Easing into Open Science

ReproducibiliTea Journal Club Geneva

01.11.2024

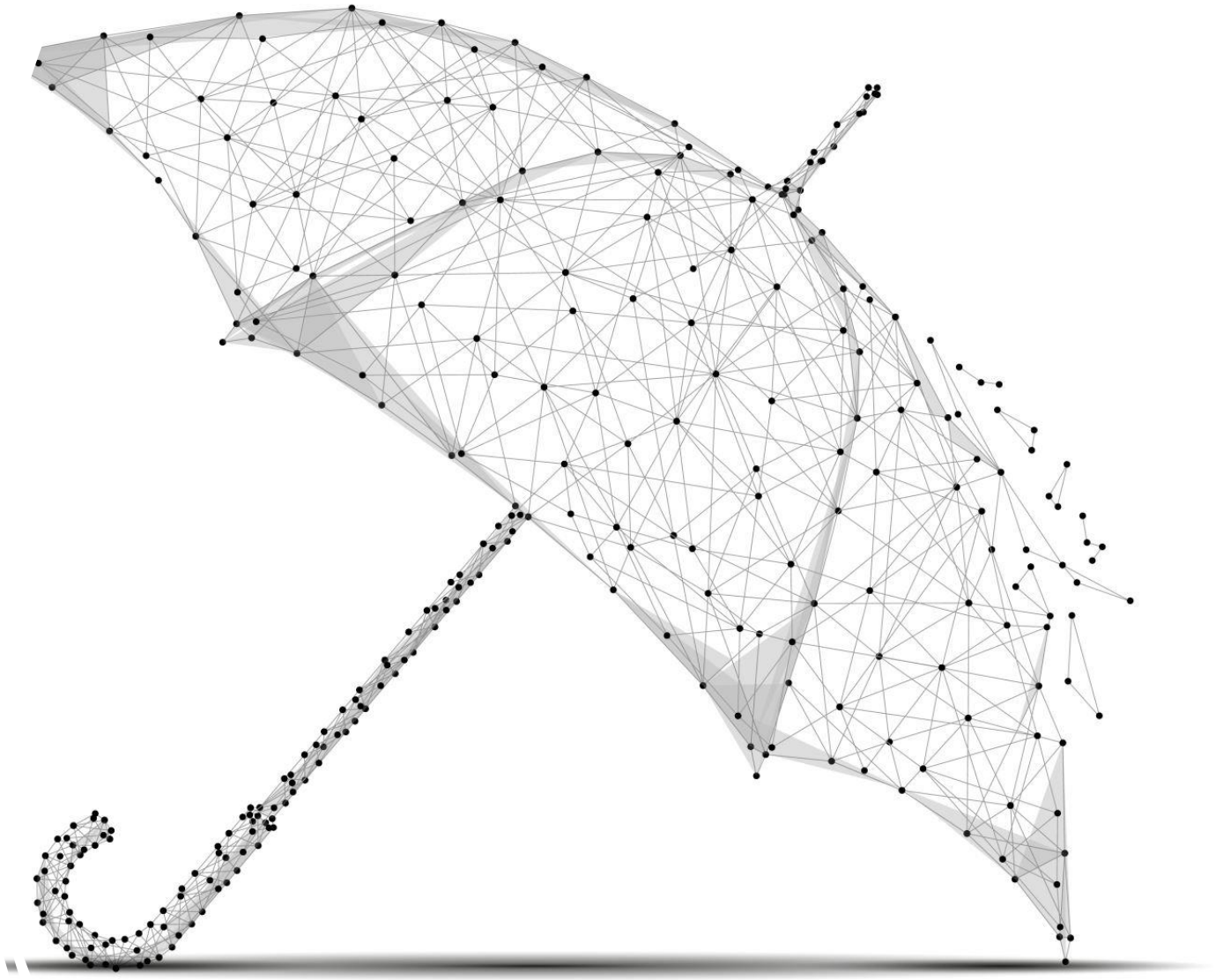
The slides were created with the help of

- 2023 CUSO Open Science Workshop by Dominique Roche: https://osf.io/4g8tk/
- SIRRO Preregistration Workshop by Caro Hautekiet and Evie Vergauwe: https://osf.io/g45nt/
- Lecture Slides “Best Practice in Pedagogy” by William Chopik and colleagues https://osf.io/mh9pe/
- ”Easing Into Open Science: A Guide for Graduate Students and Their Advisors” (Kathawalla et al., 2021)
- “Seven Easy Steps to Open Science” (Crüwell et al., 2019)

