Causal Thinking

MATH-336

Mats J. Stensrud
Chair of Biostatistics
mats.stensrud@epfl.ch

Autumn 2025

Section 1

Structure of the course

Structure

- Lectures on **Mondays 10h15**.
  I will use the iPad and the blackboard.
- I encourage you to ask questions along the way!
- Moodle is our main platform.
  - Announcements.
  - Problem sheets (and solutions).
  - Links to relevant literature.
  - Link to Ed Discussions.
  - All questions about the course should be asked on Ed Discussions.
- Slides and problem sheets will be uploaded every week.
  - Problem sheets will be made available on Mondays evenings.
  - Office hours with the teaching assistants: Monday 08h15-10h00.

Exam

- One evaluation in November, 20 % of the grade.
- Written exam, 80 % of the grade.

Goal

After the course, you should:

- understand the meaning and utility of causal models,
- understand and critically evaluate causal assumptions,
- recognize whether a research question concerns causal effect,
- design a study to answer a causal question,
- be able to translate a research question to a formal causal estimand,
- critically evaluate how causal inference is drawn in practice from data,
- suggest and implement suitable causal methods in practice.

Outline of the course

- Study the **theory for causal inference** using counterfactuals,
- See how this theory can be applied in practice,
- and study its close link to the **design of experiments**.
- Derive results for identification of causal parameters in different study designs,both experiments and observational studies.
- Causal graphs will play a key role here...
- Translate practical questions to counterfactual parameters.
- Look at examples

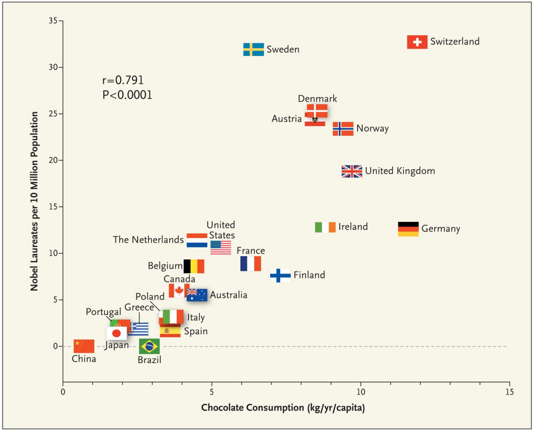
Why is this useful?

- Scientific literacy
- Understand bias in data analyses.

  This is very important for data scientists...
- Pose good (causal) questions
- Think about validation

Section 2

Motivation


r=0.791
P<0.0001
Nobel Laureates per 10 Million Population
Chocolate Consumption (kg/yr/capita)
35
30
25
20
15
10
5
0
0
5
10
15
Sweden
Switzerland
Denmark
Austria
Norway
United Kingdom
United States
The Netherlands
France
Ireland
Germany
Belgium
Canada
Finland
Poland
Australia
Greece
Portugal
Italy
Spain
Japan
China
Brazil

Warm-up example: Race and death penalty

Consider a famous data set that records the race of the defendants (D) in murder cases in Florida between 1976 and 1987.[1] The outcome is death penalty (Y).

Defendant	White	Black
Yes	53	15
No	430	176

$P(Y = 1 \mid D = w) = \frac{53}{53+430} = 0.11 > P(Y = 1 \mid D = b) = \frac{15}{15+176} = 0.08.$

Now, consider death penalty conditional on the race of the victim (V):

Victim	White	
Defendant	White	Black
Yes	53	11
No	414	37

Victim	Black	
Defendant	White	Black
Yes	0	4
No	16	139

$P(Y = 1 \mid D = w, V = w) = \frac{1}{8} < P(Y = 1 \mid D = b, V = w) = \frac{1}{5}.$

$P(Y = 1 \mid D = w, V = b) = 0 < P(Y = 1 \mid D = b, V = b) = 0.03.$

[1] From *Robin Evans*, Oxford, see also Agresti, 2002

Lessons learned from the warm-up example

- Example of Simpson's paradox (that you may be familiar with).
  By the way, paradoxes don't really exist...
- Be careful about interpreting *marginal* and *conditional* (in)dependencies.
  We will carefully (and formally) study conditional (in)dependencies in much detail in this course.
- The reason why we believe that the conditional estimates are more useful was due to **a causal story**.
  There is no statistical method that can determine the causal story from the data alone.
- How would design a study to answer the causal question "Are black defendants more likely to get death penalty just because they are black"?

