PH 252E: Lectures 2 and 3

Time-dependent confounding; Identification of the joint effects of multiple interventions

Outline

1. Specify observed data
   - With censoring, right censoring, missing data
2. Identification
   - Review – point treatment setting
3. Identification
   - Time-dependent confounding
   - Sequential Randomization Assumption/ sequential back door criteria
   - Longitudinal G computation formula

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

Quick Notation/Concept Review

- Observed data? Its distribution?
- Link to SCM?
- Statistical Model?

Linking the Observed Data to the SCM

- Defining the statistical estimation problem requires specifying the link between endogenous variables X and the observed data O
  - In other words, we specify how the observed data were generated by the data generating system encoded in our SCM
- Often, O=X
  - Ie endogenous variables in SCM correspond to observed variables O
  - Ex: point treatment: O=(W,A,Y)

Linking the Observed Data to the SCM

- We observe a sample of size n of the random variable O
  - We will work with independent samples
  - The framework is not restricted to this
- We assume our observed data were generated by sampling *n* times from the data generating system specified in our causal model
- This gives us n i.i.d. copies $O_1, O_2, \ldots, O_n$ drawn from true probability distribution P_0

The Statistical Model

- The structural causal model $\mathcal{M}^{\mathcal{F}}$ (which tells us the set of possible distributions for U,X) implies a model (set of possible distributions) for O
- We refer to this set of possible distributions as the statistical model $\mathcal{M}$
- The true distribution P_0 of O is an element of $\mathcal{M}$

The Statistical Model

- Often, a model that respects the limits of our knowledge puts no restrictions on the set of allowed distributions for O
- **In this case our statistical model is non-parametric**
- **We need to respect this fact when we frame the statistical estimation problem**

Specifying the Observed Data: Longitudinal data

- Basic longitudinal observed data structure

$$O = (L(1), A(1), L(2), A(2), ..., L(K), L(K+1))$$
$$= (\bar{A}(K), \bar{L}(K+1)) \sim P_0 \in \mathcal{M}$$

- With survival data and additional right censoring: Observe data drawn up till the minimum of end of follow up (K+1), censoring, or "failure" (ie event of interest occurs)
  - For notational convenience, much literature defines data after failure or censoring as equal to last observed value

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

Quick Notation/Concept Review

- Target Causal Parameter?
  - General notation
    - Function of what distribution?
- What do we mean (in causal context) by identification?
- Target Statistical parameter (estimand)?
  - General notation
    - Function of what distribution?
- Identification of ATE in point treatment context (assumptions and estimand)?

Review: Identifiability

- Are the assumptions in our model sufficient to express our target causal quantity ($\Psi^F(P_{U,X})$) as a parameter of the distribution P_0 of the observed data ($\Psi(P_0)$) ?
  - Distribution P_0 of O implied by distribution $P_{U,X}$: $P_0=P(P_{U,X})$
  - Need to show that $\Psi^F(P_{U,X})=\Psi(P(P_{U,X}))$ for all $P_{U,X}$ in $\mathcal{M}^{\mathcal{F}}$
- Focus here on one identifiability result:
  - "G-computation formula"
- Holds under
  - Randomization assumption
  - Backdoor criterion

Recap: Identifiability for point treatment

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

$$\Psi : \mathcal{M} \rightarrow \mathbb{R}$$

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"

Recall: Positivity Assumption

- Need E(Y|A=a,W=w) to be well-defined
- In non-parametric model, each treatment of interest must occur with some positive probability for each possible covariate history
- Let g_0(a|W) denote P_0(A=a|W)
- Positivity assumption:

$$g_0(a|W) > 0\text{- a.e.}$$

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

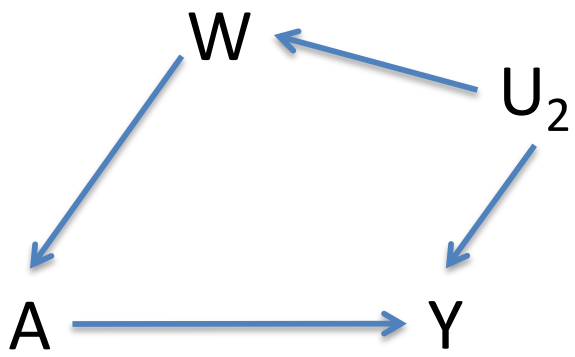
$$\underbrace{E_{U,X}(Y_a)}_{\Psi^F(P_{U,X})} = \underbrace{\sum_w E_0(Y|A=a,W=w)P_0(W=w)}_{\Psi(P_0)}$$

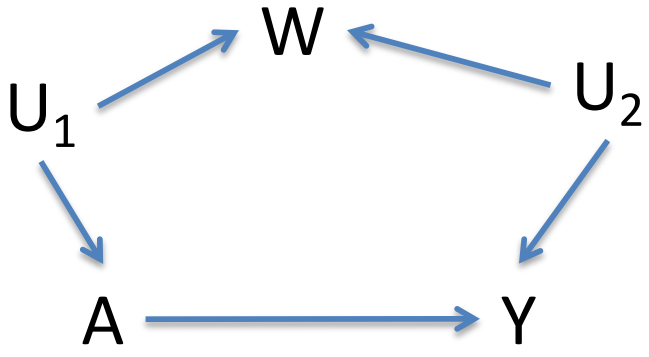
Back-door criterion

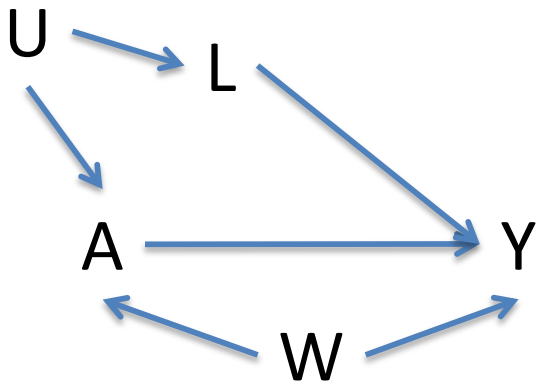
- A set of variables W satisfies the back door criterion with respect to (A,Y) if
  1. No node in W is a descendent of A
  – Motivation:
     1. Avoid blocking the path of interest
     2. Avoid introducing spurious sources of dependence
  2. W blocks all “backdoor” paths from A to Y
  – Backdoor path=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

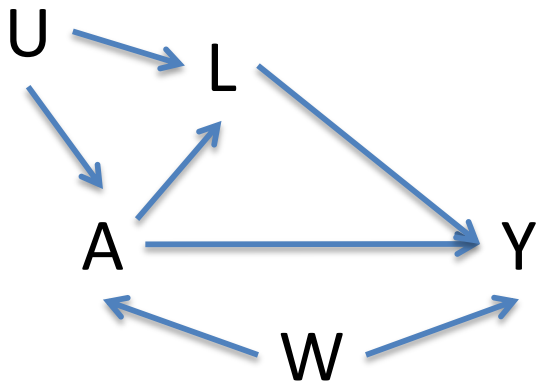
Quick Practice: Backdoor Criteria

Back door criteria holds for effect of A on Y conditional on what?



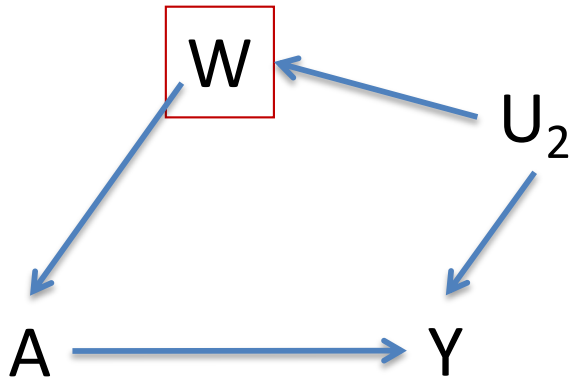


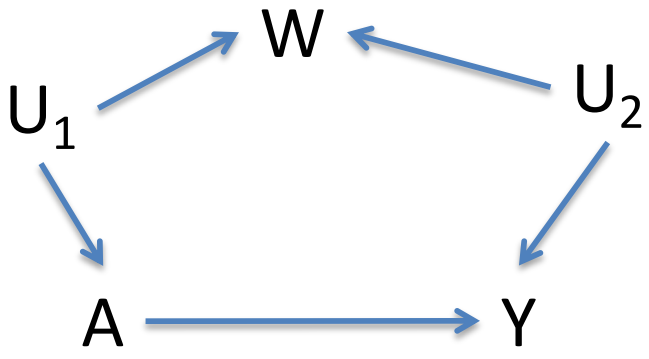




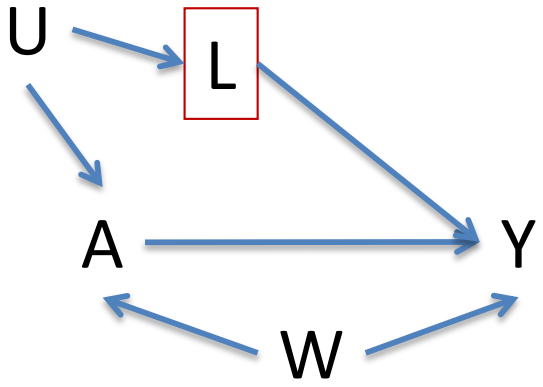
Quick Practice: Backdoor Criteria

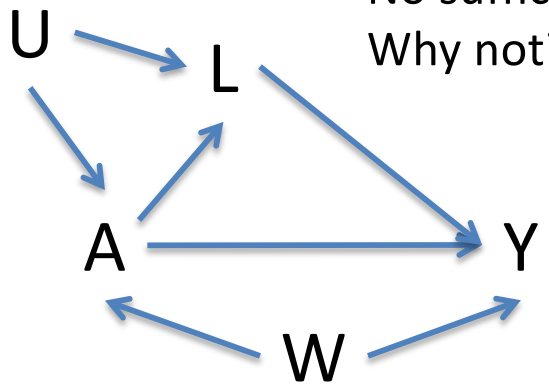
Back door criteria holds for effect of A on Y conditional on what?





Null Set
Why?





No sufficient set
Why not?

