Applied Data Analysis (CS401)

Lecture 6
Causal analysis of observational data
25 Oct 2023

EPFL

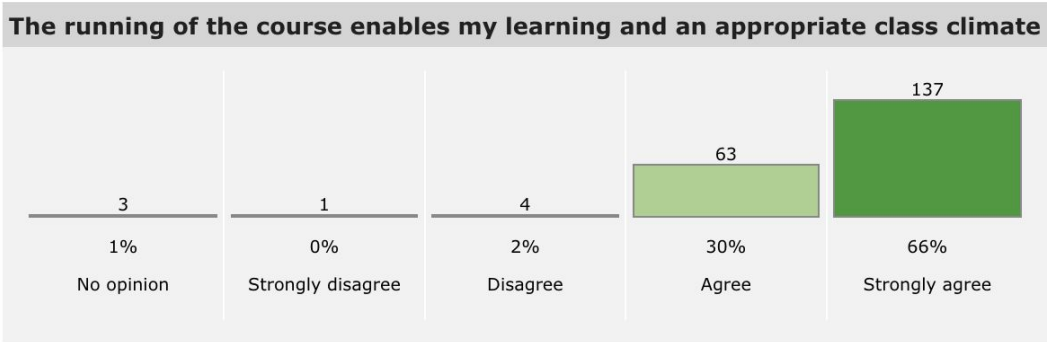
Robert West

dlab

Announcements

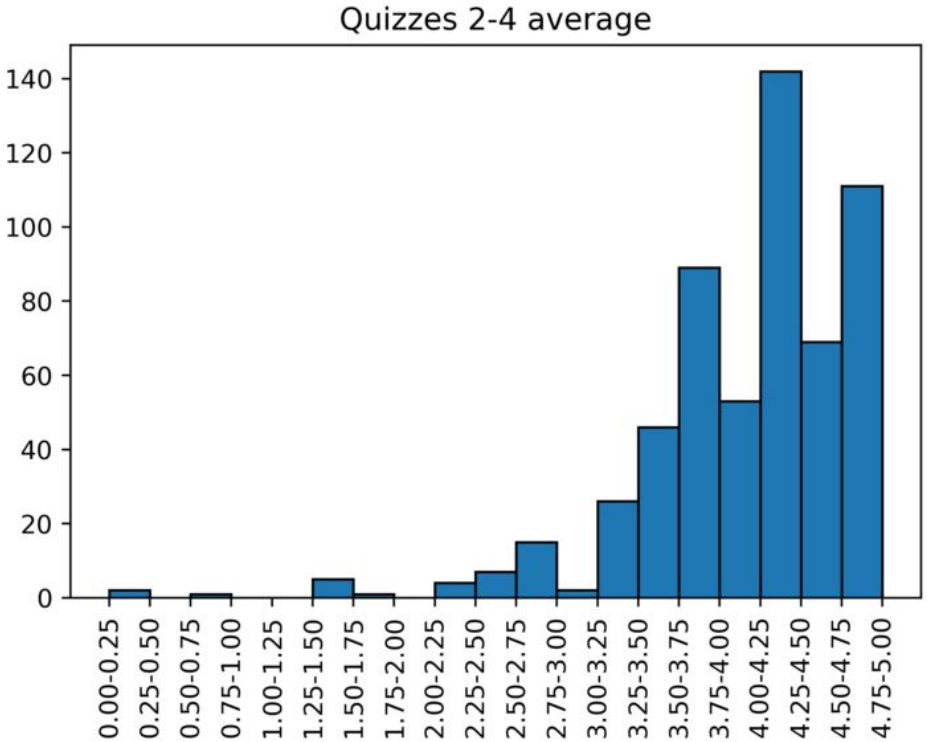
- Homework H1 due on Fri 27 Oct 23:59
  - You can ask questions until Thu 23:59; we won't answer questions asked after Thu
- Project:
  - Milestone P1 feedback has been released
  - Milestone P2: get cracking once homework H1 is done!
  - Don't use ChatGPT (one student already penalized on P1)
- Friday's lab session:
  - Exercises on causal analysis of observational data
  - Quiz 5

Course evaluation



- Thanks for your feedback! -- We'll use it to improve the class further
- Most of you like the class and what you learn
- Some concerns: class too hands-off ("more code!"); class too hands-on ("more theory!"); exercise sessions too crowded (?); quiz difficulty, wording, timing

Quiz scores so far



Feedback

Give us feedback on this lecture here:
https://go.epfl.ch/ada2023-lec6-feedback

- What did you (not) like about this lecture?
- What was (not) well explained?
- On what would you like more (fewer) details?
- …

