Causality in Biomedicine Lecture Series: Lecture 3

Ava Khamseh

5 Feb, 2020

Info

GitHub: https://github.com/avakhamseh/Causality_in_Biomedicine_Lectures

Causal machine learning master class, deadline 11 Feb 2020

Tuesday 03 Mar 2020, Time: 10:00-17:30

Organised by Alan Turing Institute, London

- Average treatment effect: $\tau = \hat{\mathbb{E}}[\tau^{(i)}] = \hat{\mathbb{E}}[y_1^{(i)} - y_0^{(i)}] = \frac{1}{N}\sum_{i=0}^{N}\left(y_1^{(i)} - y_0^{(i)}\right)$

- Targeted maximum likelihood estimation for ATE
- Doubly robust estimates: Regression + propensity score
- Machine learning considerations
- Longitudinal data with time-dependent confounding

So Far …

- **Matching:** Stratification, Propensity score, IPTW, …

$$y_1^{(i)}, y_0^{(i)} \perp\!\!\!\perp t^{(i)} \; | b(x)$$

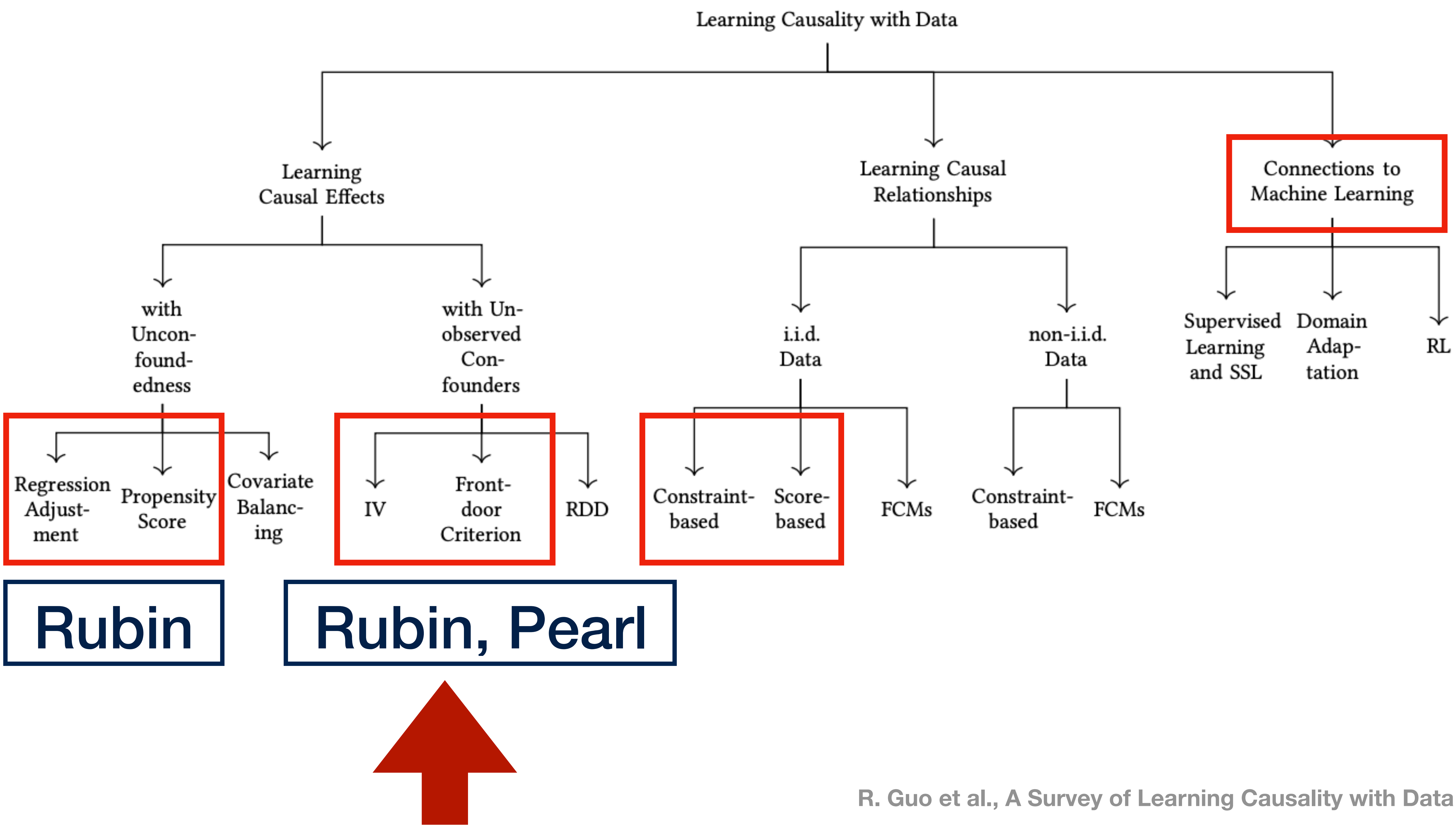
- Estimation of propensity scores directly from the data & algorithms

$$e(x) = p(t = 1|x)$$

- **Sensitivity analysis:** No guarantee that matching leads to balance on variables we did not match for, people who look comparable may differ. If there is hidden bias, how severe is it:
  - Does the conclusion change from statistically significant to not?
  - Does it change the direction of effect?

