ETH zürich

Instrumental variables, transportability and counterfactuals

Causality

Christina Heinze-Deml

Spring 2019

Last week

- Covariate adjustment
- Frontdoor criterion

ETH zürich

Property of jointly Normal random variables

- Let X and Y be jointly Normal random variables. The conditional expectation of X given Y satisfies:

$$E(X|Y) = E(X) + \rho\frac{\sigma_X}{\sigma_Y}(Y - E(Y))$$

$$\mathrm{Var}(X|Y) = (1-\rho^2)\sigma_X^2$$

where $\rho = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$

Today

- Instrumental variables
- Transportability
- Counterfactuals

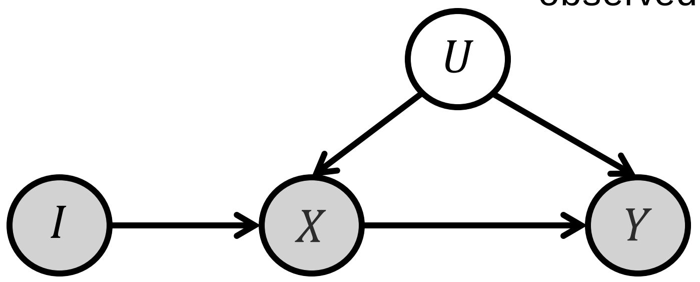
Identification strategies

- Interventional distribution is identifiable if it can be computed from the observational distribution and the graph structure
  - If there is a valid adjustment set for (X, Y), $p(y|do(x))$ is identifiable
  - Other means:
    - Frontdoor criterion
    - Instrumental variables

Instrumental variables

- X, Y and I observed, U unobserved
- Interested in causal effect of X on Y
  - Assume randomized experiment is not feasible
  - Cannot use covariate adjustment
  - Cannot use frontdoor criterion
  - Can use instrumental variable I

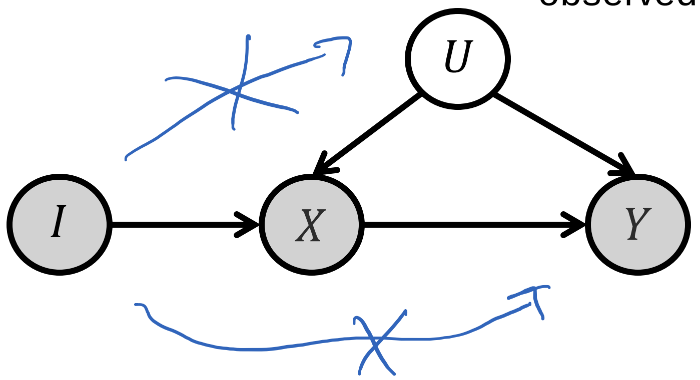
Shaded nodes are observed



Instrumental variables

- X, Y and I observed, U unobserved
- Interested in causal effect of X on Y

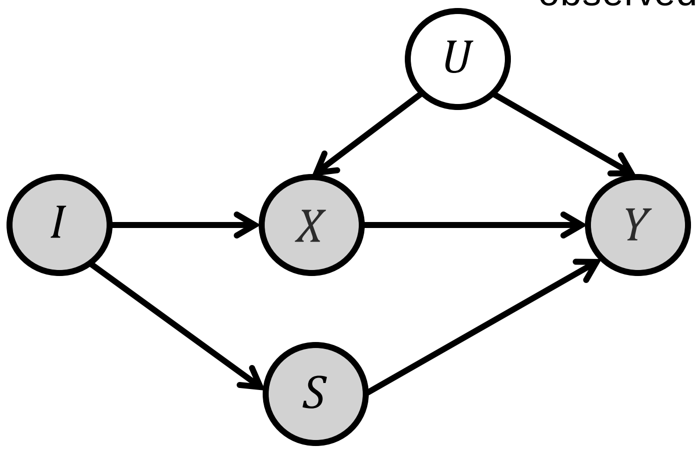
Shaded nodes are observed



- Conditions for instrumental variable (IV) I
  i. I affects X
  ii. $I \perp\!\!\!\perp U$
  iii. I affects Y only through X

