

SoK: I have the (Developer) Power!

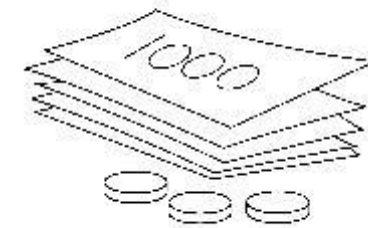
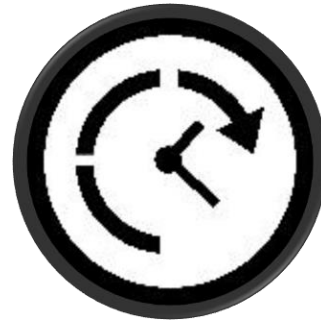
Sample Size Estimation for Fisher's Exact, Chi-Squared, McNemar's, Wilcoxon Rank-Sum, Wilcoxon Signed-Rank and t-tests in Developer-Centered Usable Security

Anna-Marie Ortloff¹, Christian Tiefenau², Matthew Smith^{1,2}

¹: University of Bonn, ²: Fraunhofer FKIE

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

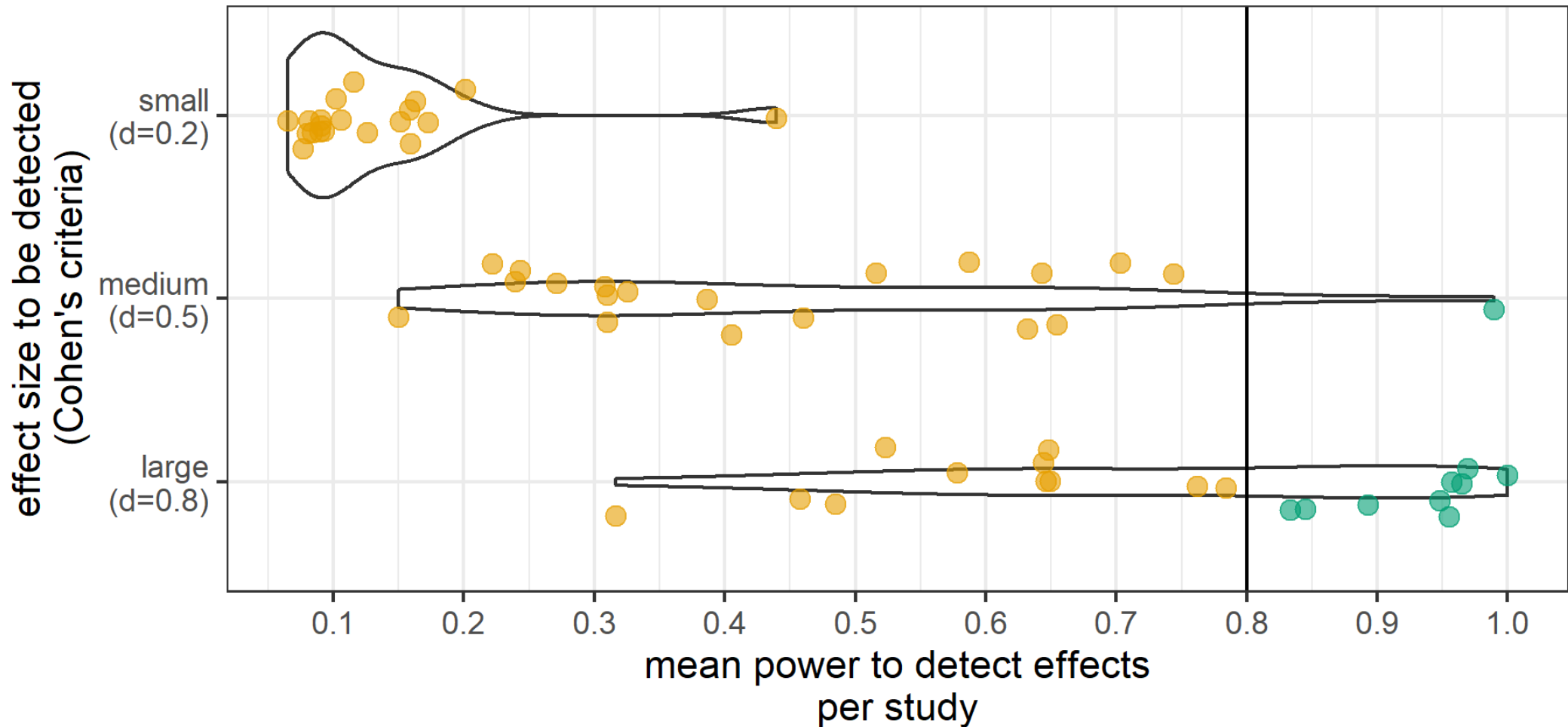
Developer-Centered Usable Security (DCUS)