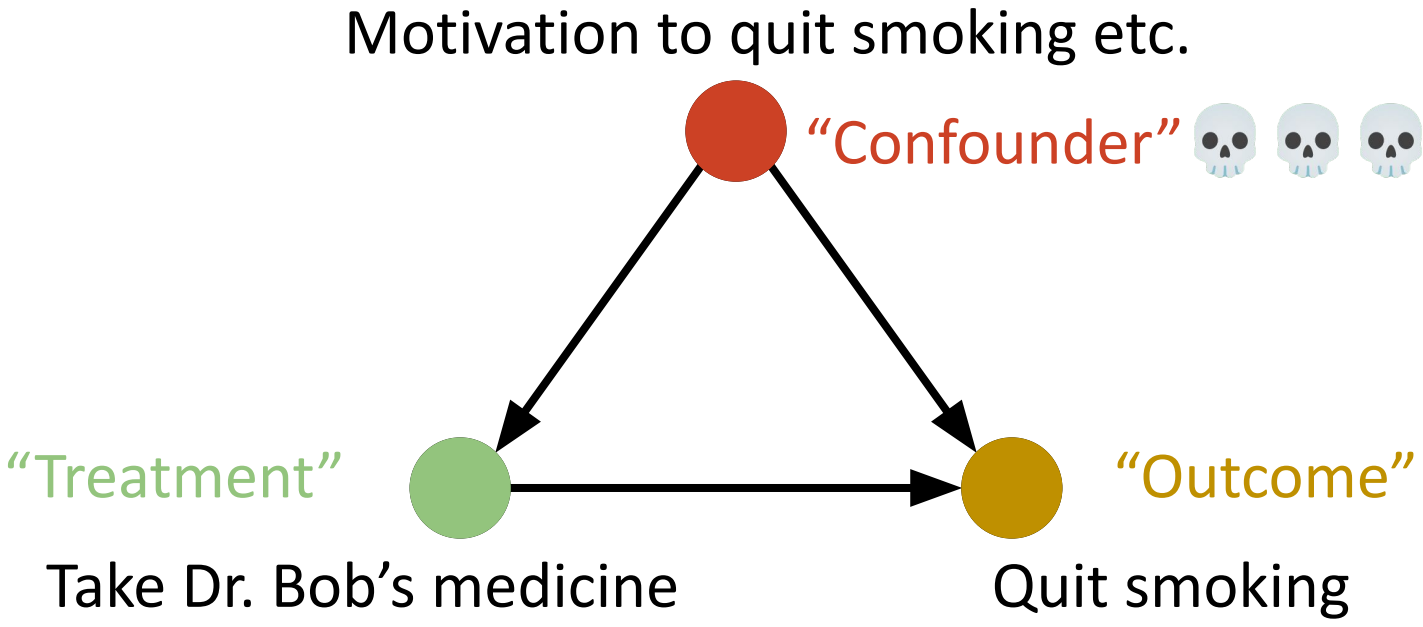
Dr. Bob's smoking cure

- I claim to have developed a medicine that helps you quit smoking
- I ask all smokers: "Do you want to try my medicine?"
- Smokers = {treated smokers}
  ∪ {untreated ("control") smokers}
- Fraction of successful quitters is higher in the treated group
- I conclude: "My medicine helps you quit smoking! Buy it!"
- Do you believe me?

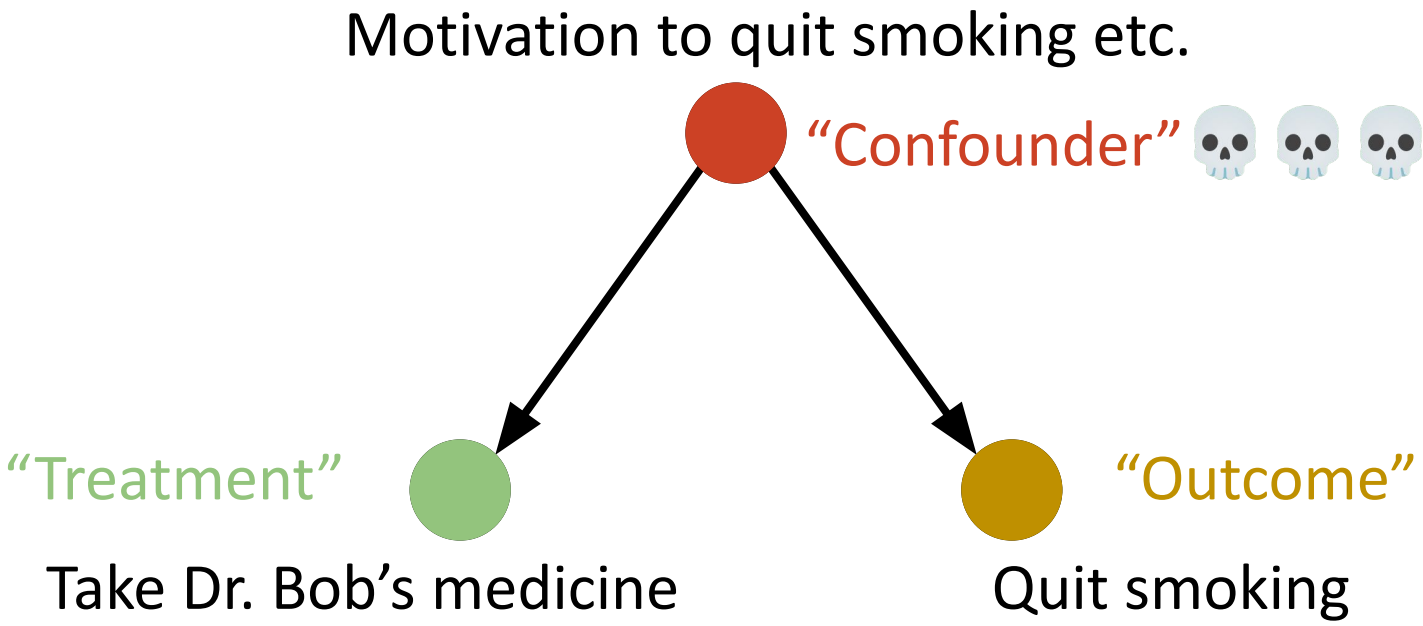
Goals of this lecture

- Clarify difference between **experimental** and **observational** studies
- Highlight **pitfalls** of observational studies
- Give you **tools** for avoiding the pitfalls, allowing you to draw valid conclusions from "found data" (very useful for project!)
- Motivate you to **read** Rosenbaum's great book "Design of Observational Studies" (in particular Chapters 1, 2, and 3; or this book) and Pearl's eye-opening "Book of Why"

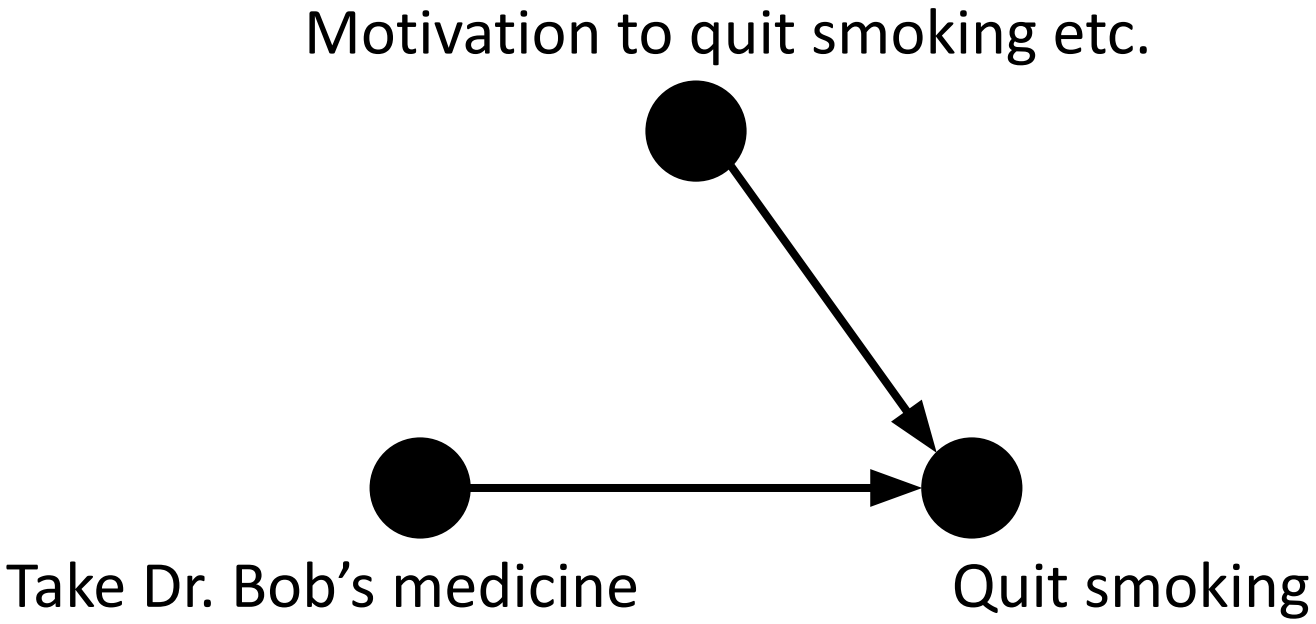
Dr. Bob's "experiment" as a causal diagram