What is Open Science

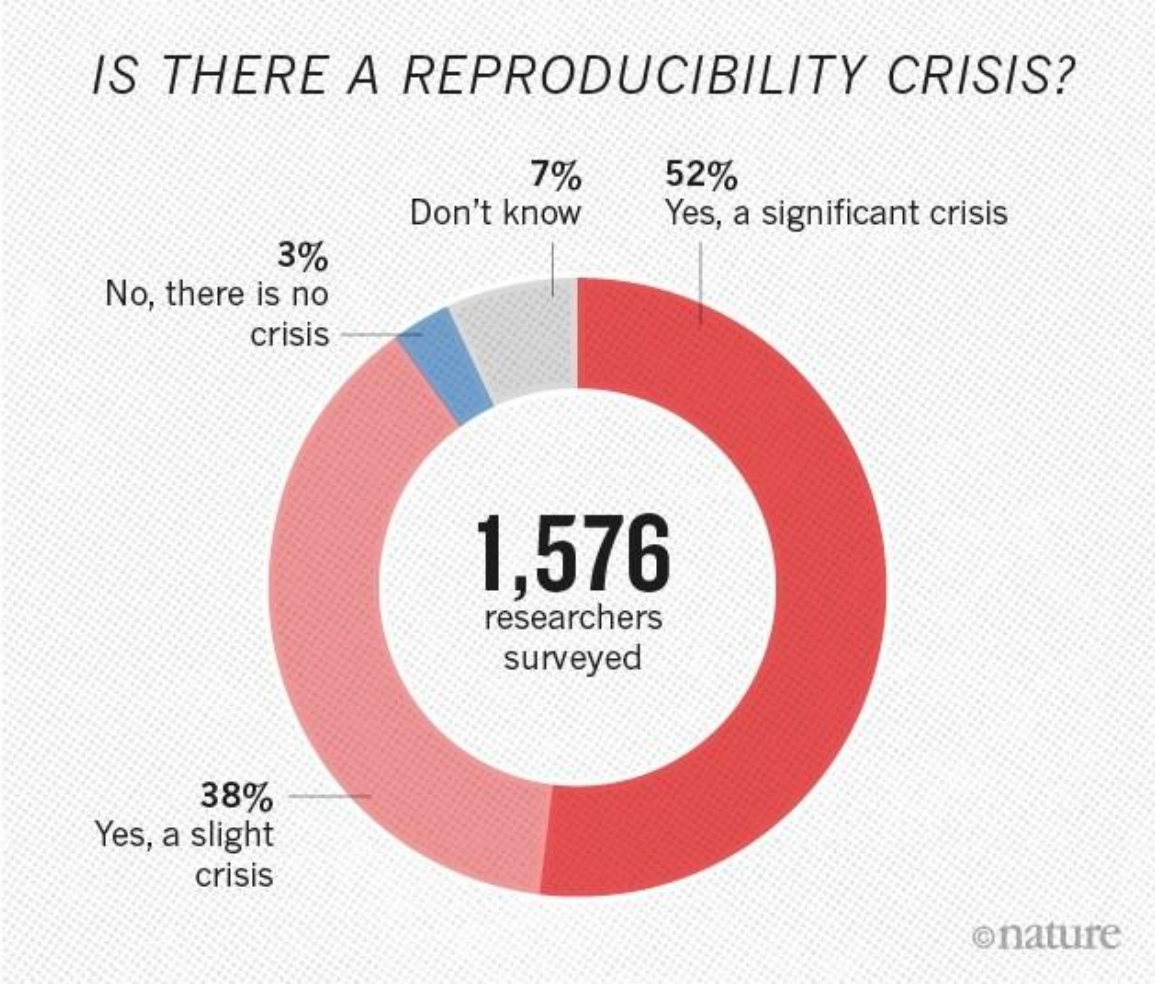
- ”umbrella term used to refer to the concepts of:
  - Openness
  - Transparency
  - Rigor
  - Reproducibility
  - Replicability
  - Accumulation of knowledge”

  (Crüwell et al, 2019, p.1)



Your experience with and worries about doing Open Science

Why do we need Open Science?



https://www.scottdmiller.com/the-replication-crisis-in-psychology-what-is-and-is-not-being-talked-about/

Baker (2016). Reproducibility crisis

Reproducibility or Replicability?

Reproducibility

Given the **same quantitative evidence** (data) and the **same statistical analysis**, can the **same result** be obtained?

Replicability

If an experiment is repeated with the **same procedures**, generating **new data**, will the **same result** be obtained?

Plesser (2018)

Types of replication

Direct replication

- Seeks to duplicate necessary elements that produced the original finding
- Are similar findings reproduced in subsequent attempts?

Conceptual replication

- Purposefully change at least one component of the original procedure (e.g., sample, measure)
- Are previous findings reproduced when tested under different conditions?
- “Alternative test” (Zwaan et al., 2018)

Zwaan et al. (2018)

Why do we have a crisis?

