Section 8

Lecture 3

Section 9

Causal inference from observational data

Observational data

Definition (Observational data)

A sample from a population where the treatment (exposure) is not under the control of the researcher.

That is, the treatment (exposure) of interest is not randomly assigned.

Following Robins[14], let's be slightly more abstract

- A dataset is a string of numbers.
- These data represent empirical measurements (for example, for each study subject, a series of treatments and outcomes).
- In an analysis, calculations are performed on these numbers.
- Based on the calculations, causal inference is drawn.
- "*Since the numerical strings and the computer algorithm applied to them are well-defined mathematical objects, it would be important to provide formal mathematical definitions for the English sentences expressing the investigator's causal inferences that agree well with our informal intuitive understanding*"[13].

[13] James M Robins. "Addendum to "a new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect"". In: *Computers & Mathematics with Applications* 14.9-12 (1987), pp. 923–945.

[14] Robins, "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect".

Observational studies

- In an observational study, treatment is not assigned according to randomisation, but according to someone's choice, for example the patient, the costumer or the medical doctor.
- People who choose to take treatment may be different from those who choose not to take treatment, in the sense that they have different risk of the outcome even before the decision is made. $Y^a \not\perp\!\!\!\perp A, \forall a \in \{0, 1\}$.
- The question is, can we find the characteristics L, which are associated with treatment and the outcome such that $Y^a \perp\!\!\!\perp A \mid L, \forall a \in \{0, 1\}$?
  In other words, exchangeability does no longer hold by design, but can we *assume* that it holds? What do we need to include in L for this to hold?
- Yet, humans have learned a lot from *observations*, and many scientific studies are not experiments. We have learned about *effects of* smoking, global warming, evolution, astrophysics etc.

Same data, different story

Suppose the data (identical numbers to the slide 70) were from an *observational* study (now A is not randomly assigned). The doctors tended to provide transplants ($A = 1$) to those with most severe disease ($L = 1$)

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Cyclope	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

Example continues

- Suppose first that L is the *only* outcome predictor unequally distributed between those with $A = 1$ and $A = 0$. Then $Y^a \perp\!\!\!\perp A \mid L, \forall a \in \{0, 1\}$.
- Now, suppose that the doctors not only used L to make treatment decisions, but also used smoking status, $S \in \{0, 1\}$, where smoking status is an outcome predictor. Then, $Y^a \not\perp\!\!\!\perp A \mid L, \forall a \in \{0, 1\}$.
- Thus, $Y^a \perp\!\!\!\perp A \mid L, \forall a \in \{0, 1\}$ may not hold in observational studies.
- Suppose the investigators did not measure S. Can they use the observed data to evaluate whether $Y^a \not\perp\!\!\!\perp A \mid L, \forall a \in \{0, 1\}$ holds? The answer is *no*.

More on consistency

- Consistency requires well-defined interventions.
- How do we reason about exchangeability for a treatment A that is ill-defined?
- Suppose now that our exposure (treatment) is obesity A.
- How can we identify common causes of obesity L and the outcome mortality Y? Difficult when we don't even have a sufficiently specified A
- And does positivity hold? There can be some Ls (say, related to exercise) for which nobody is obese.
- The target trial where obesity is the exposure seems to involve unreasonable interventions. How can we instantly make people non-obese? By forcing them to exercise? By doing surgery? By diet? All of these interventions may have different effects.

Section 10

Effect modification and conditional effects

THE PRECISION MEDICINE INITIATIVE

PRECISION MEDICINE | INITIATIVE | PRINCIPLES | STORIES

GO TO TOP

"Doctors have always recognized that every patient is unique, and doctors have always tried to tailor their treatments as best they can to individuals. You can match a blood transfusion to a blood type — that was an important discovery. What if matching a cancer cure to our genetic code was just as easy, just as standard? What if figuring out the right dose of medicine was as simple as taking our temperature?"

- President Obama, January 30, 2015

Effect modification

Definition (Effect modification)

We say that V is a modifier of the effect of A on Y when the average causal effect of A on Y varies across levels of V.

Since the average causal effect can be defined on different scales, effect modification depends on the scale.