Dr. Bob's "experiment" as a causal diagram



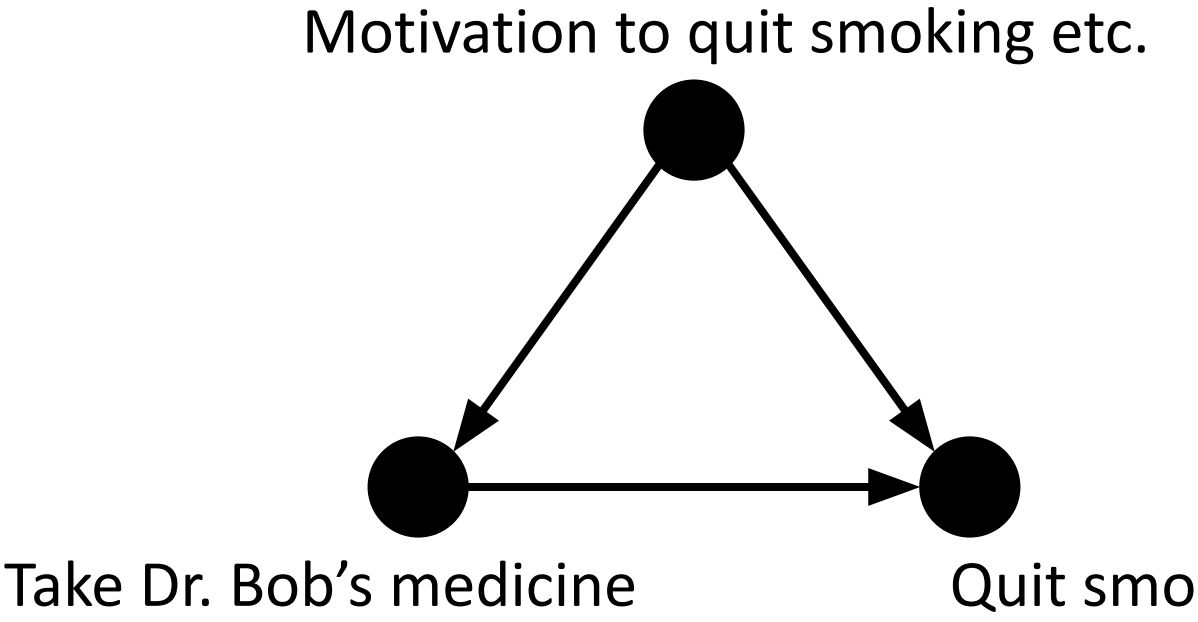
Ideal setting as a causal diagram



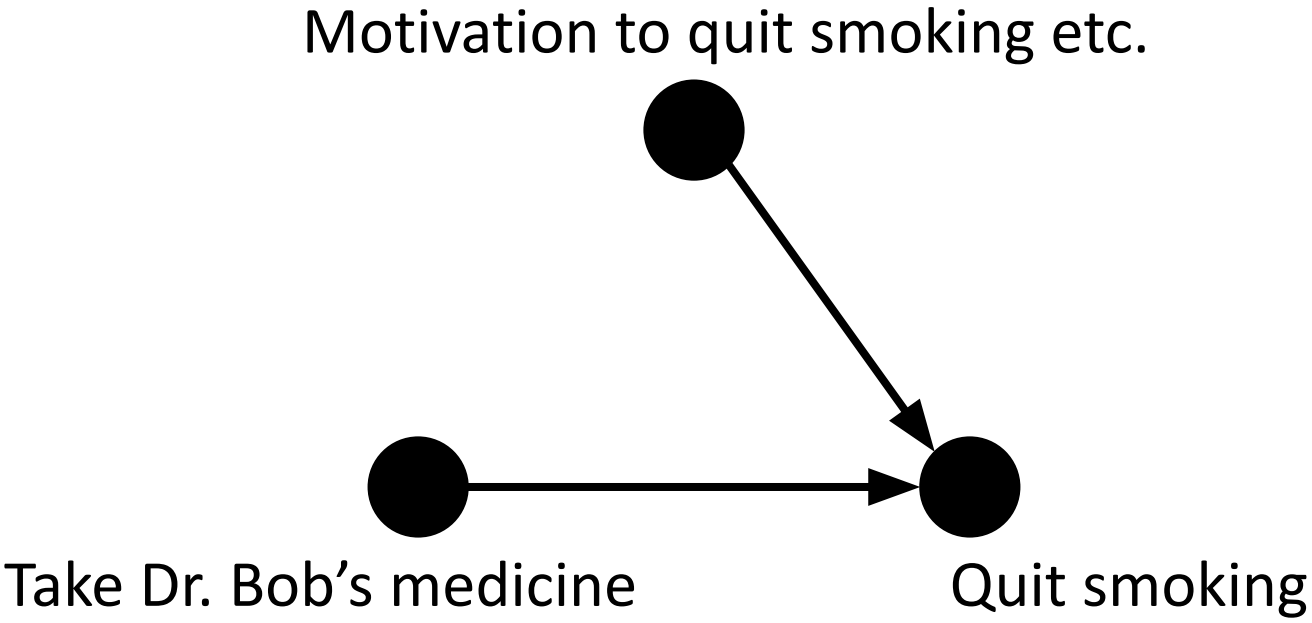
Randomized controlled experiments

- Two experimental conditions:
  - Treatment (e.g., medicine)
  - Control (e.g., placebo [fun fact])
- Assignment of participants to conditions is random
  - Probability of receiving treatment same for everyone
- Treatment and control groups are indistinguishable
  - E.g., determination to quit smoking is not systematically higher in the treated group

