SoK: I have the (Developer) Power!
Sample Size Estimation for Fisher‘s Exact, Chi-Squared, McNemar‘s, Wilcoxon Rank-Sum, Wilcoxon Signed-Rank and t-tests in Developer-Centered Usable Security

Anna-Marie Ortloff[1], Christian Tiefenau[2], Matthew Smith[1,2]

[1]: University of Bonn, [2]: Fraunhofer FKIE

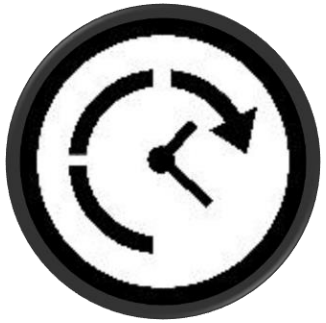
Behavioural Security Group

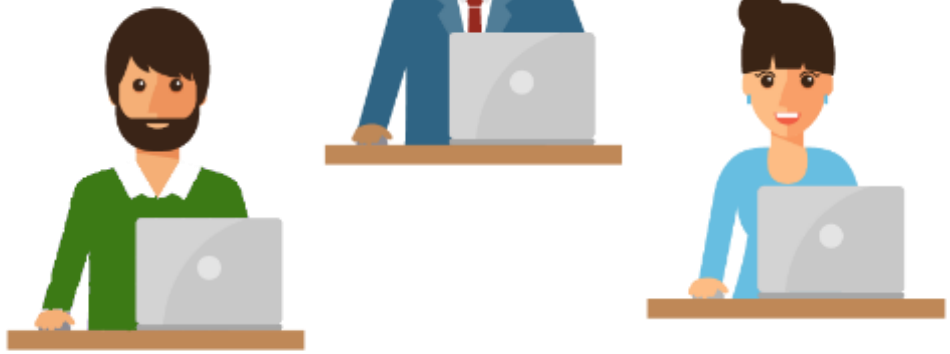
UNIVERSITÄT BONN

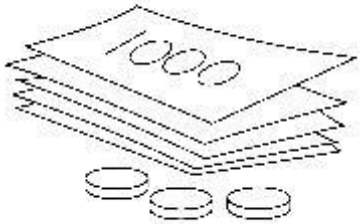
Statistical power:

The probability of detecting an effect, if a true effect exists.

Developer-Centered Usable Security (DCUS)







