Reproducibility and Experimental Design

Tamás Schauer

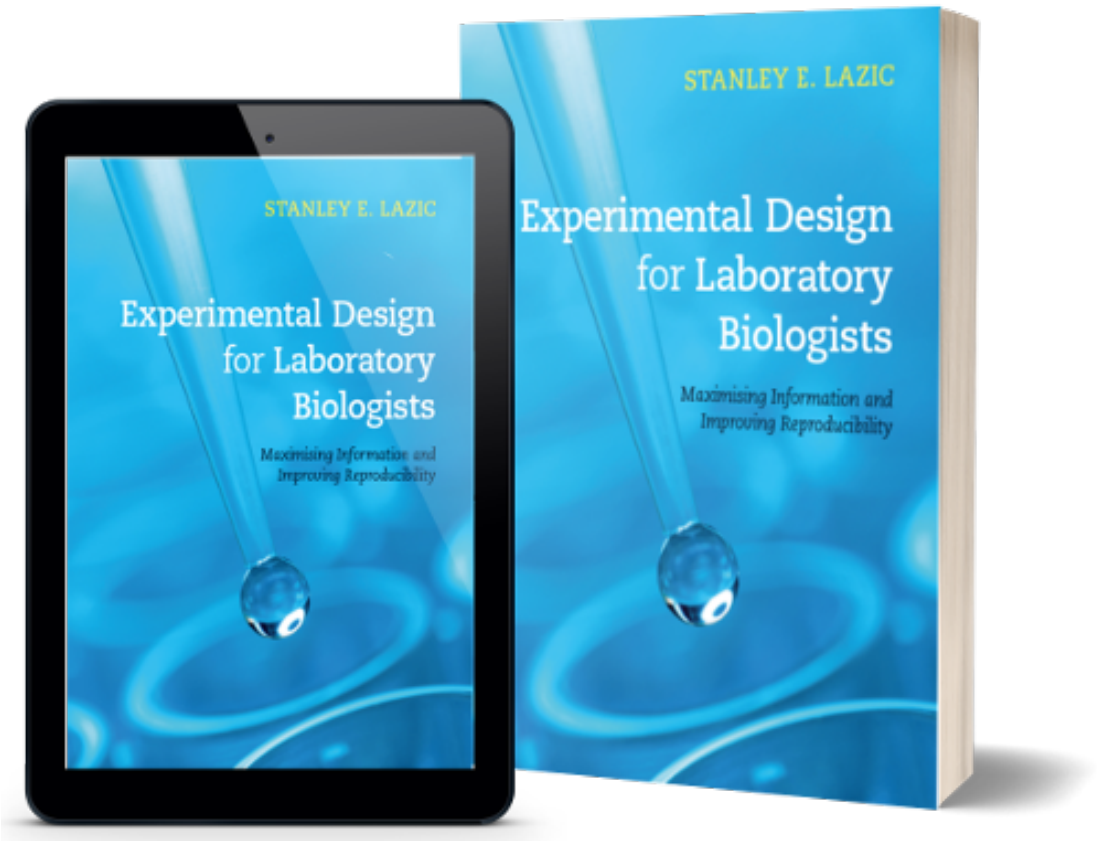
15.04.2020

Overview

- What is reproducibility?
- Source of bias
- Experimental Design
- Replication (N = ?)
- Recommendations

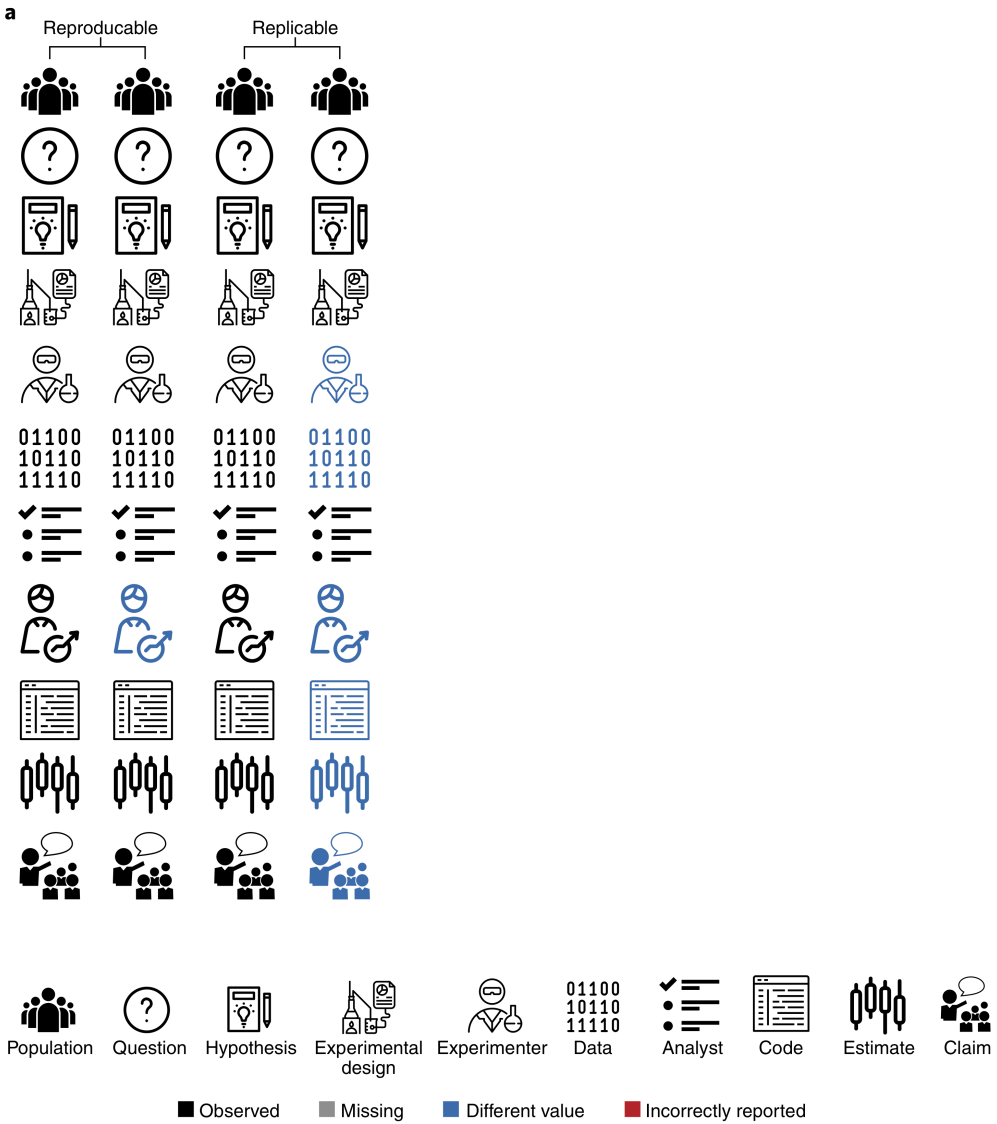
Source: https://github.com/tschauer/Reprodicibility_ExpDesign

