Lecture #5

Identification of Point Treatment Causal Parameters

A roadmap for causal inference

1. Specify **Causal Model** representing real background knowledge
2. Specify **Causal Question**
3. Specify **Observed Data** and link to causal model
4. **Identify** : Knowledge + data sufficient?
5. Commit to an **estimand** as close to question as possible, and a **statistical model** representing real knowledge.
6. **Estimate**
7. **Interpret** Results

Overview: Identifiability of point treatment effects

1. Randomization Assumption
2. Back door criterion
3. G-computation formula
4. Working SCMs or ("what to do when your knowledge is not sufficient to identify the causal effect you care about")
5. Positivity Assumption

References

- TLB. Chapter 2
- Pearl. “An Introduction to Causal Inference” *Int J Biostat*, 6(2): Article 7, 2010.
- Greenland and Pearl. “Causal Diagrams” in S. Boslaugh, editor, Encyclopedia of Epidemiology. Sage Publications, Thousand Oaks, CA, 2007
- Greenland, Pearl and Robins. “Causal Diagrams for Epidemiologic Research” *Epidemiology*, 10(1): 37-48, 1999.
- Robins and Hernan. “Estimation of the causal effects of time-varying exposures” In Fitzmaurice, Davidian, Verbeke, Molenberghs, editors, Longitudinal Data Analysis, chapter 23. Chapman & Hall/CRC Press, Boca Raton, FL, 2009
- Glymour. “Using casual diagrams to understand common problem in social epidemiology” In: Methods in Social Epidemiology
- Petersen, et al. “Diagnosing and responding to violations in the positivity assumption” *Statistical Methods in Medical Research*, 21(1):31-54, 2012

Overview of Identifiability

- What we want (target of inference): $\Psi^F(P_{U,X})$
  - Ex. $$\Psi^F(P_{U,X}) = E_{U,X}(Y_1 - Y_0)$$
- What we have: a sample from the observed data distribution
  - Ex. n i.i.d. observations of $O \sim P_0$
  - Can use this to make inferences about parameters of the observed data distribution: $\Psi(P_0)$

Overview: Identifiability

- Identifiability in a nutshell:

  Can we write $\Psi^F(P_{U,X})$ as $\Psi(P_0)$?

- Slightly more formally, we need that:

  For each $P_{U,X}$ in $\mathcal{M}^{\mathcal{F}}$ (each $P_{U,X}$ compatible with the SCM) we have that $\Psi^F(P_{U,X}) = \Psi(P_0)$ for some Ψ

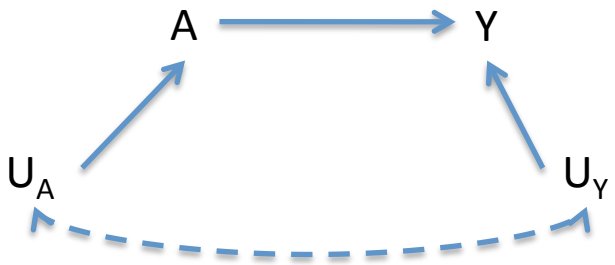
Overview: Identifiability

- Lots of work (across many disciplines) has gone into identifiability problems
  - In the SEM framework
  - In the missing data framework
  - In the causal graph framework

- Theorems exist that tell us what assumptions are sufficient for a given identifiability result to hold
  - See Shpitser 2008; Tian 2002

Simple Example: What is needed for identifiability?

- Target: $\Psi^F(P_{U,X})=E_{U,X}(Y_1-Y_0)$
- Observe: $O=(A,Y)\sim P_0$
- What do we need to assume to be able to write $\Psi^F(P_{U,X})$ as $\Psi(P_0)$?



Simple Example: Identifiability

- What assumption would give us the following $P_0(Y=y|A=a)=P_{U,X}(Y_a=y)$?

$\Psi(P_0)$ $\Psi^F(P_{U,X})$

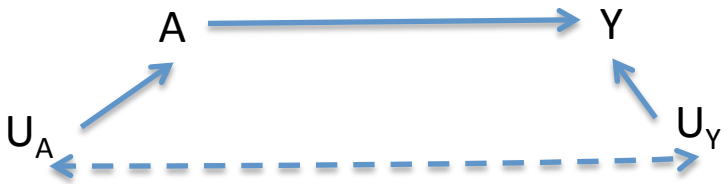
- $P_0(Y=y|A=a)=P_{U,X}(Y_a=y|A=a)$
  - By definition of counterfactuals
- $P_{U,X}(Y_a=y|A=a)=P_{U,X}(Y_a=y)$
  - If Y_a independent of A
  - When does the SCM imply that this holds?

Simple Example: Identifiability

When is Y_a independent of A?

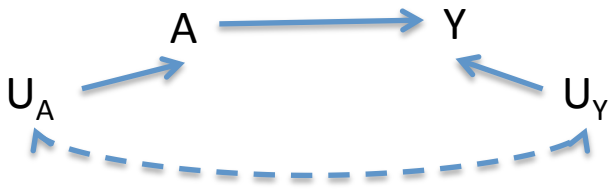
- $Y_a(U)=f_Y(a,U_Y)$
- Once you set a, Y_a is only a function of its error U_Y

- If U_Y is independent of A, then Y_a is independent of A
- If U_Y independent of U_A, then Y_a is independent of A, and $E_{U,X}(Y_a)$ is identified as a parameter of P_0



Intuition for this result

- We need that $E(Y|A=1)-E(Y|A=0) = E(Y_1-Y_0)$
- For this to hold we must be sure that <u>all</u> of any <u>observed</u> association between A and Y is due to the causal effect we are interested in
  - Thus we need that there are no additional potential sources of association between A and Y
  - In other words, no unmeasured common causes
  - This is the basic Epi concept of confounding



