INVERSE PROBABILITY WEIGHTING FOR TIME-VARYING STRATEGIES

February 25, 2022

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

LEARNING OBJECTIVES.

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

1. Define unstabilized and stabilized inverse probability of treatment weights for time-varying treatments
2. Define marginal structural models for time-varying treatments
3. Implement these methods in R

PLAN FOR TODAY.

1. Recap
2. Data example
3. Unstabilized weights for time-varying treatments
4. Stabilized weights for time-varying treatments

RECAP: 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:

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

RECAP: SEQUENTIAL EXCHANGEABILITY.

To estimate the effect of a time-varying treatment, sequential exchangeability must hold

- Exchangeability must hold at each treatment time point, conditional on past treatment and covariate history
- No unmeasured confounders for the effect of A_k on Y for all time points k

Formally:

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

RECAP: TREATMENT-CONFOUNDER FEEDBACK.

Treatment-confounder feedback occurs if:

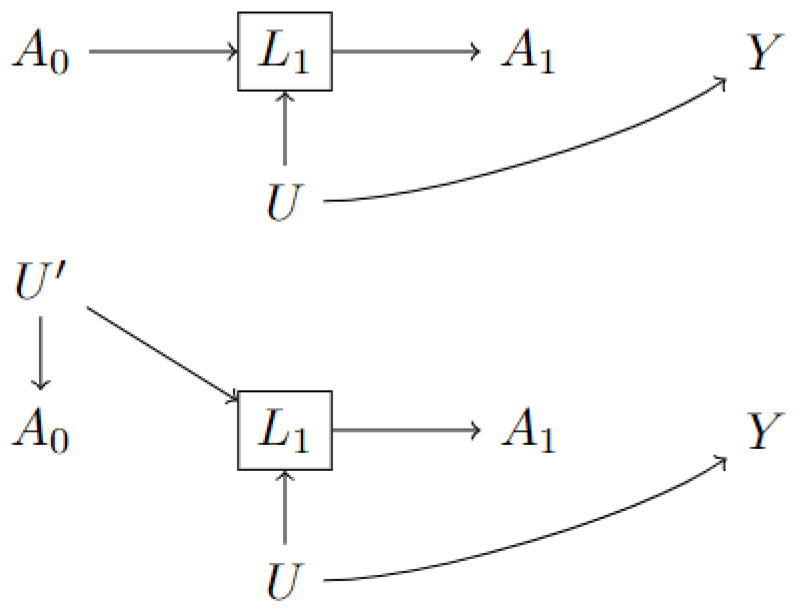
- The confounder is affected by treatment; or
- The confounder and treatment share common causes

Conditioning on L_1 will

block confounding for the effect of A_1 on Y

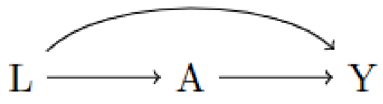
but doing so also

introduces selection bias for the effect of A_0 on Y



RECAP: INVERSE PROBABILITY WEIGHTING.

To adjust for confounding for a time-fixed treatment:



We can calculate inverse probability of treatment weights:

$$W = \frac{1}{f(A|L)}$$

For people with $A = 1$:

$$W = \frac{1}{PS} = \frac{1}{\Pr[A = 1|L]}$$

For people with $A = 0$:

$$W = \frac{1}{1 - PS} = \frac{1}{\Pr[A = 0|L]}$$

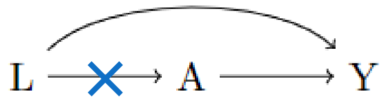
With high-dimensional data, we can estimate the denominator by using logistic regression:

$$\text{logit}\Pr[A = 1|L] = \beta_0 + \beta_1 L_1 + \beta_2 L_2 + \cdots$$

RECAP: PSEUDO-POPULATION.

The weighted population is a pseudo-population in which there is no confounding

- Weighting removes the $L \rightarrow A$ arrow from the DAG



If treatment can take on two possible values (e.g., $A = 1$ and $A = 0$):

- The size of the pseudo-population is double that of the original study population (i.e., mean of weights = 2)
- Half are assigned to $A = 0$ and half are assigned to $A = 1$
- Distribution of L is the same among those with $A = 0$ and $A = 1$

Since there is no confounding in the pseudopopulation, we use outcome regression in the weighted population without having to adjust for confounders:

$$\mathrm{E}[Y|A] = \beta_0 + \beta_1 A$$

where $\hat{\beta}_1$ is the estimate for the causal effect, $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$ (under conditional exchangeability, positivity and consistency)

RECAP: STABILIZED WEIGHTS.

