



# Why we need protocols in science

Stefan Thoma

2025-08-20



## About me

- MSc Psychology in Methods, Cognition & Perception at Unibe
- MSc in Statistics at ETHZ





## Current work

- At Roche working in a blended role as
  - Statistical programmer
  - Biostatistician
  - R package developer, e.g. {admiral}

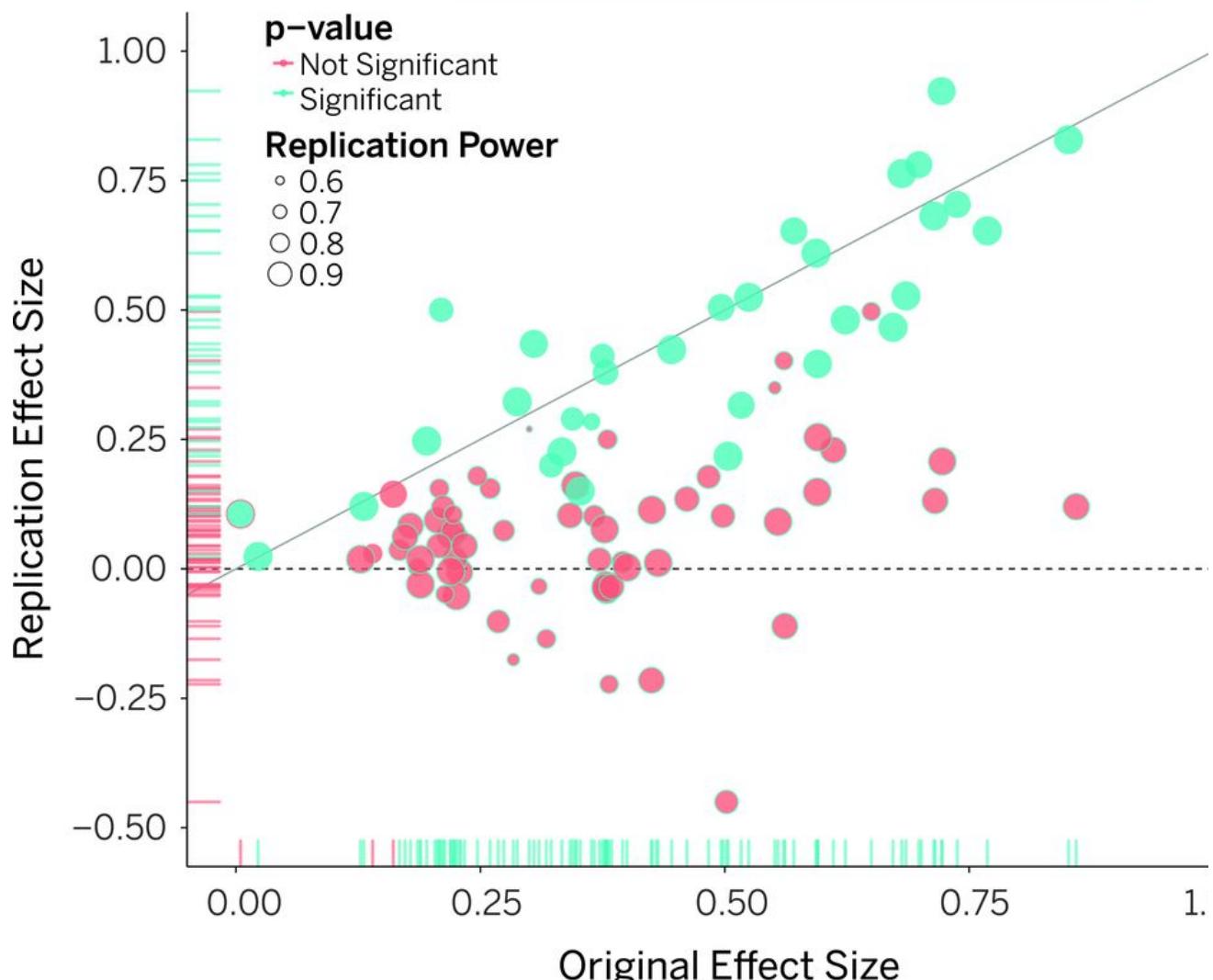




## Replication crisis in Psychology

- Researchers are unable to replicate many findings from psychological experiments.
  - Klein et al. ([2018](#)) managed to replicate only 15 / 28 studies
  - Open Science Collaboration ([2015](#)) replicated only 36 / 100 studies, with on average half of the original effect size
- Yarkoni ([2022](#)): How generalisable are findings from psychological experiments?