Publish or Perish

Giphy.com

The selection of bad science

“ Selection for high output leads to poorer methods and increasingly high false discovery rates” (Smaldino & McElreath, 2016)

Poor practices

- Publication bias
- Confirmation bias
- HARKing
- Under-reporting
- Selective reporting
- P-hacking
- Outright fraught

Publication bias

Failure to publish the results of a study on the basis of the direction or strength of the study findings

Credit: Deyan Georgiev/500Px Plus/Getty

Dickersen & Min (1993)

Confirmation bias

The widespread human tendency to preferentially search for and interpret observations as **consistent with one's beliefs** about how the world works

→Increase in **false discoveries**

Giphy.com

Selective reporting & under-reporting

Reporting only a **subset** of hypotheses tested

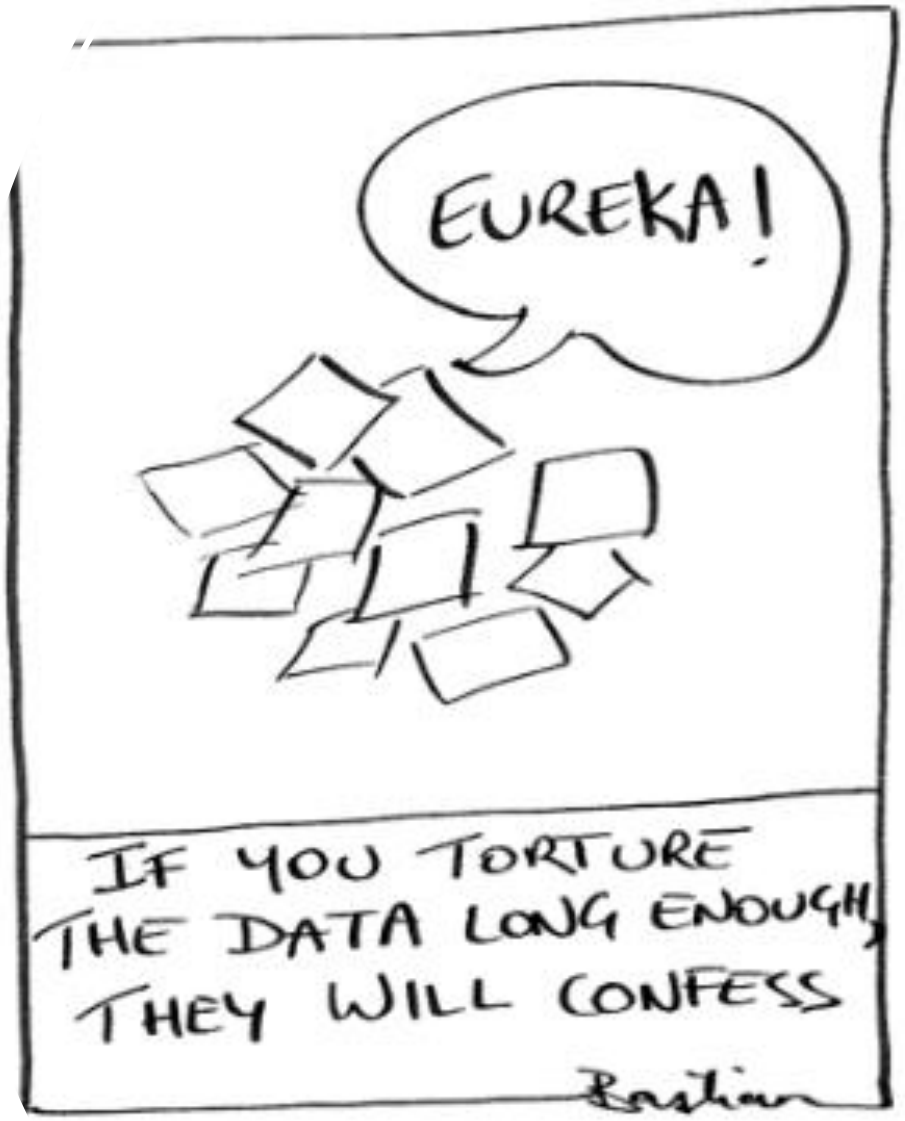
→Increase in **false positives**

Credit: Aaron Watson Photography

P-Hacking

A variety of practices that **increase the odds of finding a statistically significant results** by, for instance, conducting multiple versions of an analysis with different covariates, interations or subsets of the data

→Increase in **irreproducible results**



HARKing

= Hypothesizing after results are known

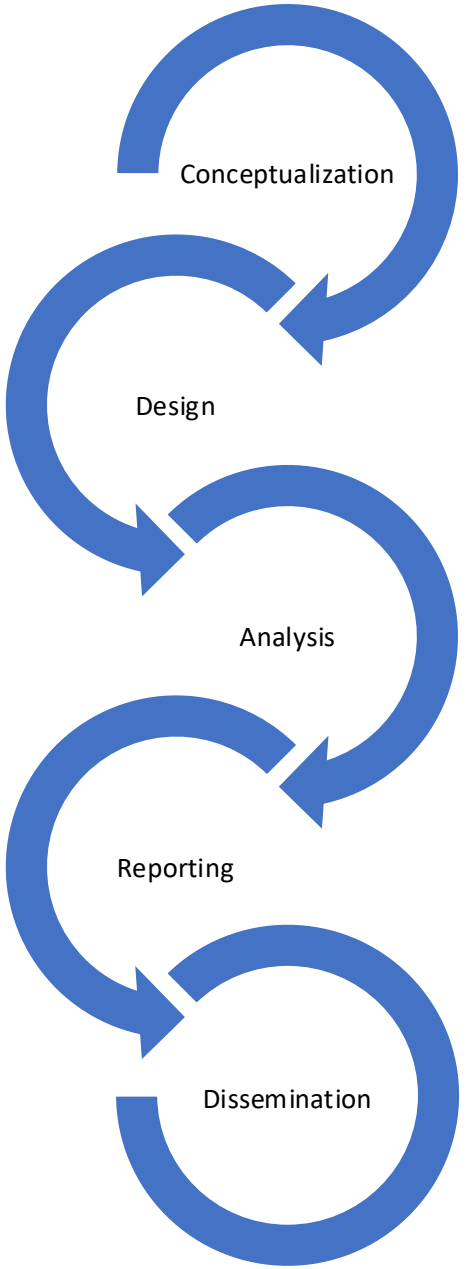
"I knew it all along!"

