Journal Club

ANNUAL REVIEWS

JOURNALS A-Z | JOURNAL INFO | PRICING & SUBSCRIPTIONS

Home / Annual Review of Public Health / Volume 42, 2021 / Peng, pp 79-93

Reproducible Research: A Retrospective

Annual Review of Public Health
Vol. 42:79-93 (Volume publication date April 2021)
First published as a Review in Advance on January 19, 2021
https://doi.org/10.1146/annurev-publhealth-012420-105110

Roger D. Peng and Stephanie C. Hicks
Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA; email: rdpeng@jhu.edu, shicks19@jhu.edu

https://github.com/richelbilderbeek/journal_club_20220602

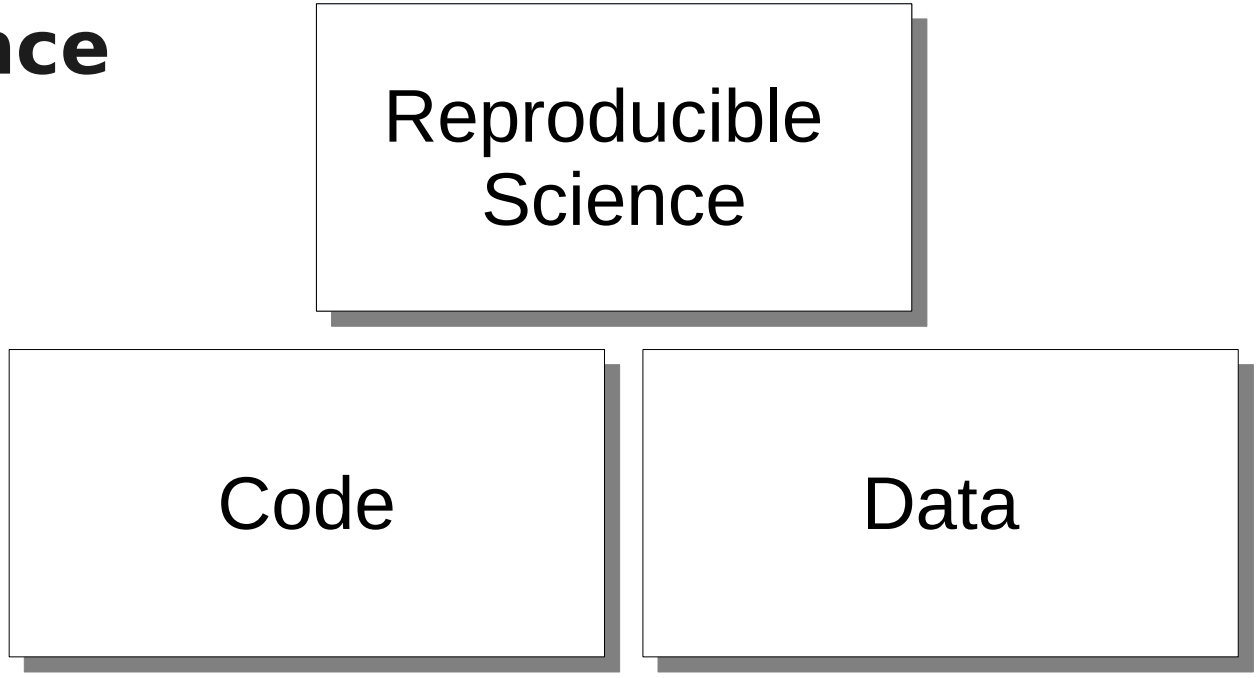
Science?

I have discovered a marvelous
proof to this theorem,
that this margin is
too narrow to contain

