# Topics in Causal Inference

STAT41530
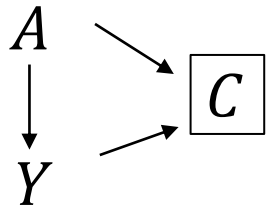
Jingshu Wang

# Lecture 5

Topic:
causal directed acyclic graph (DAG)

- Selection bias

- IPW with do-operator

# Collider bias

$$A \rightarrow C \leftarrow Y$$
$$A \rightarrow Y$$

Conditional on a collider $C$ creates non-causal association between $A$ and $Y$
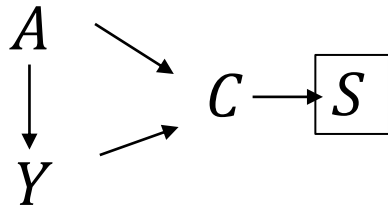
Example:

$A$: Give folic acid supplements to pregnant women shortly after conception

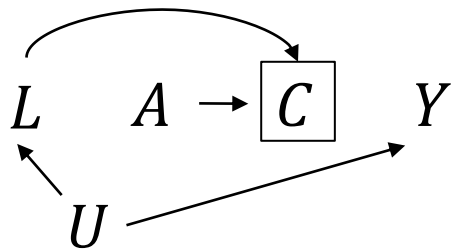$Y$: fetus's risk of developing a cardiac malformation

$C$: survival at birth

Selection bias exists even in randomized controlled experiments!

# Other structures of selection bias



$A$: Give folic acid supplements to pregnant women shortly after conception
$Y$: fetus's risk of developing a cardiac malformation
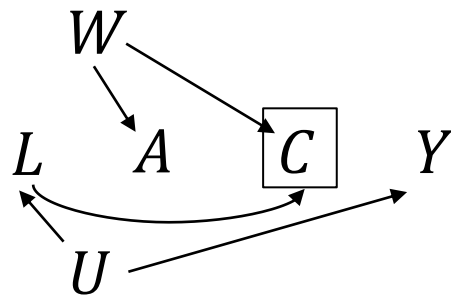$C$: survival at birth
$S$: parent grief



$A$: Antiretroviral treatment on HIV
$Y$: 3-year risk of death
$U$: high-level of immunosuppression
$L$: presence of symptoms
$C$: lost to follow-up

Healthy worker bias:
$A$: Occupational exposure to chemical
$Y$: Mortality
$U/L$: health status
$C$: being at work



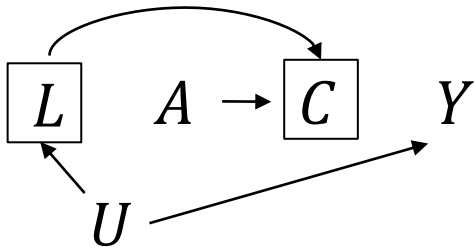$W$: unmeasured lifestyle/personality/educational variables

Volunteer bias:
$A$: cigarette smoking
$Y$: coronary heart disease
$U$: family history of heart disease
$L$: disease awareness
$C$: agree to participate in the study
$W$: healthy lifestyle

# Adjust for selection bias

- Avoid conditioning on post-treatment variables

- Selection bias is often unavoidable

- Adjustment: conditioning on both C and L

$L$ ⟶ $A$ → $C$  $Y$

$U$

$C$ is a collider that must be conditioned on. If there is $L$ satisfying:
1. $L$ satisfies the backdoor criterion
2. $(C \perp\!\!\!\perp Y \mid A, L)_G$

then

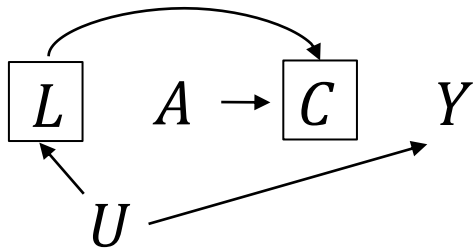$$P(Y|do(A)) = \sum_l P(Y|A, C = 1, L = l)P(L = l)$$

