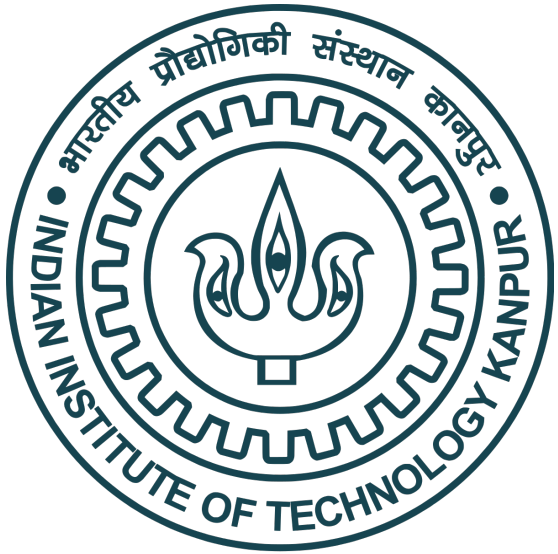
An Introduction to Causal Inference



Rahul Singh

Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Student Seminar
August 24,2020

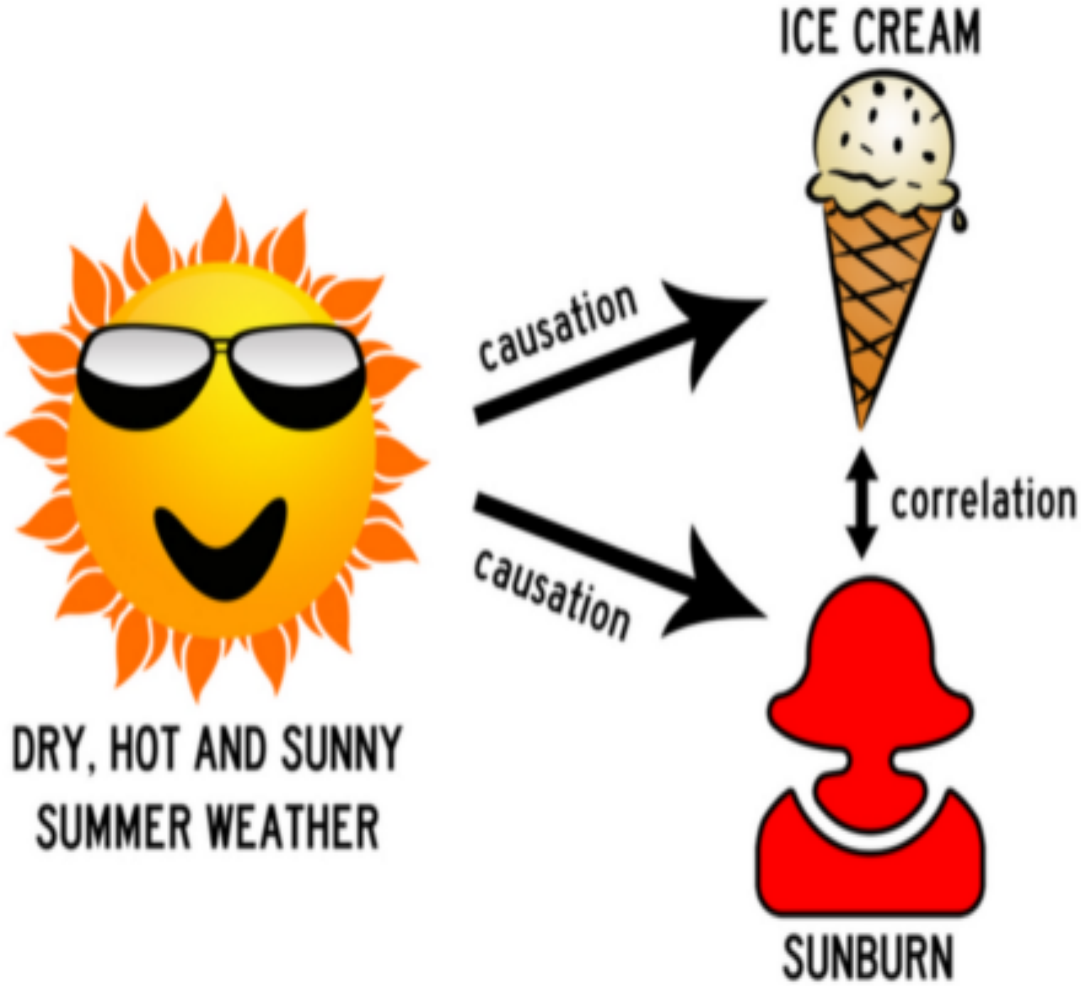
Outline

- Differentiate between causation and association.
- To understand cause and effect relationship.
- Contrast between randomised and observational study.
- Role of Balancing score in causal inference.

Superstition

Source: goodson-3-superstitions.weebly.com/cartoon.html

Higher rates of sunburn and ice cream consumption correspond. Does that mean that eating ice cream can put you at risk of sunburn? What might be the reason behind this phenomenon?



Source: towardsdatascience.com/correlation-is-not-causation-ae05d03c1f53

https://www.nejm.org/doi/full/10.1056/NEJMon1211064

The NEW ENGLAND JOURNAL of MEDICINE

OCCASIONAL NOTES FREE PREVIEW

Chocolate Consumption, Cognitive Function, and Nobel Laureates

Franz H. Messerli, M.D.

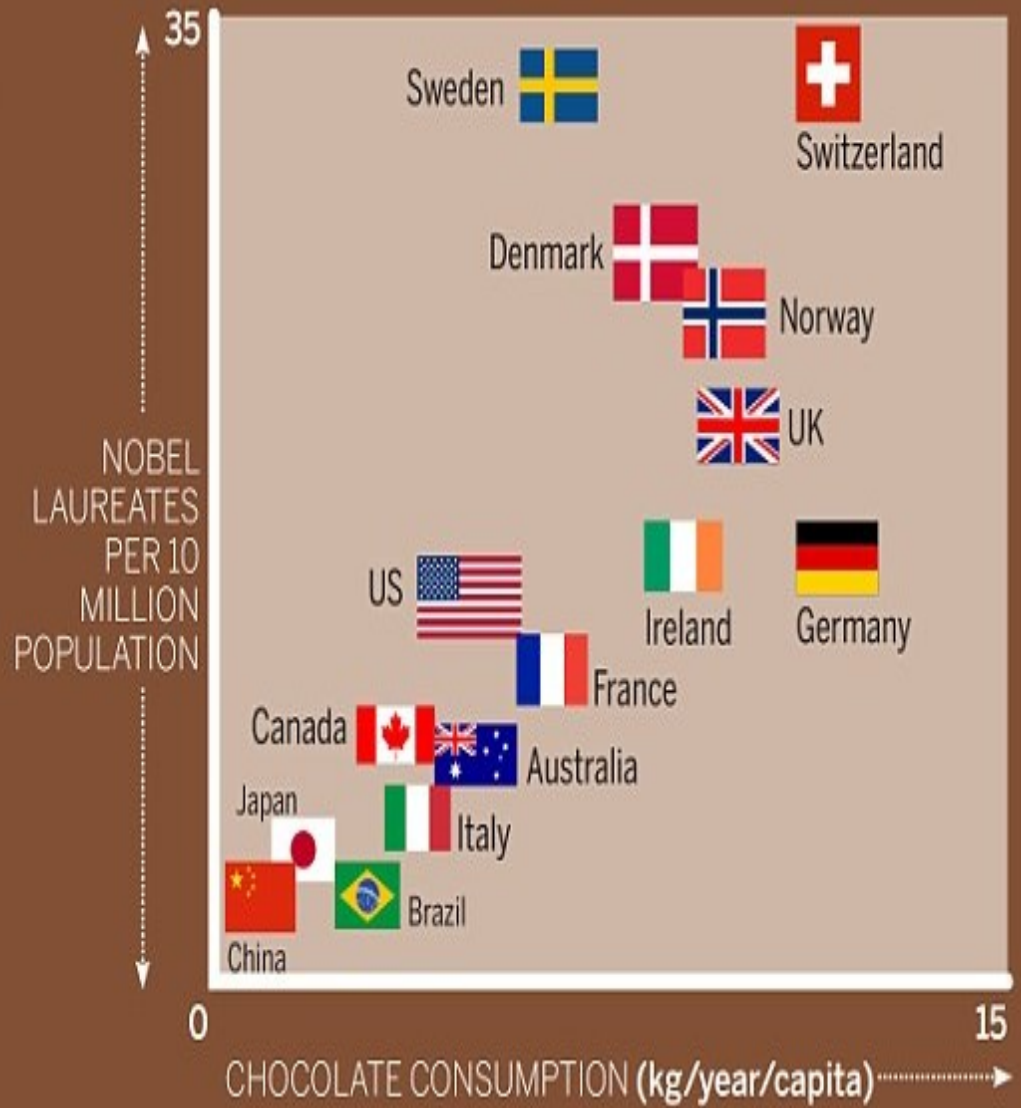
Chocolate consumption could hypothetically improve cognitive function not only in individuals but in whole populations. Could there be a correlation between a country's level of chocolate consumption and its total number of Nobel laureates per capita?