https://nesslabs.com/hindsight-bias

Behaviors to implement

See article by Kathawalla et al. (2021)

Open Science Practices



- Project Workflow
- Journal Clubs

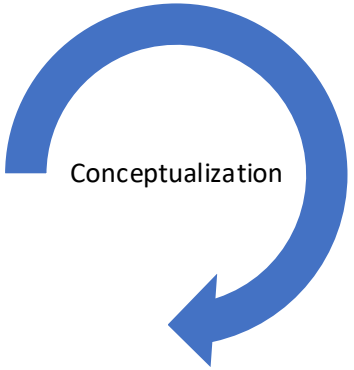
- Preregistration
- Registered Reports
- Data sharing planning

- Reproducible Code

- Transparent Writing

- Preprints
- Open Data, Materials & Code

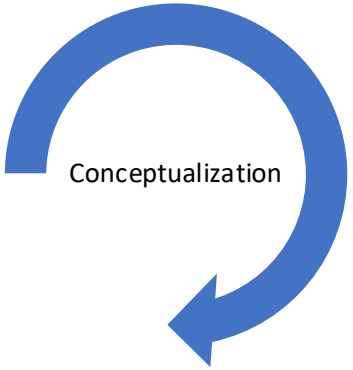
Conceptualization
Project Workflow



- Organizing projects and moving through the various stages of the research cycle
  - File folder structure
  - Document naming conventions
  - Version control
  - (Cloud) Storage
  - Who has access to what?

- One option: OSF project
  - Guidance: Soderberg et al. (2018): Using OSF to share data – a step by step guide
- Example: Turoman et al. (2021)

Conceptualization
Journal Clubs etc.

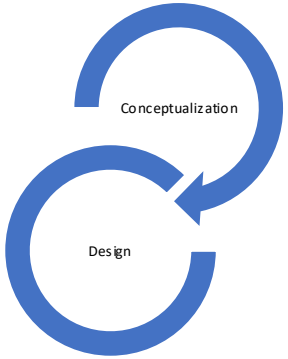


- Helps with:
  - Becoming familiar with common works and issues
  - Getting to know others interested in or knowledgeable about open science
  - There are also many online options!

  https://reproducibilitea.org

  → Reading lists and resources available

Design
Preregistration



Confirmatory vs. exploratory research

- Verifying a hypothesis vs. looking for a hypothesis
- Hypothesis-testing vs. hypothesis generation
- Both are valid and important BUT

1) One should not be presented as the other
2) The same data cannot be used to generate *and* test a hypothesis, which can happen unintentionally and reduce the credibility of your results

Taken from: https://osf.io/g45nt/

Preregistration

Principle: A timestamped, non-modifiable, online document presenting the research, its methods and goals