The weights presented earlier are unstabilized weights:

$$W = \frac{1}{f(A|L)}$$

We could also calculate stabilized weights

$$SW = \frac{f(A)}{f(A|L)}$$

For people with $A = 1$: $SW = \frac{\Pr[A=1]}{\Pr[A=1|L]}$

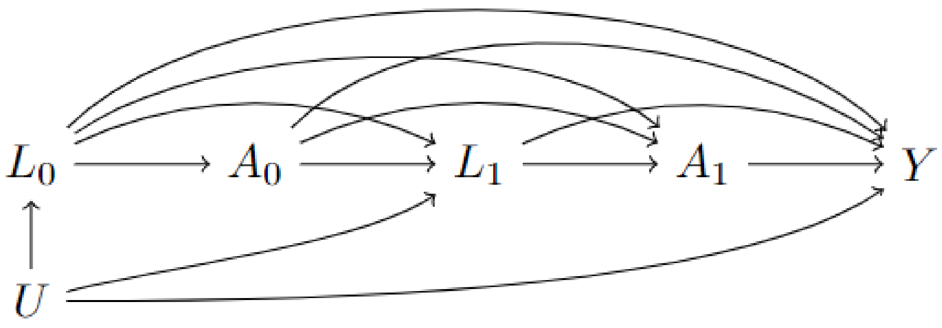
For people with $A = 0$: $SW = \frac{\Pr[A=0]}{\Pr[A=0|L]}$

- Mean of stabilized weights is 1
- People are randomly assigned to $A = 1$ or $A = 0$ according to the observed probability $\Pr[A = 1]$
- Compared to using W, using SW results in narrower 95% confidence intervals if the outcome regression model is parametric (i.e., not saturated)

PLAN FOR TODAY.

1. Recap
2. Data example
3. Unstabilized weights for time-varying treatments
4. Stabilized weights for time-varying treatments

SUPPOSE WE'RE INTERESTED IN THE EFFECT OF ANTI-DEPRESSANT USE DURING PREGNANCY ON BIRTH WEIGHT…



A_0: antidepressant use during 2nd trimester
(1: yes, 0: no)

A_1: antidepressant use during 3rd trimester
(1: yes, 0: no)

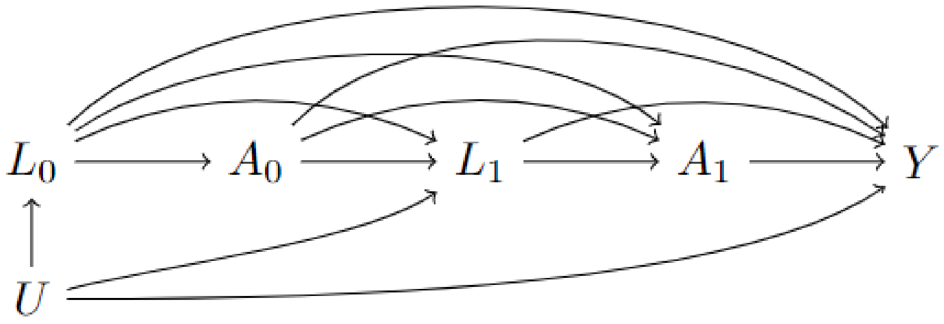
Y: birth weight

L_0: age at conception, education, urban/rural, living alone, smoking, anxiety, asthma, depressive score, hypertension

L_1: depressive score, hypertension

U: unmeasured variables

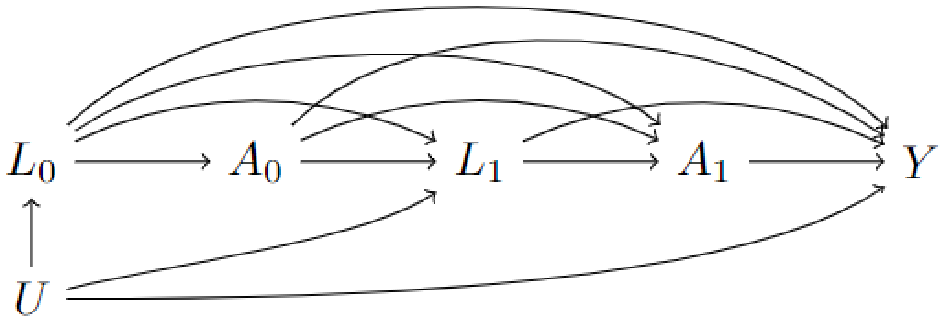
CONDITIONAL EXCHANGEABILITY FOR A_0.



Poll question: What variables do we have to condition on in order to block all backdoor paths between A_0 and Y?

A. L_0 only
B. L_0 and U
C. L_0 and L_1
D. L_0 and A_1
E. L_0, L_1 and A_1

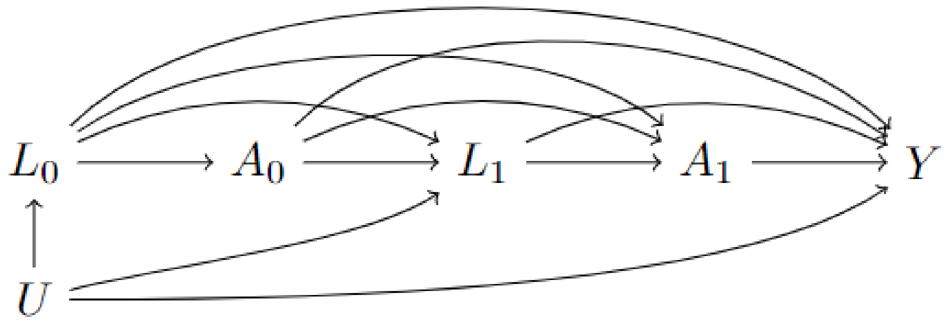
CONDITIONAL EXCHANGEABILITY FOR A_1.



