INTRODUCTION TO TIME-VARYING TREATMENT STRATEGIES

Joy Shi
CAUSALab, Department of Epidemiology
Harvard T.H. Chan School of Public Health

February 18, 2022

LEARNING OBJECTIVES.

By the end of the session, you will be able to:

1. Formulate causal questions for time-varying treatments
2. Describe treatment-confounder feedback and recognize its structure on a DAG
3. Understand why conventional methods fail in the presence of treatment-confounder feedback

PLAN FOR TODAY.

1. Recap
2. Defining causal effects for time-varying treatments
3. Time-varying treatments and confounders on a DAG
4. Sequential exchangeability
5. Treatment-confounder feedback

RECAP: DEFINING THE CAUSAL EFFECT.

We define the causal effect as a contrast between two mean counterfactual outcomes:

- $\mathrm{E}[Y^{a=1}]$ is the mean outcome that would have been observed had everyone's treatment A been set to the value 1
- $\mathrm{E}[Y^{a=0}]$ is the mean outcome that would have been observed had everyone's treatment A been set to the value 0

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

or

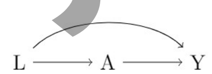
$$\frac{\mathrm{E}[Y^{a=1}]}{\mathrm{E}[Y^{a=0}]}$$

RECAP: ASSUMPTIONS FOR CAUSAL INFERENCE.

We need assumptions to estimate causal effects:

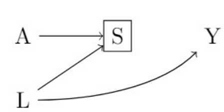
1. Conditional exchangeability: the mean outcome in the treated would have been the same as the mean outcome in the untreated, had they been treated, and vice versa, within levels of L
$$Y^a \perp\!\!\!\perp A | L \text{ for all } a$$
2. Positivity: the probability of being assigned to each treatment level is greater than 0 within levels of L
$$\Pr[A = a | L = l] > 0 \text{ for all } a, l \text{ if } \Pr[\mathrm{L} = \mathrm{l}] \neq 0$$
3. Consistency: an individual's counterfactual outcome under their observed treatment level is equal to their observed outcome
$$Y^a = Y \text{ when observed treatment } A \text{ is equal to } a$$

RECAP: THREE STRUCTURAL SOURCES OF BIAS.



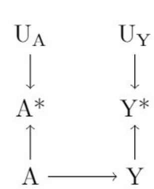
CONFOUNDING

- Open backdoor path from treatment A to outcome Y
- Common cause of A and Y



SELECTION BIAS

- Conditioning on a collider, resulting in an open path from treatment A to outcome Y

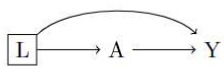


INFORMATION BIAS

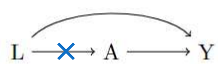
- Mismeasurement of A, Y or L
- Could be independent or dependent, differential or non-differential

RECAP: METHODS TO ADJUST FOR CONFOUNDING.

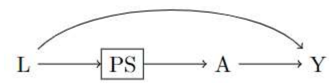
- Stratification
- Outcome regression with confounders



- Inverse probability weighting



- Propensity scores



- Standardization

These methods vary with respect to their modelling assumptions, but all of these methods address confounding by blocking or removing backdoor paths from treatment A to outcome Y.

TIME FIXED VERSUS TIME-VARYING TREATMENTS.

Up until this point, we've discussed these methods in the context of a time-fixed treatment, e.g.

- Surgery vs. no surgery
- Vaccine vs. no vaccine

However, most treatments of interest are time-varying, e.g.

- Medication use (e.g., aspirin, anti-retroviral therapy)
- Smoking
- Diet

PLAN FOR TODAY.

1. Recap
2. Defining causal effects for time-varying treatments
3. Time-varying treatments and confounders on a DAG
4. Sequential exchangeability
5. Treatment-confounder feedback

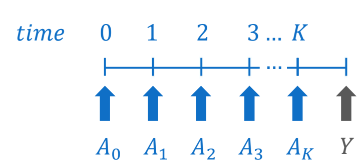
NOTATION FOR TIME-VARYING TREATMENTS.

Before, we defined the causal effect as:

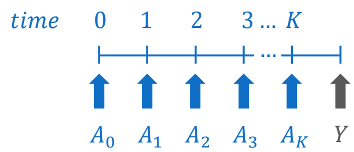
$$\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$$

- No reference to time in that notation
- Treatment status is determined at a single time point for everyone (i.e., at baseline/time zero)

When treatment status can change over time, we need to specify *when* we are assessing treatment:



TREATMENT HISTORY.



