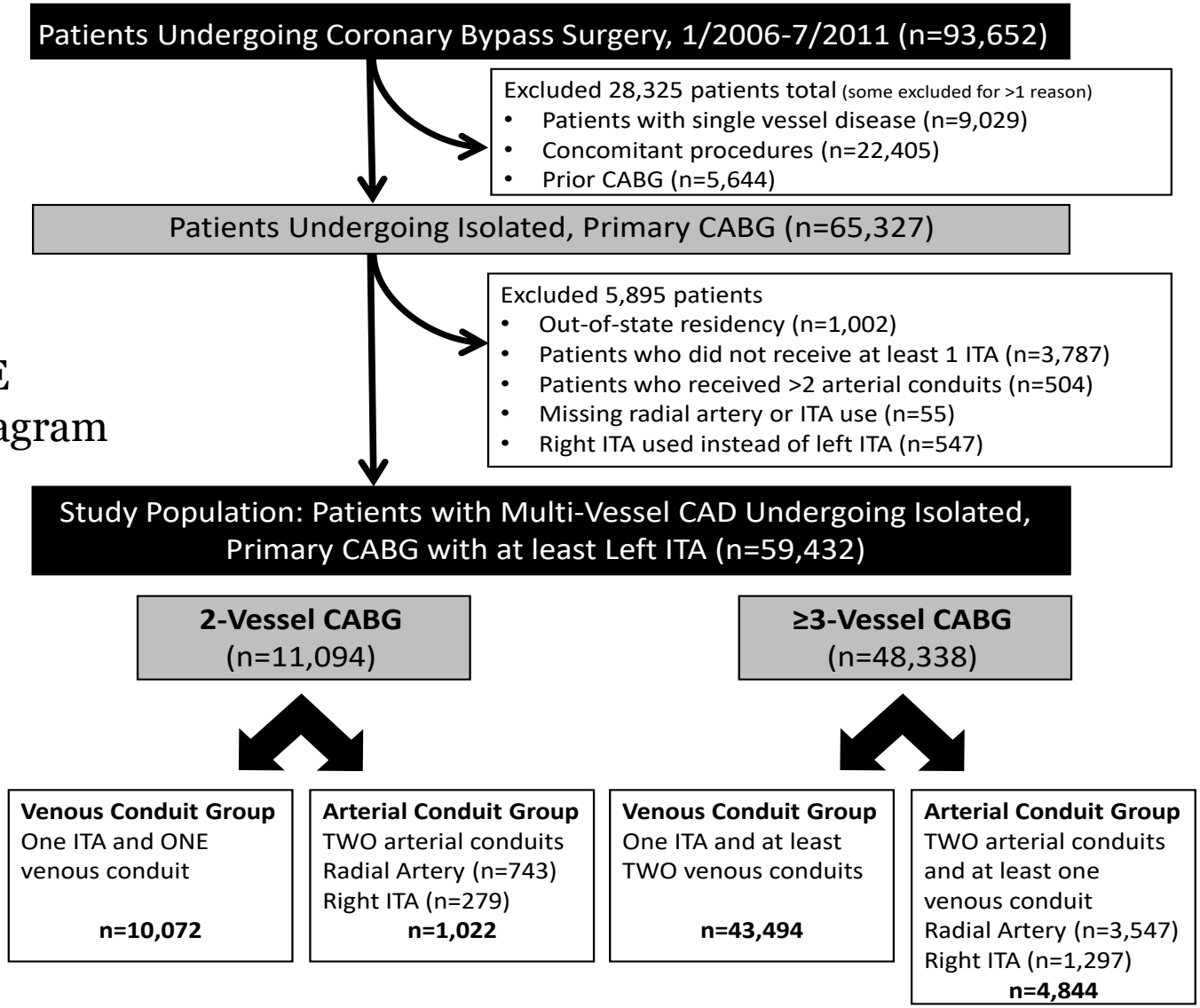


Figure 1: Participant flow diagram for this study. See "Losses and Exclusions" for more discussion.

Figure 1: **Patient selection flow diagram**

STROBE Flow Diagram



CAD, coronary artery disease; CABG, coronary artery bypass grafting; ITA, internal thoracic artery

treatment effect on [insert group]