Conditional instrumental variables

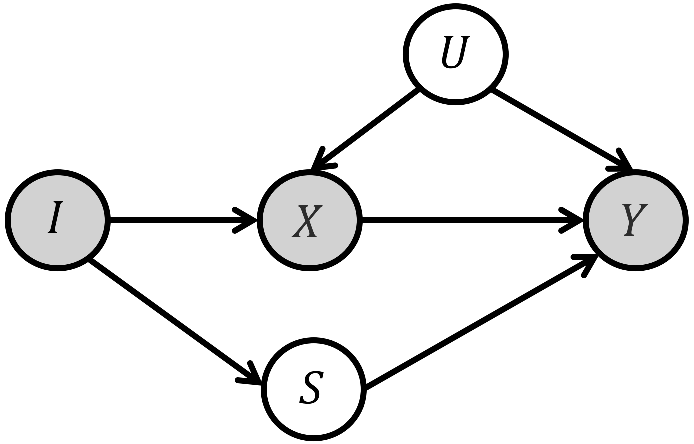
- X, Y, S and I observed, U unobserved
- Interested in causal effect of X on Y



- Conditions for (conditional) instrumental variable I not independent
  i. $I \not\perp\!\!\!\perp X | S$
  ii. Every path from I to Y that is not blocked by S has an arrow pointing into X: $I \perp\!\!\!\perp Y | S, do(X)$
     - Implies $I \perp\!\!\!\perp U | S$
     - I affects Y only through X once we control for S