Poll question: What variables do we have to condition on in order to block all backdoor paths between A_1 and Y?

A. L_1 only
B. L_1 and U
C. L_1 and L_0
D. L_1 and A_0
E. L_1, L_0 and A_0

SEQUENTIAL EXCHANGEABILITY.



To estimate the joint effects of A_0 and A_1, we need both of the following exchangeability assumptions to hold:

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

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

We can't use conventional methods. Conditioning on L_1 will:

- Block confounding for A_1, but also
- Introduce selection bias for A_0

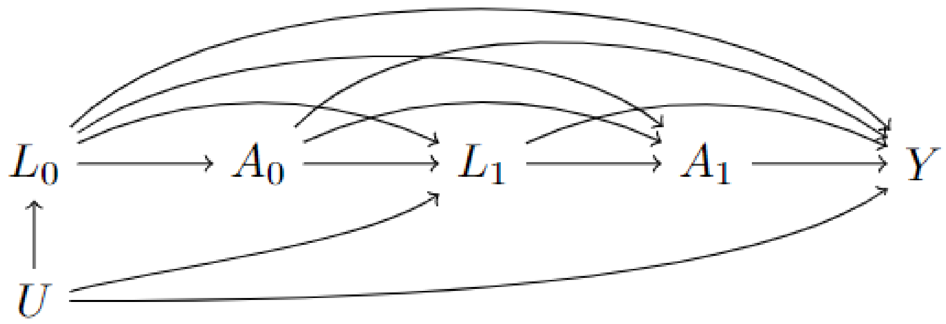
PLAN FOR TODAY.

1. Recap
2. Data example
3. Unstabilized weights for time-varying treatments
4. Stabilized weights for time-varying treatments

INVERSE PROBABILITY WEIGHTS FOR TIME-VARYING TREATMENTS.

- For time-varying treatments, we generate separate weights for treatment at each time point
  - Weights for time 0 incorporate the confounders for treatment at time 0
  - Weights for time 1 incorporate the confounders for treatment at time 1
  - etc.
- Weights address confounding by creating a pseudo-population where the confounders are balanced between the treated and untreated at each time point

DATA EXAMPLE: UNSTABILIZED WEIGHTS.



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

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

At time 0, we need to adjust for L_0:

$$W_0 = \frac{1}{f(A_0|L_0)}$$

At time 1, we need to adjust for L_1, A_0 and L_0:

$$W_1 = \frac{1}{f(A_1|L_1, A_0, L_0)}$$

We combine these weights by multiplying them together:

$$W = \frac{1}{f(A_0|L_0)} \times \frac{1}{f(A_1|L_1, A_0, L_0)}$$

DATA EXAMPLE:
UNSTABILIZED WEIGHTS NOTATION EXPLAINED.

$$W = \frac{1}{f(A_0|L_0)} \times \frac{1}{f(A_1|L_1, A_0, L_0)}$$

If someone is $A_0 = 1, A_1 = 1$:

$$W = \frac{1}{\Pr(A_0 = 1|L_0)} \times \frac{1}{\Pr(A_1 = 1|L_1, A_0, L_0)}$$

If someone is $A_0 = 0, A_1 = 0$:

$$W = \frac{1}{\Pr(A_0 = 0|L_0)} \times \frac{1}{\Pr(A_1 = 0|L_1, A_0, L_0)}$$

If someone is $A_0 = 1, A_1 = 0$:

$$W = \frac{1}{\Pr(A_0 = 1|L_0)} \times \frac{1}{\Pr(A_1 = 0|L_1, A_0, L_0)}$$

If someone is $A_0 = 0, A_1 = 1$:

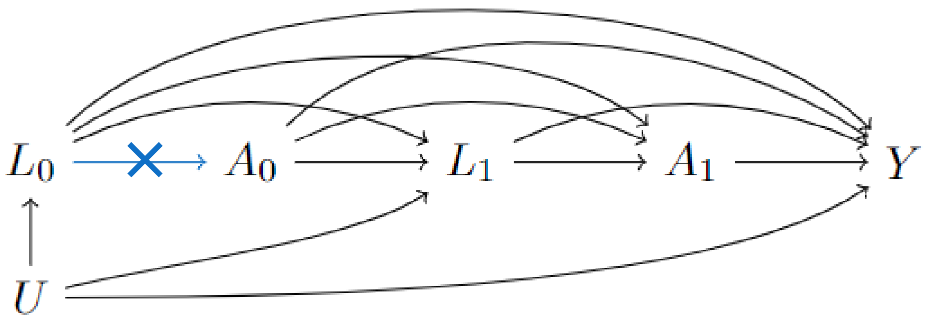
$$W = \frac{1}{\Pr(A_0 = 0|L_0)} \times \frac{1}{\Pr(A_1 = 1|L_1, A_0, L_0)}$$