DO WE HAVE THE POWER?

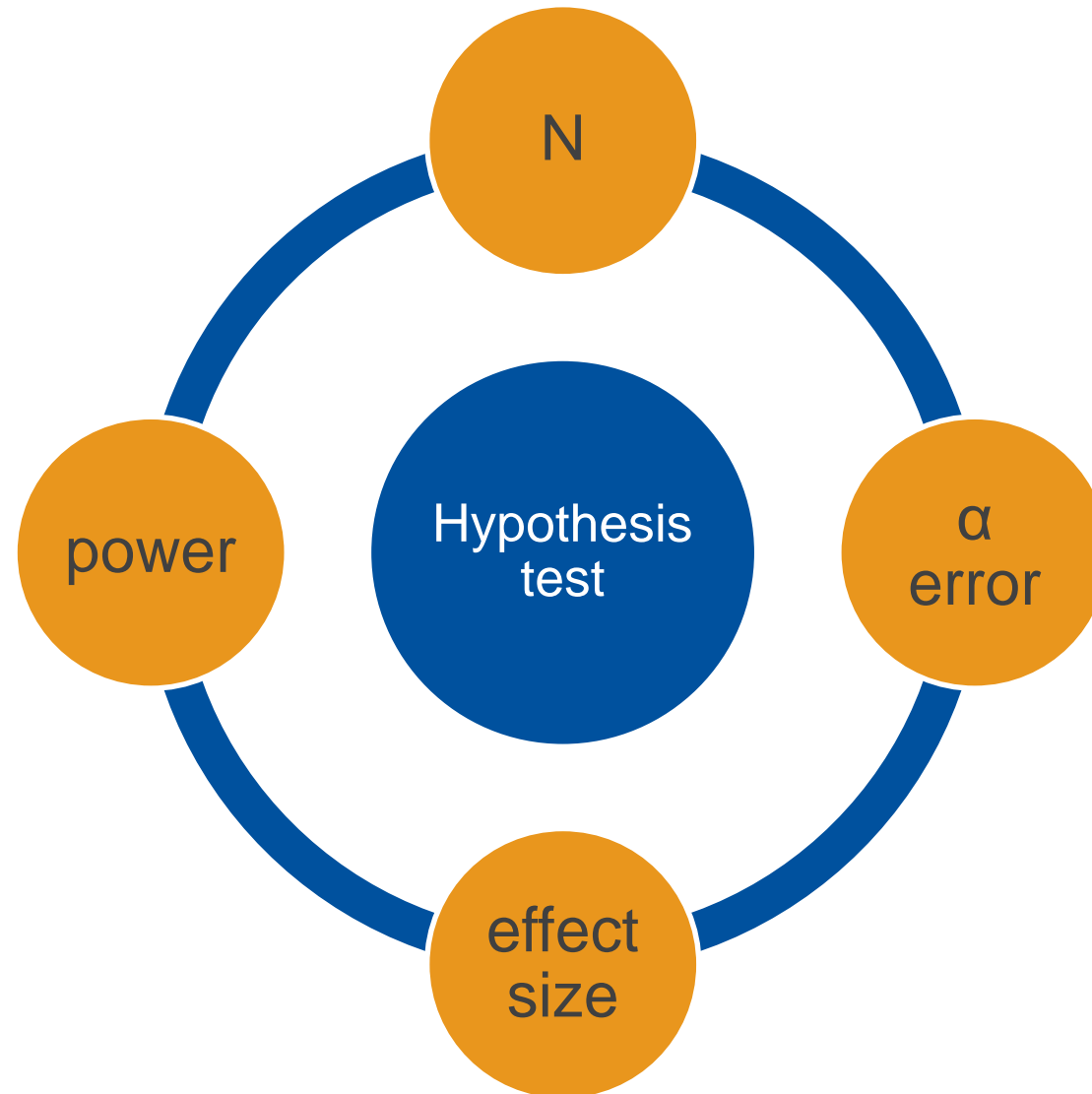
Power Meta-Analysis (simulated a-priori power analysis)