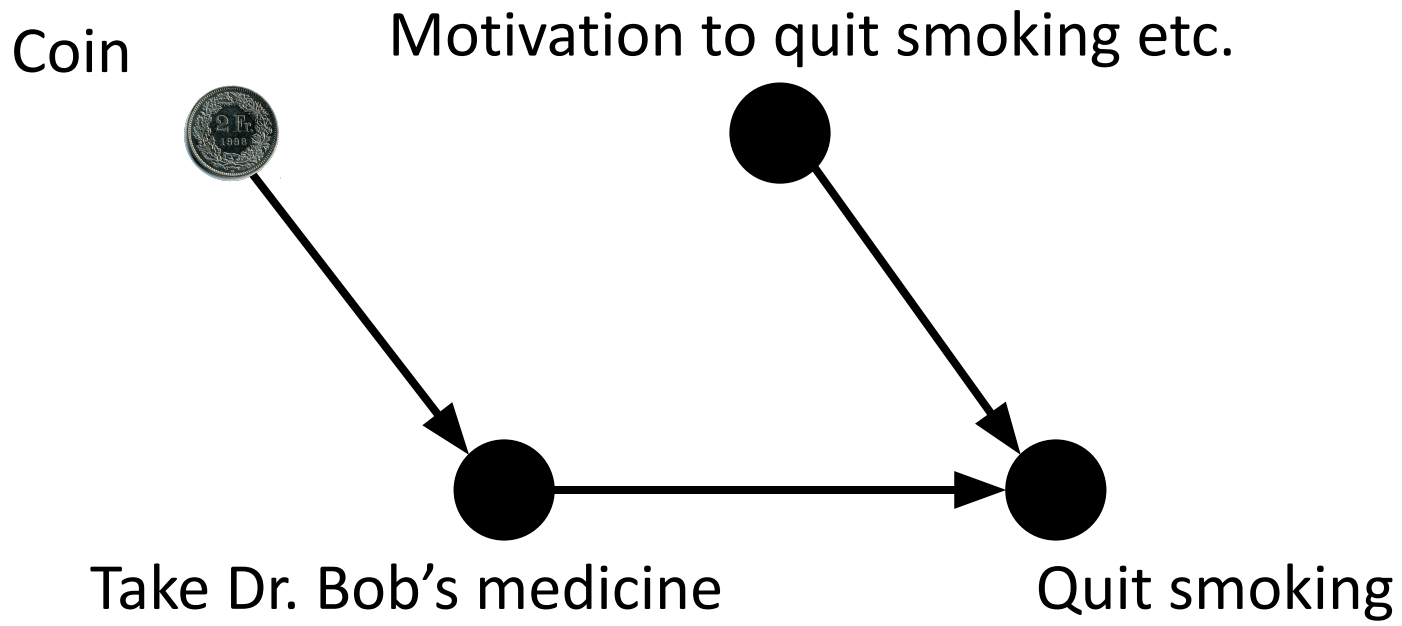
Randomized controlled experiment as a causal diagram



Randomized controlled experiment as a causal diagram



Randomized controlled experiment as a causal diagram



Limits of randomization

- Do seat belts save lives?
- Experiment:
  - Flip coin at birth to assign to treatment (always wear seat belt for entire life) or control (never wear seat belt)
  - Measure fraction of traffic deaths in each group
- Randomized experiments aren't always feasible
  - Unethical (see above), expensive, fundamentally impossible (e.g., do earthquakes decrease life spans?)
  - Most modern "big data" is "found data"
- Experiments may lead to unrealistic scenarios

Alternative: observational studies

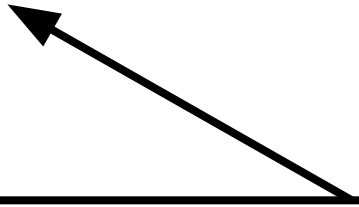
- Fundamentally different from experiment:
  - Researcher can’t control who goes to which condition
  - Researcher is merely an observer, not a tinkerer
  - Much less problematic w.r.t. ethics, price, feasibility
  - Much more problematic w.r.t. validity of conclusions
- All advantages of randomized experiment are gone
  - Subjects self-select to be treated
  - Treatment assignment and response may be caused by same hidden correlate (a.k.a. confounders; e.g., motivation to quit smoking)

Example: seat belts revisited

- Recall: experiment infeasible because unethical
- Observational study:
  - Dataset: all traffic accidents in a given time span
  - Two treatment conditions:
    - Treated: seat-belt wearers
    - Control: non-seat-belt wearers
  - Compare fraction dead in treated vs. control
- What problems do you see?

- Two treatment conditions:
  - Treated: seat-belt wearers
  - Control: non-seat-belt wearers
- Compare fraction dead in treated vs. control

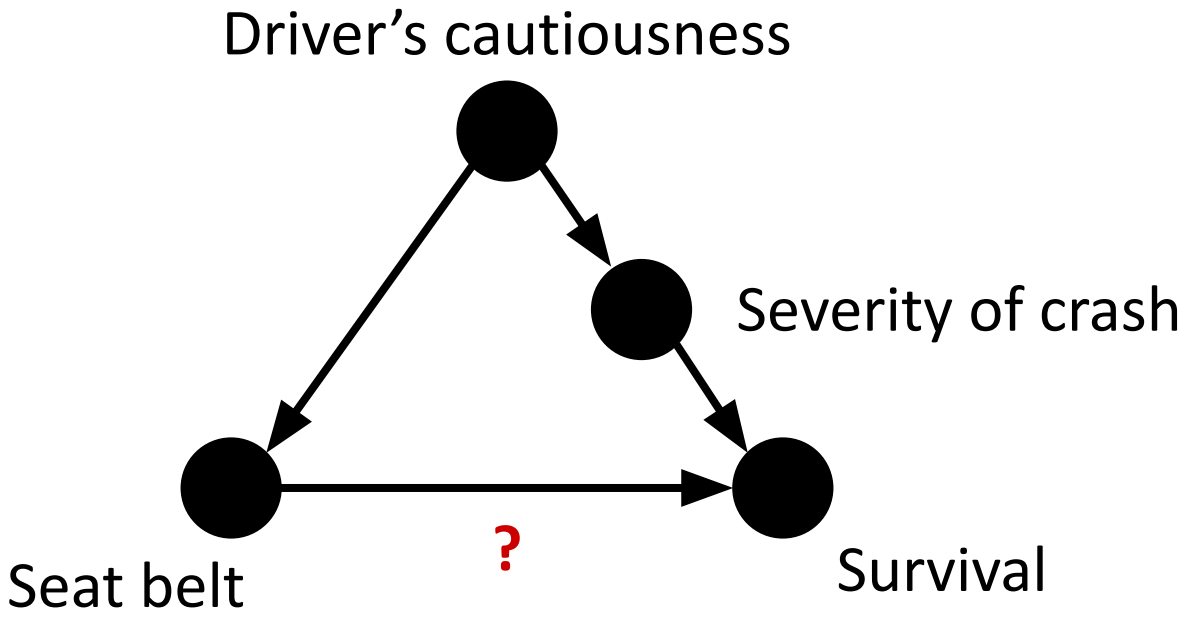
- What problems do you see?



THINK FOR A MINUTE!

(Feel free to discuss with your neighbor.)

