Sequential Causal Inference

Robin J. Evans and Vanessa Didelez

University of Oxford and BIPS, University of Bremen

APTS Week 4, Oxford
September 2024

Outline

1. Graphical Sequential Treatment
   - Multiple Regression
   - Multiple Treatments

Regression

Question

Suppose we perform a regression of an outcome Y on **several** other variables. Under what circumstances are the coefficients estimating a **causal** quantity?

If they are, then what quantity is it, exactly?

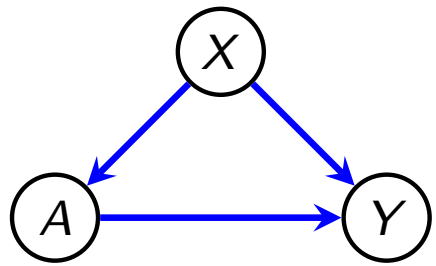
Regression

Question

Suppose we perform a regression of an outcome Y on **several** other variables. Under what circumstances are the coefficients estimating a **causal** quantity?

If they are, then what quantity is it, exactly?

If we compute
$\mathbb{E}[Y \mid A, X] = \beta_{AY \cdot X} A + \beta_{XY \cdot A} X$
under this graph, will we have:



$$p(y \mid do(a)) = \beta_{AY \cdot X} a \, ?$$

$$p(y \mid do(x)) = \beta_{XY \cdot A} x \, ?$$

$\beta_{AY \cdot \boldsymbol{C}}$ denotes the regression coefficient for A on Y in the model that includes $\boldsymbol{C}$.

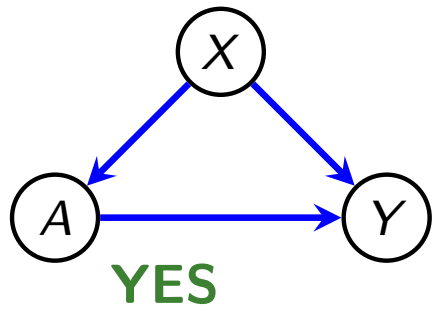
Regression

Question

Suppose we perform a regression of an outcome Y on **several** other variables. Under what circumstances are the coefficients estimating a **causal** quantity?

If they are, then what quantity is it, exactly?

If we compute
$\mathbb{E}[Y \mid A, X] = \beta_{AY \cdot X} A + \beta_{XY \cdot A} X$
under this graph, will we have:



$$p(y \mid do(a)) = \beta_{AY \cdot X} a \, ?$$

$$p(y \mid do(x)) = \beta_{XY \cdot A} x \, ?$$

$\beta_{AY \cdot \boldsymbol{C}}$ denotes the regression coefficient for A on Y in the model that includes $\boldsymbol{C}$.

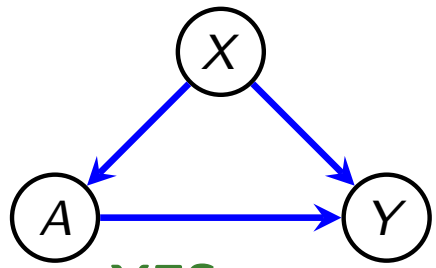
Regression

Question

Suppose we perform a regression of an outcome Y on **several** other variables. Under what circumstances are the coefficients estimating a **causal** quantity?

If they are, then what quantity is it, exactly?

If we compute
$\mathbb{E}[Y \mid A, X] = \beta_{AY \cdot X} A + \beta_{XY \cdot A} X$
under this graph, will we have:



$$p(y \mid do(a)) = \beta_{AY \cdot X} a \, ?$$

YES

$$p(y \mid do(x)) = \beta_{XY \cdot A} x \, ?$$

NO

$\beta_{XY \cdot A}$ **is** a causal effect, but it is the effect of X on Y when keeping A constant; this is called the **controlled direct effect**.

$\beta_{AY \cdot C}$ denotes the regression coefficient for A on Y in the model that includes $\boldsymbol{C}$.