Definition (Qualitative effect modification)

We say there is qualitative effect modification if the average causal effects if there exist v, v' such that the effect given $V = v$ are in the opposite direction of effects given $V = v'$.

Note that:

- V may or may not be equal to L.
- *"Effect heterogeneity across strata of V"* is often used interchangeably with *"effect modification by V"*.

Why bother with effect modification?

- So far we have focused on average causal effects.
- However, effects will often be different in different subpopulations of individuals (between men and women, Greek and Romans etc.).
- It is often of practical interest to target future intervention to subsets of the full population.

  For example, if the treatment has a positive effect in men and negative effect in women, we would like to give men and women different treatments.
- Some individuals will have different benefit of treatment than others (towards precision medicine and personalised medicine...).
- Later in the course, we will also see that this is important when we are going to *generalize* (or *transport*) effects from a study to other populations (for example, we have done an experiment in a selected population, and now we want to make decisions in another population. Therefore our question is how the intervention will work in this other population).

Illustrative experiment (trial) on heart transplant.

We may be interested in effects conditional on a baseline variable V.

	V	Y^0	Y^1
Rheia	1	0	1
Demeter	1	0	0
Hestia	1	0	0
Hera	1	0	0
Artemis	1	1	1
Leto	1	0	1
Athena	1	1	1
Aphrodite	1	0	1
Persephone	1	1	1
Hebe	1	1	0
Kronos	0	1	0
Hades	0	0	0
Poseidon	0	1	0
Zeus	0	0	1
Apollo	0	1	0
Ares	0	1	1
Hephaestus	0	0	1
Cyclope	0	0	1
Hermes	0	1	0
Dionysus	0	1	0

Here, $V = 1$ if woman, $V = 0$ if man.

Concrete example

Suppose that:

- $\mathbb{E}(Y^{a=1} \mid V = 1) = 0.6 > \mathbb{E}(Y^{a=0} \mid V = 1) = 0.4.$
- $\mathbb{E}(Y^{a=1} \mid V = 0) = 0.4 < \mathbb{E}(Y^{a=0} \mid V = 0) = 0.6.$

We conclude that there is qualitative effect modification by gender.
Treatment $A = 1$

- increases mortality in women, but
- reduces mortality in men.

Let $P(V = 0) = 0.5$. Then, the average causal effect
$\mathbb{E}(Y^{a=1}) - \mathbb{E}(Y^{a=0}) = 0.$

Identification of effects modified by V.

For simplicity suppose that V and L are disjoint; that is, they are different random variables.

1. $Y^a \perp\!\!\!\perp A \mid L, V, \forall a \in \{0, 1\}$ (Exchangeability).
2. $P(A = a \mid L = l, V = v) > 0 \ \forall a \in \{0, 1\}, \forall l \in \mathcal{L}, \forall v \in \mathcal{V}$ (Positivity).
3. $Y^a = Y$ for every unit with $A = a$ (Consistency).

How to identify effect modification

- Strategy for identification:
  1. Stratify by V.
  2. Identify the effect within each level $V = v$.
- For example, in a conditional randomised trial, an identification formula for the average causal effect of $A = a$ in the stratum defined by $V = v$ is

$$\mathbb{E}(Y^a \mid V = v) = \sum_l \mathbb{E}(Y \mid L = l, V = v, A = a) P(L = l \mid V = v).$$

Romans vs Greeks.

Consider a conditional randomised study on Heart transplant, and let V indicate whether the individual is Roman ($V = 0$) or Greek ($V = 1$)[15]

Stratum $V = 0$

	L	A	Y
Cybele	0	0	0
Saturn	0	0	1
Ceres	0	0	0
Pluto	0	0	0
Vesta	0	1	0
Neptune	0	1	0
Juno	0	1	1
Jupiter	0	1	1
Diana	1	0	0
Phoebus	1	0	1
Latona	1	0	0
Mars	1	1	1
Minerva	1	1	1
Vulcan	1	1	1
Venus	1	1	1
Seneca	1	1	1
Proserpina	1	1	1
Mercury	1	1	0
Juventas	1	1	0
Bacchus	1	1	0

[15] Hernan and Robins, *Causal inference: What if?*

Concrete example from Slide 88

Suppose that:

- $\mathbb{E}(Y^{a=1}) = 0.55$ and $\mathbb{E}(Y^{a=0}) = 0.40$.
- $\mathbb{E}(Y^{a=1} \mid V = 1) = 0.5 = \mathbb{E}(Y^{a=0} \mid V = 1) = 0.5$ (in Greeks).
- $\mathbb{E}(Y^{a=1} \mid V = 0) = 0.6 > \mathbb{E}(Y^{a=0} \mid V = 0) = 0.3$. (in Romans)

We conclude that there is effect modification by nationality.