Admission to universities in the USA

GRE scores are used in the admission process at American universities.

The Limitations of the GRE in Predicting Success in Biomedical Graduate School

Liane Moneta-Koehler, Abigail M. Brown, Kimberly A. Petrie, Brent J. Evans, Roger Chalkley

Published: January 11, 2017 • https://doi.org/10.1371/journal.pone.0166742

Article	Authors	Metrics	Comments	Media Coverage

Abstract

Historically, admissions committees for biomedical Ph.D. programs have heavily weighed GRE scores when considering applications for admission. The predictive validity of GRE scores on graduate student success is unclear, and there have been no recent investigations specifically on the relationship between general GRE scores and graduate student success in biomedical research. Data from Vanderbilt University Medical School's biomedical umbrella program were used to test to what extent GRE scores can predict outcomes in graduate school training when controlling for other admissions information. Overall, the GRE did not prove useful in predicating who will graduate with a Ph.D., pass the qualifying exam, have a shorter time to defense, deliver more conference presentations, publish more first author papers, or obtain an individual grant or fellowship. GRE scores were found to be moderate predictors of first semester grades, and weak to moderate predictors of graduate GPA and some elements of a faculty evaluation. These findings suggest admissions committees of biomedical doctoral programs should consider minimizing their reliance on GRE scores to predict the important measures of progress in the program and student productivity.

Here, **conditioning** (on admission) leads to an inappropriate comparison.
In this course, we will formalize how to *design studies* and *analyse data* to answer causal questions.

Let's be explicit

Unfortunately, the scientific literature is plagued by studies in which the causal question is not explicitly stated and the investigators' unverifiable assumptions are not declared. This casual attitude towards causal inference has led to a great deal of confusion.

Questions

- Descriptive / predictive:
  - "Is this patient at high risk of developing complications during surgery?"

  Causal:
  - "Which type of anaesthetic should this patient receive to reduce the risk of complications during surgery?"
  - "How does the amount of anaesthetic affect the risk of complications during surgery?"
  - "What can be done to reduce the risk of complications during surgery for an average / a particular type of patient?"

Questions

- Descriptive / predictive:
  - "Which type of client will buy which kind of product?"

  Causal:
  - "Should advert be at the top or bottom of website to increase the probability of viewing product?"
  - "How does the size of advert affect the probability of viewing product?"
  - "How can I get a client to buy my product?"

Questions

- Descriptive / predictive:
  - "Who is most likely to become long-term unemployed?"

  Causal:
  - "Will a minimum wage legislation increase the unemployment rate of a country?"
  - "What can be done to prevent someone from becoming unemployed?"

What's the question (Hernan et al, Chance, 2019)

How can women aged 60–80 years with stroke history be partitioned in classes defined by their characteristics?
`Hernan et al, Chance (2019)`

This question is just about description.

What's the question

What is the probability of having a stroke next year for women with certain characteristics?

This question is just about prediction.

What's the question

Will starting a statin reduce, on average, the risk of stroke in women with certain characteristics?

This question is about **causal effects***, sometimes called counterfactual prediction.*

3 tasks of data scientists

- Description
- Prediction
- Counterfactual prediction (*What would happen if...*)

Section 3

Prediction vs. causal inference

Prediction and causal inference are different exercises

- Prediction: Learn about Y after observing $X = x$.
- Causal inference: Learn about Y after ~~observing~~ setting $X = x$.

Pearl

Figure 1: Judea Pearl

"All the impressive achievements of deep learning amount to just curve fitting"

Motivation

Albert Einstein (1953):

> *"Development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment (during the Renaissance)".*

Experiments

- Experiment (biology).
- The randomized controlled trial (medicine).
- A/B testing (tech industry).

Scurvy - the first randomized trial?