Table 2 Fallacy

JOURNAL ARTICLE

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients

Daniel Westreich ✉, Sander Greenland Author Notes

American Journal of Epidemiology, Volume 177, Issue 4, 15 February 2013, Pages 292–298,
https://doi.org/10.1093/aje/kws412
Published: 30 January 2013 **Article history** ▾

Abstract

It is common to present multiple adjusted effect estimates from a single model in a single table. For example, a table might show odds ratios for one or more exposures and also for several confounders from a single logistic regression. This can lead to mistaken interpretations of these estimates. We use causal diagrams to display the sources of the problems. Presentation of exposure and confounder effect estimates from a single model may lead to several interpretative difficulties, inviting confusion of direct-effect estimates with total-effect estimates for covariates in the model. These effect estimates may also be confounded even though the effect estimate for the main exposure is not confounded. Interpretation of these effect estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. We offer suggestions to limit potential misunderstandings when multiple effect estimates are presented, including precise distinction between total and direct effect measures from a single model, and use of multiple models tailored to yield total-effect estimates for covariates.

Multiple Exposures

Often in applications not a clear distinction between specific exposure and covariates used for adjustment.

There may well be multiple questions of interest: e.g. what is effect of smoking, diet, alcohol intake all **together**.

Be clear about causal question relating to multiple exposures: what would be your **ideal target trial**?

Multiple Exposures

Often in applications not a clear distinction between specific exposure and covariates used for adjustment.

There may well be multiple questions of interest: e.g. what is effect of smoking, diet, alcohol intake all **together**.

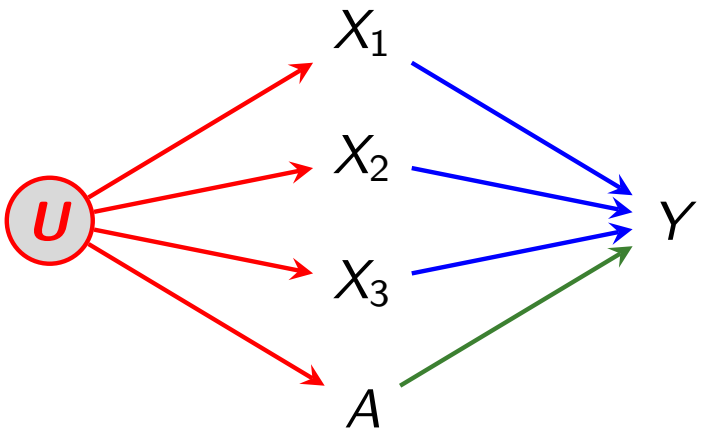
Be clear about causal question relating to multiple exposures: what would be your **ideal target trial**?

Many possible causal effects to define with multiple exposures:

- separate total effects;
- joint intervention effects;
- controlled direct effects;
- strategy for dynamic treatment effects;
- separable (or natural) direct and indirect effects.

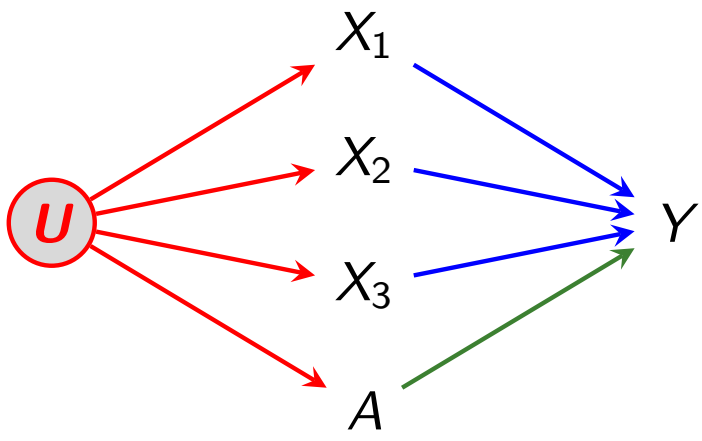
Confounding Due to Common History

Suppose the causal structure is more like below—what should we adjust for?



Confounding Due to Common History

Suppose the causal structure is more like below—what should we adjust for?



