Causal Inference & Causal Learning

Definition of Causal Effects & PO

Xiaolei Lin

School of Data Science
Fudan University

Mar 8, 2024

Causal inference

- Causal inference is the study of **counterfactuals**: what would happen if we were to change some aspect of the world (via intervention)?

- Biomedical science and Public Health theories are almost always causal in their nature. For an exposure A and an health outcome Y

  H_1: an increase in A causes Y to increase

- Knowing causal inference will help us

  1. understand when we can answer these questions, and
  2. design better studies to provide answers

Identification vs Estimation

- Identification of a quantity of interest (mean, effect, etc) tells us what we can learn about that quantity from the type of data available.
- Would we know this quantity if we had access to unlimited data? (No worrying about estimation uncertainty here.)
- A quantity is identified if, with infinite data, it can only take on a single value.
- Causal identification tells us what we can learn about a causal effect from the available data.

Note: Identification depends on assumptions, not on estimation strategies. If an effect can not be identified, no estimation method will recover it.

Association vs Causation

Causal Inference attempts to articulate the assumptions needed to move from conclusions about association to conclusions about causation

- **Association**: Two variables are associated if information about one tells you something about the likelihood of the other (statistical correlation)
- **Causation**: Two variables are causally related if an intervention on one has the potential to change the other

Association vs Causation

Association does not imply causation

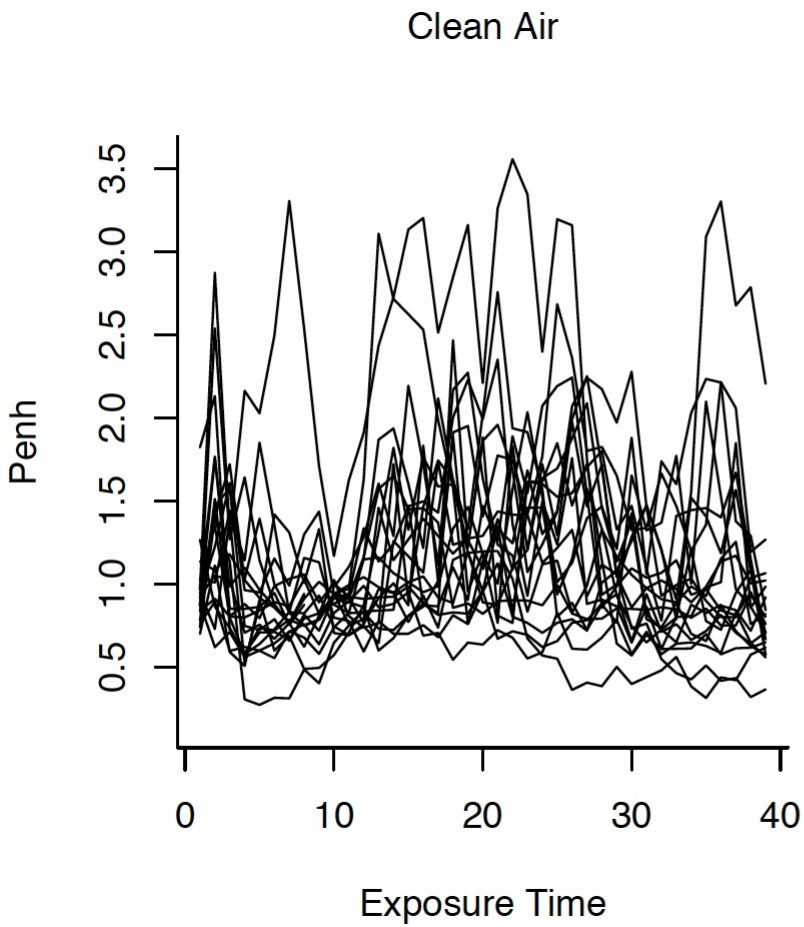
- Many research studies will appropriately qualify their findings, noting that their results concern association amongst variables and do not necessarily imply causal relationships
- However: Whenever these finding are interpreted, the interpreter will almost inevitably interpret the findings causally
- We need the discipline of causal inference to be able to articulate what is being assumed when we go about interpreting our findings causally (moving from association to causation) and to be able to discuss whether these assumptions are reasonable

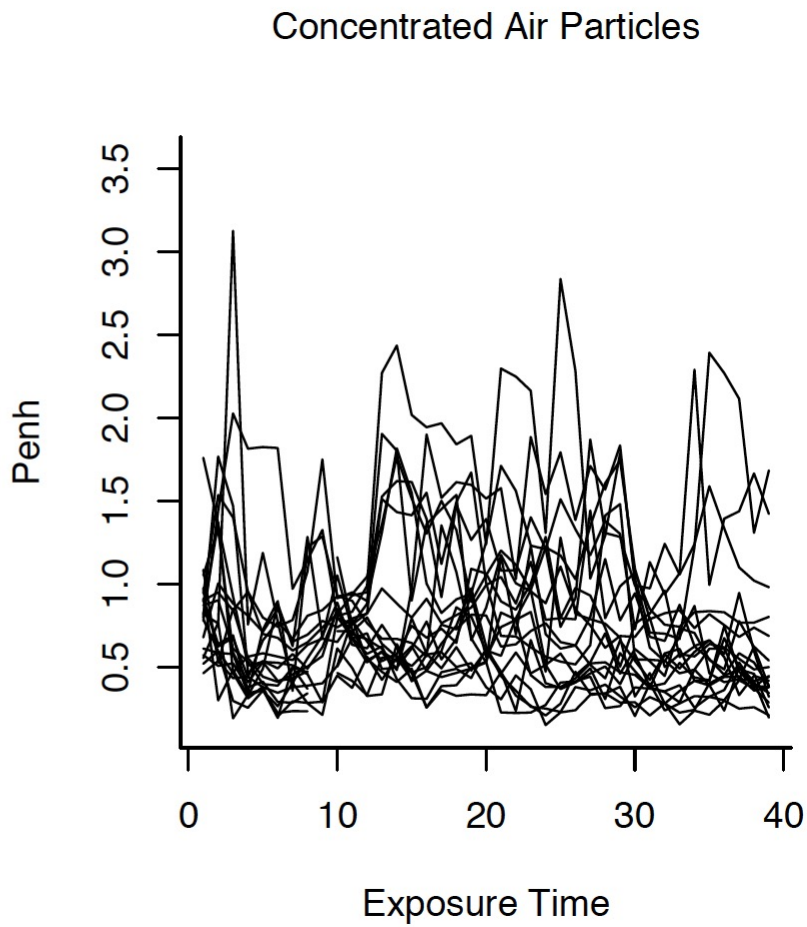
Case study 1: air pollution and cardio-pulmonary function

