Are Studies Published in Open Access Journals of Higher Statistical Quality Compared to Studies Published in Subscription Journals

François Léonard[1], Christian Monseur[1], Ezio Tirelli[1]

[1] University of Liège, francois.leonard@uliege.be

Poster presented at Metascience 2025, University College London, 30 June – 2 July.

Introduction

Publishing scientific articles in Open Access (OA) journals is often viewed as part of the solution to the replication crisis in science (Munafò et al., 2017). OA journals tend to promote greater transparency in research (Wicherts, 2016). However, to our knowledge, no empirical studies have yet compared the *statistical quality* of research published in OA journals with that in subscription-based (SB) journals as represented by statistical power.

This study aims to assess whether tests evaluating the efficacy of psychotherapies for depression published in OA and SB articles differ in terms of (I) sample sizes, (II) statistical power and (III) observed effect sizes (ES).

Methods

Studies included in Cuijpers et al. (2023) were sourced from *Metapsy* database (Miguel et al., 2022). Only primary outcomes based on a continuous measure were selected from this database.

All necessary information to test our hypotheses was available in the database, except the publication type and statistical power. The publication type was determined by accessing the journal's website via the DOI. The statistical power of each included test was calculated for a two-sided two-sample t-test with an alpha level of 5% and the three effect sizes as defined by Cohen's classification. Sample sizes, statistical power and observed ES were compared using a Mann-Whitney U test with the publication type as the independent variable. ES for Mann–Whitney tests are reported as point-biserial correlations r_{pb}.

Results

Out of the 415 studies (826 ES) in the *Metapsy* database, 357 studies (467 ES) were included. In this sample, 167 studies (199 ES) were published in OA and 190 studies (268 ES) were published in SB.

Tests published in OA exhibit larger total sample sizes than tests published in SB articles ($r_{pb} = 0.27$, $99\%CI = [0.15, 0.38]$, $p = 5.25 \times 10^{-9}$, Figure 1).

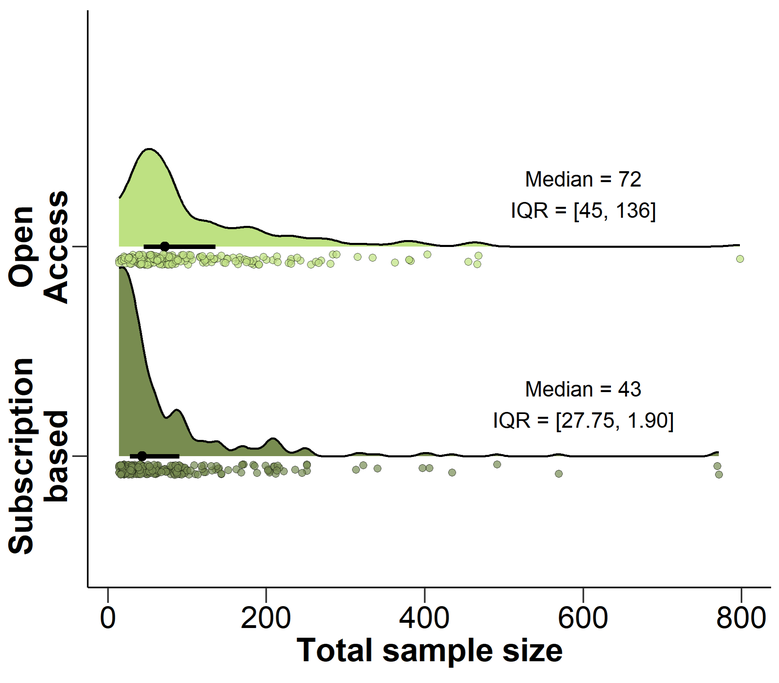

Figure 1: Distribution of total sample sizes in tests published in OA (top) and SB (bottom) articles. Black dots and lines indicate the median and the interquartile range. Colored points represent sample size in each article. Overall, the median sample size is 58, IQR=[31.5, 105]

Statistical power for the three ES in Cohen's classification is higher for tests published in OA than for tests published in SB articles (for small, medium and large ES: $r_{pb} = 0.269$, $99\%CI = [0.15, 0.38]$, Figure 2 and Table 1).

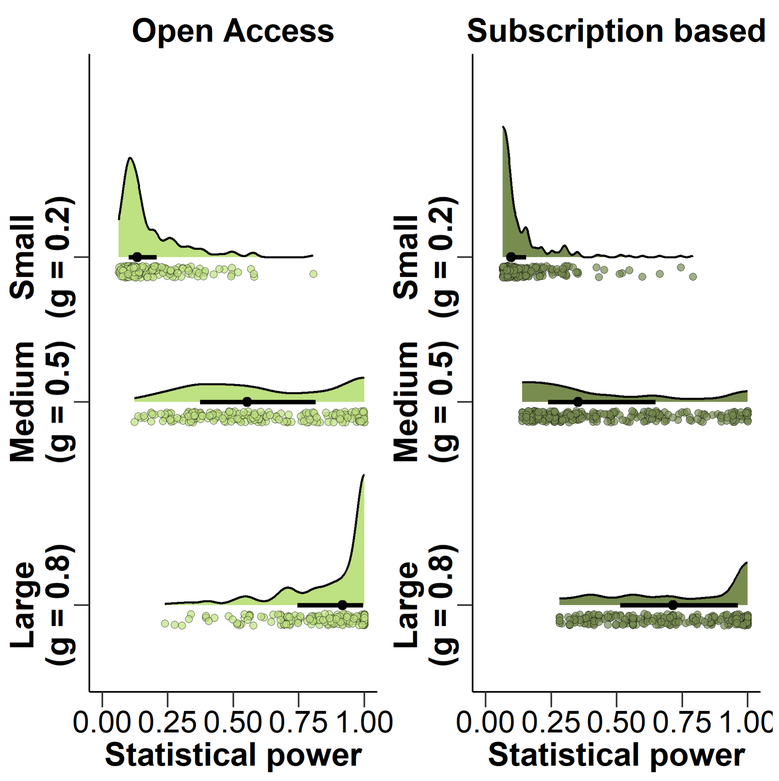

Figure 2: Distribution of statistical power for detecting small (g = 0.2), medium (g = 0.5), and large (g = 0.8) ES in tests published in OA (left) and SB (right) articles Black dots and lines indicate the median and interquartile range of the distribution. Colored points represent observed tests