Answer

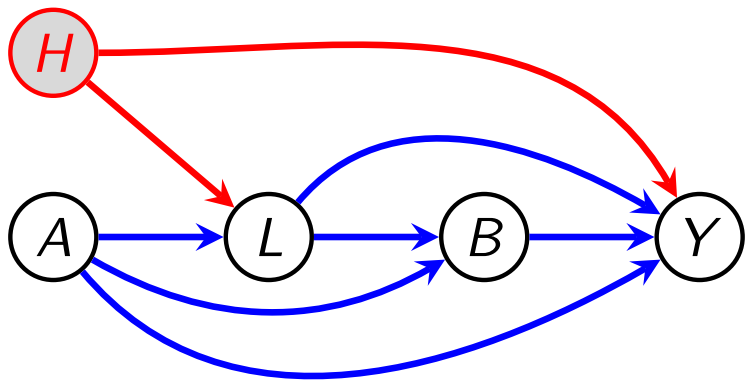
Condition on X_1, X_2, **and** X_3!

This is the only way to block all back-door paths from A to Y.

Joint Interventions

A **joint intervention** considers what would happen if we intervene on **multiple variables** at the same time.

For example, suppose we have the graph below:

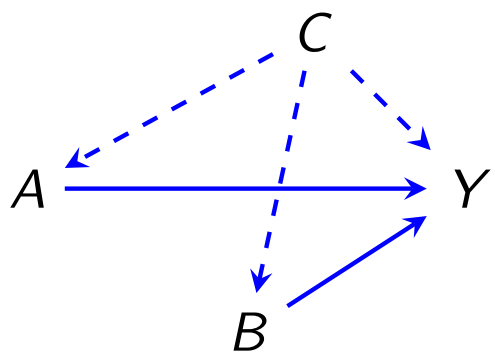


How can we identify $P(Y \mid do(A, B))$?

Examples I

What is the interpretation of β_A, β_B in the regressions?

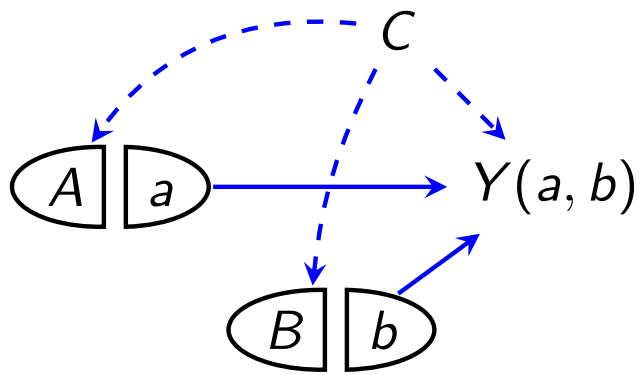
$$\mathbb{E}[Y \mid A = a, B = b, C = c] = \beta_A a + \beta_B b + \beta_C c.$$



Examples I

What is the interpretation of β_A, β_B in the regressions?

$$\mathbb{E}[Y \mid A = a, B = b, C = c] = \beta_A a + \beta_B b + \beta_C c.$$

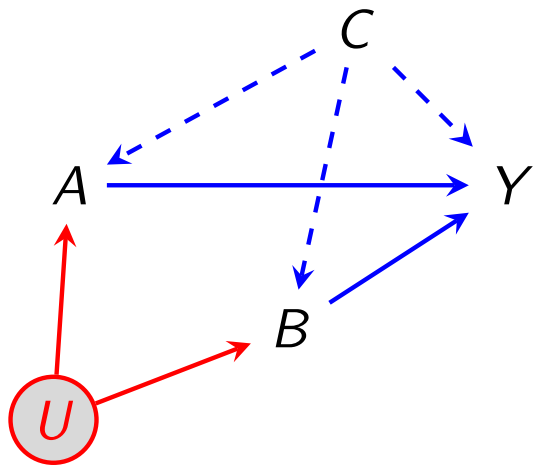


Note that $Y(a, b) \perp_d A, B \mid C$, so indeed (β_A, β_B) **are** the joint causal effects of A, B on Y.

