

# The significance gap: statistical significance rates decrease from primary literature to effects used in ecological meta-analyses

Danijela Žanko<sup>1</sup>, Antica Culina<sup>1,\*</sup>

<sup>1</sup>Laboratory for Informatics and Environmental Modelling, Division for Marine and Environmental Research, Ruđer Bošković Institute, Bijenička Cesta 54, 10000 Zagreb, Croatia

\*Corresponding author: Antica Culina ([aculina@irb.hr](mailto:aculina@irb.hr))

ORCID: Danijela Žanko [0009-0005-0436-0796](https://orcid.org/0009-0005-0436-0796); Antica Culina [0000-0003-2910-8085](https://orcid.org/0000-0003-2910-8085)

# The significance gap: statistical significance rates decrease from primary literature to effects used in ecological meta-analyses

## Abstract

Publication bias is widely considered a major threat to ecological synthesis, yet whether significance rates observed in the general literature are reflected in studies used in meta-analyses remains unknown. Using two large, independent datasets—26,547 results from 359 primary studies and 104,418 effects from 17,879 studies included in 464 ecological meta-analyses—we quantify a “significance gap” across the research-to-synthesis pipeline. In primary studies from the general literature, 49.7% of individual results were statistically significant. However, 98.9% of these studies contained at least one significant result. Among individual effects extracted for meta-analyses, 41.1% were significant, but only 63.6% of the contributing studies contained at least one significant effect. Rarefaction analysis showed that much of the disparity in study-level significance rates between the two datasets diminishes, yet persists, when equalised to the same number of effects per study. These patterns suggest that publication bias operates mainly at the study level: studies apparently need at least one significant finding to be published, but once published, they freely report null results alongside the significant ones. By extracting effects beyond the headline findings, meta-analyses may partly counteract rather than reinforce publication bias.

**Keywords:** evidence synthesis, meta-research, scientific publishing, effect size, statistical significance

## *Synthesis*

Nearly all published ecological studies contain at least one statistically significant result, yet only half of the individual results reported in these studies are significant. Meta-analyses show a sharp drop in the study-level significance rate, while the overall prevalence of significant effects is only moderately lower. This “significance gap” suggests publication bias acts as a study-level

28 filter, and that meta-analyses may partly dilute bias by extracting results beyond the headline  
29 findings.

## 30 1 Introduction

31 In ecology, meta-analyses increasingly guide conservation policy, management decisions, and  
32 theoretical development. Yet, the validity of their conclusions depends on whether the included  
33 published literature represents an unbiased sample of completed research. Thus, when statisti-  
34 cally significant results are preferentially published [Sterling, 1959], meta-analyses aggregate a  
35 distorted evidence base, potentially overestimating the effect sizes [Ioannidis, 2005, Gurevitch  
36 et al., 2018, van Zwet and Cator, 2021, Nakagawa et al., 2022]. Indeed, [Yang et al., 2023] found  
37 that meta-analytic means in ecology are overestimated by at least 0.12 standard deviations,  
38 with 66% of initially significant results becoming non-significant after correction for publication  
39 bias. What remains a critical gap is an understanding of how publication bias manifests at  
40 different stages of the research-to-synthesis pipeline—and at which level it primarily acts.

41 Meta-analyses in ecology are commonly based on published results reported in journal arti-  
42 cles. Thus, to be included in a meta-analysis, results must undergo a hierarchical process. First,  
43 researchers commonly conduct a number of analyses, but often submit only selected results to  
44 a journal, preferably those that are 'significant'—a form of selective reporting termed outcome  
45 reporting bias [Fanelli, 2012, John et al., 2012, Head et al., 2015, Fraser et al., 2018, Gopalakr-  
46 ishna et al., 2022, van Aert and Wicherts, 2024]. Some of the submitted results end up being  
47 published, and consequently, published manuscripts tend to contain at least some significant  
48 results [Sterling, 1959, Fanelli, 2012]. Meta-analysts then subsequently search for published  
49 studies and extract effects from them. In this process, they might predominantly extract sig-  
50 nificant effects—because significant effects are more likely to be reported in full [Cassey et al.,  
51 2004], or because of extraction bias, given that meta-analysts typically extract only a fraction of  
52 effects from included studies [Pollo et al., 2025]. Alternatively, meta-analyses might capture null  
53 results that exist within published studies but receive less prominence (e.g. in supplements),  
54 potentially resulting in lower overall significance rates.