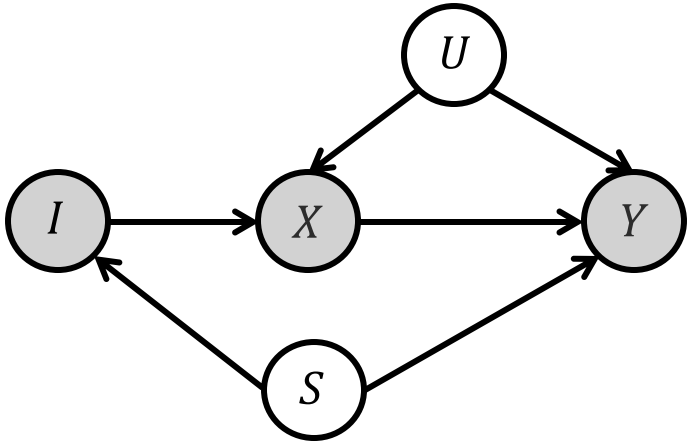
Example

Shaded nodes are observed



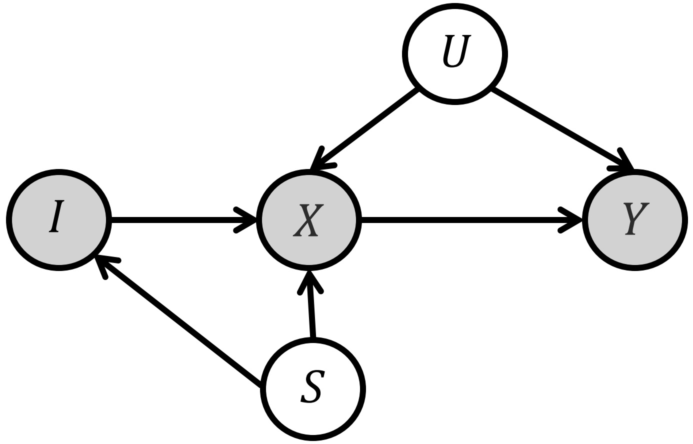
I not valid IV bc $I \to S \to Y$

I would be valid IV if we could control for S



I not valid IV bc $I \leftarrow S \to Y$

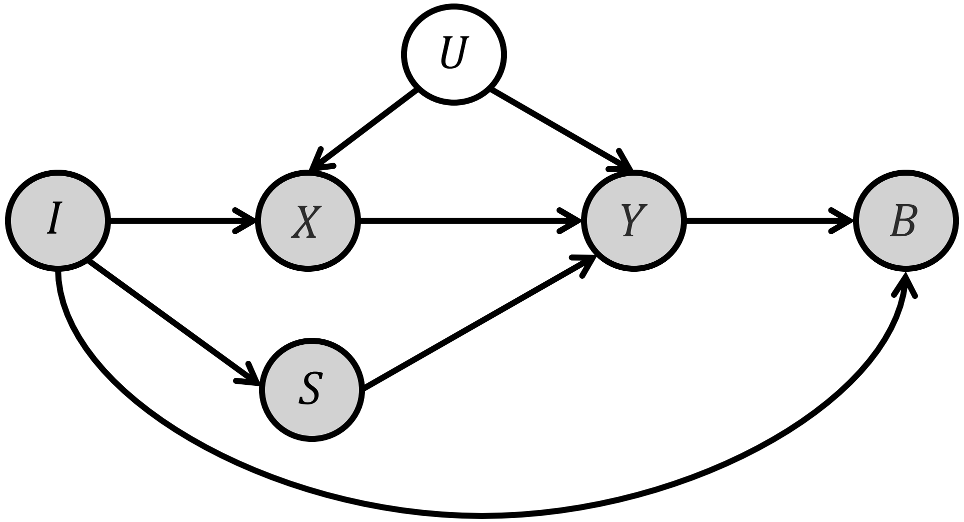
I would be valid if we could condition on S



I is a valid instrument

Example

Shaded nodes are observed

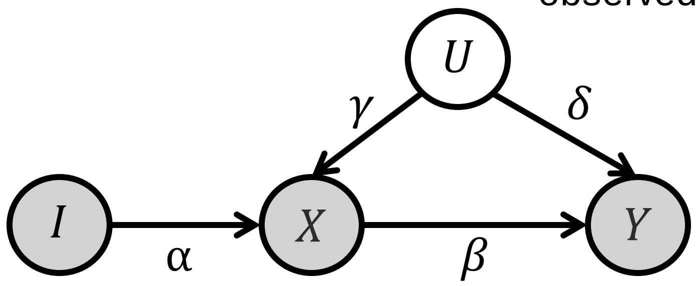


I is valid if we condition on S but not on B

Instrumental variables

- X, Y and I observed, U unobserved
- Assume a linear SEM
  - Parameter of interest is causal effect of X on Y: β

Shaded nodes are observed



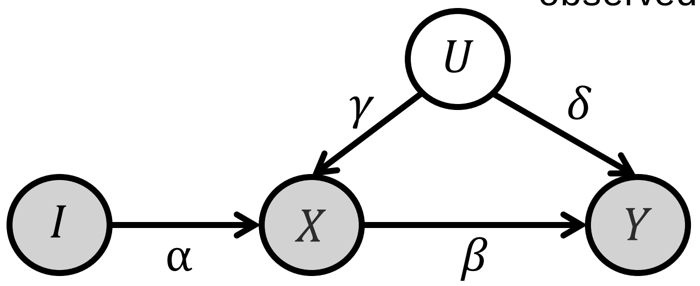
- Linear regression from Y on X (OLS):
  - Regression coefficient: $\beta^{\text{OLS}} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}$
  - Here: $\beta^{\text{OLS}} = \beta + \gamma\delta\frac{\text{Var}(U)}{\text{Var}(X)}$
  - $\hat{\beta}^{\text{OLS}}$ is biased and inconsistent

Instrumental variables

- Instrumental variable regression:
  - Regression coefficient: $\beta^{\mathrm{IV}} = \frac{\mathrm{Cov}(I,Y)}{\mathrm{Cov}(I,X)} = \beta$
  - Moment-based estimator $\hat{\beta}^{\mathrm{IV}} = \frac{\widehat{\mathrm{Cov}}(I,Y)}{\widehat{\mathrm{Cov}}(I,X)}$
  - $\hat{\beta}^{\mathrm{IV}}$ is consistent

  - If I affects only X weakly, $\mathrm{Cov}(I,X)$ is small
    - I is then called a weak instrument
    - Small estimator errors in $\widehat{\mathrm{Cov}}(I,X)$ can lead to large estimation errors in $\hat{\beta}^{\mathrm{IV}}$
  - [See R script and Series 4.]