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

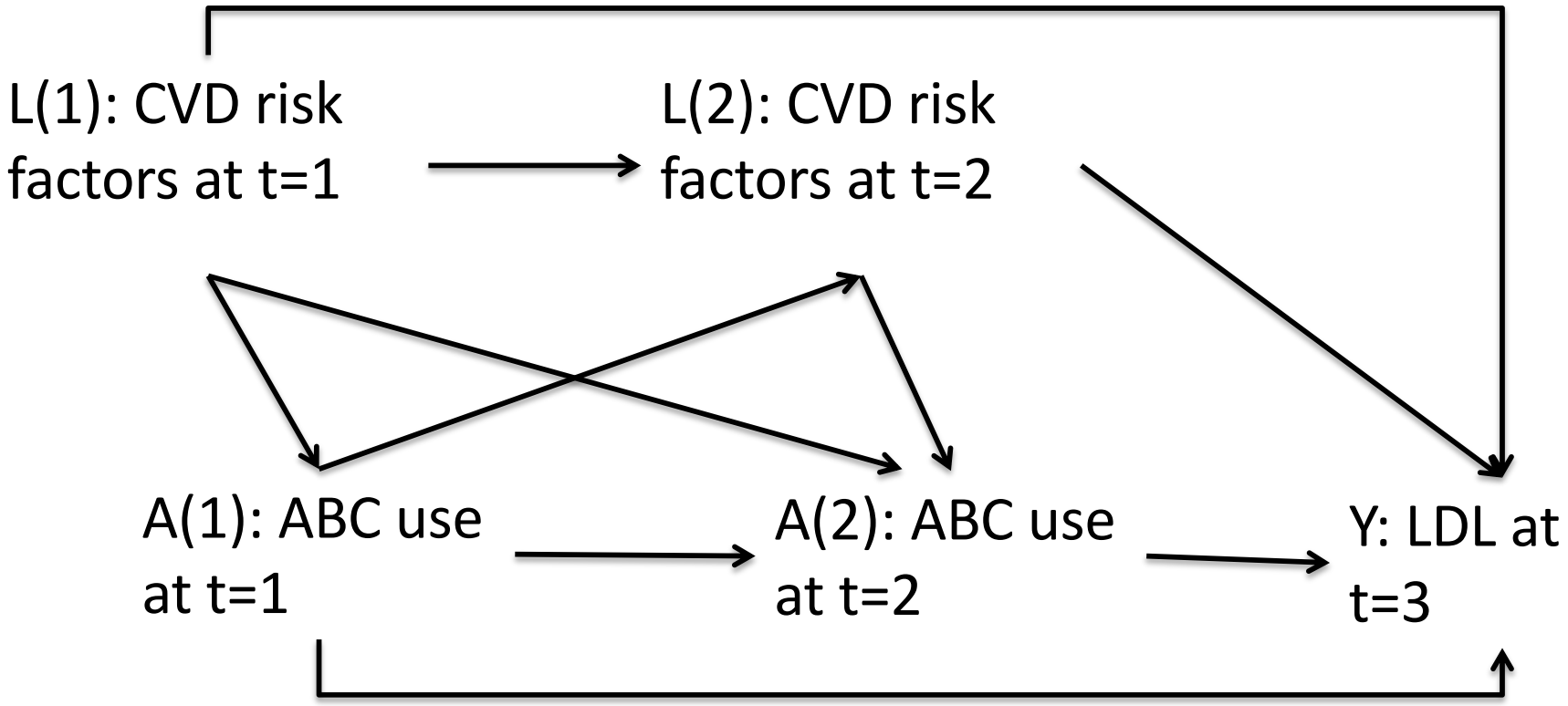
$$\underbrace{E_{U,X}(Y_a)}_{\Psi^F(P_{U,X})} = \underbrace{\sum_w E_0(Y|A=a,W=w)P_0(W=w)}_{\Psi(P_0)}$$

Identifiability for longitudinal exposures

- What causal assumptions are sufficient for our target longitudinal causal parameter to be identified as a parameter of the observed data distribution?

- Back to our simplified example for illustration
  - Effect of Abacavir use at t=1 and t=2 on LDL at t=3
  - Measure CVD risk factors at t=1 and t=2
  - Assume no deaths, censoring, or missing data

Abacavir Example: SCM/Graph



$$L(t) = f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t)}, t = 1, 2, 3$$

$$A(t) = f_{A(t)}(\bar{L}(t), \bar{A}(t-1), U_{A(t)}), t = 1, 2$$

$$Y \subset L(3)$$

Abacavir Example: Target Parameter and Observed Data

- Target causal parameter: $E_{U,X}(Y_{\bar{a}=1} - Y_{\bar{a}=0})$
- Observed data: n i.i.d. copies of
  $O = (L(1), A(1), L(2), A(2), Y) \sim P_0$
- Under what conditions can we write our target causal parameter as a parameter of the observed data distribution?
- **We need to move beyond the simple back door criterion….**

How are longitudinal parameters different?

- Our previous identifiability result relied on conditioning on some set of covariates W that were sufficient to block all back door paths from our intervention A to our outcome Y

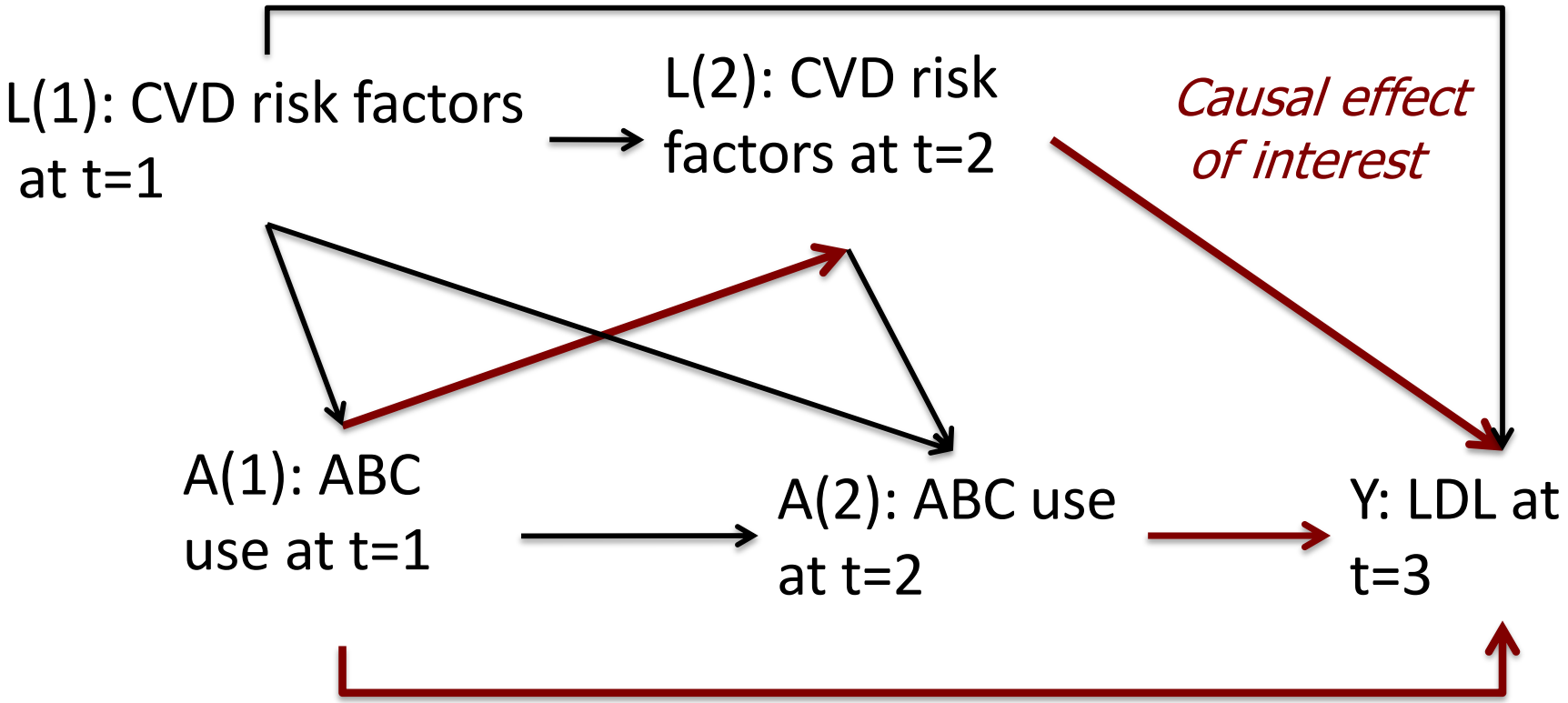
- We could <u>not</u> condition on descendents of A

How are longitudinal parameters different?

- When we are interested in intervening on multiple nodes, we are often in a situation where no one set of covariates that meet the back door criterion for all intervention nodes simultaneously exists

- However, the distribution of counterfactuals indexed by interventions on these multiple nodes may still be identified…

ABC Example: SCM/Graph



$$L(t) = f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t)}), t = 1, 2, 3$$

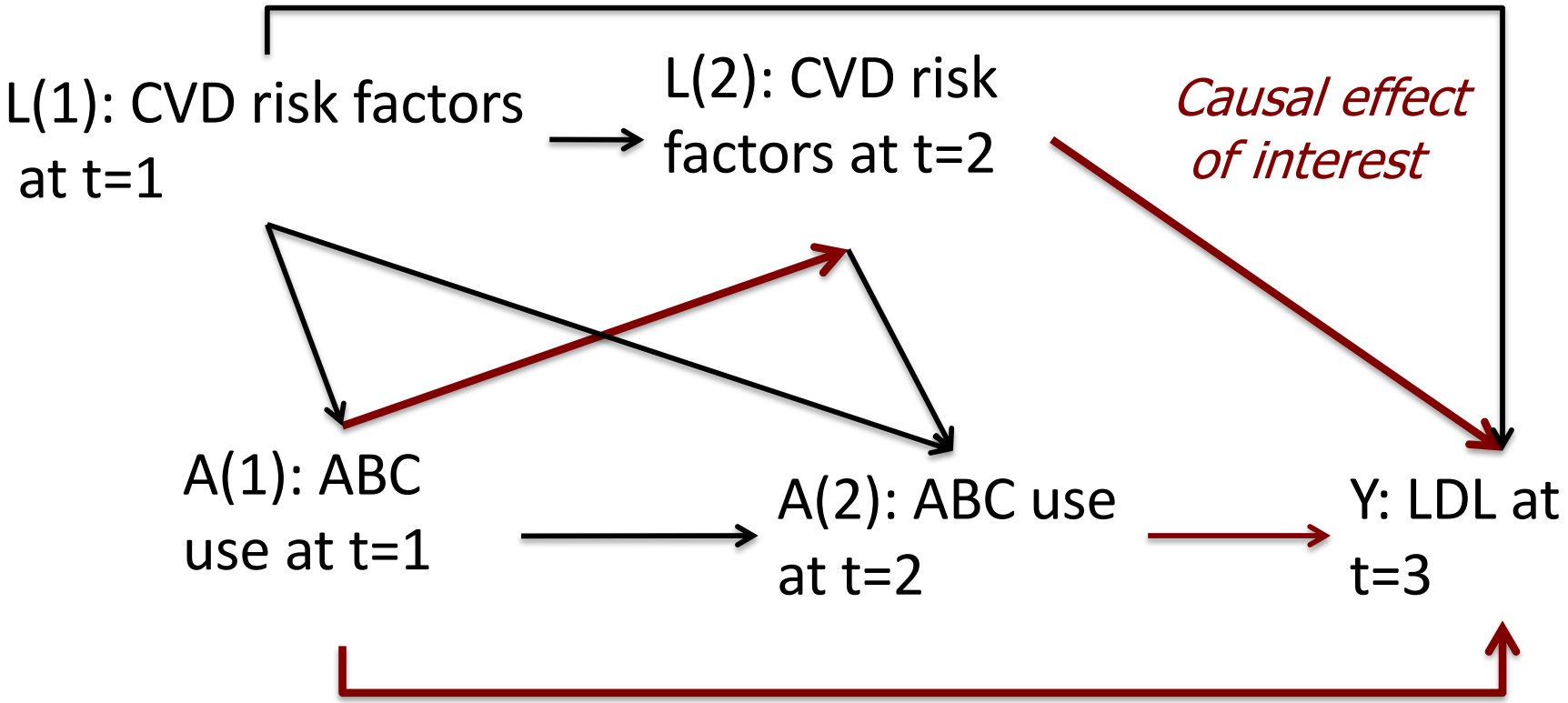
$$A(t) = f_{A(t)}(\bar{L}(t), \bar{A}(t-1), U_{A(t)}), t = 1, 2$$

$$Y \subset L(3)$$

Is $E(Y_{11} - Y_{00})$ identified using the standard (point treatment) back door criterion?