Notice: There are two sources of uncertainty (confidence interval):
1) Due to the causal estimates
2) Due to sensitivity analysis

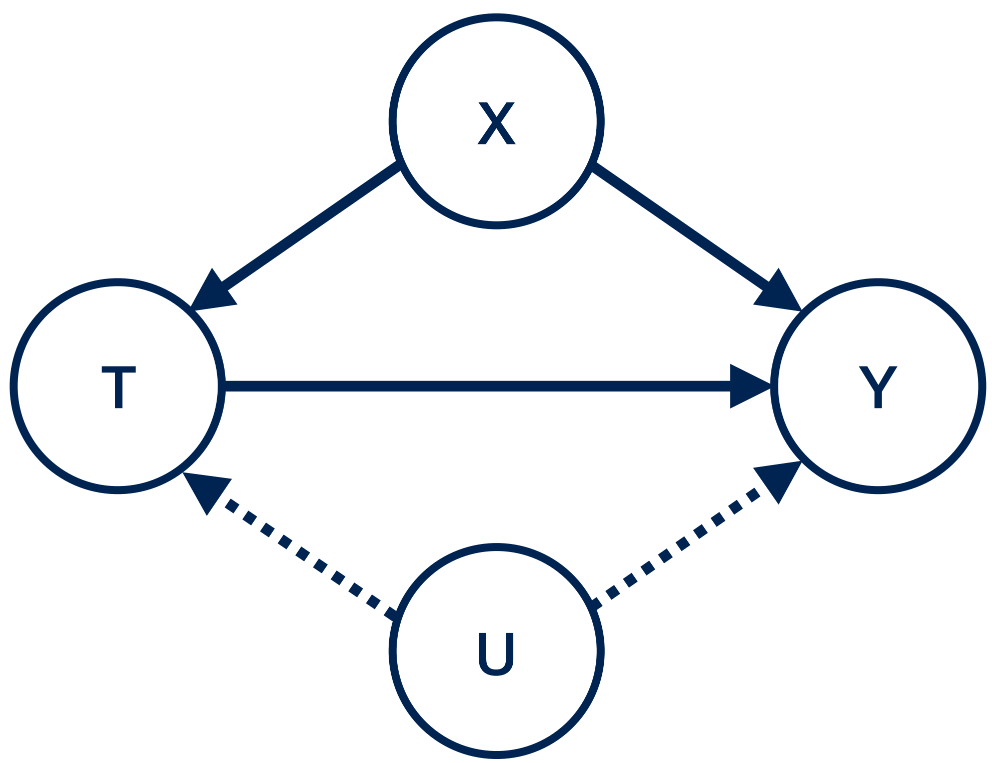
Overview of the field



R. Guo et al., A Survey of Learning Causality with Data

Instrumental Variable (Originally due to Rubin)

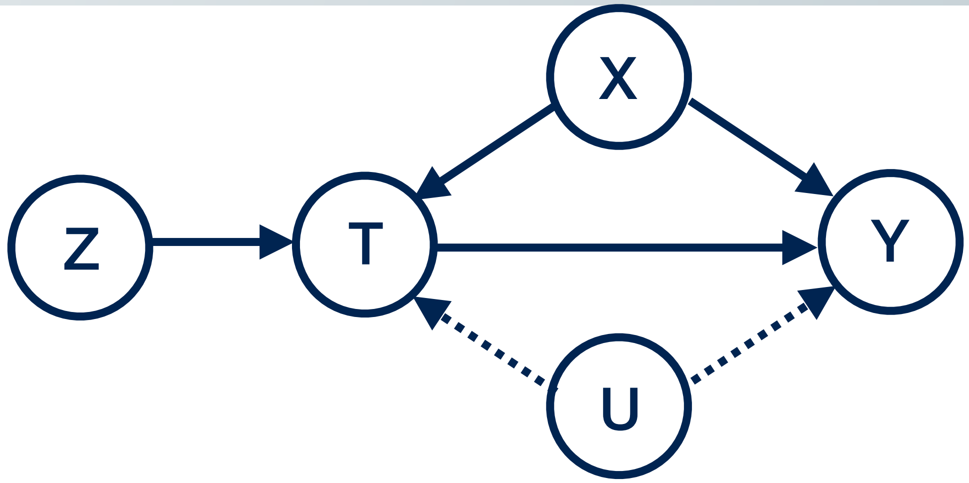
- Unobserved confounders (U), **violates unconfoundedness (ignorability),** i.e. conditioning on X alone, would not results in a randomised treatment assignment
- Unconfoundedness is fundamentally unverifiable



Rubin 1996

Instrumental Variable example

- Example 1:
  - T: smoking during pregnancy
  - Y: birthweight
  - X: parity, mother's age, weight, …
  - U: Other unmeasured confounders



- Randomise Z (intention-to-treat): either receive encouragement to stop smoking (Z=1), or receive usual care (Z=0)
- Intention-to-treat analysis gives causal effect estimator of encouragement z on outcome y:

$$\mathbb{E}(y|z=1) - \mathbb{E}(y|z=0)$$

- What can we say about the causal effect of smoking itself?

