Is reproducibility good enough?

Katinca Haavardstun

2025-09-26

Reproducibility is increasingly recognized as a cornerstone of robust and reliable science. This paper explores the state of reproducibility across disciplines, including economics, and considers whether current practices are "good enough.

1 Introduction

Reproducibility is increasingly recognized as a cornerstone of robust and reliable science. This paper explores the state of reproducibility across disciplines, including economics, and considers whether current practices are "good enough." Drawing on lectures, literature, and examples of computable documents such as Quarto, the paper argues that although progress has been made, significant challenges remain in ensuring replicable research.

2 Literature review

Scientific credibility depends on whether findings can be reproduced. The so-called "replication crisis" has revealed that many results, particularly in psychology, medicine, and social sciences, cannot be replicated (Ioannidis, 2005; Nosek et al., 2015). Reproducibility is not just about re-running an experiment—it is about transparency, data availability, and robust methodology (Jasny et al., 2011; Simmons et al., 2011).

Already in early economics, researchers such as Frisch emphasized the need for transparent models and methods (Frisch, 1933). However, aspirations and practice have often diverged, raising the question: *Is reproducibility good enough today, or do we need stronger norms and better tools?*

The literature highlights multiple perspectives on reproducibility. Ioannidis (2005) famously argued that "most published research findings are false." Others, like Nosek et al. (2015), have called for systemic reforms, including preregistration and open data. Jasny et al. (2011) highlighted the importance of replication studies in major journals, while Simmons et al. (2011) demonstrated how researcher degrees of freedom inflate false-positive rates.

In economics, early methodological contributions by Ezekiel (1933) and Frisch (1933) laid a foundation for statistical rigor. Yet audits of journal archives have shown serious gaps in reproducibility (McCullough et al., 2008). Despite explicit data archiving requirements, many published results cannot be replicated in practice.

New infrastructure is emerging. Bechhofer et al. (2013) describe **Research Objects** as digital packages that capture data, code, and context. Brase (2009) explains how **DataCite DOIs** make datasets citable, linking publications directly to the underlying evidence. These developments extend beyond replication—they enable reuse and transparency.

Computable documents are another key innovation. Building on Knuth's *literate programming* (Knuth, 1984), Gentleman & Temple Lang (2007) introduced **Sweave** to combine R code and narrative. This idea expanded to **knitr** (Xie, 2015), R Markdown, and now **Quarto**, which integrates code, text, and references in one reproducible document.

3 Discussion

Should replicability be the norm? Advocates argue that it is essential for trust, efficiency, and cumulative knowledge (Young et al., 2008). Without replication, false results spread unchecked, wasting resources and damaging credibility. However, others point out that strict replication of every study is impractical. Some experiments are costly, context-specific, or require conditions that cannot be easily reproduced.

Quarto and similar tools address one part of the problem: transparency in analysis. By combining data, code, and text in one file, results become far easier to check and reproduce. Still, tools alone are insufficient. Cultural change is needed— journals must reward replication, and researchers must see transparency as part of good science rather than a burden.

Remaining challenges include selective reporting (Iyengar & Greenhouse, 1988), lack of incentives, and the sheer complexity of modern data analysis. In economics, replication efforts show progress, but reproducibility is not yet standard practice.

4 Conclusion

Reproducibility is improving but not yet "good enough." Progress is visible through data and code archives, DOIs, and computable documents. But incentives and cultural norms continue to lag behind technical solutions. The future of robust and reliable science depends on sustained effort to align tools, norms, and rewards in favor of reproducibility.

5 Software used

```
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

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M., Dunlop, I., Gamble, M., Michaelides, D., Owen, S., Newman, D., Sufi, S., & Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, *29*(2), 599–611.

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.

Ezekiel, M. (1933). Some considerations on the analysis of the prices of competing or substitute commodities. *Econometrica*, *1*(2), 172–180.

Frisch, R. (1933). Editor's Note. *Econometrica*, *1*(1), 1–4.

Gentleman, R., & Temple Lang, D. (2007). Statistical Analyses and Reproducible Research. *Journal of Computational and Graphical Statistics*, *16*(1), 1–23.

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

Iyengar, S., & Greenhouse, J. B. (1988). Selection models and the file drawer problem. *Statistical Science*, *3*(1), 109–117.

Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Again, and again, and again. *Science*, *334*(6060), 1225–1225.

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

McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2008). Do economics journal archives promote replicable research? *Canadian Journal of Economics/Revue Canadienne d'économique*, *41*(4), 1406–1420.

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.

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

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

Young, N. S., Ioannidis, J. P. A., & Al-Ubaydli, O. (2008). Why Current Publication Practices May Distort Science. *PLoS Medicine*, *5*(10), e201.