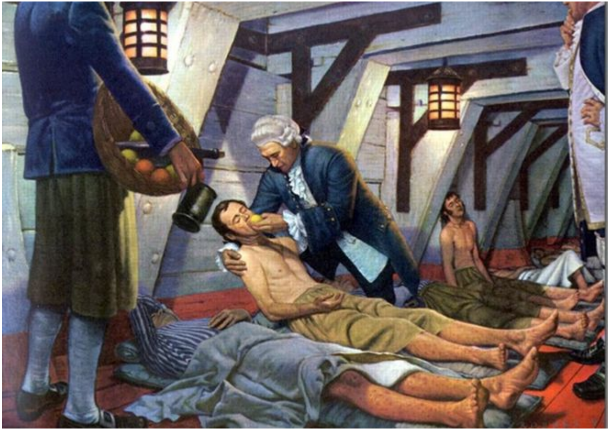


Figure 2: James Lind, the surgeon, 1753.

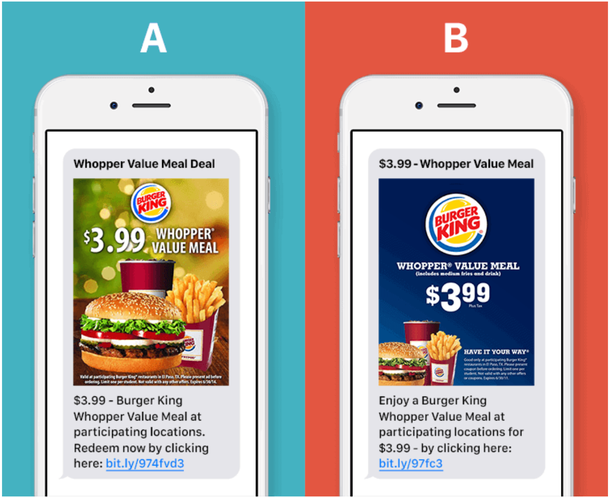
https://www.bbc.com/news/uk-england-37320399

Lind's experimental set-up (simplified)

- Recruit a bunch of sailors suffering from scurvy
- Flip a coin for each sailor to determine course of action A
  - Heads: $A = 1$ (a lemon a day)
  - Tails: $A = 0$ (elixir of vitriol a.k.a sulphuric acid)
- For each sailor note down the outcomes denoted by Y : let's say $Y = 1$ is healthy and $Y = 0$ is sick with scurvy
- This is an example of a simple randomized controlled trial (RCT), also called A/B test.

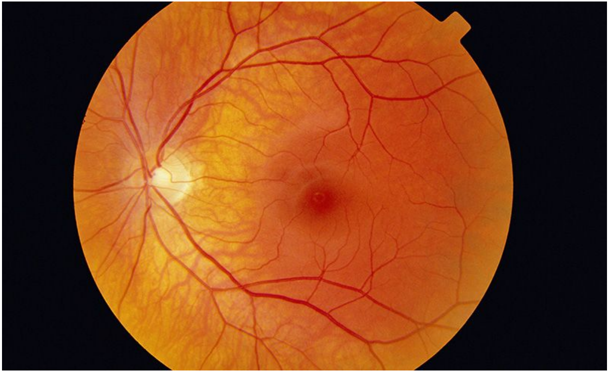
An aerial view of Rothamsted's Broadbalk field, site of the Broadbalk Wheat Experiment since 1843.

In Silicon Valley they call it AB testing.



Why bother?

Why bother?



"Can (...) predict whether someone is at risk of an impending heart attack,"
Nature Biomedical Engineering, 2018

Decisions have to be made...

What if...

- Would starting treatment A prevent a heart attack?
- Is Drug A better than Drug B?
- Would the ad get more clicks if it were green instead of red?
- Would the election campaign increase the number of votes?
- Would university education increase my future earnings?

What if questions can be assessed in experiments

... but experiments are often not available because they are

- impractical,
- expensive,
- time consuming,
- unethical,

... and experiments may not be perfectly executed.

So, what do we do?

Emulate the experiment of interest from available *observational* data.


THE LANCET Neurology
Volume 14 · Issue 12 · December 2015
www.thelancet.com/neurology
Articles
PIK3R2 mutations in polymicrogyria
See page 1182
Review
Cranial functional movement disorders
See page 1196
Personal View
Precision medicine in the epilepsies
See page 1219

Is this something we should consider seriously?

