Correlation vs. causation

Prisoners example

- Hypothetical data:

	Rearrested	Not rearrested	Rearrest rate
Participants	100	400	20%
Non participants	500	500	50%

- Possible questions:
  - non-causal
    - Can we predict whether a prisoner will be rearrested based on participation in the program?
  - causal
    - Does the program lower the rearrest rate? (causal mechanism)
    - What would be the rearrest rate if the program were compulsory for all prisoners? (prediction after new intervention)

Causal questions

- Causal questions are about the mechanism behind the data or about predictions after some outside intervention
- Some examples:
  - Does smoking cause lung cancer?
  - What is the efficacy of a new drug?
  - What is the gene regulatory network of yeast?
  - Did racial discrimination play a role in hiring processes?
  - What is the projected economic growth in the UK after the Brexit?

Causal statements in the news

- "Eat breakfast if you want to reduce your risk of coronary heart disease" (The Guardian)
- "Anesthesia may harm children's brains. Study: anesthesia before age 3 linked to later mental problems" (WebMD)
- "Breastfeeding may reduce Alzheimer's risk" (Cambridge U)
- …

- One should be careful with causal statements. Example:
  - People who eat breakfast tend to have less heart disease
  - Advice: eat breakfast?

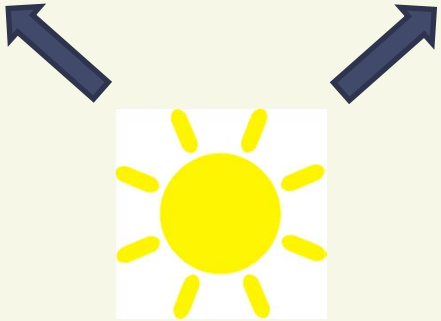
Classical example

- Days with high ice-cream sales tend to have more drowning accidents
- Advice: stop selling ice-cream ?

Correlation

Classical example

- Days with high ice-cream sales tend to have more drowning accidents
- Advice: stop selling ice-cream ?



No, hot weather is a cause for both

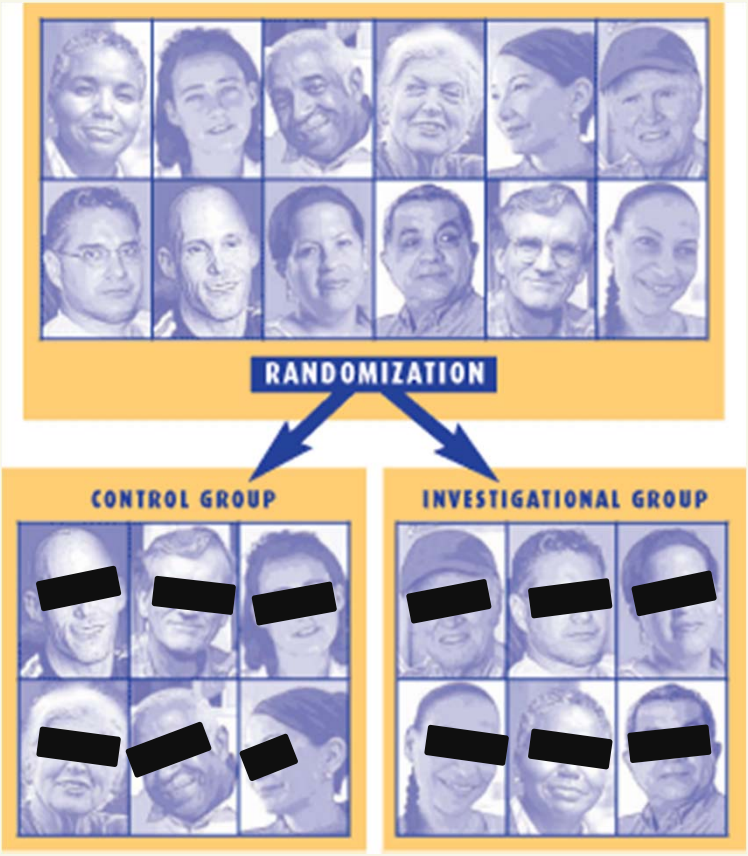
Back to the prisoners example

	Rearrested	Not rearrested	Rearrest rate
Participants	100	400	20%
Non participants	500	500	50%

- What questions can we answer?
  - non-causal
    - Can we predict whether a prisoner will be rearrested based on participation in the program?
  - causal
    - Does the program lower the rearrest rate?
    - What would be the rearrest rate if the program were compulsory for all prisoners?