- We need to find a single set of variables that
  1. Are non-descendents of (A(1),A(2)) and
  2. Block all back door paths from (A(1),A(2)) to Y…

ABC Example: SCM/Graph



$$L(t) = f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t)}, t = 1, 2, 3$$

$$A(t) = f_{A(t)}(\bar{L}(t), \bar{A}(t-1), U_{A(t)}), t = 1, 2$$

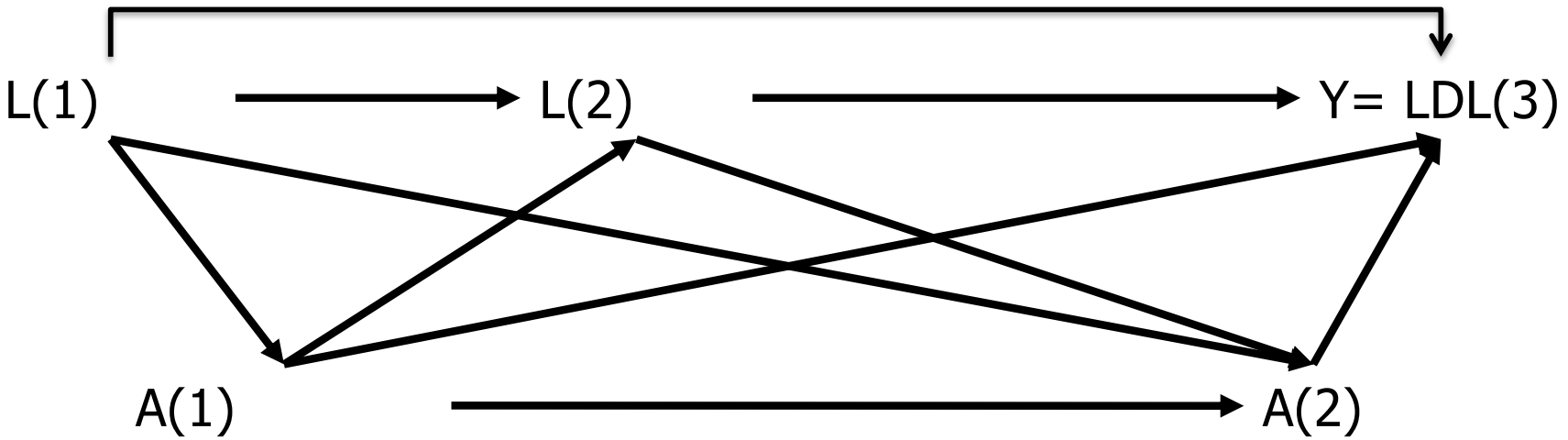
$$Y \subset L(3)$$

Why does the point treatment G-computation formula break down?

- Option #1: Does L(1) satisfy backdoor criterion for Effect A(1) and A(2) on Y?

$$E(Y_{11}) \stackrel{?}{=}$$

$$\sum_{l(1)} E(Y|A(1)=1, A(2)=1, L(1))P(L(1)=l(1))$$

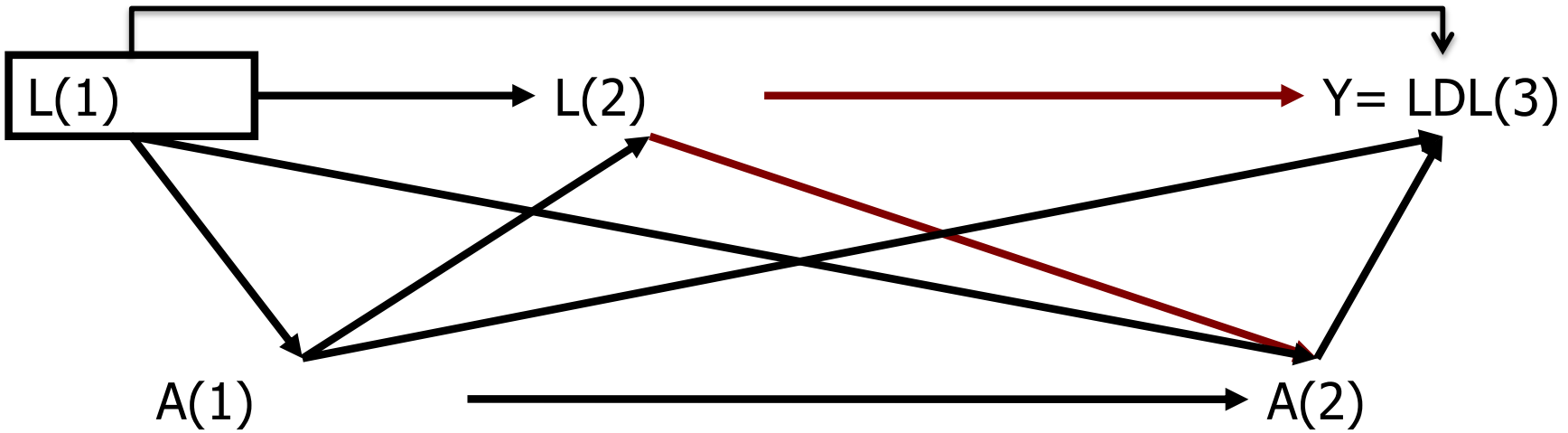


Why does the point treatment G-computation formula break down?

$$E(Y_{11}) \neq \sum_{l(1)} E(Y|A(1)=1, A(2)=1, L(1))P(L(1)=l(1))$$

- Why not?

Unblocked backdoor path

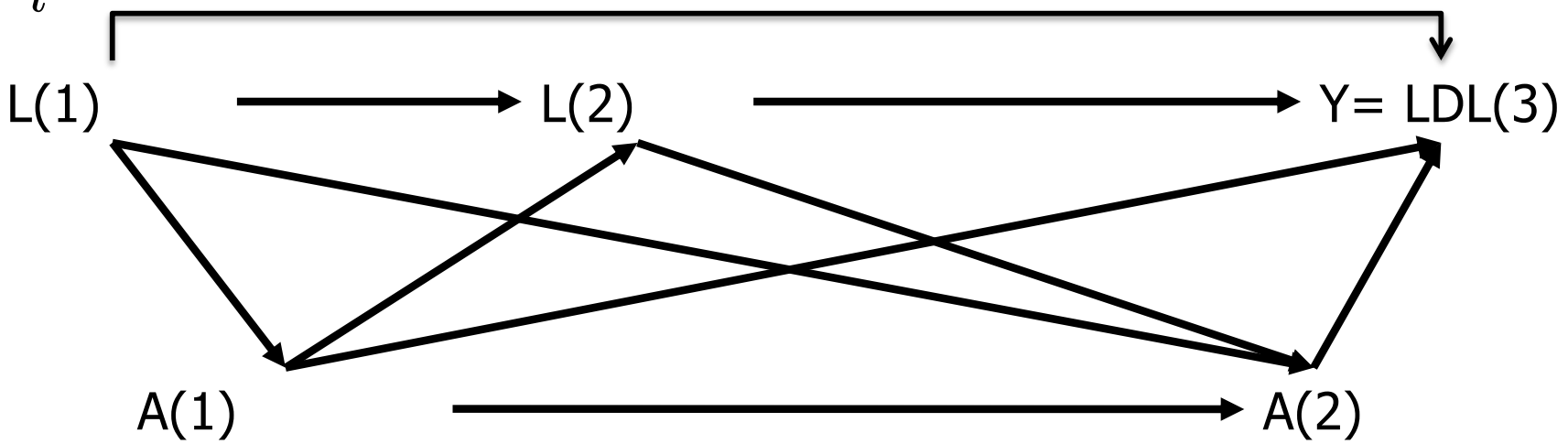


Why does the point treatment G-computation formula break down?

- Option #2: Does (L(1),L(2)) satisfy backdoor criterion for Effect A(1) and A(2) on Y?