Why this article

I care about reproducible science

Article is recent

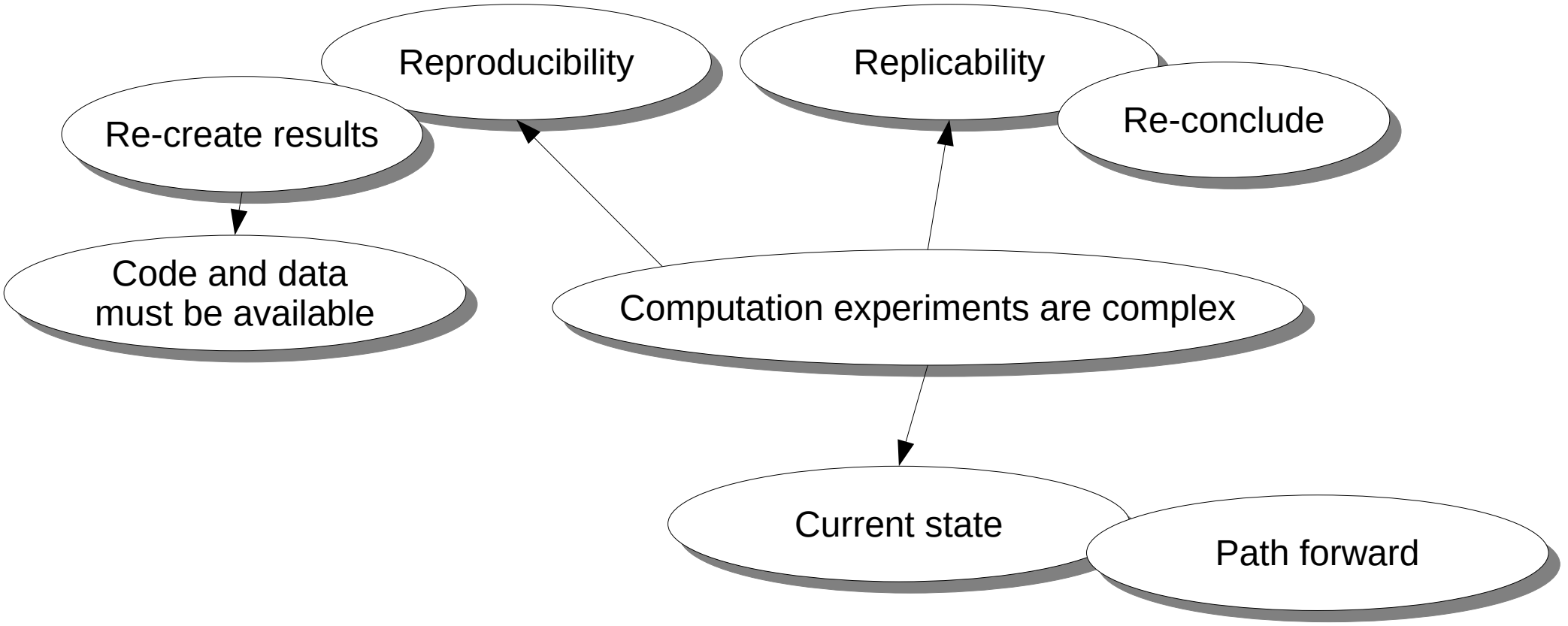


The authors

Roger D Peng

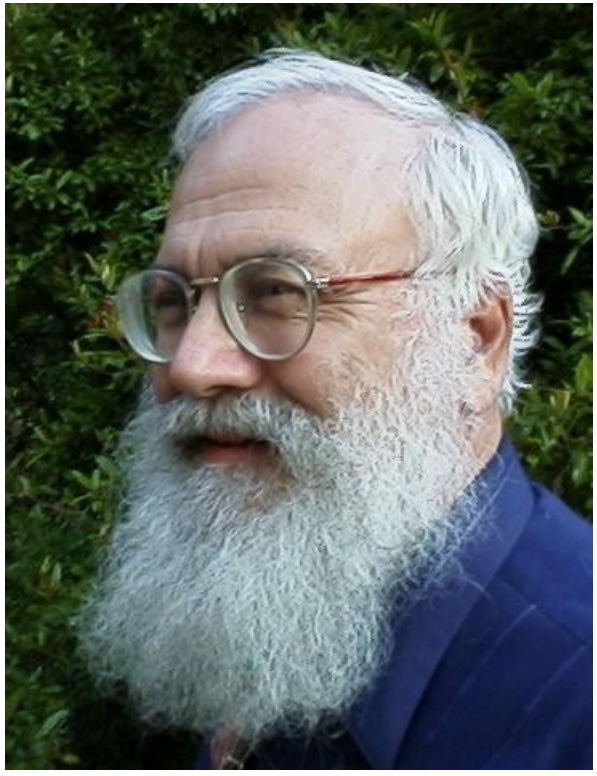
Stephanie Hicks

One-slide overview



Introduction 1/3

[...] it often seems that the greatest beneficiary of preparing the work in a reproducible form is the original author!