- Depends on how data were collected
  - Randomized controlled experiments
  - Observational studies

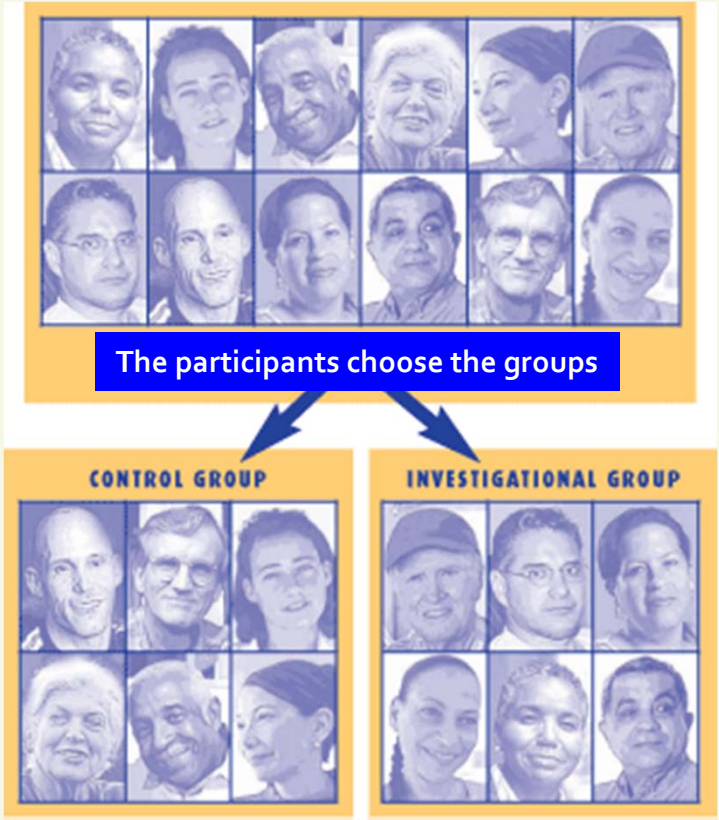
Randomized controlled experiments



- There is a control group
- The treatment is randomized
- When possible also double blind

We compare like with like.
Hence, any difference between the groups must be due to treatment.

Observational studies



- The participants decide about the "treatment". The scientists only observe what is going on.

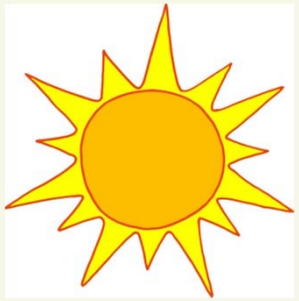
Observational studies

- Example: Breakfast or only coffee?

Less stress?
More sports?
Healthier snacks?
….

- What is the cause of the lower rate of heart disease in the group of people that eat breakfast?
- We don’t know…

Compare to experiments in biology, physics, etc



- Why do the flowers in the second group grow faster?
- We don't know, because we changed two things at a time. These two effects are confounded.

Back to the prisoners example

	Rearrested	Not rearrested	Rearrest rate
Participants	100	400	20%
Non participants	500	500	50%

- What questions can we answer?
  - non-causal
    - Can we predict whether a prisoner will be rearrested based on participation in the program?
  - causal
    - Does the program lower the rearrest rate?
    - What would be the rearrest rate if the program were compulsory for all prisoners?

- Depends on how data were collected
  - Voluntary participation → observational study → Q1
  - Randomized participation → rand. contr. exp. → Q2+Q3

Design experiments carefully

- Examples:
  - Data are collected without a clear research question in mind.
    Danger: multiple testing.
  - 100 Cows are randomly divided in two groups. Each group is in a stable. The cows in one stable get a special diet. Does the special diet change the average heart rate of the cows?
    Danger: stable may be confounded with treatment; heart rates of cows within one stable may be correlated.
  - A group of students has to do two tests: test I and test II. The tests are different but also show similarities.
    Danger: learning effect over time

Take home points

- We all need a basic understanding of causality:
  - Recognize causal versus non-causal questions
  - Realize importance of study design:
    ask how the data were collected
  - Recognize observational vs. randomized controlled studies
  - Identify potential confounders in observational studies
  - When designing an experiment, talk to a statistician in the planning phase