55 Here, we use two independent datasets—359 primary studies from the general ecological liter-  
56 ature [Kimmel and Ferraro, 2023] and 17,879 primary studies from 464 ecological meta-analyses  
57 [Fox and Costello, 2022]—to quantify significance rates at different stages of the research-to-

58 synthesis pipeline. Although the two datasets do not share the same studies, comparing their  
 59 significance rates indicates a “significance gap” and provides insight into how publication bias  
 60 manifests across hierarchical levels of the evidence base.

Table 1: Key terms and definitions used in this study.

Term	Definition
General literature dataset	Dataset of 359 primary studies published in five ecological journals (2018–2020); data from <a href="#">Kimmel et al. [2023]</a>
Meta-analytic dataset	Dataset of 17,879 primary studies included in 464 ecological meta-analyses (1991–2020); data from <a href="#">Costello and Fox [2022]</a>
Overall significance rate	Proportion of individual effects with $p < 0.05$ among all effects in each dataset
Study-level significance rate	Proportion of studies containing $\geq 1$ significant effect in each dataset
Within-study significance rate	Proportion of significant effects within a study; reported as a mean (and median) across studies

## 61 2 Methods and results

62 For the general literature dataset (Table [1](#)), we used data from studies published in five high-  
 63 impact journals (Ecology, Ecology Letters, Journal of Ecology, Science, Nature) between Jan-  
 64 uary 2018 and May 2020, originally collected by [Kimmel et al. \[2023\]](#), who extracted all es-  
 65 timated parameters and their associated standard errors from main-text and supplementary  
 66 tables. From the archived dataset of [Kimmel and Ferraro \[2023\]](#), we retained 26,547 estimates  
 67 with valid coefficients and standard errors from 359 studies (358 with calculable p-values), prior  
 68 to the trimming steps applied in their original analysis (Kimmel et al. report 354 after addi-  
 69 tional data cleaning). For the meta-analytic dataset (Table [1](#)), we used a database of ecological  
 70 meta-analyses published between 1991 and 2020 and compiled by [Costello and Fox \[2022\]](#). From  
 71 the archived dataset of [Fox and Costello \[2022\]](#), we obtained 104,418 effect sizes from 17,879  
 72 studies used in 464 meta-analyses, after removing extreme values (absolute effect sizes  $>5$  and  
 73 variances  $>2$ ) from the original set of 111,385 effect sizes from 467 meta-analyses.

74 Because neither dataset reported p-values, we derived two-tailed p-values by dividing each  
 75 reported value of the effect by its standard error and applying a normal distribution. Recom-  
 76 puting p-values with t-distributions (degrees of freedom 10–100) yielded similar estimates. For  
 77 each dataset, we then quantified significance rates at three levels (Table [1](#)): the *overall sig-*  
 78 *nificance rate*, the *study-level significance rate*, and the *within-study significance rate*. For the  
 79 meta-analytic dataset, all three rates reflect only the effects that meta-analysts extracted, which  
 80 are typically a subset of all results reported in the original primary study.

81 We compared the rates of statistical significance between the two datasets for 'raw' data and  
 82 using rarefaction analyses, where, for each dataset, we randomly drew the same number of effects  
 83 per study ( $k \in \{1, 2, 3, 5, 10, 20\}$ ), classified each study as containing at least one significant effect  
 84 or not, and repeated this 1,000 times to obtain stable estimates. We stopped at  $k = 20$  because  
 85 higher values would exclude most studies in the meta-analytic dataset, where the median number  
 86 of extracted effects per study is 2. To test whether studies reporting fewer effects have a higher  
 87 proportion of significant results, we calculated Spearman rank correlations between the number  
 88 of effects and the within-study significance rate. All analyses were conducted in R version 4.4.2  
 89 [R Core Team, 2024], using dplyr (v1.1.4) [Wickham et al., 2023], ggplot2 (v4.0.1) [Wickham,  
 90 2016], and patchwork (v1.2.0) [Pedersen, 2024].

