Is reproducibility good enough?

Frida Alendal

Sunday 21 Sep, 2025

Abstract

This paper explores the importance of reproducibility and replicability in scientific research. It reviews key challenges such as publication bias, type I errors, and the replication crisis, while also considering solutions including open science practices and computable documents like Quarto. The discussion argues that replicability is an aspirational goal, but reproducibility must serve as the minimum standard to maintain credibility and trust in science.

1 Introduction

Reproducibility and replicability lie at the core of modern scientific practice. Scientific knowledge is built on the assumption that results can be verified, either by reusing the same data and methods (reproducibility) or by conducting new studies with comparable designs (replicability). However, numerous studies and large-scale projects have demonstrated that a substantial share of published findings cannot be reproduced or replicated, as argued by Ioannidis (2005). This phenomenon has contributed to what is often referred to as the "replication crisis", initially recognized in psychology but increasingly evident across various scientific disciplines.

This assignment discusses what reproducibility entails, why it is a fundamental requirement for scientific progress, and the consequences when this ideal is not fulfilled. Furthermore, it highlights key challenges such as publication bias, type I errors, and incentive structures in academia, before turning to possible solutions. Among these, the use of open data and code archives, as well as the development of "computable documents" through tools such as R Markdown and RStudio, will be emphasized. The purpose is to demonstrate how technological and methodological innovations can support a more robust, transparent, and trustworthy scientific practice. This assignment therefore asks whether reproducibility alone is good enough, or whether replicability should be demanded at the standard.

2 Literature review

2.1 Reproducibility and the credibility of scientific evidence

Peng (2011) emphasizes that reproducibility, the ability to recompute results given the same data and code, is a central benchmark for credible science, particularly when full, independent replication is infeasible. This view is also supported by Goodman et al. (2016), who stresses that reproducibility serves as the minimum requirement for credible claims. Evidence from psychology shows substantial gaps, with large-scale projects revealing that only a fraction of classic effects replicate under high-powered, pre-registered protocols (Collaboration, 2015; Klein et al., 2018). Critics such as Ioannidis (30.aug.2005) have long argued that small studies and methodological flexibility and selective reporting inflate false-positive rates, highlighting the broader replication crisis that undermines trust in research.

2.2 Open research culture and policy responses

Nosek et al. (2015) introduce the Transparency and Openness Promotion (TOP) guidelines, which outline practices such as data sharing, preregistration, and replication. At the policy level, European initiatives have promoted access to publications and data, accelerating alignment around FAIR principles and durable citation (Starr et al., 2015). Open access growth further increases visibility and reuse, with measurable citation advantages (Piwowar et al., 2018). Funders also emphasize that reproducibility is part of research integrity, and failure to provide data or code may be treated as misconduct (National Academies of Sciences, Engineering, and Medicine, 2019).

2.3 Computational reproducibility: workflows, tooling, and literate programming

Because so much modern inference is computational, transparent workflows are essential (Peng, 2011). Literate programming, first described by Knuth (1984), weaves prose and code so that analyses are executable and inspectable. In the R ecosystem, dynamic documents such as R Markdown integrate narrative, code, and outputs (Allaire et al., 2020; Xie, 2015). These tools reduce analytic drift and align with good scientific writing practices (Gentleman, 2005). More recently, Quarto has expanded these possibilities by combining code, data, text, and results in a single source document, making analyses easier to reproduce and share across disciplines (Posit Software, PBC, 2023).

2.4 Data sharing, identifiers, and metadata

Reproducibility depends on access to the exact inputs and sufficient contextual metadata. Persistent identifiers such as DOIs, along with machine-actionable metadata, enable stable citation and automated reuse (Brase, 2009). Clear, consistent data organization further reduces clerical errors and supports secondary analysis (Broman & Woo, 2018). To improve consistency, the FAIR guidelines

highlight the importance of standardized schema and minimal reporting (Wilkinson et al., 2016).

2.5 Discipline-specific lessons and exemplar domains

High-throughput fields such as genomics emphasized early the need for sharable pipelines and transparent procedures (Golub et al., 1999). These lessons generalize to other disciplines, where reproducible compendia and executable papers help bridge the gap between static publications and live analyses (Markowetz, 2015). Community surveys underline that reproducibility should not require prohibitive effort, and reproducibility cultures differ between disciplines, with computational biology moving faster than psychology or economics (Stodden et al., 2016).

2.6 Putting it into practice in R

At the project level, using scripted data ingestion, tidy transformations, and declarative visualization encourages clarity and reduces off-pipeline edits (Wickham & Grolemund, 2016). Dynamic documents (R Markdown) compiled via knitr allow text, code, and results to be integrated in a single source, which strengthens reproducibility (Xie, 2015). This is further supported by documenting session info, pinning package versions, and using reproducible environments such as renv or containers (R Core Team, 2020). Journals have also highlighted the practical value of publishing working code, even if imperfect, as it enables others to verify and build upon results (Barnes, 2010). More recently, reproducibility has been extended through platforms that allow complete notebooks and environments to be shared online (Team, 2019).

2.7 Summary

The literature points to a set of common principles that support more credible and transparent research. These include stronger incentives for openness, reliable access to data, the use of executable and literate workflows, clear standards for data and metadata, and the adoption of reproducible tools such as version control and containerization. Together, these measures help address well-known threats such as selective reporting and analytic flexibility, and they contribute to closing the gap between nominal and actual reproducibility.

