DIRECTED ACYCLIC GRAPHS

INFORMATION BIAS AND TIME-VARYING TREATMENTS

March 4, 2021

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

LEARNING OBJECTIVES.

After this session, you should be able to:

1. Identify different types of information bias on a DAG
2. Recognize the structure of treatment-confounder feedback on a DAG.
3. Identify situations when stratification-based methods fail.
4. Devise an approach to draw your own causal DAGs.

RECAP FROM OUR LAST SESSION: D-SEPARATION.

A set of rules that allow us determine whether two variables on a DAG are associated (i.e. whether the path between them is open or blocked)

1. If there are no variables being conditioned on, a path is blocked if two arrowheads on a path collide at some variable on the path.
2. A path that contains a non-collider that is conditioned on is blocked.
3. A collider that has been conditioned on does not block a path.
4. A collider that has a descendant that has been conditioned on does not block a path.

RECAP FROM OUR LAST SESSION: DAG STRUCTURES.

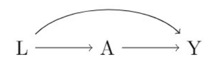
	DAG	Are we conditioning on anything?	Are A and Y associated?	Conclusion
Mediator	$A \rightarrow M \rightarrow Y$	No	Yes	A and Y are marginally associated
	$A \rightarrow \boxed{M} \rightarrow Y$	Yes	No	A and Y are independent, conditional on M
Common cause	$L \rightarrow A \quad Y$	No	Yes	A and Y are marginally associated
	$L \rightarrow \boxed{A} \quad Y$	Yes	No	A and Y are independent, conditional on L
Common effect	$A \quad Y \rightarrow L$	No	No	A and Y are marginally independent
	$A \quad Y \rightarrow \boxed{L}$	Yes	Yes	A and Y are associated, conditional on L
	$A \quad Y \rightarrow L \rightarrow \boxed{S}$	Yes	Yes	A and Y are associated, conditional on S

RECAP FROM OUR LAST SESSION: CONFOUNDING AND SELECTION BIAS.

We discussed two structural sources of bias

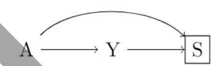
Confounding

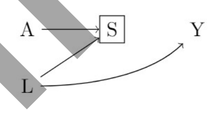
- Common cause of exposure (A) and outcome (Y)
- Open backdoor path from exposure (A) to outcome (Y)



Selection bias

- Selection (S) of participants into a study and/or analysis
- Conditioning on a common effect of treatment (or a cause of treatment) and outcome (or a cause of the outcome)



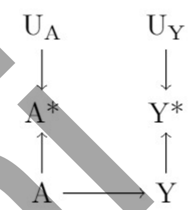


INFORMATION BIAS

INFORMATION BIAS.

- Arises from imperfect definition of study variables or flawed data collection procedures
- Also referred to as measurement bias, misclassification bias, recall bias, recall error

- Here is an example of a DAG with measurement error in the exposure and outcome:
  - Indicate a mismeasured variable with a star (A* and Y*)
  - The true value of the variable affects the measured value (arrows from A to A* and from Y to Y*)
  - U_A and U_Y are the measurement error for A and Y, respectively (i.e. factors other than A and Y that determine the value of A* and Y*)



FOUR TYPES OF STRUCTURES FOR INFORMATION BIAS.

Can classify measurement error in the treatment and outcome as being
- Independent vs. dependent
- Nondifferential vs. differential

This gives us four types of measurement error:
1. Independent nondifferential
2. Dependent nondifferential
3. Independent differential
4. Dependent differential

INDEPENDENT NONDIFFERENTIAL ERRORS.

$$
\begin{tikzcd}
U_A \arrow[d] & U_Y \arrow[d] \\
A^* & Y^* \\
A \arrow[u] \arrow[r] & Y \arrow[u]
\end{tikzcd}
$$

Measurement errors U_A and U_Y are:

- Independent: the path from U_A to U_Y is blocked by colliders (A^* and Y^*)
- Non-differential: error for the exposure, U_A, is independent of the true outcome, Y (and similarly, U_Y is independent of A)

Example:

A: Vitamin D status

A*: Self reported vitamin D intake using a food frequency questionnaire

Y: Mortality

Y*: National Death Index

DEPENDENT NONDIFFERENTIAL ERRORS.

