STRATOS – Open Science Panel

Panel chair: Sabine Hoffmann
Presenter: Kim Luijken

The Statistical Crisis in Science

Data-dependent analysis—a "garden of forking paths"— explains why many statistically significant comparisons don't hold up.

Andrew Gelman and Eric Loken

There is a growing realization that reported "statistically significant" claims in scientific publications are routinely mistaken. Researchers typically express the confidence in their data in terms of *p*-value: the probability that a perceived result is actually the result of random variation. The value of *p* (for "probability") is a way of measuring the extent to which a data set provides evidence against a so-called null hypothesis. By convention, a *p*-value below 0.05 is considered a meaningful refutation of the null hypothesis; however, such conclusions are less solid than they appear.

a short mathematics test when it is expressed in two different contexts, involving either healthcare or the military. The question may be framed nonspecifically as an investigation of possible associations between party affiliation and mathematical reasoning across contexts. The null hypothesis is that the political context is irrelevant to the task, and the alternative hypothesis is that context matters and the difference in performance between the two parties would be different in the military and healthcare contexts.

At this point a huge number of possible comparisons could be performed, all consistent with the researcher's the-

This *multiple comparisons* issue is well known in statistics and has been called "*p*-hacking" in an influential 2011 paper by the psychology researchers Joseph Simmons, Leif Nelson, and Uri Simonsohn. Our main point in the present article is that it is possible to have multiple potential comparisons (that is, a data analysis whose details are highly contingent on data, invalidating published *p*-values) without the researcher performing any conscious procedure of fishing through the data or explicitly examining multiple comparisons.

How to Test a Hypothesis


The Statistical Crisis in Science
HARKing
Limited access
Selective reporting
Researcher degrees
of freedom

Open Science

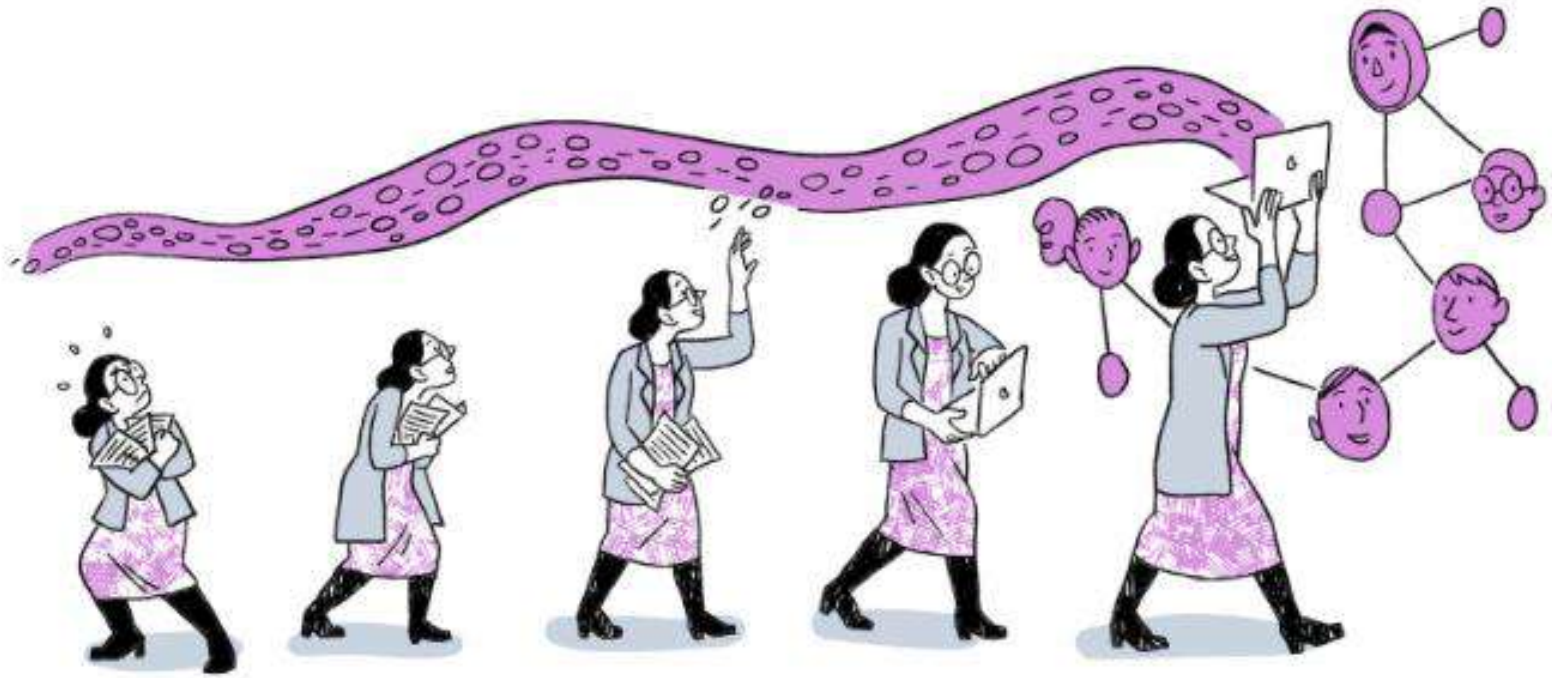
Parsons et al., 2022: A community-sourced glossary of open scholarship terms