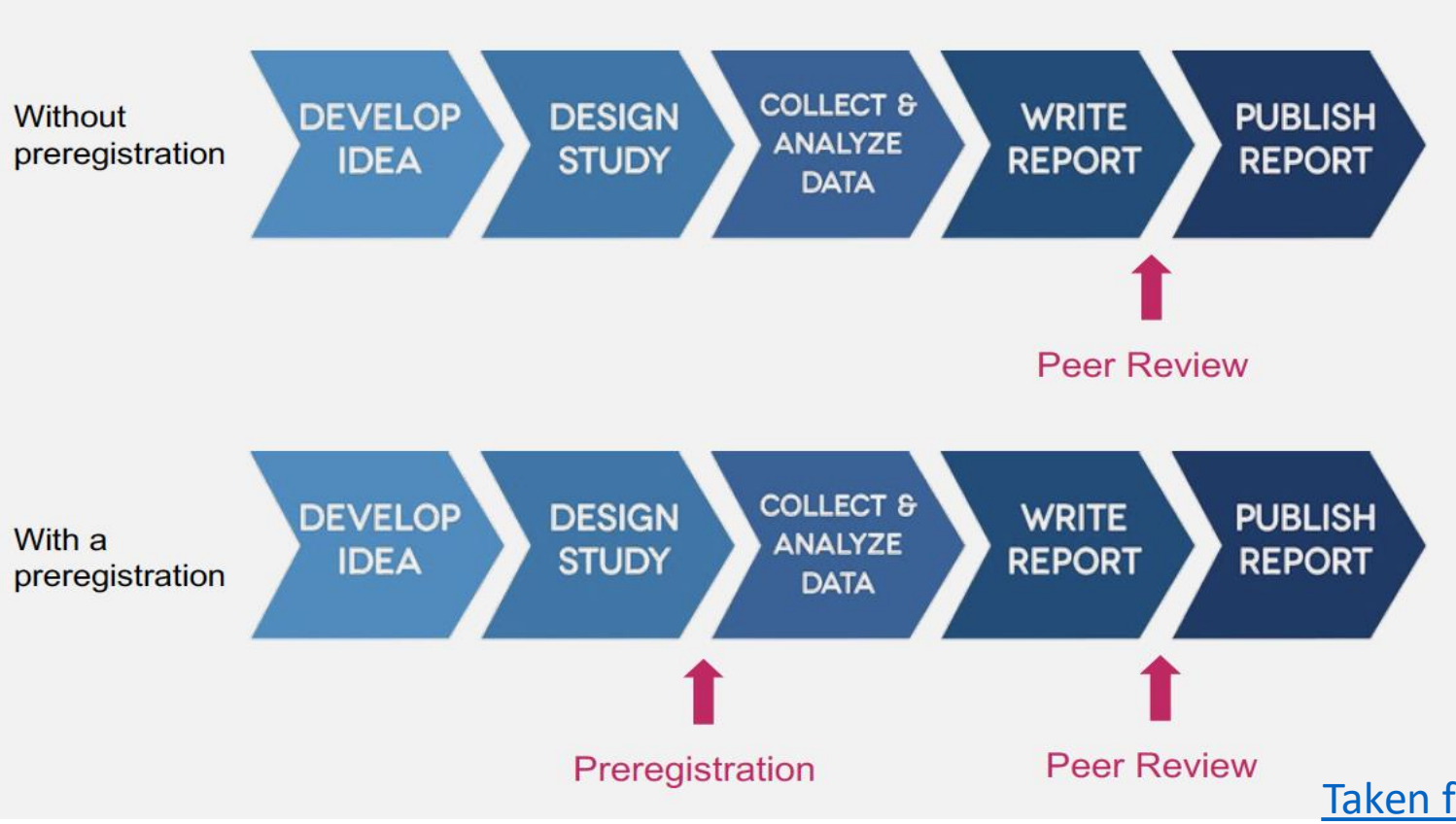
When: Before collecting/accessing/inspecting/analyzing research data

Content: Hypothesis, methods, participants (including sample size, inclusion or exclusion criteria), analysis, main outcome

Sharing: Generally on a dedicated server, allowing for a unique and permanent identifier

Taken from: https://osf.io/g45nt/

Preregistration in the **Research Workflow**



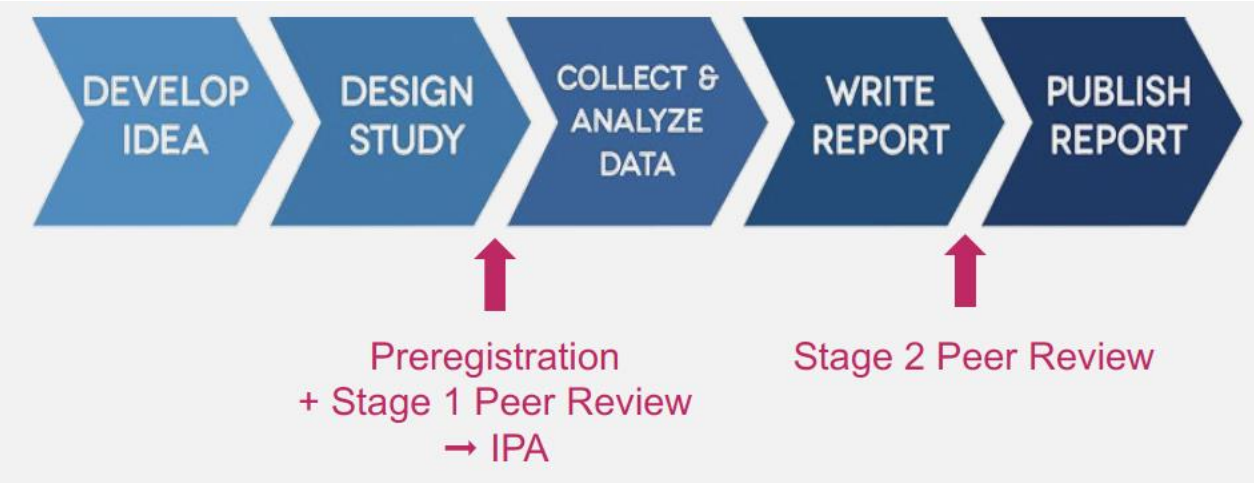
Taken from: https://osf.io/g45nt/

Tools & Implementation

Taken from: https://osf.io/g45nt/

Design
Registered Report

- Study proposal is reviewed before research is undertaken
- **In-principle-acceptance (IPA)**: acceptance before results are known, independent of results → prevents publication bias
- Explicit check of reported analyses vs. preregistration during peer review 2