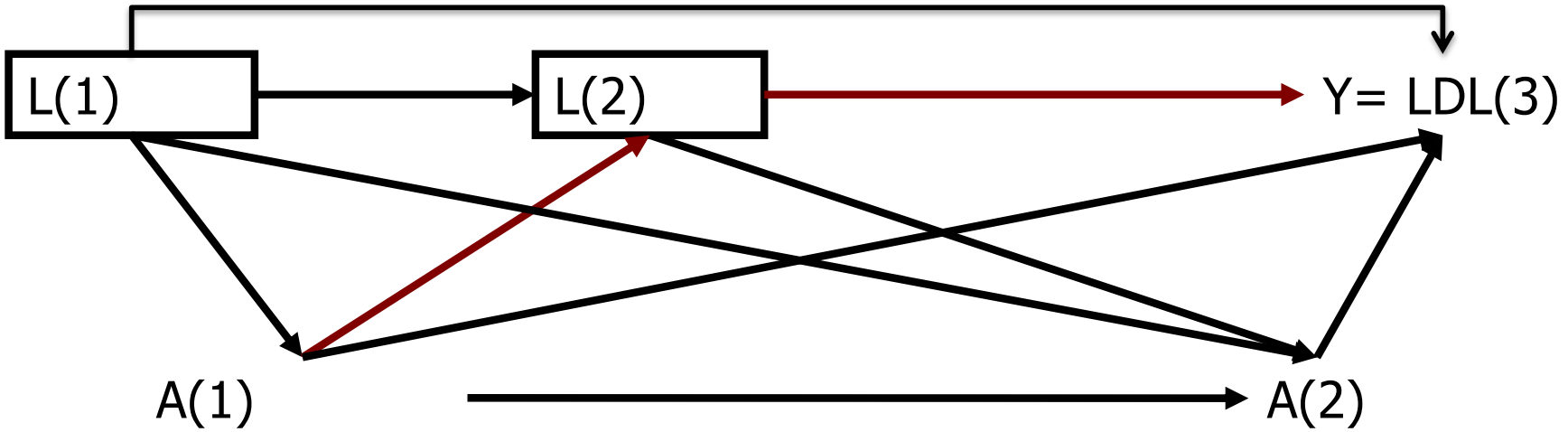
$$E(Y_{11}) \stackrel{?}{=}$$

$$\sum_{\bar{l}} E(Y|A(1)=1, A(2)=1, \bar{L}=\bar{l})P(\bar{L}=\bar{l})$$



Why does the point treatment G-computation formula break down?

$$E(Y_{11}) \stackrel{?}{=} \sum_{\bar{l}} E(Y|A(1)=1, A(2)=1, \bar{L}=\bar{l})P(\bar{L}=\bar{l})$$



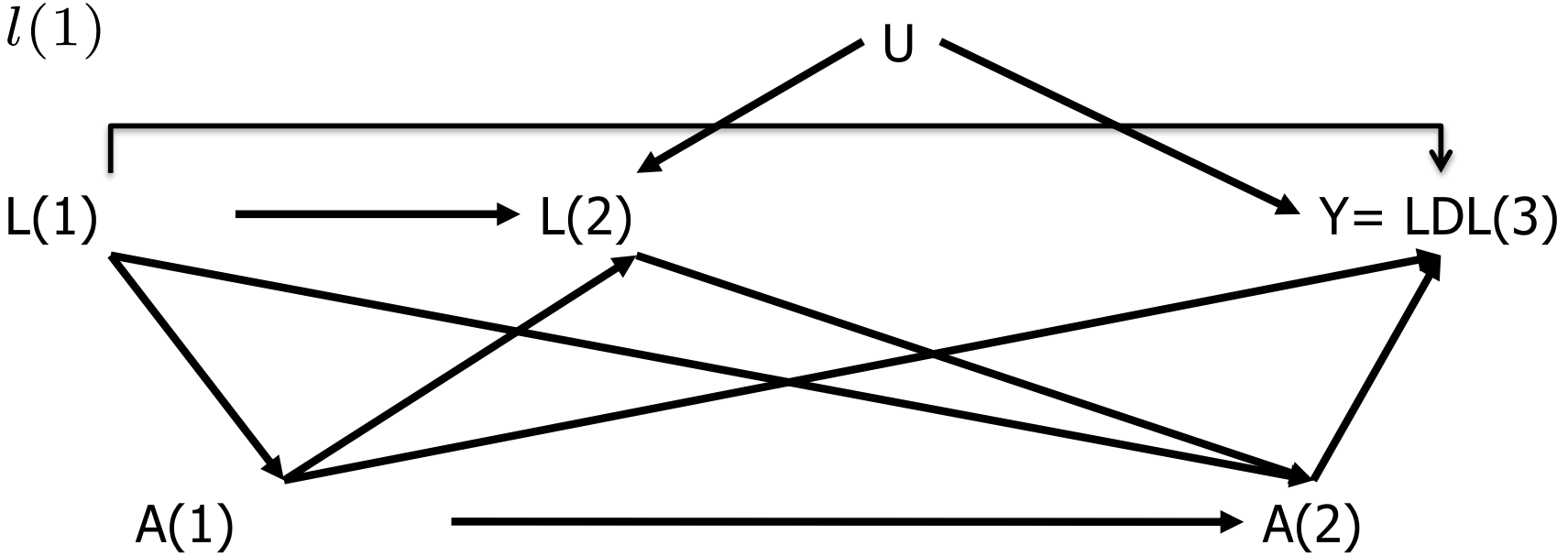
- Intuition: losing the effect of A(1) on Y via L(2)
  - Point treatment identifiability result takes expectation with respect to the wrong distribution of covariates L

Even if A does not affect Y via L(2), we may have a problem

- if L(2) is a confounder for A(2) and affected by A(1)…

$$E(Y_{11}) \stackrel{?}{=}$$

$$\sum_{l(1)} E(Y|A(1)=1, A(2)=1, L(1))P(L(1)=l(1))$$

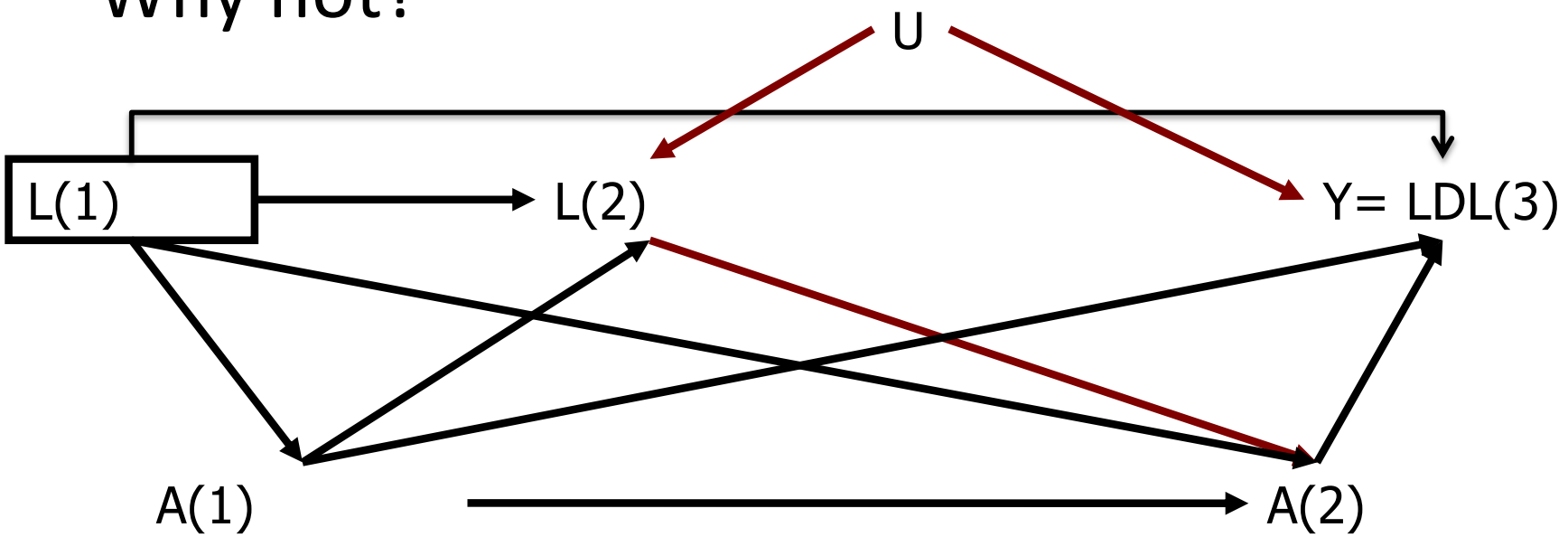


Why does the point treatment G-computation formula break down?

$$E(Y_{11}) \neq \sum_{l(1)} E(Y|A(1)=1, A(2)=1, L(1))P(L(1)=l(1))$$

Unblocked backdoor path

- Why not?

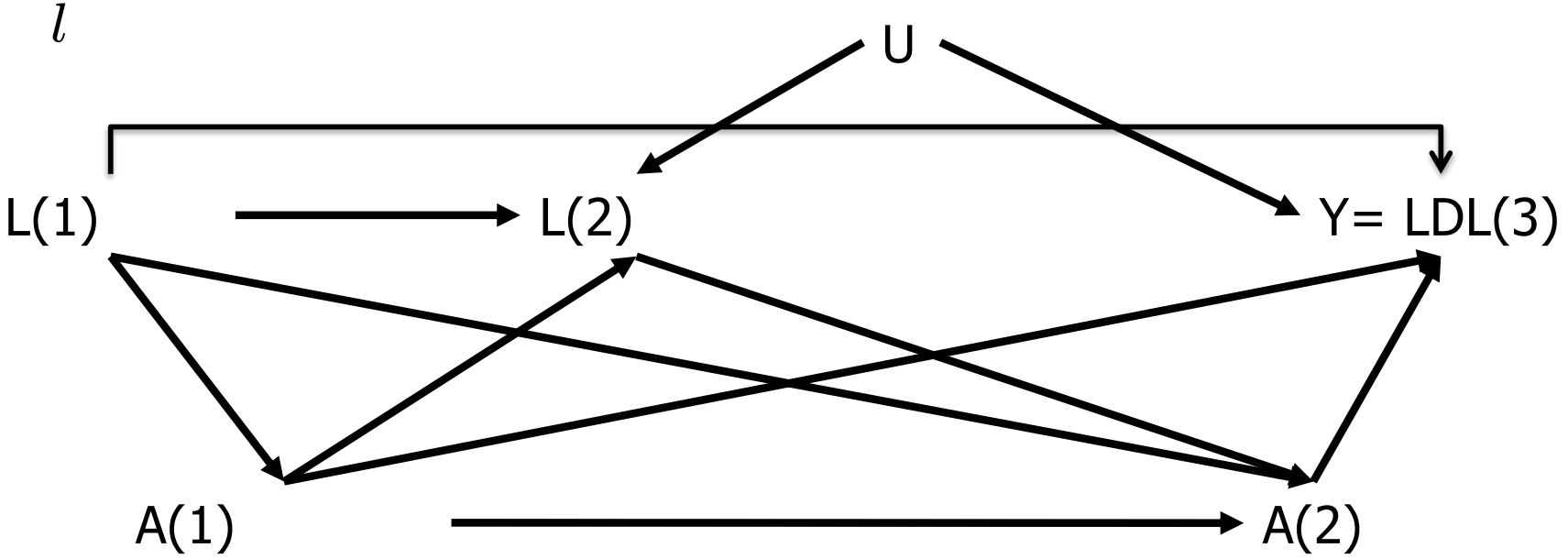


Why does the point treatment G-computation formula break down?