Reference Book



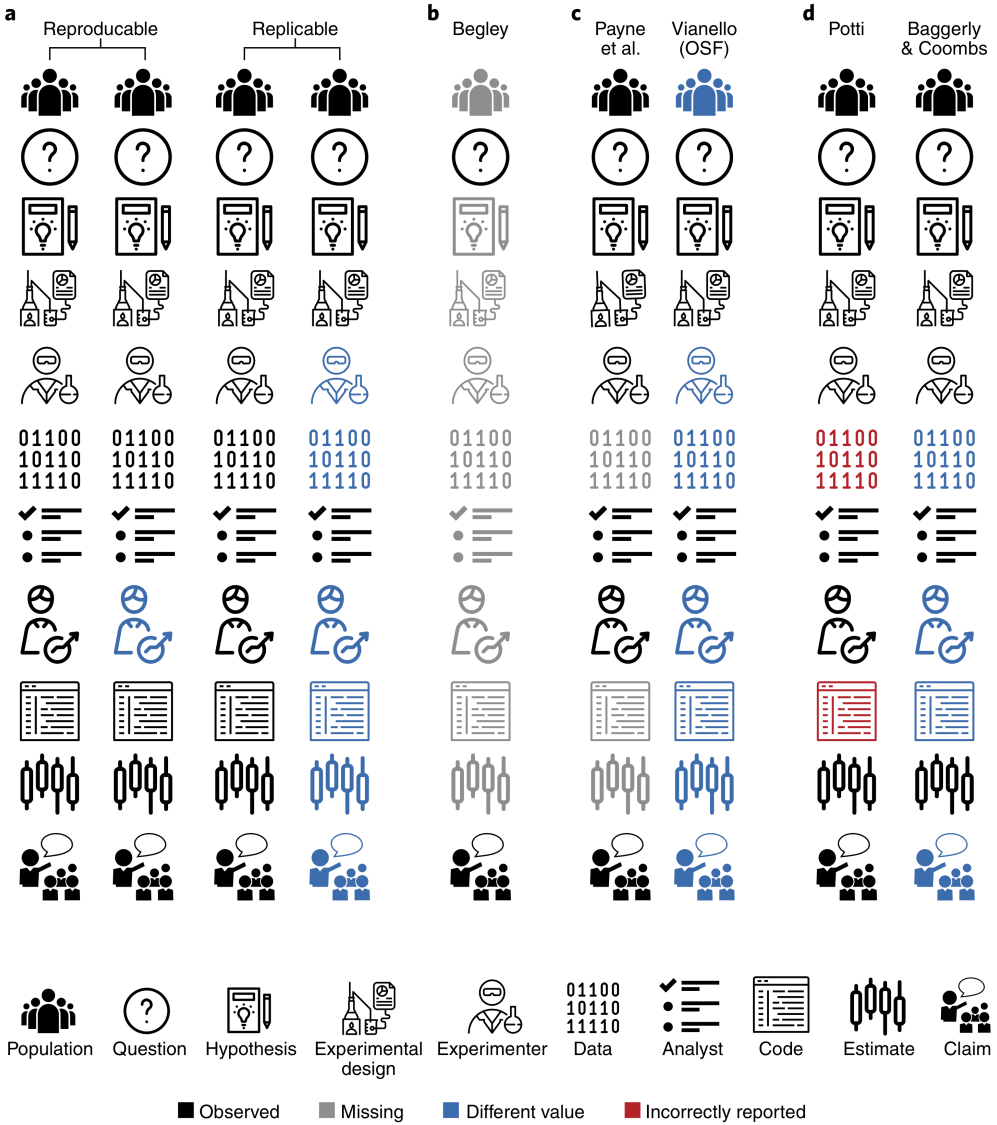
Lazic, 2016

Reproducibility vs. replicability



Patil, Peng and Leek, 2019

Reproducibility vs. replicability



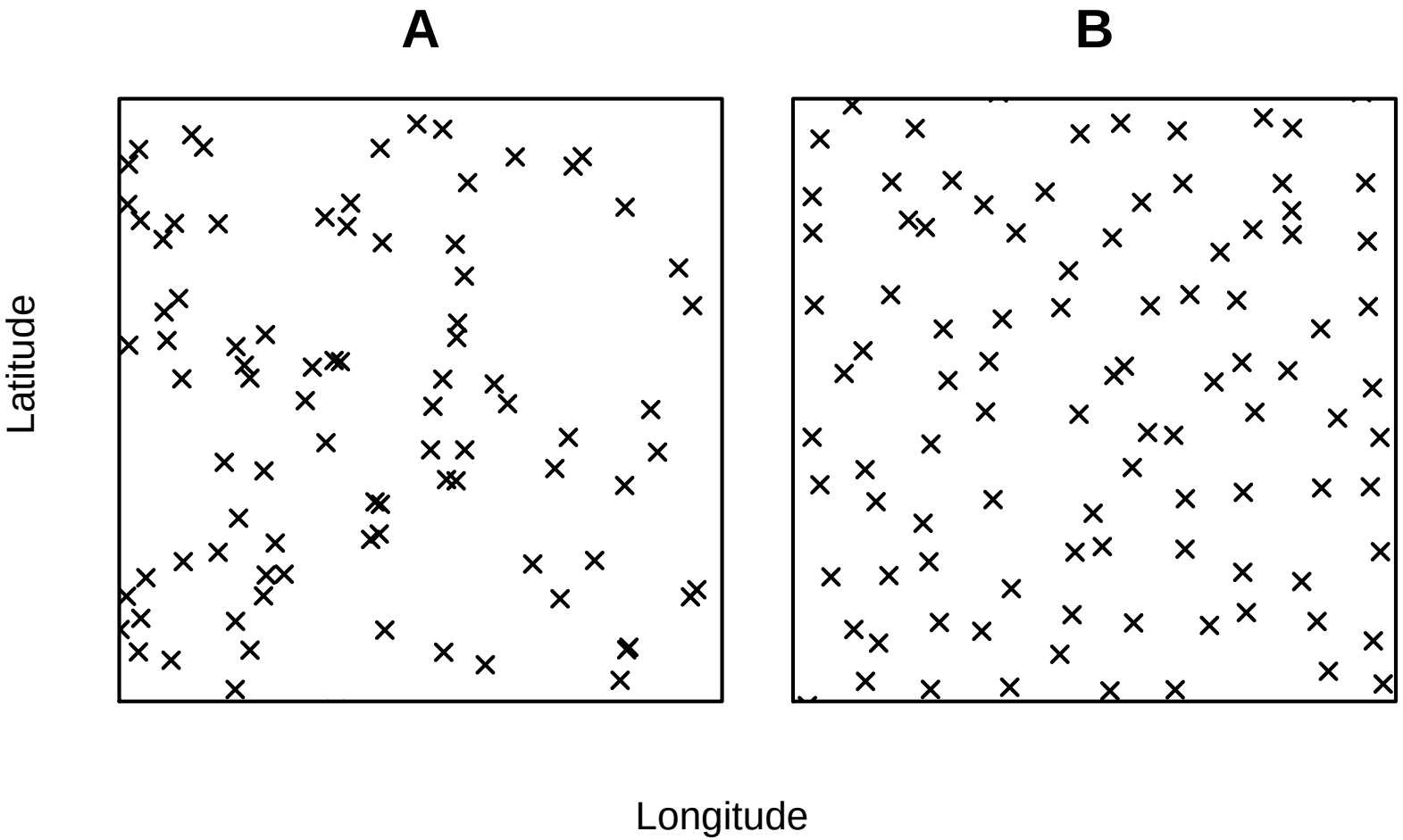
Patil, Peng and Leek, 2019

Reproducibility by Lazic

- **analytical**: original data and analysis (code!)
- **direct**: same conditions, materials, methods
- **systematic**: different conditions (e.g. cell line, KD vs drug)
- **conceptual**: general under diverse conditions (paradigm)

Source of bias

- Strategy: which location was bombed randomly?