Identifiability results

- We've derived what we need for one simple example…
- What if things get more complicated?
  - Ex. Measure more covariates
  - Ex. Multiple intervention nodes
- We can do the same sort of exercise and figure out what assumptions we need
- For many common problems, helpful theorems have been developed

Identifiability for Point Treatment

- Focus today on identifiability for the effect of a single intervention (point treatment) when baseline covariates have been measured

- We will focus on one identifiability result:
  - "G-computation formula"
- Holds under
  - Randomization assumption
  - Backdoor criterion

Example: Identifiability Problem

- SCM $\mathcal{M}^{\mathcal{F}}$:
  - $X=(W,A,Y)$; $U=(U_W,U_A,U_Y)\sim P_U$
  - No exclusion restrictions or independence assumptions
- Observe: $O=(W,A,Y)\sim P_0$
- Statistical model $\mathcal{M}$ is non-parametric
- Target: $\Psi^F(P_{U,X})=E_{U,X}(Y_1-Y_0)$
- Can we write $\Psi^F(P_{U,X,0})$ as a parameter of P_0?

What can we assume that will identify our target causal parameter as a parameter of of the observed data distribution?

- Randomization Assumption: $Y_a \perp A|W$

1. Why is this sufficient?
2. When will it hold?
   - What independence assumptions are needed?
   - How can we assess whether it holds based on the graph?

Identifiability of Point Treatment Effects under the Randomization Assumption

- Randomization Assumption (RA):

$$Y_a \perp A|W$$

- Identifiability Result

$$P_0(Y = y|A = a, W = w) = P_{U,X}(Y_a = y|A = a, W = w)$$

By definition of counterfactuals

$$= P_{U,X}(Y_a = y|W = w)$$

Under the randomization assumption

Identifiability of Point Treatment Effects under the Randomization Assumption

- If the Randomization Assumption $Y_a \perp A|W$ holds then:

$$E_{U,X}(Y_a|W=w) = E_0(Y|A=a,W=w)$$

- This gives us the G-computation formula

$$E_{U,X}(Y_a) = \sum_w E_0(Y|A=a,W=w)P_0(W=w)$$

$\Psi^F(P_{U,X})$

$\Psi(P_0)$: "estimand"

Backdoor criterion

- Plausibility of the randomization assumption can be hard to assess.
  - What variables to include in W? Are they sufficient?
- Alternative: Graphical criteria for establishing whether a given adjustment set is sufficient
  - If W satisfies backdoor criterion, the effect of A on Y is identified via the G-computation formula

$$E_{U,X}(Y_a) = \sum_w E_0(Y|A=a,W=w)P_0(W=w)$$

$\Psi^F(P_{U,X})$ $\Psi(P_0)$

The Back-door Criterion

- Conditional on W, we want to be sure that any observed association between A and Y is due to the effect of A on Y we are interested in

$$E_0(Y|A=1,W) - E_0(Y|A=0,W) = E_{U,X}(Y_1|W) - E_{U,X}(Y_0|W)$$

- This means we need W to

1. Block all spurious sources of association
2. Not create any new spurious sources of association
3. Leave the path we are interested in unperturbed

The Back-door Criterion

- Conditional on W, we want to be sure that any observed association between A and Y is due to the effect of A on Y we are interested in

- This tells us what characteristics W should have

1. W does block any association between A and Y that arises from unmeasured common causes
2. W does not create any new non-causal associations between A and Y
3. W does not block any of the effect of A on Y

Back-door criterion

- A set of variables W satisfies the back door criterion with respect to (A,Y) if
  1. No node in W is a descendent of A
  - Motivation:
    1. Avoid blocking the path of interest
    2. Avoid introducing spurious sources of dependence
  2. W blocks all “backdoor” paths from A to Y
  - Backdoor path= path with arrow into A
  - Motivation: Block all sources of spurious association between A and Y

Recall: when is a path blocked?

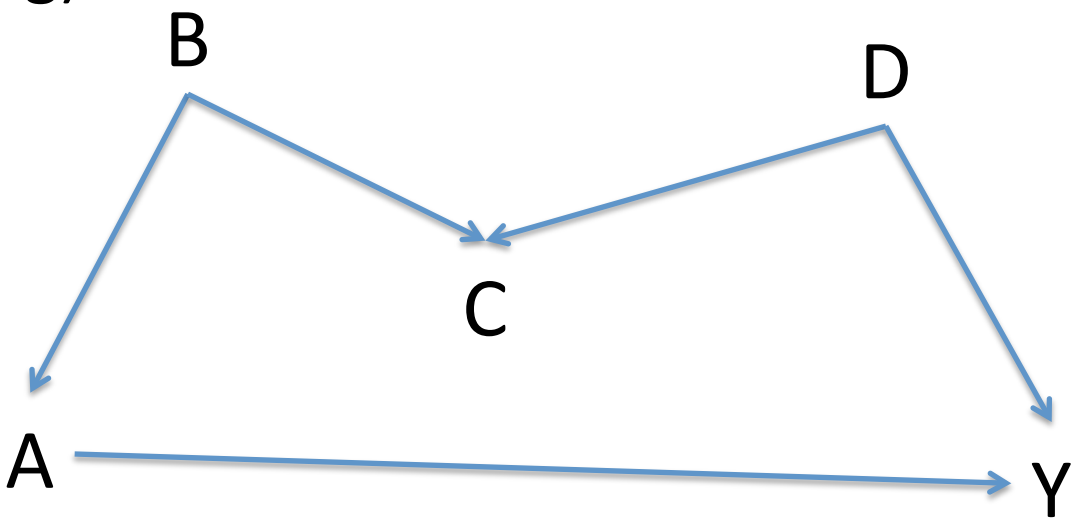
- If it has a non-collider that has been conditioned on

or

- If it has a collider *and* neither the collider nor a descendent of the collider has been conditioned on