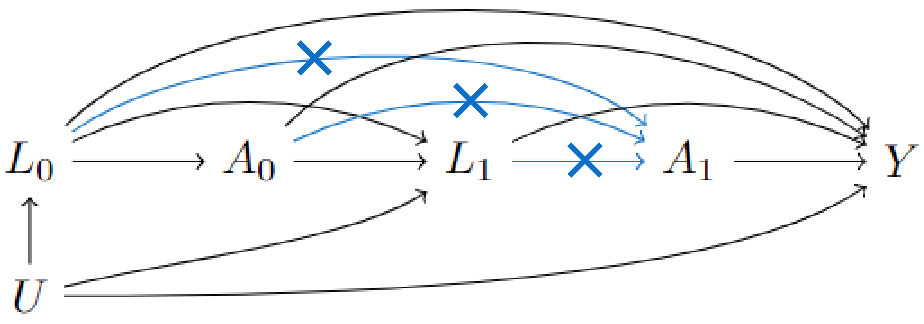
DATA EXAMPLE: PSEUDOPOPULATION WITH UNSTABILIZED WEIGHTS.

$$W = \frac{1}{f(A_0|L_0)} \times \frac{1}{f(A_1|L_1, A_0, L_0)}$$

Weights for time 0 remove the arrow $L_0 \rightarrow A_0$:

Weights for time 1 remove the arrows (1) $L_1 \rightarrow A_1$, (2) $A_0 \rightarrow A_1$, and (3) $L_0 \rightarrow A_1$:

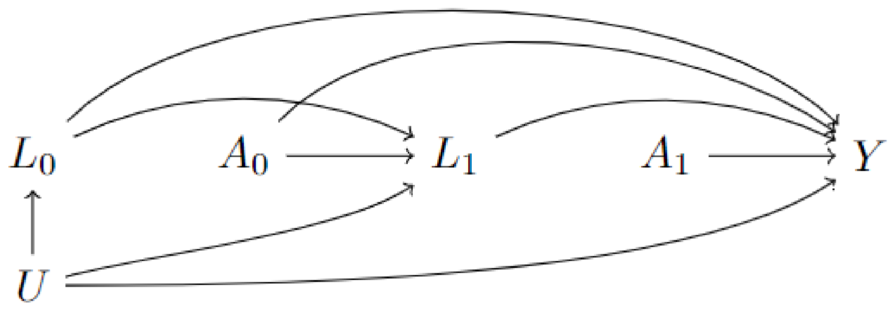




DATA EXAMPLE: PSEUDOPOPULATION WITH UNSTABILIZED WEIGHTS.

$$W = \frac{1}{f(A_0|L_0)} \times \frac{1}{f(A_1|L_1, A_0, L_0)}$$

When the weights are multiplied, this is what the DAG would look like in the pseudopopulation:



No confounding for A_0 or for A_1!

- W_0 has a mean of 2
- W_1 has a mean of 2
- W has a mean of $2 \times 2 = 4$
- i.e., the pseudopopulation is 4 times as large as the original study population because there are four possible treatment strategies:

  $(a_0 = 1, a_1 = 1)$ $(a_0 = 0, a_1 = 0)$
  $(a_0 = 1, a_1 = 0)$ $(a_0 = 0, a_1 = 1)$

DATA EXAMPLE: ESTIMATING UNSTABILIZED WEIGHTS.

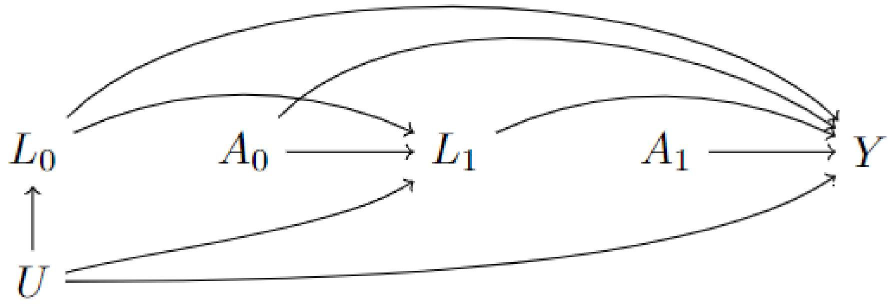
$$W = \frac{1}{f(A_0|L_0)} \times \frac{1}{f(A_1|L_1, A_0, L_0)}$$

To estimate the denominator these weights, we fit two logistic regression models.