- Investigators are interested in assessing the effects of air pollution inhaltion on cardio-pulmonary function.
- Researchers conduct laboratory studies that expose animals to either filtered (clean) air or air pollution particles.
- For each outcome, over the course of hour-long exposures, detailed data are collected on a whole host of respiratory and cardiac endpoints.
- Determining which endpoints are affected by pollution inhalation will help establish physiological mechanisms of morbidity and mortality associated with air pollution.

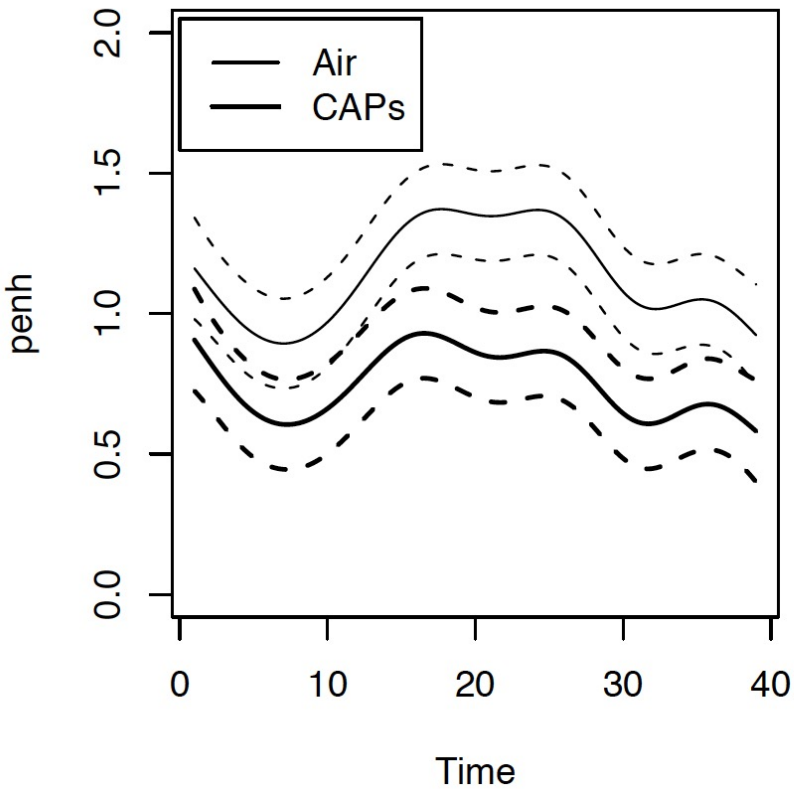
Question: How can we quantify differences between the two exposures over time?

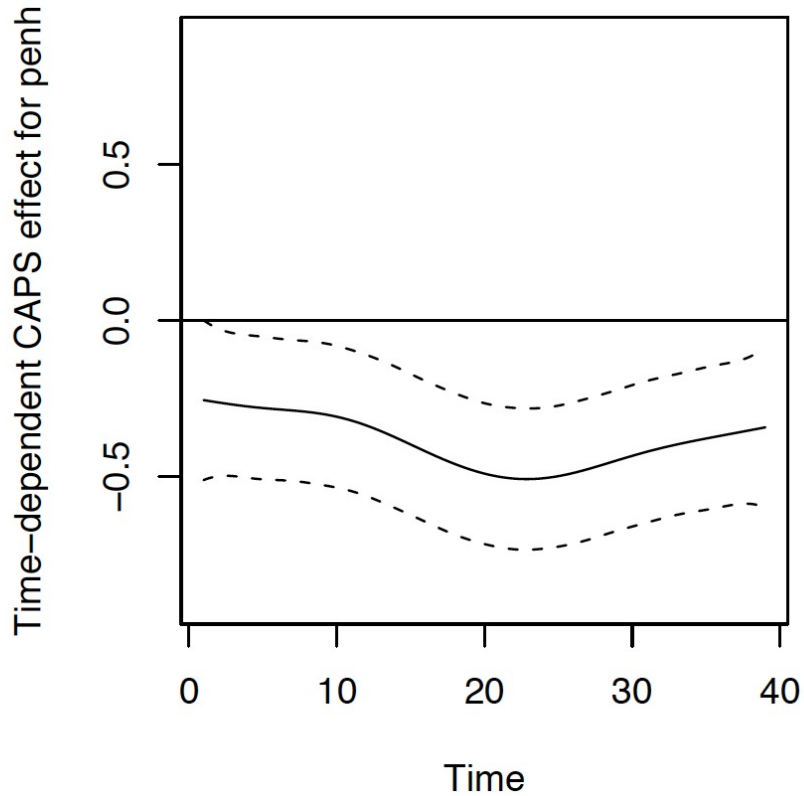
Case study 1


Clean Air
Penh
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
10
20
30
40
Exposure Time


