A White Hat's Guide to P-Hacking

Pierre LeVan, PhD
Departments of Radiology and Paediatrics
Hotchkiss Brain Institute
Alberta Children's Hospital Research Institute

December 14, 2021

UNIVERSITY OF CALGARY

Speaker name:

Dr. Pierre LeVan, University of Calgary

Conflicts of interest regarding this presentation:

Nothing to disclose

ESMRMB

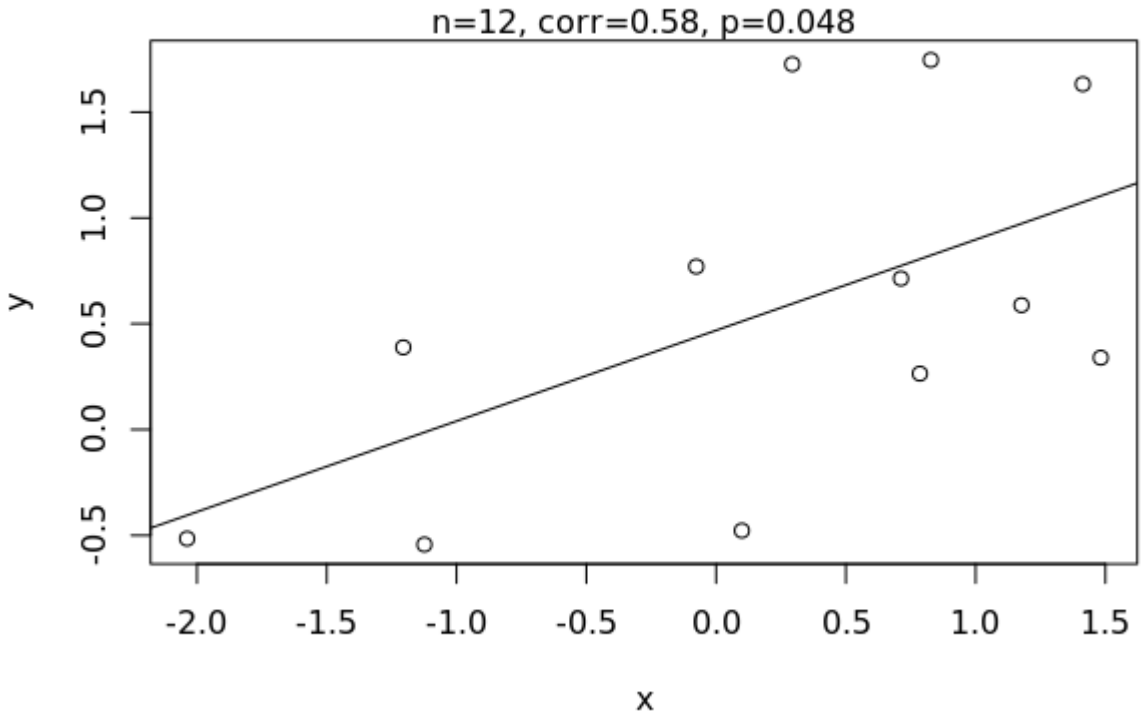
European Society for Magnetic Resonance in Medicine and Biology