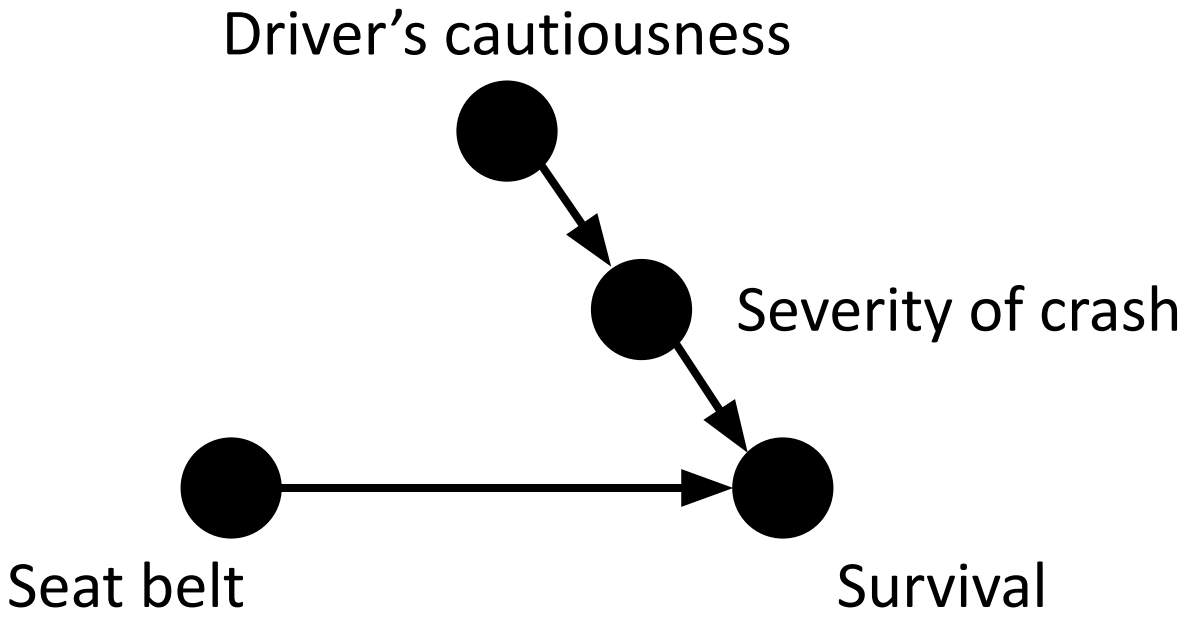
As a causal diagram



A matched observational study

- Consider only a particular subset of accident cars:
  - 2 people in car: driver + passenger
  - Exactly one of them died in accident
  - Exactly one of them wore seat belt at time of accident (i.e., 1 treated + 1 control per car)
- As before: compare fraction dead in treated vs. control
- New: many potential confounders are controlled for, incl. type of car, speed, severity of accident
- Fundamental concept: **matching**

As a causal diagram



Settling the seat-belt question

	Driver Passenger	Not Belted Belted	Belted Not Belted
Driver Died	Passenger Survived	189	153
Driver Survived	Passenger Died	111	363

Natural experiments

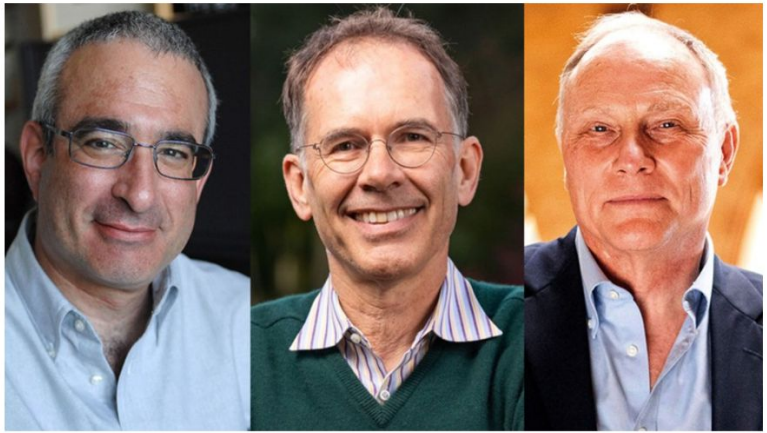
- Not researcher, but nature, “flips a coin” to decide treatment assignment
- Rosenbaum: “When investigators are especially proud, having found unusual circumstances in which treatment assignment, though not random, seems unusually haphazard, they may speak of a ‘natural experiment.’”
- Examples: twin studies, Vietnam draft, cholera in London
- Is matched seat-belt study a natural experiment?

NEWS | 13 October 2021

Nobel-winning 'natural experiments' approach made economics more robust

Joshua Angrist, Guido Imbens and David Card share the prize for finding a way to identify cause and effect in social science.

Philip Ball



Nature didn’t flip a coin for me – should I just go home and weep?

Commercial break

Nature didn’t flip a coin for me – should I go home and weep?

- No! You can still get good mileage if you’re smart about it
- Fundamental concept: **matching**
- Ideally: Pair up 2 identical people:
  - 1 treated, 1 control
  - Ex-post (vs. natural experiment: ex-ante)
- Compare outcome of treated vs. control
  - e.g., mean difference treated-minus-control
  - or regression analysis (see last lecture)

Matching

- Ideally: Pair up 2 identical people:
  - 1 treated, 1 control
- Such ideal matching usually not feasible

  - **Problem 1: Unobserved covariates:**
    You usually can't even know if two people are identical

  - **Problem 2: Combinatorial explosion:**
    (Nearly) no two people are identical

Problem 1: Unobserved covariates

- You usually can’t even know if two people are identical
- e.g., (hypothetical) gene that causes both desire to smoke and lung cancer