October 18, 2012
N Engl J Med 2012; 367:1562-1564
DOI: 10.1056/NEJMon1211064

Chocolate eaters win Nobel Prize

That's what the graph conclusively establishes, but that doesn't mean you can eat your way to a Nobel. Because correlations are not causations

Source: New England Journal of Medicine



Source: www.dailymail.co.uk/sitemap-articles-day 2014-09-24.xml

Post Hoc Fallacy amidst pandemic, COVID-19

Correlation between BCG vaccinations & the rates of COVID-19 morbidity and mortality.

"We found that countries without universal policies of BCG vaccination, such as Italy, the Netherlands, and the United States, have been more severely affected compared to countries with universal and long-standing BCG policies", noted the researchers led by Gonzalo Otazu, assistant professor of biomedical sciences at New York Institute of Technology.

The study concluded that a combination of reduced morbidity and mortality could make the BCG vaccination a revolutionary in the fight against corona virus.

Post Hoc Fallacy amidst pandemic, COVID-19

The Maharashtra government in India approached the Indian Council of Medical Research (ICMR) and the Drug Controller General (DCG) with an application seeking permission to use BCG vaccine for clinical trial on Covid-19 patients. The decision was taken following "positive results" of research carried out by experts at the Haffkine Institute in Mumbai.

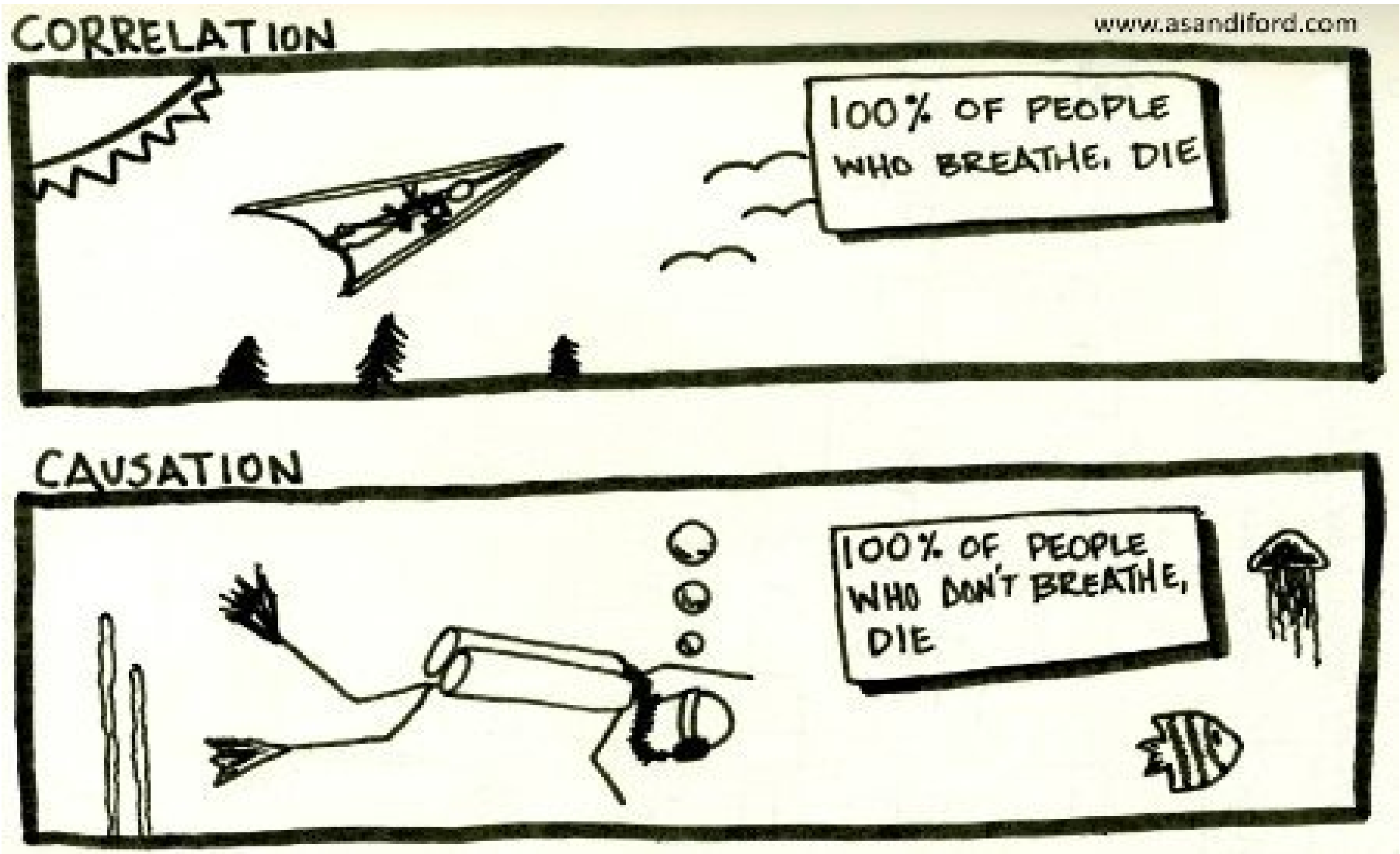
Post Hoc Fallacy amidst pandemic, COVID-19

The Maharashtra government in India approached the Indian Council of Medical Research (ICMR) and the Drug Controller General (DCG) with an application seeking permission to use BCG vaccine for clinical trial on Covid-19 patients. The decision was taken following "positive results" of research carried out by experts at the Haffkine Institute in Mumbai.

Amidst the confusion between scientists trying to find a potential conclusive base for such studies, some media houses and social media has already proved the effectiveness of BCG against COVID-19.