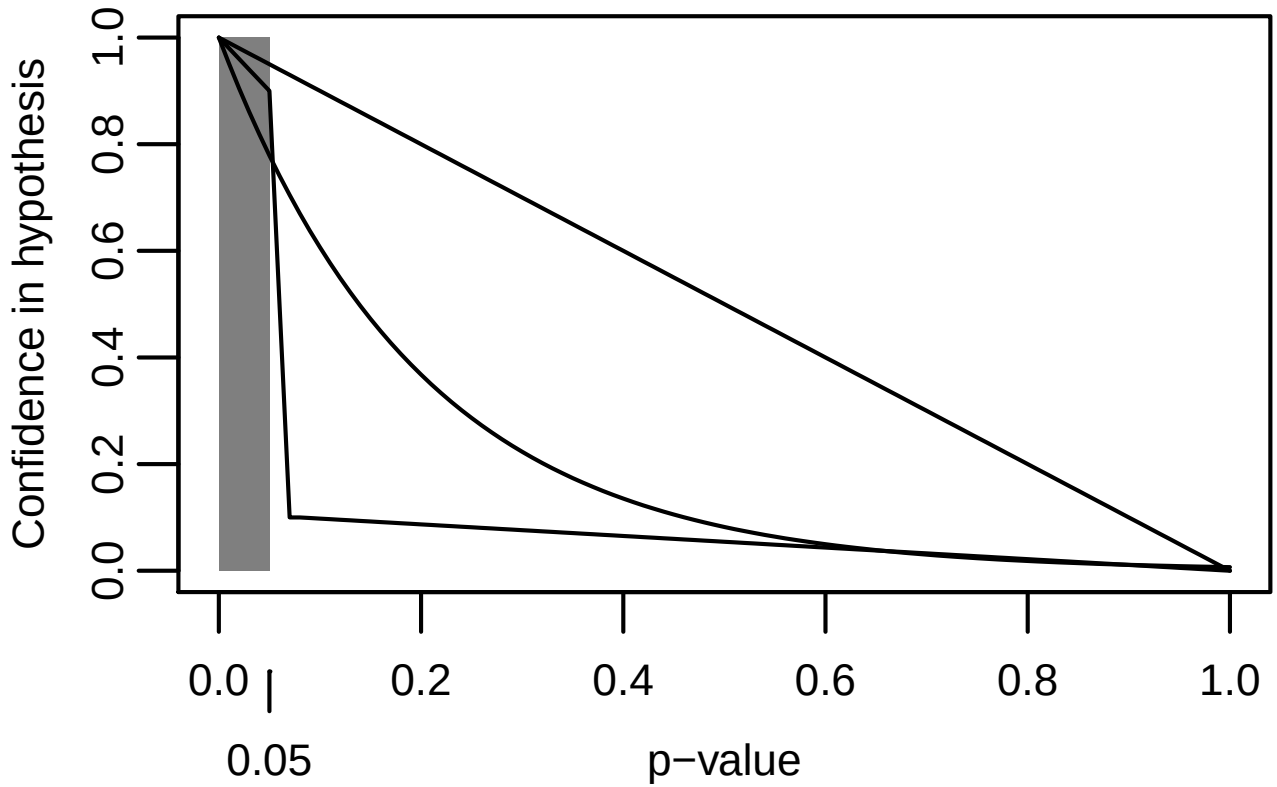


Source of bias

- seeing pattern in randomness
- not wanting to miss anything (what else can we get out?)
- *"if a hypothesis is derived from the data, then the ability of the data to support that hypothesis is diminished"*
- exploratory vs. confirmatory research

