Observational studies

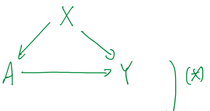
03.05.21 04:50

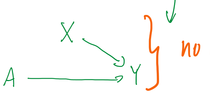
DAG 

X... pre-treatment variables

- ignorability assumption: $Y^0, Y^1 \perp\!\!\!\perp A \mid X$ implies that treatment assignment A is randomized given the set of confounders X.

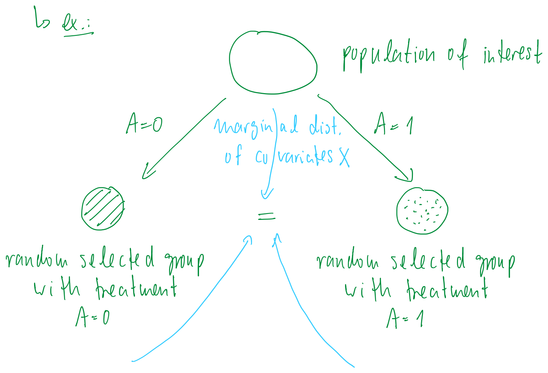
↳ in a randomized trial, treatment assignment A would be determined by a coin toss
⇒ effectively "erasing" the arrow from X to A in the DAG (*)

(1) observational: 

(2) randomization: 

} no more back-door paths from A to Y

- other benefits of randomized trial (RT)
  (1) treatment assignment A is random (unbiased e.g. like a fair coin toss)
  (2) dist. of X same in both treatment groups
  ↳ ex.:



⇒ all dist. will be almost equal : (property of a RT dataset)

⇒ in a RT, the dist. of pre-treatment variables X that affect Y are the same in both treatment groups

↳ COVARIATE BALANCE

↳ thus, if the outcome dist. ends up differing, it will not be because of differences in X

- covariate balance is dealt with at the study design phase.

Issues with RT:

a) expensive
b) treatment/exposure sometimes unethical
c) takes too much time (need to wait for outcome to happen)

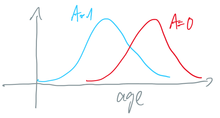
↳ Reasons for doing observational studies instead:

a) faster to get results/outcomes
b) active & passive data collection
   - prospective, active data collection
   - retrospective, passive data collection : e.g. databases (EHR, registries)
c) larger sample sizes
d) inexpensive
e) data quality lower & no uniform collection standard

observational "study" observational "dataset"

Issues with observational data:

- dist. of X differs btw. treatment groups :



↳ old people more likely to receive A=0.

Matching:

- motivation: attempt to make the observational study/data more like a randomized trial (RT) study/data.
- main idea: match individuals in treatment group (A=1) to individuals in control group (A=0) based on covariates X

⇒ b/c in RT, for any particular value of X, there should be about the same number of people in each treatment group (A=0, A=1)

↳ by matching treated to control group of the same/similar X values (e.g. age), there will be approx. same # of treated and control at any particular X (e.g. age) value.

Advantages of matching:

↳ simplifies outcome analysis a lot!
- once data is matched, we can treat the problem/data as a randomized trial (RT)
- allows us to control/deal with confounding at the design phase (i.e. w/o knowing Y)
- it can reveal a lack of overlap in the X dist.
  ↳ excluding individuals that cause lack of overlap, helps with meeting the positivity assumption of causal inference