Shaded nodes are observed

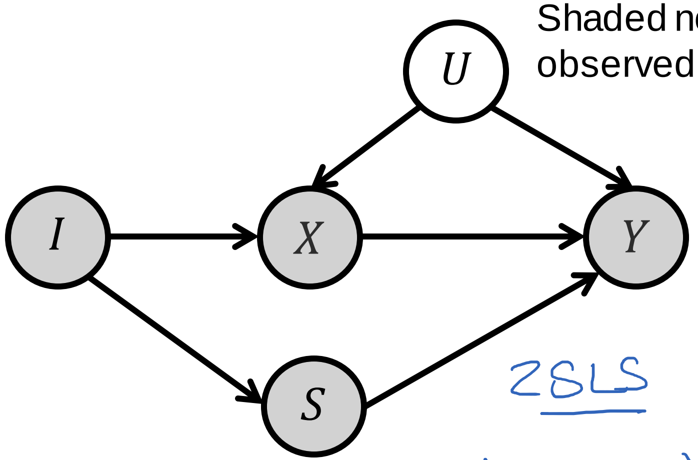


→ No control variables, everything is univariate

Instrumental variables

- Two-stage least squares (2SLS)
  i. Regress X on I and S
  ii. Construct estimate of X without influence of U: $\widetilde{\mathrm{X}}$
  iii. Regress Y on $\widetilde{\mathrm{X}}$ and S to obtain β

- Intuition
  - IV I provides a source of variation in X, uncorrelated with other common ancestors of X and Y
  - By seeing how both X and Y respond to these “perturbations” allows us to deduce something about how X influences Y
    - Assumptions needed

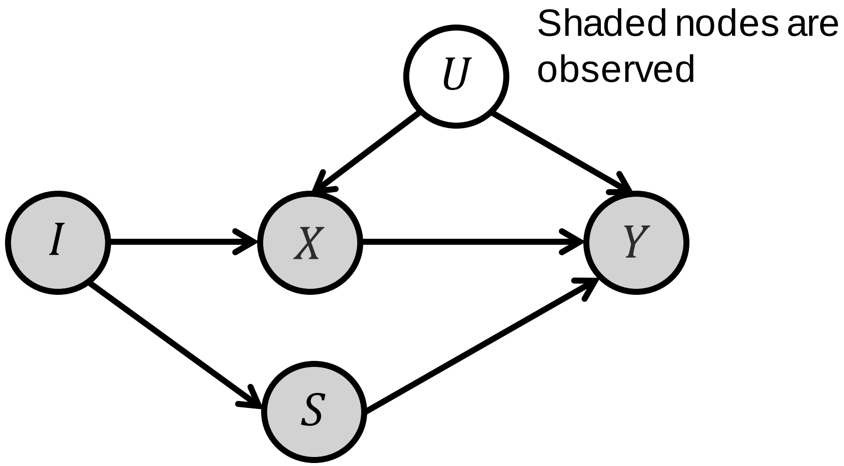


2SLS

- $\dim(y) = 1$
- can have controls
- $\dim(x)$ and $\dim(I)$ can be ≥ 1

Clicker question – Instrumental variables

- Conditions for instrumental variable (IV) I
  i. $I \not\perp\!\!\!\perp X | S$
  ii. Every path from I to Y that is not blocked by S has an arrow pointing into X: $I \perp\!\!\!\perp Y | S, do(X)$