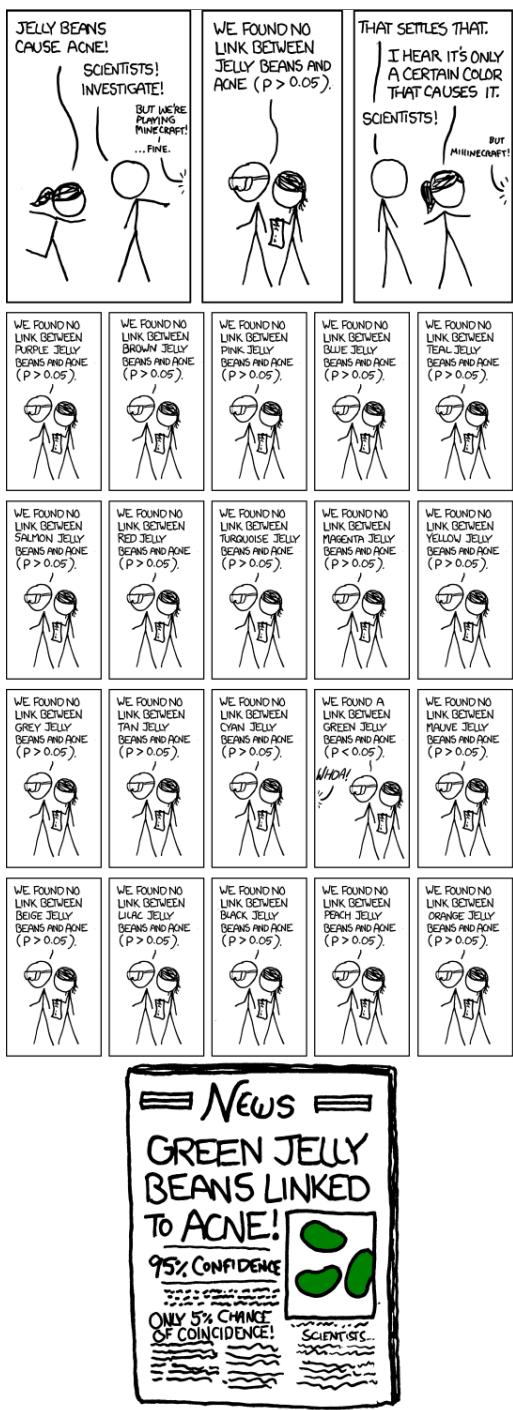
Many labs result

Results of Open Science Collaboration (2015)



## Misaligned incentives for researchers

- Publication bias
  - Harder and less likely to publish null-results
  - More likely to continue looking for something
- Questionable research practices undermine trial integrity
  - Hypothesising After Results are Known (HARK-ing)
  - P-Hacking, e.g. adding covariates until p-value is significant (Tsang (2025))
- Fraud, for an example see Data Colada posts by Simonsohn, Nelson, and Simmons (2023)
  - Hard to spot, as open data was not very common
- Doing replications is not “novel research”, less prestigious





## Realign researcher incentives

Move away from rewarding the most attention-grabbing findings

- Preregistration, see [OSF](#)
  - Support trial integrity / validity
  - Registered reports almost guarantee publication before data collection Chambers ([2019](#))
  - Published based on the merits of the hypothesis and methods, independent of results



From Science ([n.d.](#))