$$
\begin{tikzcd}
 & U_{AY} \arrow[dl] \arrow[dr] & \\
U_A \arrow[d] & & U_Y \arrow[d] \\
A^* & & Y^* \\
A \arrow[u] \arrow[rr] & & Y \arrow[u]
\end{tikzcd}
$$

Measurement errors U_A and U_Y are:

- Dependent: U_A to U_Y are associated through a common cause, U_{AY}
- Non-differential: error for the exposure, U_A, is independent of the true outcome, Y (and similarly, U_Y is independent of A)

Example:

A: Childhood chemical hair product use

A*: Retrospectively self reported via questionnaire

Y: Age at menarche

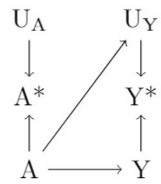
Y*: Retrospectively self-reported via questionnaire

U_{AY}: Memory

INDEPENDENT DIFFERENTIAL ERRORS.

Measurement errors U_A and U_Y are:

- Independent: the paths from U_A to U_Y is blocked by colliders (A* or Y*)
- Differential: U_A is associated with Y, or U_Y is associated with A

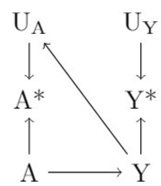


Example 1: error for the outcome is differential with respect to the exposure (i.e. U_Y is associated with A)

A: Elective surgery

Y: Quality of life

Y*: Self-reported via questionnaire



Example 2: error for the exposure is differential with respect to the outcome (i.e. U_A is associated with Y)

A: Oral contraceptive use

A*: Self-reported after knowing outcome status (e.g. case-control study)

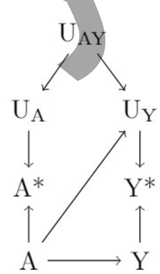
Y: Breast cancer

Note: often referred to as recall bias

DEPENDENT DIFFERENTIAL ERRORS.

Measurement errors U_A and U_Y are:

- Dependent: U_A to U_Y are associated through a common cause, U_{AY}
- Differential: U_A is associated with Y, or U_Y is associated with A

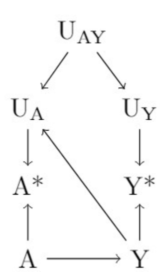


Example 1: error for the outcome is differential with respect to the exposure (i.e. U_Y is associated with A) and dependent errors

A: Chemotherapy

Y: Cancer progression

A* and Y*: Retrospectively collected using medical records



Example 2: error for the exposure is differential with respect to the outcome (i.e. U_A is associated with Y) and dependent errors

A: Cholesterol intake

A*: Retrospectively assessed via FFQ

Y: Dementia

Y*: Self-reported dementia

DIRECTION OF INFORMATION BIAS.

- Bias arises from using A* and Y* to estimate the association between A and Y
- Under certain (but not all!) scenarios, expect independent nondifferential errors to bias towards null
- Direction of bias for other types of errors can be in any direction

MEASUREMENT ERROR FOR CONFOUNDERS.

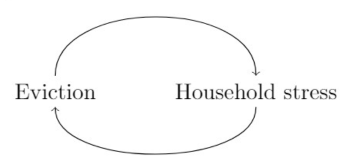
$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] \arrow[d] & A \arrow[r] & Y \\ \boxed{L^*} & & \end{tikzcd}$$

- L* acts as a surrogate confounder
- Partially blocks the backdoor path but there is still residual confounding

TIME-VARYING EXPOSURES AND CONFOUNDERS.

BUT WHAT ABOUT FEEDBACK LOOPS?

- Many pathways of interest in epidemiologic research are "cyclical"
- For example:



- Eviction can cause household stress, which can contribute to being evicted again
- However, this is not a DAG: recall that DAGs are acyclic (i.e. no feedback loops)
- How do we represent a process like this using a DAG?

TIME-VARYING VARIABLES.

- In the previous example (eviction and household stress), we can conceptualize both variables to be time-varying
- Once we start thinking about repeated measures of a time-varying variable, we need to specify the time point that these variables were measured
- Returning to our previous example, we can turn it into the following DAG:

$$\text{Eviction}_1 \longrightarrow \text{Household stress}_2 \longrightarrow \text{Eviction}_3 \longrightarrow \text{Household stress}_4$$

- This type of feedback is common with exposures and confounders (i.e. treatment-confounder feedback) and pose additional analytical challenges

EVICTION AND CHILD COGNITIVE OUTCOMES: TIME-FIXED VARIABLES.