Treatment status can change over time, e.g.,

$A_0 = 0$ if someone didn't take treatment at time 0
$A_1 = 1$ if that person then starts taking treatment at time 1
$A_2 = 1$ if that person continues to take treatment at time 2
$A_3 = 0$ if that person then stops taking treatment at time 3
…

We use an overbar over treatment A_k denote the treatment history from the beginning of the study (time 0) to time k

e.g.,

$$\bar{A}_3 = (A_0, A_1, A_2, A_3)$$

When we refer to the entire treatment history (from time 0 to time K), we denote this as

$\bar{A}_K$ or as $\bar{A}$ (without a subscript)

TREATMENT STRATEGIES.

A treatment strategy is a rule to assign treatment at each time point from the beginning to the end of the study.

Examples:

- Never treat
$$\bar{A} = (A_0 = 0, A_1 = 0, A_2 = 0, \dots, A_K = 0)$$
$$\equiv (0,0,0, \dots 0)$$
$$\equiv \bar{0}$$
- Always treat
$$\bar{A} = (1,1,1, \dots 1) \equiv \bar{1}$$
- Treat at every other time point, starting with giving treatment at time 0
$$\bar{A} = (1,0,1,0, \dots)$$
- Treat while $L_k = 0$; stop treatment when $L_k = 1$ and stay off treatment after that time

STATIC VERSUS DYNAMIC TREATMENT STRATEGIES.

The first three examples of treatment strategies on the previous slides are examples of static treatment strategies:

- Never treat: $\bar{A} = (0,0,0, \dots 0)$
- Always treat: $\bar{A} = (1,1,1, \dots 1)$
- Treat at every other time point, starting with giving treatment at time 0: $\bar{A} = (1,0,1,0, \dots)$

Treatment assignment at each time point does not depend on a time-varying covariate L_k

The last example is an example of a dynamic treatment strategy:

- Treat while $L_k = 0$; stop treatment when $L_k = 1$ and stay off treatment after that time
- E.g., let L_k represent the development of a contraindication; the strategy above says to take medication until the development of a contraindication (i.e., $L_k = 1$) and stay off treatment thereafter

Assignment of treatment in a dynamic treatment strategies relies on the evolution of a time-varying L_k

CAUSAL EFFECT FOR A TIME-VARYING TREATMENT.

A causal effect for a time-varying treatment is a contrast between the mean counterfactual outcomes under two different treatment strategies:

$$\mathrm{E}[Y^{\bar{a}}] - \mathrm{E}[Y^{\bar{a}'}]$$

For example, perhaps we want to compare the strategy "always treat" against the strategy "never treat". We can define the causal estimand as:

$$\mathrm{E}[Y^{\bar{a}=\bar{1}}] - \mathrm{E}[Y^{\bar{a}'=\bar{0}}]$$

For simplicity, we're only going to consider comparing the effects of static treatment strategies. Estimating the effects of dynamic treatment strategies adds additional complexities that are outside the scope of this lecture.

Shi – Time-varying treatment strategies 1

PLAN FOR TODAY.

1. Recap
2. Defining causal effects for time-varying treatments
3. Time-varying treatments and confounders on a DAG
4. Sequential exchangeability
5. Treatment-confounder feedback

TIME-VARYING TREATMENTS ON A DAG.

For a time-fixed treatment, we could have the following DAG:

$$A \longrightarrow Y$$

This is the DAG we would expect for an ideal randomized trial of a time-fixed treatment.

For a time-varying treatment, we could have the following DAG:

$$\begin{tikzcd} A_0 \arrow[rrrrr, bend left] & A_1 \arrow[rrrr, bend left] & A_2 \arrow[rrr, bend left] & \cdots & A_K \arrow[r] & Y \end{tikzcd}$$

Notice that each time point of the time-varying treatment is a separate node.

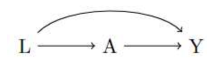
This is the DAG for a sequentially randomized trial of a time-varying treatment:

- Randomize treatment at time 0 ($A_0 = 0$ or $A_0 = 1$)
- Randomize treatment at time 1 ($A_1 = 0$ or $A_1 = 1$)
- etc.

TIME-VARYING TREATMENTS AND CONFOUNDERS ON A DAG.

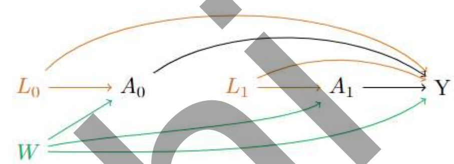
With observational data, we expect to have treatment-outcome confounders.