Concentrated Air Particles
Penh
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
10
20
30
40
Exposure Time

Case study 1


Air
CAPs
penh
0.0
0.5
1.0
1.5
2.0
0
10
20
30
40
Time


Time−dependent CAPS effect for penh
−0.5
0.0
0.5
0
10
20
30
40
Time

Case study 1: air pollution and cardio-pulmonary function

- What are the best study designs and statistical models to investigate the causal effects of air pollution on cardiovascular health?
- In what way are these distinct from statistical models to study associations?
- What are the criteria to evaluate the validity of the estimated causal effect?

Association

- Define a **population** Ω and let ω denote the units which are the object of investigation.
- Define a **variable** a real-valued function defined on every unity of Ω
- A **population** of units and **variables** defined on the units are the basic elements for both associational and causal models.
- Let $Y(\omega)$ denote the **response** variable.
- Let $A(\omega)$ denote an **attribute** of the units.

Association

- The associational parameter is determined by joint distribution of Y and A over Ω

$$P(Y = y, A = a)$$

- A typical associational parameter is the regression of Y on A

$$E(Y \mid A = a) = \int_y y \frac{P(Y = y, A = a)}{P(A = a)} dy = \int_y y P(Y = y \mid A = a) d$$

- Associational inference is addressing the scientific question of description

Causation

- The effect of a cause is always relative to another cause (e.g. treatment vs control)
- For causal inference, it is critical that each unit be potentially exposable to any one of the causes.
- Let 1 and 0 denote two levels of treatment
- Let A be the assignment variable (i.e. a variable that specifies to which cause the unit in Ω is exposed)
  - In a controlled study, A is constructed by the experimenter
  - In an uncontrolled study, A is determined by factors beyond the control of the experimenter

Causation

- For causation, the value of $A(\omega)$ indicates exposure to a specific cause and for each unit (could have been different from what observed)
- For association, the value of $A(\omega)$ indicates a characteristic of the cause and for each unit (could not have been different from what observed)
- For causation the role of time of measurement is important.

Counterfactual & Potential Outcomes

- Counterfactual: The basic idea of a counterfactual is what would have happened if, contrary to fact, we had done something other than what we did?
- E.g. what would have happened if we had given treatment 1 to a particular individual instead of treatment 0?

Lewis (1973): "If c and e are two actual events such that e would not have occurred without c, then c is a cause of e." Idea of tying causation to counterfactuals goes at least as far back as Hume (1748)

Intervention

- Causality is tied to an action (or manipulation, treatment, or intervention)
- “no causation without manipulation“ - manipulation need not be performed, but should be theoretically possible
- Treatments must be plausible as a (perhaps hypothetical) intervention (Gender? Age? Race?)

Counterfactual & Potential Outcomes

- Suppose you consider taking an aspirin for your headache, and the outcome denotes whether or not you are headache free within the next hour
- As a thought experiment, you may think of two potential outcome variables or counterfactuals, either of which may be observed depending on whether or not you decide to take the aspirin.
- Let Y denote health status and A a binary treatment assignment indicator:
  - Y^1: headache status of individual if given treatment $A = 1$
  - Y^0: headache status of individual if given treatment $A = 0$

Counterfactual & Potential Outcomes

- Let Y^a denote the (counterfactual) outcome given treatment $A = a$
- For each individual we only get to observe one of Y^1 and Y^0
  - We observe Y^1 if the individual received treatment 1
  - We observe Y^0 if the individual received treatment 0
- We have no way to observe the other counterfactual outcome

Causal effects

- Y^a is the outcome that would be observed if possibly countering to fact that you followed treatment $a = 0, 1$
- The english sentence "aspirin has no causal effect on my headache outcome Y", is equivalent to a mathematical statement about potential outcomes (i.e. the causal effect is the difference in potential outcomes under two treatment assignments)

$$Y^1 = Y^0$$

- Similarly, we can think of an individual with a beneficial causal effect of aspirin if $Y^1 > Y^0$, or one with a harmful causal effect of aspirin if $Y^1 < Y^0$
- The effect of a cause for subject ω in the population Ω is given by

$$Y^1(\omega) - Y^0(\omega)$$

Missing data problem

- The fundamental problem of causal inference is that you only observe one of the two potential outcomes
- So that if in the data sample, you happen to be a person with $A = 1$, we observe Y^1 and Y^0 is missing, and for a person with $A = 0$, we observe Y^0 and Y^1 is missing
- Therefore, it is impossible to evaluate individual causal effects. This is fundamentally a missing data problem. The only difference is that the full data is never observed with probability one
- Since simultaneous observation of Y^1 and Y^0 is impossible, should we give-up? How do we obtain a substitute for the missing potential outcome?

Statistical solution

- Under some assumptions, we can still say something about population causal effects. For instance, consider the following finite population version of the previous headache example.

