

Reproducibility: Promoting scientific rigor and transparency

Roma Konecky, PhD
Editorial Quality Advisor

What does reproducibility mean?

- Reproducibility is the ability to generate similar results each time an experiment is duplicated.
- Data reproducibility enables us to validate experimental results.
- Reproducibility is a key part of the scientific process; however, many scientific findings are not replicable.

The Reproducibility Crisis

- ~2010 as part of a growing awareness that many scientific studies are not replicable, the phrase “Reproducibility Crisis” was coined.
- An initiative of the Center for Open Science conducted replications of 100 psychology experiments published in prominent journals. (Science, 349 (6251), 28 Aug 2015)
 - Out of 100 replication attempts, only 39 were successful.

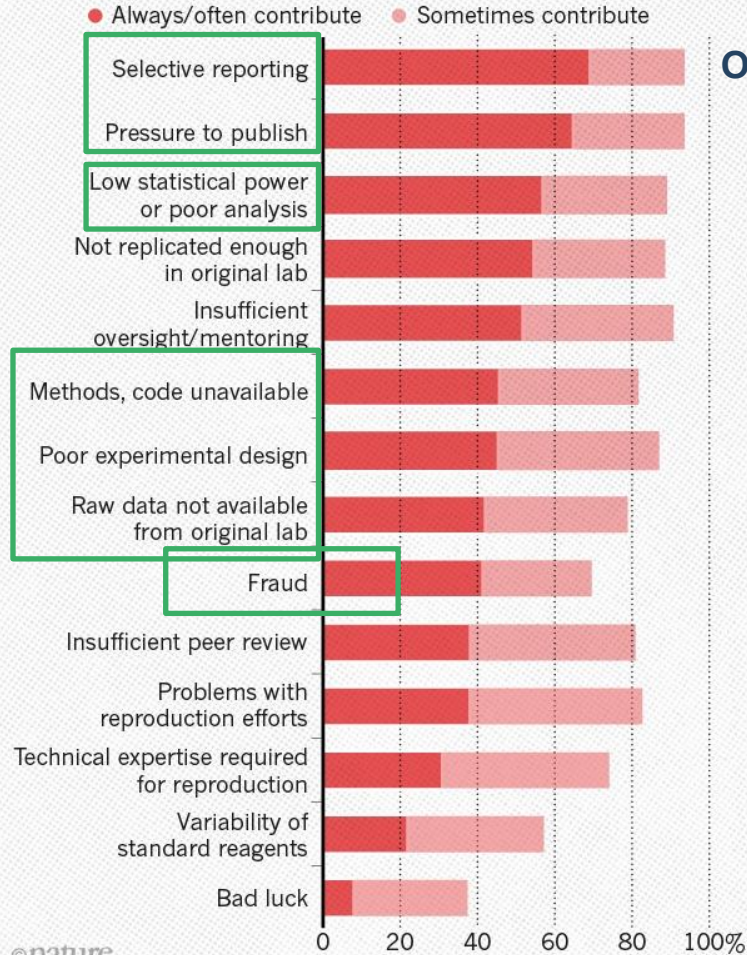
The Reproducibility Crisis

- According to a poll of over 1,500 scientists, 70% had failed to reproduce at least one other scientist's experiment or their own. (Nature 533 (437), 26 May 2016)
- Irreproducible research is a major concern because in valid claims:
 - slow scientific progress
 - waste time and resources
 - contribute to the public's mistrust of science