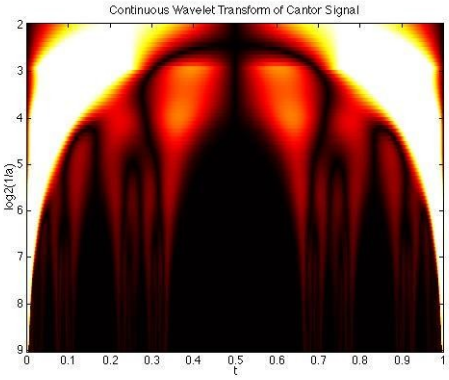


Claerbout J, Schwab M. 2001. CD-rom versus the Web. Rep. 84, Stanford Explor. Proj., Stanford, CA.

Introduction 2/3

An article [...] is merely
advertising of [..] scholarship.

The ***actual*** scholarship
is the complete software
development environment
and the complete set of instructions
which generated the figures.



Buckheit JB, Donoho DL. 1995. Wavelab and reproducible research. In Wavelets and Statistics, ed. A Antoniadis, pp. 55–81. New York: Springer-Verlag

Introduction 3/3

Our reviewers were not able
to conduct an independent
and private peer review
and therefore notified us
of their withdrawal
from the peer-review process.

THE LANCET

TRUMP IS TAKING HYDROXYCHLOROQUINE: WHY EXPERTS SAY YOU SHOULDN'T

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7274621/

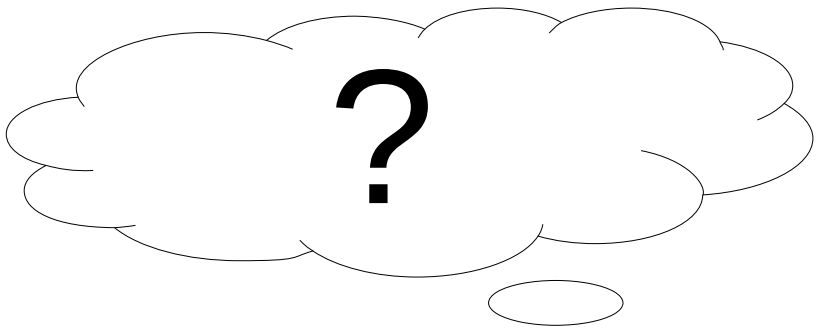
Definition of reproducible research

A published data analysis is reproducible
if the analytic **data** sets
and the computer **code** used to create the data analysis
are made available to others for independent study and analysis

Subset of data set

Analysis and visualization

Tuning parameters



Peng RD. 2011. Reproducible research in computational science. Science 334(6060):1226–27

Definition of reproducible research

A published data analysis is reproducible
if the analytic **data** sets
and the computer **code** used to create the data analysis
are made available to others for independent study and analysis

Subset of data set

Analysis and visualization

Tuning parameters

Sensitive data set?

How to make available?