Open science is “an umbrella term that reflects the idea that scientific knowledge of all kinds, where appropriate, should be openly accessible, transparent, rigorous, reproducible, replicable, accumulative and inclusive”

Open Science

The Turing Way: Open Research

“Open research aims to transform research by making it more reproducible, transparent, reusable, collaborative, accountable, and accessible to society”



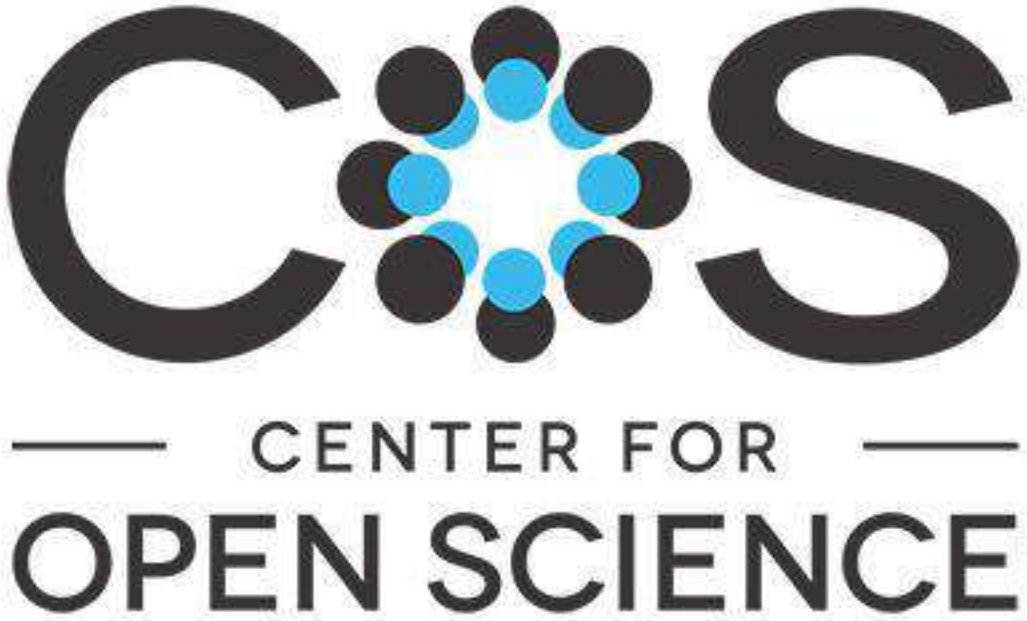
EVOLVING TOWARDS AN ERA OF OPEN RESEARCH