91 In the general literature dataset, the overall significance rate was 49.7% (13,187 of 26,547  
 92 individual estimates were significant at  $p < 0.05$ ). However, the study-level significance rate was  
 93 98.9% (354 of 358 studies contained at least one significant result). If individual estimates within  
 94 a study are independent, an assumption that is likely violated in the dataset, the probability  
 95 that a study contains at least one significant result is  $1 - (1 - 0.497)^n$ , where  $n$  is the number of  
 96 estimates in that study. Averaging across all studies gives an expected rate of 99.0%, indicating  
 97 that this high study-level rate is expected given the number of estimates per study, though  
 98 both quantities may reflect publication pressure. The within-study significance rate averaged  
 99 60.2% (median = 61.7%), and studies reporting fewer estimates had higher significance rates  
 100 (Spearman  $\rho = -0.253$ ,  $p < 0.001$ ; Fig. 1A).

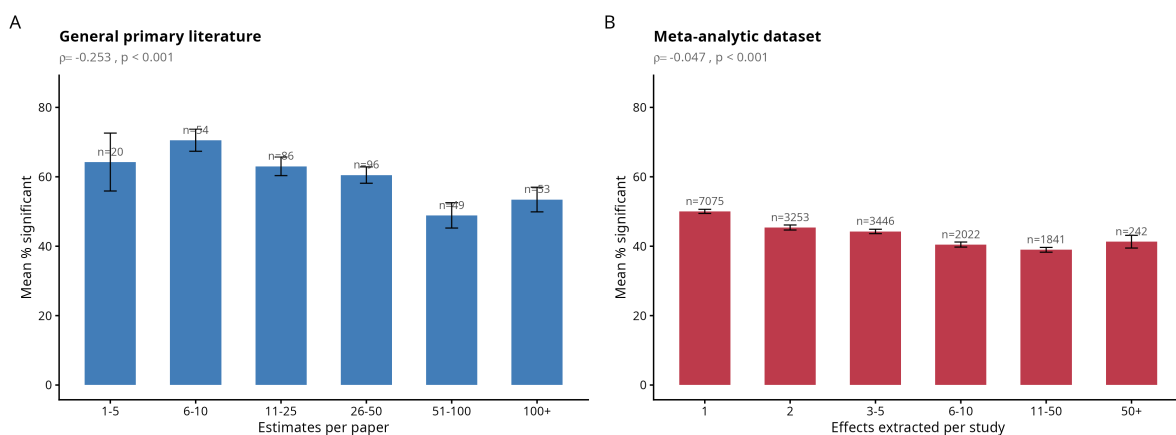


Figure 1: Within-study significance rate by number of reported results. (A) Primary studies from the general literature ( $n = 358$  studies; median 30 estimates per study). (B) Primary studies from meta-analyses ( $n = 17,879$  studies). Error bars show standard errors. Studies reporting fewer results have higher within-study significance rates, consistent with selective prominence of significant findings.

101 In the meta-analytic dataset, the overall significance rate was 41.1% (42,888 of 104,418  
 102 individual effects were significant)—of comparable magnitude to that we detected in the general  
 103 literature. This rate was broadly consistent across effect size measures (log response ratios:  
 104 47.8%, Hedges’  $d$ : 35.6%, Fisher’s  $z$ : 43.4%), suggesting that the result is not driven by any  
 105 single effect size type. However, the study-level significance rate was only 63.6% (11,374 of  
 106 17,879), substantially lower than the 98.9% we detected in the general literature. Using the same  
 107 calculation as above ( $1 - (1 - 0.411)^n$ , averaged across studies), the expected rate would be 67.2%,  
 108 suggesting that meta-analytic studies are, if anything, slightly less likely to contain a significant  
 109 effect than chance alone would predict. The within-study significance rate averaged 45.7%  
 110 (median = 42.9%), and its distribution was bimodal with peaks near 0% and 100% (Fig. 2B),  
 111 contrasting with the right-skewed distribution among primary studies from the general literature  
 112 (Fig. 2A). As in the general literature, studies with fewer extracted effects had slightly higher  
 113 significance rates, but this relationship was much weaker (Spearman  $\rho = -0.047$ ,  $p < 0.001$ ;  
 114 Fig. 1B), consistent with meta-analysts extracting effects more uniformly regardless of their  
 115 significance.