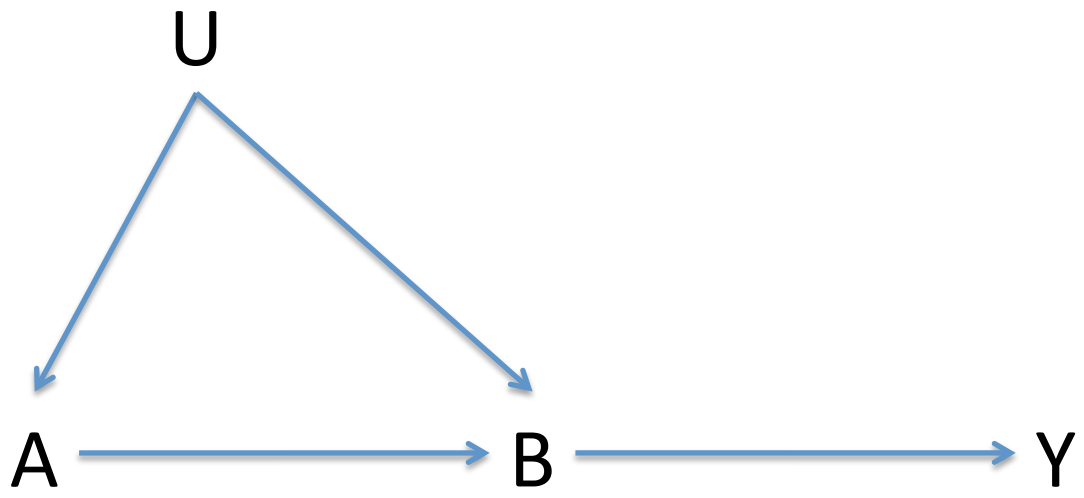
Example

- Back door criterion satisfied for the effect of A on Y by:
  - {} (nothing)?
  - {C}?
  - {B,C}?



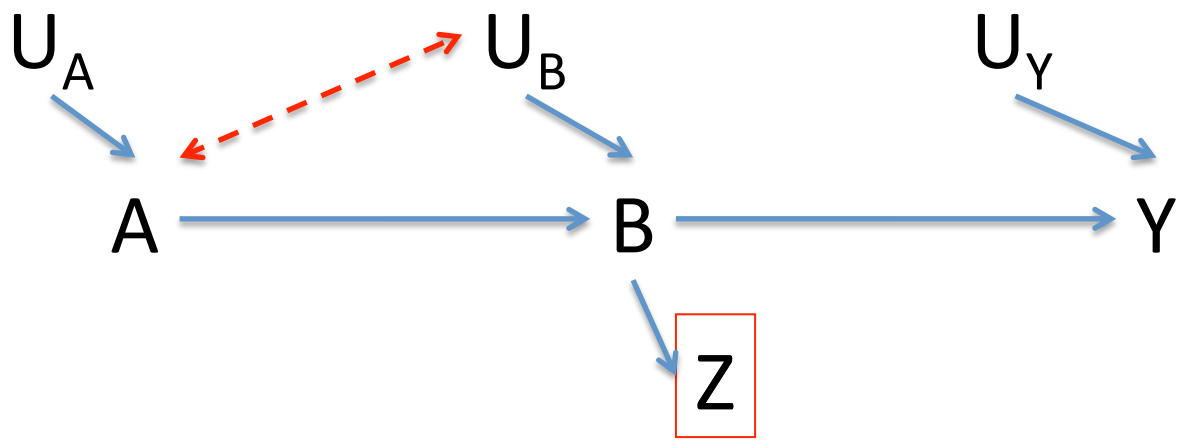
Why are descendents of A excluded?

1. Avoid blocking the path we are interested in
   - Does {} satisfy the back-door criterion?
   - Does B satisfy the back door criterion?



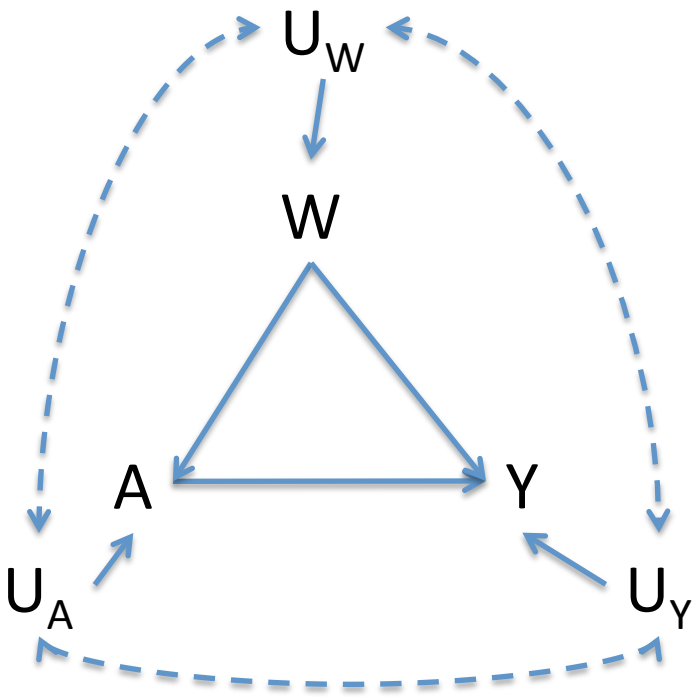
Why are descendents of A excluded?