Scriberia

Center for Open Science

Founded in 2013 by
Brian Nosek & Jeffrey Spies

Launched Open Science Framework



STRATOS Open Science Panel

Components of Open Science highlighted by STRATOS

- Open Access manuscripts
- Reproducibility
- Transparency
- Replicability

STRATOS Components of Open Science

Open Access manuscripts

Reproducibility

- Open code
- Open data / data availability statement

Transparency

- Open code
- Open data / data availability statement

Replicability

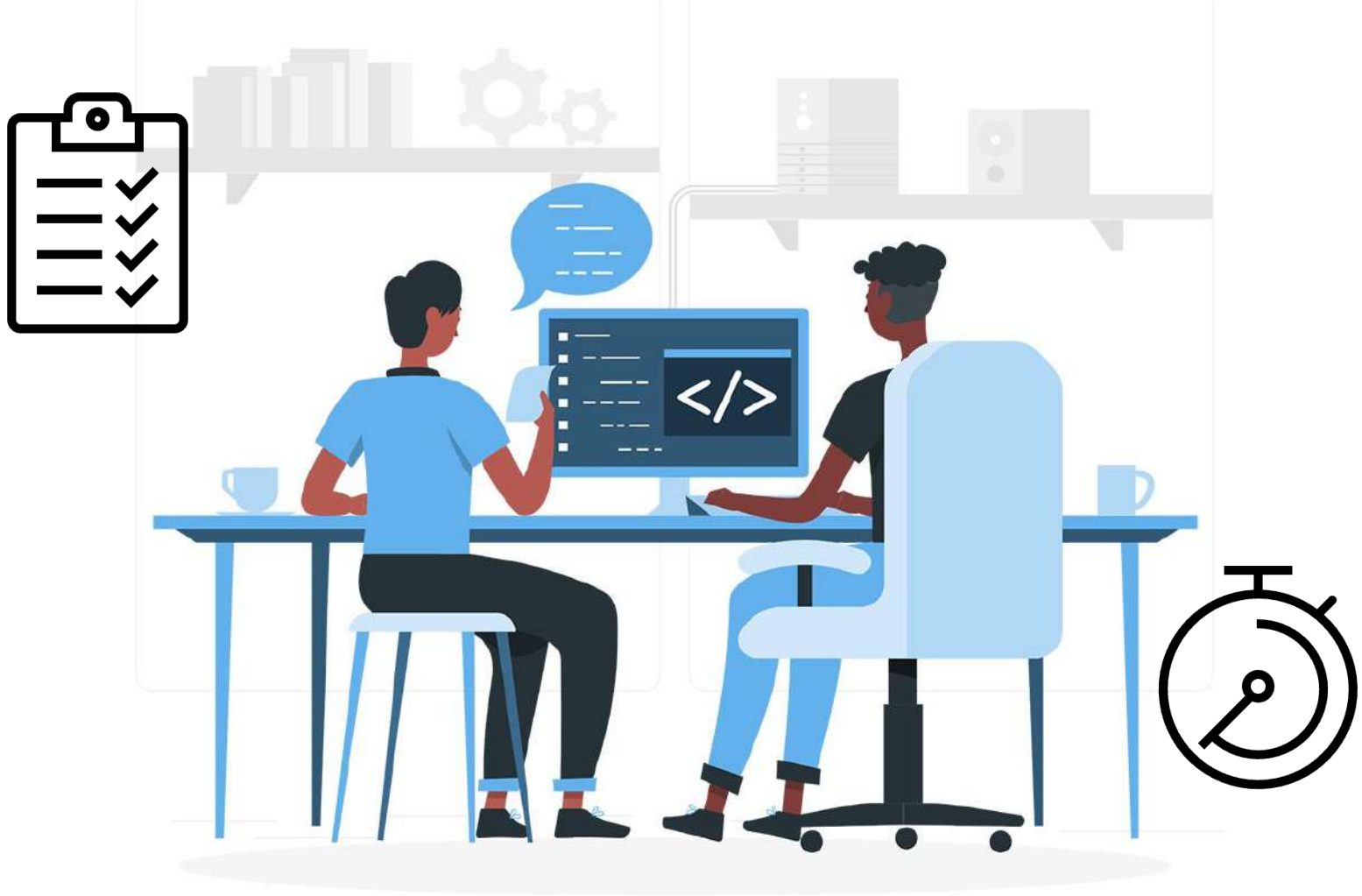
- Dealing with “researcher degrees of freedom” (Wicherts et al., 2016)
- Awareness of result-dependent selective reporting