Source of bias

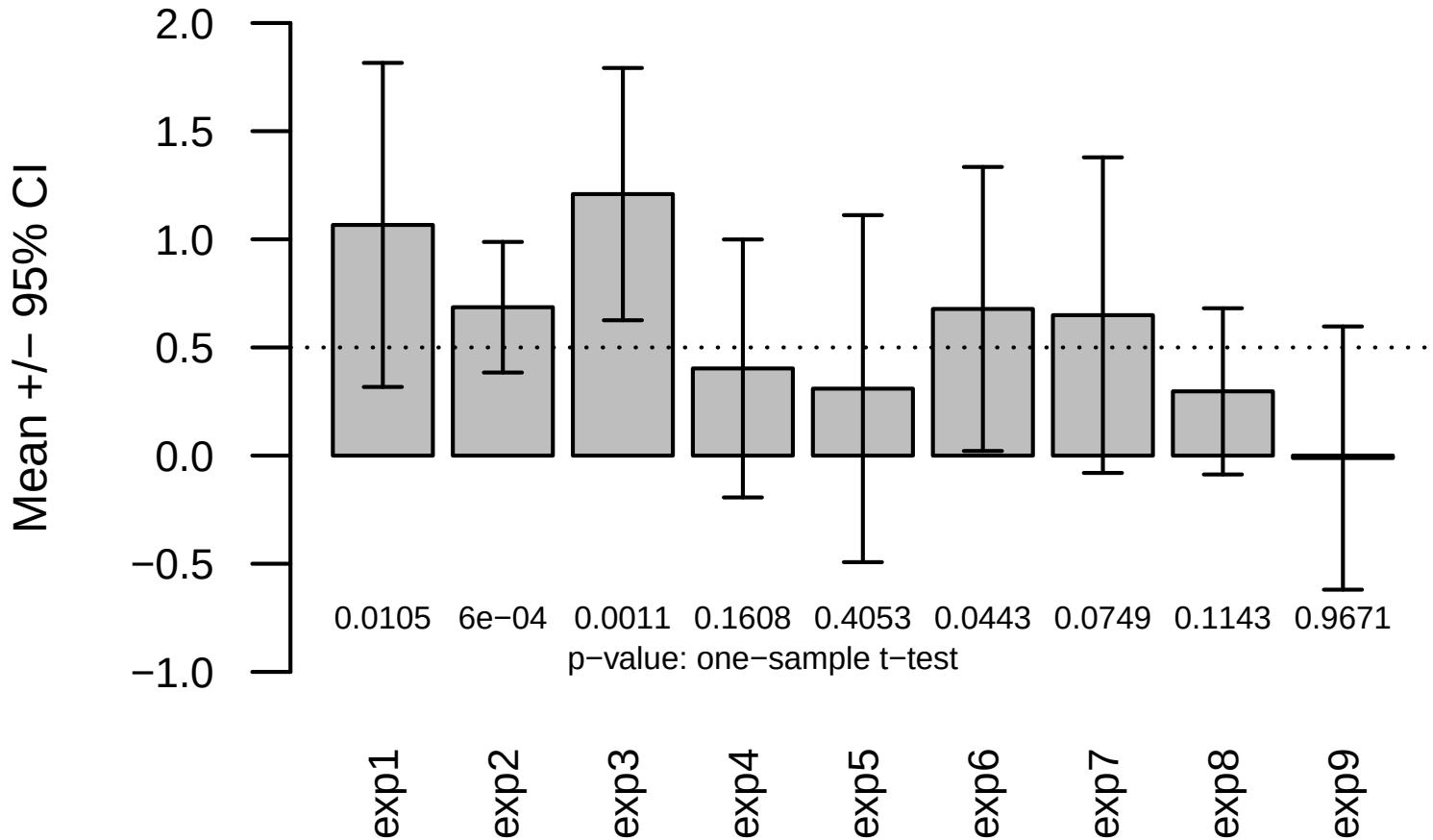
- psychological cliff at p = 0.05



original study: Poitevineau and Lecoutre, 2001

Source of bias

- neglect of sampling variability



- True mean = 0.5, standard deviation = 1, n=10 each

Source of bias

- lack of independence
  - repeated measures
    - observations are close together in space or time
    - same animal, litter, cell culture dish, fly vial
  - correlated variables
    - different measures of a single underlying effect
    - co-regulated genes, proteins, metabolites
    - disease severity

Source of bias

- confirmation bias
  - Pubmed search: disease + gene name
    - what about studies which do not find the association?
    - neglecting negative results
  - data transformation until it "gets" significant
  - selecting data to tell the story (data that do not fit excluded)

Source of bias

- expectancy effects (measurements are influenced)
- hindsight bias ('I knew it all along')
- herding effect (scientific inbreeding)

Common problems

- Experimental Design