Instrumental Variable assumptions

- **SUTVA**: Potential outcomes for each individual i are unrelated to the treatment status of other individuals:

$$Y^{(i)}(\mathbf{Z}, \mathbf{T}) = Y^{(i)}(Z^{(i)}, T^{(i)}) \ , \ |\mathbf{Z}| = |\mathbf{T}| = N \text{ individuals}$$

- Treatment assignment Z (**associated** with the treatment) is **random**:

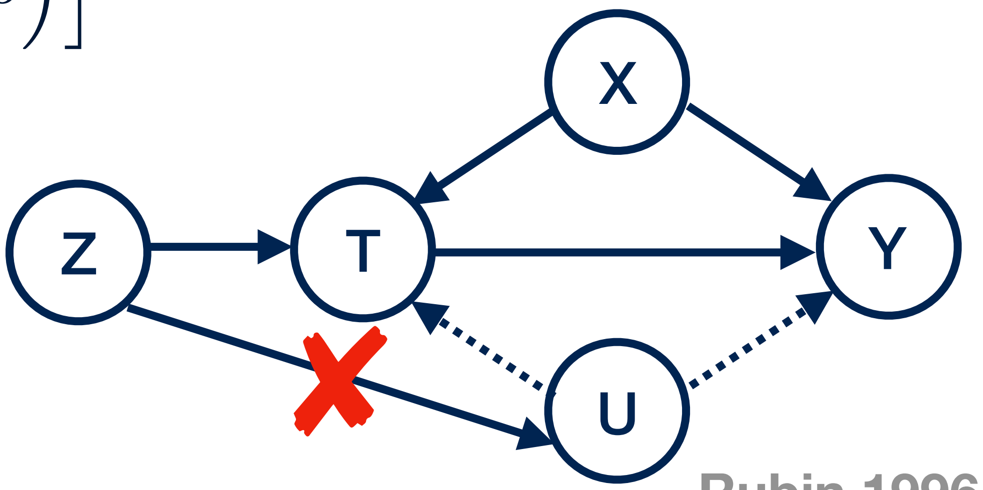
$$P(Z^{(i)} = 0) = P(Z^{(i)} = 1) \ , \ \forall i$$

- **Exclusion Restriction**: Any effect of Z on Y is via an effect of Z on T, i.e., Z should not affect Y when T is held constant $\left(Y^{(i)}|z = 1, t\right) = \left(Y^{(i)}|z = 0, t\right)$

- **Non-zero Average:** $\mathbb{E}\left[\left(T^{(i)}|z = 1\right) - \left(T^{(i)}|z = 0\right)\right]$

- **Monotonicity** (increasing encouragement "dose" increases probability of treatment, no defiers)**:**

$$\left(T^{(i)}|z = 1\right) \geq \left(T^{(i)}|z = 0\right)$$



Rubin 1996

Instrumental Variable: Potential values of T

Population	T|z=0	T|z=1	Desription
Never-takers	0	0	Causal effect of Z on Y is zero, since $\left(T^{(i)} \vert z=1\right) - \left(T^{(i)} \vert z=0\right) = 0$
Compliers	0	1	Treatment received is randomised, $\left(T^{(i)} \vert z=1\right) - \left(T^{(i)} \vert z=0\right) = 1$ **causal effect inference**: $\left(Y^{(i)} \vert T^{(i)}=1\right) - \left(Y^{(i)} \vert T^{(i)}=0\right)$
Defiers	1	0	Rule out by **monotonicity**, since $\left(T^{(i)} \vert z=1\right) - \left(T^{(i)} \vert z=0\right) = -1$
Always-takers	1	1	Causal effect of Z on Y is zero, since $\left(T^{(i)} \vert z=1\right) - \left(T^{(i)} \vert z=0\right) = 0$

Rubin 1996

Instrumental Variable: The estimand

Want ATE:

$$\mathbb{E}\left[\left(Y^{(i)}\left(t^{(i)}=1\right)-Y^{(i)}\left(t^{(i)}=0\right)\right)\right]$$

"Almost"

Will estimate:

$$\boxed{\hat{\tau}=\frac{\mathbb{E}\left[\left(Y^{(i)}|z=1\right)-\left(Y^{(i)}|z=0\right)\right]}{\mathbb{E}\left[\left(T^{(i)}|z=1\right)-\left(T^{(i)}|z=0\right)\right]}}$$

Rubin 1996

Instrumental Variable: The estimand

$$\mathbb{E}\left[\left(Y^{(i)}\left(t^{(i)}=1\right)-Y^{(i)}\left(t^{(i)}=0\right)\right)\right]$$

$$\boxed{\hat{\tau}=\frac{\mathbb{E}\left[\left(Y^{(i)}|z=1\right)-\left(Y^{(i)}|z=0\right)\right]}{\mathbb{E}\left[\left(T^{(i)}|z=1\right)-\left(T^{(i)}|z=0\right)\right]}}$$

Derivation:

$$\left(Y^{(i)}|T^{(i)}(z=1)\right)-\left(Y^{(i)}|T^{(i)}(z=0)\right)$$ t is either t=0 or t=1, and exclusion restriction

$$=\left[Y^{(i)}\left(t^{(i)}=1\right)\cdot\left(t^{(i)}|z=1\right)+Y^{(i)}\left(t^{(i)}=0\right)\cdot\left(1-\left(t^{(i)}|z=1\right)\right)\right]$$

$$-\left[Y^{(i)}\left(t^{(i)}=1\right)\cdot\left(t^{(i)}|z=0\right)+Y^{(i)}\left(t^{(i)}=0\right)\cdot\left(1-\left(t^{(i)}|z=0\right)\right)\right]$$

$$=\left(Y^{(i)}\left(t^{(i)}=1\right)-Y^{(i)}\left(t^{(i)}=0\right)\right)\cdot\left(\left(t^{(i)}|z=1\right)-\left(t^{(i)}|z=0\right)\right)$$

Hence, the causal effect of Z on Y for individual i, is the product of the causal effect of Z on T, and, the casual effect of T on Y.