3 Discussion of the research question

As outlined in the literature review, reproducibility and replicability are closely connected but not identical. A central question is whether replicability should be the norm. From a philosophical perspective, replication is essential to establish generalisable knowledge. Without it, findings risk remaining anecdotal. However, demanding universal replicability may be too ambitious in the short term. Contextual differences, limited resources, and methodological diversity make replication challenging in fields such as psychology and biomedicine. It may therefore be more practical to promote transparency and reproducibility as immediate priorities, while striving for replicability as a long-term goal.

Technological tools such as Quarto documents offer significant progress for reproducibility. By integrating code, data, text, and results in a single file, Quarto ensures that analyses can be rerun and verified by others. This reduces the risks of selective reporting, or errors that weaken credibility. Moreover, Quarto´s ability to output to multiple formats makes it easier to share reproducible research across disciplines. However, while Quarto improves computational reproducibility, it cannot guarantee replicability, which requires new data and independent confirmation.

Several challenges remain unresolved. Publication bias continues to distort the scientific record, while incentive structures in academia reward novelty over verification. Replication studies are rarely valued or published, reducing researchers´ motivation to attempt them. Technical barriers such as missing metadata, proprietary software, and inadequate infrastructure further hinder accessibility. Addressing these issues requires both cultural change and technical solutions. Policies mandating data and code sharing, preregistration of studies, and recognition of replication work as valuable contributions are crucial steps forward.

Ultimately, a combination of open science practices and technological innovation offers the best path forward. While replicability may not yet be universal, strengthening reproducibility through transparency and tools like Quarto can help rebuild trust and credibility in research.

4 Conclusion

Reproducibility and replicability are fundamental to scientific progress, yet the replication crisis has revealed major shortcomings across disciplines. Problems such as publication bias, type I errors, and inadequate data sharing undermine trust in research, while cultural and institutional incentives often work against replication. At the same time, new practices and technologies are helping to address these challenges. Open data initiatives, data citation standards, and dynamic documents in R and Quarto make it easier to reproduce research, thereby increasing transparency and accountability.

This paper concludes that while replicability should remain an aspirational goal, reproducibility must be treated as the minimum standard. By combining cultural reforms with technological tools, the scientific community can take meaningful steps to restore credibility and ensure that research provides a reliable foundation for future knowledge. Future improvements in training, incentives, and infrastructure will be crucial to ensure that reproducibility becomes not just a requirement, but a natural part of scientific culture.

4.1 Appendix: R Session Info

```
sessionInfo()
```

```
R version 4.5.1 (2025-06-13)
Platform: aarch64-apple-darwin20
Running under: macOS Sequoia 15.5
```

```
Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/lib/libRlapack.dylib;

locale:
[1] no_NO/no_NO/no_NO/C/no_NO/en_US.UTF-8

time zone: Europe/Oslo
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

loaded via a namespace (and not attached):
 [1] compiler_4.5.1    fastmap_1.2.0    cli_3.6.5        tools_4.5.1
 [5] htmltools_0.5.8.1 rstudioapi_0.17.1 yaml_2.3.10      rmarkdown_2.29
 [9] knitr_1.50        jsonlite_2.0.0   xfun_0.52        digest_0.6.37
[13] rlang_1.1.6       evaluate_1.0.4
```

References

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2020). *Rmarkdown: Dynamic documents for r.*

Barnes, N. (2010). Publish your computer code: It is good enough. *Nature*, *467*(7317), 753–753.

Brase, J. (2009). DataCite - A Global Registration Agency for Research Data. *2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology*, 257–261.

Broman, K. W., & Woo, K. H. (2018). *Data organization in spreadsheets* (No. e3183v2). PeerJ Inc.

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251).

Gentleman, R. (2005). Reproducible Research: A Bioinformatics Case Study. *Statistical Applications in Genetics and Molecular Biology*, *4*(1).

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., & Lander, E. S. (1999). Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science (New York, N.Y.)*, *286*(5439), 531–537.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*(341), 341ps12–341ps12.

Ioannidis, J. P. A. (30.aug.2005). Why Most Published Research Findings Are False. *PLOS Medicine*, *2*(8), e124.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Reginald B. Adams, Jr., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra,

R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490.

Knuth, D. E. (1984). Literate Programming. *The Computer Journal*, *27*(2), 97–111.

Markowetz, F. (2015). Five selfish reasons to work reproducibly. *Genome Biology*, *16*(1), 274.

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press. https://doi.org/10.17226/25303

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, *334*(6060), 1226–1227.

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, *6*, e4375.

Posit Software, PBC. (2023). *Quarto user guide*.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R. R., Duerr, R., Haak, L. L., Haendel, M., Herman, I., Hodson, S., Hourclé, J., Kratz, J. E., Lin, J., Nielsen, L. H., Nurnberger, A., Proell, S., Rauber, A., Sacchi, S., Smith, A., … Clark, T. (2015). Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science*, *1*, e1.

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., Heroux, M. A., Ioannidis, J. P. A., & Taufer, M. (2016). Enhancing reproducibility for computational methods. *Science*, *354*(6317), 1240–1241. https://doi.org/10.1126/science.aah6168

Team, P. J. (2019). *Binder 2.0 – reproducible, interactive, sharable environments for science at scale*.

Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data* (pp. XXV, 492). O'Reilly.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., Silva Santos, L. B. da, Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Xie, Y. (2015). *Dynamic documents with R and knitr* (Second). Chapman and Hall/CRC.