  - confounding (conditions $\sim$ biological, technical effects)
  - experimental unit (replicates)
  - lack of randomization
  - low statistical power
- Conducting experiments
  - lack of blinding
  - lack of randomization
  - optional stopping

Common problems

- Analysis
  - experimental unit (inflated sample size)
  - inapproriate model (normal distribution)
  - incorrect interpretation
  - selective reporting

Experimental Goal

	Exploratory	Confirmatory
Question	General	Specific
Hypothesis	Generating	Before
Order	Before	After (independent data)
Analysis	Data dependent	Data independent
Minimize	False Negatives	False Positives
P-value	No Diagnostic Value	Diagnostic Value
Power Analysis	Rarely	Yes

Experimental Goal

	Exploratory	Confirmatory
Subjects	Heterogeneous	Homogeneous
Environment	Varied	Standardized
Treatments	Many	Few
Levels	Many	Few
Time points	Many	Few
Outcome	Many	Few
Controls	Few	Many
Blinding	Possibly	Yes
Randomization	Yes	Yes
Blocking	Yes	Yes

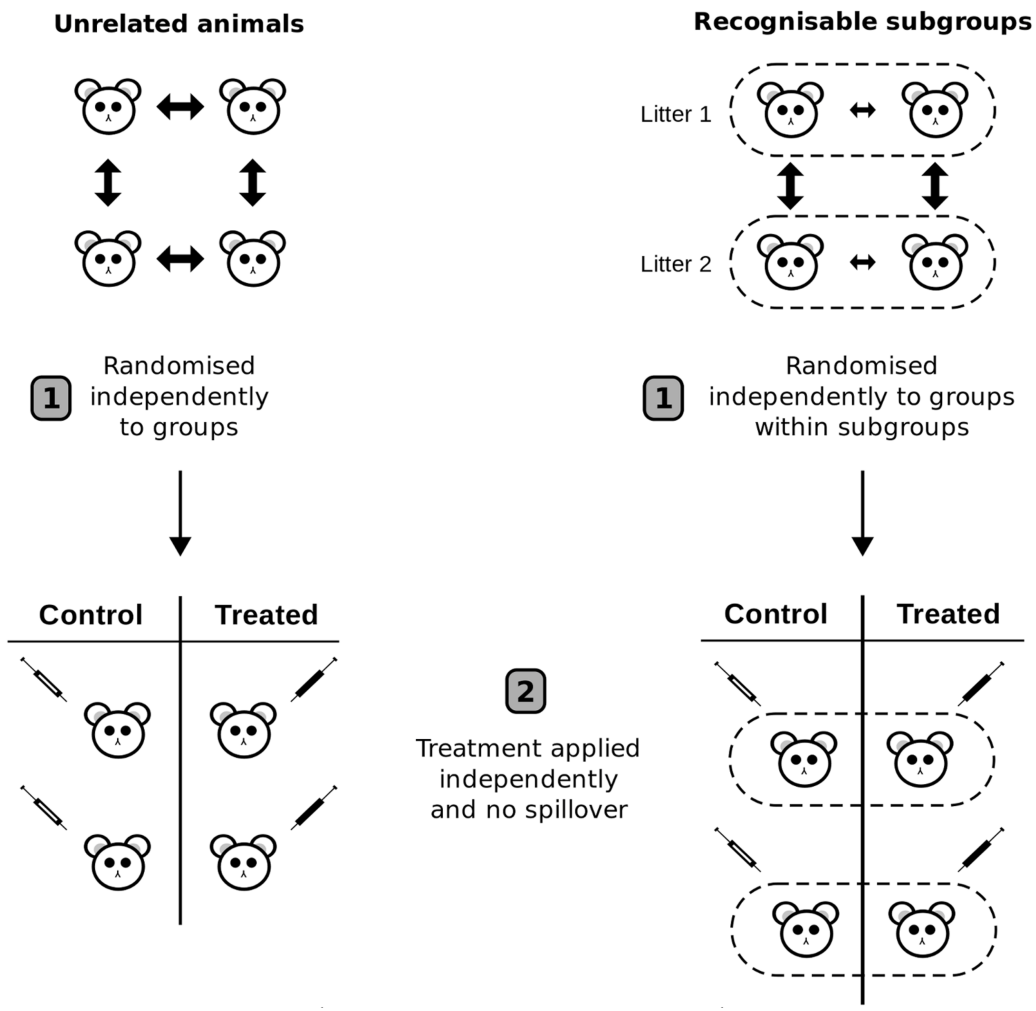
Hypothesis testing?