Rubin 1996

Instrumental Variable: The estimand

To continue the derivation, we use the fact that:

$$\mathbb{E}\left[xy\right] = \int\int xy\ p(x,y)dxdy = \int dy\ y\ p(y)\int dx\ x\ p(x|y) = \int dy\ y\ p(y)\mathbb{E}[x|y]$$

and write,

$$\mathbb{E}\left[\left(Y^{(i)}|T^{(i)}(z=1)\right) - \left(Y^{(i)}|T^{(i)}(z=0)\right)\right]$$

0, 1, -1

$$= \mathbb{E}\left[\left(Y^{(i)}\left(t^{(i)}=1\right) - Y^{(i)}\left(t^{(i)}=0\right)\right)\cdot\left(\left(t^{(i)}|z=1\right) - \left(t^{(i)}|z=0\right)\right)\right]$$

Rubin 1996

Instrumental Variable: The estimand

To continue the derivation, we use the fact that:

$$\mathbb{E}\left[xy\right] = \int\int xy\ p(x,y)dxdy = \int dy\ y\ p(y) \int dx\ x\ p(x|y) = \int dy\ y\ p(y)\mathbb{E}[x|y]$$

and write,

$$\mathbb{E}\left[\left(Y^{(i)}|T^{(i)}(z=1)\right) - \left(Y^{(i)}|T^{(i)}(z=0)\right)\right]$$

0, 1, -1

$$= \mathbb{E}\left[\left(Y^{(i)}\left(t^{(i)}=1\right) - Y^{(i)}\left(t^{(i)}=0\right)\right)\cdot\left(\left(t^{(i)}|z=1\right) - \left(t^{(i)}|z=0\right)\right)\right]$$

$$= \mathbb{E}\left[\left(Y^{(i)}\left(t^{(i)}=1\right) - Y^{(i)}\left(t^{(i)}=0\right)\right) \mid \left(\left(t^{(i)}|z=1\right) - \left(t^{(i)}|z=0\right)\right)=1\right]\cdot$$

$$P\left(\left(t^{(i)}|z=1\right) - \left(t^{(i)}|z=0\right)=1\right)$$

$$-\mathbb{E}\left[\left(Y^{(i)}\left(t^{(i)}=1\right) - Y^{(i)}\left(t^{(i)}=0\right)\right) \mid \left(\left(t^{(i)}|z=1\right) - \left(t^{(i)}|z=0\right)\right)=-1\right]\cdot$$

$$P\left(\left(t^{(i)}|z=1\right) - \left(t^{(i)}|z=0\right)=-1\right)$$

0, by monotonicity

Rubin 1996

Instrumental Variable: The estimand

$$\frac{\mathbb{E}\left[\left(Y^{(i)}|T^{(i)}(z=1)\right)-\left(Y^{(i)}|T^{(i)}(z=0)\right)\right]}{\mathbb{E}\left[\left(t^{(i)}|z=1\right)-\left(t^{(i)}|z=0\right)\right]}$$

$$=\mathbb{E}\left[\left(Y^{(i)}\left(t^{(i)}=1\right)-Y^{(i)}\left(t^{(i)}=0\right)\right)\Big|\left(\left(t^{(i)}|z=1\right)-\left(t^{(i)}|z=0\right)\right)=1\right]$$

i.e. restricting to *compliers*, the average casual effect of Z on Y is proportional to the average causal effect of T on Y. Rubin 1996

- In this example, Z was randomly assigned as part of the study
- IV can also be randomised in nature (nature randomiser):
    - Mendelian randomisation
    - Quarter of birth

Pearl’s framework

Graphical models & Do-calculus

Causal Inference: DoWhy (a unifying language)

- **Model** a causal inference problem using assumptions, [**Pearl**'s Causal Graphical Models]

- **Identify** an expression for the causal effect under these assumptions ("causal estimand"), [**Pearl**'s Causal Graphical Models]

- **Estimate** the expression using statistical methods such as matching or instrumental variables, [**Rubin**'s Potential Outcomes] ✓

- **Verify** the validity of the estimate using a variety of robustness checks. ✓