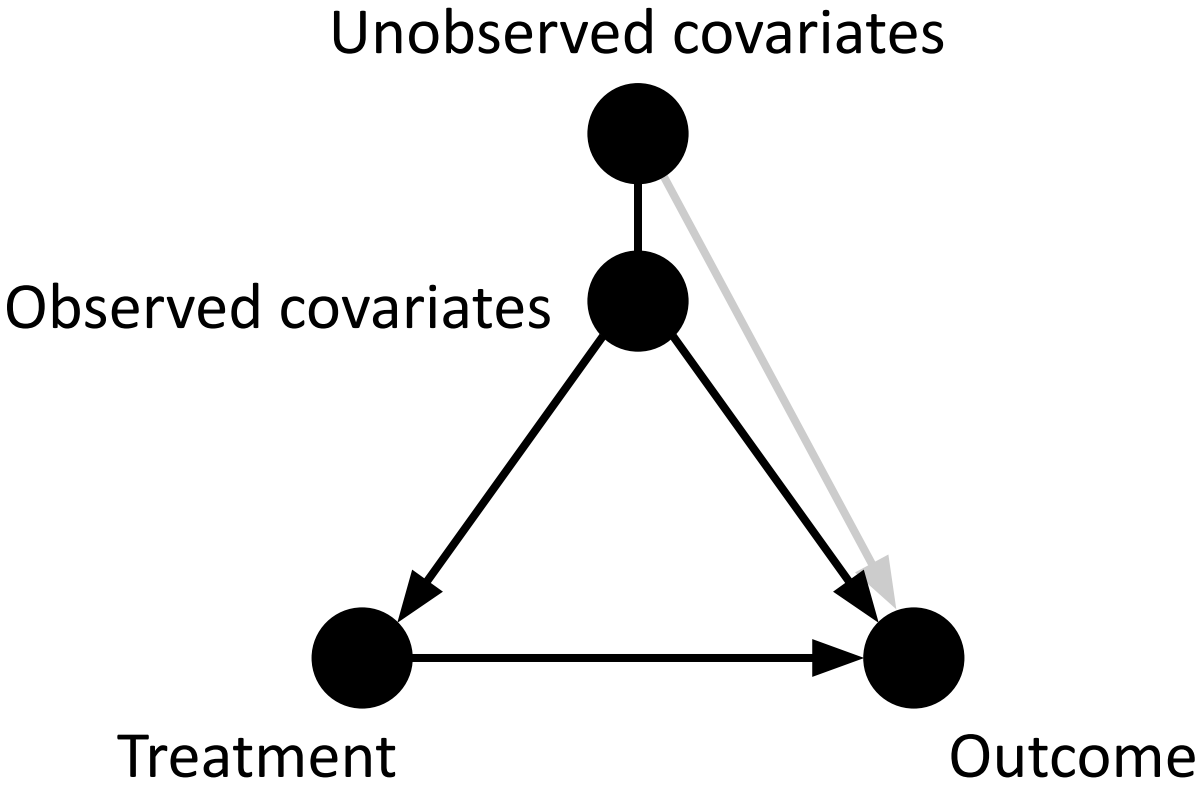
Addressing Problem 1 by ignoring it: A naive model

"People who look comparable are comparable"

or equivalently:

"Only observed covariates determine treatment assignment"

Naive model as a causal diagram



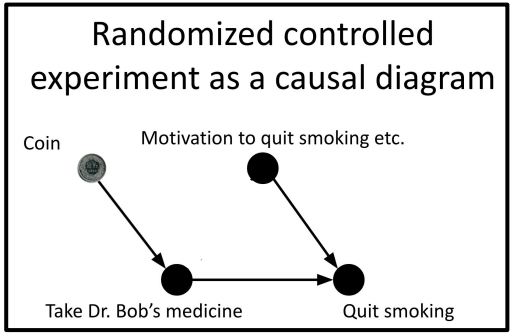
If the naive model were true…

- … you could "simulate" a randomized experiment:
  - Simply match subjects with identical observed covariates (1 treated, 1 control)
  - Subjects in a pair have the same probability to treat
  - So who gets treated is up to chance, as in experiment
  - Analysis: compare outcome for treated to outcome of control (e.g., mean difference treated-minus-control)

Problem 2!

Problem 2: Combinatorial explosion

- (Nearly) no two people are identical
- So finding two people to match is often impossible
  - Even when considering only observed covariates (as in the naive model)
- Do we really need to match people with identical **covariates?**
  - Recall “holy grail”: randomized controlled trials
  - Coin makes sure everyone has identical **probability to be treated**
  - ⟶ Let’s mimic what the coin does!



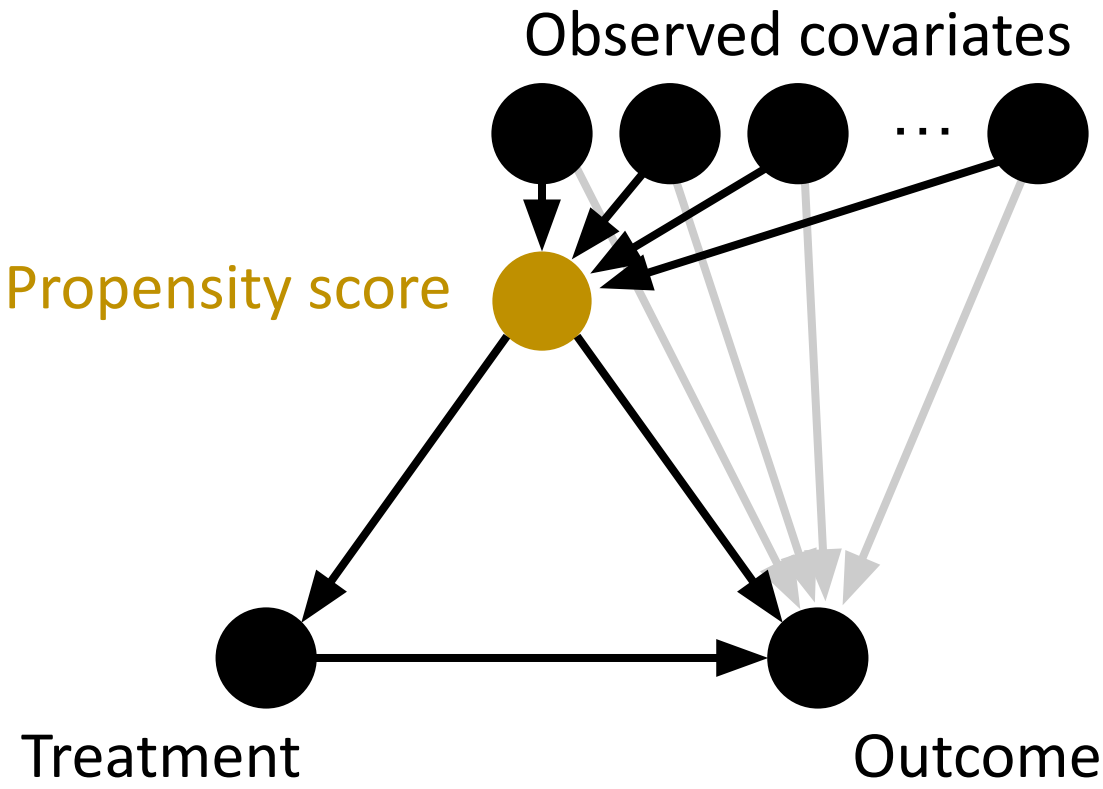
Addressing Problem 2: Propensity score

- Compress the (potentially many) observed covariates into a single number: the probability to receive the treatment (a.k.a. **propensity score**):

  Pr(subject is treated | observed covariates)

- Can be estimated from the data
  - E.g., via logistic regression (see next lecture)
  - Input: observed covariates
  - Output: treatment indicator (1 if treated, 0 if control)