STRATOS OS Panel

Chair	Sabine Hoffmann
Members	Anne-Laure Boulesteix, Daniela Dunkler, Roman Hornung, Michael Kammer, Kim Luijken, Willi Sauerbrei, Fabian Scheipl, Ewout Steyerberg

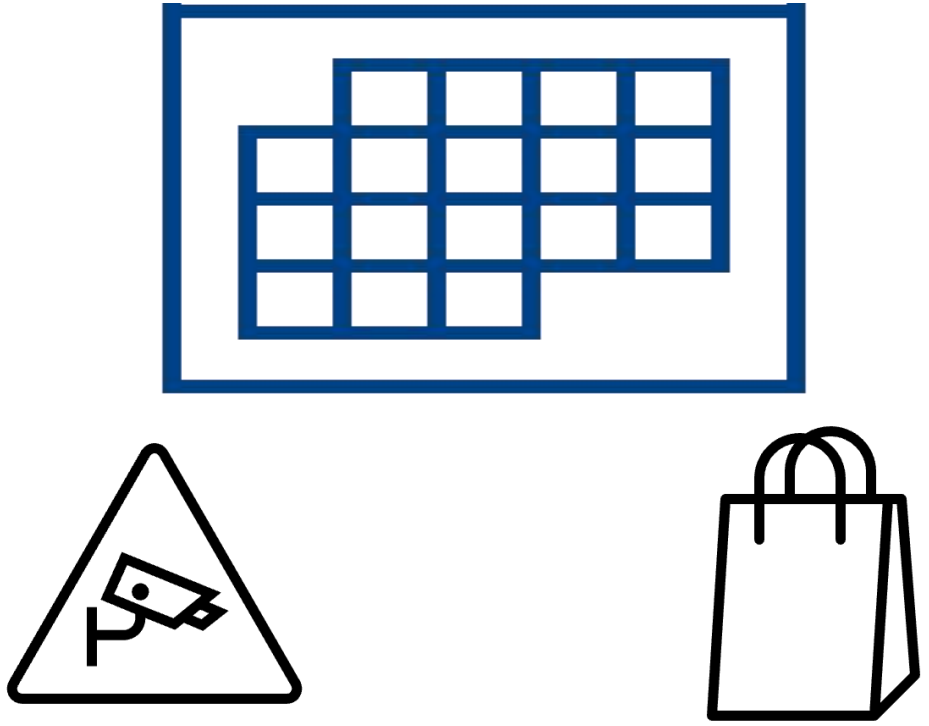
Barriers to Open Science

Barriers to Open Science



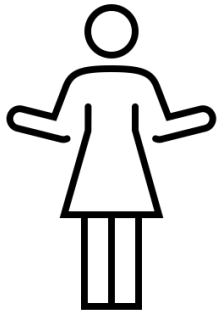
https://distantjob.com/blog/2017-08-16-pair-programming-why-you-should-care-about-it-and-how-to-do-it-remotely/