For time 0: $\operatorname{logit}\Pr[A_0 = 1|L_0] = \beta_0 + \beta_1 L_0$

For time 1: $\operatorname{logit}\Pr[A_1 = 1|L_1, A_0, L_0] = \beta_0 + \beta_1 L_1 + \beta_2 A_0 + \beta_3 L_0$

DATA EXAMPLE: SATURATED OUTCOME MODEL.



Since there is no confounding in the pseudopopulation, we can identify the mean outcome in each of the four strata:

$$(A_0 = 1, A_1 = 1) \quad (A_0 = 0, A_1 = 0)$$
$$(A_0 = 1, A_1 = 0) \quad (A_0 = 0, A_1 = 1)$$

Or, we could fit a saturated/non-parametric outcome model:

$$E[Y|A_0, A_1] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 A_0 A_1$$

Poll question 1: Which of the following corresponds to $\mathrm{E}[Y^{a_0=1,a_1=0} - Y^{a_0=0,a_1=0}]$?

A. β_0

B. β_1

C. β_2

D. β_3

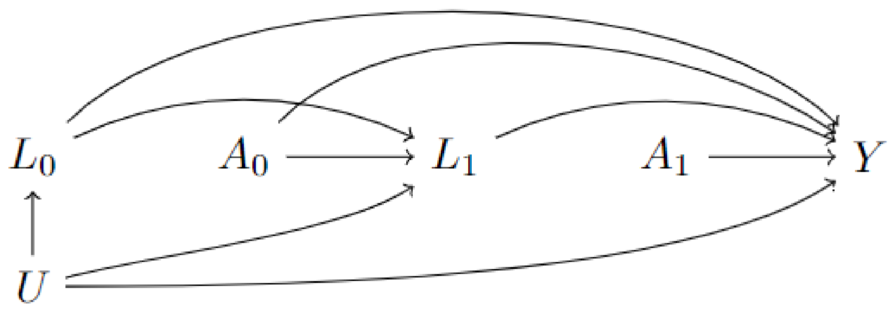
E. $\beta_1 + \beta_2$

F. $\beta_1 + \beta_3$

G. $\beta_2 + \beta_3$

H. $\beta_1 + \beta_2 + \beta_3$

DATA EXAMPLE: SATURATED OUTCOME MODEL.



Since there is no confounding in the pseudopopulation, we can identify the mean outcome in each of the four strata:

$$(A_0 = 1, A_1 = 1) \quad (A_0 = 0, A_1 = 0)$$
$$(A_0 = 1, A_1 = 0) \quad (A_0 = 0, A_1 = 1)$$

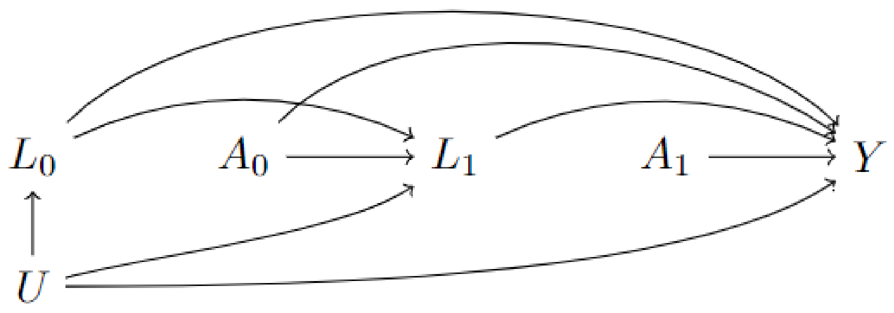
Or, we could fit a saturated/non-parametric outcome model:

$$E[Y|A_0, A_1] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 A_0 A_1$$

Poll question 2: Which of the following corresponds to $\mathrm{E}[Y^{a_0=1,a_1=1} - Y^{a_0=1,a_1=0}]$?

A. β_0
B. β_1
C. β_2
D. β_3
E. $\beta_1 + \beta_2$
F. $\beta_1 + \beta_3$
G. $\beta_2 + \beta_3$
H. $\beta_1 + \beta_2 + \beta_3$

DATA EXAMPLE: SATURATED OUTCOME MODEL.



Since there is no confounding in the pseudopopulation, we can identify the mean outcome in each of the four strata:

$(A_0 = 1, A_1 = 1)$ $(A_0 = 0, A_1 = 0)$
$(A_0 = 1, A_1 = 0)$ $(A_0 = 0, A_1 = 1)$

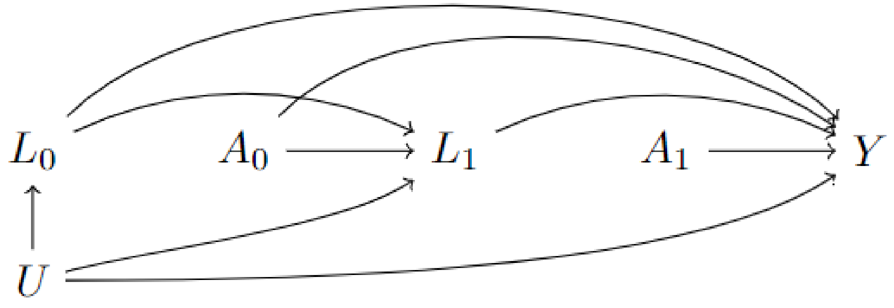
Or, we could fit a saturated/non-parametric outcome model:

$$E[Y|A_0, A_1] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 A_0 A_1$$

Poll question 3: Which of the following corresponds to $\mathrm{E}[Y^{a_0=1,a_1=1} - Y^{a_0=0,a_1=0}]$?