- Option #2:

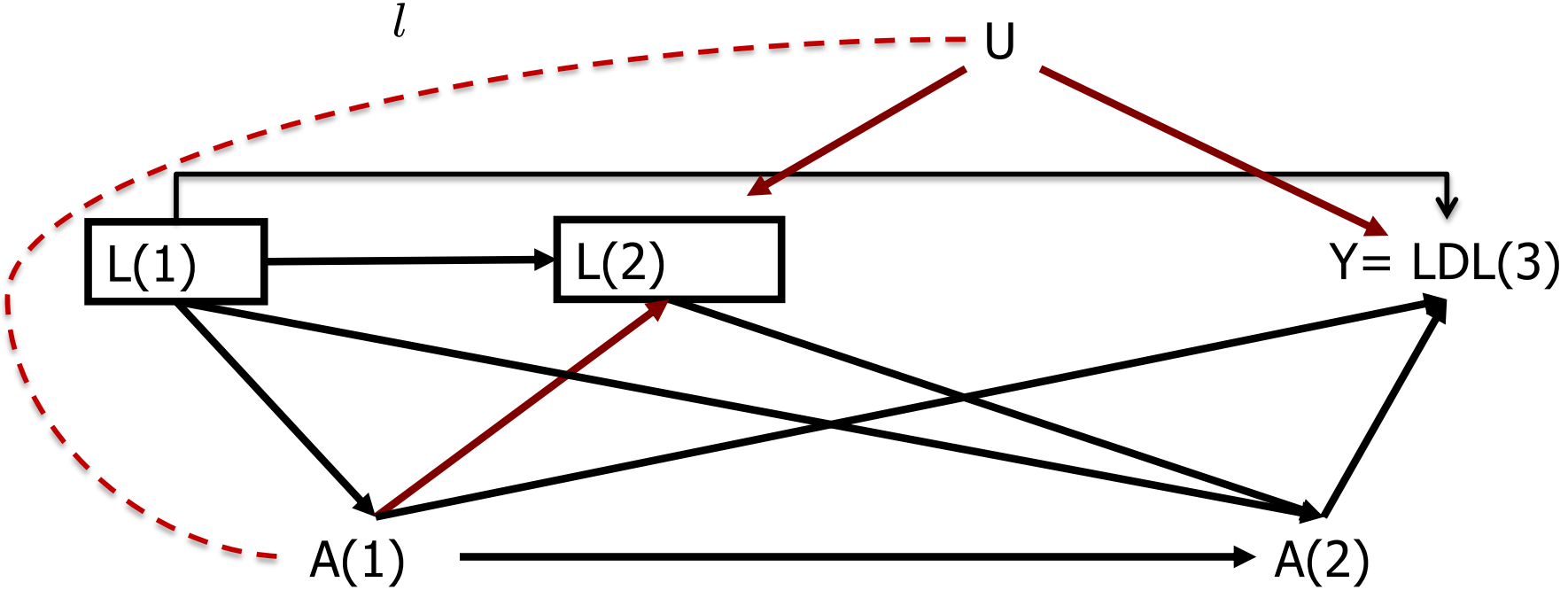
$$E(Y_{11}) \stackrel{?}{=}$$

$$\sum_{\bar{l}} E(Y|A(1)=1, A(2)=1, \bar{L}=\bar{l})P(\bar{L}=\bar{l})$$



Why does the point treatment G-computation formula break down?

$$E(Y_{11}) \stackrel{?}{=} \sum_{\bar{l}} E(Y|A(1)=1, A(2)=1, \bar{L}=\bar{l})P(\bar{L}=\bar{l})$$



- Conditioning on collider creating new confounding
  - Point treatment identifiability result takes expectation with respect to the wrong distribution of covariates

The Dilemma of Time –Dependent Confounding

- No subset of covariates for which the simple back door criterion holds
  - We need L(2) to block the back door path from A(2) to Y
  - But L(2) is a descendent of A(1)
- Is our target parameter really unidentified?

- Not necessarily! But we do need a new identifiability assumption -> new estimand

How can we modify the point treatment G-computation formula?

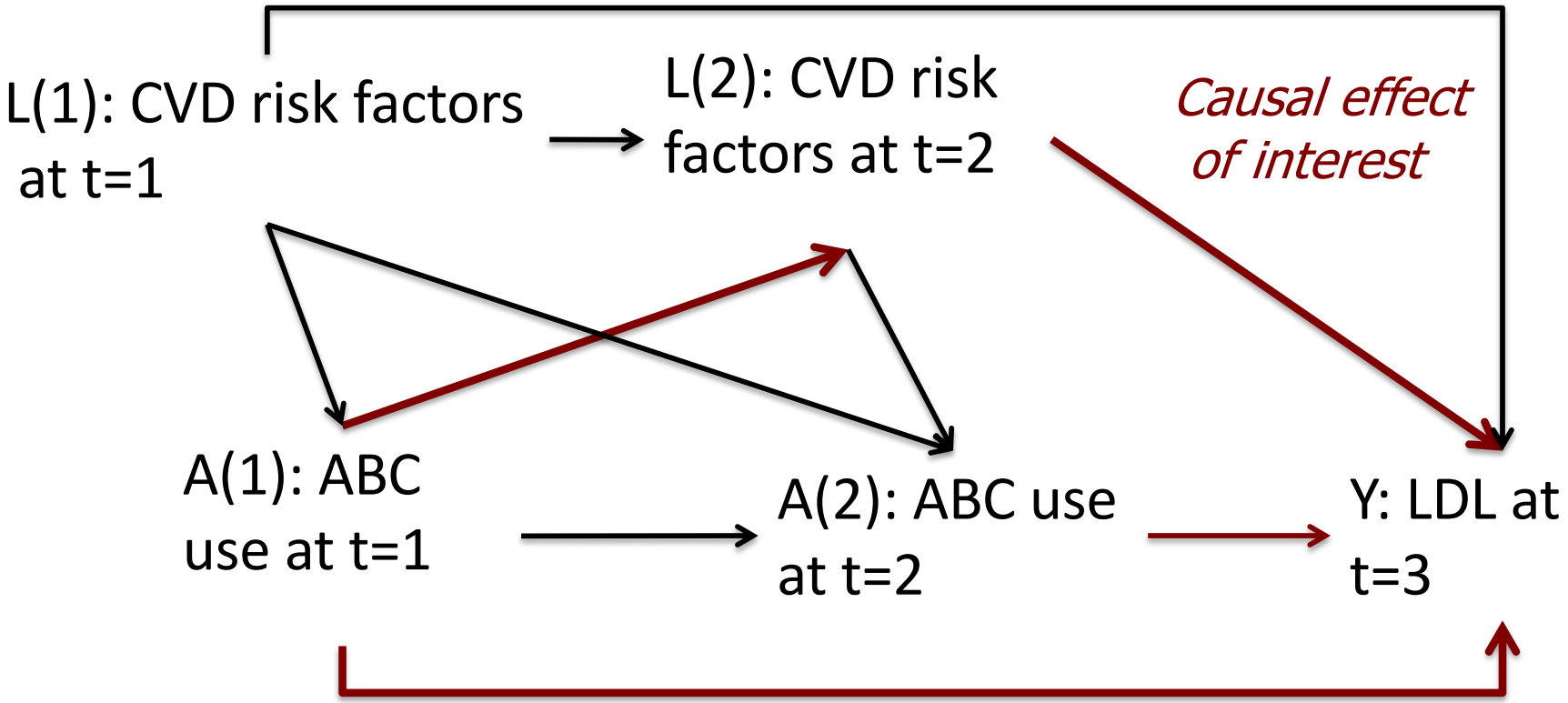
– Point treatment identifiability result takes expectation over the wrong distribution of L(2)

$$\sum_{\bar{l}} E(Y|A(1)=1, A(2)=1, \bar{L}=\bar{l}) P(\bar{L}=\bar{l})$$

Distribution of CVD risk factors under "**non- intervened on" SCM**

- Instead: that expectation with respect to the distribution of covariates (L(1),L(2)) under intervention to set A(1)=1,A(2)=1

ABC Example: SCM/Graph

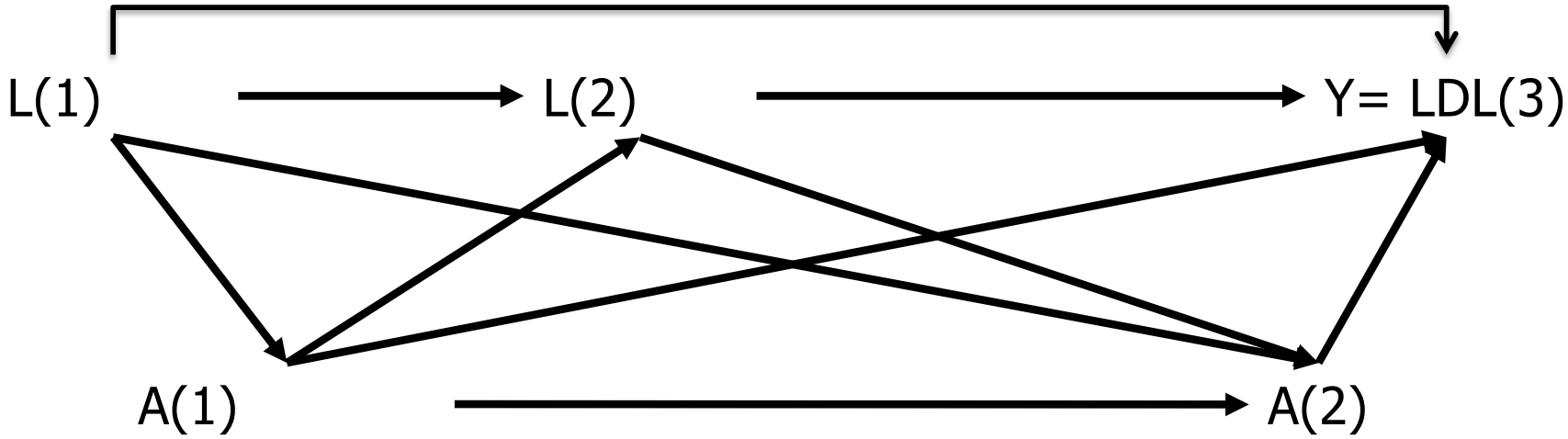


$$L(t) = f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t)}), t = 1, 2, 3$$

$$A(t) = f_{A(t)}(\bar{L}(t), \bar{A}(t-1), U_{A(t)}), t = 1, 2$$

$$Y \subset L(3)$$

Example: Longitudinal G-computation Formula



$$E(Y_{11}) = \sum_{l(1),l(2)} \left(\begin{array}{l} E(Y|A(1)=1, A(2)=1, L(1)=l(1), L(2)=l(2)) \times \\ P(L(2)=l(2)|A(1)=1, L(1)=l(1)) \times P(L(1)=l(1))) \end{array} \right)$$

Distribution of CVD risk factors under **post-intervention SCM**

Longitudinal G computation Formula

$$E(Y_{\bar{a}}) = \sum_{\bar{l}} E(Y|\bar{A}(K) = \bar{a}(K), \bar{L}(K) = \bar{l}(K)) \times$$

$$\prod_{t=1}^{K} P(L(t) = l(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1))$$

Distribution of covariates under **post-intervention SCM**

Key insight: we don't need to adjust for everything all at once

- Instead, we can think of simulating our data sequentially from our set of structural equations
- This lets us consider the problem of identifiability sequentially
  - For each A(t) in sequence, ask if its effect on Y can be identified by conditioning on some subset of the observed past.