Individual	Y^1	Y^0
1	1	0
2	0	1
3	1	1
4	0	0
5	1	1
6	0	0
7	1	1
8	1	0
total	5	4

- Each individual has two counterfactual outcomes: Y^1 and Y^0
  - $Y = 1$ if individual is cured
  - $Y = 0$ if individual is not cured

Counterfactuals: a simple example

Individual	Y^1	Y^0
1	1	0
2	0	1
3	1	1
4	0	0
5	1	1
6	0	0
7	1	1
8	1	0
total	5	4

- how many are cured if everyone is given treatment 1
- how many are cured if everyone is given treatment 0

Counterfactuals: a simple example

Individual	Y^1	Y^0	$Y^1 - Y^0$
1	1	0	1
2	0	1	-1
3	1	1	0
4	0	0	0
5	1	1	0
6	0	0	0
7	1	1	0
8	1	0	1

- $E(Y^1) - E(Y^0) = 5/8 - 4/8 = 1/8$
- Treatment 1 is better on average. This is the average causal effect (**ACE**)

Counterfactuals: observed data

Individual	Y^1	Y^0	Trt
1	1	?	$A = 1$
2	0	?	$A = 1$
3	?	1	$A = 0$
4	0	?	$A = 1$
5	?	1	$A = 0$
6	?	0	$A = 0$
7	?	1	$A = 0$
8	1	?	$A = 1$
total	?	?	
observed	2	3	

- $E(Y \mid A = 1) - E(Y \mid A = 0) = 2/4 - 3/4 = -1/4$
- possible explanation: those who get treatment 0 may be healthier

ACE

- A commonly used population causal effect is given by the average causal effect (ACE):

$$ACE = E[Y^1 - Y^0] = 5/8 - 4/8 = 1/8$$

- We shall sometimes refer to this quantity as estimand of interest
- Note that this mean difference can be written as a difference of two means without requiring the joint distribution of counterfactuals. This is going to be key to identifying ACE
- Information on different units that can be observed can be used to gain knowledge about the causal effect

ACE

- Note that $E(Y^a), a = 0, 1$ are population quantities that are computed by taking an average of a potential outcome among all individuals in the population
- This estimand is well defined even before patients select or are selected to take or not to take the active intervention
- This is in contrast to $E(Y \mid A = a)$ which is only defined post selection, because A must be well defined
- Furthermore $E(Y \mid A = a)$ is computed by taking an average of observed outcomes only in the subset of the population with $A = a$

Population and Sample ACE

- Say that we are sampling $i = 1, ..., n$ units from our population. We can average them over a sample of units to obtain sample average causal effect (SACE)

$$SACE = \frac{1}{n} \sum_{i=1}^{n} \{Y_i^1 - Y_i^0\}$$

- or population average causal effect (PACE)

$$PACE = E(Y_i^1 - Y_i^0)$$

- Goal: use assumptions and statistical methods to use available information to recover the average causal effect

Note: SACE is also called sample average treatment effect (SATE) and PACE is also called population average treatment effect (PATE).

Why focus on the sample

- SACE is the in-sample version of the PACE.
- SACE varies over samples from the population, whereas the PACE is fixed.
- SACE still unknown because we only observe Y_i^1 or Y_i^0 for unit i
- Estimators for the SACE have lower variance

Case Study

Zeus is a patient waiting for a heart transplant. On January 1, he received a new heart. Five days later, he died. Had Zeus not received a heart transplant on January 1, he would have been alive five days later.

Another patient, Hera, also received a heart transplant on January 1. Five days later she was alive. Had Hera not received the heart on January 1, she would still have been alive five days later.

- Let A=1 denote receiving heart transplant, A=0 not receiving heart transplant
- Let Y=1 denote death, and Y=0 alive
- For Zeus, $Y^{A=1} = 1$ and $Y^{A=0} = 0$
- For Hera, $Y^{A=1} = 0$ and $Y^{A=0} = 0$
- Is there an causal effect for Zeus and Hera? Which one is potential outcome / counterfactual?

Case Study

We needed three pieces of information to define an individual causal effect

- an outcome of interest Y
- the actions $a = 1$ and $a = 0$ to be compared
- individual whose counterfactual outcomes $Y^{a=0}$ and $Y^{a=1}$ are to be compared

In this example, which outcomes cannot be observed in reality?

To define a population average causal effect, we need

- an outcome of interest Y
- the actions $a = 1$ and $a = 0$ to be compared
- a well-defined population of individuals whose outcomes $Y^{a=0}$ and $Y^{a=1}$ are to be compared

Case study

Individual	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Polyphemus	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0
total	10	10

Case Study

Take this population as of interest

- half would have died if they had received a heart transplant $Pr(Y^{a=1}) = 10/20 = 0.5$
- half would have died if they had not received a heart transplant $Pr(Y^{a=0}) = 10/20 = 0.5$
- what is the population average causal effect? $Pr(Y^{a=1}) - Pr(Y^{a=0})$

Absence of an average causal effect does not imply absence of individual causal effects.

- $Pr(Y^{a=1})$ and $Pr(Y^{a=0})$ differ for 12 patients
- 6 were harmed by heart transplantation $Pr(Y^{a=1}) - Pr(Y^{a=0}) = 1$
- another 6 were helped by heart transplantation $Pr(Y^{a=0}) - Pr(Y^{a=1}) = 1$
- sharp null: $Pr(Y^{a=0}) - Pr(Y^{a=1}) = 0$ for all individuals

Measures of causal effect

For binary outcome, the causal effects can be measured by

- causal risk difference $Pr[Y^{a=1} = 1] - Pr[Y^{a=0} = 1]$
- causal risk ratio $\frac{Pr[Y^{a=1}=1]}{Pr[Y^{a=0}=1]}$
- causal odds ratio $\frac{Pr[Y^{a=1}=1]/Pr[Y^{a=1}=0]}{Pr[Y^{a=0}=1]/Pr[Y^{a=0}=0]}$

These quantify the strength of the same causal effect on different scales and were used for different purposes.