2. Avoid introducing new sources of spurious association between A and Y
   - Does {} satisfy the back-door criterion?
   - Does Z satisfy the back door criterion?
   - What happens if you condition on Z?
     - Not part of the causal pathway of interest



Example

O=(W,A,Y)

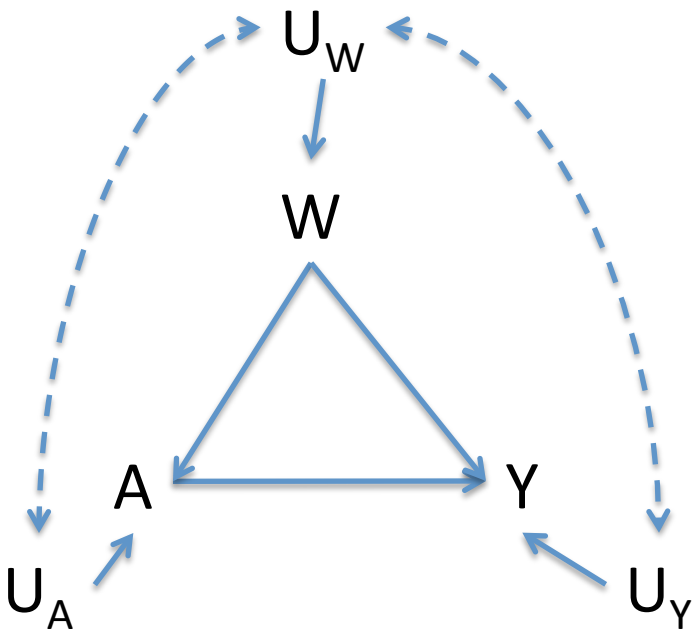


- Is there a subset of the observed covariates that satisfies the back door criterion?
- What is it?

- $E[Y_1-Y_0]$ identified?
- If, so, what is the target parameter of the observed data distribution (estimand)?

Example

O=(W,A,Y)

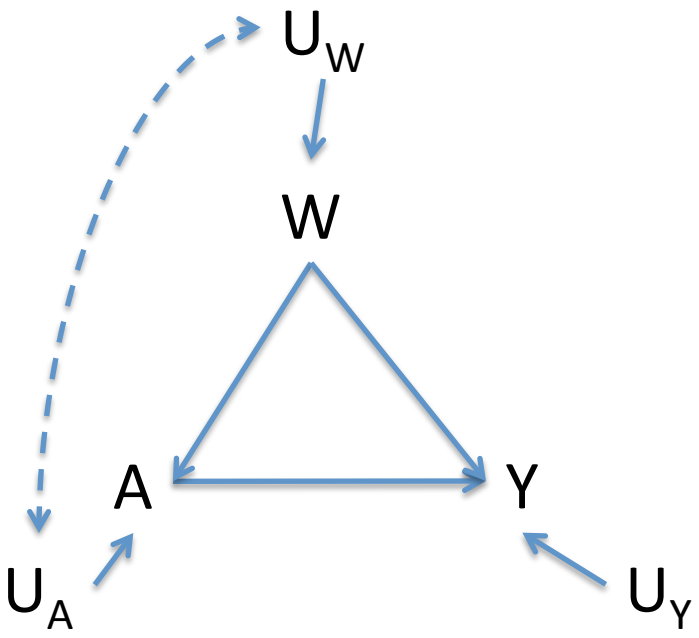


- Is there a subset of the observed covariates that satisfies the back door criterion?
- What is it?

- $E[Y_1-Y_0]$ identified?
- If, so, what is the target parameter of the observed data distribution (estimand)?

Example

O=(W,A,Y)

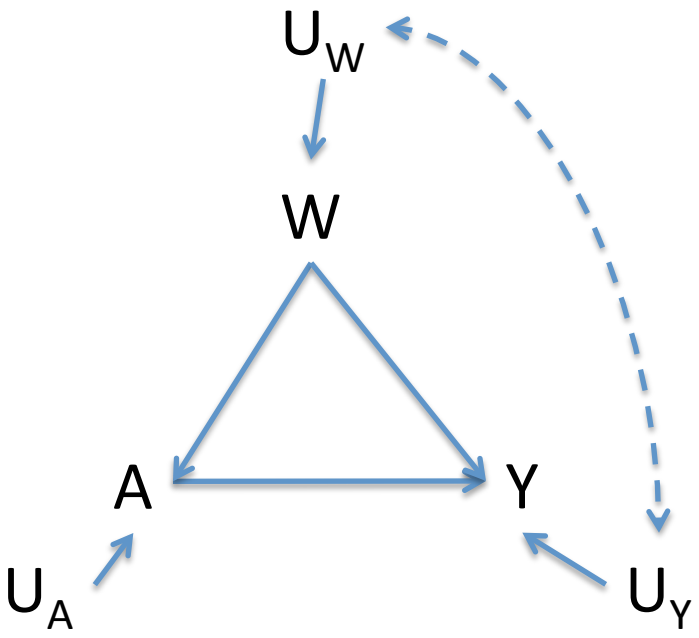


- Is there a subset of the observed covariates that satisfies the back door criterion?
- What is it?

- $E[Y_1-Y_0]$ identified?
- If, so, what is the target parameter of the observed data distribution (estimand)?

Example

O=(W,A,Y)