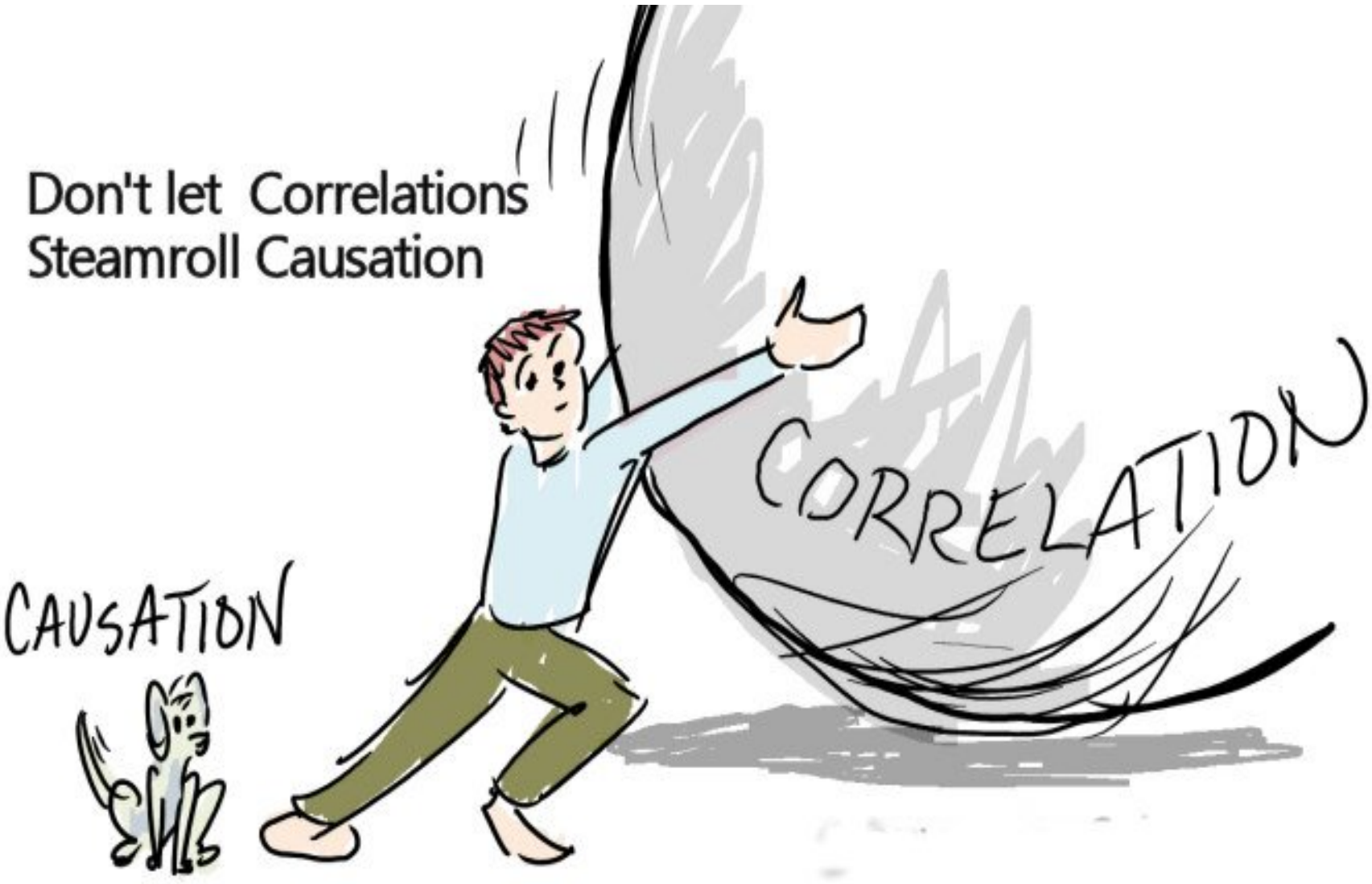
However, later, WHO clarified the doubts.

Even Researchers get trapped into such fallacies, Why ?



Source: verstaresearch.com/blog/the-dubious-roi-of-customer-satisfaction-surveys/

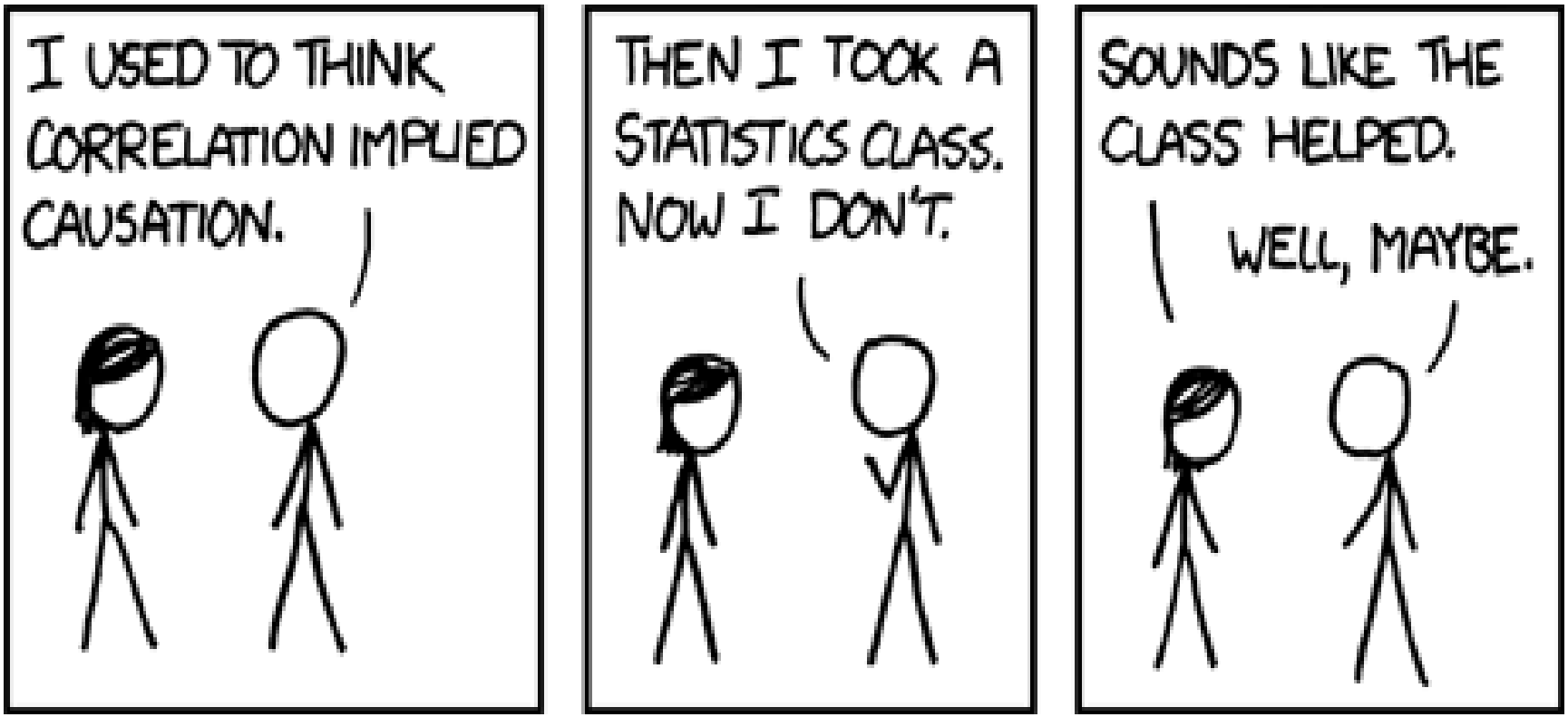
How to identify causation ?



Source: conversionsciences.com/correlation-causation-impact-ab-testing/

Counterfactual

"what would have happened if"



Source: xkcd.com/552/

Counterfactual

Consider an example:

- I have a headache;
- I take an aspirin;
- my headache goes away.
- Is it because I took the aspirin?

Counterfactual

Consider an example:

- I have a headache;
- I take an aspirin;
- my headache goes away.
- Is it because I took the aspirin?

- It is impossible to know for sure.
- We could be certain only if we could have also observed what happened to me if I had not taken the aspirin. But this control condition is impossible to observe for a single individual.

Counterfactual

Consider an example:

- I have a headache;
- I take an aspirin;
- my headache goes away.
- Is it because I took the aspirin?