Barriers to Open Science

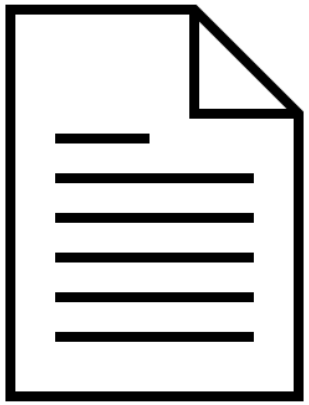


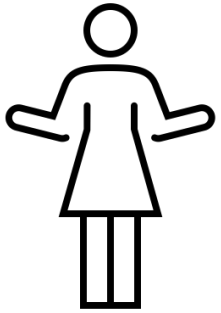
Barriers to Open Science

Research question

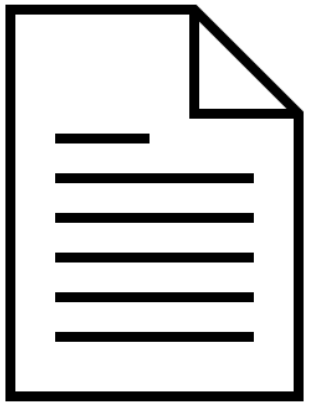


Methods

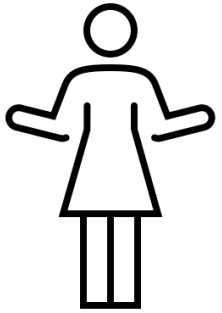




Results



Discussion



Barriers to Open Science

Open Access manuscripts

- Open access **fees**

Reproducibility

- **Burdensome** to create well-structured and well-documented code
- Reproducing results using someone else's code can be **time consuming**

Transparency

- **Privacy** issues for medical data sets
- Question of **ownership of shared data sets**

Replicability

- **Research incentives** favor significant and innovative results

	STRATOS members	Research community
Open Access	STRATOS publications ideally Open Access	Underlining importance of Open Access publications
Reproducibility	STRATOS publications should be reproducible STRATOS publications ideally uses Open Access data sets	Guidance on reproducibility for level-1 audience Guidance on sharing data while preserving privacy protection and validity of statistical inference
Transparency	STRATOS publications ideally specify a study protocol and ask community for feedback	Illustrating benefits of open simulation setup
Replicability		Guidance on dealing with uncertain choices for level-1 audience

STRengthening Analytical Thinking for Observational Studies (STRATOS): Introducing the Open Science Panel

Sabine Hoffmann[1], Kim Luijken[2], Willi Sauerbrei[3], Pamela Shaw[4], Anne-Laure Boulesteix[5]

[1] Department of Statistics, LMU Munich, Germany

[2] Department of Epidemiology, Julius Center for Health Sciences and Primary Care, Utrecht University Medical Center, University Utrecht, The Netherlands

[3] Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

[4] Division of Biostatistics, Kaiser Permanente Washington Health Research Institute, USA

[5] Institute for Medical Information Processing, Biometrics and Epidemiology (IBE), Faculty of Medicine, LMU Munich, Germany

Concerning open science practices within the STRATOS initiative, there is a general consensus that STRATOS publications should be open access. Moreover, they should use open access data sets (or a synthetic resemblance of it) to make results more easily reproducible. The Open Science and the Publication panels will work together to promote Open Science best practices for STRATOS publications, including development of an Open Science review process for STRATOS publications to undergo prior to submission to a journal.

Concerning guidance on adopting open science practices in observational studies, we plan to develop a paper for biomedical researchers which will outline approaches to dealing with uncertain choices in the analysis of observational studies (also referred to "researcher degrees of freedom" [Simmons et al., 2011]). Although there is increasing awareness of the dangers related to questionable research practices including "HARKing" [Kerr, 1998] and "p hacking", many researchers are unaware of the consequences of seemingly innocuous decisions concerning data pre-processing and model choice that may often occur after considering several possible results on the analyzed data sets. It is important to increase awareness of the problems caused by result-dependent selective reporting and to give an overview of solutions that exist to deal with researcher degrees of freedom without invalidating statistical inference [Hoffmann et al., 2021]. We additionally plan to provide researchers with practical advice to improve the transparency about decisions during data processing and analysis made in their work through a tutorial paper and videos on the STRATOS website about how to make their analysis code readable and reproducible.

