Metascience

https://doi.org/10.1038/s41559-023-02156-z

Exaggerated effects in ecology

Timothy H. Parker & Yefeng Yang

Check for updates

A study of over 18,000 effect sizes from more than 350 published studies in ecology finds clear evidence of selective reporting and exaggeration of effect sizes.

Growing concerns about the reliability of published research have led scientists in a variety of disciplines to direct their scientific lenses onto science itself. This research on research, known as 'meta-science', has produced evidence that the published literature is often plagued by biased and unreliable results[1–3]. Writing in *Nature Ecology & Evolution*, Kimmel and colleagues[4] make an important contribution to the rapidly growing body of evidence that ecologists often conduct underpowered studies and engage in selective reporting and, in so doing, exaggerate the size of effects in the published literature. Their conclusions — based on 18,917 effect sizes from 354 published empirical ecology studies in the journals *Ecology*, *Ecology Letters*, *Journal of Ecology*, *Science* and *Nature* — are consistent with conclusions from two other recent meta-science studies based on different data and methodologies[5,6].

The simplest and so maybe most convincing result from Kimmel and colleagues was their observation that the distribution of P values in the published literature is what we would expect owing to reporting and publication bias. Instead of a unimodal frequency distribution of P values (converted to t statistics in their analysis), there are two modes caused by a rarity of P values just above 0.05 (the common threshold of statistical significance). This pattern is absent among P values derived from supplements, which demonstrates the value of supplementary materials for reducing reporting bias. However, when supplementary results were combined with results from the main text, the rarity of

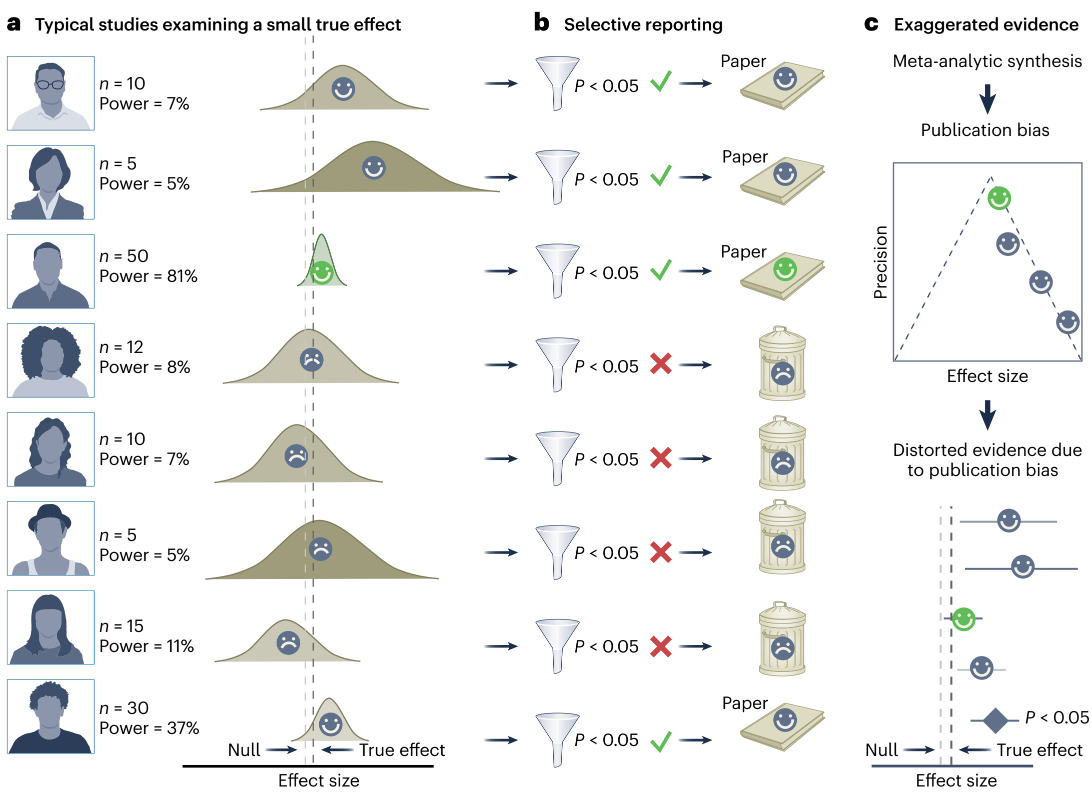

Fig. 1 | How low power and disproportionate publishing of 'significant' effects leads to exaggerated effects in the literature. a, When statistical power is low owing to small samples (depicted in shades of tan), as is common in ecology, studies will reach statistical significance thresholds only when sampling error produces a strong pattern by chance. We see this in **a** where the strongest effects are found in the studies with low power. For example, a low-powered study (7%) with a small sample size of $n = 10$ may result in a sixfold exaggeration of the effect (known as a type M error). By contrast, a well-powered study (depicted in green) can achieve statistical significance without an exaggerated effect owing to its small sampling error and high precision. **b**, The publication filter often leads to selective reporting (also known as reporting and publication bias), in which effects with $P > 0.05$ are often filtered (frowning face to represent bias) and those with $P < 0.05$ are more likely to be published. **c**, This selective reporting leads to a disproportionate representation by exaggerated effects and so distorts the evidence derived from meta-analyses that is used for informing scientific research and decision-making in ecology and evolutionary biology.

non-significant P values remained evident, which demonstrates that supplements have not eliminated reporting bias. These results complement another recent finding that the proportion of statistically significant P values in the ecology literature may be twice as high as would be expected in the absence of bias[5].