- Consider a study which examines the effect of eviction on child cognitive outcomes at age 10:
  - Exposure (A): whether or not eviction occurred before the age of 10
  - Outcome (Y): child cognition at age 10
  - Confounder (L): household stress
- If we were thinking about these variables as time-fixed, we might have drawn the following DAG:

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

EVICTION AND CHILD COGNITIVE OUTCOMES: TIME-VARYING VARIABLES.

- However, suppose we measured the exposure and confounder at two time points:
  - Exposure: occurrence of eviction between ages 6 and <8 (A_1), occurrence of eviction between ages 8 and <10 (A_2)
  - Outcome (Y): child cognition at age 10
  - Confounder: household stress at age 6 (L_1) and household stress at age 8 (L_2)
- How do we expand our DAG to include multiple time points for the exposure and confounder?

DAG WITH TIME-VARYING VARIABLES: STEP 1.

First, let's consider just the exposure and outcome:

- We can add arrows from each exposure time point to the outcome
- We can add an arrow from A_1 to A_2 (prior eviction may affect later eviction)

$$\begin{tikzcd} A_1 \arrow[r, bend left] \arrow[rr, bend left] & A_2 \arrow[r] & Y \end{tikzcd}$$

L_1: household stress at age 6

A_1: occurrence of eviction between ages 6 and <8

L_2: household stress at age 8

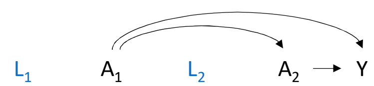
A_2: occurrence of eviction between ages 8 and <10

Y: child cognition at age 10

DAG WITH TIME-VARYING VARIABLES: STEP 2.

Next, we add in household stress as a confounder

- For simplicity, we consider household stress as the only confounder here (in practice, L may represent a vector of covariates for multiple confounders)
- Household stress is also time-varying



L_1: household stress at age 6

A_1: occurrence of eviction between ages 6 and <8

L_2: household stress at age 8

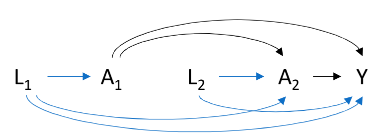
A_2: occurrence of eviction between ages 8 and <10

Y: child cognition at age 10

DAG WITH TIME-VARYING VARIABLES: STEP 3.

Now, we add arrows from confounders to exposure and from confounders to outcome

- Arrows from L_1 to A_1 and A_2, and from L_2 to A_2 because prior household stress can affect eviction
- Arrows from L_1 to Y and L_2 to Y because household stress can affect child cognition



L_1: household stress at age 6

A_1: occurrence of eviction between ages 6 and <8

L_2: household stress at age 8

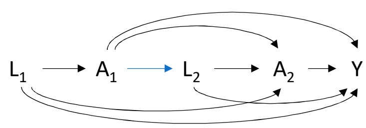
A_2: occurrence of eviction between ages 8 and <10

Y: child cognition at age 10

DAG WITH TIME-VARYING VARIABLES: STEP 4.

We also add an arrow from A_1 to L_2:

- Prior eviction could affect later household stress
- This introduces treatment-confounder feedback; without this arrow, we simply having time-varying exposures and confounders



L_1: household stress at age 6

A_1: occurrence of eviction between ages 6 and <8

L_2: household stress at age 8

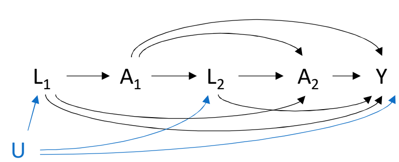
A_2: occurrence of eviction between ages 8 and <10

Y: child cognition at age 10

DAG WITH TIME-VARYING VARIABLES: STEP 5.

Last, we add any other common causes of variables on the graph

- A DAG is only considered a causal DAG if common causes of any pair of variables on the graph are also included
- Add potential common causes for L_1, L_2 and Y



L_1: household stress at age 6

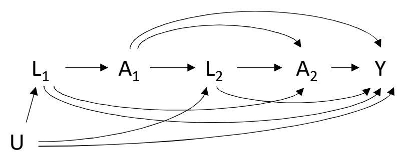
A_1: occurrence of eviction between ages 6 and <8

L_2: household stress at age 8

A_2: occurrence of eviction between ages 8 and <10

Y: child cognition at age 10

ESTIMATING JOINT EFFECTS.