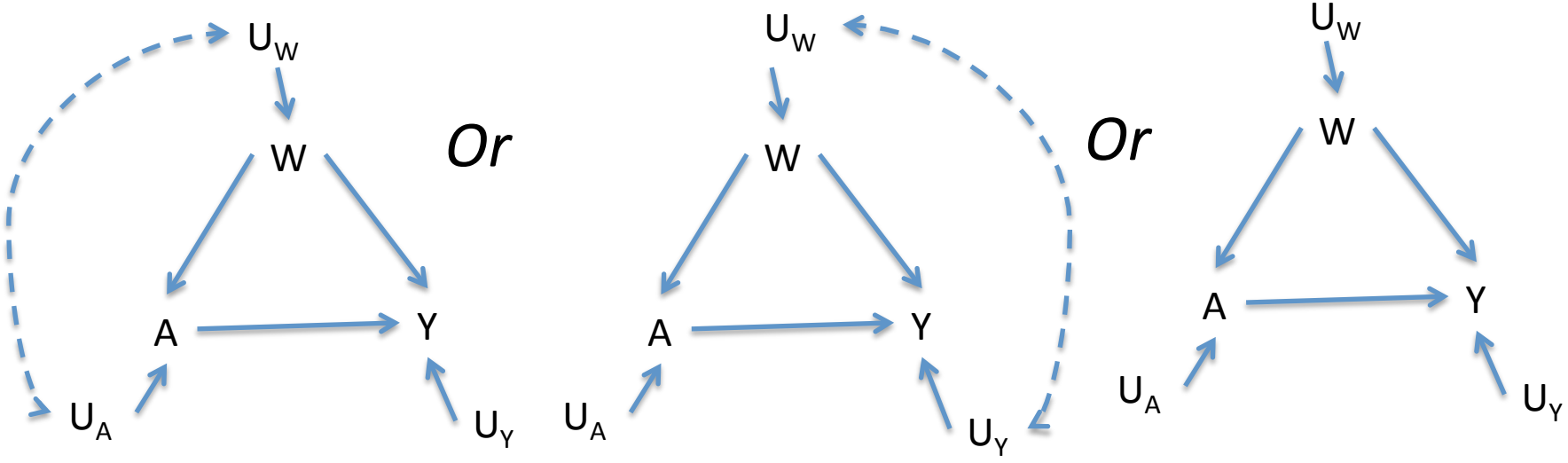
- Is there a subset of the observed covariates that satisfies the back door criterion?
- What is it?

- $E[Y_1-Y_0]$ identified?
- If, so, what is the target parameter of the observed data distribution (estimand)?

Summary: Identifiability for Point Treatment Effects with basic structure

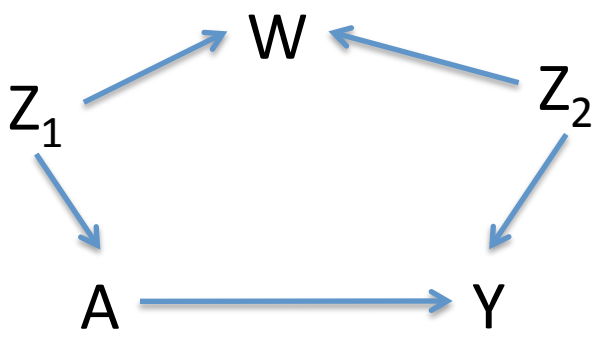
- Under what sets of independence assumptions will the G-computation identifiability result hold?

$$E_{U,X}(Y_a) = \sum_w E_0(Y|A=a, W=w)P_0(W=w)$$



Example

O=(W,A,Y)



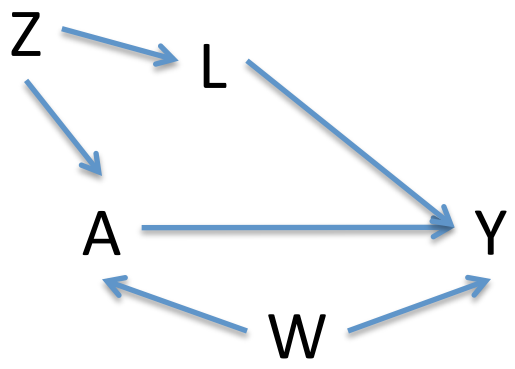
- Is there a subset of the observed covariates that satisfies the back door criterion?
- What is it?

- $E[Y_1-Y_0]$ identified?
- If so, what is the target parameter of the observed data distribution?

Example

O=(W,A,L,Y)

(L after A)

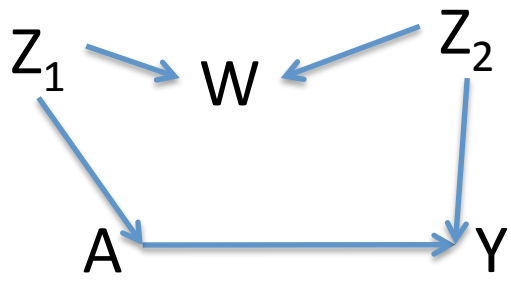


- Is there a subset of the observed covariates that satisfies the back door criterion?
- What is it?

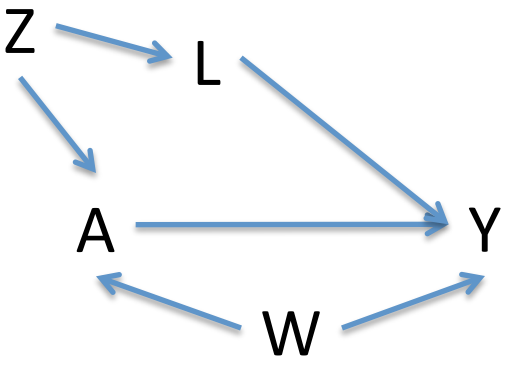
- $E[Y_1-Y_0]$ identified?
- If so, what is the target parameter of the observed data distribution?

In Summary: Conditioning on the whole past and only the past is not always a good idea…

- Ex 1. O=(W,A,Y); Z occurs before A
  - RA fails conditional on W
  - RA holds conditional on {}



