

Journal Analysis

Elia H. Thiele-Evans, Jennifer L. Beaudry

14 June, 2021

As we were interested in how journal “prestige” aligns with open science policy, we conducted a cross-comparison of journal prestige and open science initiatives. We took the top 10% of journals¹ from a Scopus database ($SD = 12$; $n = 9742$; measured through “CiteScore”, Scopus’ version of impact factor) in order to examine the distribution of open science practices in “prestigious” journals. We then measured open science initiatives through a combination of open science (TOP factor) and open access policies (Sherpa/ROMEO). We gave each journal an open science policy score, which was an aggregation of the **total** number of open science policies a journal adopts (*not* the level of policy). We then categorised the open science policy scores into quintiles: very low (0 - 2), low (3), medium (4), high (5 - 7), and very high (8 - 12). We reached the final journal sample ($n = 191$) by performing the following:

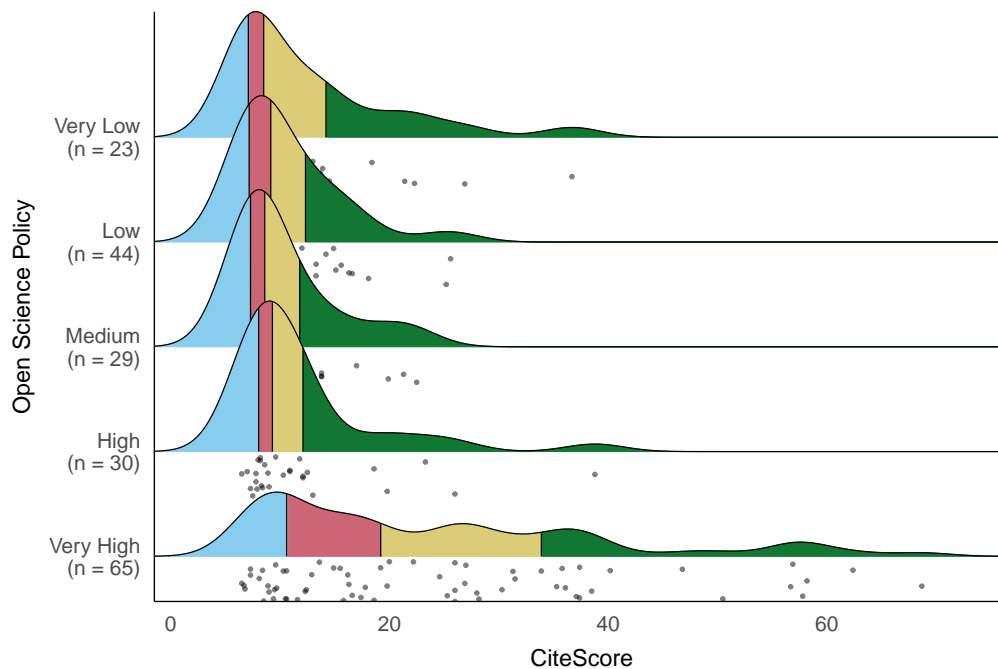
1. Retrieve CiteScore ($N = 59344$) and TOP Factor ($N = 758$) data files²
2. Filter the journals to those in the top 10% of CiteScores on Scopus
3. Remove journals from TOP that do not have an associated CiteScores on the modified CiteScore file
4. Retrieve SHERPA/RoMEO policies through their API, removing journals that did not have any OA policies listed

Figure 1 displays the distribution of open science policy score, compared to CiteScore.