- It is impossible to know for sure.
- We could be certain only if we could have also observed what happened to me if I had not taken the aspirin. But this control condition is impossible to observe for a single individual.

Estimating counterfactual is the fundamental problem in causal inference.

Standard Statistical techniques vs. Causal inference

- The aim of standard statistical analysis is to assess parameters of a distribution from samples drawn of that distribution;
- With the help of such parameters, associations among variables can be inferred, which permits the researcher to estimate probabilities of past and future events and update those probabilities in light of new information;
- These tasks are managed well by standard statistical analysis so long as experimental conditions remain the same.

Standard Statistical techniques vs. Causal inference

- The aim of standard statistical analysis is to assess parameters of a distribution from samples drawn of that distribution;
- With the help of such parameters, associations among variables can be inferred, which permits the researcher to estimate probabilities of past and future events and update those probabilities in light of new information;
- These tasks are managed well by standard statistical analysis so long as experimental conditions remain the same.
- Causal analysis goes one step further; its aim is to infer probabilities under conditions that are changing, for example, changes induced by treatments or external interventions.

Standard Statistical techniques vs. Causal inference

- There is nothing in a distribution function to tell us how that distribution would differ if external conditions were to change say from observational to experimental setup because the laws of probability theory do not dictate how one property of a distribution ought to change when another property is modified.

Standard Statistical techniques vs. Causal inference

- There is nothing in a distribution function to tell us how that distribution would differ if external conditions were to change say from observational to experimental setup because the laws of probability theory do not dictate how one property of a distribution ought to change when another property is modified.

- This information must be provided by causal assumptions which identify relationships that remain invariant when external conditions change.

Demarcation line between associational and causal concepts

- An associational concept is any relationship that can be defined in terms of a joint distribution of observed variables, and a causal concept is any relationship that cannot be defined from the distribution alone.

Demarcation line between associational and causal concepts

- An associational concept is any relationship that can be defined in terms of a joint distribution of observed variables, and a causal concept is any relationship that cannot be defined from the distribution alone.

- This distinction further implies that causal relations cannot be expressed in the language of probability and, hence, that any mathematical approach to causal analysis must acquire new notation - probability calculus is insufficient.

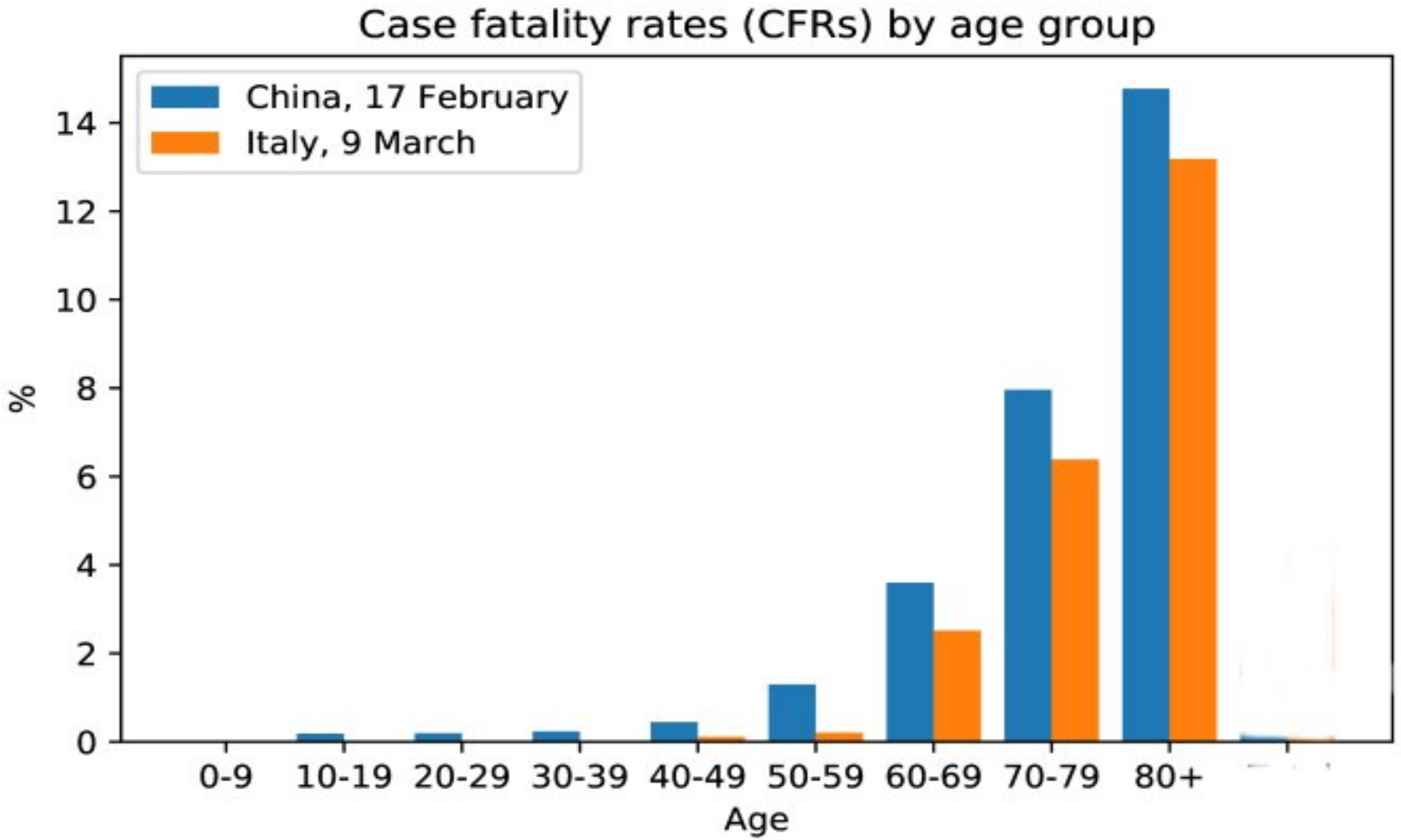
Demarcation line between associational and causal concepts

- An associational concept is any relationship that can be defined in terms of a joint distribution of observed variables, and a causal concept is any relationship that cannot be defined from the distribution alone.

- This distinction further implies that causal relations cannot be expressed in the language of probability and, hence, that any mathematical approach to causal analysis must acquire new notation - probability calculus is insufficient.