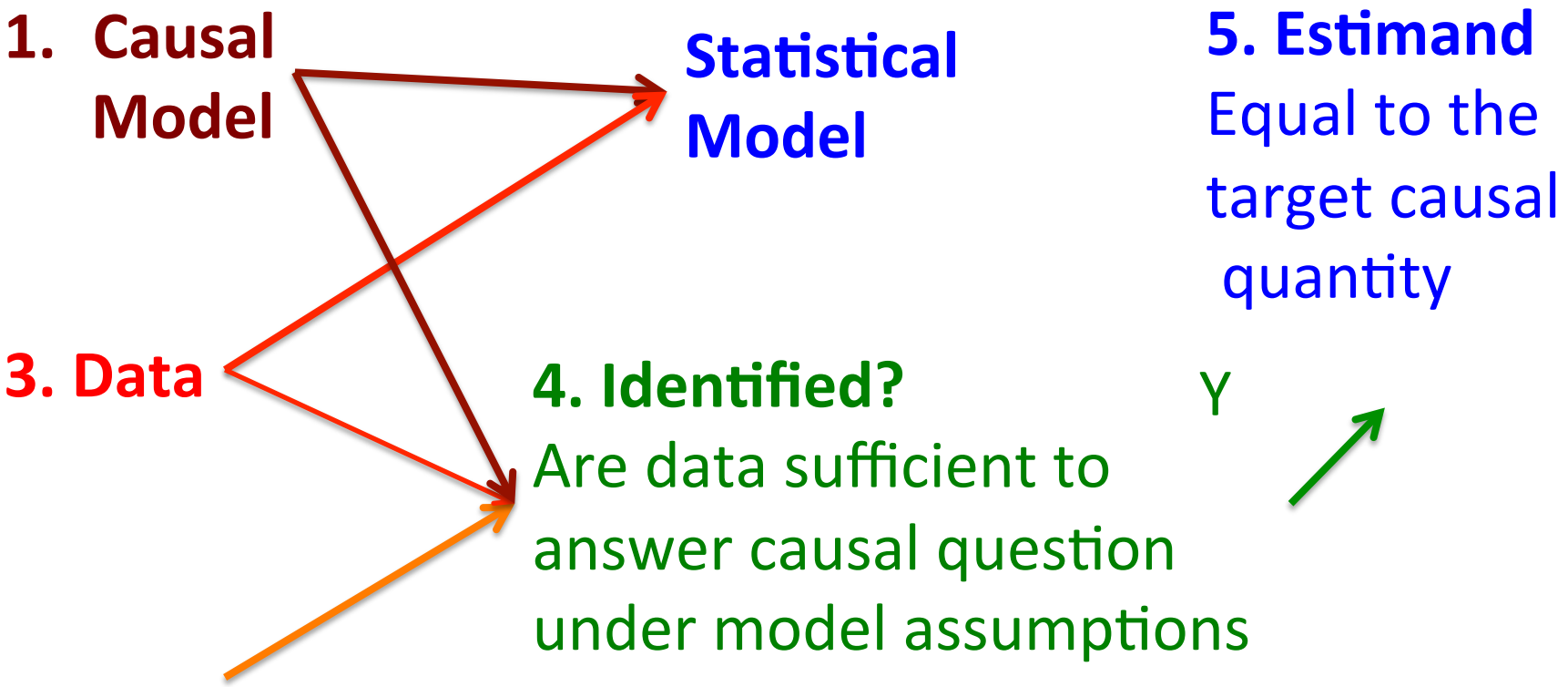
- Ex 2. O=(W,A,L,Y); L occurs after A
  - RA fails conditional on W
  - RA holds conditional on (W,L)



Where are we?

- We specified a structural causal model: $\mathcal{M}^{\mathcal{F}}$
  - $(U,X) \sim P_{U,X}$ in $\mathcal{M}^{\mathcal{F}}$
- We specified counterfactuals and a target parameter of their distribution
  - Ex. $\Psi^F(P_{U,X})=E_{U,X}(Y_1-Y_0)$
- We specified our observed data and statistical model
  - $O \sim P_0$ in $\mathcal{M}$
- We established assumptions under which $\Psi^F(P_{U,X})=\Psi(P_0)$ for some Ψ
  - For one class of target parameter indexed by a static intervention on a single variable

A Roadmap….


1. Causal Model
Statistical Model
5. Estimand
Equal to the target causal quantity
3. Data
4. Identified?
Are data sufficient to answer causal question under model assumptions
Y
2. Question
Translate the scientific question into a formal causal quantity (using counterfactuals)

Our initial model assumptions are not sufficient. Now what?

- If we are honest with ourselves about the limits of what we know, this happens a lot!
  - $\Psi^F(P_{U,X})$ is not identified under $\mathcal{M}^{\mathcal{F}}$
- Options
  - Go get some more data/background research
  - Give up
- But…. Lots of questions require a timely "best guess" to inform ongoing decisions !?!
  - Goal: Get the best answer you can and be honest and transparent when interpreting results

Our initial model assumptions are not sufficient. Now what?

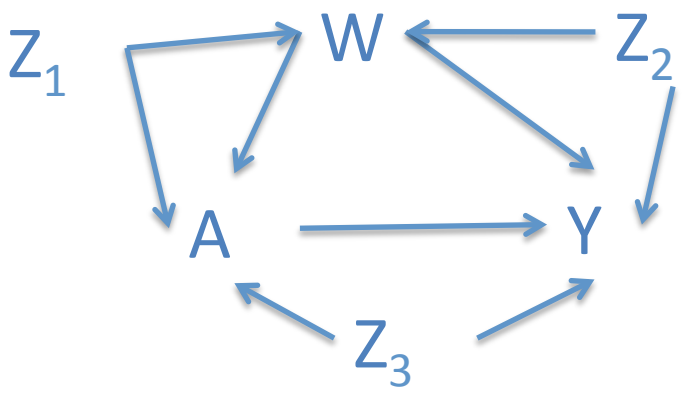
- $\Psi^F(P_{U,X})$ is not identified under $\mathcal{M}^{\mathcal{F}}$
  - We know which additional assumptions would serve to identify $\Psi^F(P_{U,X})$
- We will use $\mathcal{M}^{\mathcal{F}^*}$ to refer to the original SCM + these additional assumptions
- This gives us a way to proceed, while keeping separate our real knowledge and our wished for identifiability assumptions
  - Useful in the interpretation stage!