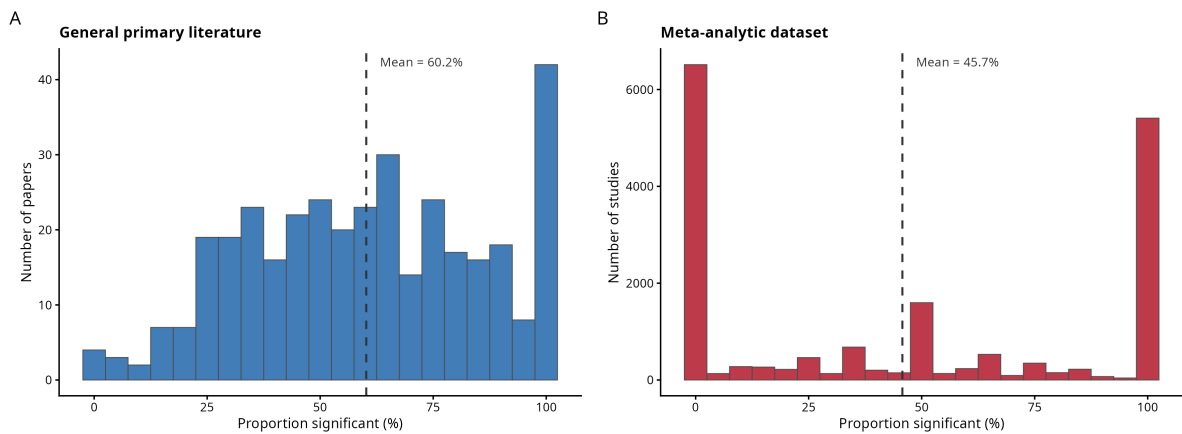


Figure 2: The distribution of the proportion of the effects that were significant ( $p < 0.05$ ) for: (A) Primary studies from the general literature ( $n = 358$  studies; median 30 estimates per study). Most studies have 40–80% of their results significant, producing a right-skewed distribution. (B) Primary studies from meta-analyses ( $n = 17,879$  studies; median 2 effects per study). The distribution is bimodal, with peaks near 0% and 100%, because most studies contribute only one or two extracted effects—a study with one non-significant effect scores 0%, and a study with one significant effect scores 100%. Dashed lines indicate unweighted means.

116 Rarefaction analysis, in which  $k$  effects were randomly subsampled per study, revealed that  
 117 much of the difference in the study-level significance rate between the two datasets reflects  
 118 differences in the number of effects per study (Figure 3). At  $k = 1$ , where each study contributes

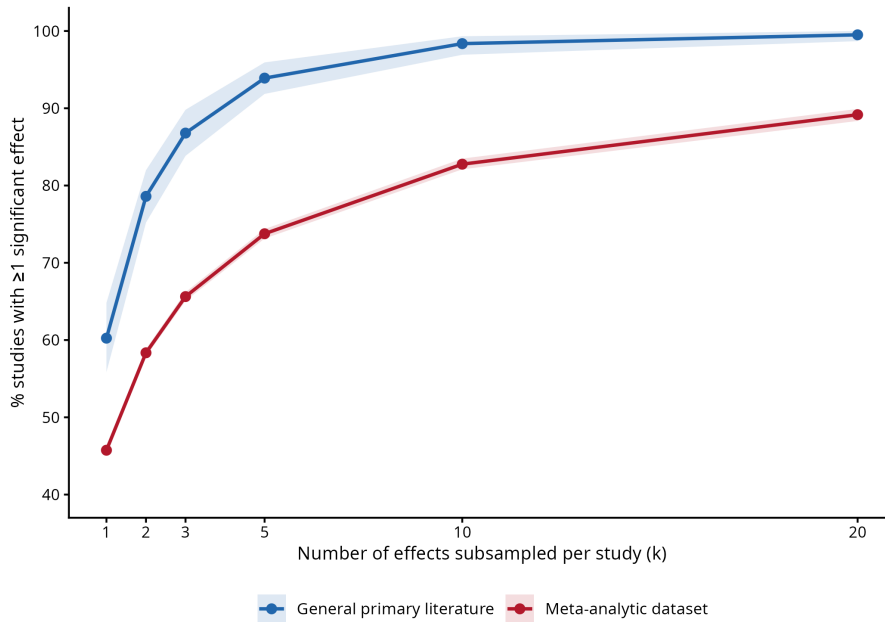


Figure 3: Rarefaction analysis: the proportion of studies with at least one significant effect ( $p < 0.05$ ) when subsampling  $k$  effects per study. Shaded bands show 95% confidence intervals from 1,000 iterations. The gap persists at all  $k$  values but is substantially smaller than the raw study-level comparison (98.9% vs. 63.6%).