Propensity score as a causal diagram



Balancing property of propensity score

- Fact: all subjects (treated and control) with equal propensity score (PS) *p* have equal distribution of observed covariates **x**:

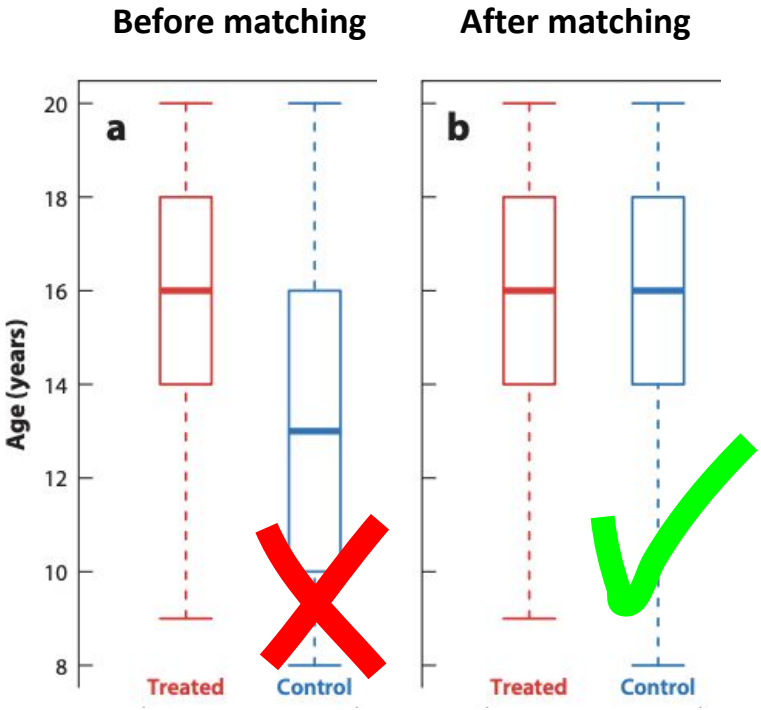
$$\Pr(\mathbf{x} \mid \text{treated} = 1, \text{PS} = p) = \Pr(\mathbf{x} \mid \text{treated} = 0, \text{PS} = p)$$

- Subjects in a matched pair might not have equal **x**, but treated and control groups will have similar distributions of **x**

Balancing property is propensity score's *raison d'être*

- There are many other methods for achieving balance
  - e.g., exact matching, Mahalanobis distance matching, coarsened exact matching, …
- You can mix and match methods (e.g., match exactly on gender, use propensity scores for other covariates)
- What eventually matters is whether you achieve balance
  - Regardless of *how you try to achieve* balance, you need to verify *that you managed to achieve* balance (p.t.o.)

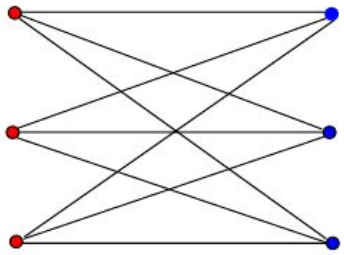
Assessing covariate balance



	Treated	Control	Unmatched
Female %	67.3	67.3	46.3
Age (mean)	14.9	14.7	13.5
Black %	14.3	15.1	28.1
Hispanic %	14.3	14.9	39.9
BMI (mean)	22.3	22.4	22.5
BMI missing %	4.1	2.2	0.0
Family income/poverty (mean)	2.4	2.4	2.1
Family income/poverty missing %	0.0	0.6	7.0
$\log_{10}$ (1 + cotinine)	0.4	0.4	0.3
Cotinine missing	2.0	2.7	11.1
Propensity score (mean, as a %)	1.7	1.7	0.7

Matching algorithms

- Goal: Match subjects into pairs (1 treated, 1 control), with identical propensity scores within each pair
- Unlikely that 2 subjects have identical propensity scores
- → Use approximate matching (remember your algo class!)
- Bipartite graph: each subject connected to all other subjects
- Edge weights: absolute (or squared) difference of propensity scores (or other matching criterion)
- Find minimum matching,
  e.g., via Hungarian algorithm



Ok, so are we done?

Let's ignore Problem 1!

Problem 1: Unobserved covariates:

You usually can't even know if two people are identical

We've been assuming the naive model:

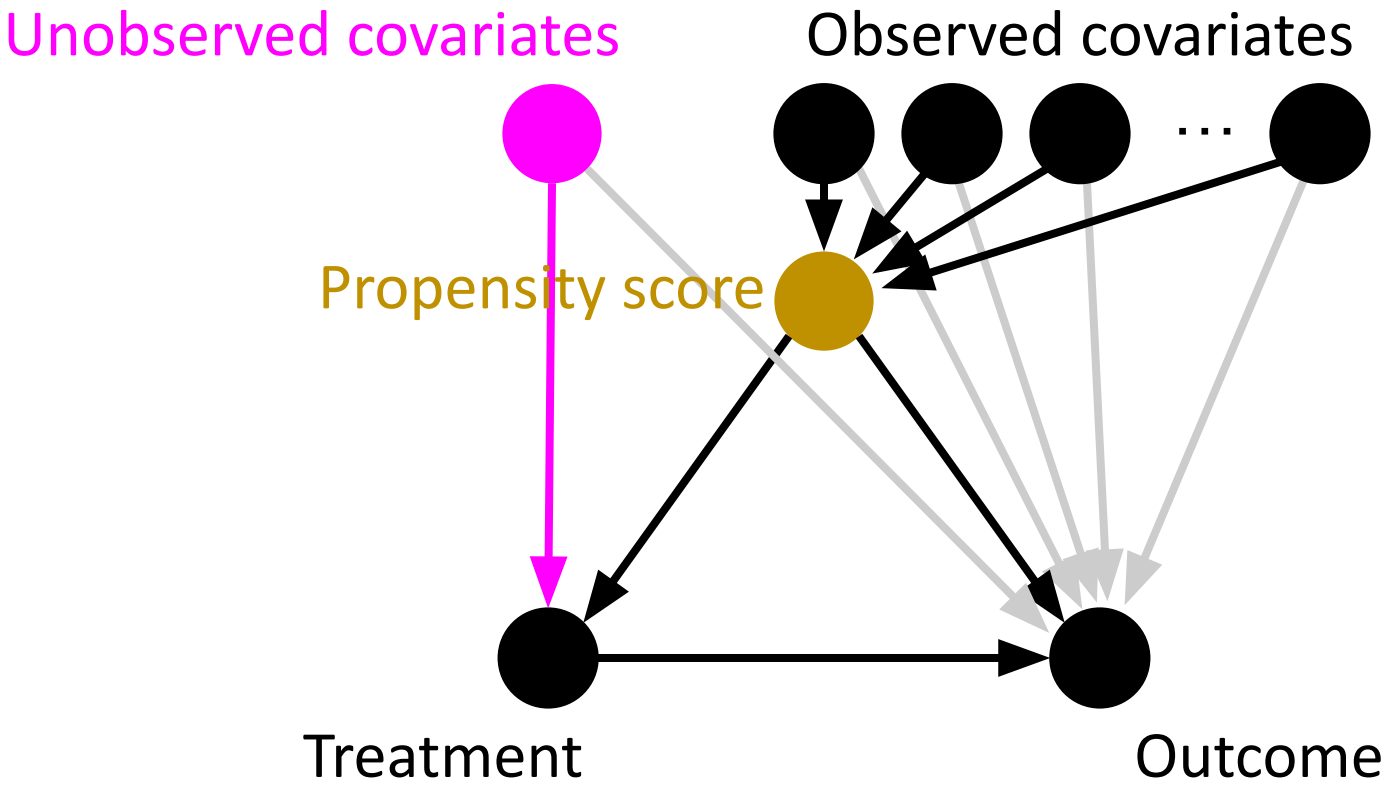
"Only observed variables determine treatment assignment"