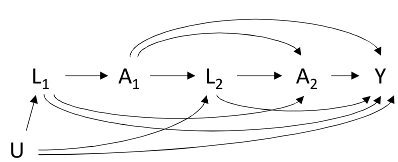


We may be interested in the joint effect of both exposure time points on child cognition, e.g. effect of:

- $A_1=1$, $A_2=1$: Getting evicted at both age periods (from ages 6 to <8 *and* from ages 8 to <10), vs.
- $A_1=0$, $A_2=0$: Not getting evicted during either age period, vs.
- $A_1=0$, $A_2=1$: Not getting evicted during age 6 to <8, but getting evicted during ages 8 to <10, vs.
- $A_1=1$, $A_2=0$: Getting evicted during ages 6 to <8, but not getting evicted during age 8 to <10

To estimate joint effects, we need to consider sources of bias for *both* the effect of A_1 on Y and for the effect of A_2 on Y

ESTIMATING JOINT EFFECTS.



We may be interested in the joint effect of both exposure time points on child cognition, e.g. effect of:

- $A_1=1$, $A_2=1$: Getting evicted at both age periods (from ages 6 to <8 *and* from ages 8 to <10), vs.
- $A_1=0$, $A_2=0$: Not getting evicted during either age period, vs.
- $A_1=0$, $A_2=1$: Not getting evicted during age 6 to <8, but getting evicted during ages 8 to <10, vs.
- $A_1=1$, $A_2=0$: Getting evicted during ages 6 to <8, but not getting evicted during age 8 to <10

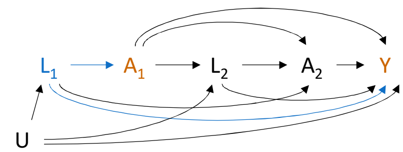
To estimate joint effects, we need to consider sources of bias for *both* the effect of A_1 on Y and for the effect of A_2 on Y

ESTIMATING JOINT EFFECTS: SOURCES OF BIAS FOR THE EFFECT OF A_1 ON Y.

There are multiple paths from A_1 to Y.

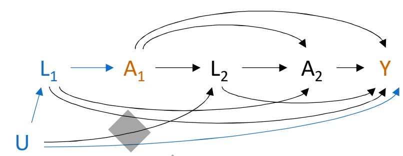
Path 1:

A_1 to L_1 to Y



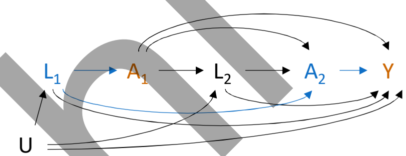
Path 2:

A_1 to L_1 to U to Y

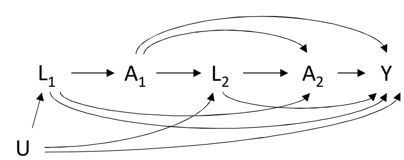


Path 3:

A_1 to L_1 to A_2 to Y



ESTIMATING JOINT EFFECTS: POLL QUESTION 1.



In the DAG above, which variables do we need to condition on in order to block all backdoor paths from A_1 to Y?

A. L_1

B. U

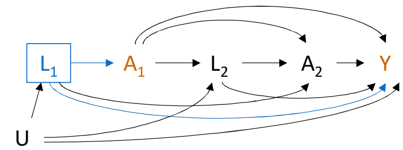
C. A_2

D. All of the above

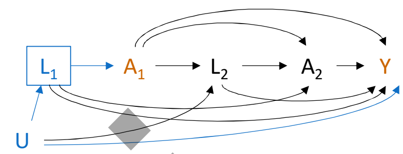
ESTIMATING JOINT EFFECTS: SOURCES OF BIAS FOR THE EFFECT OF A_1 ON Y.