- Effect estimates: ATE, TonT, TonC and CACE
- Causal effect of the treatment

$$Y_i(D_i = 1) - Y_i(D_i = 0) = \Delta_i$$

- Average Treatment Effect

$$E_i[Y_i(D_i = 1) - Y_i(D_i = 0)] = \overline{\Delta}_i$$

- Treatment effect on the Treated

$$E_i[Y_i(D_i = 1) - Y_i(D_i = 0)|d_i = 1] = \overline{\Delta}_i^{\,T}$$

- Treatment effect on the Control

$$E_i[Y_i(D_i = 1) - Y_i(D_i = 0)|d_i = 0] = \overline{\Delta}_i^{\,C}$$

- Complier average causal effect

$$E_i[Y_i(D_i = 1) - Y_i(D_i = 0)|i \in complier] = \overline{\Delta}_i^{\,IV}$$

treatment effect on [insert group]

observation	Y(d=1)	Y(d=0)	delta	dose
1	8	9	-1	0
2	5	3	2	1
3	4	5	-1	0
4	6	7	-1	0
5	10	11	-1	0
6	3	4	-1	0
7	3	1	2	1
8	-1	0	-1	0
9	5	6	-1	0
10	2	0	2	1

treatment effect on [insert group]

observation	Y(d=1)	Y(d=0)	delta	dose	included in effect
1	8	9	-1	0	-1
2	5	3	2	1	2
3	4	5	-1	0	-1
4	6	7	-1	0	-1
5	10	11	-1	0	-1
6	3	4	-1	0	-1
7	3	1	2	1	2
8	-1	0	-1	0	-1
9	5	6	-1	0	-1
10	2	0	2	1	2
				average:	0

Average Treatment Effect

treatment effect on [insert group]

observation	Y(d=1)	Y(d=0)	delta	dose	included in effect
1	8	9	-1	0	
2	5	3	2	1	2
3	4	5	-1	0	
4	6	7	-1	0	
5	10	11	-1	0	
6	3	4	-1	0	
7	3	1	2	1	2
8	-1	0	-1	0	
9	5	6	-1	0	
10	2	0	2	1	2
				average:	2

Treatment Effect on the Treated

treatment effect on [insert group]

observation	Y(d=1)	Y(d=0)	delta	dose	included in effect
1	8	9	-1	0	-1
2	5	3	2	1	
3	4	5	-1	0	-1
4	6	7	-1	0	-1
5	10	11	-1	0	-1
6	3	4	-1	0	-1
7	3	1	2	1	
8	-1	0	-1	0	-1
9	5	6	-1	0	-1
10	2	0	2	1	
				average:	-1

Treatment Effect on the Control

a second outcome

Design of Observational Studies: chapter 5.2.3 and 5.2.4
Rosenbaum, "The Role of Known Effects in Observational Studies"

the structure of the argument: two outcomes

- If your theory is well developed then you might be able to locate multiple outcomes that will support your understanding of the mechanism of the intervention.*
- Two ways this can happen:
  - The second outcome can be compatible (show violation)
  - The confirmation of a "null effect" can help rebuff claims of unobserved biases

*Keep this idea separate from "intermediate effects," not because there's a deep fundamental difference in these concepts but rather conflating them will tend to confuse discussions.

coherence

- (Rough) Definition: A claim is made that an intervention must have a certain form (i.e., there's a detailed hypothesis). In this situation, coherence means a pattern of observed associations compatible with this anticipated form, and incoherence means a pattern of observed associations incompatible with this form.
- Claims of coherence or incoherence are arguable to the extent that the anticipated form of treatment effect is arguable.
- If you want to see the technical details of how to build a statistical argument around this then check out *Observational Studies*, section 17.2 (coherent signed rank statistic).

the structure of the argument: null effect outcomes

- Basic idea: Suppose that a treatment is known to not change a particular outcome. Then if we see differences between the treatment and control groups on this particular outcome, this must mean that there are differences between the treatment and control group on unmeasured covariates and thus there is hidden bias.

example: methylmercury fish

- Example: Skerfving (1974) studied whether eating fish contaminated with methylmercury causes chromosome damage. The outcomes of interest was the percentage of cells exhibiting chromosome damage. Pairs were matched for age and sex.

example: methylmercury fish

- Example: Skerfving (1974) studied whether eating fish contaminated with methylmercury causes chromosome damage. The outcomes of interest was the percentage of cells exhibiting chromosome damage. Pairs were matched for age and sex.

```
control.cu.cells <- c(2.7,.5,0,0,5,0,0,1.3,0,1.8,0,0,1,1.8,0,3.1)
exposed.cu.cells <- c(.7,1.7,0,4.6,0,9.5,5,2,2,2,1,3,2,3.5,0,4);
library(exactRankTests)

wilcox.exact(exposed.cu.cells,control.cu.cells,paired=TRUE)

        Exact Wilcoxon signed rank test

data:  exposed.cu.cells and control.cu.cells
V = 84, p-value = 0.04712

alternative hypothesis: true mu is not equal to 0
```

example: methylmercury fish

- In the absence of hidden bias, there's evidence that eating large quantities of fish containing methylmercury causes chromosome damage.
- Going further, *Skerfving* described other health conditions of these subjects including other diseases such as (i) hypertension, (ii) asthma, (iii) drugs taken regularly, (iv) diagnostic X-rays over the previous three years, (v) and viral diseases such as influenza.
- These can be considered outcomes since they describe the period when the exposed subjects were consuming contaminated fish.
- However, it is difficult to imagine that eating fish contaminated with methylmercury causes influenza or asthma, or prompts X-rays of the hip or lumbar spine.