Pearl’s Causal Model

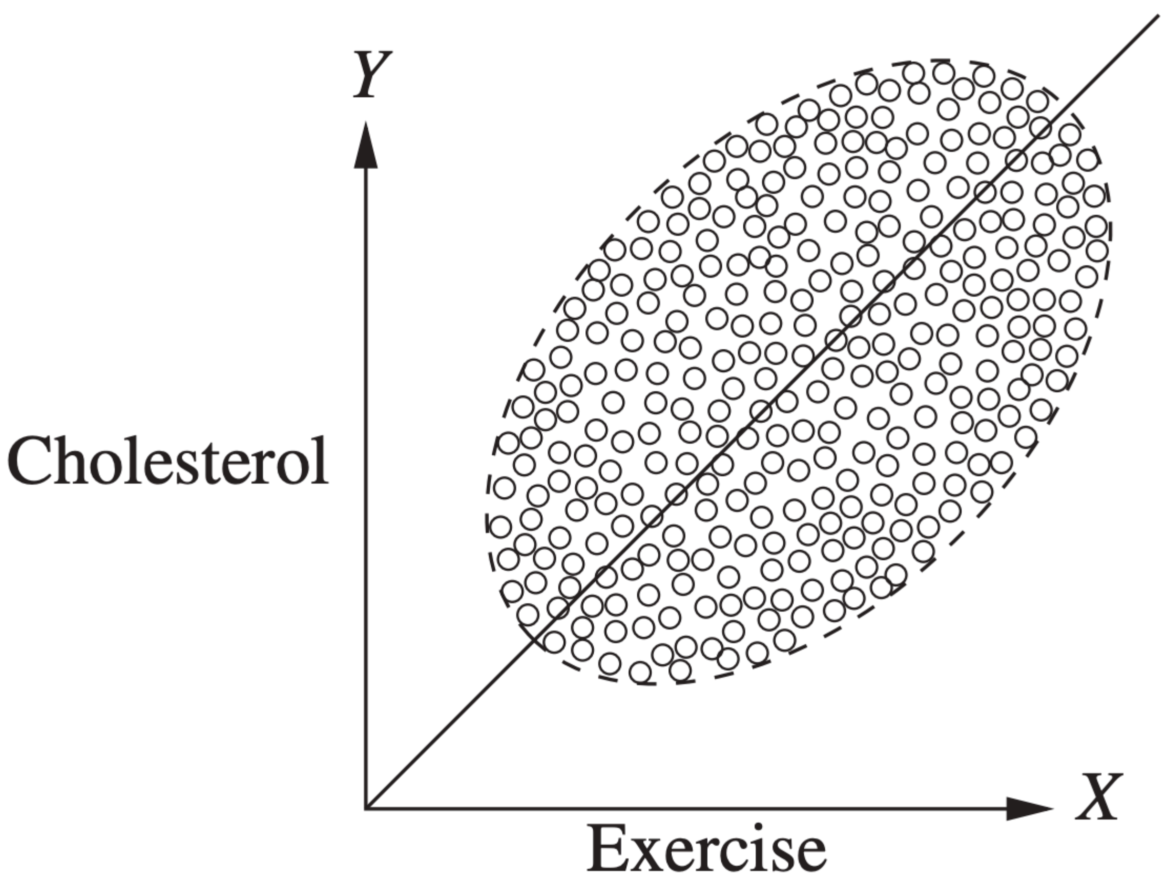
- Ladder of causation:
  - **Association:** What does a symptom tell me about a disease?
  - **Intervention (perturbation):** If I take aspirin will my headache be cured?
  - **Counterfactual:** Was it the aspirin that stopped the headache? (alternative versions of past events, strongest causal statements e.g. **physical laws**)

- Aim: To **model** and **identify** the causal estimand

- Causal graphical models + structural equations

Simpson’s Paradox

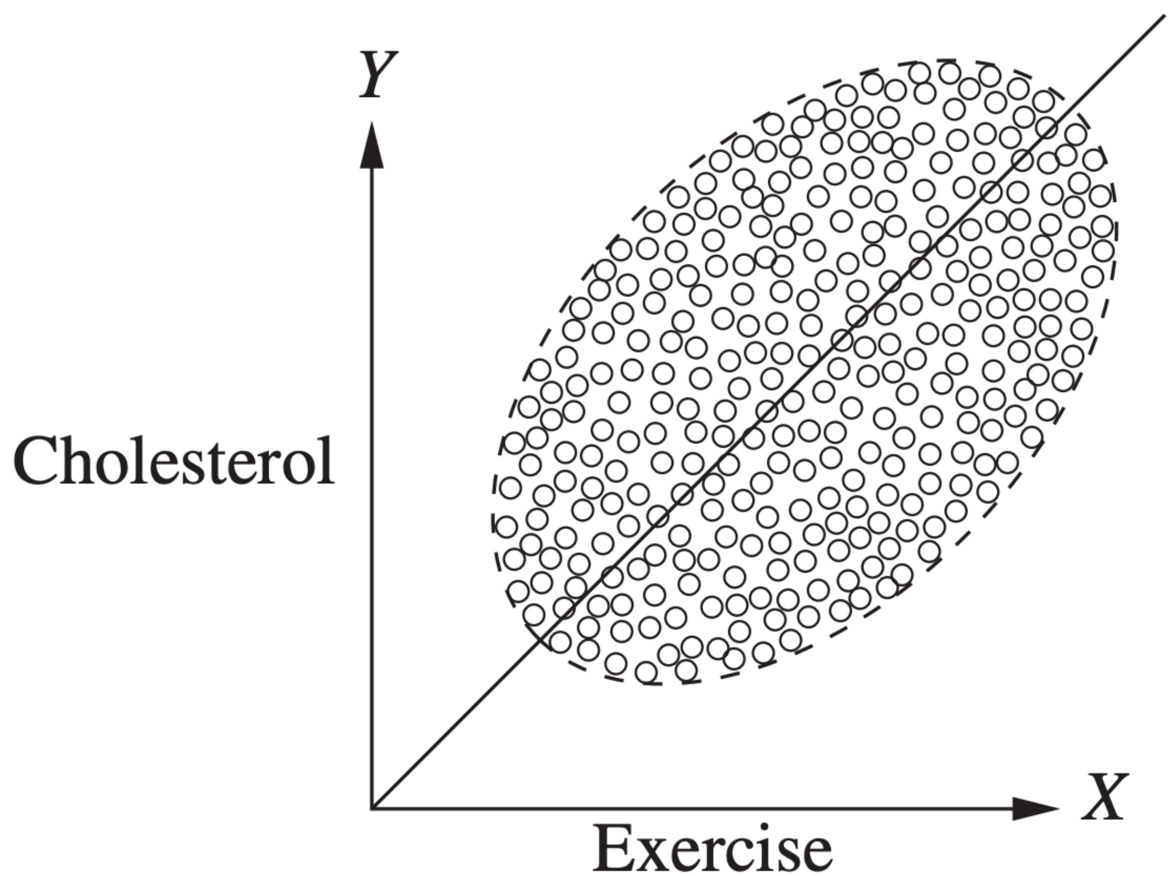
- Why concluding causality from purely associational measures, i.e. correlation, can be **very wrong** (not just neutral): “It would have better not to make any statements!”