- p-value?
  - A: "Given these data what is the probability that the null hypothesis is true"
  - B: "Given that the null hypothesis is true, what is the probability of these (or more extreme) data"
- "The Earth is round ($p < .05$)" Cohen, 1994
- Solution? Pre-registration: https://cos.io/prereg/

Experimental Design Equation

Outcome =	Treatment +	Biological +	Technical +	Error
Gene exp.	Environment	Sex	Person	Experimental
Protein	Compound	Age	Batch	Treatment
Cell counts	Inhibitor	Weight	Flask	Sampling
	siRNA	Litter	Cage	Measurement
	Dose	Genotype	Day	
	Time	Cell line	Incubator	

Randomization



- Completely randomized vs. randomized blocked design

Lazic, 2018

2-factor design

- crossed

	Control	Treated
Day 1	*****	*****
Day 2	*****	*****

- nested

	Control	Treated
Cage 1	*****	
Cage 2	*****	
Cage 3		*****
Cage 4		*****

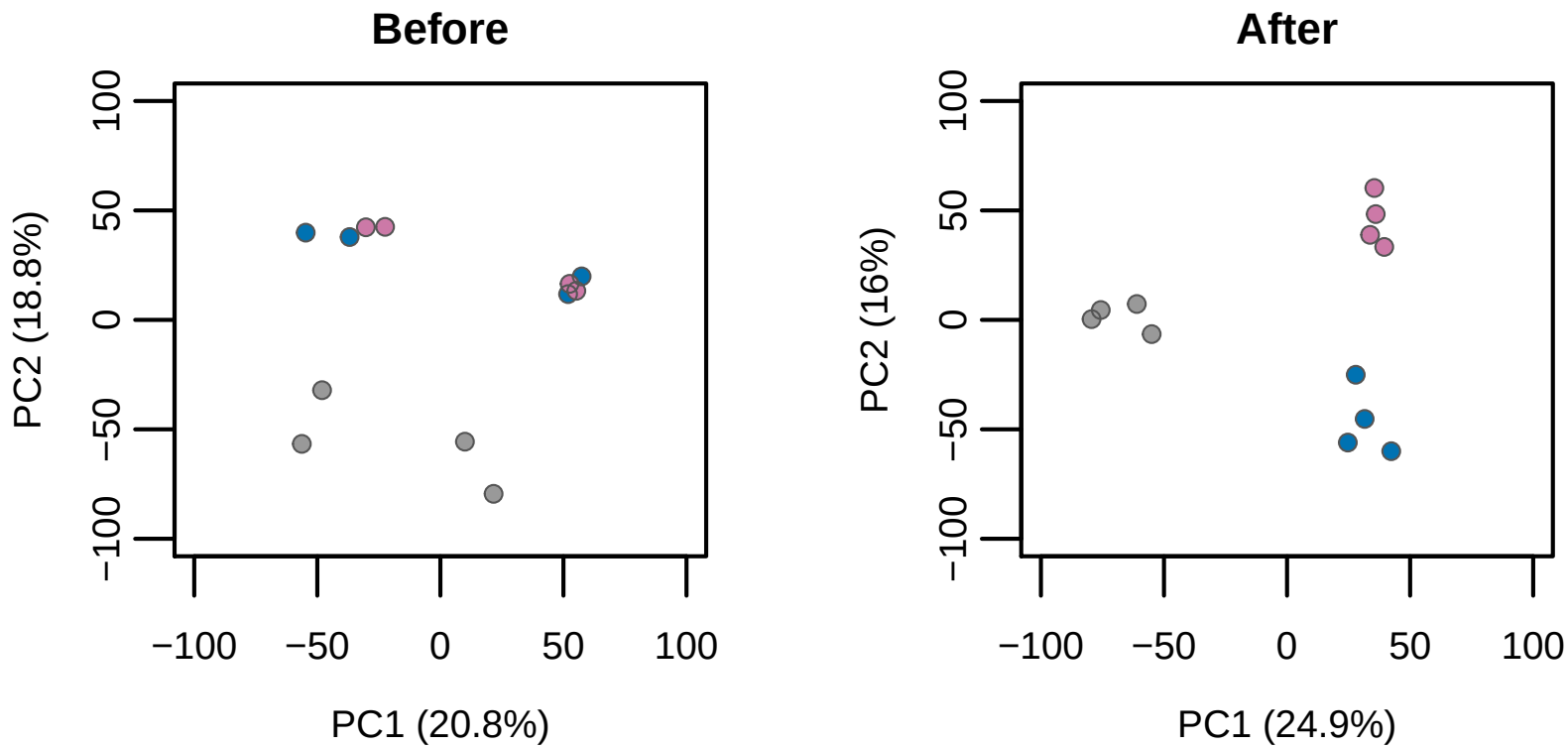
2-factor design

- confounded

	Control	Treated
Batch 1	**********	
Batch 2		**********

Batch effects

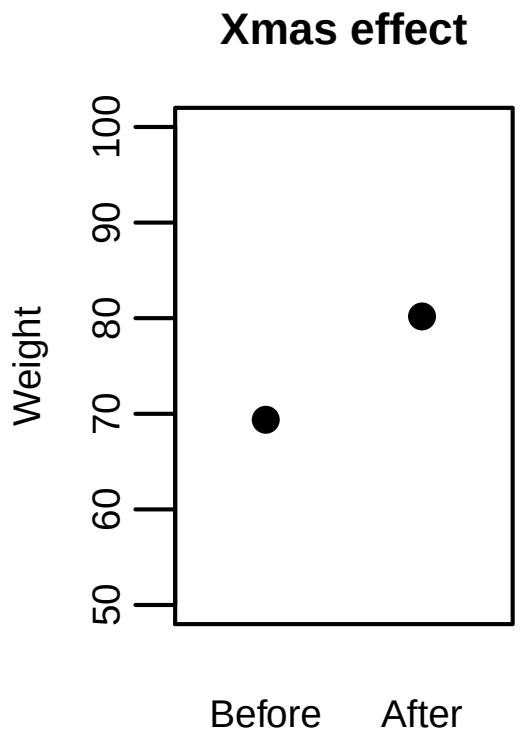
- 3 conditions, 4 reps in 2 batches



Data: Catherine Regnard

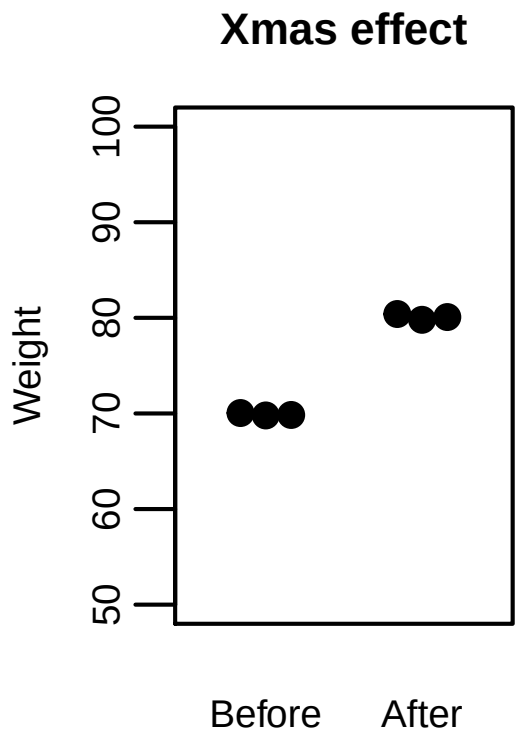
Example

- How does Christmas affect human body weight?



Example

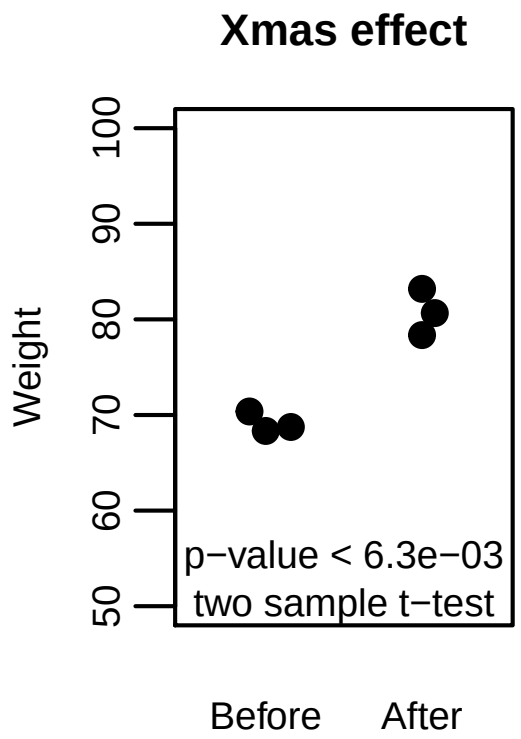
- measurement error (3x within minutes)