Imagine a large population in which 3 in a million individuals would develop the outcome if treated, and 1 in a million individuals would develop the outcome if untreated.

- causal risk difference $= 0.000002$
- causal risk ratio $= 3$

Random Variability

In this example, we only had a population of 20 individuals. However, the populations of interest are typically much larger. In practice, investigators only collect information on a sample of the population of interest.

Potential sources of random error

- sampling variability ($\hat{Pr}(Y^{a=1})$ vs $Pr(Y^{a=1})$)
- non-deterministic counterfactual (Zeus has a 90% chance of dying if treated, and a 10% chance of dying if untreated)

Assignment mechanism

Individualistic

- The probability a unit is assigned to the active treatment does not depend on the covariates or potential outcomes of the other units
- Example of non-individualistic assignment: adaptive trials

Probabilistic

- Every unit has some chance of being in either treatment group, based on covariates and potential outcomes $0 < p_i(C_i, Y_i^1, Y_i^0) < 1$
- Example of non-probabilistic: perfect doctor

Unconfounded (exchangeability)

- Assignment mechanism does not depend on potential outcomes $Pr(A \mid C, Y^1, Y^0) = Pr(A \mid C)$ or $(Y^1, Y^0) \perp A \mid C$
- Example of Not unconfounded: perfect doctor

Known and controlled

- Randomized experiments: assignment mechanism is known and controlled
- Observational studies: assignment mechanism not known or controlled

Randomization

- Randomization enforces the assumption of unconfoundedness or exchangeability marginally across covariates
- Identification through randomization: Suppose we randomize our population of patients with a headache to either aspirin or no aspirin with equal probability $1/2$
- The crude association between A and Y will coincide with the average causal effect of A on Y
- This is not a coincidence but by design

Assumptions

To identify ACE, the following assumptions are required

1. Consistency
2. Stable Unit Treatment Value Assumption (SUTVA)
3. Exchangeability
4. Positivity

Consistency assumptions

- The fundamental assumption in causal inference links the observed data to the latent counterfactuals

$$Y = AY^1 + (1 - A)Y^0$$

- So that if in the data sample, you happen to be a person with $A = 1$, we observe Y^1, and vice versa for a person with $A = 0$
- The observed outcome is the counterfactual corresponding to the treatment you did indeed take
- Y^A is the factual outcome, and Y^{1-A} is the counterfactual

Consistency assumptions $Y^A = Y$

The intervention is well defined and therefore there is only one version of the potential outcome.

For example, taking aspirin is defined as

- taking, while alive, 150 mg of aspirin by mouth (or nasogastric tube if need be) daily for 5 years"

SUTVA assumption

A person's outcome is not influenced by another person's exposure or treatment

$$Y_i(A_1, A_2, ..., A_n) = Y_i(A_i)$$

1. Can you think of a possible violation of the first part of SUTVA?

2. When will the second part of SUTVA not satisfied in practice?

An individualisic and controlled assignment to a well defined intervention satisfies SUTVA and consistency.

Exchangeability assumption

- The assignment to A does not depend on the potential outcomes
- If we had the perfect doctor, we would violate this assumption

$$\{Y^1, Y^0\} \perp A$$

- unconfounded assignment satisfies the exchangeability assumption

Note: $Y^1, Y^0 \perp A \mid C$ formally defines conditional randomization, which is also called conditional exchangeability or no unmeasured confounding

Positivity assumption

- Positivity assumption holds when $Pr(A = a \mid C = c)$ is positive for all C such that $Pr(A = a)$ is non-zero
- A probabilistic assignment satisfies the positivity assumption

Note: This positivity condition is guaranteed to hold in conditionally randomized experiments.

Assumptions to identify ACE

- (CA) Consistency Assumption: $Y = Y^A$ w.p.1
- (SUTVA) Stable Unit Treatment Value Assumption: there is only one version of the potential outcome and $Y_i(A_1, A_2, ..., A_n) = Y_i(A_i)$
- (RA) Randomization Assumption: $\{Y^1, Y^0\} \perp A$ (also called exchangeability, or no unmeasured confounding)
- Positivity: $Pr(A = a)$ is positive

ACE

If CA, SUTVA, RA and positivity hold, then

$$
\begin{aligned}
ACE &= E[Y^1 - Y^0] \\
&\overset{\text{RA}}{=} E[Y^1 \mid A = 1] - E[Y^0 \mid A = 0] \\
&\overset{\text{CA}}{=} E[Y \mid A = 1] - E[Y \mid A = 0]
\end{aligned}
$$

and

$$
\begin{aligned}
E(Y^a) &= \sum_c E(Y^a \mid C = c) Pr(C = c) \\
&\overset{\text{RA}}{=} \sum_c E(Y^a \mid A = a, C = c) Pr(C = c) \\
&\overset{\text{CA+SUTVA}}{=} \sum_c E(Y \mid A = a, C = c) Pr(C = c)
\end{aligned}
$$

Violation of the assumptions

- feedback effects (cause and effect loop)
- spill over effects
- multiple versions of treatment
- continuous exposure and non-positivity

Case study: smoking and lung cancer

By the mid-1940s, it had been observed that lung cancer cases had tripled over the previous three decades. But the cause for the increase in lung cancer was unclear and not agreed upon. Possible explanations including

- Changes in air quality due to the introduction of automobile;
- Widespread expansion of paved roads that contained many carcinogens;
- Aging of the population;
- The advent of radiography;
- Better clinical awareness of lung cancer and better diagnostic methods;
- Smoking.