Among the studies in their sample, Kimmel and colleagues found worryingly low power (13%) to detect the average effect size – far short of the oft-cited standard of 80% for high power. This means that most ecology studies have an insufficient sample size to detect the average effect with the common standard of $P < 0.05$. But even if the average effect that researchers were attempting to distinguish from zero was 2.5 times larger than the average calculated by Kimmel and colleagues, more than 50% of statistical tests in their dataset would still fall short of 80% power. This observation of low power is typical of recent observations of the literature in ecology, as well as in disciplines that range from medicine and neuroscience to economics[6].

With insufficient statistical power, reaching statistical significance thresholds typically occurs when sampling error creates an inflated effect size due to chance (Fig. 1). When many 'significant' effects are inflated in this way and reporting or publication bias is common, this leads to the disproportionate publication of inflated effects[7]. Kimmel and colleagues found that most effects from the studies with low (<80%) power were much larger than the effects from studies with high (>80%) power. Over 60% of effects from low-powered studies were at least twice as large (and often much larger) than the average effect from high-powered studies. The bias toward publishing these inflated effect sizes creates misleading representations of nature in the typical ecology paper and in later meta-analyses. Estimates of how much meta-analytic averages are inflated range from 10% when not accounting for publication bias[6,8] to between 30% and 150% when accounting for bias[6]. This bias has also caused a large fraction of published meta-analytic P values in ecology to incorrectly cross the traditional threshold of statistical significance[6].

It is important to acknowledge that the results of Kimmel and colleagues regarding statistical power and the exaggeration of effect size depend on an assumption that, to some extent, is probably wrong: that ecologists do not adjust their sample sizes in response to anticipated effect size (sampling more when small effects are anticipated, and less when anticipating large effects). However, there are at least three reasons not to be concerned by this assumption. First, ecologists rarely report conducting power analysis[4], so if they are adjusting sampling effort, they do not appear to be using a rigorous process. Second, sample sizes in ecological studies are often constrained by resources and so the option to increase power is often unavailable[9]. Finally, two other recent studies that are arguably less dependent on this assumption found even stronger patterns of low power[5,6] and inflated effects[6] in ecology.

Although many meta-science questions in ecology remain to be investigated, the recent work by Kimmel et al. and other groups provides compelling evidence that publication and reporting bias are common in ecology and that this may often influence the conclusions we draw from the literature. Given this, we might ask what we should do. Kimmel and colleagues argue that we need a cultural change that includes both a rethinking of what ecologists consider high-quality work and the adoption of policies that can lead to greater reliability. As they point out, these cultural changes are already underway and supported by SORTEE (the Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology). Kimmel and colleagues also provide a helpful outline of some of the proposed policy ideas. For instance, they mention registered reports, in which study and analysis plans are peer-reviewed by journals before data collection (see ref. 6 for a recent example). Among other benefits, registered reports appeared to reduce reporting bias in a recent meta-science study in psychology[10].

But will ecologists embrace efforts to incentivize or require practices to improve the reliability of the discipline? A look at mandatory data archiving gives us a clue. More than 10 years after the first journals in ecology and evolutionary biology began requiring data archiving, there is widespread consensus regarding its value. So, we know we can implement bold new practices for the good of the discipline. And given what we now also know about publication and reporting bias in ecology and evolutionary biology from Kimmel et al. and other recent papers[5,6], the potential benefits of adopting policies to reduce this bias are clear. Now we just need to follow the lesson of data archiving and step up our game for reliable science.

Timothy H. Parker[1] ✉ **& Yefeng Yang**[2] ✉
[1]Department of Biology, Whitman College, Walla Walla, WA, USA.
[2]Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, Australia.
✉e-mail: parkerth@whitman.edu; yefeng.yang1@unsw.edu.au

Published online: 03 August 2023

References

1. Fanelli, D. *PLoS ONE* **5**, e10068 (2010).
2. Open Science Collaboration. **349,** aac4 716 (2015).
3. Parker, T. H. et al. *Trends Ecol. Evol.* **31**, 711–719 (2016).
4. Kimmel, K., Avolio, M. L. & Ferraro, P. J. *Nat. Ecol. Evol.*, https://doi.org/10.1038/s41559-023-02144-3 (2023).
5. Deressa, T. et al. Preprint at *EcoEvoRxiv*, https://doi.org/10.32942/X24G6Z (2023).
6. Yang, Y. et al. *BMC Biol.* **21**, 71 (2023).
7. Gelman, A. & Loken, E. *Am. Sci.* **102**, 460–465 (2014).
8. Fox, J. W. *Ecol. Evol.* **12**, e9521 (2022).
9. Lemoine, N. P. et al. *Ecology* **97**, 2554–2561 (2016).
10. Scheel, A. M., Schijen, M. R. M. J. & Lakens, D. *Adv. Methods Pract. Psychol. Sci.* **4**, https://doi.org/10.1177/25152459211007467 (2021).

Competing interests

T.H.P. is a co-founder and an elected member of the board of directors of SORTEE. Y.Y. declares no competing interests.