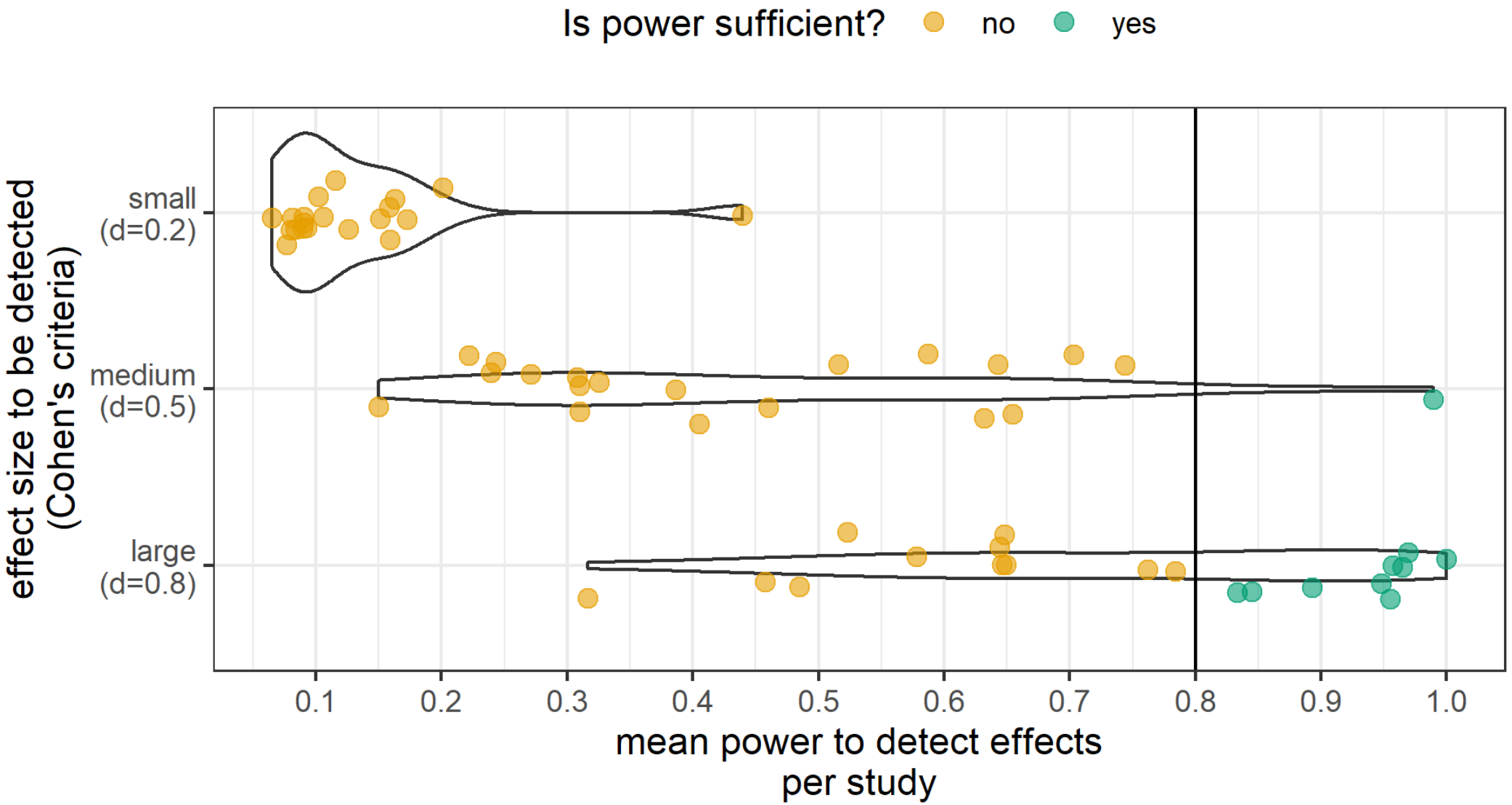
Behavioural Security Group

UNIVERSITÄT BONN

DO WE HAVE THE POWER?

Power Meta-Analysis (simulated a-priori power analysis)

Power Meta-Analysis (simulated a-priori power analysis)

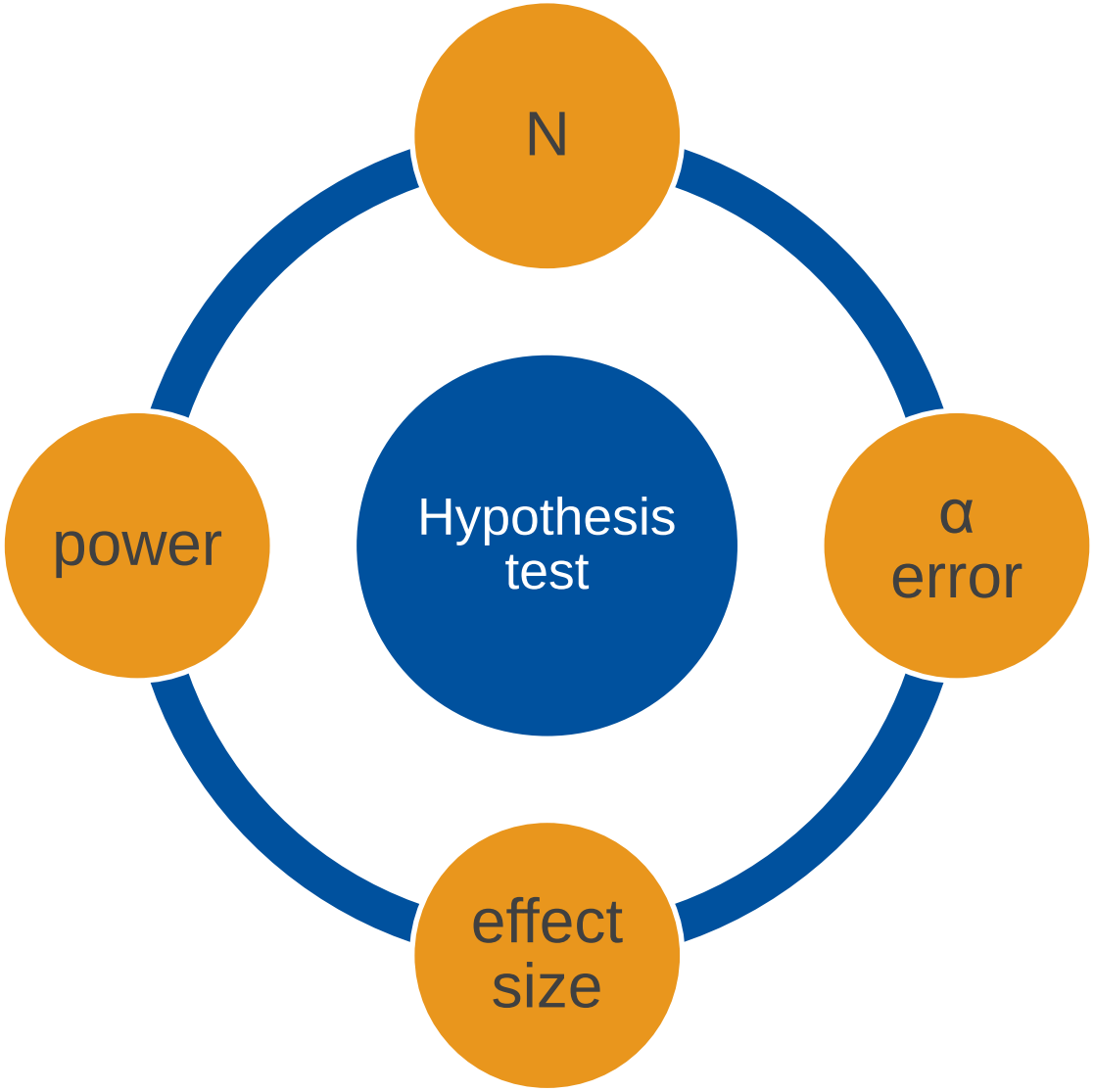


We‘re not using power analysis.