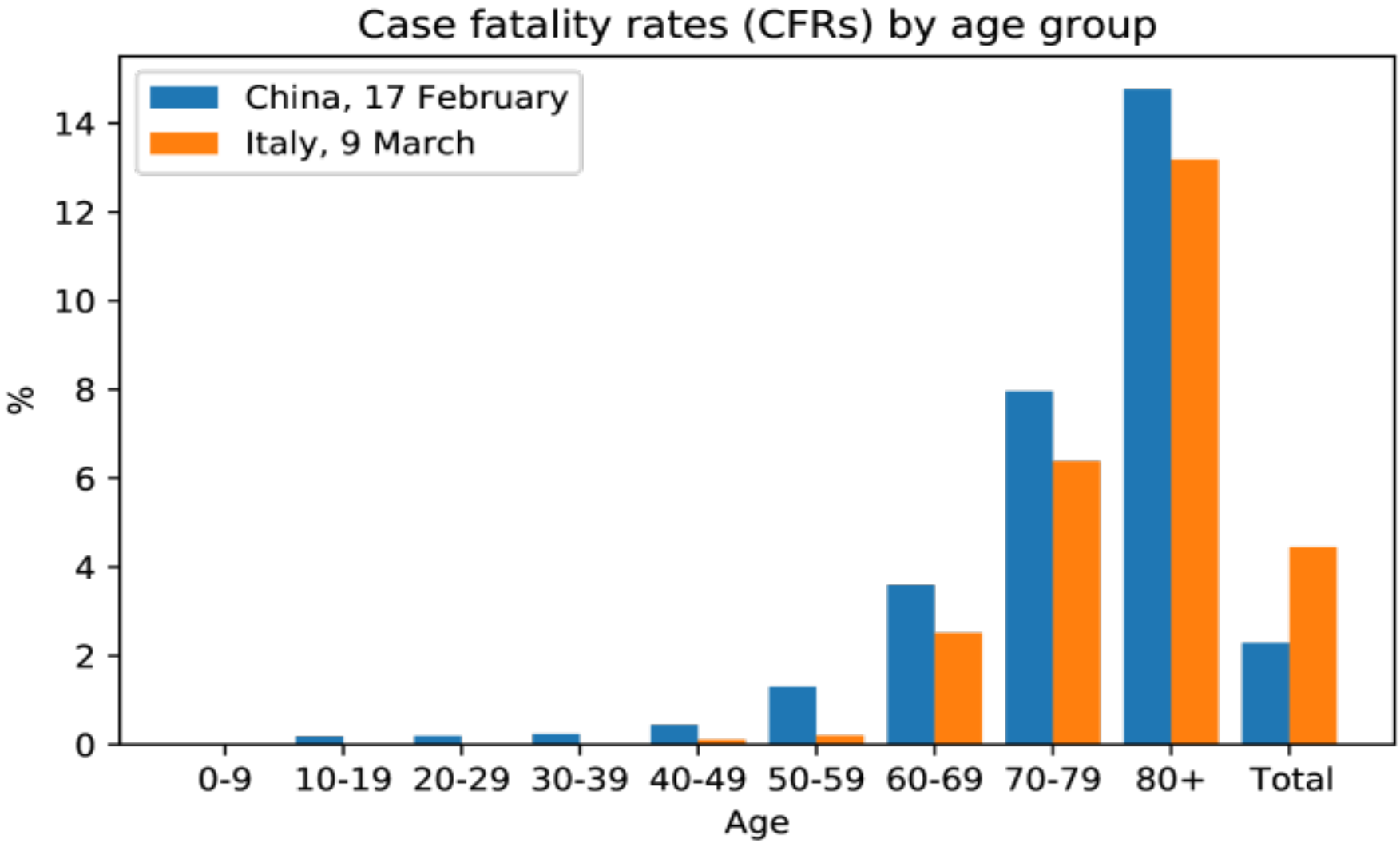
- All we can say is that two events are dependent - meaning that if we find one, we can expect to encounter the other, but we cannot distinguish statistical dependence, quantified by the conditional probability P(disease| symptom) from causal dependence, for which we have no expression in standard probability calculus.

Simpsons Paradox and Causality.



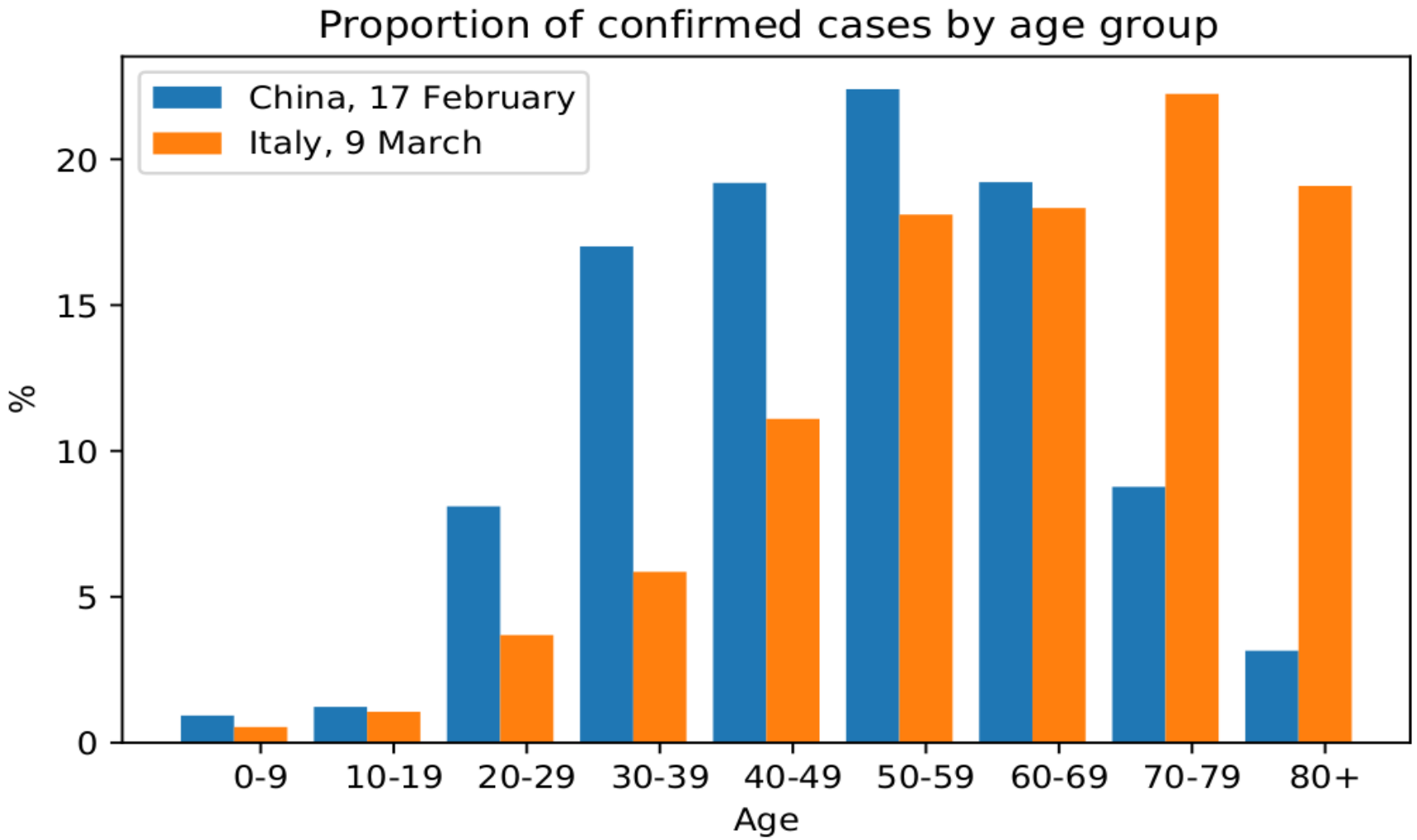
Source: von Kgelgen, J., Gresele, L., & Schlkopf, B. (2020). Simpson's paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effects. arXiv preprint arXiv:2005.07180.

Simpsons Paradox and Causality.