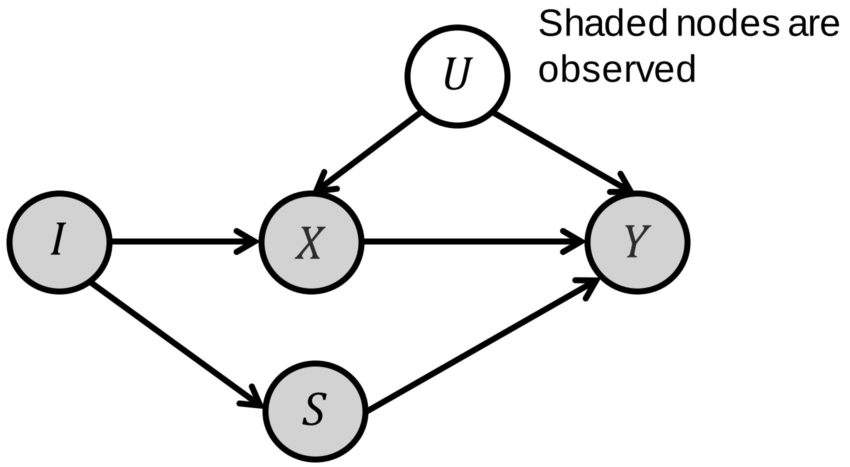
- Can we test for condition i. ($I \not\perp\!\!\!\perp X | S$)? Yes

- Can we test for condition ii. by looking at whether $I \perp\!\!\!\perp Y | X, S$? No

X is a collider on the path
$I \to X \leftarrow U \to Y$

Instrumental variables

- Conditions for instrumental variable (IV) I
  i. $I \perp\!\!\!\perp X | S$
  ii. Every path from I to Y that is not blocked by S has an arrow pointing into X: $I \perp\!\!\!\perp Y | S, do(X)$



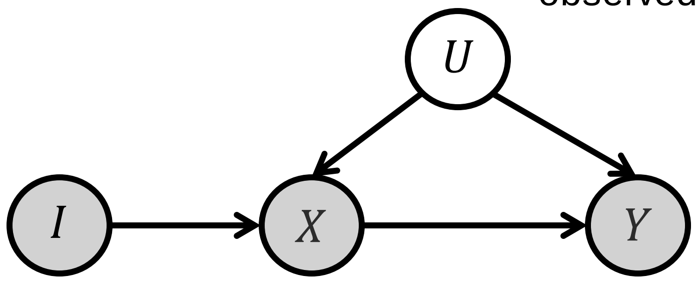
- Can test i.
- Have to argue ii.
  - After controlling for S, every mechanism by which I influences Y is mediated by X
  - No unobserved confounders between I and Y
  - Most easily defended when I is random

Instrumental variables

- Examples

I	*X*	*Y*
Random treatment assignment	Taking treatment	Health outcome
Lottery number for draft	Military service	Wage
Mosquito net supplied	Mosquito nets used	Health status

Shaded nodes are observed



Identification strategies

- Experimentation (RCT)
- Covariate adjustment
- Frontdoor criterion
- Instrumental variables

External validity and transportability

- External validity: Generalizability of empirical findings to new environments, settings or populations
  - Threats:
    - Only particular subpopulation (e.g. college students, volunteers, …)
    - Situation (e.g. lighting, noise, treatment administration, investigator, timing, …)
    - …

- Transportability: “license” to transfer causal effects learned in experimental studies to a new population
  - Sometimes possible under causal assumptions (Pearl and Bareinboim (2014))
    - Recalibration, transport formulas
  - Need to characterize commonalities and differences between populations

Transportability

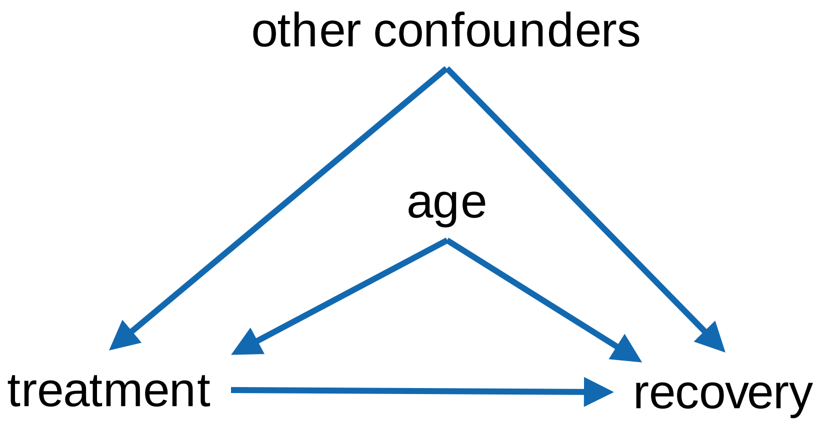
- Want to generalize findings obtained in one “environment”
  - E.g., environment could be a laboratory setting
  - Want to ”transport” conclusions to a new environment
  - New environment may differ in many aspects from that of the lab
  - If target environment is arbitrary, nothing can be learned
  - Usually target environment is deemed sufficiently similar