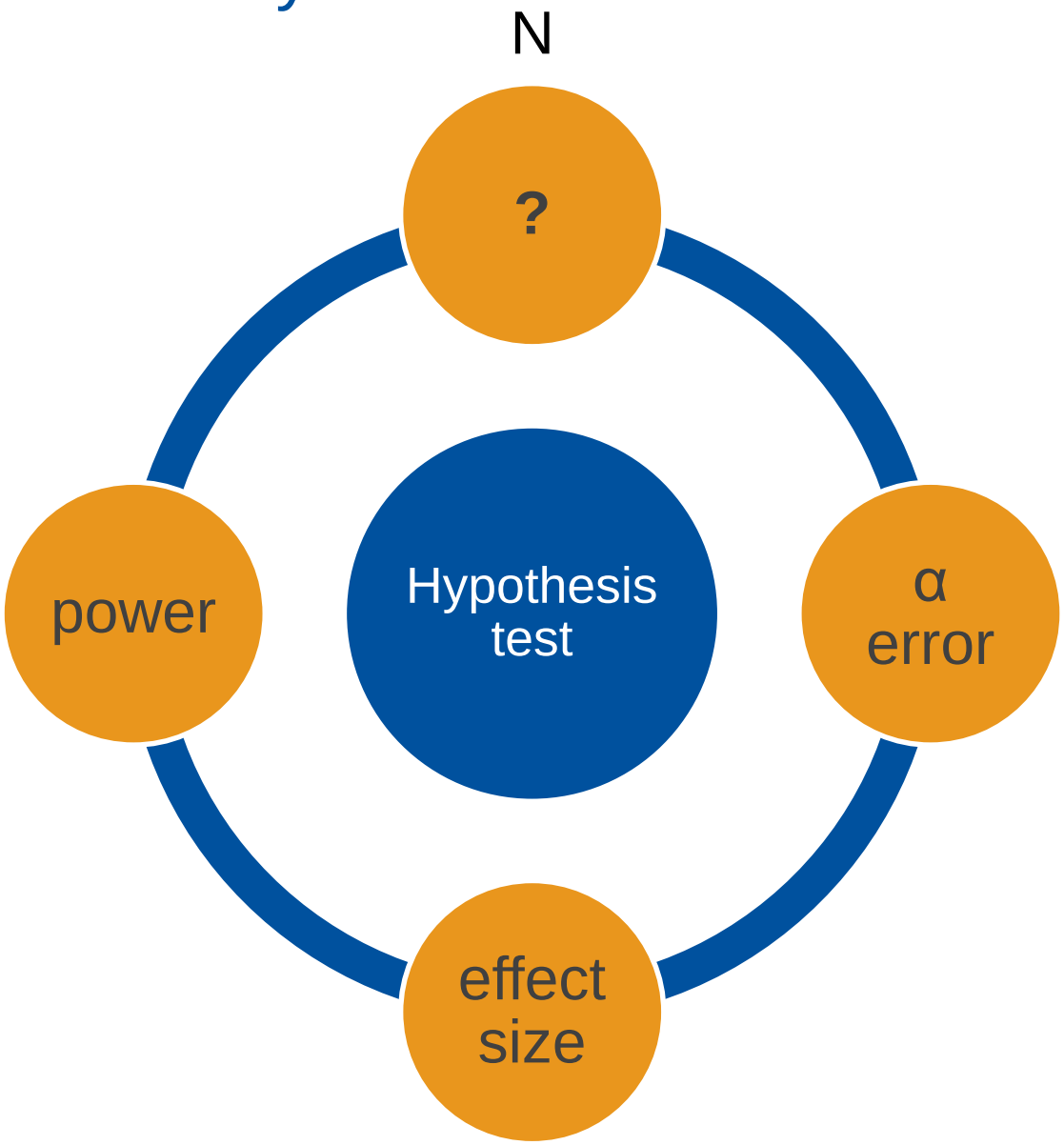
In SOUPS and CHI USP publications from 2020/2021
only 8 of 74 (10.8%) quantitative papers used a priori power analysis

HOW CAN WE GET THE POWER?