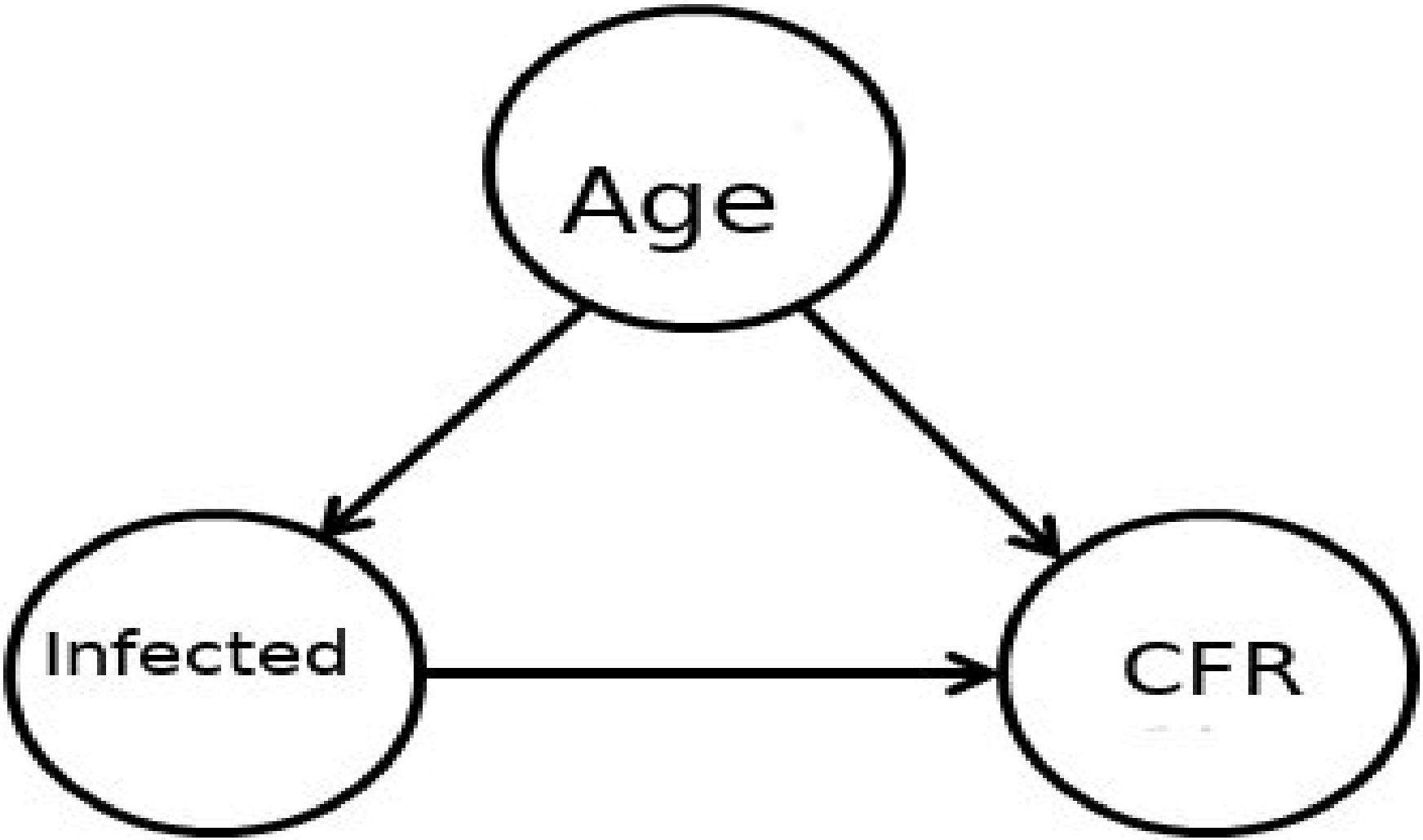
Source: von Kgelgen, J., Gresele, L., & Schlkopf, B. (2020). Simpson's paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effects. arXiv preprint arXiv:2005.07180.

Simpsons Paradox and Causality.



Source: von Kgelgen, J., Gresele, L., & Schlkopf, B. (2020). Simpson's paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effects. arXiv preprint arXiv:2005.07180.

Simpsons Paradox and Causality.



Simpsons paradox explained.

Theorem

An action A that increases the probability of an event B in each subpopulation (of C) must also increase the probability of B in the population as a whole, provided that the action does not change the distribution of the subpopulations.

In such situations, theres usually a bias that have been overlooked.

Pearl, J. (2009). Causality: Models, Reasoning, and Inference. 2nd ed. Cambridge University Press, New York.

Experimental studies vs Observational studies

	Experimental Studies	Observational Studies
Pros	• MORE validity • Can determine causality • Randomized and blinded	• May require LESS resources and/or time • LESS ethical concerns when dealing with potentially harmful exposures • Good if outcome of interest is rare
Cons	• May require MORE resources and/or time • Ethical concerns for certain exposures • Difficult if outcome being studied is rare	• LESS validity → difficult to determine causality • No randomization or blinding

Causal inference

Any conception of causation worthy of the title theory must be able to

1. represent causal questions in some mathematical language,
2. provide a precise language for communicating assumptions under which the questions need to be answered,
3. provide a systematic way of answering at least some of these questions and labeling others unanswerable, and
4. provide a method of determining what assumptions or new measurements would be needed to answer the unanswerable questions.

Notations

- Control$\rightarrow$ 0; Treatment$\rightarrow$ 1;
- In principle i^{th} unit has both responses r_{0i} and r_{1i}, $i = 1, 2, \ldots, N$;
- Causal effect is some function of r_{0i} and r_{1i}, eg. $r_{1i} - r_{0i}$, $\frac{r_{1i}}{r_{0i}}$, etc;
- Practically only one of r_{0i} and r_{1i}, is observed;
- Usually the average Causal effect of the treatment is defined as

$$\mathbb{E}(r_1) - \mathbb{E}(r_0) \tag{1}$$

Causal assumptions

Identifiability of causal effects require making some untestable assumptions.

- Stable Unit Treatment Values Assumption (SUTVA):
  The potential outcomes for any unit do not vary with the treatments assigned to other units;

Causal assumptions

Identifiability of causal effects require making some untestable assumptions.

- Stable Unit Treatment Values Assumption (SUTVA):
  The potential outcomes for any unit do not vary with the treatments assigned to other units;
- Consistency:
  The potential outcome of treatment $Z = a$, r_a, is equal to the observed outcome if the actual treatment received is $Z = a$;

Causal assumptions

Identifiability of causal effects require making some untestable assumptions.

- Stable Unit Treatment Values Assumption (SUTVA):
  The potential outcomes for any unit do not vary with the treatments assigned to other units;
- Consistency:
  The potential outcome of treatment $Z = a$, r_a, is equal to the observed outcome if the actual treatment received is $Z = a$;

- Positivity:
  For every set of covariates X, treatment assignment was not deterministic, that is $\mathbb{P}(Z = a|\ X = x) > 0,\ \forall x$;

Causal assumptions

Identifiability of causal effects require making some untestable assumptions.

- Stable Unit Treatment Values Assumption (SUTVA):
  The potential outcomes for any unit do not vary with the treatments assigned to other units;
- Consistency:
  The potential outcome of treatment $Z = a$, r_a, is equal to the observed outcome if the actual treatment received is $Z = a$;
- Positivity:
  For every set of covariates X, treatment assignment was not deterministic, that is $\mathbb{P}(Z = a|\ X = x) > 0,\ \forall x$;
- Strong ignorability:
  Given $X = x$ treament assignment mechanism does not matter.

Ignorability assumption

- It is the most important (untestable) assumption for causal inference.
- Treatment assignment is ignorable given a vector of covariates X if

$$(r_1,\ r_0) \perp\!\!\!\perp Z|\ X,\ 0 < \mathbb{P}(Z = 1|\ X) < 1,\ \text{for all } X.$$

- That is given pre treatment covariates treatment assignmenmet is stochastically independent of potential outcomes.
- This assumption is satisfied by a randomised experiment.

We can put these assumptions together to identify causal effect.

$\mathbb{E}(r|\ Z = a, X = x)$ involves only observed data.

We can put these assumptions together to identify causal effect.

$\mathbb{E}(r|\ Z = a, X = x)$ involves only observed data.

$$\begin{aligned}\mathbb{E}(r|\ Z = a, X = x) &= \mathbb{E}(r_a|\ Z = a, X = x), \text{ by consistency} \\ &= \mathbb{E}(r_a|\ X = x), \text{ by strong ignorability.}\end{aligned}$$

Balancing score

Definition

A balancing score, $b(x)$, is a function of the observed covariates x such that the conditional distribution of x given $b(x)$ is the same for treated $(z = 1)$ and control $(z = 0)$ units.