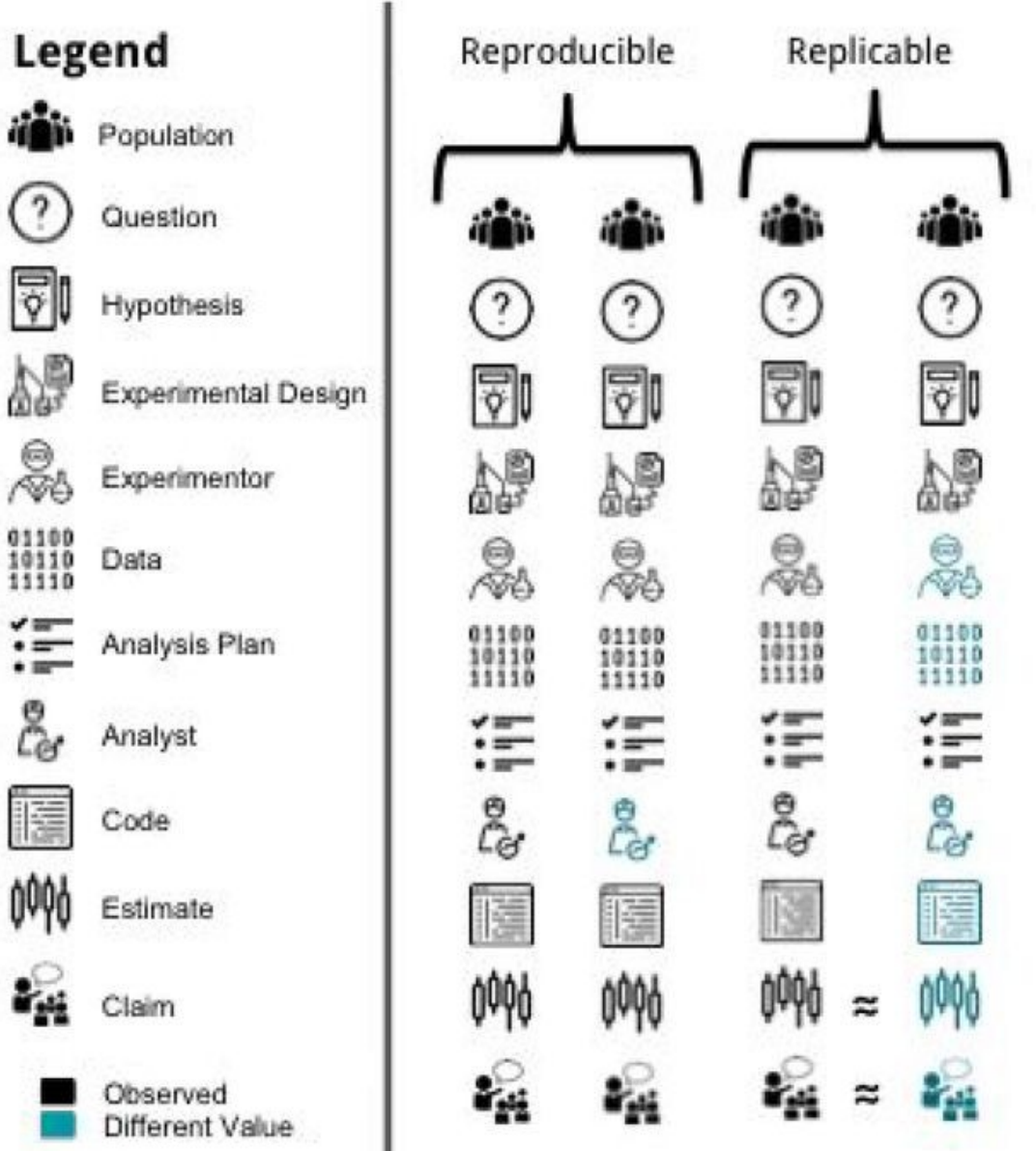
Testing data set?

Requirements for code?

Peng RD. 2011. Reproducible research in computational science. Science 334(6060):1226–27