Is power sufficient? ● no ● yes

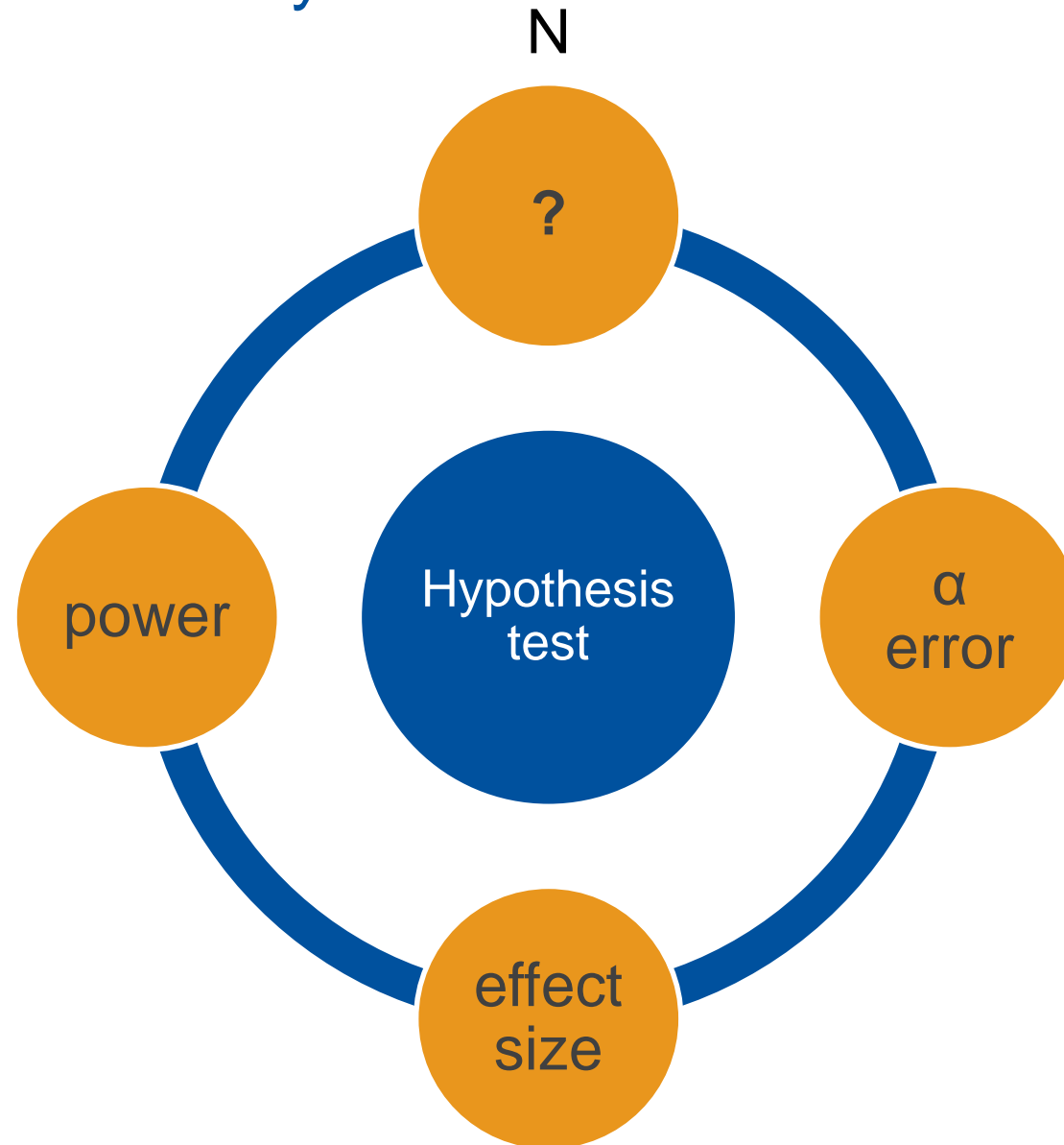


In SOUPS and CHI USP publications from 2020/2021