Taken from: https://osf.io/g45nt/

Registered Report

- Scheel et al. (2021): compared 152 standard reports with 71 RRs
- Standard reports: 95% found support for their hypothesis
- RRs: only 45%

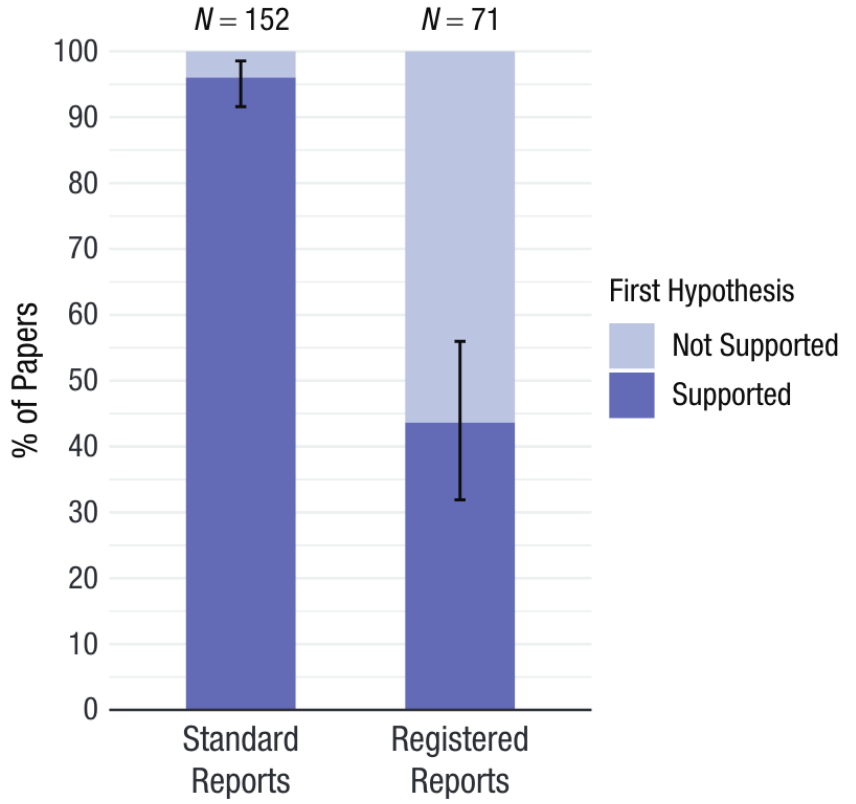


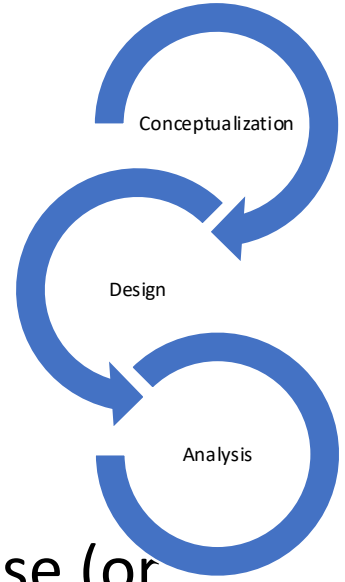
Fig. 2. Positive result rates for standard reports and Registered Reports. Error bars indicate 95% confidence intervals around the observed positive result rate.

Taken from: https://osf.io/g45nt/

Design
Plan Data Sharing

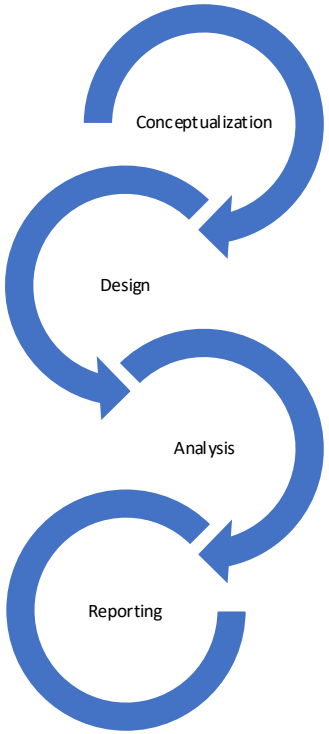
- Plan how to organize data
- Consent Forms
- Data Management Plan (FNS)
- Think about anonymization of data beforehand
- Meyer (2018): “Practical Tips for Data Sharing”

Analysis
Reproducible Code



- Detailed, written version of your code that would allow someone else (or your future self!) to generate the same output
- Allows for detection of errors, facilitates sharing and collaborations (for example with GitLab)
- How?:
  - Write the code yourself
  - Or also: Point-and click programs (e.g., SPSS, JAPS) also allow for script saving!
  - Recommended: open source software
  - Clear and extensive comments!
  - Coding buddy!