Section 11

Interaction is different from effect modification

Interaction requires multiple interventions

- Consider two binary treatments $A \in \{0, 1\}$ and $E \in \{0, 1\}$.
  For example, chemotherapy and surgery.
- For each individual we can imagine 4 potential outcomes, that is, $Y^{a=0,e=0}$, $Y^{a=1,e=0}$, $Y^{a=0,e=1}$ and $Y^{a=1,e=1}$.

Definition (Additive interaction)

There is additive interaction if

$$\mathbb{E}(Y^{a=0,e=0}) - \mathbb{E}(Y^{a=1,e=0}) \neq \mathbb{E}(Y^{a=0,e=1}) - \mathbb{E}(Y^{a=1,e=1}).$$

Additive interaction is symmetric wrt. A and E,

$$\begin{aligned} & \mathbb{E}(Y^{a=0,e=0}) - \mathbb{E}(Y^{a=1,e=0}) \neq \mathbb{E}(Y^{a=0,e=1}) - \mathbb{E}(Y^{a=1,e=1}) \\ \implies & \mathbb{E}(Y^{a=0,e=0}) - \mathbb{E}(Y^{a=0,e=1}) \neq \mathbb{E}(Y^{a=1,e=0}) - \mathbb{E}(Y^{a=1,e=1}). \end{aligned}$$

`Remember that, unlike interactions, effect heterogeneity did only involve interventions on` A, `not the modifier` V.

Multiplicative interaction

Definition (Multiplicative interaction)

There is multiplicative interaction if

$$\frac{\mathbb{E}(Y^{a=0,e=0})}{\mathbb{E}(Y^{a=1,e=0})} \neq \frac{\mathbb{E}(Y^{a=0,e=1})}{\mathbb{E}(Y^{a=1,e=1})}.$$

Example: Interaction

- A chemotherapy, E radiation therapy, Y being cured of cancer.
- Interaction question: Is there interaction between the effect of receiving both A chemotherapy and E radiation therapy?

	E = 0	E = 1
A = 0	0.02	0.05
A = 1	0.04	0.10

Table 1: Experiment where A and E are randomised[16]

[16] Tyler J VanderWeele and Mirjam J Knol. "A tutorial on interaction". In: *Epidemiologic Methods* 3.1 (2014), pp. 33–72.

Conceptual example

- Let Y indicate being cured. There is additive interaction because

$$\mathbb{E}(Y^{a=0,e=0}) - \mathbb{E}(Y^{a=1,e=0}) \neq \mathbb{E}(Y^{a=0,e=1}) - \mathbb{E}(Y^{a=1,e=1})$$
$$0.02 - 0.04 \neq 0.05 - 0.10,$$

  but no multiplicative interaction because $\frac{0.02}{0.04} = \frac{0.5}{0.10}$.
- Suppose we had 100 versions of drug E after A was randomly assigned. Then, we would expect to cure 3 additional persons if we used all of the drug supply among those with $A = 0$. However, we would expect to cure 6 additional people if we used all the supply among those with $A = 1$.

Interaction and its relation to factorial experiments[17]

- How would you conduct an experiment to evaluate interactions between varialbes?
- We need a *factorial* design.
  - Each treatment (A and E in our example) has different levels ($A, E \in \{0, 1\}$ in our example). A factorial design consists of an equal number of replicates of all possible combinations of the levels of the factors.
  - In our Example from Slide 97, there are $2^2 = 4$ different combination of treatment levels.

[17] David Roxbee Cox and Nancy Reid. *The theory of the design of experiments*. CRC Press, 2000.

Interaction summary

- Just to say that there is an interaction on some scale is uninteresting; all it means is that both exposures have some effect on the outcome.
- Additive interaction is more relevant to public health.