- Only need to condition on L_1 to block all backdoor paths
- Can't condition on U because it is unmeasured
- Don't want to condition on A_2 because this would block some of the effect of A_1 on Y (e.g. the path A_1 to A_2 to Y)

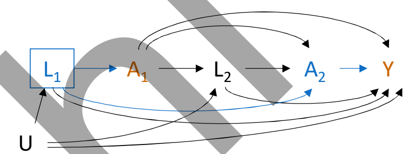
Path 1:
A_1 to L_1 to Y



Path 2:
A_1 to L_1 to U to Y



Path 3:
A_1 to L_1 to A_2 to Y

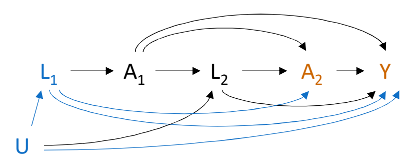


ESTIMATING JOINT EFFECTS: SOURCES OF BIAS FOR THE EFFECT OF A_2 ON Y.

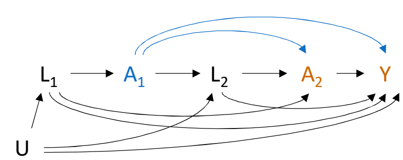
There are even more backdoor paths from A_2 to Y.

There are three arrows going into A_2
- L_1 to A_2
- A_1 to A_2
- L_2 to A_2

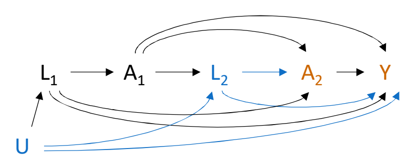
Paths starting with arrow from L_1:
A_2 to L_1 to Y
A_2 to L_1 to U to Y

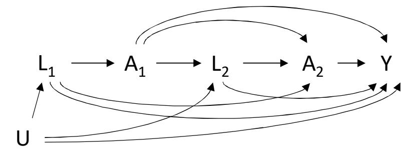


Paths starting with arrow from A_1:
A_2 to A_1 to Y
...plus more with A_2 to A_1 to L...



Paths starting with arrow from L_2:
A_2 to L_2 to Y
A_2 to L_2 to U to Y
...plus more...





ESTIMATING JOINT EFFECTS: POLL QUESTION 2.

In the DAG above, which set of variables do we need to condition on in order to block all backdoor paths from A_2 to Y?

A. L_2 only

B. L_1 and L_2 only

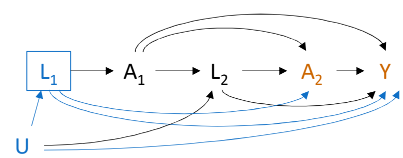
C. L_1, A_1 and L_2 only

D. L_1, A_1, L_2 and U only

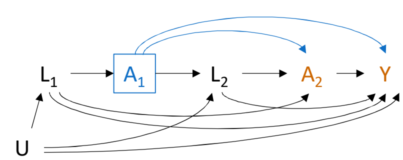
ESTIMATING JOINT EFFECTS: SOURCES OF BIAS FOR THE EFFECT OF A_2 ON Y.

Need to condition on L_1, A_1 and L_2 to block all backdoor paths

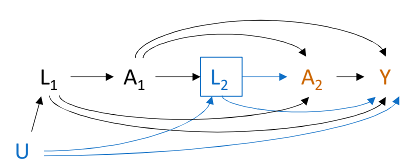
Paths starting with arrow from L_1:
A_2 to L_1 to Y
A_2 to L_1 to U to Y



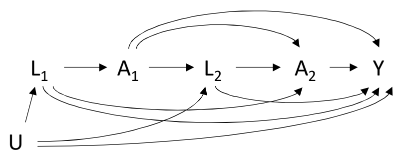
Paths starting with arrow from A_1:
A_2 to A_1 to Y
...plus more with A_2 to A_1 to L_1...



Paths starting with arrow from L_2:
A_2 to L_2 to Y
A_2 to L_2 to U to Y
...plus more...



ESTIMATING JOINT EFFECTS: SUMMARY.



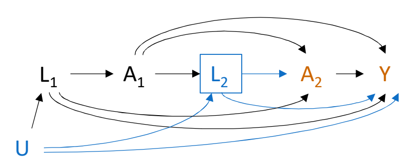
To block all backdoor paths between A_1 and Y:
need to condition on L1

To block all backdoor paths between A_2 on Y:
need to condition on L_1, A_2 and L_2

We can't have any open backdoor paths from A_1 to Y or from A_2 to Y to estimate the joint effects of A_1 and A_2 on Y

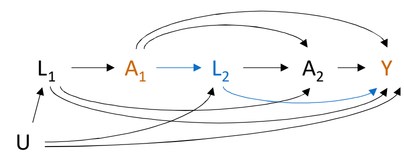
LET'S FURTHER CONSIDER WHAT HAPPENS WE CONDITION ON L_2.