119 a single randomly drawn effect, 60% of studies in the general literature dataset and 46% of  
 120 studies in the meta-analytic dataset contained a significant result—a gap of approximately 14%,  
 121 substantially smaller than the unadjusted study-level comparison (98.9% vs. 63.6%). The gap  
 122 persisted across all  $k$  values: at  $k = 20$ , the maximum number of effects subsampled per study  
 123 in our analysis, study-level significance reached approximately 99% ( $n = 229$  studies retained)  
 124 in the general literature dataset, and approximately 89% ( $n = 1,031$  studies retained) in the  
 125 meta-analytic dataset. Thus, the gap narrows but persists when studies from both datasets are  
 126 compared at the same number of effects, indicating that the unadjusted study-level comparison  
 127 overstates the difference, but that a real difference in the overall significance rate between the  
 128 two datasets still remains.

### 129 3 Discussion

130 Our results reveal several gaps in the overall statistical significance rate between the general  
 131 literature dataset and the meta-analytic dataset. These gaps indicate the presence of several  
 132 processes that determine the flow of results from those that get published to those that are  
 133 included in the meta-analyses.

134 First, our results indicate that a study needs to report at least one significant effect to be  
135 published. Although the near-universal study-level significance rate we detected in the general  
136 literature dataset (98.9%) is expected given the number of estimates per study, the number of  
137 estimates per study is likely itself partly shaped by publication pressure. Researchers might  
138 conduct additional analyses, test alternative model specifications, or explore secondary outcomes  
139 precisely because of the need for at least one significant finding for publication [Head et al.,  
140 2015, Fraser et al., 2018]. Our result that studies with a lower number of effects show higher  
141 significance rates within the study supports this assumption. Similar study-level significance  
142 rates have been observed in biomedical research [Turner et al., 2008, Chavalarias et al., 2016,  
143 Cristea and Ioannidis, 2018], suggesting that this pattern may extend beyond ecology. Small  
144 sample sizes, which are typical in ecology [Fox, 2025], compound the problem by increasing the  
145 proportion of significant results that are exaggerated in magnitude [Gelman and Carlin, 2014].  
146 Simulations confirm that outcome reporting bias can severely inflate meta-analytic estimates  
147 [Friese and Frankenbach, 2020, van Aert and Wicherts, 2024].

148 Second, our findings suggest that by extracting multiple effects per study, meta-analysis  
149 may partly counteract publication bias. While the study-level significance rate is near 100% in  
150 the general literature, it is much lower in the meta-analytic dataset (63.6%). Our rarefaction  
151 analysis confirms that the gap between the two datasets is not solely driven by differences in  
152 the number of effects per study. Finally, at the effect level, the general literature studies con-  
153 tain more significant effects than the meta-analytic dataset (49.7% vs. 41.1%). Since main-text  
154 results have higher significance rates than supplementary results (60.1% vs. 48.5%, calculated  
155 from the general literature dataset), and meta-analyses miss on average 16% of relevant effect  
156 sizes from included studies [Pollo et al., 2025], meta-analysts appear to extract beyond the head-  
157 line findings. The study-level significance rate in the meta-analytic dataset (63.6%) fell slightly  
158 below the expected rate under independence (67.2%), consistent with, though not uniquely  
159 explained by, the recovery of non-significant effects. The bimodal distribution of within-study  
160 significance rates in the meta-analytic dataset largely reflects low extraction density (57.8% of  
161 studies contributed only one or two effects), as also noted in Fig. 2

162 Our study has three limitations. First, both datasets consist exclusively of published work,  
163 so the rates we report should not be interpreted as unbiased baselines. Second, the two datasets  
164 contain different statistical quantities—statistically estimated parameters reported in tables  
165 (e.g., regression coefficients, means, treatment effects) among primary studies from the general

166 literature versus standardised effect sizes (Hedges'  $d$ , log response ratios, Fisher's  $z$ ) in the  
167 meta-analytic dataset. Thus, the difference in the overall significance rates might partly reflect  
168 this compositional difference. Third, our sample of primary studies from the general literature  
169 covers five high-impact journals over three years, which may not represent the full breadth of  
170 ecological research.