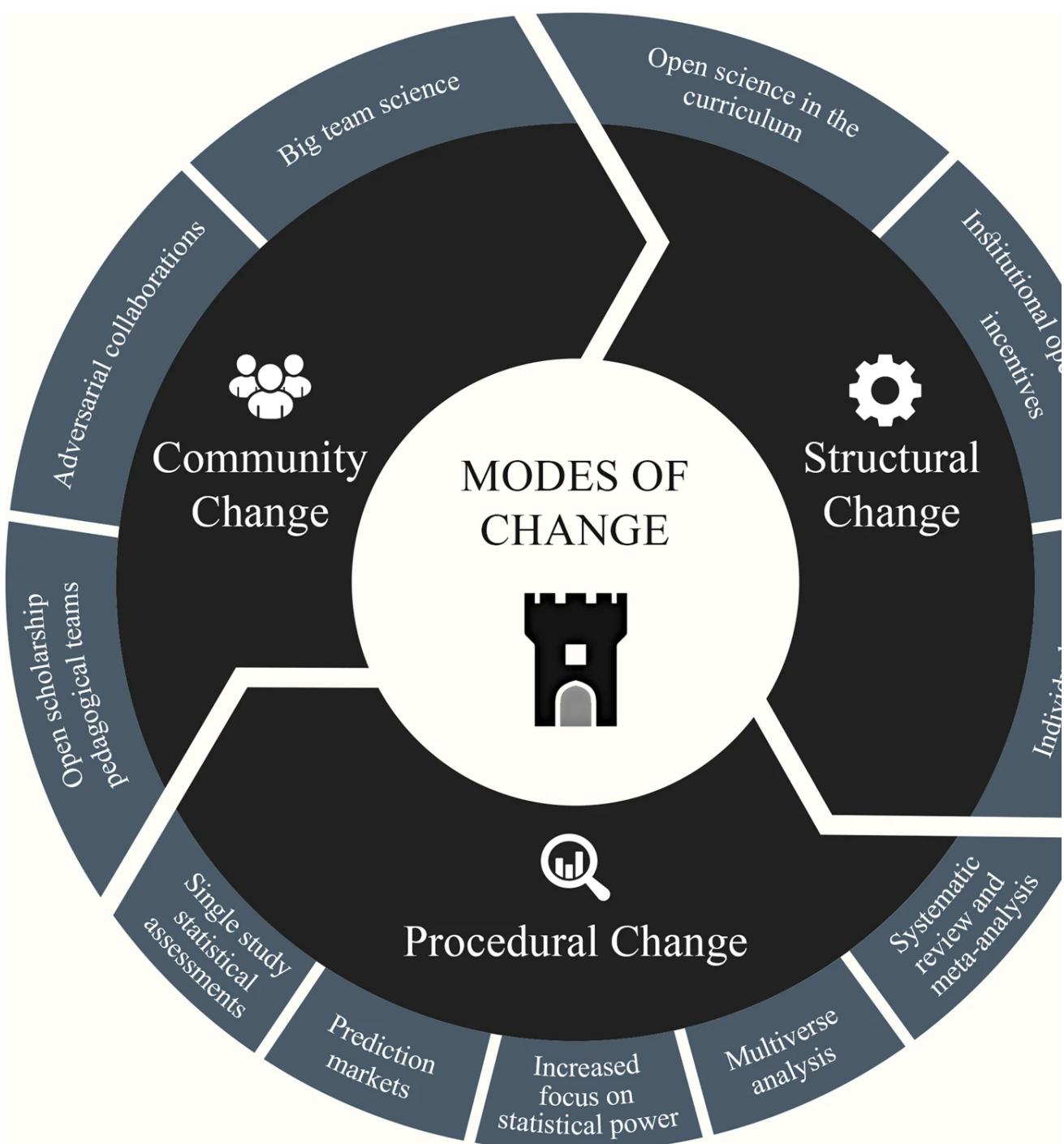
## Open research practices

- Openly available data and methods
- Encourage collaboration
- Reduce ease of fraud
- Openly available publications
- Data as publication rewards data collection
  - Increases data quality
  - Treated as legitimate research



## Things are changing

Structural changes are happening!