Case study: smoking and lung cancer

A series of observational studies reported overwhelming association between smoking and lung cancer. Some data: 36,975 pairs of heavy smoker and nonsmoker, matched by age, race, nativity, rural versus urban residence, occupational exposures to dust and fumes, religion, education, marital status, ...Of the 36,975 pairs, there were 122 discordant pairs in which exactly one person died of lung cancer. Among them

- 12 pairs in which nonsmoker died of lung cancer;
- 110 pairs in which smoker died of lung cancer.

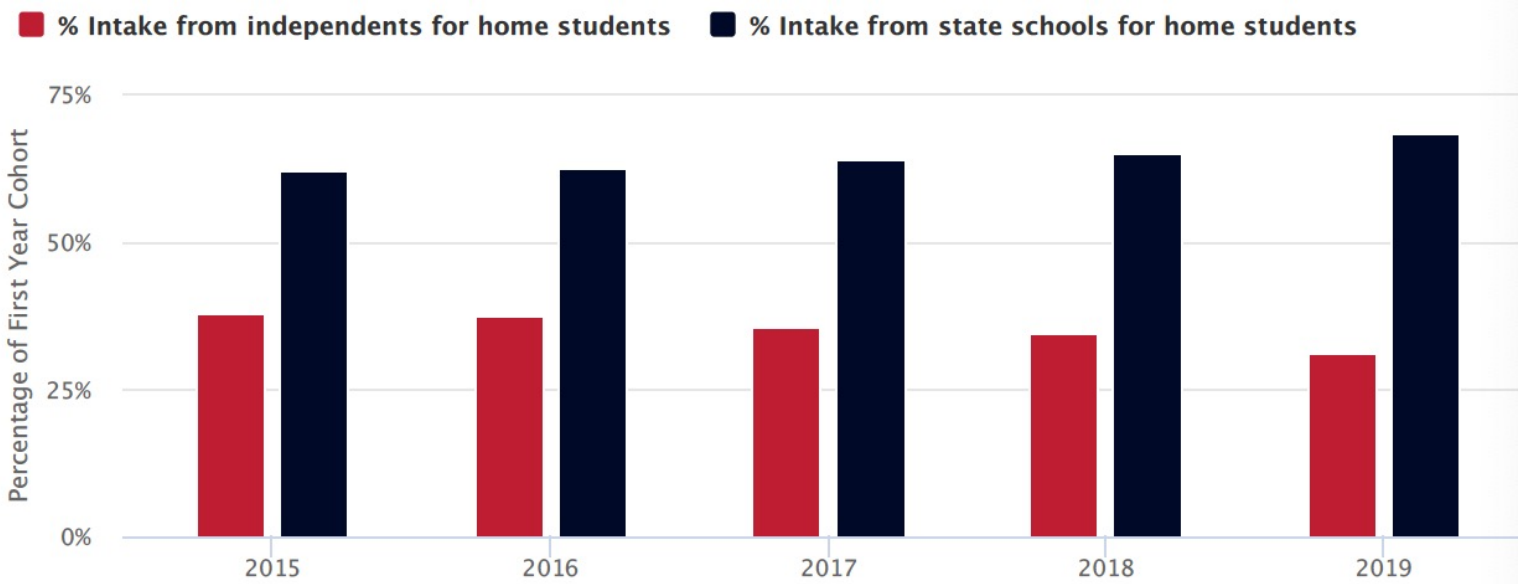
So smoking is very strongly associated with lung cancer.

Case study: smoking and lung cancer

However, Fisher strongly objected the idea that smoking is carcinogenic. Fisher demonstrated evidence of a gene that is associated with both smoking and lung cancer. He argued that differences in genetic make-up between those classes would naturally be associated with differences of disease incidence without the disease being causally connected with smoking.

We now know Fisher was wrong. His criticism was logical, but the association between smoking and lung cancer is simply too strong to be explained away by different genetic make-ups. Some believe that his views may have been influenced by personal and professional conflicts, by his work as a consultant to the tobacco industry, and by the fact that he was himself a smoker.