Furthermore, data sharing is another pressing topic in biomedical research for which guidance is urgently needed [Mansmann et al., 2023]. If data is FAIR (Findable, Accessible, Interoperable, Reusable), science becomes more efficient, collaborative and transparent. Although data sharing is increasingly encouraged by journals and funding agencies, biomedical researchers remain hesitant to share

Guidance on reproducibility (level-1)

Invitation to improve code writing and sharing practices in biostatistical research

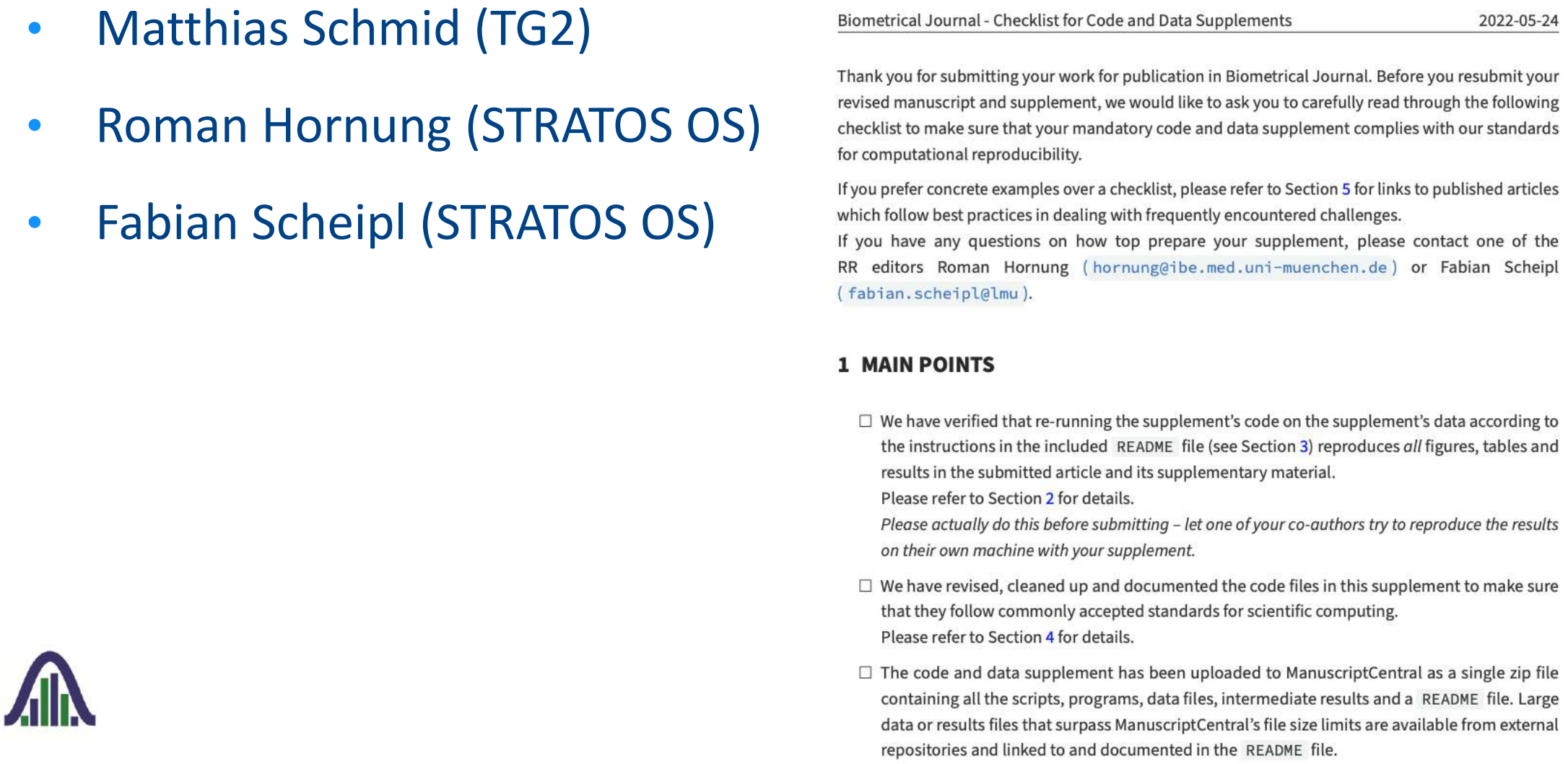
- How to write code that facilitates easy reproduction of the analysis
- How to review code
- How to make code publicly available