UNIVERSITY OF CALGARY

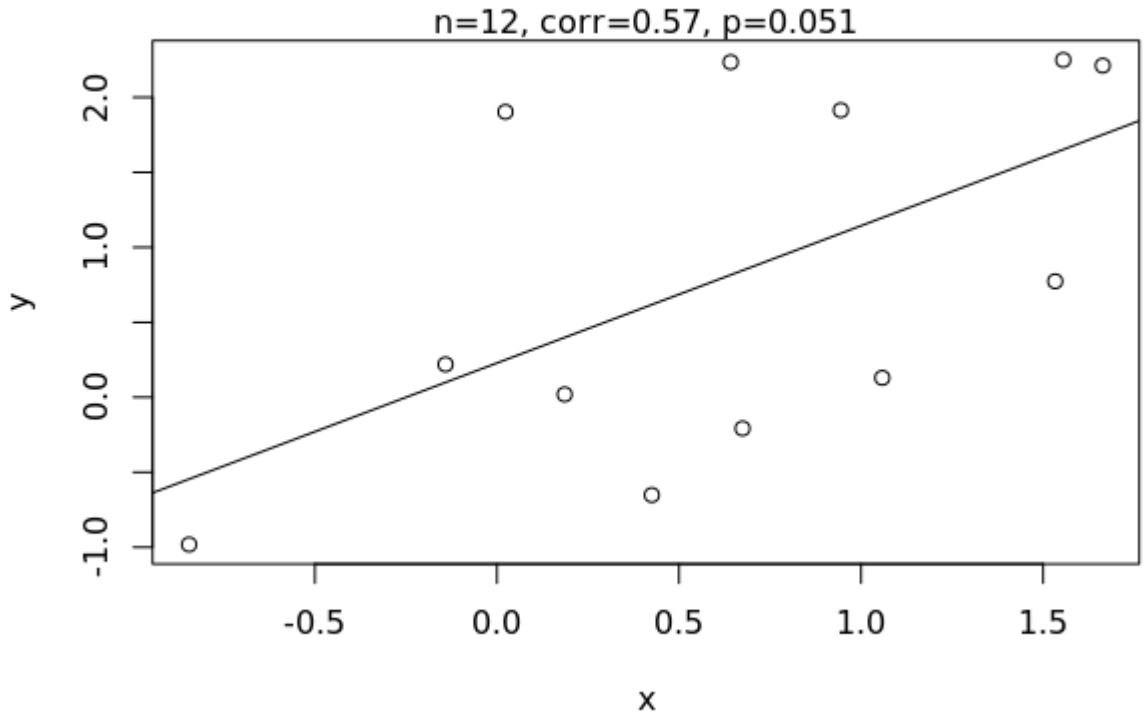
Introduction

- The unfortunate, yet necessary obsession with p-values