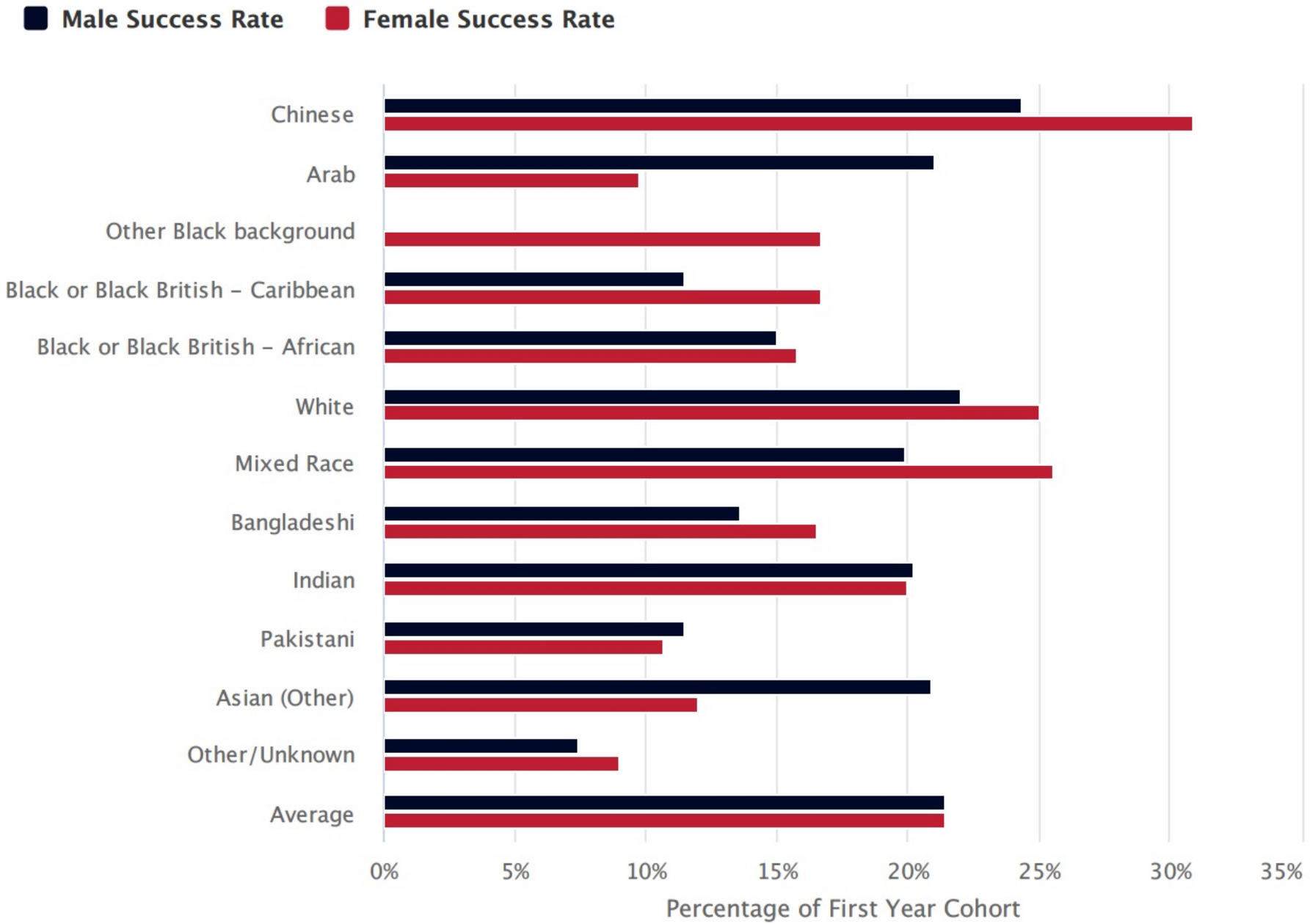
Case study: Undergraduate admissions



- 93% of pupils in England are taught in state schools,but only 68.7% were admitted into Cambridge
- Cambridge's admission biased against state schools?

Causal inference can be used to understand fairness in decisions made by human and computer algorithms.

Case study: Racial disparities

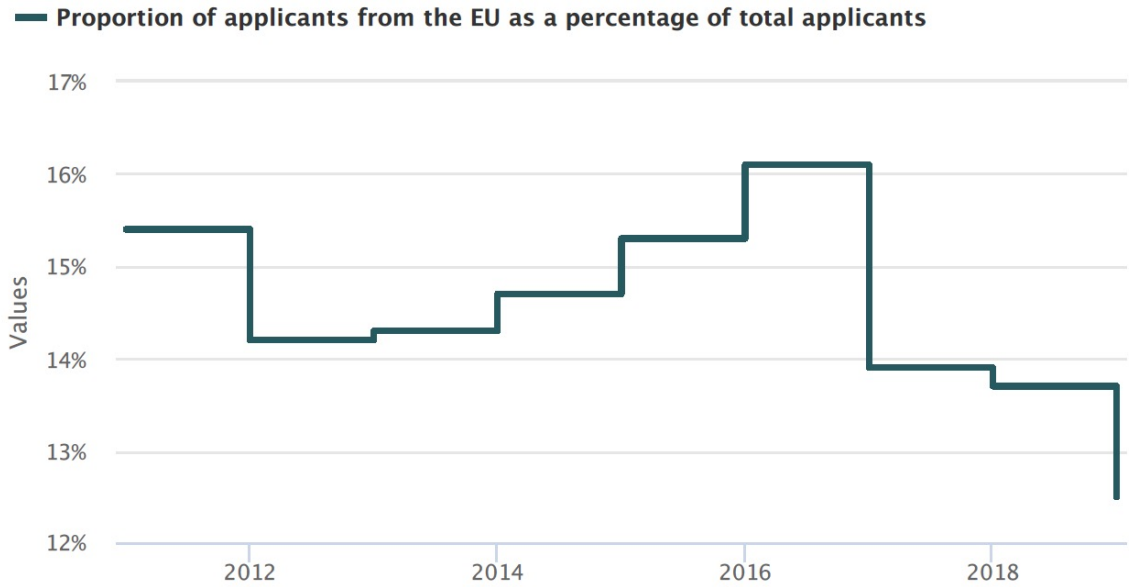


Case study: Racial disparities

- 15.1% acceptance ratio for black students, while an average 21.4% across all groups
- possible explanations for the association?

In general it is not straightforward to discuss the causal effect of race, because it is hard to "manipulation" race at birth.

Case study: Impact of Brexit



- Percentage of EU applicants declines to 12.5% (at the same time, Chinese applications increase by 33%)
- Is the steep decline in EU applications caused by Brexit (or Brexit dubiety)?

It is possible to answer this question by using a concept in causal inference called probability of causation, which is quite useful in law.