Identifiability for the effects of multiple interventions

- What do we need for identifiability in this case?
- Intuition: Sequentially Randomized Trial
  - **If** at each time point we could randomize A(t) within strata of (some subset of) covariates and treatment observed up until then
  - **Then**, at each time point the effect of A(t) on future nodes would be identified
    - We know we measured enough of the past to estimate the effect of intervening on that node
  - We could then estimate the effect of setting each A(t) sequentially

Identifiability for multiple interventions

- Sequential Randomization Assumption

$$Y_{\bar{a}} \perp A(t)|\bar{L}(t), \bar{A}(t-1) = \bar{a}(t-1), t = 1, ..., K$$

  - If A(t) is randomly assigned at each time point, given the observed past, this will hold
  - This is sometimes called a sequentially randomized trial or sequential multiple assignment randomized trial (SMART)
- Counterpart to the Randomization Assumption for a single intervention

$$Y_a \perp A|W$$

A graphical criterion for identifiability

- For the single time point case, the back door criterion:

1. Allowed us to evaluate if our target parameter was identified by the G computation formula
2. Helped us to decide what to condition on
   - What to include in "W"
   - Recall- not always a good idea to include all pre-treatment variables…

- Is there an equivalent criterion for target parameters with multiple interventions?

Identifying the effects of interventions on multiple nodes

- The sequential back door criterion (counterpart to the back door criterion for a single intervention node):

  For each intervention node A(t), every "intervention avoiding" backdoor path from A(t) to Y must be blocked by some subset of non-descendents of A(t)

  – Intervention avoiding= path with no arrows into a future "A" node (i.e. a intervention node after A(t))

Sequential Back Door Criterion

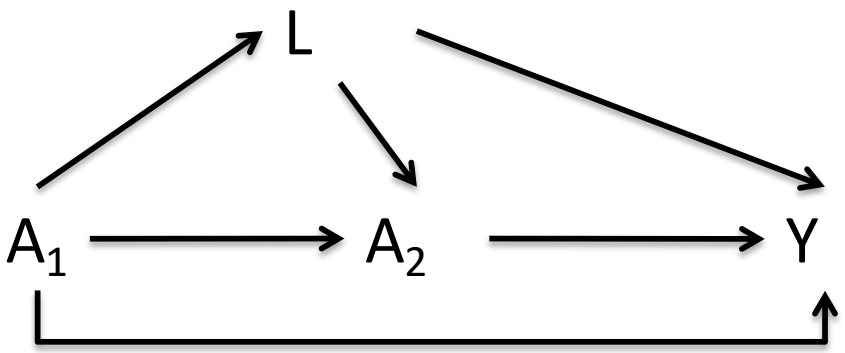
- Essentially we just want to apply the usual back door criterion, for each intervention node A(t) in series:
  1. We are looking for set of covariates (+ past treatment) that will block all back door paths from A(t) to the outcome
  2. These covariates cannot be descendents of A(t)
- Same justification: Want to remove any sources of association between each A(t) and the outcome other than those that we are interested in

Sequential Back door Criterion

- Just the standard back door criterion applied to each intervention node is sequence except
- Now it is OK if there is an unblocked back door path that goes through a future intervention node
- Why?
  - Any paths through future A nodes will already be blocked because we are intervening to set those nodes
  - We don't need to worry about blocking them

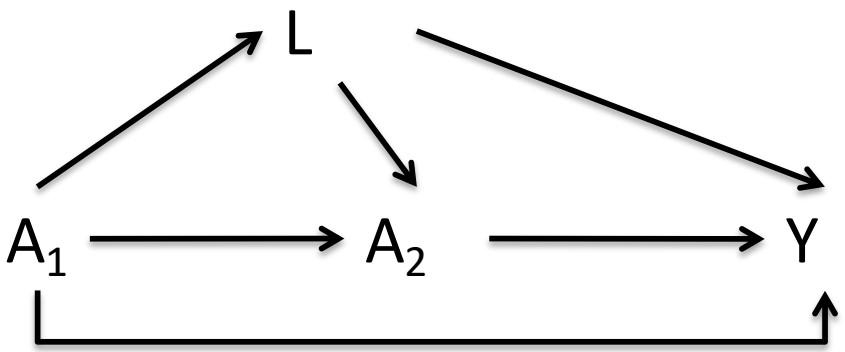
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what?
  - For A_2 given what?



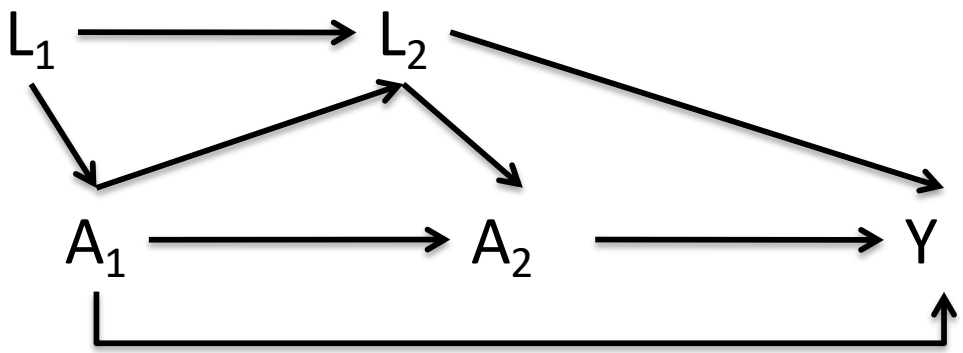
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what? **nothing**
  - For A_2 given what? **A1,L**



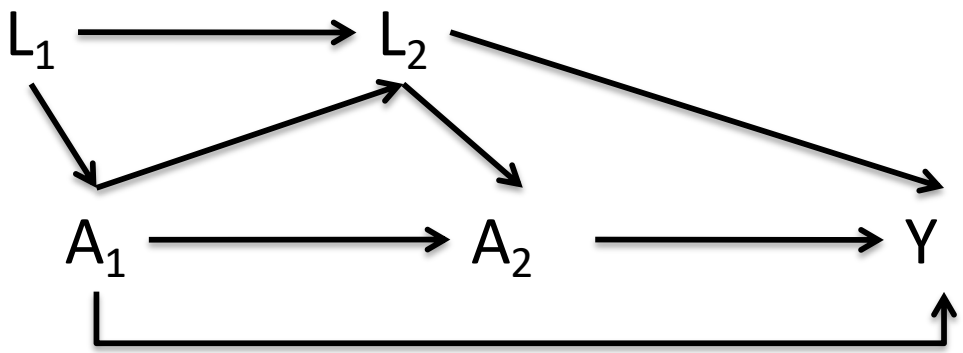
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what?
  - For A_2 given what?



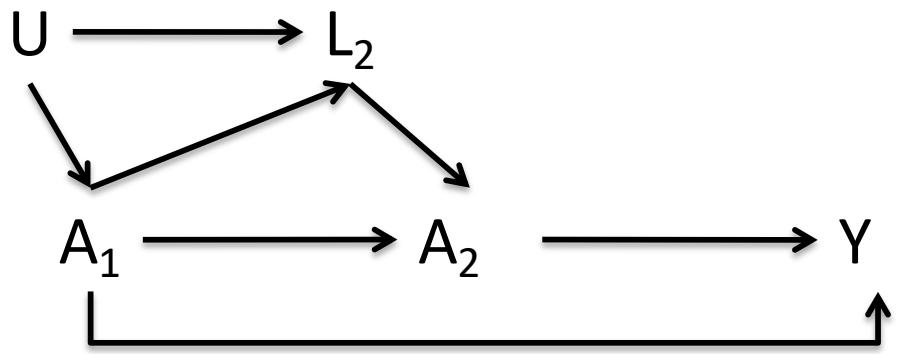
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what? **L1**
  - For A_2 given what? **A1, L2**



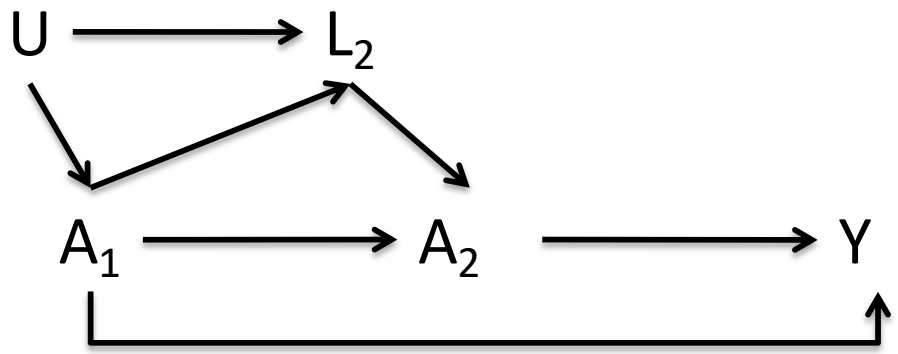
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what?
  - For A_2 given what?



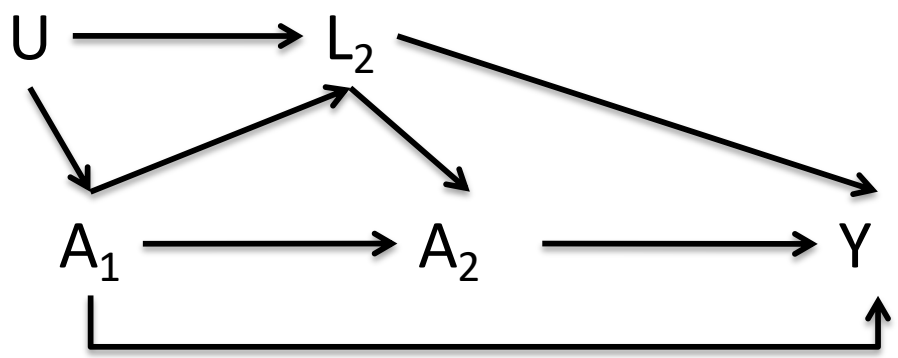
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what? **Nothing**
  - For A_2 given what? **A1**
  - *Unconfounded*



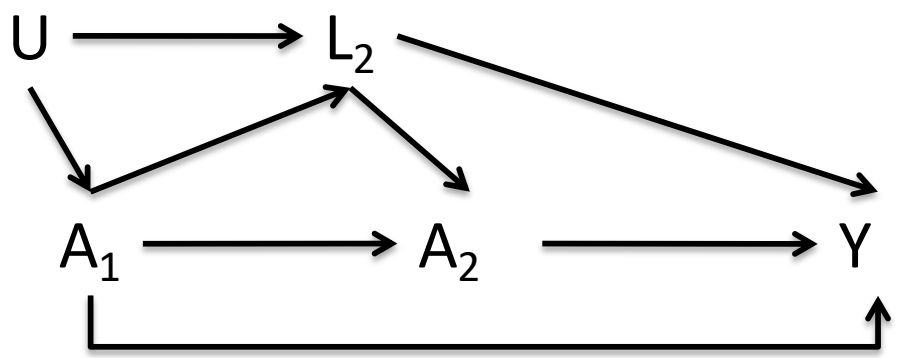
Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what?
  - For A_2 given what?