The most trivial balancing score is $b(x) = x$.

Rosenbaum & Rubin (1983)

Definition (Propensity Score)

Conditional probability of assignment to a particular treatment on a given vector of observed covariates.

Results

- Any score that is 'finer' than the propensity score is a balancing score; moreover, x is the finest balancing score and the propensity score is the coarsest;
- Treatment assignment is strongly ignorable given X, then it is strongly ignorable given any balancing score $b(X)$;

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Rosenbaum & Rubin (1983)

Results

- At any value of a balancing score, the difference between the treatment and control means is an unbiased estimate of the average treatment effect at that value of the balancing score if treatment assignment is strongly ignorable.
    - Consequently, with strongly ignorable treatment assignment, pair matching on a balancing score, subclassification on a balancing score and covariance adjustment on a balancing score can all produce unbiased estimates of treatment effects;
- Using sample estimates of balancing scores can produce sample balance on x.

Rosenbaum & Rubin (1983)

The response r, to treatment a is observed only if the unit receives treatment a, that is if $z = a$. Thus, if a randomly selected treated unit, $z = 1$, is compared to a randomly selected control unit, $z = 0$, the expected difference in response is

$$\mathbb{E}(r_1|\ z = 1) - \mathbb{E}(r_0|\ z = 0) \tag{2}$$

Observe that, $\mathbb{E}(r_1) - \mathbb{E}(r_0) \neq \mathbb{E}(r_1|\ z = 1) - \mathbb{E}(r_0|\ z = 0)$

Rosenbaum & Rubin (1983)

Suppose a specific value of the vector of covariates x is randomly sampled from the entire population of units, that is, both treated and control units together, and then a treated unit and a control unit are found both having this value for the vector of covariates. In this two-step sampling process, the expected difference in response is

$$\mathbb{E}_x\Big(\mathbb{E}(r_1|\ x, z=1) - \mathbb{E}(r_0|\ x, z=0)\Big) \tag{3}$$

If treatment assignment is strongly ignorable with $X = x$, then

$$\mathbb{E}_x\Big(\mathbb{E}(r_1|\ x, z=1) - \mathbb{E}(r_0|\ x, z=0)\Big) = \mathbb{E}_x\Big(\mathbb{E}(r_1|\ x) - \mathbb{E}(r_0|\ x)\Big)$$

Rosenbaum & Rubin (1983)

Suppose a specific value of the vector of covariates x is randomly sampled from the entire population of units, that is, both treated and control units together, and then a treated unit and a control unit are found both having this value for the vector of covariates. In this two-step sampling process, the expected difference in response is

$$\mathbb{E}_x\Big(\mathbb{E}(r_1|\ x, z=1) - \mathbb{E}(r_0|\ x, z=0)\Big) \tag{3}$$

If treatment assignment is strongly ignorable with $X = x$, then

$$\begin{aligned}\mathbb{E}_x\Big(\mathbb{E}(r_1|\ x, z=1) - \mathbb{E}(r_0|\ x, z=0)\Big) &= \mathbb{E}_x\Big(\mathbb{E}(r_1|\ x) - \mathbb{E}(r_0|\ x)\Big) \\ &= \mathbb{E}(r_1) - \mathbb{E}(r_0).\end{aligned}$$

Balancing score matching

- This is a method that attempts to make an observational study more like randomised study;
- In some study, suppose older people are more likely to get treatment.
  - There will be more younger people with control ($Z = 0$)
  - There will be more older people with treament ($Z = 1$)
- Matching reveals lack of overlap in covariate distribution;
- There are many matching methods.

Balancing score matching example

One covariate greedy matching

Treated subjects	Available control
45	72
38	44
41	60
	63
	35
	65
	47
	54
	36
	71
	56
	27

Matches

Treated	Control
45	44
38	36
41	47

$$\text{Total distance} = 1 + 2 + 6 = 9$$

Balancing score matching example

Greedy matching is not optimal

Treated subjects	Available control
45	72
38	44
41	60
	63
	35
	65
	47
	54
	36
	71
	56
	27

Matches

Treated	Control
45	47
38	36
41	44

$$\text{Total distance} = 2 + 2 + 3 = 7$$

Propensity score example

We would like to obtain unbiased results from two groups of patients (treated and untreated) which are different in their gender structure; that is a typical example for the application of propensity score.

The computation of propensity score using contingency tables consists of the following steps:

There are 60 male patients who are treated (Z=1). So the probability that the patient is a treated man P(Z=1 and Men) is 60/200 and probability that this patient is man P(men) is 100/200.

	Treated group (Z=1)	Untreated group (Z=0)	Overall
Men	60	40	100
Women	20	80	100
Overall	80	120	200

The conditional probability (propensity score) for treated men is computed by the following formula:

$$P(Z=1 \mid \text{Men}) = \frac{P(Z=1 \text{ and Men})}{P(\text{Men})} = \frac{\frac{60}{200}}{\frac{100}{200}} = 0.6$$

The propensity score for treated women is computed in the same way:

$$P(Z=1 \mid \text{Women}) = \frac{P(Z=1 \text{ and Women})}{P(\text{Women})} = \frac{\frac{20}{200}}{\frac{100}{200}} = 0.2$$

The propensity score for untreated men:

$$P(Z=0 \mid \text{Men}) = \frac{P(Z=0 \text{ and Men})}{P(\text{Men})} = \frac{\frac{40}{200}}{\frac{100}{200}} = 0.4$$

The propensity score for untreated women:

$$P(Z=0 \mid \text{Women}) = \frac{P(Z=0 \text{ and Women})}{P(\text{Women})} = \frac{\frac{80}{200}}{\frac{100}{200}} = 0.8$$

The computation of propensity score using contingency tables have mostly educational importance but its practical usage is limited. The propensity score for effective adjustment of confounding factors should contain more than one variable and/or continuous covariates. Therefore Fig. 2a and 2b involve a more complicated example of propensity score computation, which is applicable in practical analyses and uses the approach of logistic regression.

Source: Littnerova, S., Jarkovsky, J., Parenica, J., Pavlik, T., Spinar, J., & Dusek, L. (2013). Why to use propensity score in observational studies? Case study based on data from the Czech clinical database AHEAD 200609. Cor et Vasa, 55(4), e383-e390.

Balancing score matching with multiple covariates

- NN matching may be used based on Mahalanobis distance.
  - NN matching is not optimal;
  - Optimal matching is computationally expensive.

Balancing score matching with multiple covariates

- NN matching may be used based on Mahalanobis distance.
  - NN matching is not optimal;
  - Optimal matching is computationally expensive.
- Propensity score matching may be used based on logit model;
- There are other matching methods.