When we condition on L_2, we block some of the backdoor paths between A_2 and Y:



BUT let's consider what happens to some of the paths from A_1 to Y.

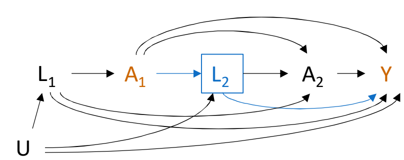
CONDITIONING ON L_2: POLL QUESTION 1.



Consider the path A_1 to L_2 to Y in the DAG above. What happens when we condition on L_2?

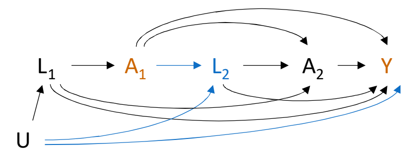
A. We block this path after conditioning on L_2.
B. We open this path after conditioning on L_2.
C. Nothing happens, the path stays open.
D. Nothing happens, the path stays blocked.

CONDITIONING ON L_2: POLL QUESTION 2.



Consider the path A_1 to L_2 to Y in the DAG above. What is the consequence of blocking this path after conditioning on L_2?

A. We block a non-causal path from A_1 to Y.
B. We block a causal path from A_1 to Y.
C. We eliminate some of the bias for the effect of A_1 on Y.
D. Nothing, we can still estimate all of the effect of A_1 on Y (that does not go through A_2).



CONDITIONING ON L_2: POLL QUESTION 3.

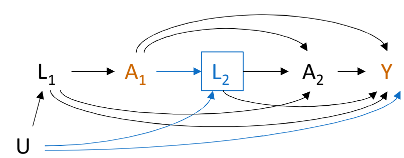
Consider the path A_1 to L_2 to U to Y in the DAG above. What happens when we condition on L_2?

A. We block this path after conditioning on L_2.

B. We open this path after conditioning on L_2.

C. Nothing happens, the path stays open.

D. Nothing happens, the path stays blocked.



CONDITIONING ON L_2: POLL QUESTION 4.

Consider the path A_1 to L_2 to U to Y in the DAG above. Consider the path A_1 to L_2 to Y in the DAG above. What is the consequence of opening this path after conditioning on L_2?

A. We eliminate some of the bias for the effect of A_1 on Y.

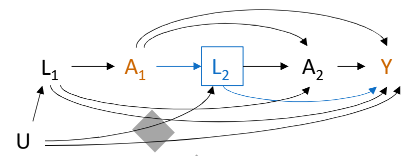
B. We introduce bias for the effect of A_1 on Y.

C. Nothing.

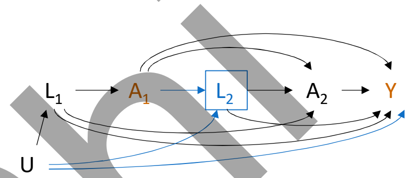
CONSEQUENCES OF CONDITIONING ON L_2.

When we consider some of the paths from A_1 to Y:

- The path that is A_1 to L_2 to Y gets blocked: this prevents us from capturing all of the effect of A_1 on Y that is independent of A_2



- The path that is A_1 to L_2 to U to Y: L_2 is a collider on this path; by conditioning on L_2, we've introduced collider-stratification bias



ANALYTIC STRATEGY IN THE PRESENCE OF TREATMENT-CONFOUNDER FEEDBACK.

We need to deal with the open backdoor paths from A_2 to Y to estimate the effect of A_2 on Y.

- Condition on L_2

BUT

Conditioning on L_2 introduces bias for the effect of A_1 on Y.

This problem arises because we have treatment-confounder feedback.

Stratification-based methods fail because they rely on conditioning on confounders to block backdoor paths

- Outcome regression
- Propensity score
- Matching
- Stratification
- Restriction

Need to use g-methods in the presence of treatment-confounder feedback:

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

More information in What If (Hernán and Robins, 2020)

LEARNING OBJECTIVES.

After this session, you should be able to:

1. Identify different types of information bias on a DAG
2. Recognize the structure of treatment-confounder feedback on a DAG.
3. Identify situations when stratification-based methods fail.
4. Devise an approach to draw your own causal DAGs.