For a time-fixed treatment, we only had to worry about confounders at baseline:



For a time-varying treatment, we could have:

- Time-fixed confounders (W) at baseline, <u>and</u>
- Time-varying confounders (L_k)

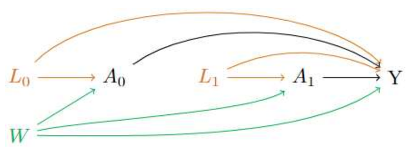


Note: We'll start using just 2 time points for simplicity but everything we'll discuss also applies to scenarios with >2 time points

PLAN FOR TODAY.

1. Recap
2. Defining causal effects for time-varying treatments
3. Time-varying treatments and confounders on a DAG
4. Sequential exchangeability
5. Treatment-confounder feedback

SEQUENTIAL EXCHANGEABILITY.



Suppose we're interested in estimating the effect of always taking versus never taking treatment on outcome Y:

$$\mathrm{E}\left[Y^{(a_0=1,a_1=1)}\right] - \mathrm{E}\left[Y^{(a_0=0,a_1=0)}\right]$$

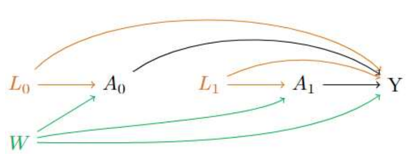
To do so, we need conditional exchangeability to hold at both time 0 and time 1.

Poll question 1: Based on the DAG to the left, what variables do we need to condition on for exchangeability to hold at time 0 ?

A. L_0
B. W
C. All of the above

Shi – Time-varying treatment strategies 1

SEQUENTIAL EXCHANGEABILITY.



Suppose we're interested in estimating the effect of always taking versus never taking treatment on outcome Y:

$$\mathrm{E}\left[Y^{(a_0=1,a_1=1)}\right] - \mathrm{E}\left[Y^{(a_0=0,a_1=0)}\right]$$

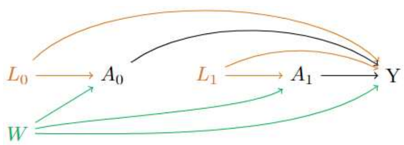
To do so, we need (conditional) exchangeability to hold at both time 0 and time 1.

Poll question 2: Based on the DAG to the left, what variables do we need to condition on for exchangeability to hold at time 1 ?

A. L_0 and L_1
B. L_0 and W
C. L_1 and W
D. L_0, L_1 and W
E. L_0, L_1, W and A_0

Shi – Time-varying treatment strategies 1

SEQUENTIAL EXCHANGEABILITY.



Suppose we're interested in estimating the effect of always taking versus never taking treatment on outcome Y:

$$\mathrm{E}\left[Y^{(a_0=1,a_1=1)}\right] - \mathrm{E}\left[Y^{(a_0=0,a_1=0)}\right]$$

To do so, we need (conditional) exchangeability to hold at <u>both</u> time 0 and time 1.

Two conditional exchangeability conditions need to hold:

$$Y^{a_0,a_1} \perp\!\!\!\perp A_0 | L_0, W$$
$$Y^{a_0,a_1} \perp\!\!\!\perp A_1 | L_1, W$$

If, at each time point, the exposure is independent of the counterfactual outcome (conditional on confounders), then we have conditional sequential exchangeability

Shi – Time-varying treatment strategies 1

SEQUENTIAL EXCHANGEABILITY: FORMAL DEFINITION.

The treated and untreated are exchangeability at each time k conditional on prior covariate history $\bar{L}_k$ and treatment history $\bar{A}_{k-1}$:

$$Y^{\bar{a}} \perp\!\!\!\perp A_k \mid \bar{L}_k, \bar{A}_{k-1}$$

e.g.,

For $k = 0$: $Y^{\bar{a}} \perp\!\!\!\perp A_0 \mid L_0$

For $k = 1$: $Y^{\bar{a}} \perp\!\!\!\perp A_1 \mid L_1, L_0, A_0$

For $k = 2$: $Y^{\bar{a}} \perp\!\!\!\perp A_2 \mid L_2, L_1, L_0, A_1, A_0$

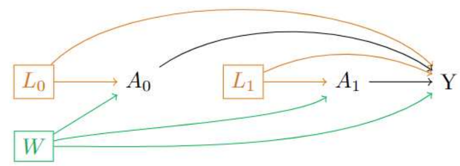
etc.

Shi – Time-varying treatment strategies 1

ESTIMATING THE JOINT EFFECT OF A_0 AND A_1.