- qPCR well-replicates
- sequencing the same library

Example

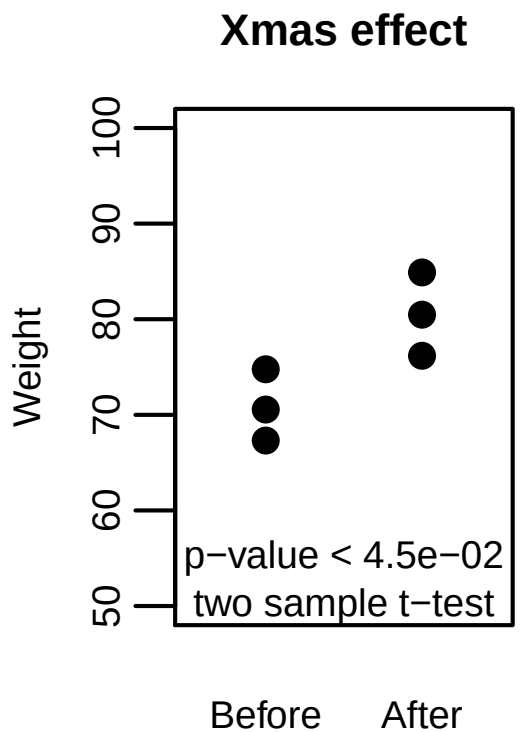
- Different days same person



- Christmas signficantly increases human body weight ???
- e.g. cell culture experiments (generalizable?)

Example

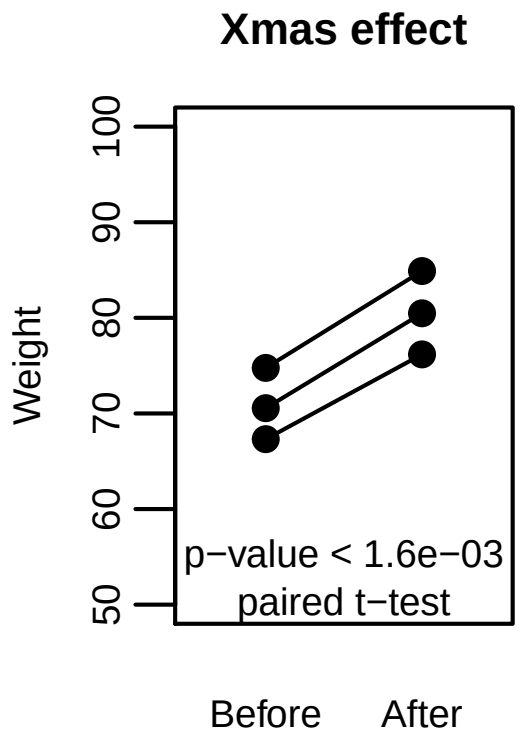
- Different years same person



- year can be used as grouping factor

Example

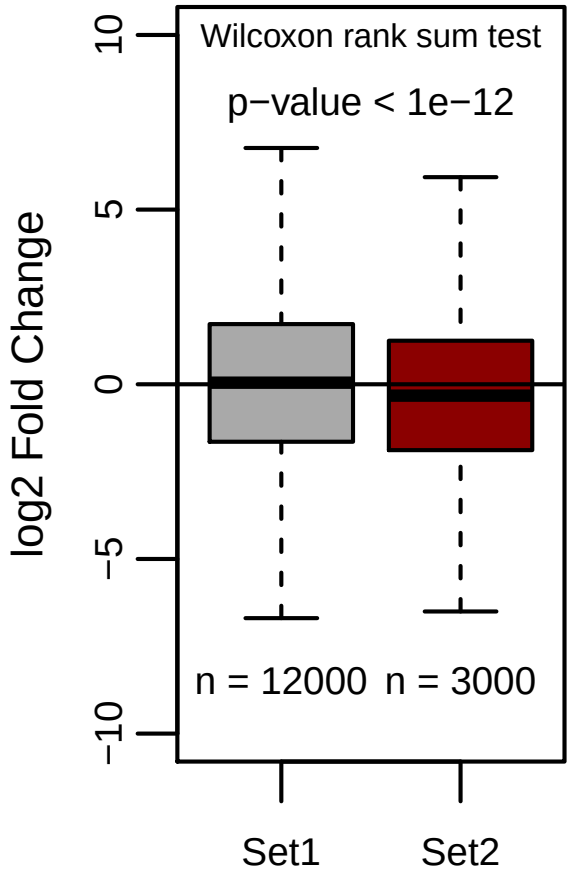
- Different years same person



- it is still just a single person
- different years not always applicable

What is N?

- within sample testing (comparing set of genes)



- True difference: -0.25
- large number of genes inflates p-value (no diagnostic value)

What is N?

- Biological unit (BU): the entity about which inferences are made.
- Experimental unit (EU): the entity that is randomly and independently assigned. Sample size N = EU.
  - BU of interest
  - groups of BUs
  - parts of a BU
  - sequence of observations on a BU
- Replicate EU to increase N!

definitions by Lazic, 2016

What is N?

- Observational unit (OU): the entity on which measurements are taken.
- More OUs do not increase N
  - e.g. cells from a single aliquot, well, slide
- Multiple OUs should be averaged (or use hierarchical model)!
- Report: what is EU, OU.

definitions by Lazic, 2016

Recommendations

- **Ask:** exploratory or confirmatory?
- **Design:** conditions, batches, confounding
- **Estimate:** effect size, variance, power
- **Report:** what is your sampling N?
- **Share:** lab protocol, raw data and analysis code

Final Thought

Acknowledgements

- BMC, Bioinformatics
  - Tobias Straub
- BMC, Molecular Biology
  - collaborators
- Funding: SFB 1064