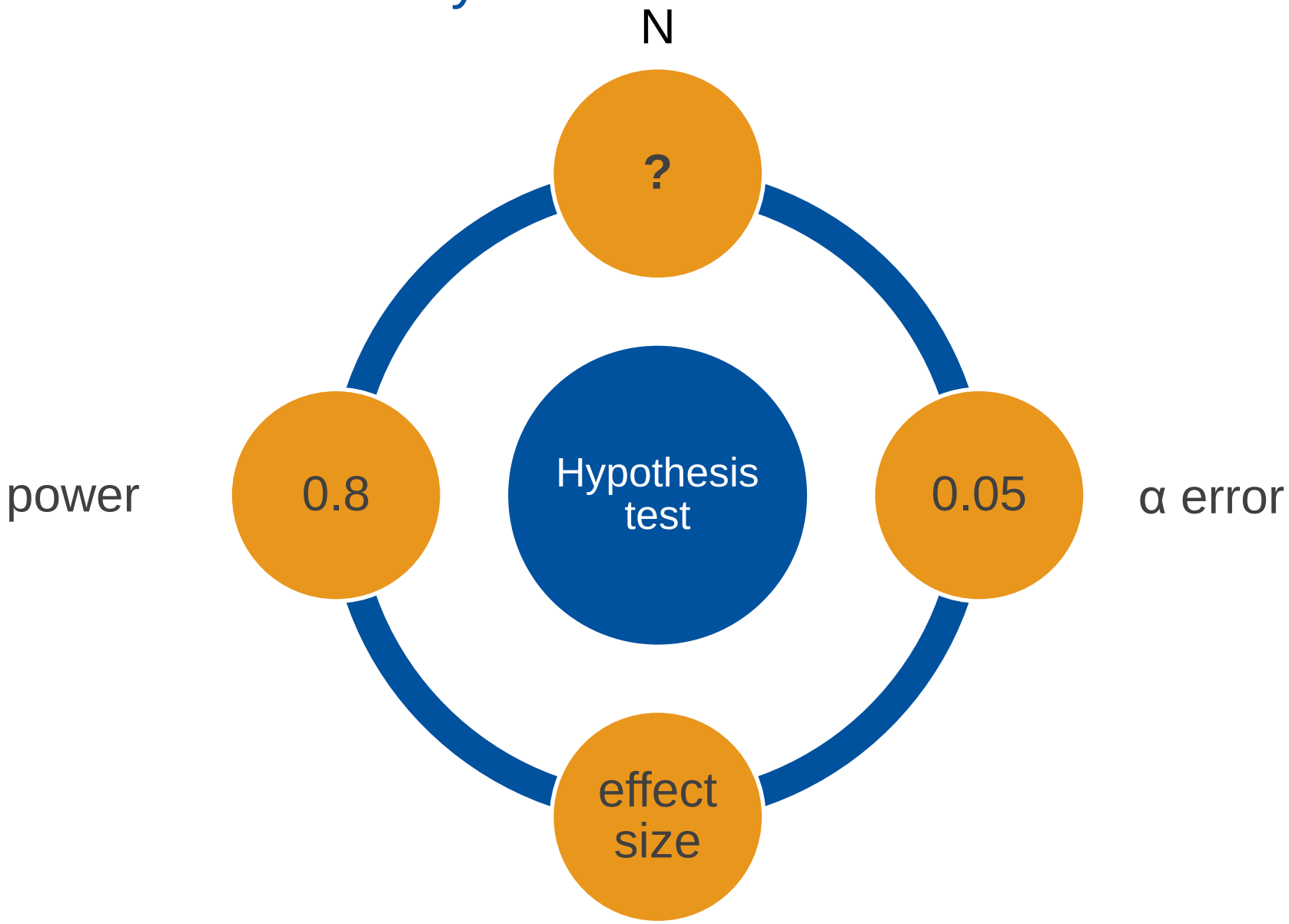
Power Analysis



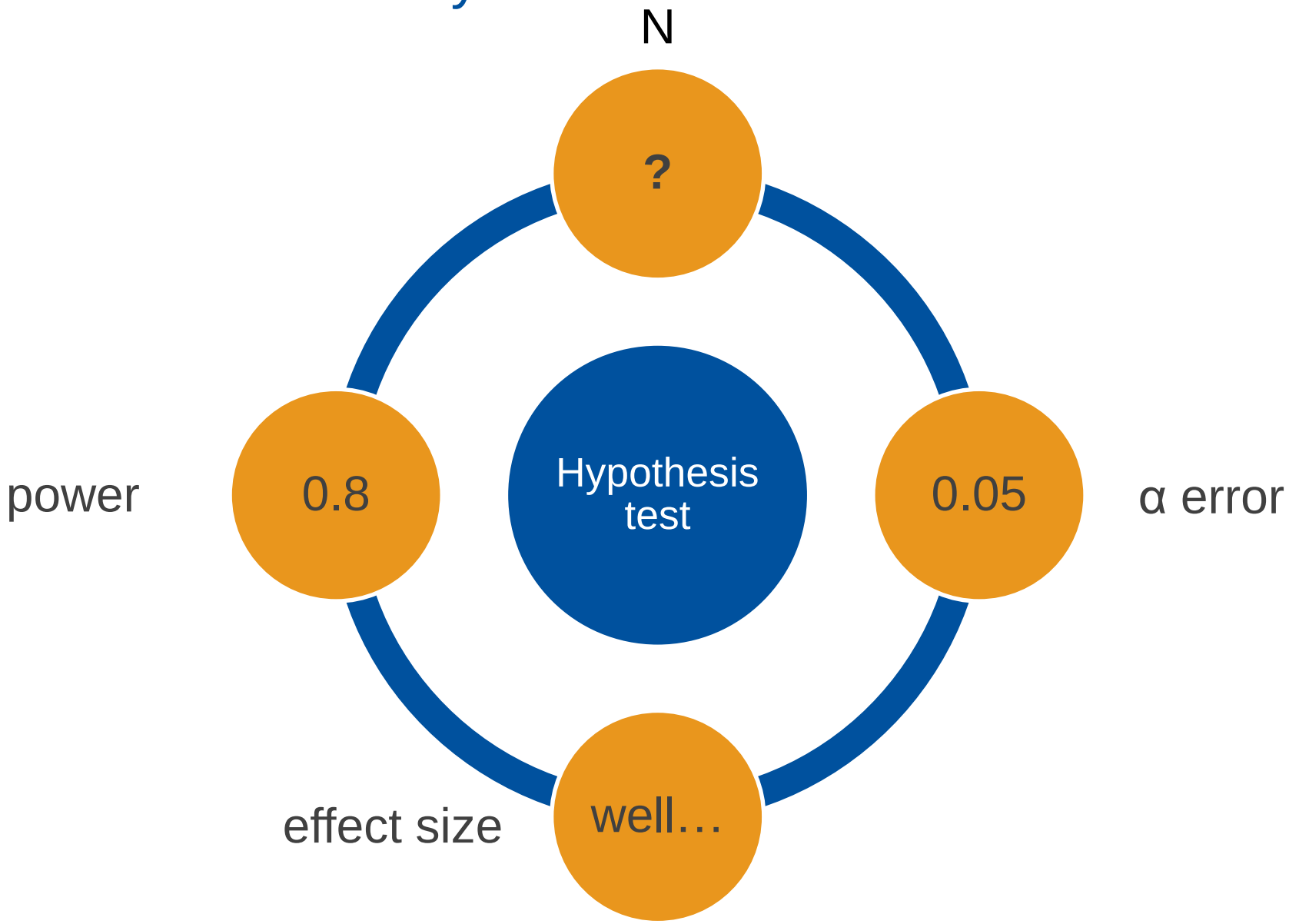
A Priori Power Analysis



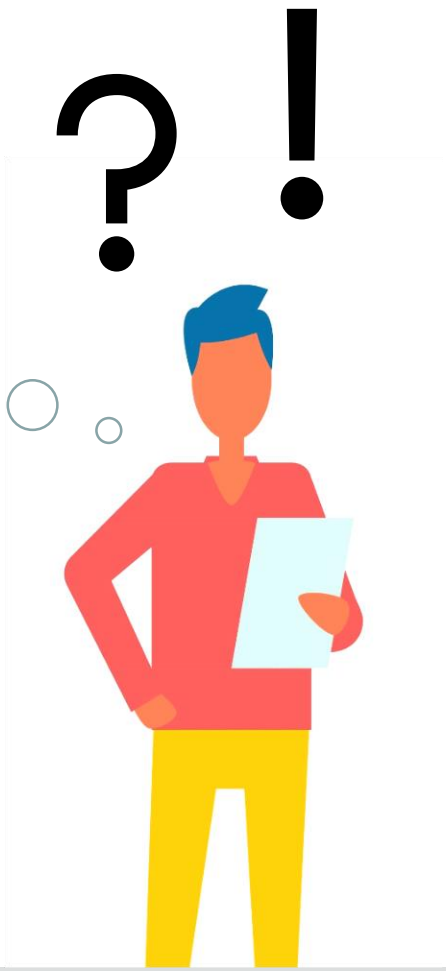
A Priori Power Analysis



A Priori Power Analysis



Which effect size should I use?



- Use (arbitrary) guidelines for large, medium, small effects
- Literature research
- Do a pilot study
- Decide on the smallest effect size of interest

Method

Literature Collection

- SOUPS, USENIX Security, S&P, CCS, ICSE, USP Tracks of CHI
- 2010 - 2021
- Include user study
- Participants: software developers, similar expert users, or proxies
- Domain of usable security and privacy