Factors contributing to irreproducibility

WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



Over 80% of respondents

Factors contributing to irreproducibility

Data dredging/
p-hacking



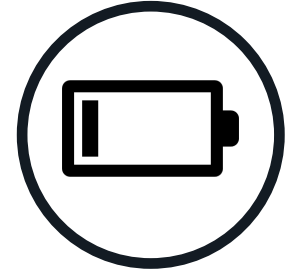
Technical
errors



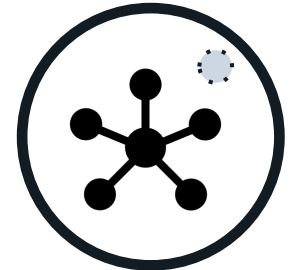
Underspecified
methods



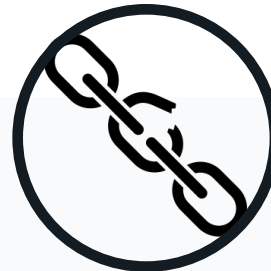
Low statistical
power



Bias - omitting
null results



Weak experimental
design



Factors contributing to irreproducibility

Data dredging/
p-hacking



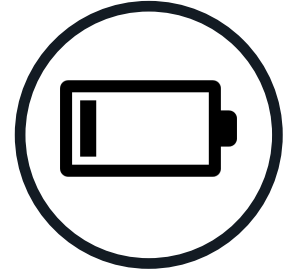
Technical
errors



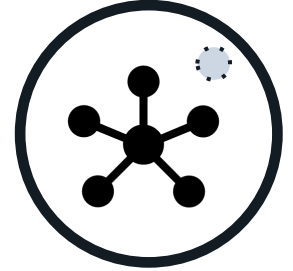
Underspecified
methods



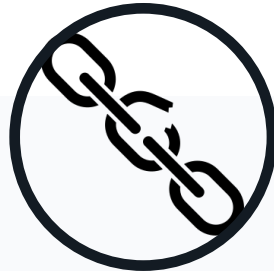
Low statistical
power



Bias - omitting
null results



Weak experimental
design



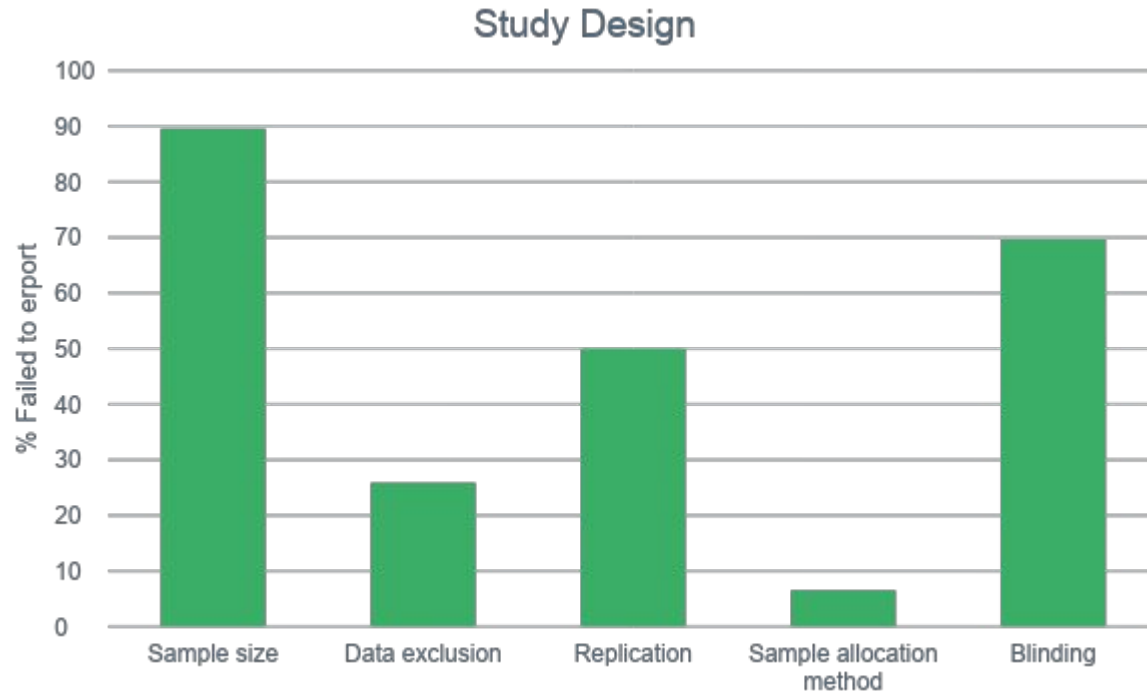


Underspecified methods

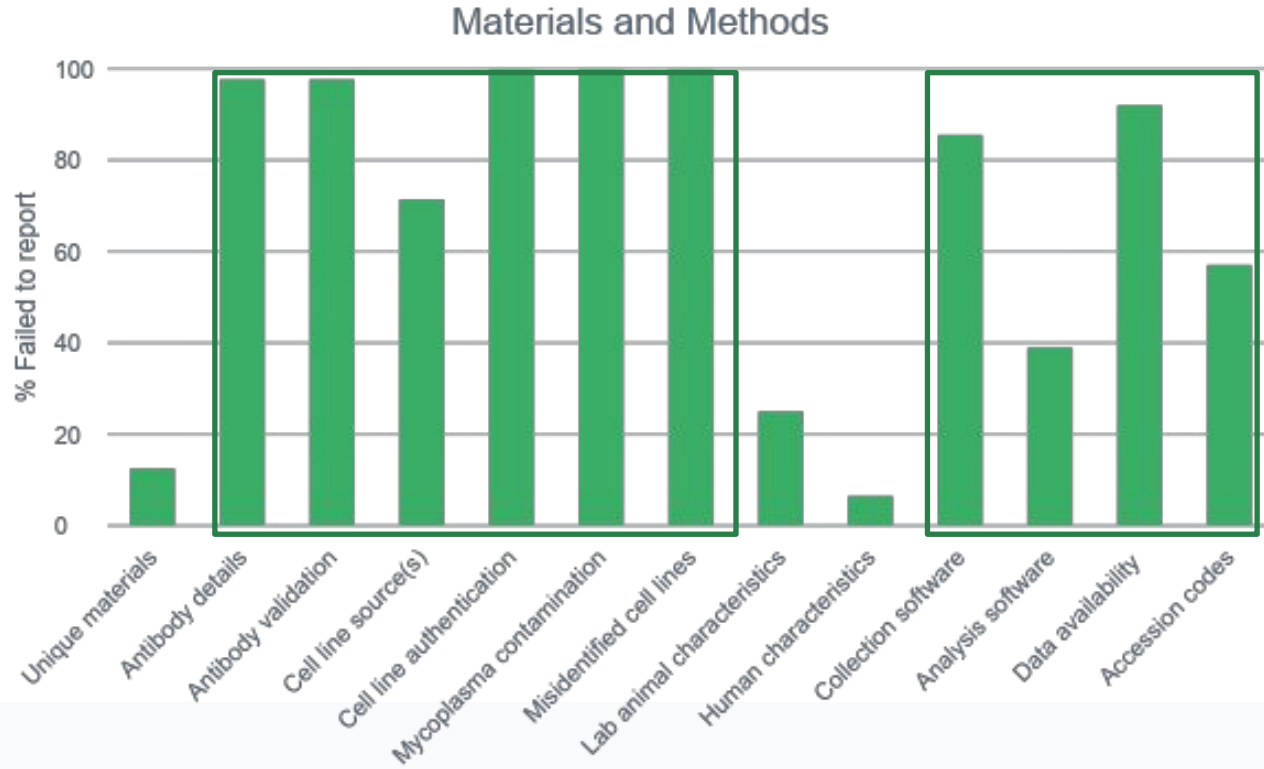
- When experimental details are omitted, the procedure needed to reproduce a study isn't clear.
- Underspecified methods are like providing only part of a recipe.