Example

- Target: $E(Y_{a1a2})$
- Sequential back door holds?
  - For A_1 given what? **No sufficient set**
  - For A_2 given what? **A1, L2**



Identifiability Result

- Under the Sequential Randomization Assumption (or if full measured history sufficient to satisfy the sequential back door criterion):

$\Psi(P_{U,X})$: Causal Parameter of Interest

$$P(Y_{\bar{a}} = y) =$$

$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y|\bar{A} = \bar{a}, \bar{L} = \bar{l}) \times \\ \prod_{t=1}^{K} P\left(L(t) = l(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)\right) \end{array} \right)$$

$\Psi(P_0)$: Target statistical parameter/estimand

- Generalization of point treatment G-computation formula
  Proof Pearl Causality p. 123 (or Robins 1986 using counterfactual framework)

Positivity Assumption

- Analogous to point treatment case, need some positive probability of following regime of interest at each time point regardless of covariate history

$$g_0(A(t) = a(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t)) > 0, t = 1, ..., K, \text{- a.e.}$$

$\Psi(P_{U,X})$: Causal Parameter of Interest

$$P(Y_{\bar{a}} = y) =$$

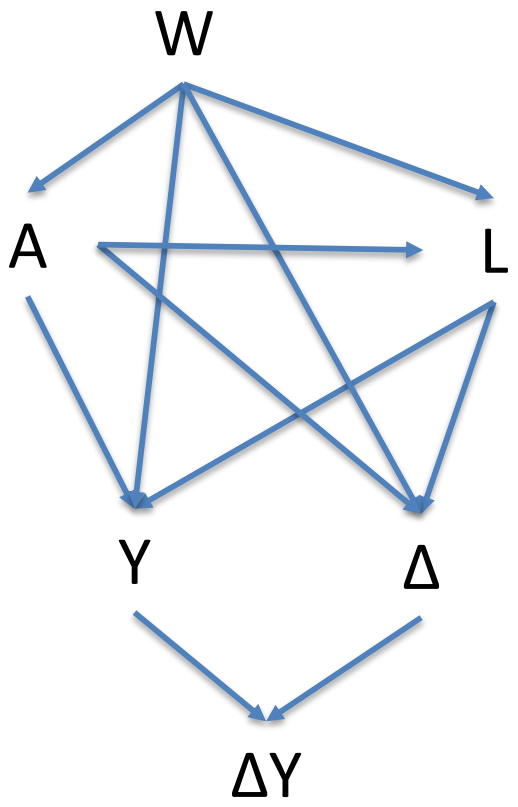
$$\sum_{\bar{l}} \left(\begin{array}{l} P(Y = y|\bar{A} = \bar{a}, \bar{L} = \bar{l}) \times \\ \prod_{t=1}^{K} P\left(L(t) = l(t)|\bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t-1) = \bar{l}(t-1)\right) \end{array} \right)$$

$\Psi(P_0)$: Target statistical parameter/estimand

Identification- Back to missing data

Missing outcome- One Approach

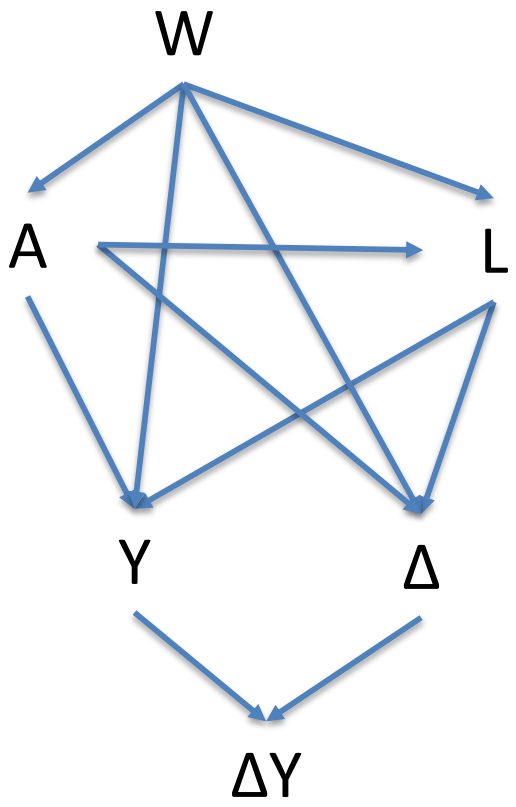
- Endogenous variables:
  X={W, A,L, Δ, Y, ΔY}
  - W=baseline CHD risk factors
  - A=ABC use
  - L= post-baseline risk factors
  - Δ= Indicator LDL is measured
  - Y= true LDL cholesterol
    - (not always observed)
  - ΔY= Observed LDL



- **Observed Data?**

Missing outcome- One Approach

- Endogenous variables:
  X={W, A,L, Δ, Y, ΔY}
  - W=baseline CHD risk factors
  - A=ABC use
  - L= post-baseline risk factors
  - Δ= Indicator LDL is measured
  - Y= true LDL cholesterol
    - (not always observed)
  - ΔY= Observed LDL



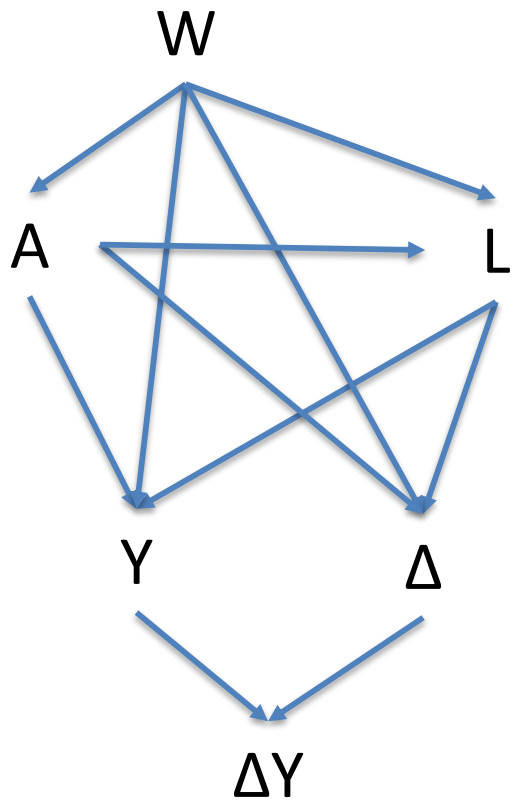
- Observed Data: n iid copies of $O=(W,A,L,\Delta,\Delta Y)\sim P_0$
- Target Causal Parameter: $E(Y_a)$
- **Identified? Under what assumptions? Estimand?**

Identification

- Target Causal Parameter: $E(Y_a)$
- Identification
  - Is W sufficient to satisfy backdoor criteria for effect of A on Y?

$$E(Y_a) = E_W E(Y|A = a, W)?$$

  - What is wrong with this approach? Are we done?

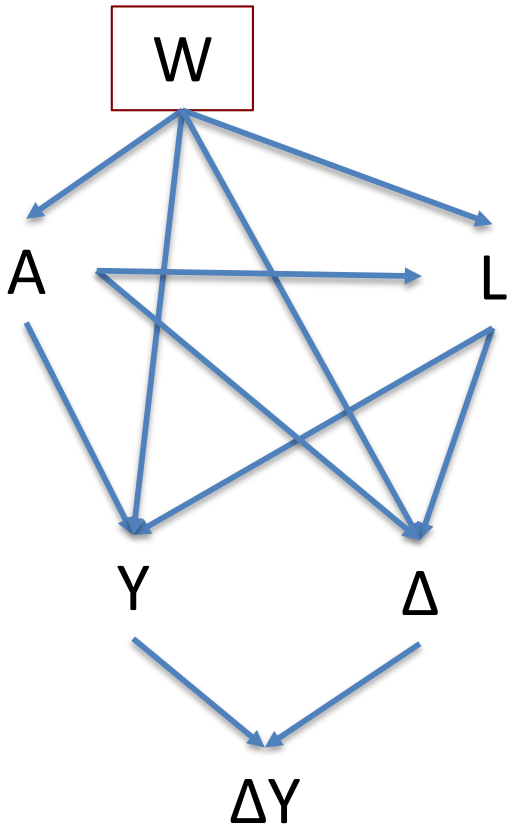


Identification

- Target Causal Parameter: $E(Y_a)$
- Identification
  - Is W sufficient to satisfy backdoor criteria for effect of A on Y? **Yes**