If the naive model isn't true...

- … propensity score may differ from true probability to treat:

$$\underbrace{\Pr(\text{treated} \mid \text{observed covariates})}_{\textit{have this}} \neq \underbrace{\Pr(\text{treated} \mid \text{all covariates})}_{\textit{need this}}$$

Violated naive model as a causal diagram



If the naive model isn't true...

… you may end up matching

Person 1	Person 2
▪ Born 1948; grew up in England ▪ Married the second time ▪ 2 children ▪ successful in business ▪ Wealthy ▪ Spend their winter holidays in the Alps ▪ Likes dogs	▪ Born 1948; grew up in England ▪ Married the second time ▪ 2 children ▪ successful in business ▪ Wealthy ▪ Spend their winter holidays in the Alps ▪ Likes dogs

The naive model is easily attacked

Rosenbaum:

> It is common for a critic to argue that, in a particular study, the naïve model may be false. Indeed, it may be false. Typically, the critic accepts that the investigators matched for the observed covariates, **x**, so treated and control subjects are seen to be comparable in terms of **x**, but the critic points out that the investigators did not measure a specific covariate u, did not match for u, and so are in no position to assert that treated and control groups are comparable in terms of u. This criticism could be dismissed in a randomized experiment — randomization does tend to balance unobserved covariates — but the criticism cannot be dismissed in an observational study. This difference in the unobserved covariate u, the critic continues, is the real reason outcomes differ in the treated and control groups: it is not an effect caused by the treatment, but rather a failure on the part of the investigators to measure and control imbalances in u. Although not strictly necessary, the critic is usually aided by an air of superiority: "This would never happen in my laboratory."

The sensitivity analysis model

- Idea: Quantify the degree to which the naive model may be wrong without you having to change your (causal) conclusions
- Assume that treatment odds of identical-looking subjects (i.e., identical observed covariates **x**) may differ by up to a factor Γ
- Then reason in spirit of proof by contradiction: “To change the conclusions of my study, two identical-looking people (1 treated, 1 control) would have to have hugely different treatment odds (i.e., huge Γ). Common sense (or domain knowledge) suggests that this is not the case, so my conclusions stand.”

The sensitivity analysis model

$$\frac{1}{\Gamma} \leq \frac{\pi_k/(1-\pi_k)}{\pi_\ell/(1-\pi_\ell)} \leq \Gamma \text{ whenever } \mathbf{x}_k = \mathbf{x}_\ell\,. \qquad \Gamma \geq 1.$$

subject ℓ's (true) probability to treat

- Bounded odds ratio (OR)
  - Reason for using OR:
    OR = Pr(*k* treated | **either *k* or ℓ treated**) / Pr(ℓ treated | **either *k* or ℓ treated**)
- Sensitivity $\Gamma = 1 \longrightarrow$ naive model is true
- Sensitivity $\Gamma = 2 \longrightarrow$ subject with same observed covariates **x** up to twice as likely to be the one to receive treatment
- Sensitivity $\Gamma = \infty \longrightarrow$ void statement (a.k.a. tautology)

Example: smoking and lung cancer

- Under naive model: matching on observed covariates gives a very small *p*-value for the null hypothesis that smoking does not increase lung cancer risk (using an appropriate hypothesis test), i.e., data hard to explain w/o a causal effect
- Tobacco lobby: “The naive model isn’t true! There may be hidden (e.g., genetic) correlates that increase both the probability to enjoy smoking and the probability of lung cancer. They, not smoking, cause cancer!”

Example: smoking and lung cancer

- Under sensitivity analysis model, increasing sensitivity Γ increases the *p*-value for null hypothesis
- Anti-tobacco lobby: “But making $p > 0.05$ would require $\Gamma > 6$; i.e., the odds of being a smoker would need to be six times higher for one of two people with the exact same observed features (age, gender, education, income, ...). It’s unlikely that any unobserved covariate would have such a large effect on smoking habits. So smoking causes cancer!”


Test the West!
West
West
West
Super Geschmack. Super Preis.
Der Bundesgesundheitsminister: Rauchen gefährdet Ihre Gesundheit. Der Rauch einer Zigarette dieser Marke enthält: West 1,0 mg Nikotin und 15 mg Kondensat (Teer).
West <100> 1,2 mg N. und 15 mg K. (Durchschnittswerte nach DIN).


„... und dann fand ich ein Angebot nach meinem Geschmack.“
9,90 € 34 STK
West ORIGINAL
West SILVER
Rauchen ist tödlich
Made for good times
EXAS
ist tödlich

Two parts: mechanical vs. scientific

- Mechanical part:
  - Create pairs (1 treated + 1 control) with similar observed covariates (using exact or propensity-score matching)

- Scientific (i.e., fun) part:
  - Mitigate concerns that your findings might be caused by unobserved covariates, rather than treatment (e.g., using sensitivity analysis, ad-hoc arguments, natural experiments)

Summary

- Holy grail: randomized experiment
- When experiment not possible: observational study
- Crucial problem: treatment assignment not random (biases!)
- Semi-holy grail: natural experiment
- Matched studies: pair up treated/control based on observed covariates
- Problem: still, treatment assignment not random (biases via unobserved covariates)
- Solution: sensitivity analysis
- Keep this lecture (more here) in mind for your projects!

Feedback

Give us feedback on this lecture here:
https://go.epfl.ch/ada2023-lec6-feedback

- What did you (not) like about this lecture?
- What was (not) well explained?
- On what would you like more (fewer) details?
- …

Credits

- Much of the material is based on Paul Rosenbaum's amazing book "Design of Observational Studies", available for free here