171 Overall, based on our results, we conclude that the number of results reported in primary  
172 studies and the number of effects meta-analysts extract from a single study can amplify or  
173 dampen the bias inherited from the primary literature [van Aert and Wicherts, 2024]. Extraction  
174 protocols that deliberately target supplementary and secondary results could function as a  
175 partial corrective, complementing statistical corrections for outcome reporting bias [van Aert  
176 and Wicherts, 2024] without requiring access to unpublished data.

## 177 References

- 178 P. Cassey, J.G. Ewen, T.M. Blackburn, and A.P. Møller. A survey of publication bias within  
179 evolutionary ecology. *Proceedings of the Royal Society B: Biological Sciences*, 271(suppl\_6):  
180 S451–S454, 2004. doi: 10.1098/rsbl.2004.0218.
- 181 D. Chavalarias, J.D. Wallach, A.H.T. Li, and J.P.A. Ioannidis. Evolution of reporting p values  
182 in the biomedical literature, 1990–2015. *JAMA*, 315(11):1141–1148, 2016. doi: 10.1001/jama.  
183 2016.1952.
- 184 L. Costello and J.W. Fox. Decline effects are rare in ecology. *Ecology*, 103(6):e3680, 2022. doi:  
185 10.1002/ecy.3680.
- 186 I.A. Cristea and J.P.A. Ioannidis. P values in display items are ubiquitous and almost invariably  
187 significant: A survey of top science journals. *PLOS ONE*, 13(5):e0197440, 2018. doi: 10.1371/  
188 journal.pone.0197440.
- 189 D. Fanelli. Negative results are disappearing from most disciplines and countries. *Scientomet-*  
190 *rics*, 90(3):891–904, 2012. doi: 10.1007/s11192-011-0494-7.
- 191 J.W. Fox. The sample size of the typical ecological correlation coefficient is small and slowly  
192 declining. *Oikos*, 2025:e11430, 2025. doi: 10.1002/oik.11430.

193 J.W. Fox and L. Costello. Decline effects are rare in ecology: a meta-meta-analysis, 2022. URL  
194 <https://doi.org/10.5061/dryad.zkh1893b7>. Accessed 15 December 2025.

195 H. Fraser, T. Parker, S. Nakagawa, A. Barnett, and F. Fidler. Questionable research practices in  
196 ecology and evolution. *PLOS ONE*, 13(7):e0200303, 2018. doi: 10.1371/journal.pone.0200303.

197 M. Friese and J. Frankenbach. p-hacking and publication bias interact to distort meta-analytic  
198 effect size estimates. *Psychological Methods*, 25(4):456–471, 2020. doi: 10.1037/met0000246.

199 A. Gelman and J. Carlin. Beyond power calculations: assessing type S (sign) and type M  
200 (magnitude) errors. *Perspectives on Psychological Science*, 9(6):641–651, 2014. doi: 10.1177/  
201 1745691614551642.

202 G. Gopalakrishna, G. ter Riet, G. Vink, I. Stoop, J.M. Wicherts, and L.M. Bouter. Prevalence of  
203 questionable research practices, research misconduct and their potential explanatory factors:  
204 a survey among academic researchers in The Netherlands. *PLOS ONE*, 17(2):e0263023, 2022.  
205 doi: 10.1371/journal.pone.0263023.

206 J. Gurevitch, J. Koricheva, S. Nakagawa, and G. Stewart. Meta-analysis and the science of  
207 research synthesis. *Nature*, 555(7695):175–182, 2018. doi: 10.1038/nature25753.

208 M.L. Head, L. Holman, R. Lanfear, A.T. Kahn, and M.D. Jennions. The extent and conse-  
209 quences of p-hacking in science. *PLOS Biology*, 13(3):e1002106, 2015. doi: 10.1371/journal.  
210 pbio.1002106.