Optimising clinical decision-making with medical algorithms (Roche, 2023)

some nuances?

https://becominghuman.ai/summary-of-the-alphago-paper-b55ce24d8a7c

Section 4

Counterfactuals

What would happen if...

Counterfactuals 1973:

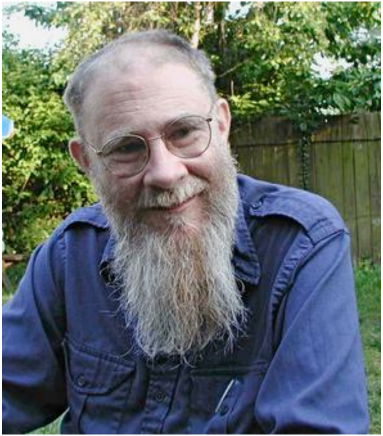


Figure 3: David Lewis

https://en.wikipedia.org/w/index.php?curid=58724625

Prediction and causal inference are different exercises

- Prediction: Learn about Y after observing $A = a$.
- Previously you have studied random variables *conditional* on parameters,

$$Y \sim \text{Distribution}\{g(a)\}.$$

Such (conditional) associations are not necessarily easy to interpret (see the next example).

Two fundamental questions of causality (Pearl, 2009)

1. What empirical evidence is required for legitimate inference of cause–effect relationships?
2. Given that we are willing to accept causal information about a phenomenon, what inferences can we draw from such information, and how?

In this course, we will consider **mathematical tools** for **casting causal questions** or **deriving causal answers**.

Section 5

Prediction vs. causal inference

Prediction and causal inference are different exercises

- Prediction: Learn about Y after observing $A = a$.
  That is, infer properties of the law P that generated the observations Y.
- Causal inference: Learn about Y after ~~observing~~ fixing $A = a$.
  That is, infer properties of a *counterfactual* law, say, P^a,that would generate data when a is fixed.

What is a causal effect (in a simple setting)

Consider the following observed random variables:

- A binary treatment $A \in \{0, 1\}$.
- An outcome $Y \in \mathcal{Y}$.
- A vector of baseline covariates $L \in \mathcal{L}$.

Define the *counterfactual* or *potential outcome* variables

- $Y^a \in \mathcal{Y}$.
  The outcome variable that would have been observed under the treatment value a (the superscript denotes the counterfactual).
- Often we will specifically instantiate a, i.e. set a to a value:
  $Y^{a=0} \in \mathcal{Y}$.
  The outcome variable that would have been observed under the treatment value $a = 0$.
  $Y^{a=1} \in \mathcal{Y}$.
  The outcome variable that would have been observed under the treatment value $a = 1$.

Individual level causal effect

Definition (Individual level causal effect)

A causal effect for individual (unit) i is $Y_i^{a=0}$ vs $Y_i^{a=1}$.

Remarks on $Y^{a=0}$ and $Y^{a=1}$

The fundamental problem of causal inference:

- Suppose $A = 1$. Then $Y = Y^{a=1}$ is observed, but $Y^{a=0}$ is unobserved...
- Suppose $A = 0$. Then $Y = Y^{a=0}$ is observed, but $Y^{a=1}$ is unobserved...

The consequence is that individual level effect cannot be identified.[2]

[2] We will consider a possible exception later.

Intervening is not the same as conditioning

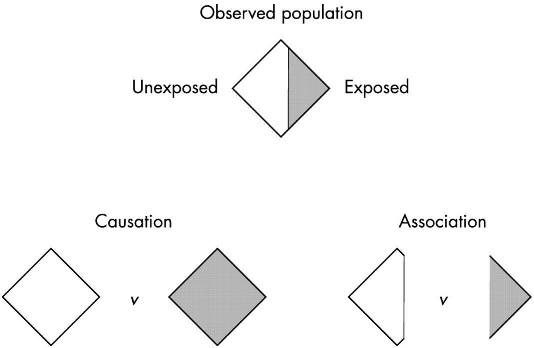


Figure taken from Hernan, 2014, BMJ.