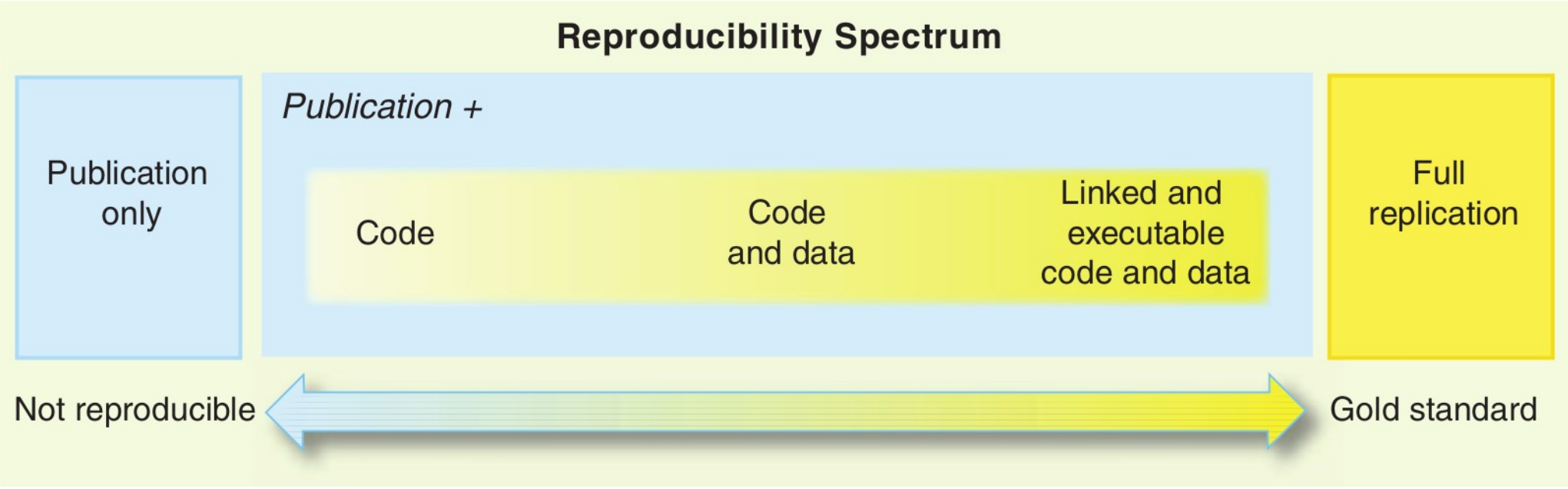
Reproduction:
re-do experiment

Replication:
re-conclude



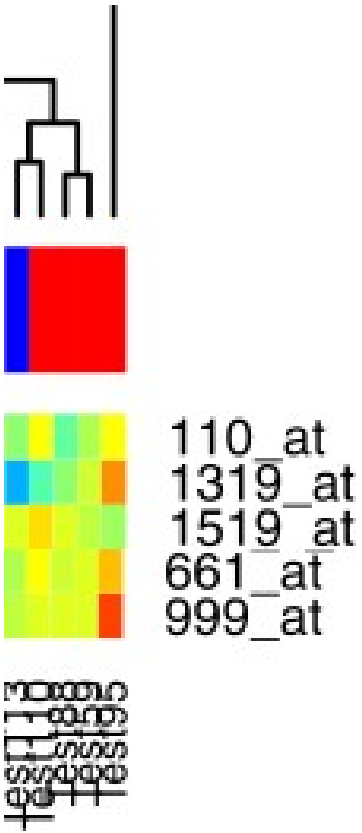
Patil, Prasad, **Roger D. Peng**, and Jeffrey T. Leek. "A visual tool for defining reproducibility and replicability." Nature human behaviour 3.7 (2019): 650-652.

Looking back



Peng, Roger D. "Reproducible research in computational science." Science 334.6060 (2011): 1226-1227.