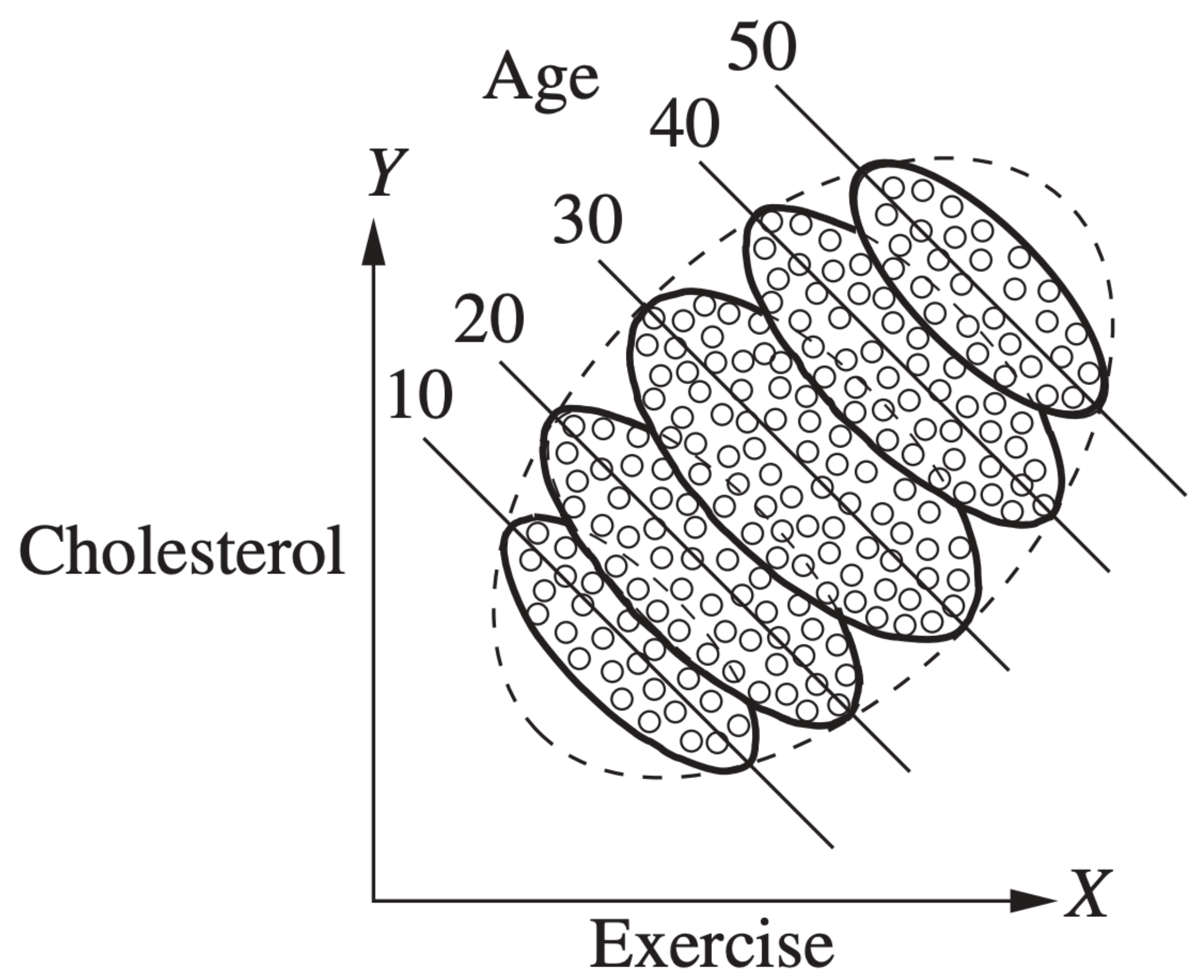


Causal Inference in Statistics, Pearl (2016)

Simpson's Paradox

- Why concluding causality from purely associational measures, i.e. correlation, can be **very wrong** (not just neutral): "It would have better not to make any statements!"