Example: Identifiability Problem

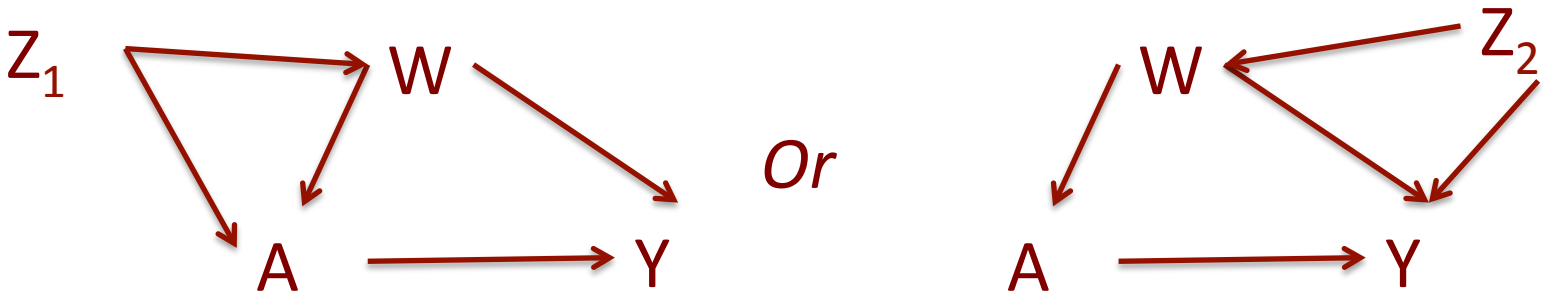
- $\mathcal{M}^{\mathcal{F}}$: X=(W,A,Y), U=($U_W$,$U_A$,$U_Y$)
  - No exclusion restrictions or independence assumptions: non-parametric model
- Observe: O=(W,A,Y)~P_0
- Target: $\Psi^F(P_{U,X})=E_{U,X}(Y_1-Y_0)$
- Do we have $\Psi^F(P_{U,X})=\Psi(P_0)$ for some Ψ?

Example: A "working" SCM; O=(W,A,Y)

Original SCM: $\mathcal{M}^{\mathcal{F}}$



$\mathcal{M}^{\mathcal{F}}$ + additional assumptions= $\mathcal{M}^{\mathcal{F}*}$



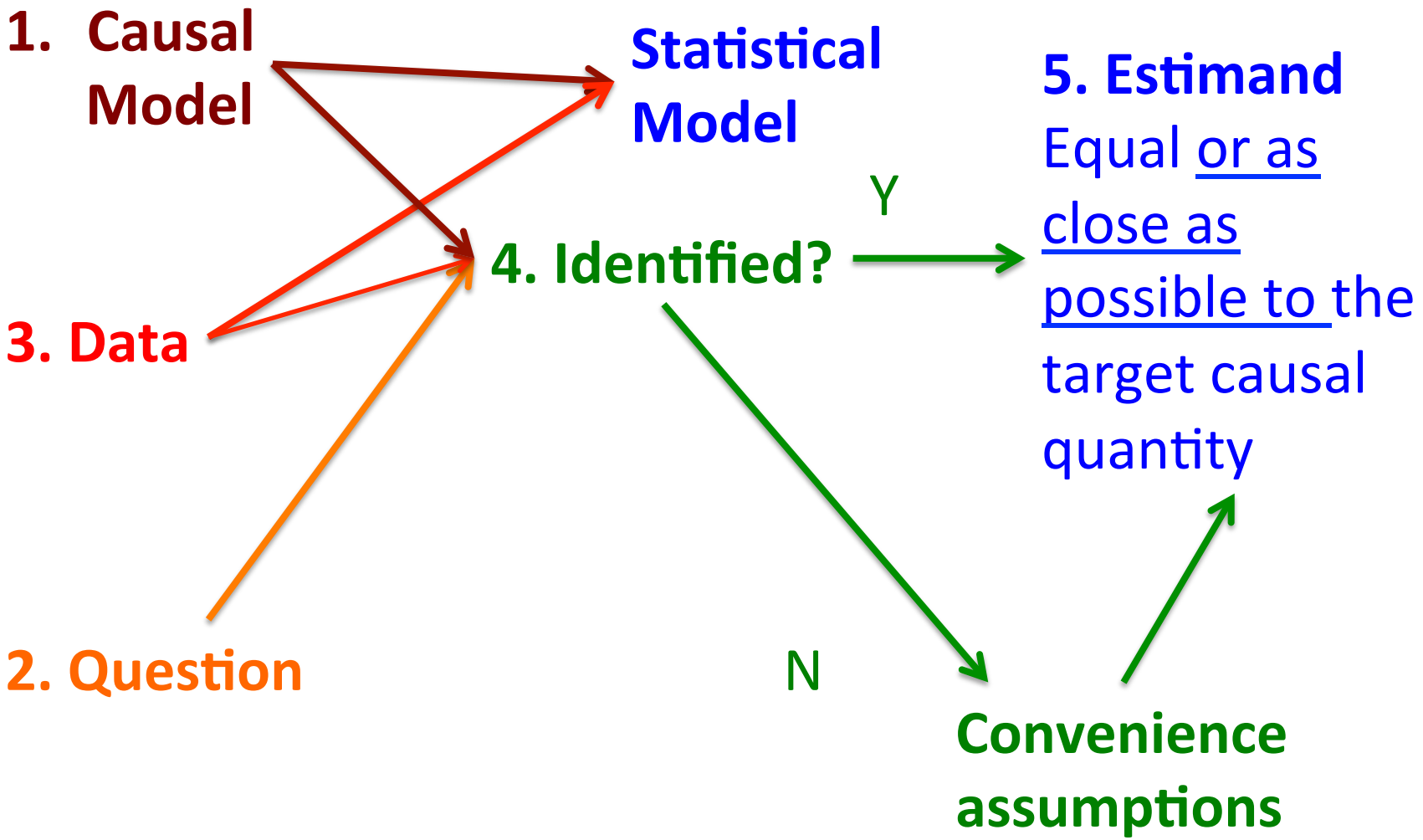
Example: Working SCM

- $\Psi^F(P_{U,X})$ is identified under $\mathcal{M}^{\mathcal{F}^*}$ using the G computation formula

$$\underbrace{E_{U,X}(Y_a)}_{\Psi^F(P_{U,X})} = \underbrace{\sum_w E_0(Y|A=a,W=w)P_0(W=w)}_{\Psi(P_0)}$$