¹The other 90% of CiteScores in the database fall between 0 and 6; $N = 59344$

²A helper R package was created for the handling of the TOP dataset `topapi`

Figure 1

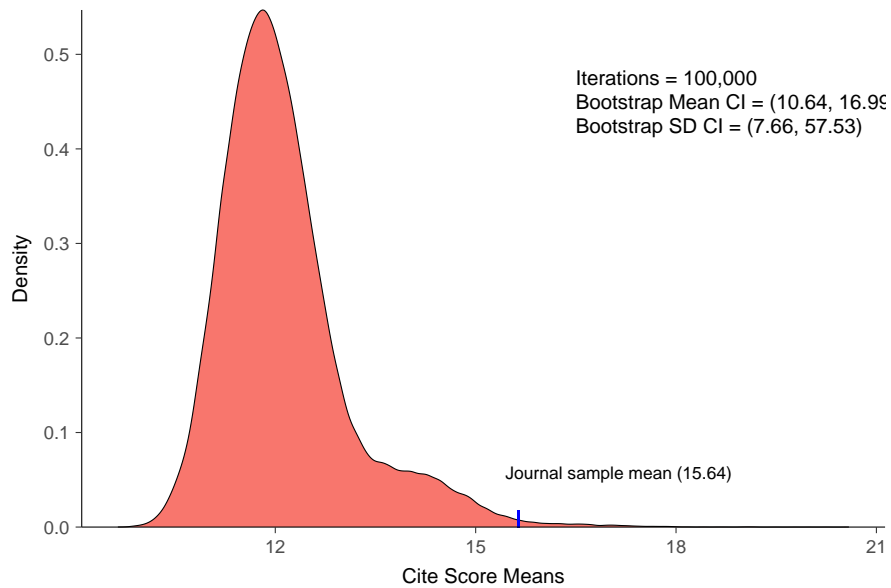


As per Figure 1, the journals in the very high group show considerable variation in their CiteScores ($SD = 16.29$). The journals of all other categories of open science policy scores show similar distributions to each other ($SD = 4.03 - 5.88$). It is conceivable that, in light of high-profile replication failures, higher impact journals have begun to adopt open science policy in some form, explaining the larger standard deviation of the “very high” group.

Did our sample reflect the CiteScore distributions of other journals?

We performed bootstrap resampling to compare the similarity of the analysis sample to a random sample from the CiteScore database. We did this in order to ensure that the journal sample had similar properties to a random sample taken from the top 10% of CiteScores. Each bootstrapped sample was of identical size to the analysis sample ($n = 191$), and was limited to the top 10% of CiteScores in the Scopus database. Figure 2 displays the distribution of bootstrapped sample means, and where in the distribution the analysis sample mean falls.

Figure 2



The bootstrap resampling indicates that the journals sampled from the TOP database fall in the 99% CI for journal means, albeit on the higher end of bootstrapped samples.

Future Directions

As the TOP database is expanded, our analysis can be expanded to include more journals. Moreover, future research could explore the level of open science policy adoption, as the current work did not differentiate whether journals were adopting a broad range of low or high level policies.

Data Sources

Center for Open science (2021). TOP Factor [Data set]. <https://osf.io/qatzk/>

Jisc. (2021). SHERPA/RoMEO [Data set]. <https://v2.sherpa.ac.uk/romeo/>

Scopus. (2021). CiteScore [Data set]. <https://www.scopus.com/sources>

Packages

Aust, F. & Barth, M. (2020). papaja: Prepare reproducible APA journal articles with R Markdown. R package version 0.1.0.9997, retrieved from <https://github.com/crsh/papaja>

Claus O. Wilke (2021). ggribes: Ridgeline Plots in ‘ggplot2’. R package version 0.5.3. <https://CRAN.R-project.org/package=ggribes>

David Robinson (2020). fuzzyjoin: Join Tables Together on Inexact Matching. R package version 0.1.6. <https://CRAN.R-project.org/package=fuzzyjoin>

Konrad Rudolph (2021). box: Write Reusable, Composable and Modular R Code. R package version 1.0.2. <https://CRAN.R-project.org/package=box>

Kropko, Jonathan and Harden, Jeffrey J. (2020). coxed: Duration-Based Quantities of Interest for the Cox Proportional Hazards Model. R package version 0.3.3. <https://CRAN.R-project.org/package=coxed>

Kun Ren (2016). rlist: A Toolbox for Non-Tabular Data Manipulation. R package version 0.4.6.1. <https://CRAN.R-project.org/package=rlist>

Landau, W. M., (2021). The targets R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. Journal of Open Source Software, 6(57), 2959, <https://doi.org/10.21105/joss.02959>

Luke Tierney, A. J. Rossini, Na Li and H. Sevcikova (2018). snow: Simple Network of Workstations. R package version 0.4-3. <https://CRAN.R-project.org/package=snow>

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.1093/josf/fkz006>

[//doi.org/10.21105/joss.01686](https://doi.org/10.21105/joss.01686)