Figure: Bad support

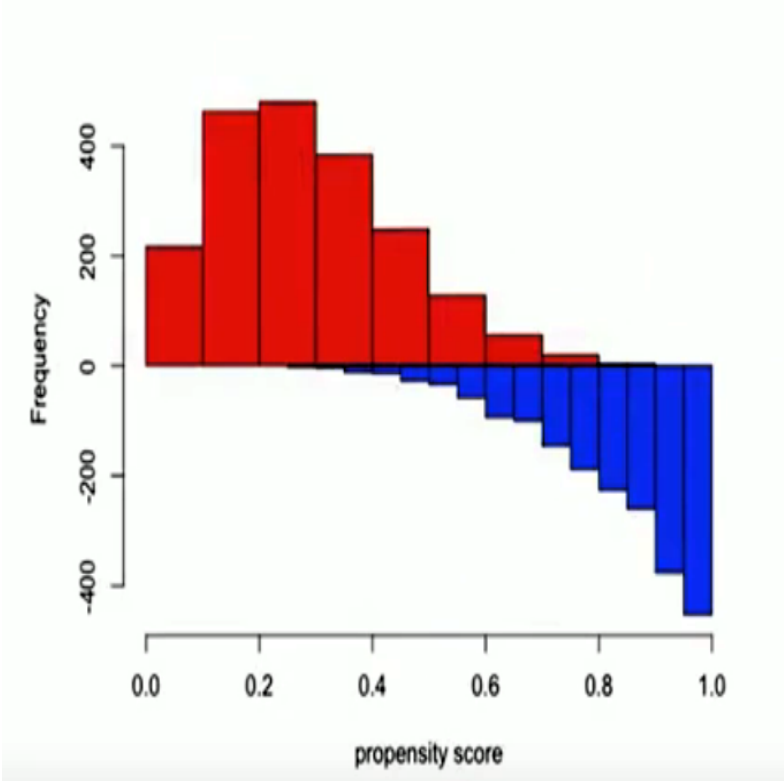
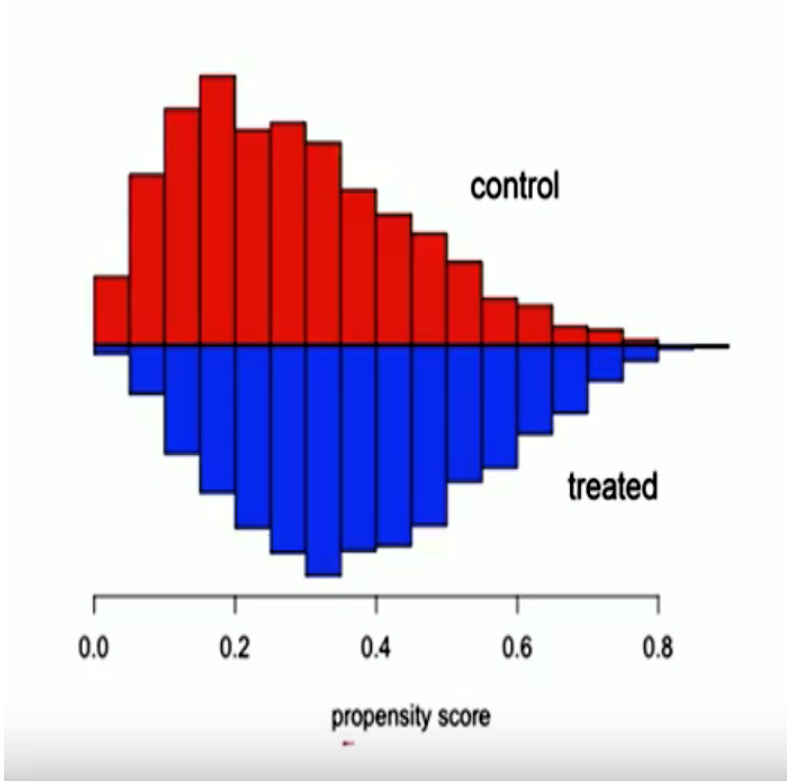
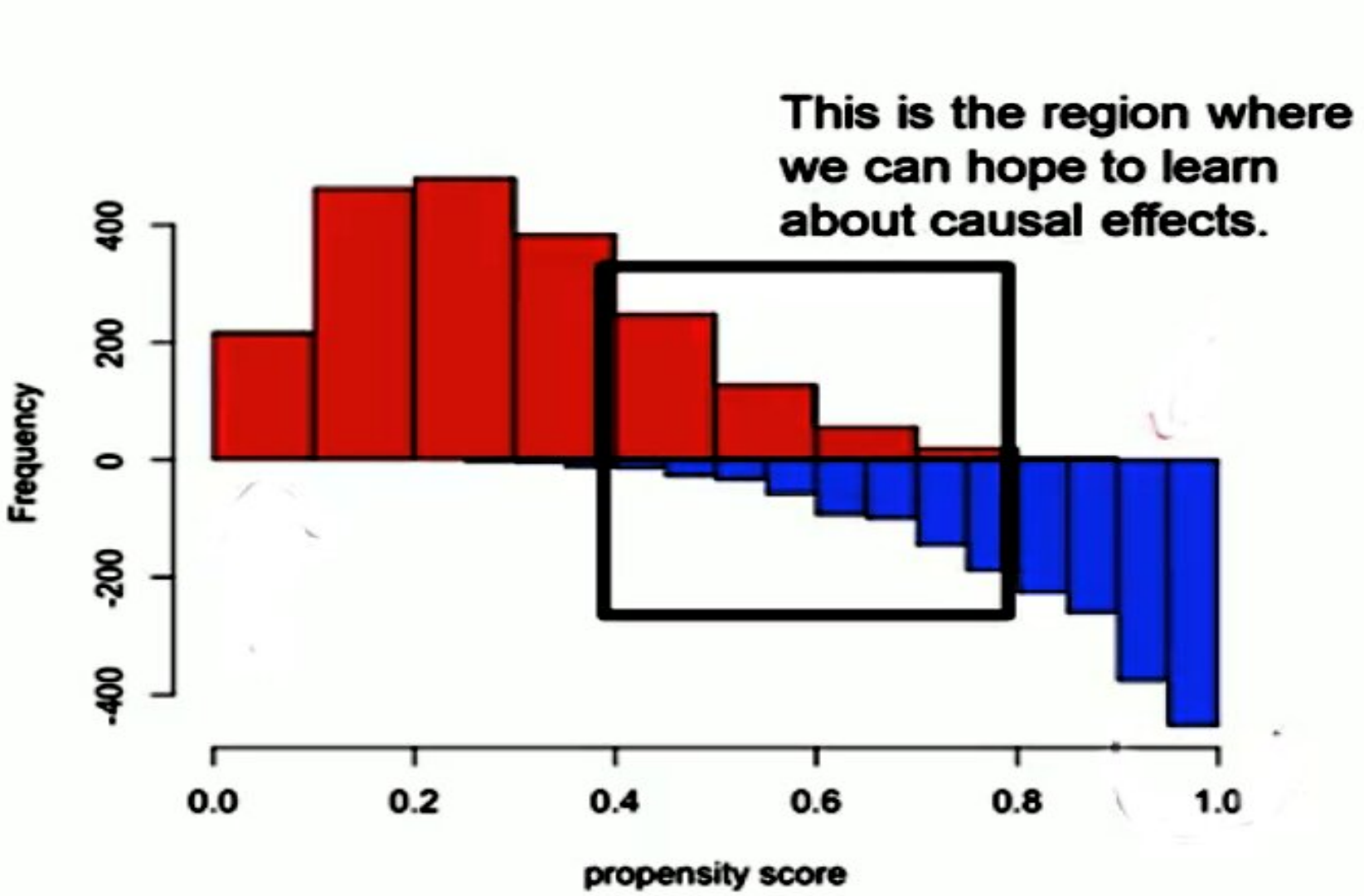


Figure: Good support




This is the region where we can hope to learn about causal effects.
Frequency
400
200
0
-200
-400
0.0
0.2
0.4
0.6
0.8
1.0
propensity score

Trimming the tails

- If there is lack of overlap, trimming the tails is an option.
  - Means removing subjects who have extreme value of propensity score;
  - For example, removing:
    * control subjects whose propensity score is less than minimum value in the treatment group,
    * treatment subjects whose propensity score is larger than maximum value in the control group.
- Trimming of tails prevents extrapolation.

Possible hidden bias

- Matching aims to achieve balance on observed covariates;
- There is no gaurantee that matching will result in balance on covariates that we didn't match on
    - If there unobserved variables are confounders then we have the hidden bias, i.e., ignorability assumption is violated.

Source: medium.com/analytics-vidhya/identify-causality-by-fixed-effects-model-585554bd9735

Thank you !