In this particular example, we can estimate the joint effect of A_0 and A_1 by conditioning on:

L_0, L_1 and W



For example, we could fit the following outcome regression model:

$$\mathrm{E}[Y|A_0, A_1, L_0, L_1, W] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 L_0 + \beta_4 L_1 + \beta_5 W$$

Poll question 1: Using outcome regression, we can estimate:

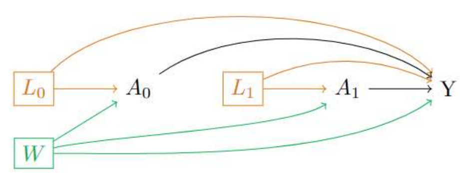
A. A marginal causal effect of A_0 and A_1 (e.g., $\mathrm{E}\left[Y^{(a_0=1,a_1=1)}\right] - \mathrm{E}\left[Y^{(a_0=0,a_1=0)}\right]$)

B. A conditional causal effect of A_0 and A_1 (e.g., $\mathrm{E}\left[Y^{(a_0=1,a_1=1)}|L_0, L_1, W\right] - \mathrm{E}\left[Y^{(a_0=0,a_1=0)}|L_0, L_1, W\right]$)

ESTIMATING THE JOINT EFFECT OF A_0 AND A_1.

In this particular example, we can estimate the joint effect of A_0 and A_1 by conditioning on:

L_0, L_1 and W



For example, we could fit the following outcome regression model:

$$\mathrm{E}[Y|A_0, A_1, L_0, L_1, W] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 L_0 + \beta_4 L_1 + \beta_5 W$$

Poll question 2: $\mathrm{E}\left[Y^{(a_0=1,a_1=1)}|L_0, L_1, W\right] - \mathrm{E}\left[Y^{(a_0=0,a_1=0)}|L_0, L_1, W\right]$ is equal to

A. β_1

B. β_2

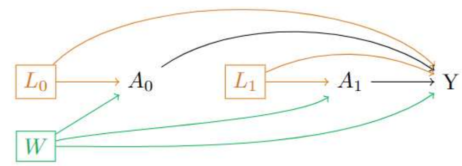
C. $\beta_1 + \beta_2$

D. $\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5$

ESTIMATING THE JOINT EFFECT OF A_0 AND A_1.

In this particular example, we can estimate the joint effect of A_0 and A_1 by conditioning on:

L_0, L_1 and W



For example, we could fit the following outcome regression model:

$$\mathrm{E}[Y|A_0, A_1, L_0, L_1, W] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 L_0 + \beta_4 L_1 + \beta_5 W$$

β_1 represents the effect of A_0, conditional on A_1, L_0, L_1 and W

β_2 represents the effect of A_1, conditional on A_0, L_0, L_1 and W

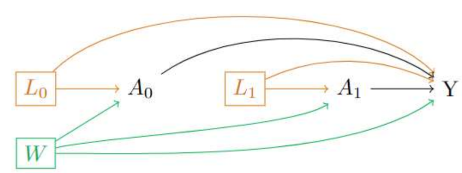
$\beta_1 + \beta_2$ represents the joint effect of A_0 and A_1, conditional on L_0, L_1 and W

Shi – Time-varying treatment strategies 1

ESTIMATING THE JOINT EFFECT OF A_0 AND A_1.

In this particular example, we can estimate the joint effect of A_0 and A_1 by conditioning on:

L_0, L_1 and W



We could also have interaction between A_0 and A_1:

$$\mathrm{E}[Y|A_0, A_1, L_0, L_1, W] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 A_0 A_1 + \beta_4 L_0 + \beta_5 L_1 + \beta_6 W$$

i.e., the effect of A_0 and A_1 together is more than the sum of their individual contributions.

Poll question 3: $\mathrm{E}\left[Y^{(a_0=1,a_1=1)}|L_0, L_1, W\right] - \mathrm{E}\left[Y^{(a_0=0,a_1=0)}|L_0, L_1, W\right]$ is equal to

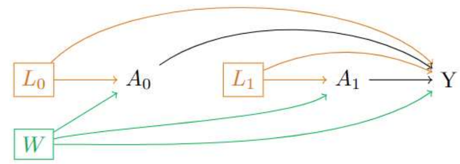
A. β_1
B. β_2
C. $\beta_1 + \beta_2$
D. $\beta_1 + \beta_2 + \beta_3$
E. $\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6$

Shi – Time-varying treatment strategies 1

ESTIMATING THE JOINT EFFECT OF A_0 AND A_1.