- What conditions permit transport?
  - See Pearl and Bareinboim (2014) for formal treatment
  - Now: look at two simple cases

Working example – Case 1

- Want to estimate the effect of treatment on recovery
- Age is one confounder
  - Older people seek treatment more often than younger people
  - Older people respond differently to treatment than younger people
- There might be other confounders

- Two populations: LA and NYC
  - Average age in NYC is significantly higher
  - Assume that the distribution of the other confounders does not differ

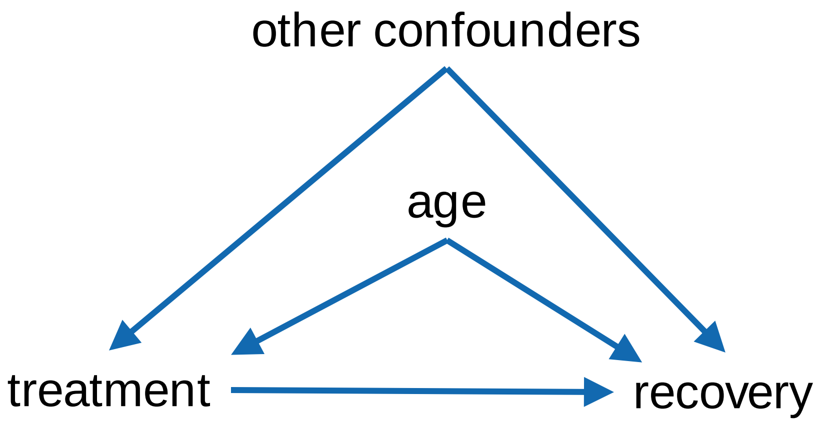


Working example – Case 1

- Perform a randomized study in LA
- Estimate causal effect of treatment on recovery for every age group $\mathrm{P}^{\mathrm{LA}}(y|do(t), \mathrm{age})$

- Want to generalize results to population of NYC
  - $\mathrm{P}^{\mathrm{LA}}(\mathrm{age}) \neq \mathrm{P}^{\mathrm{NYC}}(\mathrm{age})$
  - $\mathrm{P}^{\mathrm{LA}}(\text{other confounders}) = \mathrm{P}^{\mathrm{NYC}}(\text{other confounders})$

- How do we estimate $\mathrm{P}^{\mathrm{NYC}}(y|do(t))$?



Working example – Case 1

- Assume age-specific effects are invariant across cities, i.e.,

$$\mathrm{P}^{\mathrm{LA}}(y|do(t), \mathrm{age}) = \mathrm{P}^{\mathrm{NYC}}(y|do(t), \mathrm{age}) = \mathrm{P}(y|do(t), \mathrm{age})$$

- Then

$$\mathrm{P}^{\mathrm{NYC}}(y|do(t)) = \sum_{\mathrm{age}} \mathrm{P}(y|do(t), \mathrm{age})\ \mathrm{P}^{\mathrm{NYC}}(\mathrm{age})$$

- Transport formula: combines experimental results from LA with observational aspects of NYC population to obtain experimental claim about NYC

Working example – Case 1

- Transport formula:

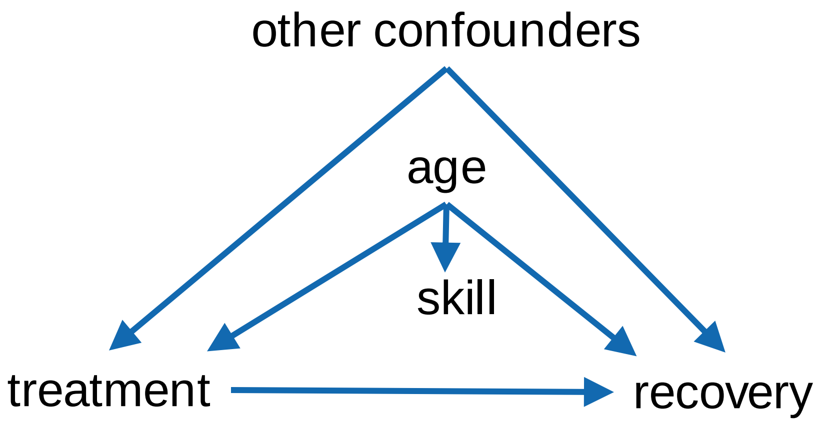
$$P^*(y|do(t)) = \sum_z P(y|do(t), z)\, P^*(z)$$

combines experimental results from source population with observational aspects of target population to obtain experimental claim about target population

Westergaard, 1916; Yule, 1934; Lane and Nelder, 1982; Cole and Stuart, 2010

Working example – Case 2

- What if age is not measured and we have only a proxy for age, e.g., language proficiency (skill)?
  - $\mathrm{P}^{\mathrm{LA}}(\mathrm{skill})$ different from $\mathrm{P}^{\mathrm{NYC}}(\mathrm{skill})$
  - Have estimated $\mathrm{P}^{\mathrm{LA}}(y|do(t), \mathrm{skill})$
- How do we estimate $\mathrm{P}^{\mathrm{NYC}}(y|do(t))$?

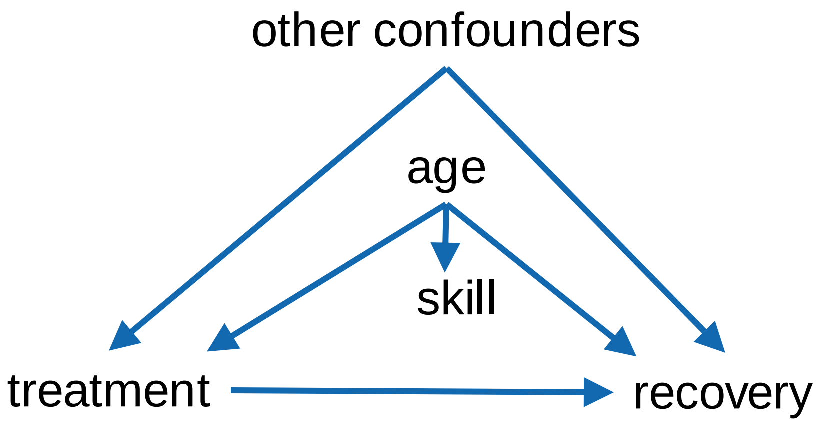


Working example – Case 2

- $\mathrm{P}^{\mathrm{LA}}(\mathrm{skill})$ different from $\mathrm{P}^{\mathrm{NYC}}(\mathrm{skill})$
  - Assume
    - difference is due to how skill depends on age
    - age distributions identical
  - Transport formula