Forensic bioinformatics



Relatedness of expression profile

Resistant or sensitive

Expression level

What is being expressed

Cell line

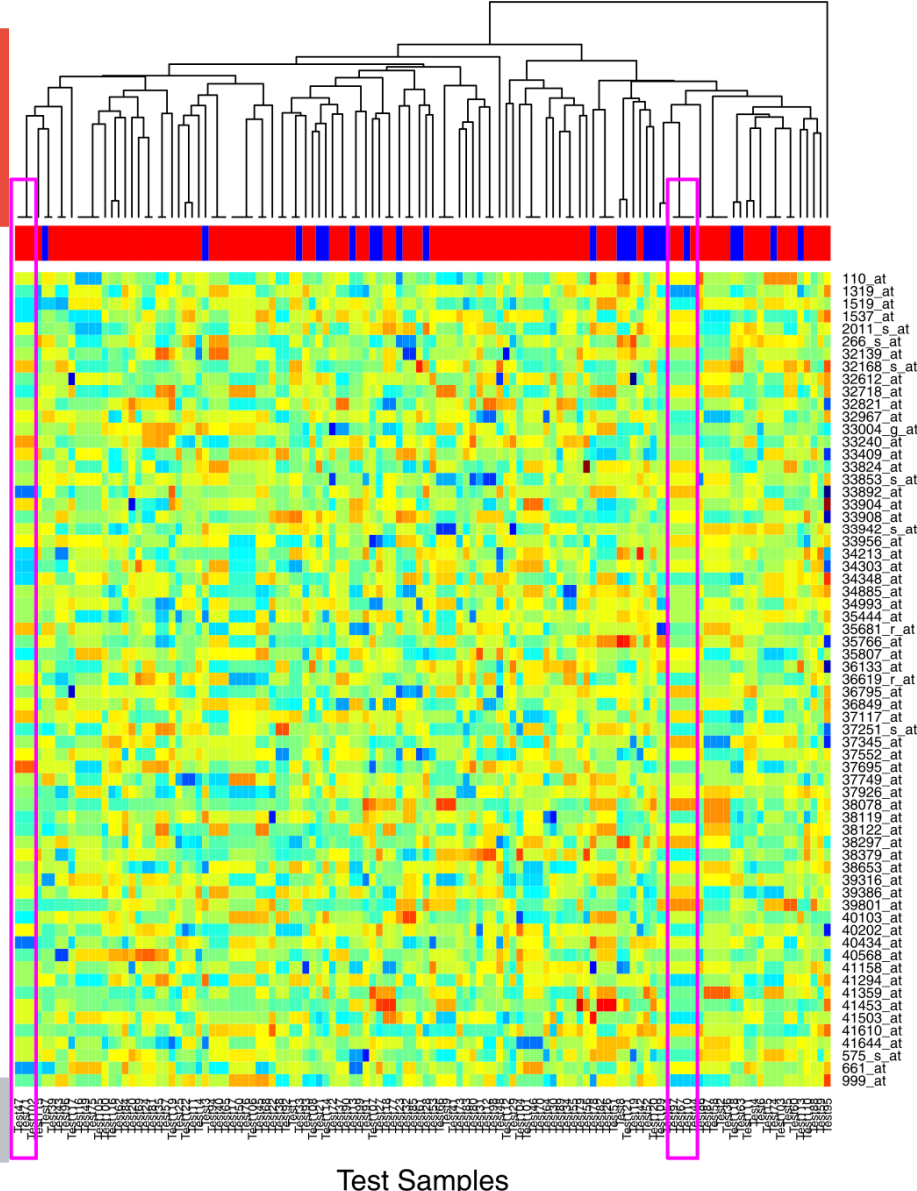
Baggerly KA, Coombes KR. 2009. Deriving chemosensitivity from cell lines: forensic bioinformatics and reproducible research in high-throughput biology. Ann. Appl. Stat. 3(4):1309–34

Forensic bioinformatics

An off-by-one error

Most common errors are simple

Simple errors are common

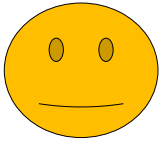


Baggerly KA, Coombes KR. 2009. Deriving chemosensitivity from cell lines: forensic bioinformatics and
reproducible research in high-throughput biology. Ann. Appl. Stat. 3(4):1309–34

Refining reproducibility

Data

Code



Refining reproducibility

Data

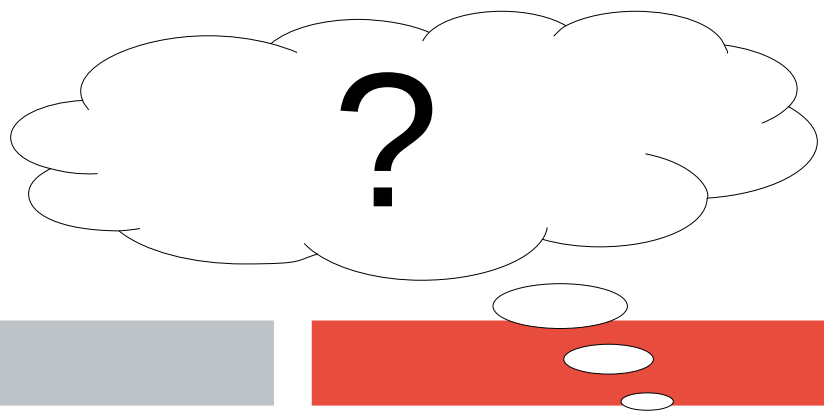
Code

More commonly shared

Big datasets are/can be uploaded

Stripping data from original context may be problematic

Best practices emerge: Tidy Data



Refining reproducibility

Data

Code

Software development has become easier

Easier to put code in packages

Increase of testing and test-based development

Refining reproducibility

Data

- Sensitive data set?
- Testing data set?
- Benchmark data set?
- Homomorphic algorithms
- Send code to data