A. β_0
B. β_1
C. β_2
D. β_3
E. $\beta_1 + \beta_2$
F. $\beta_1 + \beta_3$
G. $\beta_2 + \beta_3$
H. $\beta_1 + \beta_2 + \beta_3$

DATA EXAMPLE: UNSATURATED OUTCOME MODEL.



If we think there is no interaction between A_0 and A_1, we could remove the interaction term.

We could fit an unsaturated/parametric outcome model:

$$E[Y|A_0, A_1] = \beta_0 + \beta_1 A_0 + \beta_2 A_1$$

Poll question 1: Which of the following corresponds to $\mathrm{E}[Y^{a_0=1,a_1=0} - Y^{a_0=0,a_1=0}]$?

A. β_0

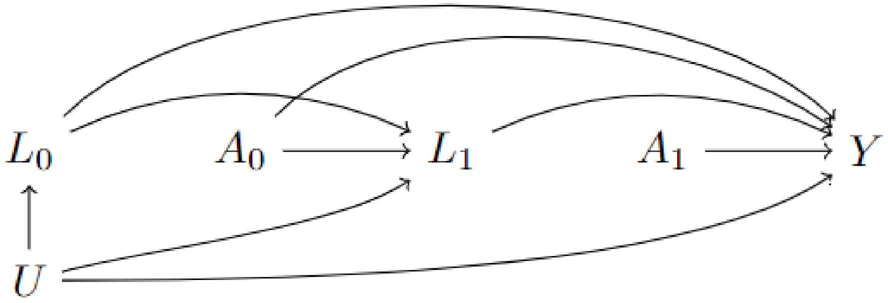
B. β_1

C. β_2

D. $\beta_1 + \beta_2$

E. $\beta_0 + \beta_1 + \beta_3$

DATA EXAMPLE: UNSATURATED OUTCOME MODEL.



If we think there is no interaction between A_0 and A_1, we could remove the interaction term.

We could fit an unsaturated/parametric outcome model:

$$E[Y|A_0, A_1] = \beta_0 + \beta_1 A_0 + \beta_2 A_1$$

Poll question 2: Which of the following corresponds to $\mathrm{E}[Y^{a_0=1,a_1=1} - Y^{a_0=1,a_1=0}]$?

A. β_0

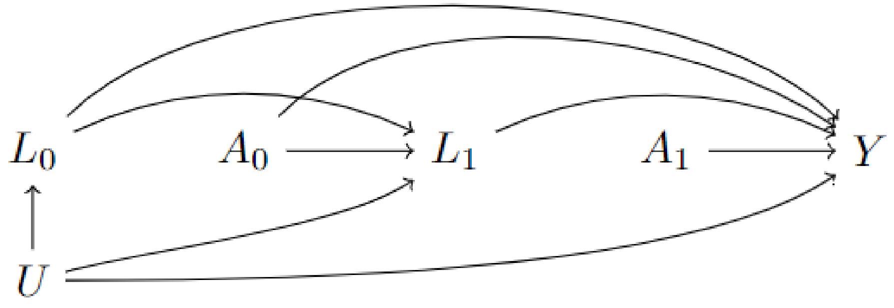
B. β_1

C. β_2

D. $\beta_1 + \beta_2$

E. $\beta_0 + \beta_1 + \beta_3$

DATA EXAMPLE: UNSATURATED OUTCOME MODEL.



If we think there is no interaction between A_0 and A_1, we could remove the interaction term.

We could fit an unsaturated/parametric outcome model:

$$E[Y|A_0, A_1] = \beta_0 + \beta_1 A_0 + \beta_2 A_1$$

Poll question 3: Which of the following corresponds to $\mathrm{E}[Y^{a_0=1,a_1=1} - Y^{a_0=0,a_1=0}]$?

A. β_0

B. β_1

C. β_2

D. $\beta_1 + \beta_2$

E. $\beta_0 + \beta_1 + \beta_3$

ROBUST STANDARD ERRORS.