Examples II

What is the interpretation of β_A, β_B in the regressions?

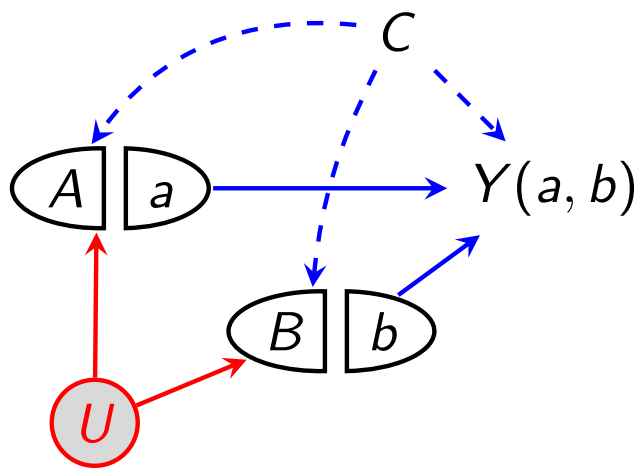
$$\mathbb{E}[Y \mid A = a, B = b, C = c] = \beta_A a + \beta_B b + \beta_C c.$$



Examples II

What is the interpretation of β_A, β_B in the regressions?

$$\mathbb{E}[Y \mid A = a, B = b, C = c] = \beta_A a + \beta_B b + \beta_C c.$$



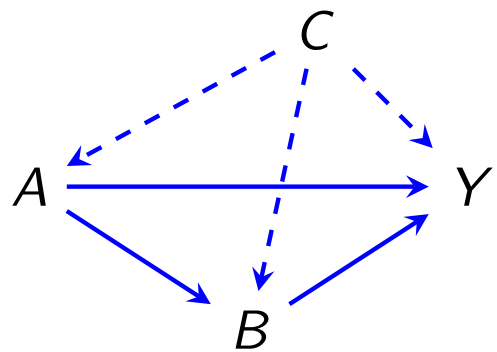
Is $Y(a, b) \perp_d A, B \mid C$?

Yes!

Examples III

What is the interpretation of β_A, β_B in the regressions?

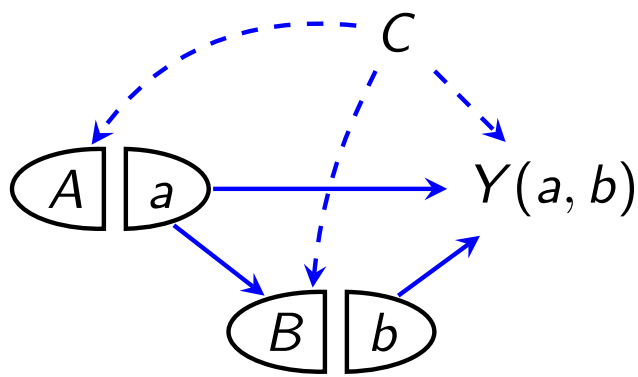
$$\mathbb{E}[Y \mid A = a, B = b, C = c] = \beta_A a + \beta_B b + \beta_C c.$$



Examples III

What is the interpretation of β_A, β_B in the regressions?

$$\mathbb{E}[Y \mid A = a, B = b, C = c] = \beta_A a + \beta_B b + \beta_C c.$$



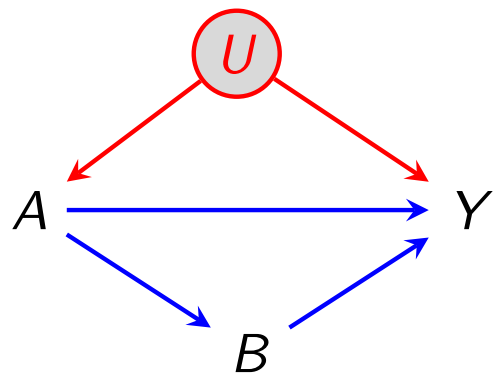
Is $Y(a, b) \perp_d A, B \mid C$?