PS: level-1 definition is slightly different from general meaning in STRATOS

Related existing initiatives – Open Science Badges

Related existing initiatives – Badge “Reproducible Research” in Biometrical Journal

- Matthias Schmid (TG2)
- Roman Hornung (STRATOS OS)
- Fabian Scheipl (STRATOS OS)



Biometrical Journal - Checklist for Code and Data Supplements 2022-05-24

Thank you for submitting your work for publication in Biometrical Journal. Before you resubmit your revised manuscript and supplement, we would like to ask you to carefully read through the following checklist to make sure that your mandatory code and data supplement complies with our standards for computational reproducibility.

If you prefer concrete examples over a checklist, please refer to Section 5 for links to published articles which follow best practices in dealing with frequently encountered challenges.
If you have any questions on how top prepare your supplement, please contact one of the RR editors Roman Hornung (hornung@ibe.med.uni-muenchen.de) or Fabian Scheipl (fabian.scheipl@lmu).

1 MAIN POINTS

☐ We have verified that re-running the supplement's code on the supplement's data according to the instructions in the included `README` file (see Section 3) reproduces *all* figures, tables and results in the submitted article and its supplementary material.
Please refer to Section 2 for details.
Please actually do this before submitting – let one of your co-authors try to reproduce the results on their own machine with your supplement.

☐ We have revised, cleaned up and documented the code files in this supplement to make sure that they follow commonly accepted standards for scientific computing.
Please refer to Section 4 for details.

☐ The code and data supplement has been uploaded to ManuscriptCentral as a single zip file containing all the scripts, programs, data files, intermediate results and a `README` file. Large data or results files that surpass ManuscriptCentral's file size limits are available from external repositories and linked to and documented in the `README` file.

Related existing initiatives – Replication of simulation studies

- 8 Highly cited simulation studies
- Teams of replicators

Direct *process* replicability:

- Simulation code based on information from the original publications

Replicability varied greatly:

- 2 Studies perfectly replicable
- 5 Studies partially replicable
- 1 Study not replicable at all

Illustrating benefits of open simulation setup

Received: 12 August 2022 | Revised: 9 December 2022 | Accepted: 22 January 2023

DOI: 10.1002/bimj.202200222

Biometrical Journal

RESEARCH ARTICLE

Phases of methodological research in biostatistics—Building the evidence base for new methods

Georg Heinze[1] | **Anne-Laure Boulesteix**[2] | **Michael Kammer**[1,3] | **Tim P. Morris**[4] | **Ian R. White**[4] | **on behalf of the Simulation Panel of the STRATOS initiative**

[1]Center for Medical Data Science, Institute of Clinical Biometrics, Medical University of Vienna, Vienna, Austria

[2]Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians University of Munich, Munich, Germany

[3]Department of Medicine III, Division of Nephrology, Medical University of Vienna, Vienna, Austria

[4]MRC Clinical Trials Unit, Institute of Clinical Trials & Methodology, University College London, London, UK

ABSTRACT

Although new biostatistical methods are published at a very high rate, many of these developments are not trustworthy enough to be adopted by the scientific community. We propose a framework to think about how a piece of methodological work contributes to the evidence base for a method. Similar to the well-known phases of clinical research in drug development, we propose to define four phases of methodological research. These four phases cover (I) proposing a new methodological idea while providing, for example, logical reasoning or proofs, (II) providing empirical evidence, first in a narrow target setting, then (III) in an extended range of settings and for various outcomes, accompanied by appro-

Guidance on (level-1)

Many researche ningly
innocuous decisi odel choice
that may often c lts on
the analyzed dat
- "Researcher
- Questionable