$$\mathrm{P}^{\mathrm{NYC}}(y|do(t)) = \mathrm{P}^{\mathrm{LA}}(y|do(t))$$

$$P(skill) = \sum_{age} P(age)\, \underbrace{P(skill|age)}$$

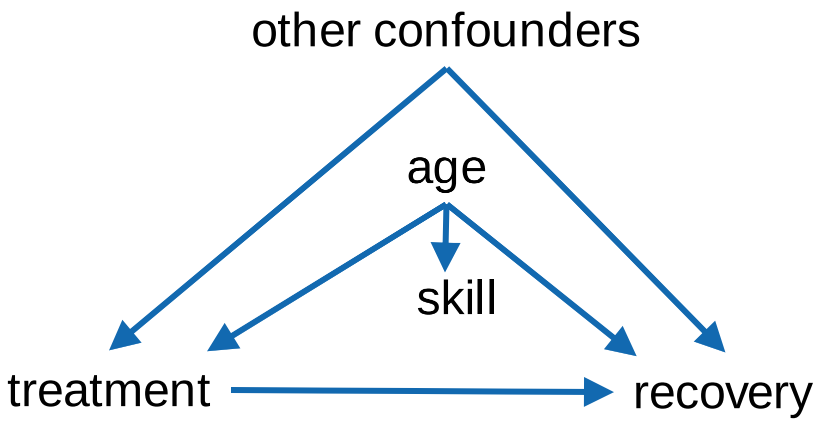


Working example – Case 2

- $\mathrm{P}^{\mathrm{LA}}(\mathrm{skill})$ different from $\mathrm{P}^{\mathrm{NYC}}(\mathrm{skill})$
  - Assume
    - $\mathrm{P}^{\mathrm{LA}}(\mathrm{skill}|\mathrm{age}) = \mathrm{P}^{\mathrm{NYC}}(\mathrm{skill}|\mathrm{age})$
    - $\mathrm{P}^{\mathrm{LA}}(\mathrm{skill}) \neq \mathrm{P}^{\mathrm{NYC}}(\mathrm{skill})$ reflects age differences
  - Previous transport formula does not hold

$$\mathrm{P}^{\mathrm{NYC}}(y|do(t)) \neq \mathrm{P}^{\mathrm{LA}}(y|do(t))$$

    - Age difference may be a critical factor in determining how people react to the treatment



Working example – Case 2

- Recall transport formula from case 1:

$$\mathrm{P}^*(y|do(t)) = \sum_z \mathrm{P}(y|do(t), z)\, \mathrm{P}^*(z)$$

- Invariance of “Z-specific” causal effects should not be taken for granted

- Case 2:
  - Cannot use this transport formula for $Z = \text{skill}$ if $\mathrm{P}^{\mathrm{LA}}(\text{age}|\text{skill}) \neq \mathrm{P}^{\mathrm{NYC}}(\text{age}|\text{skill})$ since then skill-specific effects are not invariant across cities, i.e.,

$$\mathrm{P}^{\mathrm{LA}}(y|do(t), \text{skill}) \neq \mathrm{P}^{\mathrm{NYC}}(y|do(t), \text{skill}) \neq \mathrm{P}(y|do(t), \text{skill})$$

  - NYC resident with skill level $Z = z$ likely to be in different age group than LA resident with $Z = z$

Transportability

- Recall transport formula from case 1:

$$\mathrm{P}^*(y|do(t)) = \sum_z \mathrm{P}(y|do(t), z)\, \mathrm{P}^*(z)$$

- Invariance of "Z-specific" causal effects should not be taken for granted

- Proper transport formula depends on the causal context in which population differences are embedded

Counterfactuals

- Not all causal questions can be expressed with $p(y|do(x))$
  - E.g., what fraction of the healthy untreated population would have gotten the disease had they been treated?
  - Retrospective thinking
- Consider SEM

$$
\begin{aligned}
Z &\leftarrow f_Z(N_Z) \\
X &\leftarrow f_X(Z, N_X) \\
Y &\leftarrow f_Y(X, N_Y)
\end{aligned}
$$

- Counterfactual reasoning analyzes relations such as "Y would be y had X been x in situation $N = n$"
  - Unit-level reasoning

Examples

Counterfactuals

- Noise as “unobserved uncertainty-producing variables” or “background variables”
- Counterfactuals may earn predictive value
  - When noise remains constant; or
  - When noise can be observed sometime in the future
    - See eye-doctor example
- Often this is not the case and many counterfactual statements cannot be falsified
  - Cannot observe $Y|T = 1$ and $Y|T = 0$ for the same individual

Recap

- Concepts to know:
  - Instrumental variables
  - Transportability
    - Simple cases
  - Counterfactuals

References and acknowledgments

- Instrumental variables
  - Shalizi (2019). Chapter 22.
- Transportability
  - Pearl and Barenboim. External Validity: From Do-Calculus to Transportability Across Populations. Statistical Science, 2014.
- Counterfactuals
  - Peters, Janzing and Schölkopf (2017). Elements of Causal Inference.