„Strong correlation between x and y (r=0.58, p<0.05)“

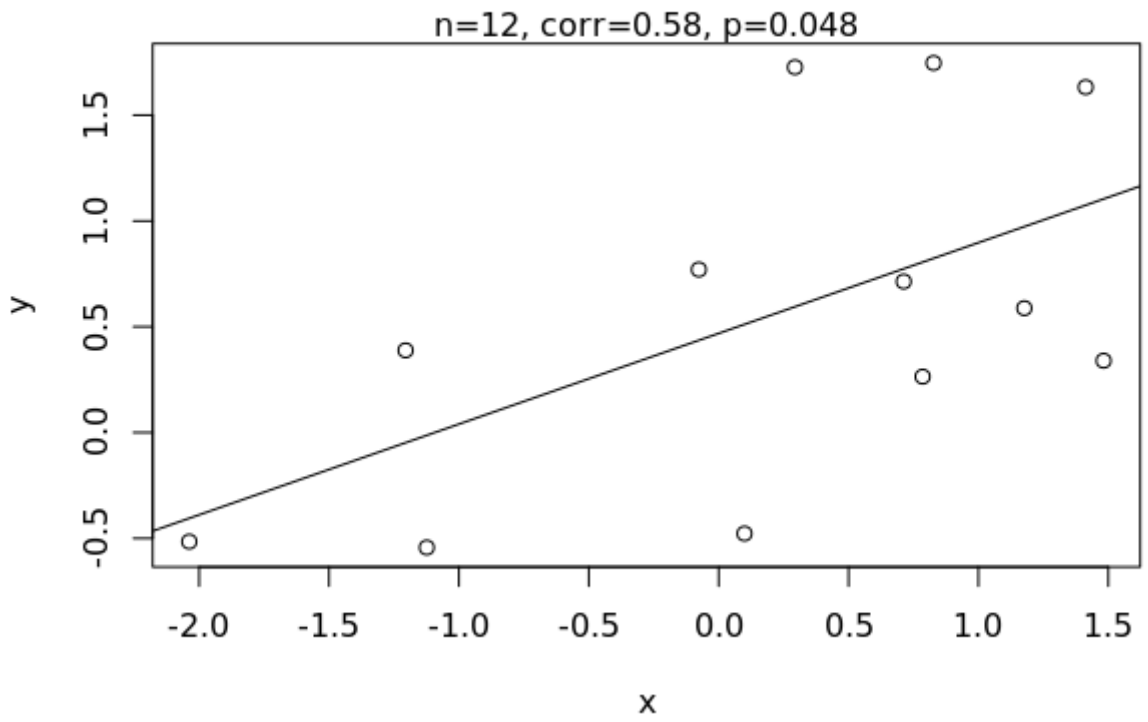


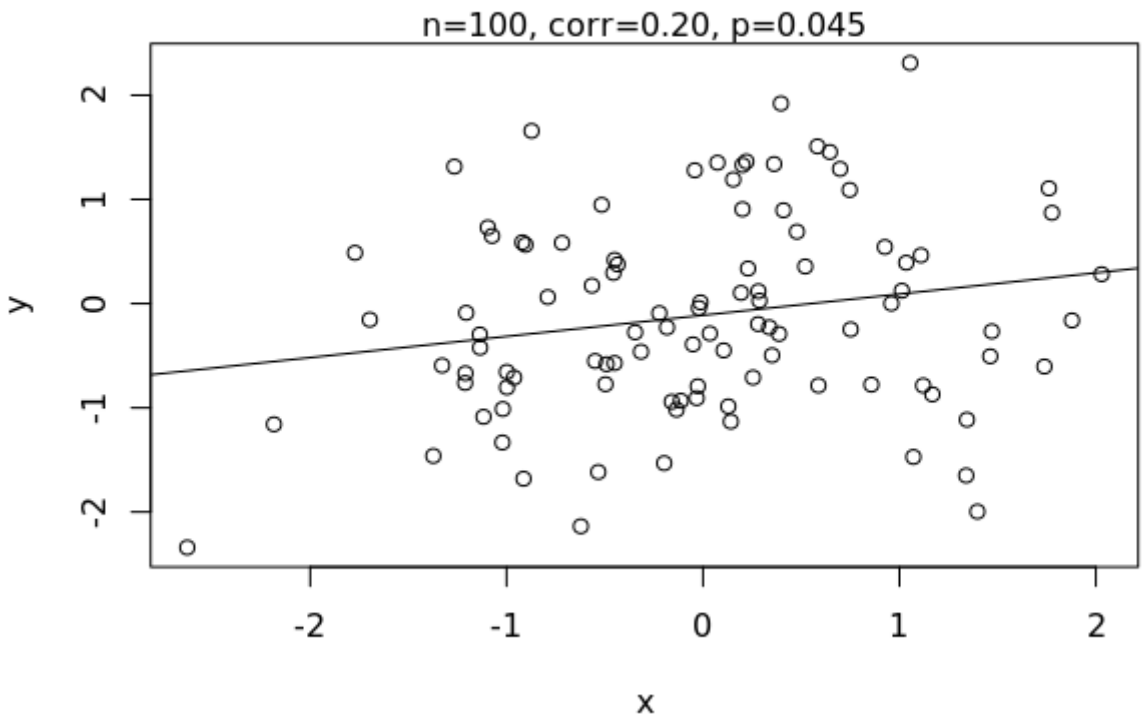
„No correlation between x and y (p>0.05)“