Code

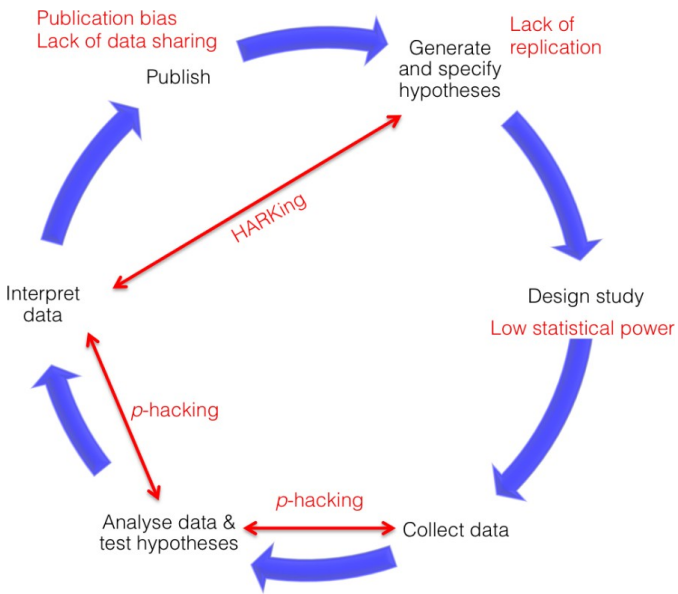
- How to make available?
- Requirements for code?
- Supply as runnable container?
- Proof of building?
- Proof of correctness?

Refining reproducibility

Paper

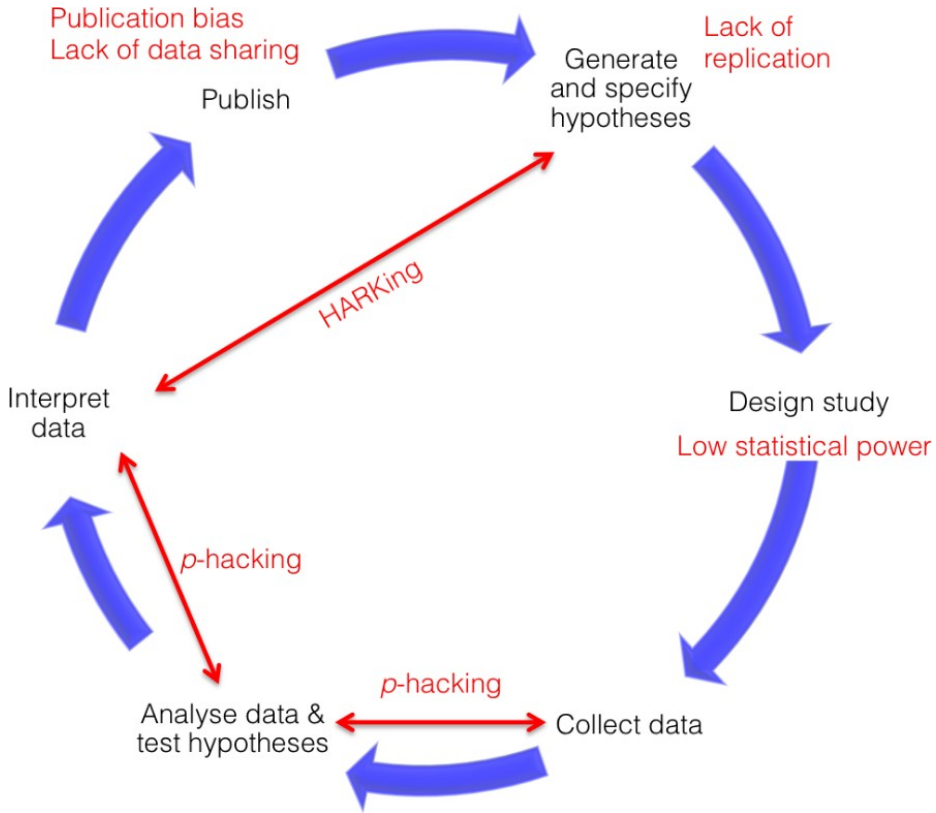
Pre-registration?

Requirements for history?