In this particular example, we can estimate the joint effect of A_0 and A_1 by conditioning on:

L_0, L_1 and W



We could also have interaction between A_0 and A_1:

$$\mathrm{E}[Y|A_0, A_1, L_0, L_1, W] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 A_0 A_1 + \beta_4 L_0 + \beta_5 L_1 + \beta_6 W$$

i.e., the effect of A_0 and A_1 together is more than the sum of their individual contributions.

β_1 represents the effect of A_0 when $A_1 = 0$, conditional on L_0, L_1 and W
$\beta_1 + \beta_3$ represents the effect of A_0 when $A_1 = 1$, conditional on L_0, L_1 and W
β_2 represents the effect of A_1 when $A_0 = 0$, conditional on L_0, L_1 and W
$\beta_2 + \beta_3$ represents the effect of A_1 when $A_0 = 1$, conditional on L_0, L_1 and W
$\beta_1 + \beta_2 + \beta_3$ represents the joint effect of A_0 and A_1, conditional on L_0, L_1 and W

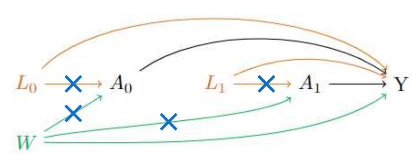
ESTIMATING THE JOINT EFFECT OF A_0 AND A_1.

We could also use g-methods, e.g.,

- Inverse probability weighting (more on this tomorrow)
- g-formula (generalization of standardization for time-varying treatments)

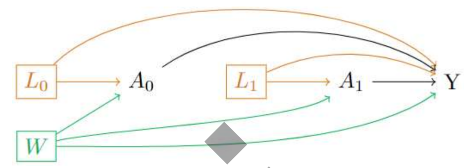
to estimate the joint effect of A_0 and A_1

With IPW, we're removing the arrows from the confounders (L_0, L_1 and W) to treatment (A_0 and A_1)



CONVENTIONAL METHODS FOR TIME-VARYING EXPOSURES.

In this particular example, conventional methods (e.g., outcome regression, propensity scores) work:



Next, we'll discuss how conventional methods fail when there is treatment-confounder feedback.

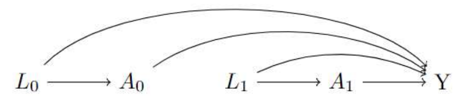
- Can only use g-methods (i.e., IPW, g-formula or g-estimation)

PLAN FOR TODAY.

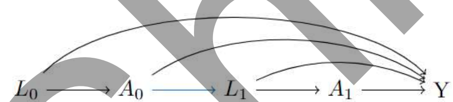
1. Recap
2. Defining causal effects for time-varying treatments
3. Time-varying treatments and confounders on a DAG
4. Sequential exchangeability
5. Treatment-confounder feedback

STRUCTURE OF TREATMENT-CONFOUNDER FEEDBACK.

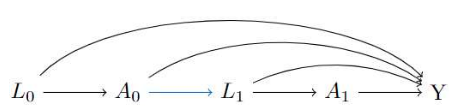
- Treatment-confounder feedback occurs when there is time-varying confounding
- We'll use the same DAG as before, but I'll remove the time-fixed confounder W to simplify the structure of the DAG:



- Treatment-confounder feedback arises once treatment affects the confounder at a later time
- i.e., we have treatment-confounder feedback once we add an arrow from A_0 to L_1



ESTIMATING THE JOINT EFFECT OF A_0 AND A_1 WITH TREATMENT-CONFOUNDER FEEDBACK.



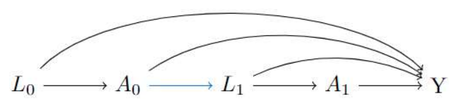
Poll question 1: In the DAG to the left, which variable(s) do we need to condition on in order to block all backdoor paths from A_0 to Y?

A. L_0
B. L_1
C. A_1
D. All of the above

Poll question 2: In the DAG to the left, which variable(s) do we need to condition on in order to block all backdoor paths from A_1 to Y?

A. L_0
B. L_1
C. A_0
D. All of the above

ESTIMATING THE JOINT EFFECT OF A_0 AND A_1 WITH TREATMENT-CONFOUNDER FEEDBACK.