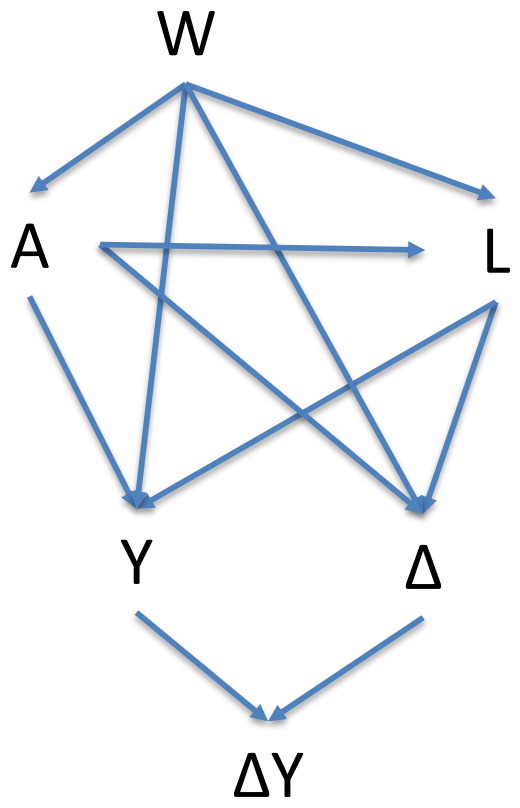
$$E(Y_a) = E_W E(Y|A=a,W)?$$

  - What is wrong with this approach?
    - **RHS a function of P_0? NO**



Identification

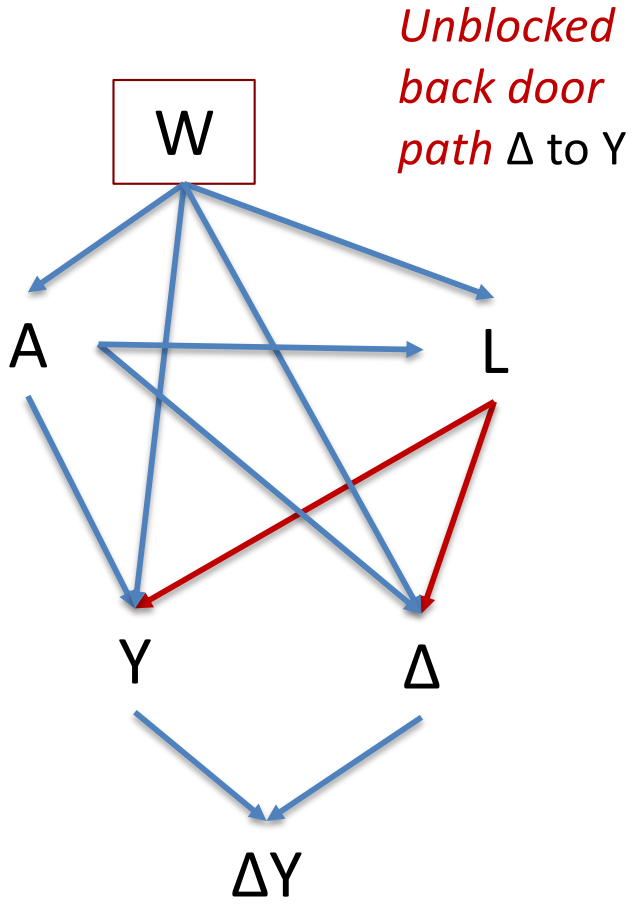
- Target Causal Parameter: $E(Y_a)$
- Identification
  - Intervene on A and Δ
  - Point treatment G-comp adjusting for W?
    - Is W sufficient to satisfy backdoor criteria for both interventions Y?
    - Why or why not?



$$E(Y_a) = E(Y_{a,\delta=1}) = E_W E(Y|A=a,\Delta=1,W)?$$

Identification

- Target Causal Parameter: $E(Y_a)$
- Identification
  - Intervene on A and Δ
  - Point treatment G-comp adjusting for W?
    - Is W sufficient to satisfy backdoor criteria for both interventions Y? **NO**
    - Why or why not?



$$E(Y_a) = E(Y_{a,\delta=1}) = E_W E(Y|A=a,\Delta=1,W)?$$

Identification

- Target Causal Parameter:

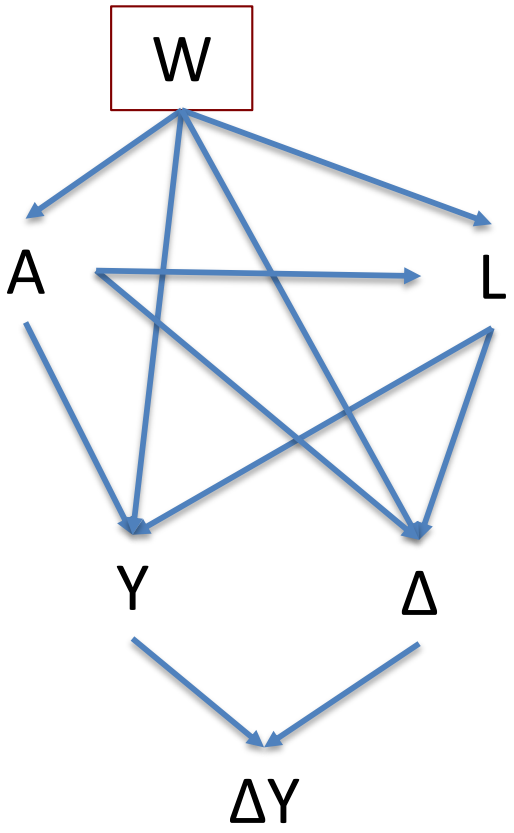
$$E(Y_a) = E(Y_{a,\delta=1})$$

- Sequential Randomization Assumption:

$$\boxed{Y_{a,\delta=1} \perp A|W}$$

$$Y_{a,\delta=1} \perp \Delta|A = a, W, L$$

- Estimand?



Identification

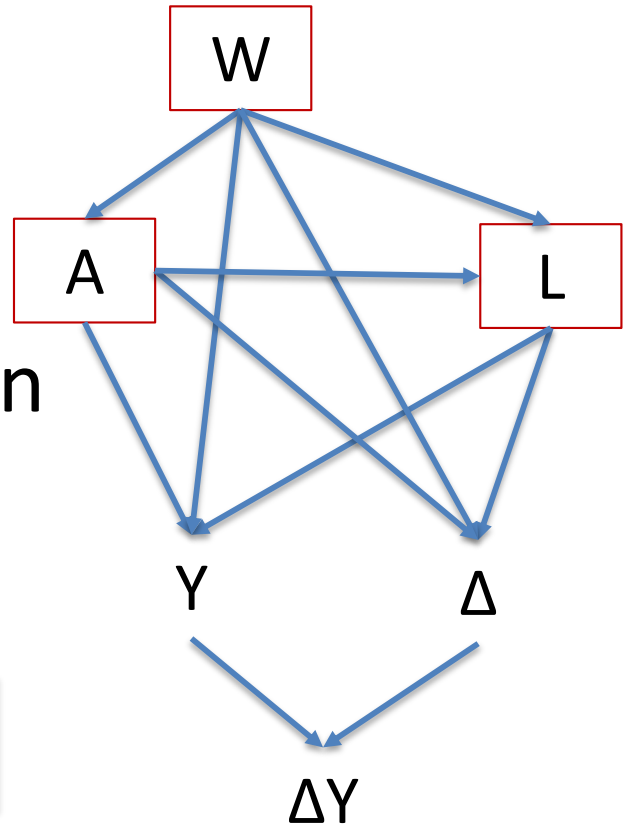
- Target Causal Parameter:

$$E(Y_a) = E(Y_{a,\delta=1})$$

- Sequential Randomization Assumption:

$$Y_{a,\delta=1} \perp A|W$$

$$\boxed{Y_{a,\delta=1} \perp \Delta|A=a,W,L}$$



- Estimand:

$$E(Y_{a,\delta=1})$$

$$= \sum_{w,l} E(Y|A=a,\Delta=1,W=w,L=l)P(L=l|A=a,W=w)P(W=w)$$

Identification

- Target Causal Parameter:

$$E(Y_a) = E(Y_{a,\delta=1})$$

- Estimand:

$$E(Y_{a,\delta=1})$$
$$= \sum_{w,l} E(Y|A=a,\Delta=1,W=w,L=l)P(L=l|A=a,W=w)P(W=w)$$

  - RHS a function of P0?

- Positivity assumption?

Identification

- Estimand:

$$E(Y_{a,\delta=1})$$

$$= \sum_{w,l} E(Y|A=a,\Delta=1,W=w,L=l)P(L=l|A=a,W=w)P(W=w)$$

  - RHS a function of P_0? **yes**
- Positivity assumption:

$$P(A=a|W) > 0$$

$$P(\Delta=1|A=a,W,L) > 0$$

  - Example where it might not hold?

Summary (1)

- Key concept: “time dependent confounding”
  - Simple back door criteria/Randomization assumption and corresponding point treatment G computation formula may fail for target parameters indexed by interventions on multiple nodes
  - In particular if there is a variable affected by one intervention node but needed to control for confounding (block backdoor paths to outcome) for another
  - Parameters that may seem to be “point treatment” may still have this problem
    - Ex: missing data

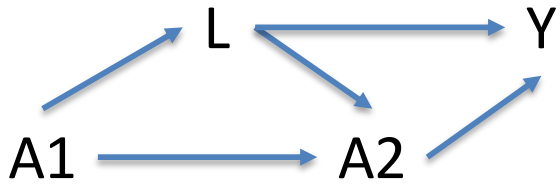
Summary (2)

- In setting of time-dependent confounding need

1. New identification assumptions:
   - Sequential randomization assumption/backdoor criteria
2. New estimands: Longitudinal G-comp formula
   - Point treatment G comp formula is a special case
3. New estimators: Coming up next!

Longitudinal-G comp formula (Simple Example)

- SCM & Observed data:
  - O=X=(A1,L,A2,Y)
  - All Us independent

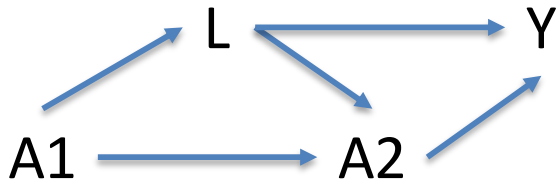


- Target Causal Parameter: $E(Y_{11})$
- Why can't we just Identify $E(Y_{11})$ as follows?

$$E(Y_{11}) = E(E(Y_{11}|L))$$
$$= EE(Y|A1 = 1, L, A2 = 1))$$

Longitudinal-G comp formula (Simple Example)

- SCM & Observed data:
  - O=X=(A1,L,A2,Y)
  - All Us independent



- Target Causal Parameter: $E(Y_{11})$
- Why can't we write ? $E(Y_{11}) = E(E(Y_{11}|L))$

$$\neq E E(Y|A1 = 1, L, A2 = 1))$$

- Problem: $E(Y|A1 = 1, L, A2 = 1) \neq E(Y_{11}|L)$

$$E(Y|A1 = 1, L, A2 = 1) = E(Y_{11}|A1 = 1, L_1, A2 = 1)$$

Definition of c.f.

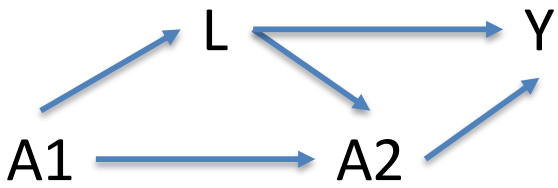
$$= E(Y_{11}|L_1)$$

$Y_{11} \perp (A1,A2) \mid L_1$

Counterfactual value of L setting A1=1

Proof: Longitudinal-G comp formula (Simple Example)

- SCM & Observed data:
  - O=X=(A1,L,A2,Y)
  - All Us independent



- Target Causal Parameter: $E(Y_{11})$

$$E(Y_{11}) = E[E(Y_{11}|A1)]$$ *iterated expectation*

$$= E[E(Y_{11}|A1 = 1)]$$ *$Y_{11} \perp A1$*

$$= E\{E[E(Y_{11}|A1 = 1, L, A2)|A1 = 1]\}$$ *iterated expectation*

$$= E\{E[E(Y_{11}|A1 = 1, L, A2 = 1)|A1 = 1]\}$$ *$Y_{11} \perp A2 \mid A1=1,L$*

$$= E\{E[E(Y|A1 = 1, L, A2 = 1)|A1 = 1]\}$$ *Definition of c.f.*

$$= \sum_l E(Y|A1 = 1, L, A2 = 1)P(L = l|A1 = 1)$$ *Definition*