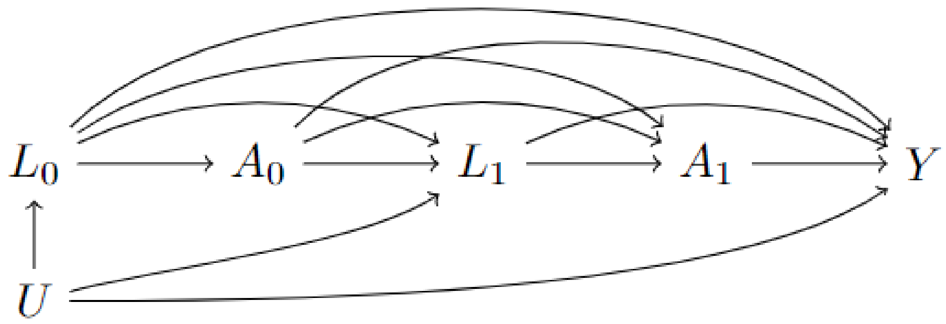
Regardless of which outcome model we fit, we need to account for the weighting process in our calculation of the variance.

Need to estimate robust standard errors.

PLAN FOR TODAY.

1. Recap
2. Data example
3. Unstabilized weights for time-varying treatments
4. Stabilized weights for time-varying treatments

DATA EXAMPLE: STABILIZED WEIGHTS.



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

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

Recall that our unstabilized weights were:

$$W = \frac{1}{f(A_0|L_0)} \times \frac{1}{f(A_1|L_1, A_0, L_0)}$$

For stabilized weights, we include functions of the treatment (and only the treatment) in the numerator:

$$SW = \frac{f(A_0)}{f(A_0|L_0)} \times \frac{f(A_1|A_0)}{f(A_1|L_1, A_0, L_0)}$$

DATA EXAMPLE:
STABILIZED WEIGHTS NOTATION EXPLAINED.

$$SW = \frac{f(A_0)}{f(A_0|L_0)} \times \frac{f(A_1|A_0)}{f(A_1|L_1, A_0, L_0)}$$

If someone is $A_0 = 1, A_1 = 1$:

$$SW = \frac{\Pr(A_0 = 1)}{\Pr(A_0 = 1|L_0)} \times \frac{\Pr(A_1 = 1|A_0)}{\Pr(A_1 = 1|L_1, A_0, L_0)}$$

If someone is $A_0 = 0, A_1 = 0$:

$$SW = \frac{\Pr(A_0 = 0)}{\Pr(A_0 = 0|L_0)} \times \frac{\Pr(A_1 = 0|A_0)}{\Pr(A_1 = 0|L_1, A_0, L_0)}$$

If someone is $A_0 = 1, A_1 = 0$:

$$SW = \frac{\Pr(A_0 = 1)}{\Pr(A_0 = 1|L_0)} \times \frac{\Pr(A_1 = 0|A_0)}{\Pr(A_1 = 0|L_1, A_0, L_0)}$$

If someone is $A_0 = 0, A_1 = 1$:

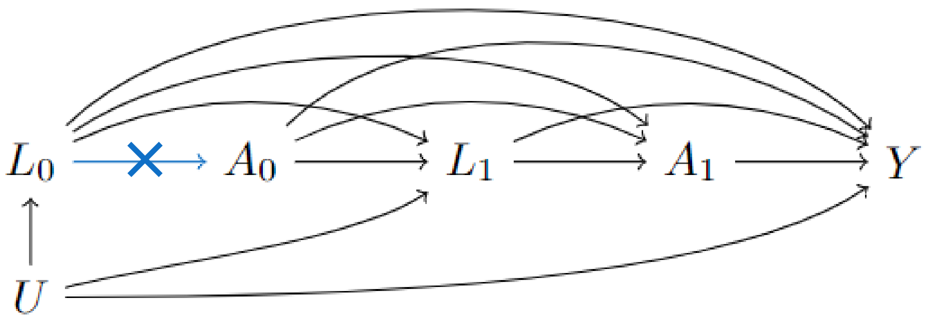
$$SW = \frac{\Pr(A_0 = 0)}{\Pr(A_0 = 0|L_0)} \times \frac{\Pr(A_1 = 1|A_0)}{\Pr(A_1 = 1|L_1, A_0, L_0)}$$

DATA EXAMPLE: PSEUDOPOPULATION WITH STABILIZED WEIGHTS.

$$SW = \frac{f(A_0)}{f(A_0|L_0)} \times \frac{f(A_1|A_0)}{f(A_1|L_1, A_0, L_0)}$$

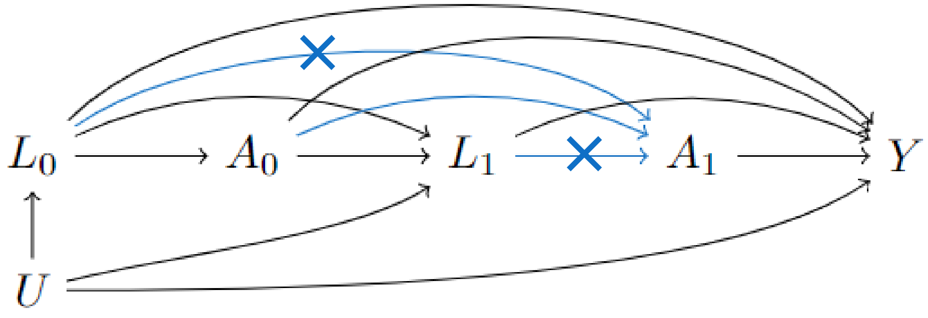
Conditioning in the denominator removes arrows. Conditioning in the numerator adds arrows.