Need positivity assumption:
$P(A = a, C = 1, L = l) > 0$ for all $l$ where $P[L = l] > 0$
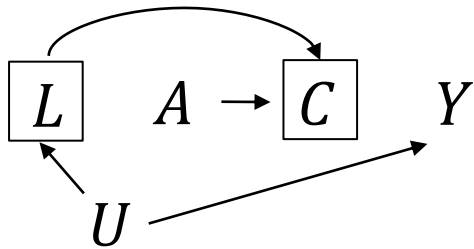- $P(A = a \mid L = l) > 0$
- $P(C = 1 \mid A = a, L = l) > 0$

$$\mathbb{P}[Y \mid do(A)] = \sum_l \mathbb{P}[Y \mid L = l, A] \, \mathbb{P}[L = l]$$

$$= \sum_l \mathbb{P}[Y \mid L = l, A, C = 1] \, \mathbb{P}[L = l]$$
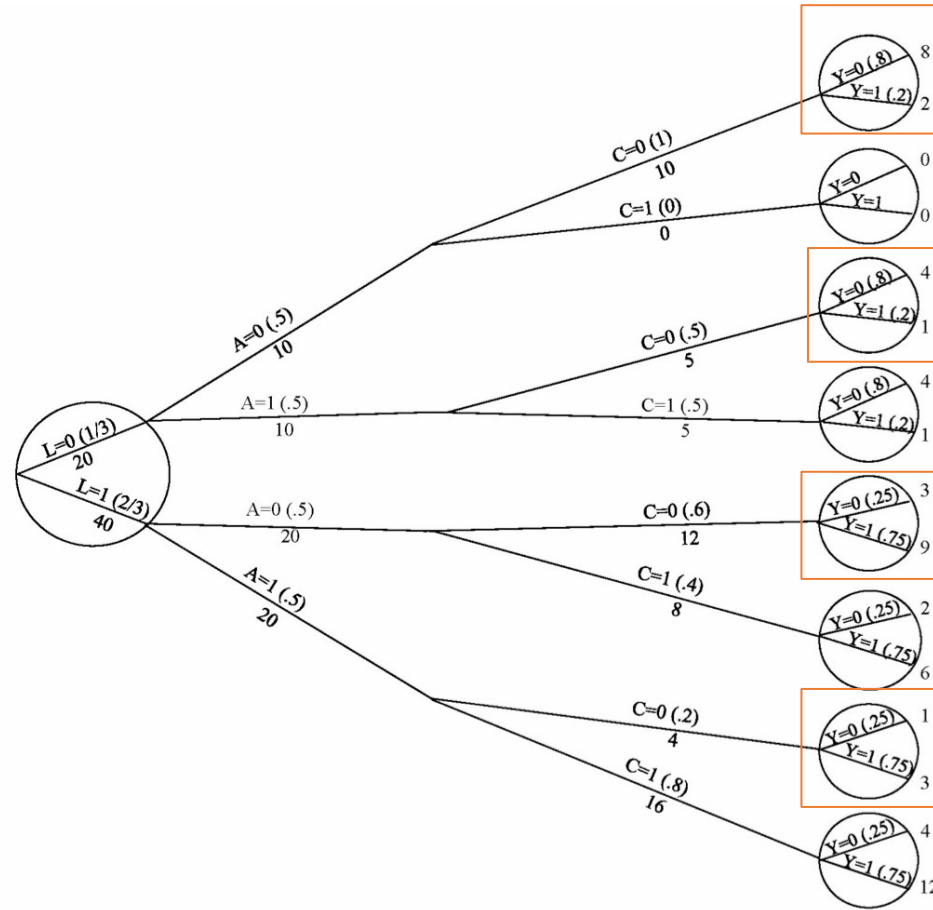
# IPW in selection bias adjustment

$$\mathbb{P}\left[Y \mid do(A) = a\right] = \sum_l \mathbb{P}\left[Y \mid L = l, A = a, C = 1\right]$$

$$= \sum_l \frac{\mathbb{P}\left[Y, C = 1 \mid L = l, A = a\right]}{\mathbb{P}\left[C = 1 \mid A = a, L = l\right]} \mathbb{P}\left[L = l\right]$$

$$= \sum_l \frac{\mathbb{P}\left[Y, C = 1, A \mid L = l\right] \mathbb{P}\left[L = l\right]}{\mathbb{P}\left[A = a \mid L = l\right] \mathbb{P}\left[C = 1 \mid A = a, L = l\right]}$$

$$= \sum_l W_a(l) \mathbb{P}\left[Y, C = 1, A = a, L = l\right]$$

# A completely randomized experiment example

$$W_a = \frac{1}{P[C=0 \mid A=a, L=l]}$$

$$[P(A = a \mid L = l) \equiv 0.5]$$



1

2

$\frac{5}{3}$

5

A: Wasabi intake
Y: one-year risk of death
L: heart disease
C: lost to follow-up ($C = 1$)

Figure 8.10

Risk ratio: $\dfrac{P(Y=1 \mid do(A)=1)}{P(Y=0 \mid do(A)=0)} = \dfrac{1*2+3*5}{2*1+9*5/3} = 1$