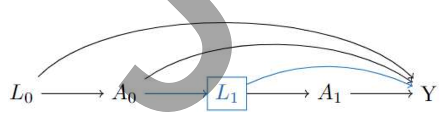
We have two exchangeability conditions:

$$Y^{a_0,a_1} \perp\!\!\!\perp A_0 | L_0$$
$$Y^{a_0,a_1} \perp\!\!\!\perp A_1 | L_1$$

Poll question 3: However, what happens to the effect of A_0 on Y when we condition on L_1?

A. We block some of the effect of A_0 on Y
B. We introduce selection bias because L_1 is a collider
C. Nothing

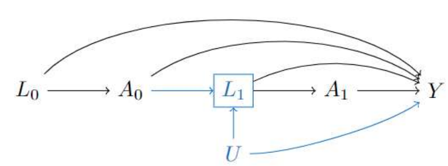
CONDITIONING ON L_1.



- Conditioning on L_1 blocks the path
$$A_0 \rightarrow L_1 \rightarrow Y$$
- We aren't capturing the effect of A_0 that is mediated through L_1

And it gets worse…

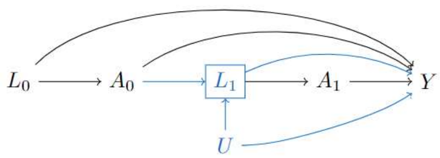
If there are unmeasured confounders (U) between L_1 and Y:



- L_1 is now a collider
- Conditioning on L_1 creates selection bias because it opens the biasing path
$$A_0 \rightarrow L_1 \leftarrow U \rightarrow Y$$

TO STRATIFY OR NOT TO STRATIFY?

If we stratify on L_1:

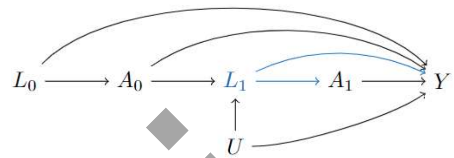


- We block the backdoor path: A_1 to L_1 to Y

BUT...

- We block some of the effect of A_0 on Y
- We introduce selection bias because L_1 is a collider: the path A_0 to L_1 to U to Y is open

If we don't stratify on L_1:



- We avoid selection bias through L_1
- We estimate the total effect of A_0 on Y (not through A_1)

BUT...

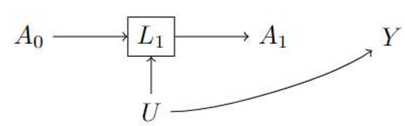
- The effect of A_1 on Y is confounded by L_1

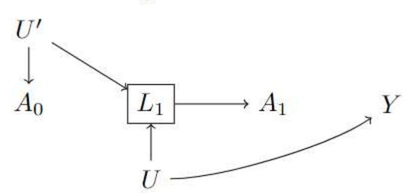
ADDITIONAL DAG STRUCTURES FOR TREATMENT-CONFOUNDER FEEDBACK.

Treatment-confounder feedback occurs if:

- The confounder is affected by treatment; or
- The confounder and treatment share common causes

Notice that there is an issue of selection bias when conditioning on L_1 even if A_0, A_1 and L_1 do not have a direct effects on Y





In general, be very cautious about conditioning on post-baseline variables – doing so could introduce selection bias.

METHODS TO HANDLE TREATMENT-CONFOUNDER FEEDBACK.

Conventional methods which eliminate confounding by conditioning on L (or functions of L, i.e., propensity scores) will fail in the presence of treatment-confounder feedback, e.g.

- Restriction
- Outcome regression (e.g., linear, logistic, Cox proportional hazards)
- Propensity score adjustment

Need to use g-methods

- G-formula
- Inverse probability weighting
- G-estimation

TAKEAWAYS.

- With time-varying treatments, we introduce new notation where the treatment is indexed by time (e.g., A_0, A_1, …)
- Causal effects for time-varying treatments are contrasts between the mean counterfactual outcomes under different treatment strategies:

$$\mathrm{E}[Y^{\bar{a}}] - \mathrm{E}[Y^{\bar{a}'}] \text{ where } \bar{a} = (a_0, a_1, a_2, \dots)$$

- On a DAG, we indicate time-varying treatments by having separate nodes for each time point
- To estimate the effects of time-varying treatments, we need (conditional) exchangeability to hold for treatment at each time point (i.e. sequential exchangeability)
- In the presence of treatment-confounder feedback, we cannot use conventional methods; we must use g-methods

LEARNING OBJECTIVES.

By the end of the session, you will be able to:

1. Formulate causal questions for time-varying treatments
2. Describe treatment-confounder feedback and recognize its structure on a DAG
3. Understand why conventional methods fail in the presence of treatment-confounder feedback