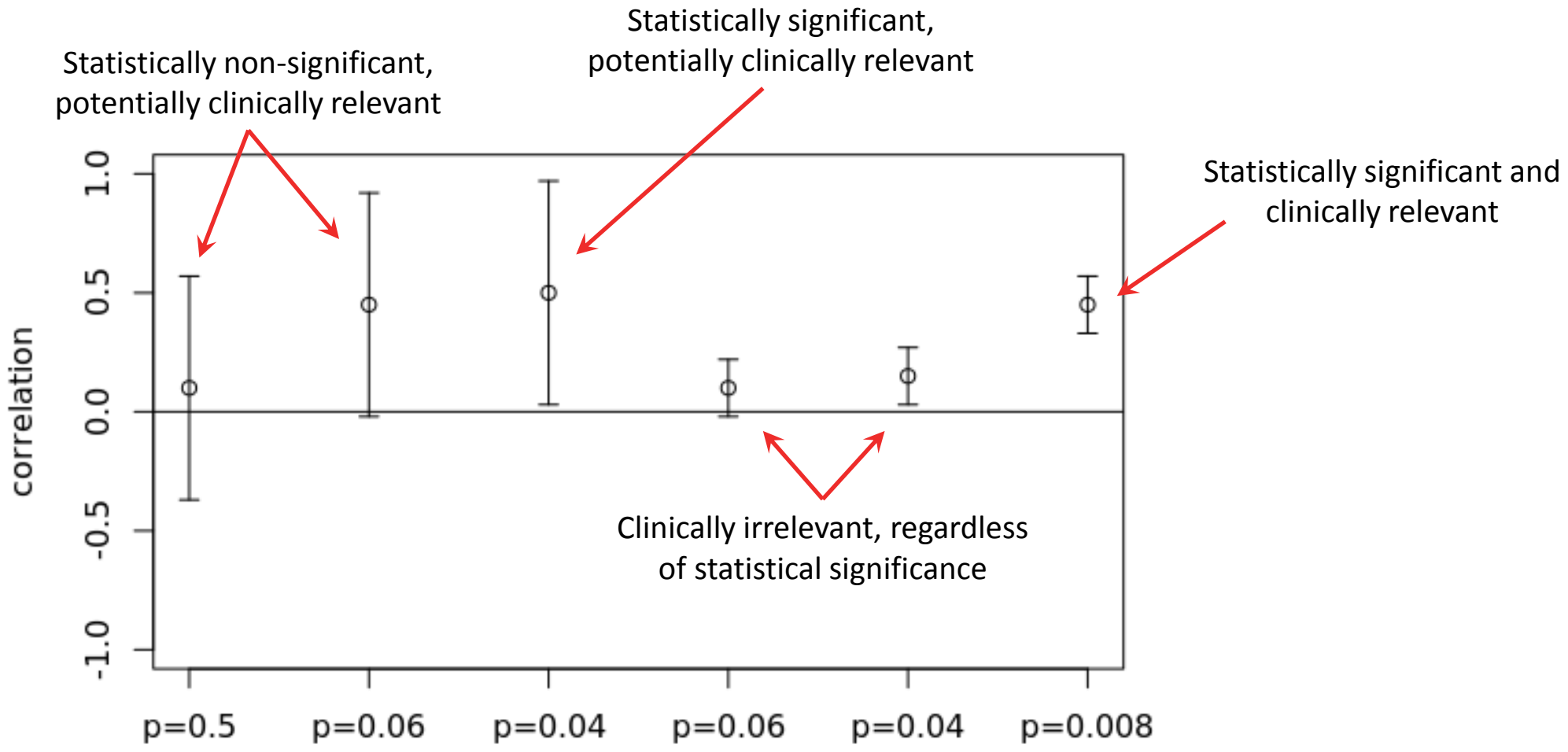
Introduction

- Which side provides the most evidence for a relevant effect between x and y?