only 8 of 74 (10.8%) quantitative papers used a priori power analysis

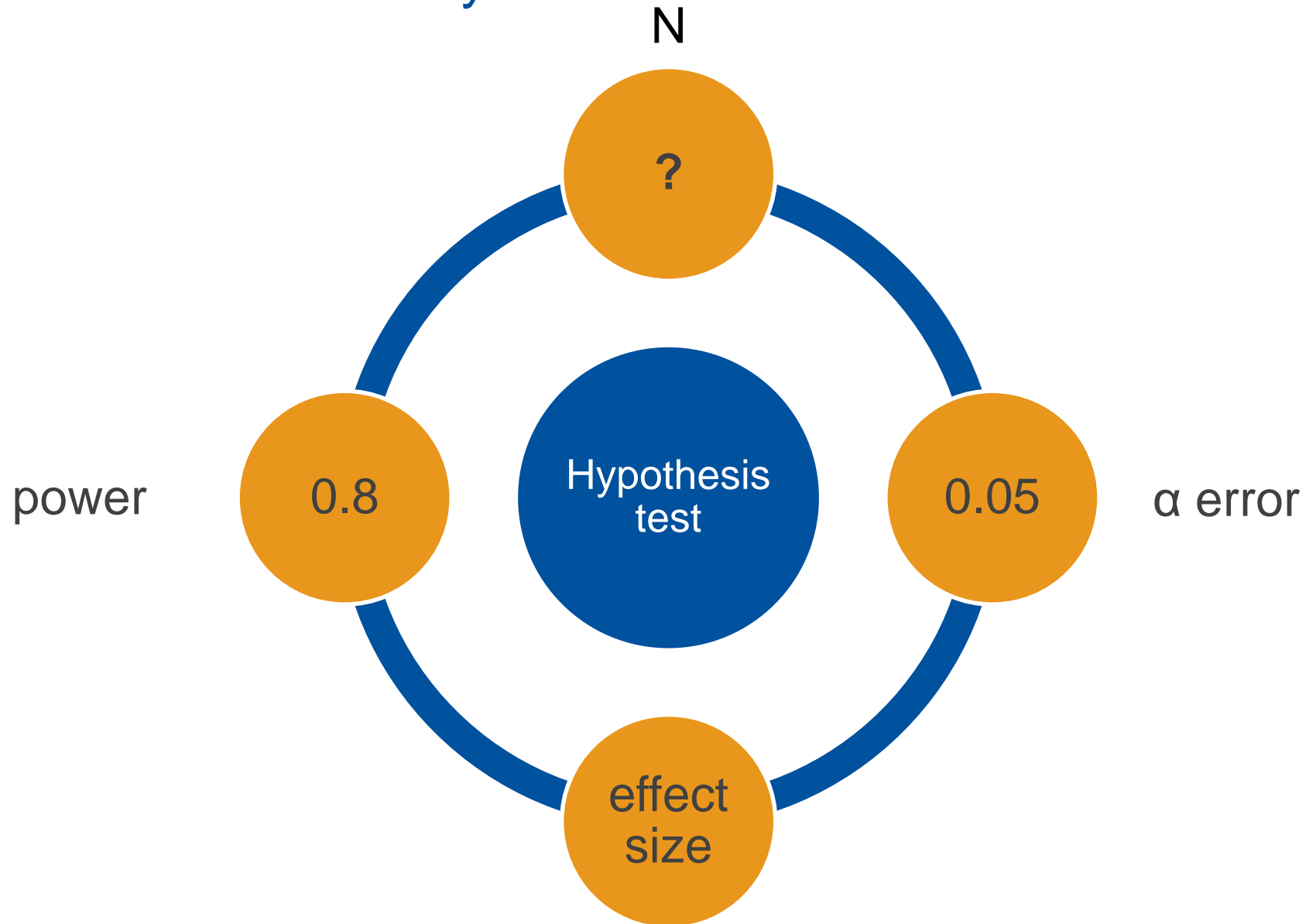
HOW CAN WE GET THE POWER?