Chambers, Christopher D., et al. "Instead of" playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond." AIMS Neuroscience 1.1 (2014): 4-17.

Refining reproducibility



Chambers, Christopher D., et al. "Instead of" playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond." AIMS Neuroscience 1.1 (2014): 4-17.

Refining reproducibility

Paper

Pre-registration?

Requirements for history?

https://www.cos.io/initiatives/registered-reports

What do you think?

https://github.com/richelbilderbeek/journal_club_20220602

Responses from 2022-06-02 (n=2)

Reproducible research takes more time, which is unaccounted for

The administrative burden of reproducible research is not given enough attention

Reproducible research limits creativity (but note quote and reference below)

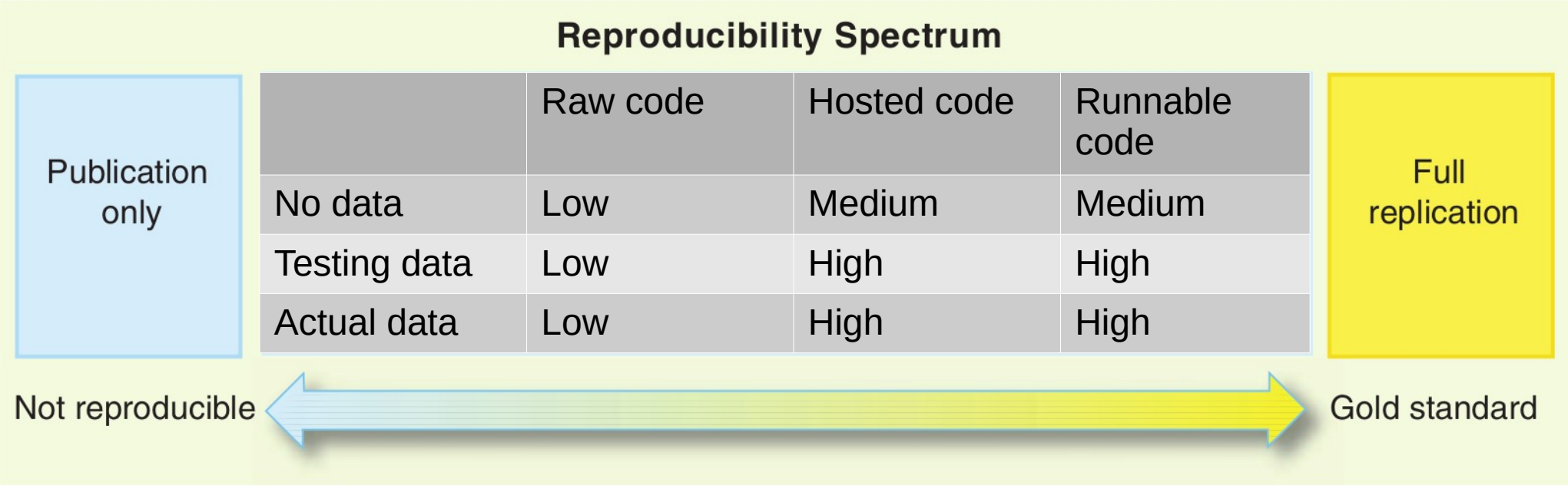
There is no incentive to do reproducible research

What happens to the copyright when code is re-used to make a figure on other data: should/must the author be attributed on that figure? Does a software license protect/help?

From Soderberg et al., 2021:
Registered Reports numerically outperformed comparison papers on all 19 criteria (mean difference 0.46, scale range −4 to +4) with effects ranging from RRs being statistically indistinguishable from comparison papers in [...] creativity (0.22, [−0.14, 0.58]) [...]

Soderberg, Courtney K., et al. "Initial evidence of research quality of registered reports compared with the standard publishing model." Nature Human Behaviour 5.8 (2021): 990-997.

Code and data are seperate components



Peng, Roger D. "Reproducible research in computational science." Science 334.6060 (2011): 1226-1227.

In school and university

Is there a solution for
$a^n + b^n = c^n$
for every n>2 ?

Yes!

Is there a solution for
$a^n + b^n = c^n$
for every n>2 ?

[complex proof here]

In academia

Complex computational pipeline

No code

Complex computational pipeline

Code public