Yes! **But:** β_A is a **controlled direct effect**, not a total effect.

Examples IV

What is the interpretation of β_A, β_B in the regression:

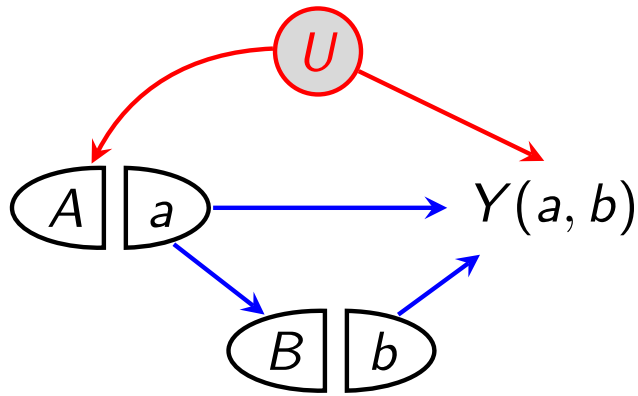
$$\mathbb{E}[Y \mid A = a, B = b] = \beta_A a + \beta_B b \quad ?$$



Examples IV

What is the interpretation of β_A, β_B in the regression:

$$\mathbb{E}[Y \mid A = a, B = b] = \beta_A a + \beta_B b \quad ?$$



Is $Y(a, b) \perp_d A, B$?

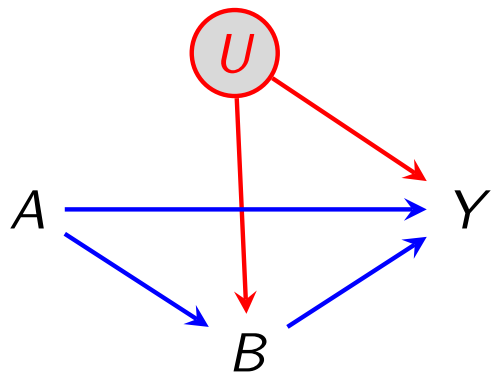
No! β_A has no causal interpretation.

However, $Y(b) \perp_d B \mid A$, so β_B **is** total effect of B on Y.

Examples V

What is the interpretation of β_A, β_B in the regression:

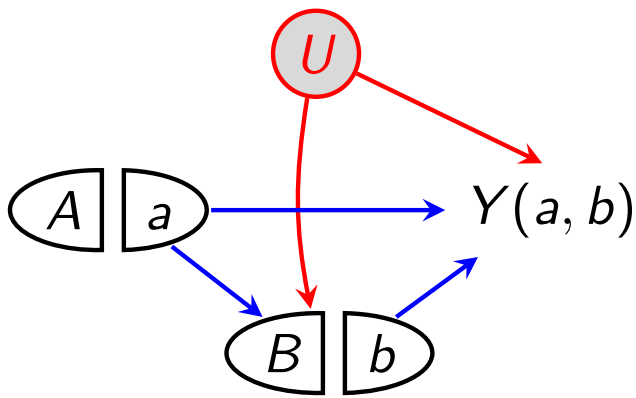
$$\mathbb{E}[Y \mid A = a, B = b] = \beta_A a + \beta_B b \quad ?$$



Examples V

What is the interpretation of β_A, β_B in the regression:

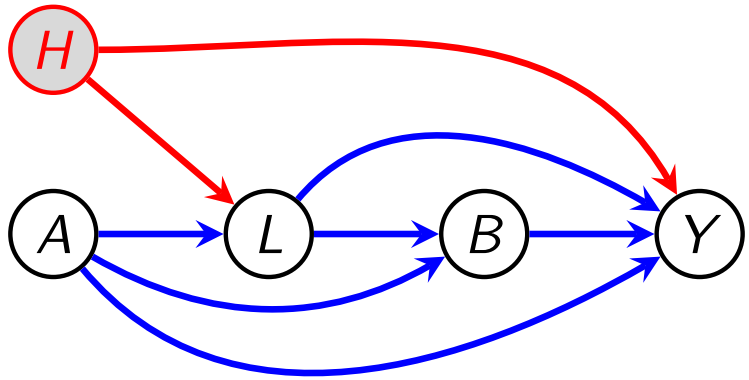
$$\mathbb{E}[Y \mid A = a, B = b] = \beta_A a + \beta_B b \quad ?$$



Is $Y(a, b) \perp_d A, B$?