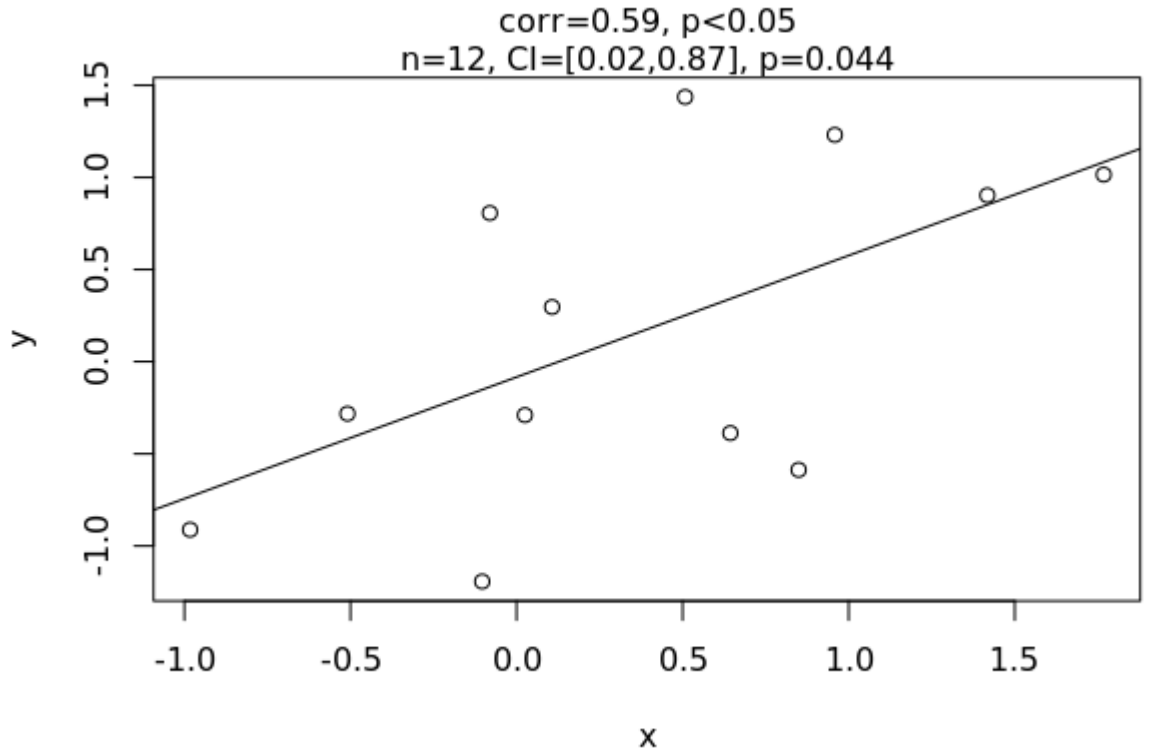
Statistical significance vs Clinical relevance



UNIVERSITY OF CALGARY

Introduction

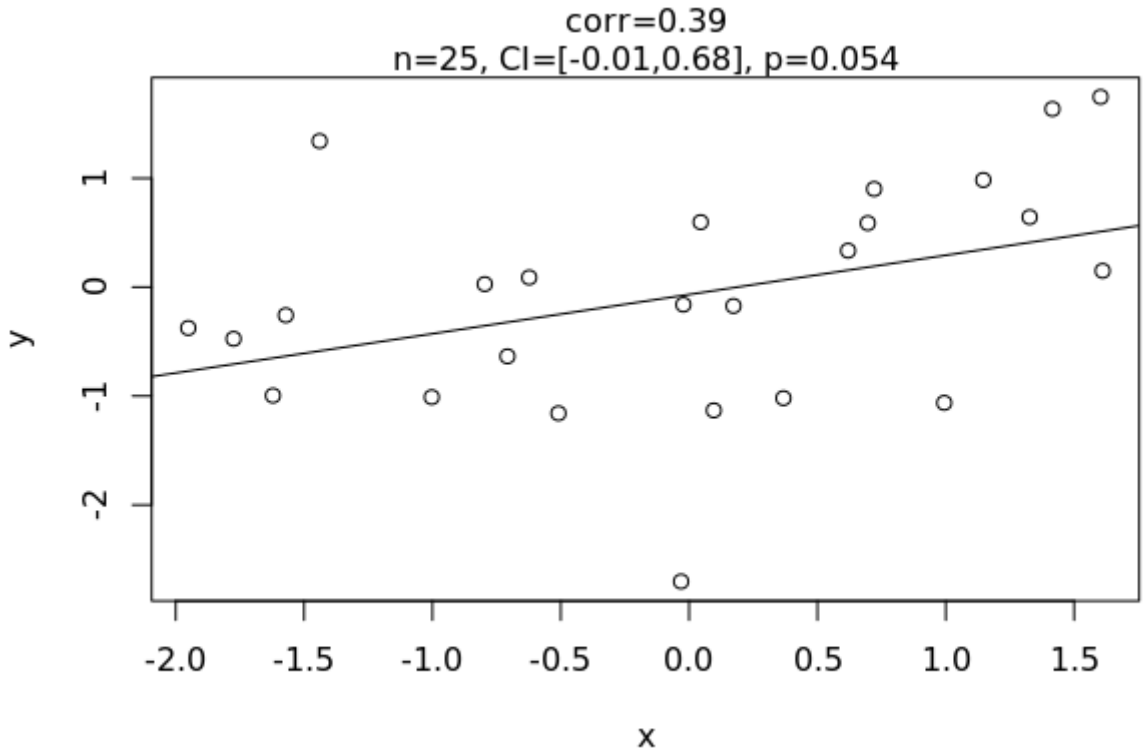
- P-values do not tell the whole story
  - Preferable to also show the data and confidence intervals