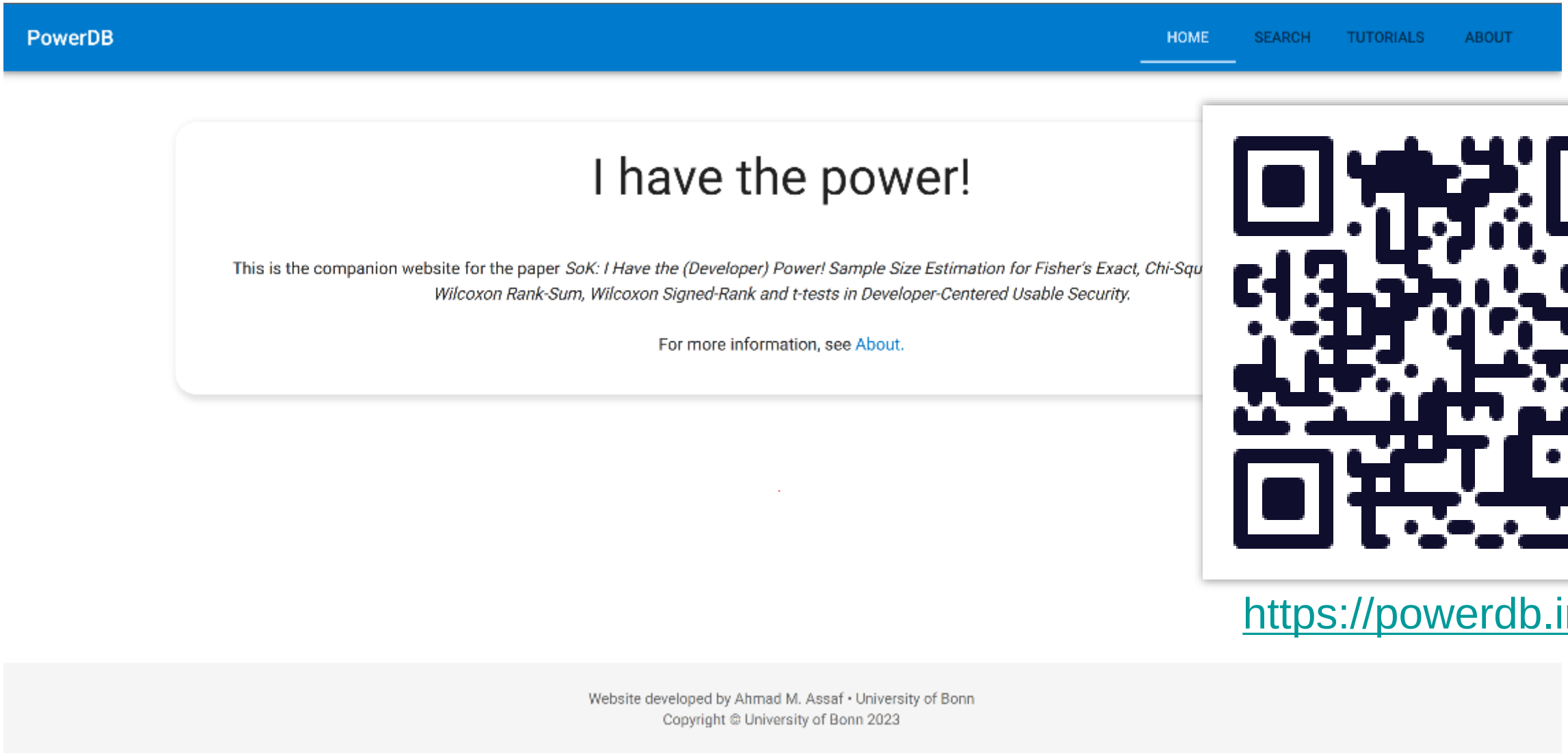
- 54 papers
- including 64 studies, 467 hypothesis tests, 413 variables