Weights for time 0 remove the arrow $L_0 \rightarrow A_0$:



Weights for time 1 remove the arrows (1) $L_1 \rightarrow A_1$, (2) $A_0 \rightarrow A_1$, and (3) $L_0 \rightarrow A_1$

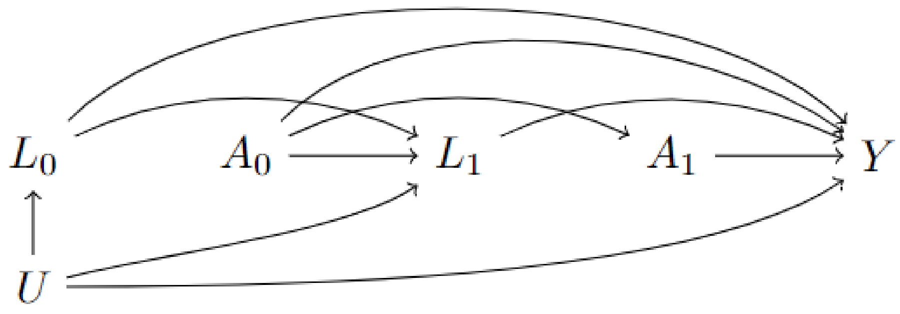
BUT the numerator adds the arrow $A_0 \rightarrow A_1$ back:



DATA EXAMPLE: PSEUDOPOPULATION WITH STABILIZED WEIGHTS.

$$SW = \frac{f(A_0)}{f(A_0|L_0)} \times \frac{f(A_1|A_0)}{f(A_1|L_1, A_0, L_0)}$$

When the weights are multiplied, this is what the DAG would look like in the pseudopopulation:



No confounding for A_0

A_1 is confounded by A_0 (but that's OK because we condition on A_0 in the outcome model)

- SW_0 has a mean of 1
- SW_1 has a mean of 1
- SW has a mean of $1 \times 1 = 1$

DATA EXAMPLE: ESTIMATING STABILIZED WEIGHTS.

$$SW = \frac{f(A_0)}{f(A_0|L_0)} \times \frac{f(A_1|A_0)}{f(A_1|L_1, A_0, L_0)}$$

Recall, we fit the following logistic regression models for the denominator of the weights:

For time 0: $\text{logit}\Pr[A_0 = 1|L_0] = \beta_0 + \beta_1 L_0$

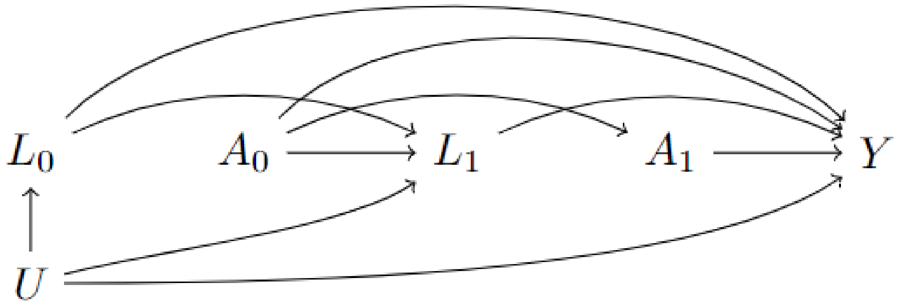
For time 1: $\text{logit}\Pr[A_1 = 1|L_1, A_0, L_0] = \beta_0 + \beta_1 L_1 + \beta_2 A_0 + \beta_3 L_0$

We also need to fit models for the numerator of the weights:

For time 0: $\text{logit}\Pr[A_0 = 1] = \beta_0$

For time 1: $\text{logit}\Pr[A_1 = 1|A_0] = \beta_0 + \beta_1 A_0$

DATA EXAMPLE: OUTCOME MODEL.



The outcome model is the same, regardless of whether we used stabilized or unstabilized weights:

Saturated/non-parametric outcome model:

$$E[Y|A_0, A_1] = \beta_0 + \beta_1 A_0 + \beta_2 A_1 + \beta_3 A_0 A_1$$

Unsaturated/parametric outcome model:

$$E[Y|A_0, A_1] = \beta_0 + \beta_1 A_0 + \beta_2 A_1$$

Why use stabilized weights?

They are more efficient (i.e., tighter 95% confidence intervals) when the outcome model is unsaturated/parametric.

LEARNING OBJECTIVES.

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

1. Define unstabilized and stabilized inverse probability of treatment weights for time-varying treatments
2. Define marginal structural models for time-varying treatments
3. Implement these methods in R