From Korbmacher et al. (2023)



## Protocol - Study blueprint

- Includes:
  - Motivation and objectives
  - Primary and key secondary endpoints, as estimands
  - Methodology
  - Organisation
  - Some statistical considerations, e.g. sample size & power, etc.
- Details should allow recreation of the study



## Estimands

An **estimand** (Pohl et al. (2021)) is a precise description of the treatment effect to be estimated in a clinical trial. - “What exactly do we want to learn from the trial?”

- **Key Attributes:**

- **Population:** The specific group of patients of interest.
- **Treatment:** The intervention or comparison of interest.
- **Outcome:** The specific outcome variable being measured.
- **Intercurrent Events:** Events that occur after treatment initiation that affect the interpretation of the outcome (e.g., use of rescue medication, treatment discontinuation).
- **Population-level Summary:** The summary measure that will be used to quantify the treatment effect (e.g., difference in means, hazard ratio).



## Statistical analysis plan

- Describes the precise statistical analysis for each estimand.
- Written before looking at the data
- Confirmatory analysis with less researchers degree of freedom
- Exploratory analysis still encouraged, but must be declared as such



## Regulation - adversary?

- Encourages good scientific practice
- Ensures transparency and reproducibility
- Encourages good scientific practice
- Sets standards in data quality and documentation
- Hurdle that every project needs to pass



## Summary

- Protocols or Preregistrations ensure study integrity
  - Your shield!
- Estimands define the question to be answered
  - Take a closer look at Estimands for your own project
- Statistical Analysis Plans ensure validity of confirmatory results
- Regulators ensure good scientific practice
  - Same objective - to ensure the safety and efficacy of the drug



## Resources

- Chambers, Chris. 2019. “What’s Next for Registered Reports?” *Nature* 573 (7773): 187–89. <https://doi.org/10.1038/d41586-019-02674-6>.
- Klein, Richard A., Michelangelo Vianello, Fred Hasselman, Byron G. Adams, Reginald B. Adams, Sinan Alper, Mark Aveyard, et al. 2018. “Many Labs 2: Investigating Variation in Replicability Across Samples and Settings.” *Advances in Methods and Practices in Psychological Science* 1 (4): 443–90. <https://doi.org/10.1177/2515245918810225>.
- Korbmacher, Max, Flavio Azevedo, Charlotte R. Pennington, Helena Hartmann, Madeleine Pownall, Kathleen Schmidt, Mahmoud Elsherif, et al. 2023. “The Replication Crisis Has Led to Positive Structural, Procedural, and Community Changes.” *Communications Psychology* 1 (1): 1–13. <https://doi.org/10.1038/s44271-023-00003-2>.
- Open Science Collaboration. 2015. “Estimating the Reproducibility of Psychological Science.” *Science* 349 (6251): aac4716. <https://doi.org/10.1126/science.aac4716>.
- Pohl, Moritz, Lukas Baumann, Rouven Behnisch, Marietta Kirchner, Johannes Krisam, and Anja Sander. 2021. “Estimandsa Basic Element for Clinical Trials.” *Deutsches Ärzteblatt International* 118 (51-52): 883–88. <https://doi.org/10.3238/arztebl.m2021.0373>.
- Science, Center for Open. n.d. “Registered Reports: Peer Review Before Results Are Known to Align Scientific Values and Practices.” <https://www.cos.io/initiatives/registered-reports>.
- Simonsohn, Uri, Leif Nelson, and Joe Simmons. 2023. “[109] Data Falsificada (Part 1): “Clusterfake”.” <https://datacolada.org/109>.
- Tsang, Benjamin. 2025. “P Hacking Five Ways It Could Happen to You.” *Nature*, May. <https://doi.org/10.1038/d41586-025-01246-1>.
- Yarkoni, Tal. 2022. “The Generalizability Crisis.” *Behavioral and Brain Sciences* 45 (January): e1. <https://doi.org/10.1017/S0140525X20001685>.