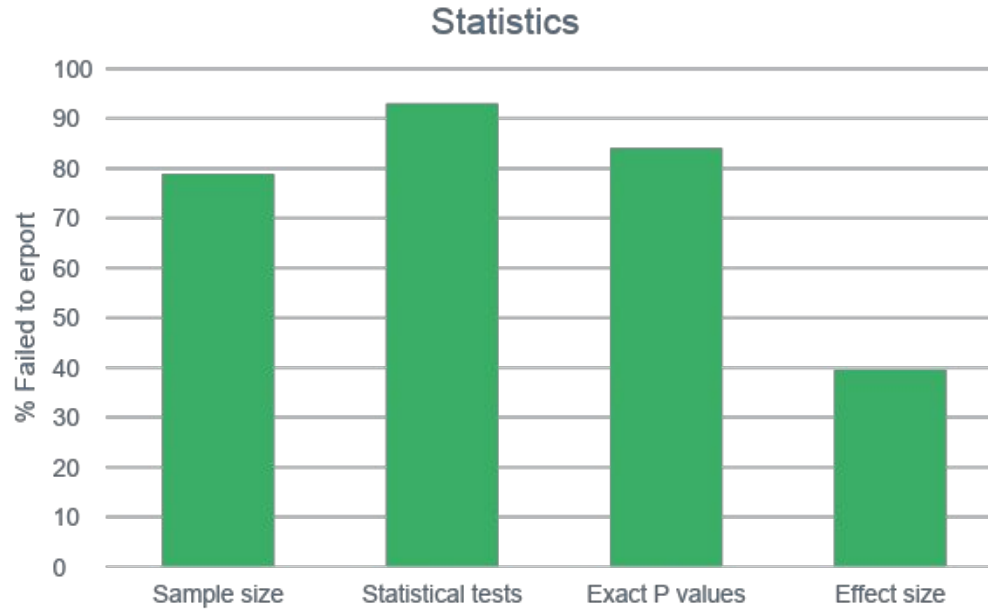
Underspecified methods



Underspecified methods



Underspecified methods



Underspecified methods



- Like baking a loaf of bread, a “scientific recipe” should include all the details needed to reproduce the study.
 - materials
 - organisms
 - instruments
 - procedures
- Without these details, we don’t know if an irreproducible finding was due to procedural differences or if the result was incorrect.

Underspecified methods



- Issues related to underspecified methods can be alleviated by:
 - decreasing journal constraints on the methods
 - publishing study protocols

Factors contributing to irreproducibility

Data dredging/
p-hacking



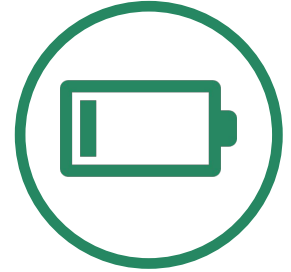
Technical
errors



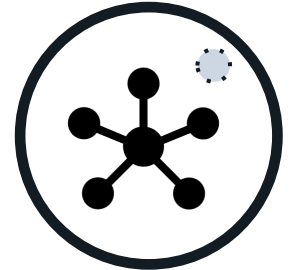
Underspecified
methods



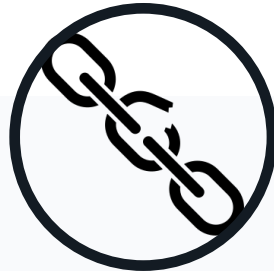
Low statistical
power



Bias - omitting
null results



Weak experimental
design





Low statistical power

- Statistical power refers to the ability of an analysis to detect a true effect.
- Pressure to publish can lead to cutting corners – using a smaller sample size than needed to detect an effect.

Low statistical power



- Underpowered studies are less likely to detect a true effect and are at a greater risk of being biased. (*Ioannidis, JPA; PLoS Medicine. 2005;2:e124*)
 - produce more false negatives
 - true effects are often exaggerated

Factors contributing to irreproducibility

Data dredging/
p-hacking



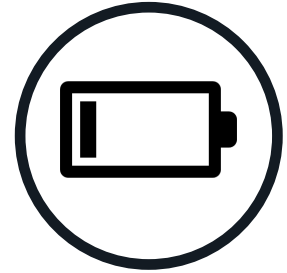
Technical
errors



Underspecified
methods



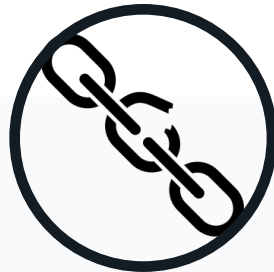
Low statistical
power



Bias - omitting
null results



Weak experimental
design





Bias – omitting null results

- Novel, statistically significant results are more likely to be published.
- Pressure to publish can lead to cherry-picking positive results and ignoring negative results.
- Consequences
 - no one learns from the null findings
 - time, money, resources are wasted