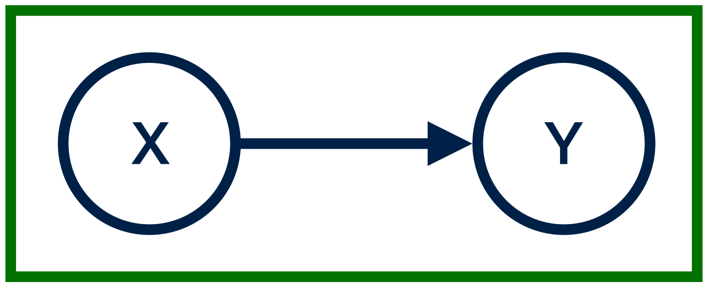


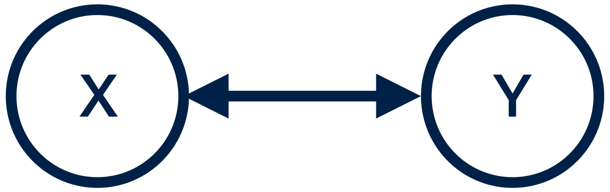


Causal Inference in Statistics, Pearl (2016)

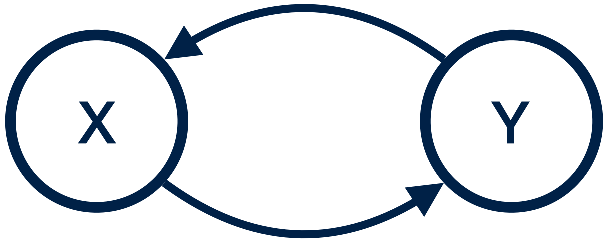
Causal Graphical Models

- Diagrammatic representation of probability distributions + **causal info**
- **Graph:** Consists of a set of **vertices** V (nodes), **edges** E
- V are the variables and E contains information between the variables
- Graphs can be directed, undirected and bidirectional (confounder?)





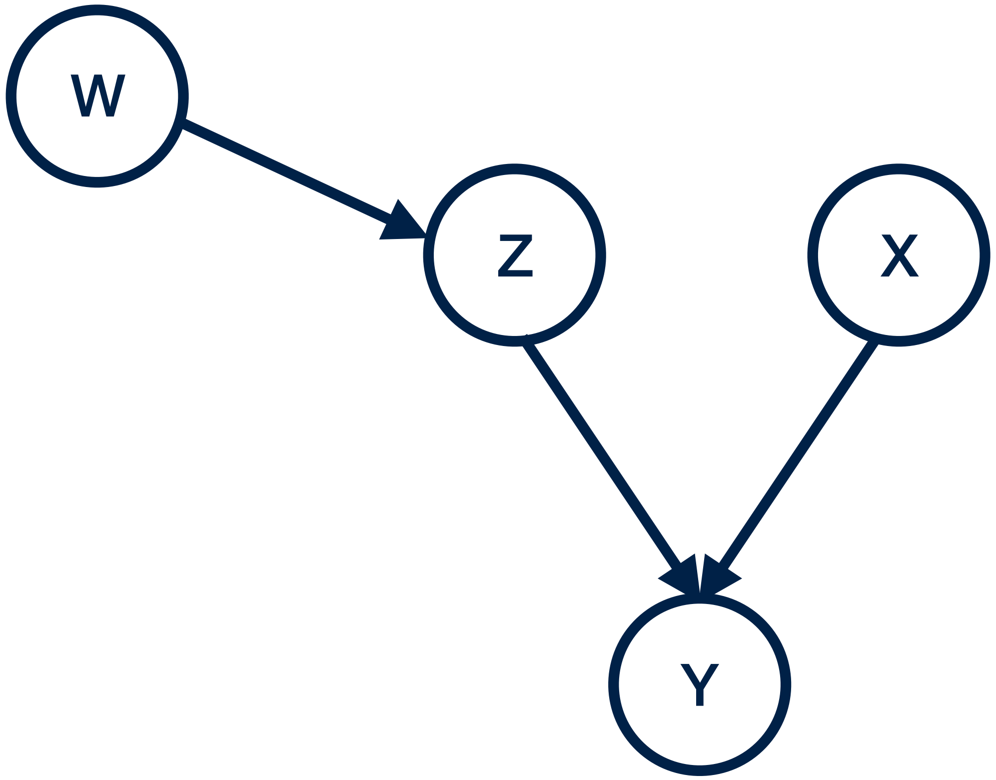
- Directed graphs may include directed cycles, i.e., mutual causation/feed-back process.



- A graph with no directed cycles is an **acyclic** graph.