Case study: Gender difference

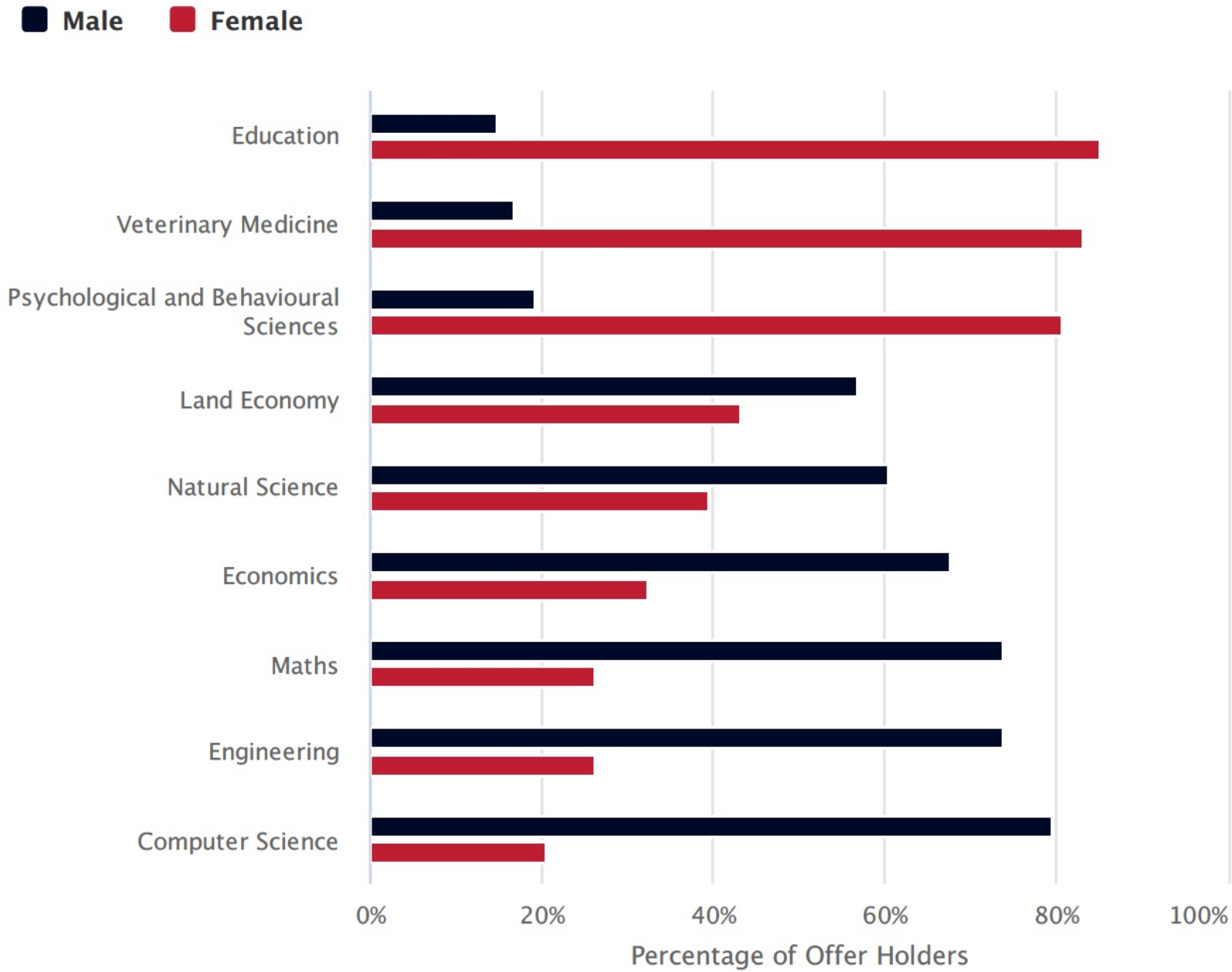
University of California, Berkeley. The admission figures for the fall of 1973 showed that men applying were more likely than women to be admitted, and the difference was so large that it was unlikely to be due to chance.[14][15]

	Men		Women	
	Applicants	**Admitted**	**Applicants**	**Admitted**
Total	8442	**44%**	4321	35%

However, when examining the individual departments, it appeared that six out of 85 departments were significantly biased against men, whereas four were significantly biased against women. In fact, the pooled and corrected data showed a "small but statistically significant bias in favor of women".[15] The data from the six largest departments are listed below, the top two departments by number of applicants for each gender italicised.

Department	**Men**		**Women**	
	Applicants	**Admitted**	**Applicants**	**Admitted**
A	***825***	62%	108	**82%**
B	***560***	63%	25	**68%**
C	325	**37%**	***593***	34%
D	417	33%	375	**35%**
E	191	**28%**	***393***	24%
F	373	6%	341	**7%**

Case study: Gender difference



Case study: Gender difference

- also called Simpson's paradox
- statistical approaches used: stratification
- This paradox is first discovered by Pearson (1899), who offered a causal explanation: "To those who persist on looking upon all correlation as cause and effect, the fact that correlation can be produced between two quite uncorrelated characters A and B by taking an artificial mixture of the two closely allied races, must come as rather a shock".

Summary

- Causal effects are function of potential outcomes
- Fundamental problem of causal inference: only one potential outcome is observed
- Causal inference as a missing data problem
- Scientific assumptions allow us to recover individual causal effects
- Statistical methods can be applied to recover PACE or SACE
- ACE is identified under a set of assumptions (some testable some not testable)
- To understand the identifiability of the causal estimand, investigators need to have clear the type of intervention and the assignment mechanism