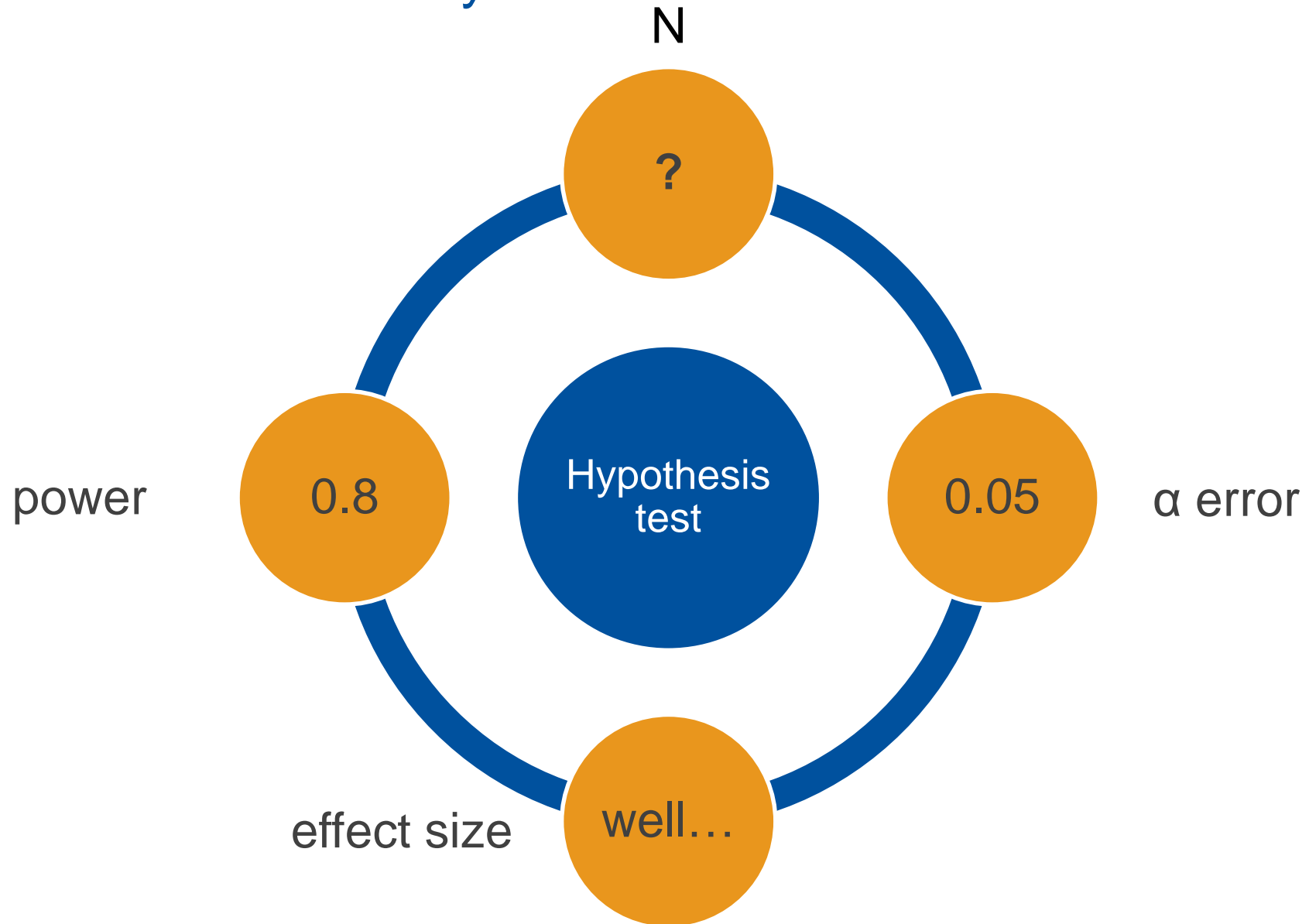
A Priori Power Analysis

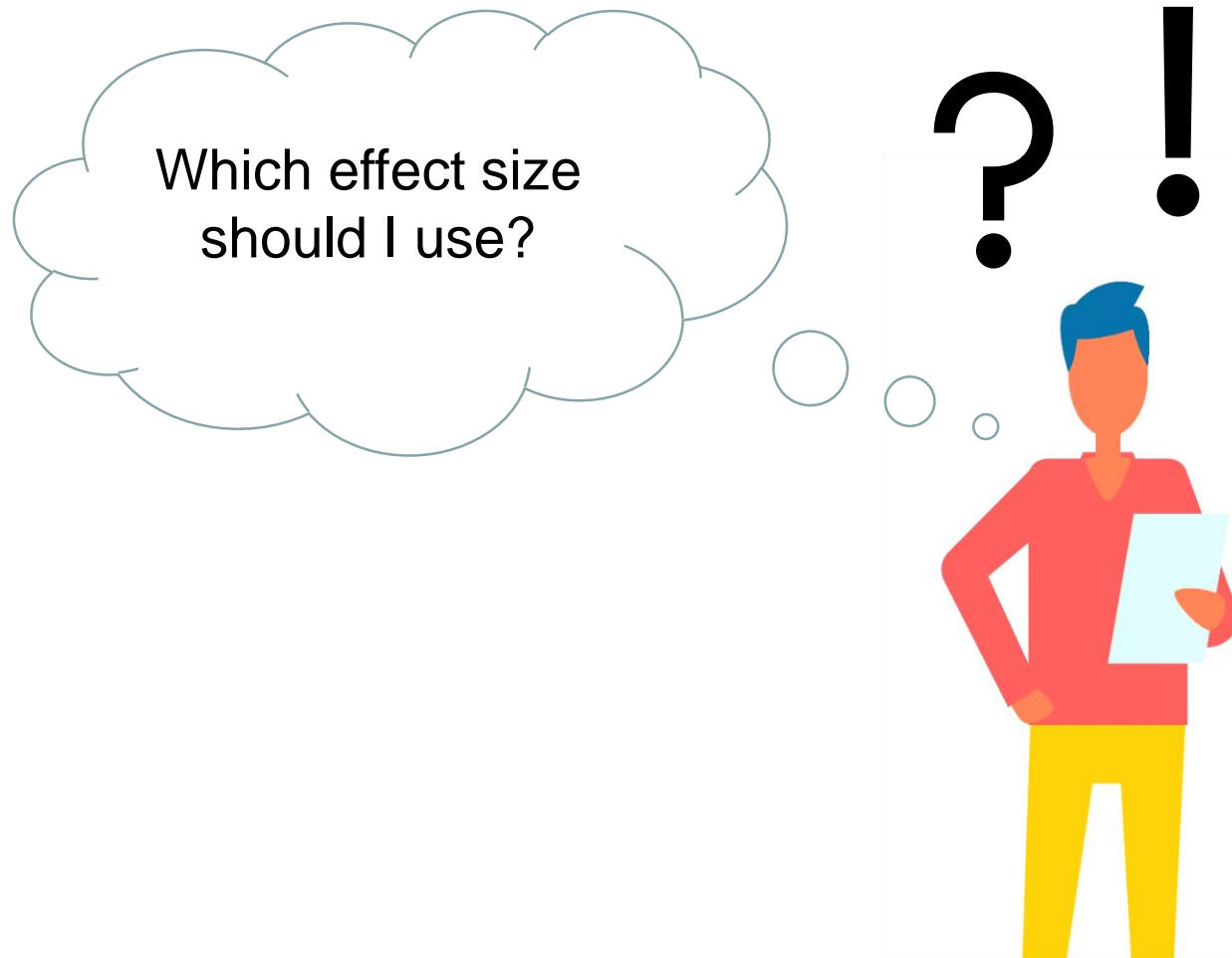


A Priori Power Analysis



A Priori Power Analysis





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

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

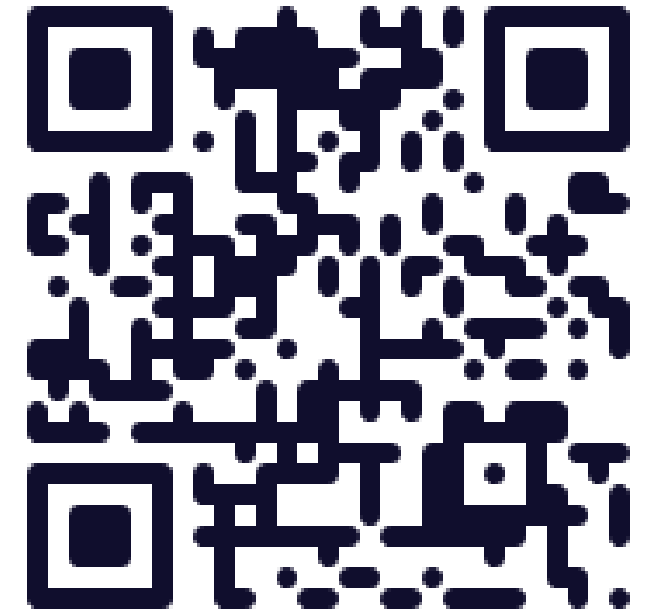


USING THE POWER DATABASE

I have the power!

This is the companion website for the paper *SoK: I Have the (Developer) Power! Sample Size Estimation for Fisher's Exact, Chi-Square, Wilcoxon Rank-Sum, Wilcoxon Signed-Rank and t-tests in Developer-Centered Usable Security*.

For more information, see [About](#).



<https://powerdb.info/>



Behavioural Security Group

Searching the database

PowerDB

HOME

SEARCH

TUTORIALS

ABOUT

[Click here to download the database as a SQLite file](#)

Operator

All selected (AND) Any selected (OR)

Variable

Variable category

security

Participants type

Test

fishers

Data collection method

SEARCH

Website developed by Ahmad M. Assaf • University of Bonn
Copyright © University of Bonn 2023



Results:

1. Test: Fisher's Exact Test

Paper: *On Conducting Security Developer Studies with CS Students: Examining a Password-Storage Study with CS Students, Freelancers, and Company Developers* (2020) Naiakshina et al.