- Also have us a statistical model $\mathcal{M}$
  - Prefer model implied by $\mathcal{M}^{\mathcal{F}}$ versus $\mathcal{M}^{\mathcal{F}^*}$
  - Ensure statistical model contains the truth
- We can now proceed to estimation…

A Roadmap….


1. Causal Model
Statistical Model
5. Estimand
Equal or as close as possible to the target causal quantity
Y
4. Identified?
3. Data
2. Question
N
Convenience assumptions

Summary: G-computation formula for point treatment effects

- When the backdoor criterion is satisfied by some subset of observed variables W we have the G-computation formula:

$$E_{U,X}(Y_a) = \sum_w E_0(Y|A=a, W=w)P_0(W=w)$$

  - Also holds under the randomization assumption

$$Y_a \perp A|W$$

- Covers a large category of target causal parameters
  - Parameters of counterfactual distributions indexed by a static intervention on a single node
    - "point treatment" effects
  - Not limited to the average treatment effect

Positivity Assumption

- In order to identify $E(Y_a)$ using the G computation formula, need $E(Y|A=a,W=w)$ to be well-defined for all possible values (a,w)

$$E_{U,X}(Y_a) = \sum_w E_0(Y|A=a,W=w)P_0(W=w)$$

  - In non-parametric model, each treatment of interest must occur with some positive probability for each possible covariate history

- Positivity Assumption
  - Assumption of experimental treatment assignment (ETA)

$$\min_{a \in \mathcal{A}} P_0(A=a|W=w) > 0,$$

$$\text{for all } w \text{ for which } P_0(W=w) > 0$$

Example: Positivity Violation

- We want to know the average treatment effect of an exposure in a population that contains men and women
- $O=(W,A,Y)\sim P_0$
  - $W=I(\text{woman})$; $0<P_0(W=1)<1$
- Target causal quantity: $E(Y_1-Y_0)$
- Assume: W blocks all backdoor paths from A to Y
- Estimand: $E_{0w}[E_0(Y|A=1,W)-E_0(Y|A=0,W)]$

  $= [E_0(Y|A=1,W=1)-E_0(Y|A=0,W=1)]*P_0(W=1)+$

  $[E_0(Y|A=1,W=0)-E_0(Y|A=0,W=0)]*P_0(W=0)$

Example: Positivity Violation

- We want to know the average treatment effect of an exposure in a population that contains men and women
- Estimand: $E_{0w}[E_0(Y|A=1,W)-E_0(Y|A=0,W)]$
  - $[E_0(Y|A=1,W=1)-E_0(Y|A=0,W=1)]*P_0(W=1)+$
    $[E_0(Y|A=1,W=0)-E_0(Y|A=0,W=0)]*P_0(W=0)$
- No women in this population get the exposure
  - $P_0(A=1|W=1)=0$
- No information about outcomes of exposed women
  - $E_0(Y|A=1,W=1)=????$

Example: Positivity Violation

- When would the ATE be identified in such a case?

1. Different target parameter
   - Ex. the effect of the exposure among men
     $E(Y|A=1,W=0)-E(Y|A=0,W=0)$
   - Changes target population….

2. Additional model assumptions
   - Ex. exposure has the same effect in women as in men

$E(Y|A=1,W=0)-E(Y|A=0,W=0)$

$=E(Y|A=1,W=1)-E(Y|A=0,W=1)$

$=E_W(E(Y|A=1,W)-E(Y|A=0,W))$

"Practical violations" of the positivity assumption

- Even if the positivity assumption holds for P_0, it may be "practically violated" for a given sample P_n from P_0
- Example
  - Probability of getting the exposure if you are a woman is positive in the underlying population

    $P_0(A=1|W=1)>0$
  - In a given sample from P_0, by chance there are no women who were exposed

    $P_n(A=1|W=1)=0$
  - No information in our sample about the outcomes of exposed women

Positivity Assumption

- The extent to which the identifiability of the target parameter is threatened by positivity violations will depend on
  - The observed data distribution P_0
  - The statistical model $\mathcal{M}$
  - The parameter $\Psi(P_0)$

- The impact of positivity violations and near violations on estimator performance also depends on the estimator…
  - More coming up…

Identifiability results are parameter-specific

- New identifiability results are needed for target parameters indexed by
  - Interventions on more than one node
    - Longitudinal treatment effects
    - Direct effects
  - Interventions that respond to patient characteristics
    - Dynamic regimes

A given parameter can have more than one identifiability result

- A point treatment effect may still be identified when there are no sets of observed covariates that satisfy the back door criterion
- Examples:
  - Instrumental Variables
  - Front Door criterion ("mediating instrumental variables")

Key Points 1

- The distribution of a counterfactual outcome under an intervention on a single node is identified by the G-computation formula
  - Under the Randomization Assumption
  - Graphically, under the backdoor criterion
- Often our knowledge alone is not sufficient for identifiability
  - Could stop there
  - If want to proceed, figure out what minimal assumptions are needed

Key Points 2

- Identifiability also relies on having sufficient support in the data
  - Positivity assumption: some positive probability of each treatment level of interest, given each possible covariate value
  - In practice, in finite samples may still have sparse or empty cells
    - How different estimators handle this problem: coming up