Table 1: median statistical power for small, medium and large hypothetical ES in all included tests, tests published in OA and tests published in SB.

ES	Overall	OA	SB
Small	0.115	0.133	0.097
Medium	0.458	0.552	0.353
Large	0.843	0.917	0.716

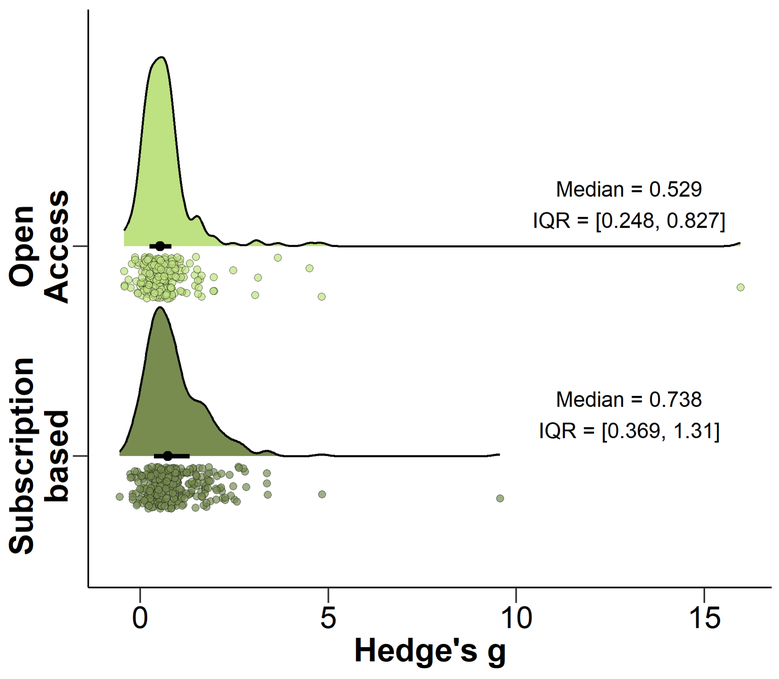

Figure 3: Distribution of observed ES (Hedge's g) in tests published in OA (top) and SB (bottom) articles. Black dots and lines indicate the median and the interquartile range. Colored points represent observed effect sizes. Overall, the median ES is 0.646, IQR=[0.312, 1.051].

In other words, articles in this literature fall short by approximately 730 participants for detecting a small ES, and by 70 participants for a medium ES. For a large ES, they include 6 more participants than required to achieve the recommended 80% power. In OA articles, the shortfall is 716 participants for a small ES and 56 for a medium ES. For a large ES, OA articles include 20 more participants than required. In SB articles, the sample size falls short by 745 participants for a small ES, 85 for a medium ES, and 9 for a large.

Finally, ES published in OA articles are statistically smaller than those published in SB articles ($r_{pb} = 0.187$, $99\%CI = [0.07, 0.29]$, $p = 5.41 \times 10^{-5}$, Figure 3).

Discussion

Our results suggest that, on average and based on our indicators, studies published in OA journals exhibit higher statistical quality than those published in SB journals.

Studies published in OA journals tend to include significantly larger total sample sizes than those in SB journals. This results in higher statistical power. Consequently, tests published in OA journals are able to detect smaller effect sizes.

Although our results do not allow causal inference, they raise the possibility that OA studies may follow more rigorous design or reporting practices. These differences may also reflect variations in submission standards, editorial policies, or peer-review processes. To account for the potential non-independence of effect sizes originating from the same articles, we plan to conduct multilevel analyses in future work. Nonetheless, statistical power remains suboptimal for small and medium effect sizes across both publication models.

Finally, this pattern may be influenced by structural factors within the scientific ecosystem, such as funding policies, institutional expectations, researcher profiles, or access to publication resources. Exploring these factors could help clarify the mechanisms behind our findings. This pattern may differ in fields where OA publishing is less common or less promoted.

Conclusion

In the scientific literature devoted to the evaluation of the efficacy of psychotherapies for depression, studies published in Open Access journals show larger sample sizes, higher statistical power and smaller effect sizes compared to those published in Subscription-Based journals, suggesting higher statistical quality for these articles.

Further research is needed to examine whether these differences generalize to other fields.

References

Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Ebert, D., & Karyotaki, E. (2023). Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. *World Psychiatry*, *22*(1), 105–115. https://doi.org/10.1002/wps.21069

Miguel, C., Harrer, M., & Plessen, C. Y. (2022). *Metapsy-project/data-depression-psyctr: Version 22.0.2*. https://doi.org/10.5281/zenodo.7255037

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9. https://doi.org/10.1038/s41562-016-0021

Wicherts, J. M. (2016). Peer review quality and transparency of the peer-review process in open access and subscription journals. *PLOS ONE*, *11*(1), e0147913. https://doi.org/10.1371/journal.pone.0147913