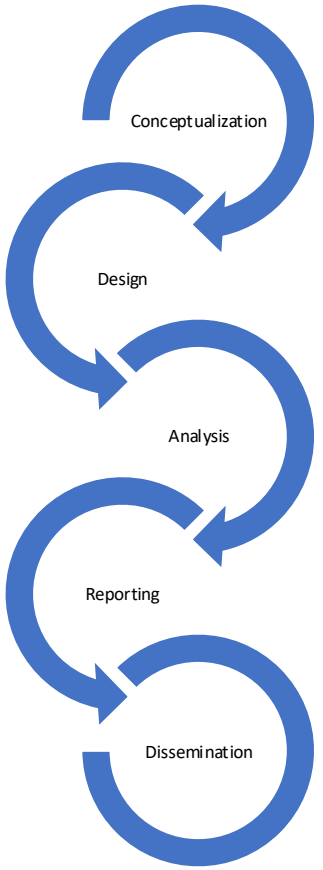
Reporting
Clear and transparent writing



- Follow manuscript writing guidelines
  - E.g. Gernsbacher (2018)
  - Suggest the level of detail for each section
  - Being transparent
  - Stating and justifying decisions (stopping rule, analyses etc.)
- Writing transparently facilitates replication attempts
- Allows for a better calibration of findings

Dissemination
Preprints

- Version of a manuscript publicly available prior to being submitted for peer-review
- But also: manuscripts currently under review, author-formatted version
- Little effort, many benefits:
  - Improve paper through feedback
  - Keep record of paper before the review-process
  - Put findings out sooner
  - Share work that does not get published
  - Clear temporal record (scooping not possible)
  - Possibly higher number of citations (Fraser et al., 2020)



Preprint Servers

Preprint Versions

- Sherpa Romeo tracks restrictions and rules for most journals, indicating whether they allow:
  - Posting of preprints pre-submission (usually yes)
  - Posting of author-formatted accepted articles (sometimes)
  - Posting of publisher-formatted accepted articles (usually no)

Dissemination
Open Data and Materials

Data and Code Sharing:

- Post data on data repository (e.g., OSF)
- Provide data codebook
- If sensitive data: possible to restrict access to full data set and sharing only metadata
- License your data:
  - E.g. CC-BY license: anyone using your data us obligated to attribute the data to you (by citing the paper, giving the DOI for the OSF project etc.)

FAIR Data Sharing

Findable

Accessible

Where is data uploaded? To be considered: DOIs, metadata, tracking of data-reuse, licensing, access control, long term availability

Interoperable

Reusable

How is data uploaded? To be considered: Data Format (proprietary (e.g., MS Word) vs. non-proprietary (e.g., txt file), how might formats change in the future?

Crüwell et al., 2019

Dissemination
Open Access

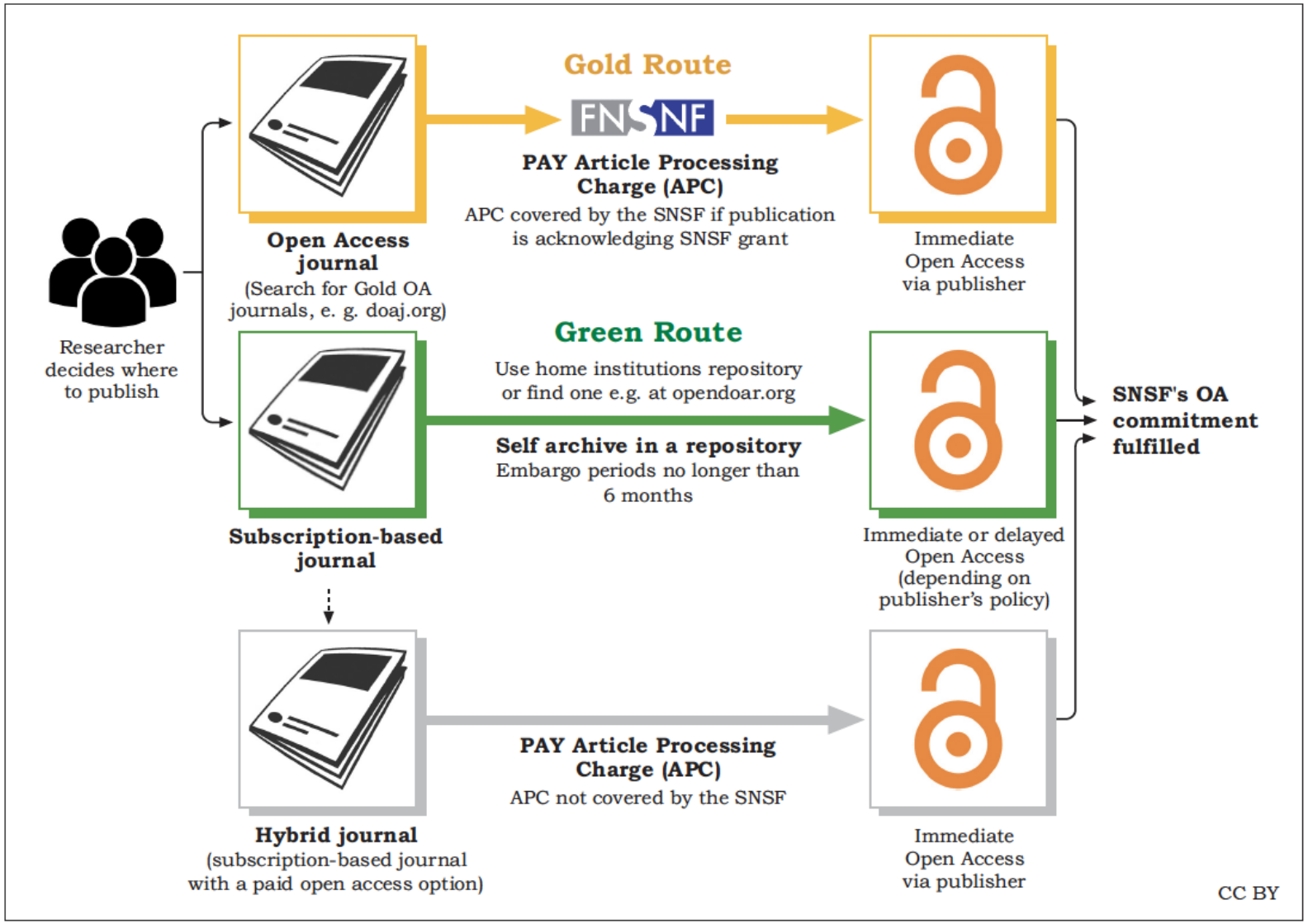
= free, public availability of a research product on the internet for distribution and re-use with acknowledgement (Budapest Open Access Initiative, Chan et al., 2002)