example: methylmercury fish

- The data

```
control.other.health.conditions <- c(rep(0,8),2,rep(0,3),2,1,4,1)
exposed.other.health.conditions <- c(0,0,2,0,2,0,0,1,1,2,0,9,0,0,1,0)
```

```
> wilcox.exact(control.other.health.conditions,exposed.other.health.conditions)

        Exact Wilcoxon rank sum test

data:  control.other.health.conditions and exposed.other.health.conditions
W = 112.5, p-value = 0.5257
alternative hypothesis: true mu is not equal to 0
```

- There is no evidence of hidden bias.
- But absence of evidence is not evidence of absence.

example: methylmercury fish

- Questions:

  (1) When does such a test have a reasonable prospect of detecting hidden bias?

  (2) If no evidence of hidden bias is found, does this imply reduced sensitivity to bias in the comparisons involving the outcomes of primary interest?

  (3) If evidence of bias is found, what can be said about its magnitude and its impact on the primary comparisons?

null effect outcomes

- Power of the test of hidden bias: Let ***y*** denote the outcome for which there is a known effect of zero. For a particular unobserved covariate ***u***, what unaffected outcome ***y*** would be useful in detecting hidden bias from ***u***?
- Precise statement of results in:
  - "The Role of Known Effects in Observational Studies"
- Basic result: The power of the test of whether ***y*** is affected by the treatment increases with the strength of the relationship between ***y*** and ***u***. If one is concerned about a particular unobserved covariate ***u***, one should search for an unaffected outcome ***y*** that is strongly related to ***u***.

takeaway

- Having a detailed understanding of how your intervention functions, what the causal pathway includes and excludes, will give you more data sources that may validate or refute your hypothesis.
- Coherence is trying to flesh out your hypothesis.
- Known null effects may help to address unobserved confounding

a second control group

TWO PROBLEMS
TWO CONTROLS

Design of Observational Studies: chapter 5.2.2
Rosenbaum, "The Role of a Second Control Group in an Observational Study"

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs).
- In an observational study there are possibly many different reasons for people to have not received the treatment.
- In some situations there are discernable subgroups within the non-treatment group, each subgroup being identifiable by the reason for the subgroup not receiving the treatment.
- In some subset of these situations these subgroups will be open to critiques of bias when compared to the treatment group, but at least two of the subgroups will differ in the nature of their bias.
- The contrast of these two control groups with the treatment group may strengthen your analysis.

second control group: army toxicity

- Example: The army is interested in the long term effects of exposure to a list of specific chemical agents that were suspected of being toxic. Relatively few soldiers were exposed to these chemicals.
- At first pass, one might think to compare these exposed ("treated") service members to service members who were not exposed at all.
- Complicating that comparison, though, is that the army sorted people into jobs which exposed them or to jobs which did not.
- The army used medical examinations – which were not well documented – to sort some individuals out of high-exposure jobs. This leaves the comparison between exposed and strictly unexposed potentially biased due to baseline conditions.

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects.

second control group: army toxicity

- A second control group was constructed using service members who were in jobs which exposed them to chemical agents, but not the specific list of chemical agents under consideration. These other chemical agents were thought to have little or no longer term effects. Thus this group is thought to have received an “ineffective dose” of the exposure.
- Each of these control groups is problematic: the first group is open to critiques of baseline differences in medical conditions; the second group has individuals who were potentially exposed to actively toxic chemical agents.
- But the first control group is unlikely to suffer from the bias encounter in the second control group, and vice versa.

second control group: army toxicity

- The hope is that the two control groups will not differ from each other in a meaningful way.
- A rejection of a test of equivalency between the control groups is a strong warning sign of potential bias.
- A non-rejection may arise for several reasons. A false-negative would be problematic.
- The hope is that the control reservoir (i.e., the ratio of controls to treated observations) is large enough that we can reach adequate levels of statistical power for our tests.
- Precise statements of how this argument works statistically, as well as a couple more examples from the literature, can be found in "The Role of a Second Control Group in an Observational Study"

structure of argument

- In an RCT the control and treatment groups are created from a pool of study participants. The assignment to C or T is due to a researcher-directed mechanism (e.g., flipping a coin, or matched pairs). All participants can receive either T or C.
- In an observational study there are possibly many different reasons for people to have not received the treatment.
- In some situations there are discernable subgroups within the non-treatment group, each subgroup being identifiable by the reason for the subgroup not receiving the treatment.
- In some subset of these situations these subgroups will be open to critiques of bias when compared to the treatment group, but at least two of the subgroups will differ in the nature of their bias.
- The contrast of these two control groups with the treatment group may strengthen your analysis.

fin.