UNIVERSITY OF CALGARY

Introduction

- However, we recognize that it is sometimes necessary to make a binary decision: is a given hypothesis true or not?
  - P-value=0.05 provides an objective (although arbitrary) decision threshold

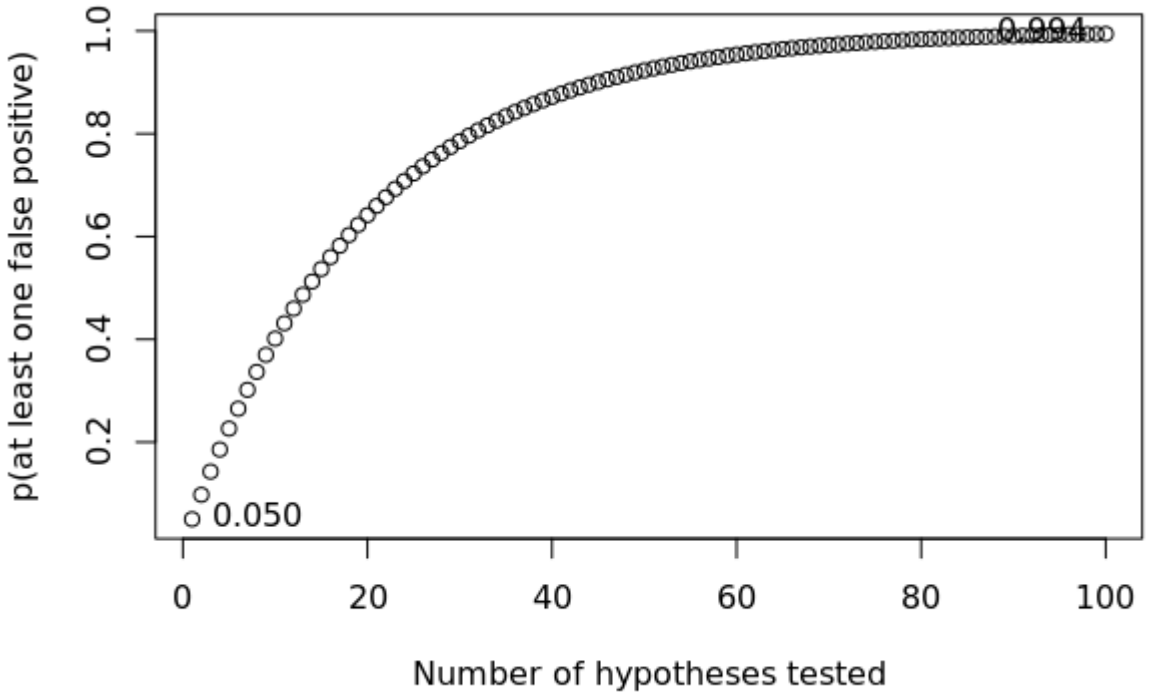


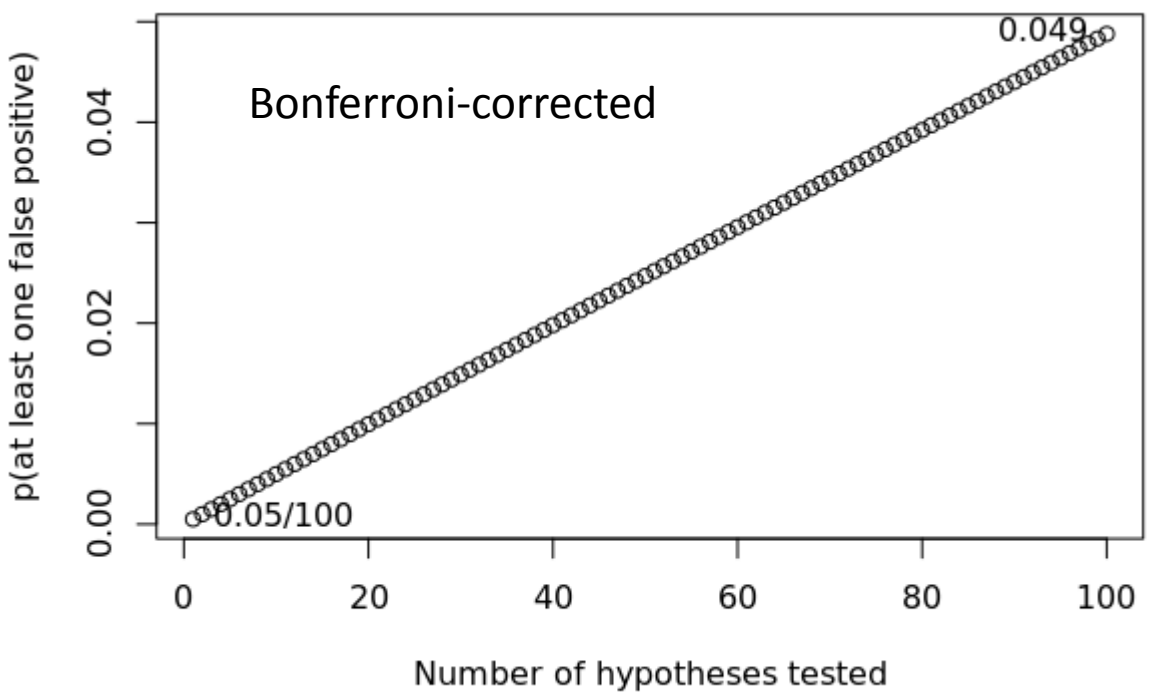
- But this leads to frustration when p is marginally non-significant
  - Probability that this occurred by chance is only 5.4%
  - Something must be done to justify all the time and grant money invested on this study!

UNIVERSITY OF CALGARY

P-Hacking

- Everyone does it, you are missing out if you don‘t!
  - Often it is done unintentionally or unknowingly
- It works thanks to the problem of multiple comparisons





P-Hacking tutorial in R

- GitHub repository: pierrelevan/mritogether

P-Hacking

- Adding data will always (eventually) give you the result you want
  - p-values follow a random walk
  - Do not stop experiments after partial data collection unless you understand the effect on p-values
- You can always find a good reason to remove undesired data points
  - Criteria for data rejection should be defined prior to data collection
- If you search long enough, you can always find a statistical test that gives a desired conclusion
  - Data analysis should be defined prior to data collection

P-Hacking

- If you make decisions after looking at the data, you have already implicitly made multiple tests

Control CBF

Treatment CBF

„Images reveal a CBF reduction in piriform cortex. This reduction was confirmed to be statistically significant (t-test, $p<0.05$)“

Confirmatory Studies vs Exploratory Studies

- Goal of confirmatory studies:
  - Validate a defined hypothesis
  - Everything (hypotheses, analysis steps) has to be defined prior to data collection
- Goal of exploratory studies:
  - Identify interesting new hypotheses
  - Statistical significance is less relevant (even marginally non-significant results may warrant further investigation)
  - Be transparent: show the data, report confidence intervals, the goal is to estimate potentially relevant effect sizes

Thank you for your attention!