The importance of no evidence

Publication bias threatens the ability of science to self-correct. It's time to change how null or negative findings are perceived and offer incentives for their publication.

Positive or statistically significant findings are much more likely to see the light of day than null or negative findings. Publication bias—the tendency of authors or journals to prioritize for publication positive findings—is not a new phenomenon. In a 1959 article, Sterling described the potential threat of the bias towards statistically significant results for fields that rely on frequentist statistics: it is possible that the literature in these fields largely consists of false conclusions¹. Writing in 1979, Rosenthal coined the term ‘file drawer problem’, describing its most extreme version conceivable as “journals are filled with the 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results”².

Although this extreme version is not true, evidence of the extent of publication bias and the file drawer problem has been unsettling. For instance, Fanelli analysed over 4,600 papers across all disciplines published between 1990 and 2007 to determine the proportion of papers reporting positive support for the hypotheses tested³. He found that publication bias across all disciplines increased by over 22% during the period studied, exceeding 80% for all years after 1999 and peaking at 88.6% in 2005. Publication bias did not seem to affect all disciplines equally: the proportion of positive results was significantly greater when moving from the physical sciences to the biological sciences to the social sciences.

As support for them amasses, scientific findings come to be accepted as facts. However, if only positive findings are published, the risk of these ‘facts’ being false is substantial. A modelling study showed that unless a sufficient proportion of negative findings are published, false claims can indeed become accepted as fact⁴. The problem is also worsened by questionable research practices, such as *P*-hacking, data dredging and HARK-ing (hypothesizing after the results are known), all of which have a documented high prevalence.

A variety of solutions for publication bias have been identified, from a range of

tools aimed at determining the extent of publication bias in meta-analyses, to the launch of journals dedicated to publishing negative results, to mandatory registration for clinical trials. None of these solutions, however, fully addresses the problem. Although tools for determining the extent of bias in a corpus of studies are invaluable, failure to disclose null or negative findings leads to significant waste of resources in unknowingly repeating failed studies. ‘Segregating’ negative results in dedicated journals does little to elevate their status in the research and publishing world. And, although mandatory preregistration helps address a host of questionable research practices, it does not fully address publication bias unless publication following preregistration becomes mandatory across all the sciences.

Publication bias cannot be effectively addressed unless our collective attitude towards null or negative findings changes. Journals—especially selective journals—have a primary role to play in this. At *Nature Human Behaviour*, we welcome the submission of studies reporting null or negative findings, provided that they address an important question of broad significance and are methodologically highly robust. An example of such a study is published in this issue, by Forscher and colleagues (<https://www.nature.com/articles/s41562-018-0517-y>).

The authors of this study asked an important question: do grant reviewers discriminate on the basis of the gender or race of the grant proposal's principal investigator? Participants in their study were actual researchers who had previously submitted R01 grant proposals to the NIH and many of whom had also acted as reviewers on such proposals. Their participants were collectively asked to evaluate 48 actual NIH proposals that differed only in one respect among conditions in the experiment: the name of the principal investigator was manipulated to be stereotypically white male, white female, black male or black female. The authors found no evidence of pragmatically important bias (defined

as half a point or more on the 9-point assessment scale).

Not all null findings are interpretable or meaningful, of course. Poorly motivated or poorly executed research is not informative, regardless of the direction of the results. Studies yielding no evidence for the hypothesized effect are not interpretable unless they are sufficiently powered to detect the effect of interest. Forscher et al. preregistered their analysis plan and ensured that their study was sufficiently powered to detect pragmatically important bias. They also subjected their analyses to several robustness checks.

A positive finding in this case would likely have made more sensational news headlines and it would have provided the NIH and other funders with an immediate target for intervention. However, the absence of evidence of bias is also informative and important: although bias may exist in other aspects of the granting process, this experiment suggests that the initial peer review process is likely not affected by it; or, if bias exists, the effect is so small as to be negligible.

A focus on confirmatory, statistically significant results undermines the aims of science and its ability to self-correct. Publication continues to be the key vehicle for researcher recognition and career advancement. It's time to shift the emphasis from positive outcomes to the importance of the question and the methodological rigour of studies, regardless of their outcome. That's why at *Nature Human Behaviour* we strongly encourage the submission of Registered Reports (where a decision for publication is made before the results are available) and welcome methodologically rigorous studies on broadly important questions that report null or negative findings. □

Published online: 12 March 2019
<https://doi.org/10.1038/s41562-019-0569-7>

References

1. Sterling, T. D. *J. Am. Stat. Assoc.* **54**, 30–34 (1959).
2. Rosenthal, R. *Psychol. Bull.* **86**, 638–641 (1979).
3. Fanelli, D. *Scientometrics* **90**, 891–904 (2012).
4. Nissen, S. B., Magidson, T., Gross, K. & Bergstrom, C. T. *eLife* **5**, e21451 (2016).