Causality, Pearl (2009)

Directed Acyclic Graphs (DAGs)



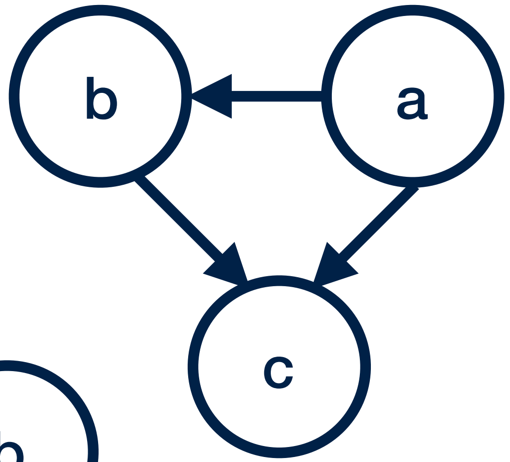
Z, X are parents of Y
Z, X, W are ancestors of Y
Y has no children
X has no parents

- DAG in which every node has at most one parent is a **tree**
- A tree in which every node has a most one child is a **chain**
- **DAG:**
  - Expresses **model assumptions** explicitly
  - Represent **joint probability** functions
  - Provides **efficient inference** of observations

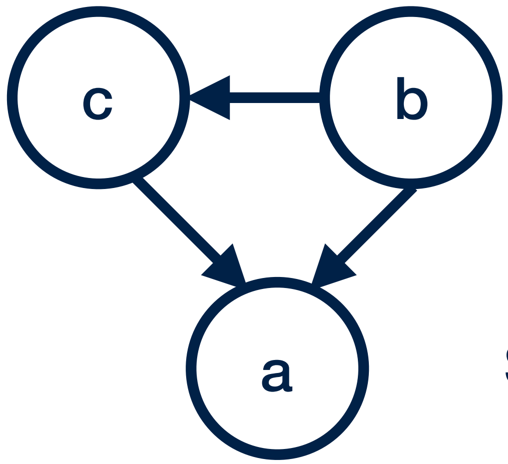
Causality, Pearl (2009)

DAG contains more info than joint probability

$$p(a,b,c) = p(c|a,b)p(a,b) = p(c|a,b)p(b|a)p(a)$$



$$p(a,b,c) = p(a|b,c)p(b,c) = p(a|b,c)p(c|b)p(b)$$



Symmetric in a, b, c

NOT symmetric

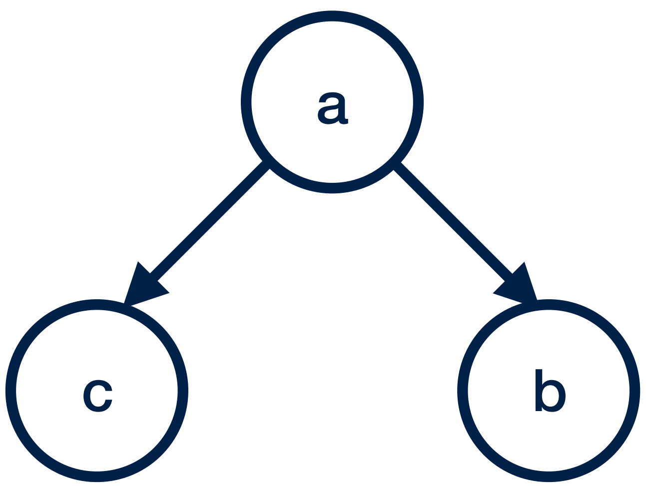
- Probabilistic notations are not enough to describe causal aspects
- Using repeated application of Bayes' rule, one can write any joint probability distribution in terms of its marginals
- A graph is **fully connected** if there is a link between every pair of nodes
- The interest lies in the **absence** of a link and link **direction**.

Next time

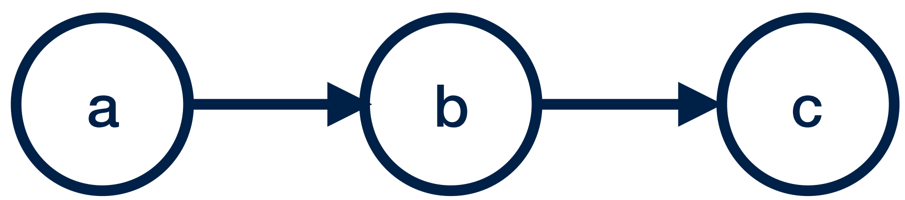
- Conditional independence via graphs

- D-separation

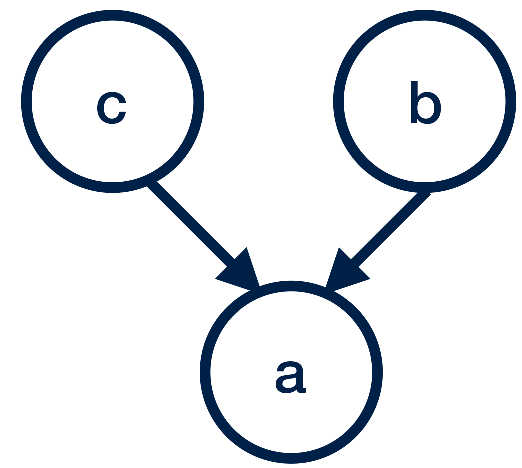
- Discuss the 3 main graph structures:



Fork



Chain



Collider

- **Do-calculus** and **identification**

Causality in Biomedicine Lecture Series: Lecture 3

Ava Khamseh

5 Feb, 2020