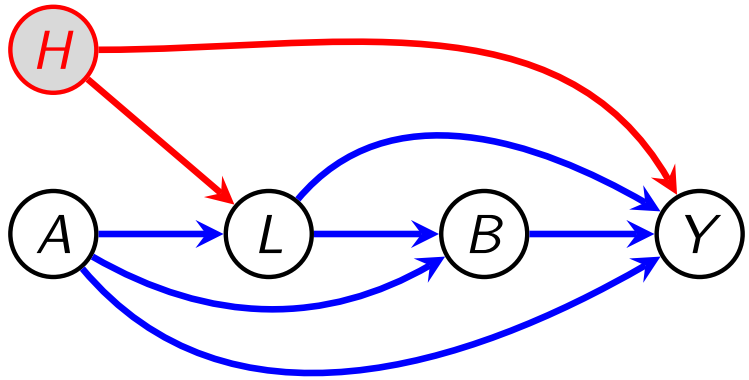
No! Neither coefficient has a causal interpretation.

Sequentially randomized experiment



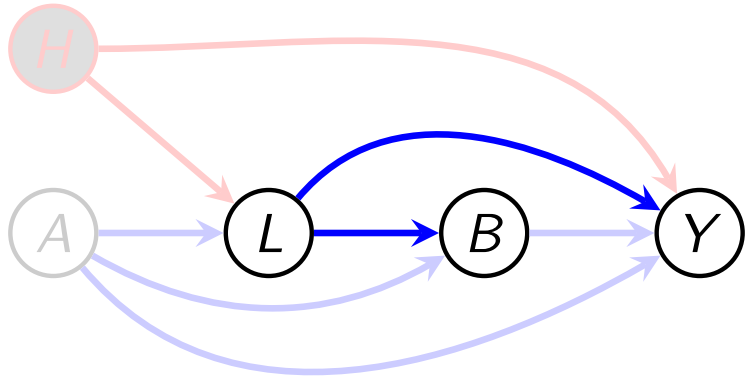
- A and B are treatments;
- H is unobserved;
- L is a time varying confounder;
- Y is the final response;
- Treatment B is assigned randomly conditional on the observed history, A and L;
- Want to know $P(Y(\tilde{a}, \tilde{b}))$.

Time-Varying Confounding



Should we adjust for L?

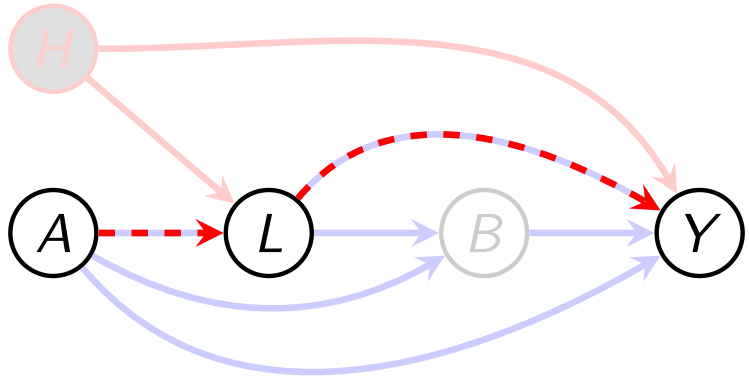
Time-Varying Confounding



Should we adjust for L?

From B's perspective, L needs to be adjusted for to control for confounding.