- Typically used in reference to published journal articles, but any output can be OA (materials, data, code,…)

Roads to Open Access

- Gold Route: OA in the formal publication system → works are made publicly available at the point of publication (by publishers themselves)
- Green Route: Authors self-archive work and make it publicly available
  - Includes peer-reviewed works and preprints pre-review
- Diamond Route: Published works without fees for the reader nor the authors (Platinum Open Access)

Harnad et al., 2008



SNSF Infographic (2018), https://issuu.com/snsf/docs/snsf_roads_to_oa_articles_951c6a8060e9c7

Effects of Open Access

1. **OA works are used more**
   - Availability to everyone
   - Inside academia: OA works are cited 36% - 600% more than non-OA works
   - Outside academia: more coverage by media, discussed more in non-scientific settings
2. **OA works facilitate meta-research**
   - increased findablity
3. **Societal Impact**
   - Availability to everyone → strengthening of work between global south and north

Downsides of Open Access

1. **Geographical wealth imbalance and effects of OA**
   - **High APCs** resulting from OA affecting mainly early career researchers, late career researchers looking to publish essays etc., self-funded researchers, low and middle income countries, whose research environments rarely provide funds for APCs (Powell et al., 2020)
   - Smith et al. (2021): Global Analysis of publishing in non-OA and OA journals: the poorer the region, the less likely OA

   → **Reinforcement of dichotomy between resource-wealthy and resource-poor researchers**, creating ”sub-standard journals” (Dudley et al., 2021)
2. **Predatory Journals**
   - Maximize profits by unselectively publishing a high proportion of submitted manuscripts
   - Come at a **massive cost for more naïve ECRs or resource-poor areas**
   - Drawback: papers do not get cited

What now?

- These are behaviors that you ***can*** implement
- Open Science is an evolving process
- You can learn and accumulate your practices over time
- Get in touch with others!

What now?

- These are behaviors that you ***can*** implement
- Open Science is an evolving process
- You can learn and accumulate your practices over time
- Get in touch with others!
- Drinks! ☺

References

Baker, M. (2016). Reproducibility crisis. *Nature, 533*(26), 353-66.

Chan, L., Cuplinskas, D., Eisen, M., Friend, F., Genova, Y., Guédon, J. C., ... & Velterop, J. (2002*). Budapest open access initiative. ARL Bimonthly*, 48

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., ... & Schulte-Mecklenbeck, M. (2019). *Seven easy steps to open science.* Zeitschrift für Psychologie.

Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints, citations and altmetrics. *Quantitative Science Studies, 1*(2), 618-638.

Gernsbacher, M. A. (2018). Writing empirical articles: Transparency, reproducibility, clarity, and memorability. *Advances in methods and practices in psychological science, 1*(3), 403-414.

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., ... & Hilf, E. R. (2008). The access/impact problem and the green and gold roads to open access: An update. *Serials review, 34(1*), 36-40

Kathawalla, U. K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. *Collabra: Psychology, 7*(1), 18684.

Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in methods and practices in psychological science, 1*(1), 131-144

Plesser, H. E. (2018). Reproducibility vs. replicability: a brief history of a confused terminology. *Frontiers in neuroinformatics, 11*, 76.

Scheel, A. M., Schijen, M. R., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. Advances in Methods and *Practices in Psychological Science, 4*(2), 25152459211007467.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society open science, 3*(9), 160384.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41,* e120.