Bias – omitting null results

- Publication of null findings contributes to scientific progress by:
 - providing information that we can learn from
 - preventing others from duplicating experiments

similar

Factors contributing to irreproducibility

Data dredging/
p-hacking



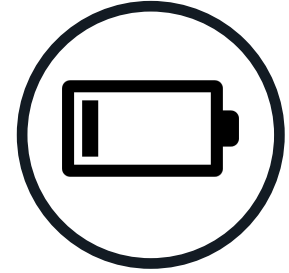
Technical
errors



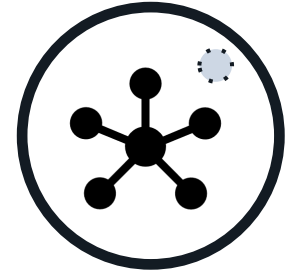
Underspecified
methods



Low statistical
power

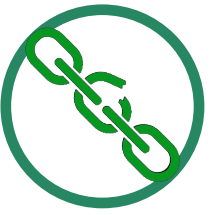


Bias - omitting
null results



Weak experimental
design

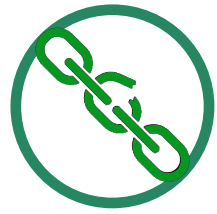




Weak experimental design

- The pressure to produce results quickly and publish them can lead to rushed experiments and cutting corners:
 - use smaller than adequate sample sizes
 - omit important control experiments
 - start an experiment before sufficient technical expertise has been acquired

Weak experimental design



“By failing to prepare, you are preparing to fail.”
— Benjamin Franklin

Take time to carefully plan out your experiments to ensure a rigorous and thorough research plan.

Factors contributing to irreproducibility

Data dredging/
p-hacking



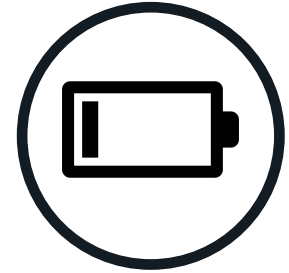
Technical
errors



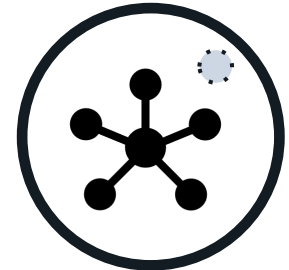
Underspecified
methods



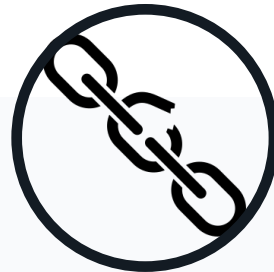
Low statistical
power



Bias - omitting
null results



Weak experimental
design





Technical errors

- Ways to minimize technical errors
 - automating your workflow
 - journal reporting guidelines and checklists
 - be critical of positive results (not just the negative ones)

Factors contributing to irreproducibility

Data dredging/
p-hacking



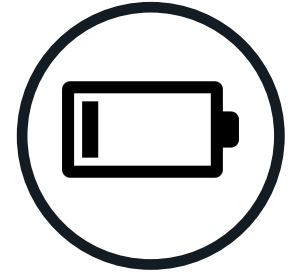
Technical
errors



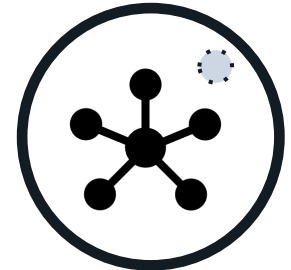
Underspecified
methods



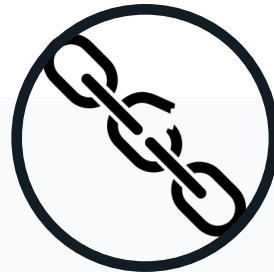
Low statistical
power



Bias - omitting
null results



Weak experimental
design



Data dredging /p-hacking



- Data dredging and p-hacking refer to the practice of repeatedly analyzing a dataset until a significant effect is found.
 - selectively reporting only significant results
 - deciding to collect more data only after a significant effect was found
 - excluding data after checking impact on p value (defining outliers post-hoc)

Data dredging /p-hacking



- Data dredging /p-hacking practices should be avoided because:
 - not hypothesis driven
 - not statistically sound
 - severely bias results

Conclusions

- Irreproducible findings are prevalent and are a major problem for science.
- Many common scientific practices contribute to the Reproducibility Crisis.
- By promoting more rigorous scientific practices, we can overcome this challenge.

Thank you!