Time-Varying Confounding

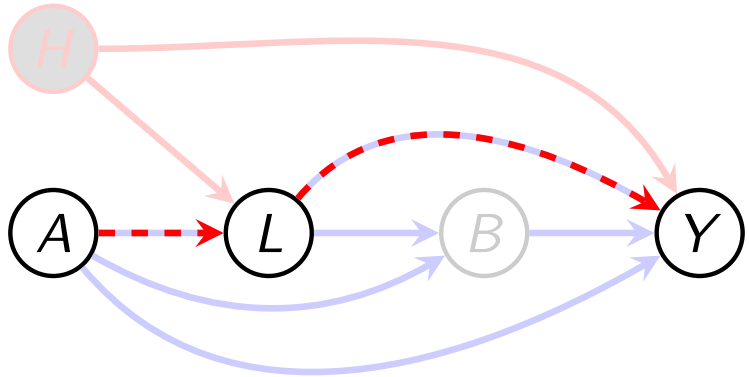


Should we adjust for L?

From B's perspective, L needs to be adjusted for to control for confounding.

From A's perspective, L is on the causal path and conditioning would open a spurious path.

Time-Varying Confounding



Should we adjust for L?

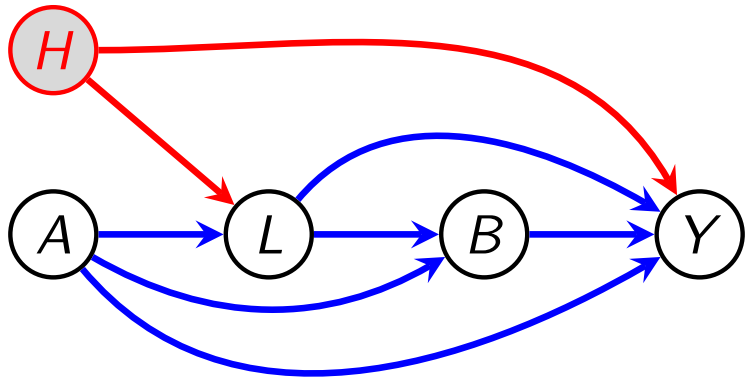
From B's perspective, L needs to be adjusted for to control for confounding.

From A's perspective, L is on the causal path and conditioning would open a spurious path.

Neither regression makes sense!

Need to **break** B's dependence on L in order to estimate the effect.

Sequentially randomized experiment



If the following holds:

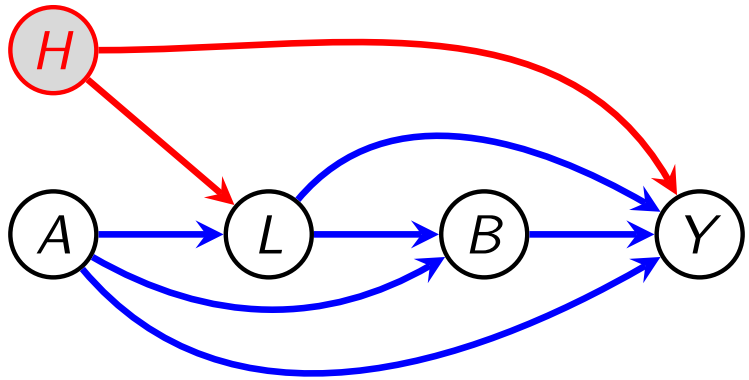
$$\begin{aligned} A &\perp\!\!\!\perp Y(a,b) \\ B(a) &\perp\!\!\!\perp Y(a,b) \mid L(a), A \end{aligned}$$

General result of Robins (1986) then implies:

$$P(Y(a,b)=y) = \sum_l P(L=l \mid A=a) \cdot P(Y=y \mid A=a, L=l, B=b).$$

Exercise: can you show this?

Sequentially randomized experiment



If the following holds:

$$\begin{aligned} A &\perp\!\!\!\perp Y(a,b) \\ B(a) &\perp\!\!\!\perp Y(a,b) \mid L(a), A \end{aligned}$$

General result of Robins (1986) then implies:

$$P(Y(a,b)=y) = \sum_l P(L=l \mid A=a) \cdot P(Y=y \mid A=a, L=l, B=b).$$

Exercise: can you show this?