211 J.P.A. Ioannidis. Why most published research findings are false. *PLoS Medicine*, 2(8):e124,  
212 2005. doi: 10.1371/journal.pmed.0020124.

213 L.K. John, G. Loewenstein, and D. Prelec. Measuring the prevalence of questionable research  
214 practices with incentives for truth telling. *Psychological Science*, 23(5):524–532, 2012. doi:  
215 10.1177/0956797611430953.

216 K. Kimmel and P.J. Ferraro. Replicability in ecology. Dataset, OSF, 2023. URL <https://osf.io/9yd2b>.  
217 Accessed 15 December 2025.

218 K. Kimmel, M.L. Avolio, and P.J. Ferraro. Empirical evidence of widespread exaggeration bias  
219 and selective reporting in ecology. *Nature Ecology & Evolution*, 7(9):1525–1536, 2023. doi:  
220 10.1038/s41559-023-02144-3.

221 S. Nakagawa, M. Lagisz, M.D. Jennions, J. Koricheva, D.W.A. Noble, T.H. Parker, A. Sánchez-  
222 Tójar, Y. Yang, and R.E. O’Dea. Methods for testing publication bias in ecological and  
223 evolutionary meta-analyses. *Methods in Ecology and Evolution*, 13(1):4–21, 2022. doi: 10.  
224 1111/2041-210X.13724.

225 T.L. Pedersen. *patchwork: The composer of plots*, 2024. URL [https://CRAN.R-project.org/  
226 package=patchwork](https://CRAN.R-project.org/package=patchwork). R package version 1.2.0. Accessed 15 December 2025.

227 P. Pollo, M. Lagisz, R.C. Macedo-Rego, A. Mizuno, Y. Yang, and S. Nakagawa. Reliability  
228 of meta-analyses in ecology and evolution: (mostly) good news from a case study on sexual  
229 signals. *Proceedings of the Royal Society B: Biological Sciences*, 292(2047):20242782, 2025.  
230 doi: 10.1098/rspb.2024.2782.

231 R Core Team. *R: A language and environment for statistical computing*. R Foundation for Sta-  
232 tistical Computing, Vienna, Austria, 2024. URL <https://www.R-project.org/>. Accessed  
233 15 December 2025.

234 T.D. Sterling. Publication decisions and their possible effects on inferences drawn from tests of  
235 significance—or vice versa. *Journal of the American Statistical Association*, 54(285):30–34,  
236 1959. doi: 10.1080/01621459.1959.10501497.

237 E.H. Turner, A.M. Matthews, E. Linardatos, R.A. Tell, and R. Rosenthal. Selective publication  
238 of antidepressant trials and its influence on apparent efficacy. *The New England Journal of*  
239 *Medicine*, 358(3):252–260, 2008. doi: 10.1056/NEJMsa065779.

240 R.C.M. van Aert and J.M. Wicherts. Correcting for outcome reporting bias in a meta-analysis:  
241 A meta-regression approach. *Behavior Research Methods*, 56(3):1994–2012, 2024. doi: 10.  
242 3758/s13428-023-02132-2.

243 E.W. van Zwet and E.A. Cator. The significance filter, the winner’s curse, and the need to  
244 shrink. *Statistica Neerlandica*, 75(4):437–452, 2021. doi: 10.1111/stan.12241.

245 H. Wickham. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag, New York, 2016.  
246 ISBN 978-3-319-24277-4.

247 H. Wickham, R. François, L. Henry, K. Müller, and D. Vaughan. *dplyr: A grammar of data*  
248 *manipulation*, 2023. URL <https://CRAN.R-project.org/package=dplyr>. R package version  
249 1.1.4. Accessed 15 December 2025.

250 Y. Yang, A. Sánchez-Tójar, R.E. O’Dea, D.W.A. Noble, J. Koricheva, M.D. Jennions, T.H.  
251 Parker, M. Lagisz, and S. Nakagawa. Publication bias impacts on effect size, statistical power,  
252 and magnitude (Type M) and sign (Type S) errors in ecology and evolutionary biology. *BMC*  
253 *Biology*, 21(1):71, 2023. doi: 10.1186/s12915-022-01485-y.