Data Structure

- Relevant information on power and effect sizes in these studies

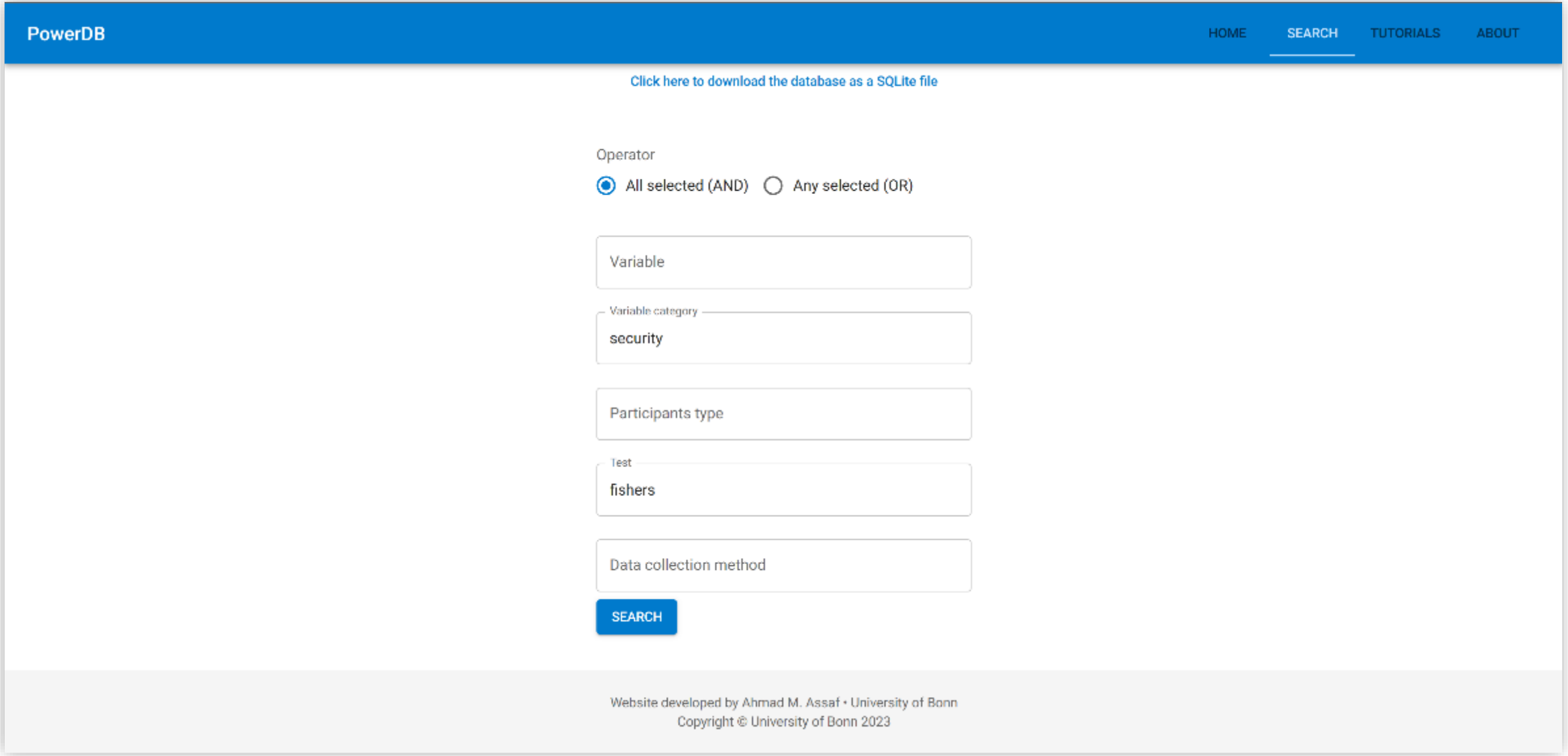
USING THE POWER DATABASE

Database + Companion Website

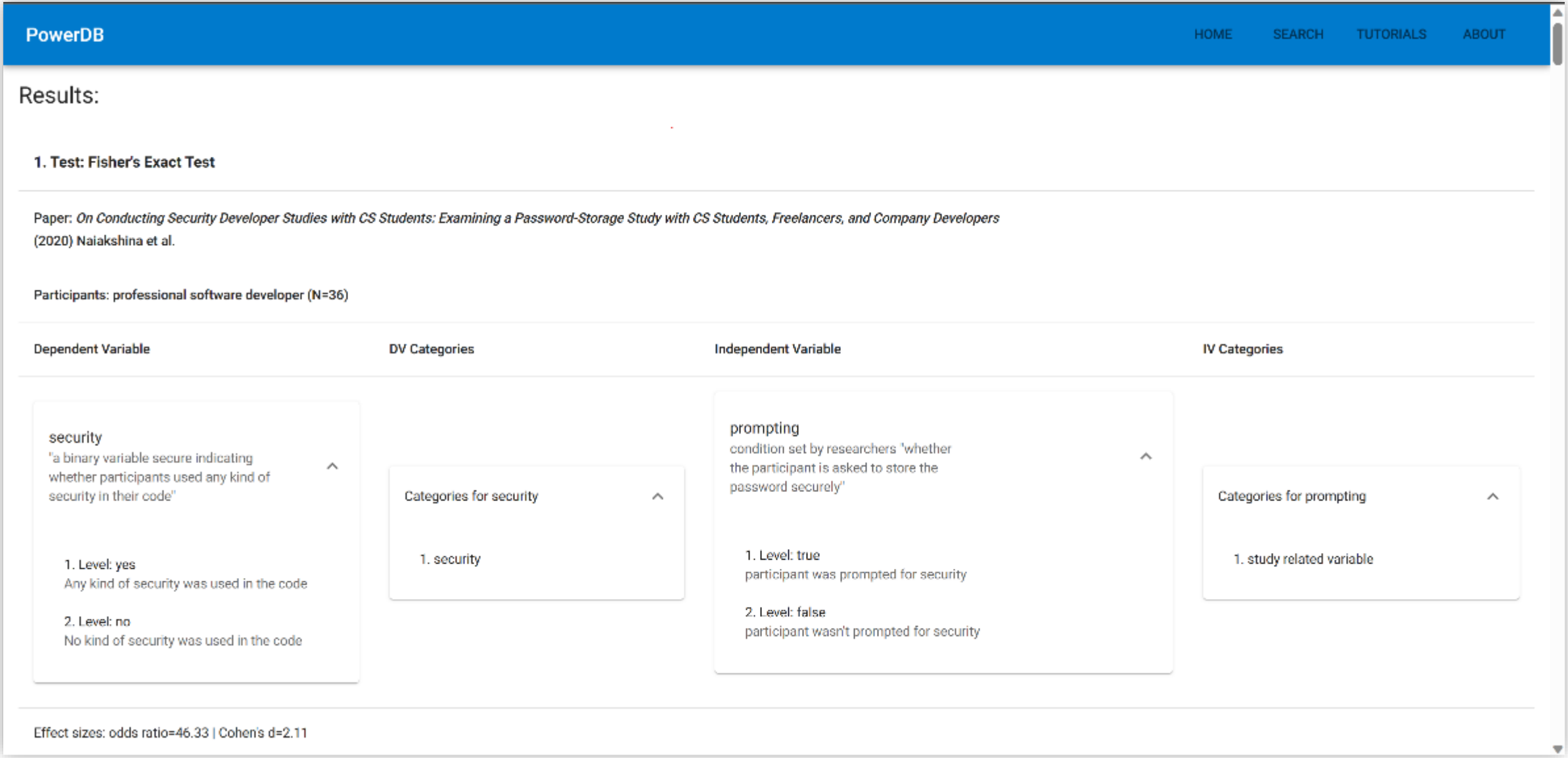


https://powerdb.info/

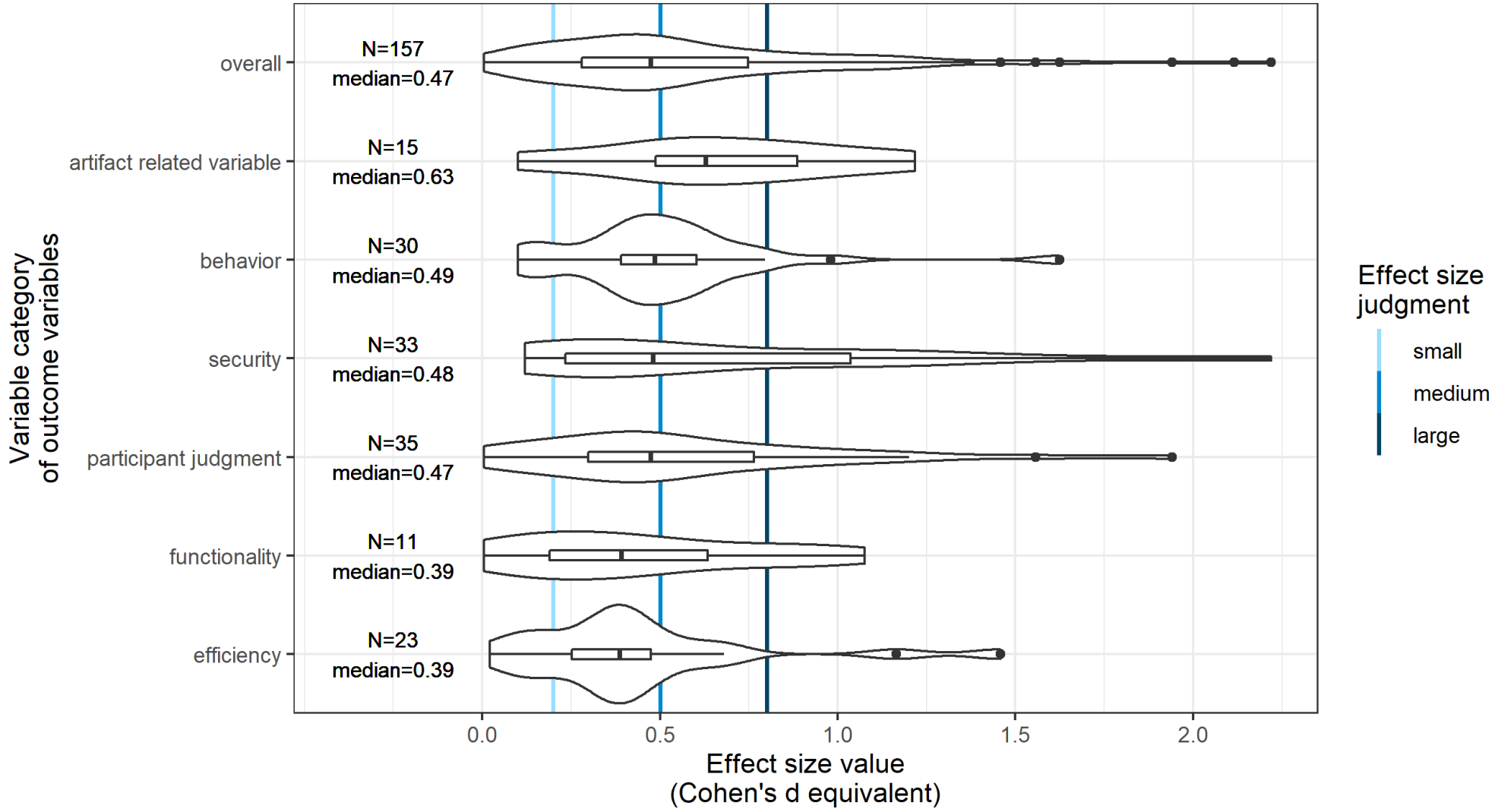
Searching the database



Searching the database



Effect sizes in DCUS



More Meta Analysis?

- How do **you** interpret effect sizes?

Come **talk to me** at SOUPS about:

- Power Analysis (this work)
- Effect sizes (on going)
- Anything else meta
- …

Behavioral Security and Privacy Group

ortloff@cs.uni-bonn.de
https://amortloff.github.io