Do the independences hold?

Marginal Structural Models

Models of the quantity

$$P(Y \mid do(A, B)) = \sum_{l} P(Y \mid A, L = l, B) \cdot P(L = l \mid A)$$

are called **marginal structural models** (Robins et al., 2000).

They have various nice properties:

- can be modelled semi-parametrically (so no need to fully specify rest of the distribution)
- either the propensity scores ($P(A)$ and $P(B \mid A, L)$) **or** the outcome models $P(Y \mid A, L, B)$ and $P(L \mid A)$ can be used to identify parameters.
- there is also a doubly robust version!

Marginal Structural Models

Models of the quantity

$$P(Y \mid do(A, B)) = \sum_{l} P(Y \mid A, L = l, B) \cdot P(L = l \mid A)$$

are called **marginal structural models** (Robins et al., 2000).

They have various nice properties:

- can be modelled semi-parametrically (so no need to fully specify rest of the distribution)
- either the propensity scores ($P(A)$ and $P(B \mid A, L)$) **or** the outcome models $P(Y \mid A, L, B)$ and $P(L \mid A)$ can be used to identify parameters.
- there is also a doubly robust version!

Hard to simulate from, though Evans and Didelez (2024) provide a solution!

IPW Identification

The model suffers from time-dependent confounding, so weights must reflect this.

Idea of IPW is to obtain the g-formula by 'removing' pieces not in it. So:

$$P(A, L, B, Y) = P(A) \cdot P(L \mid A) \cdot P(B \mid A, L) \cdot P(Y \mid A, L, B)$$

IPW Identification

The model suffers from time-dependent confounding, so weights must reflect this.

Idea of IPW is to obtain the g-formula by 'removing' pieces not in it. So:

$$P(A, L, B, Y) = \underbrace{P(A)}_{\text{remove}} \cdot P(L \mid A) \cdot \underbrace{P(B \mid A, L)}_{\text{remove}} \cdot P(Y \mid A, L, B)$$

So create a **pseudo-population** by fitting models for $P(A)$ and $P(B \mid A, L)$ and then reweight ith observation by

$$w_i = \frac{1}{\widehat{P}(A = a_i)} \cdot \frac{1}{\widehat{P}(B = b_i \mid A = a_i, L = l_i)}$$

Common to use logistic regression if A, B are binary.

IPW Remarks

Conditions are similar to single treatment case; see Robins et al. (2000) for details.

General form of weights with treatments $A_0, \ldots, A_K$ and covariates $L_0, \ldots, L_K$ is

$$\widehat{w}_i = \prod_{k=0}^{K} \frac{1}{\widehat{P}(A_k \mid \overline{A}_{k-1}, \overline{L}_k)},$$

where $\overline{A}_j = (A_0, \ldots, A_j)$.

IPW Remarks

Conditions are similar to single treatment case; see Robins et al. (2000) for details.

General form of weights with treatments $A_0, \ldots, A_K$ and covariates $L_0, \ldots, L_K$ is

$$\widehat{w}_i = \prod_{k=0}^{K} \frac{1}{\widehat{P}(A_k \mid \overline{A}_{k-1}, \overline{L}_k)},$$

where $\overline{A}_j = (A_0, \ldots, A_j)$.

- Models for $A_k | \overline{A}_{k-1}, \overline{L}_k$ must be correctly specified;
- in practice people use stabilised weights;
- can also estimate optimal dynamic treatments using *Q-learning* (Chakraborty and Moodie, 2013).

References

Chakraborty, B. and Moodie, E.M. Statistical Methods for Dynamic Treatment Regimes. *Reinforcement Learning, Causal Inference, and Personalized Medicine*. Springer, 2013.

Evans, R.J. and Didelez, V. Parameterizing and simulating from causal models (with discussion). *JRSS-B*, 2023.

Robins, J.M., Hernan, M.A. and Brumback, B. Marginal structural models and causal inference in epidemiology. *Epidemiology*, pp.550-560, 2000.