Participants: professional software developer (N=36)

Dependent Variable

DV Categories

Independent Variable

IV Categories

security

"a binary variable secure indicating whether participants used any kind of security in their code"

- 1. Level: yes
Any kind of security was used in the code
- 2. Level: no
No kind of security was used in the code

Categories for security

- 1. security

prompting

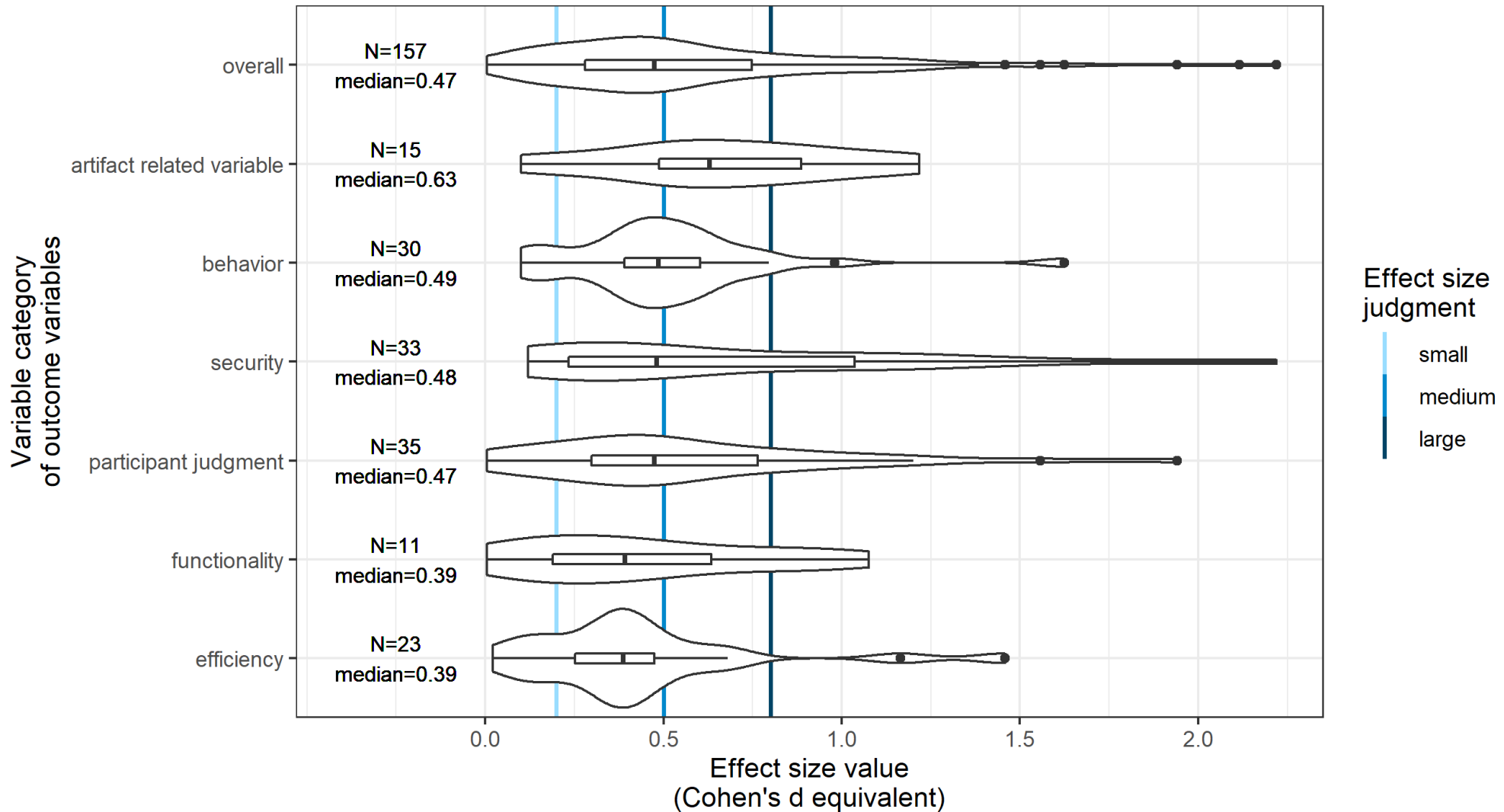
condition set by researchers 'whether the participant is asked to store the password securely'

- 1. Level: true
participant was prompted for security
- 2. Level: false
participant wasn't prompted for security

Categories for prompting

- 1. study related variable

Effect sizes: odds ratio=46.33 | Cohen's d=2.11



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>