Overview of:
- Problems cau
- Solutions tha g
  invalidating s eedom without

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Hoffmann S, Schönbrodt F, Elsas R, Wilson R, Strasser U, Boulesteix A-L. 2021 The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. *R. Soc. Open Sci.* **8**: 201925. https://doi.org/10.1098/rsos.201925

Received: 2 November 2020
Accepted: 22 March 2021

Subject Category:
Mathematics

Subject Areas:
statistics

Keywords:
replicability crisis, uncertainty, open science, interdisciplinary perspective, metaresearch

The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines

Sabine Hoffmann[1,2], Felix Schönbrodt[1,3], Ralf Elsas[1,4], Rory Wilson[6], Ulrich Strasser[7] and Anne-Laure Boulesteix[1,2,5]

[1]LMU Open Science Center, [2]Institute for Medical Information Processing, Biometry, and Epidemiology, Medical School, [3]Department of Psychology, Psychological Methods and Assessment, [4]Institute for Finance and Banking, Munich School of Management, and [5]Department of Statistics, Faculty of Mathematics, Computer Science and Statistics, Ludwig-Maximilians-Universität München, Munich, Germany
[6]Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg, Germany
[7]Department of Geography, University of Innsbruck, Innsbruck, Austria

SH, 0000-0001-6197-8801; FS, 0000-0002-8282-3910; RE, 0000-0001-8117-1175; US, 0000-0003-4776-2822; ALB, 0000-0002-2729-0947

For a given research question, there are usually a large variety of possible analysis strategies acceptable according to the scientific standards of the field, and there are concerns that this multiplicity of analysis strategies plays an important role in the non-replicability of research findings. Here, we define a general framework on common sources of uncertainty arising in computational analyses that lead to this multiplicity, and apply this framework within an overview of approaches proposed across disciplines to address the issue. Armed with this framework, and a set of recommendations derived therefrom,

What is your opinion?

Please join the conversation

Go to www.wooclap.com

Use event code:

LKGULF

Please note that wooclap is one of the many available tools for interaction with the audience, but it is not a STRATOS recommendation

STRATOS OS Panel

Chair	Sabine Hoffmann
Members	Anne-Laure Boulesteix, Daniela Dunkler, Roman Hornung, Michael Kammer, Kim Luijken, Willi Sauerbrei, Fabian Scheipl, Ewout Steyerberg

References

Allen C and Mehler DMA. Open science challenges, benefits and tips in early career and beyond. PLoS Biology, 17(5):e3000246, 2019.

Gelman A and Loken E. The statistical crisis in science data-dependent analysis—a "garden of forking paths"—explains why many statistically significant comparisons don't hold up. American Scientist, 102(6):460–465, 2014.

Hoffmann S, Schönbrodt F, Elsas R, Wilson R, Strasser U, Boulesteix AL. The multiplicity of analysis strategies jeopardizes replicability: lessons learned across disciplines. Royal Society Open Science, 8(4):201925, 2021. DOI: 10.1098/rsos.201925.

Kerr NL. HARKing: Hypothesizing after the results are known. Personality and Social Psychology Review, 2(3):196–217, 1998. DOI: 10.1207/s15327957pspr0203-4.

Mansmann U, Locher C, Prasser F, Weissgerber T, Sax U, Posch M, Decullier E, Cristea IA, Debray TPA, Held L, et al. Implementing clinical trial data sharing requires training a new generation of biomedical researchers. Nature Medicine, Volume 29, 2023.

Parsons S, Flavio Azevedo F, Elsherif MM, Guay S, Shahim ON, Govaart GH, Norris E, Aoife O'Mahony A, Parker AJ, Todorovic A, et al. A community-sourced glossary of open scholarship terms. Nature Human Behaviour, 6(3):312–318, 2022.

Seibold H, Charlton A, Boulesteix AL, and Hoffmann S. Statisticians, roll up your sleeves! there's a crisis to be solved. Significance, 18(4):42–44, 2021.

Simmons JP, Nelson LD, and Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science, 22(11):1359–1366, 2011.

Wallach JD, Boyack KW, and Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. PLOS Biology, 16(11):e2006930, 2018.