RESEARCH · RESEARCH INTEGRITY

Publication Bias and the File-Drawer Problem

What publication bias is, how the file-drawer problem distorts the evidence, the research that measured it, its consequences for reviews and decisions, methods for detecting and correcting it, and the reforms that reduce it.

ModQu research article · 2,920 words · 29 references (APA 7th edition)

Suppose twenty research teams test the same idea, and the idea is false. By chance alone, one team may obtain a statistically significant result. If that team publishes and the other nineteen file their results away, anyone reading the literature will see one study supporting the idea and none against it. This is the essence of **publication bias**: the tendency for the results of research to influence whether, when and where it is published. It is one of the most important and best documented threats to the reliability of published evidence. This article explains what it is, how it has been measured, why it matters, how to detect it and what can be done to reduce it.

Definitions

Publication bias occurs when the likelihood that a study is published depends on the nature or direction of its results. Most commonly, studies with statistically significant or positive results are more likely to be published than studies with null or negative results.

Rosenthal (1979) gave the problem its most memorable name. He described the **file drawer problem**: in the extreme case, journals publish the 5 per cent of studies that show false positive results by chance, while the 95 per cent that find nothing stay in researchers' file drawers. Few believed the extreme case literally applied, but the image captured how the published record could differ systematically from the research actually done.

Publication bias is one of a family of **dissemination biases**. Song et al. (2010), reviewing the evidence, distinguished several related forms, including time-lag bias, where positive results are published faster; language bias, where positive results are more likely to appear in English-language journals; citation bias, where positive results are cited more often; and **outcome reporting bias**, where a study is published but only some of its outcomes are reported, chosen according to the results.

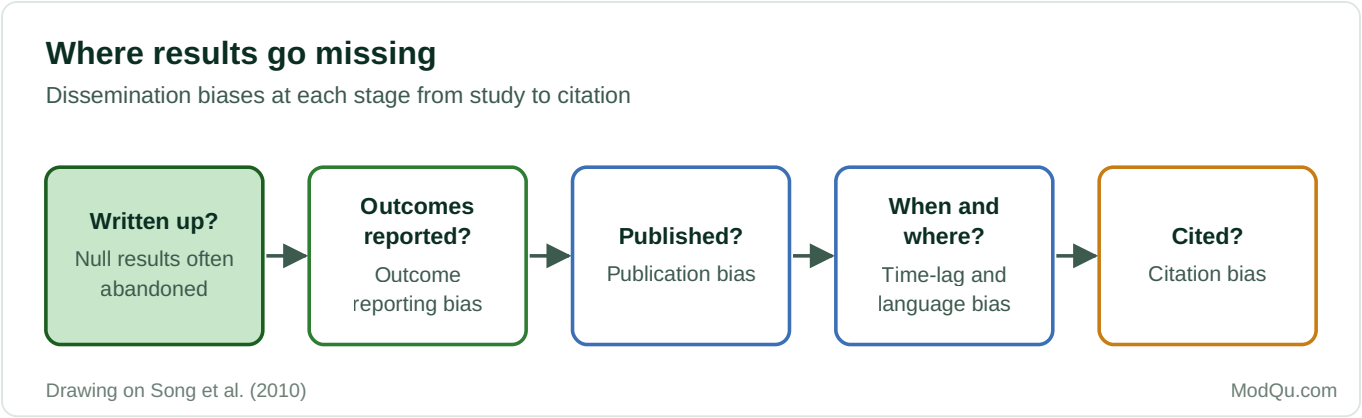


Figure 1. Stages at which the results of research can be lost or distorted. Diagram by ModQu, drawing on Song et al. (2010).

The early evidence

Concern about publication bias is old. Sterling (1959) examined articles in four psychology journals and found that, of those using tests of significance, about 97 per cent reported rejecting the null hypothesis. Either almost every hypothesis tested in psychology was true, or studies that failed to reject the null were not being published. Sterling et al. (1995) repeated the exercise more than three decades later and found that little had changed.

Greenwald (1975) surveyed authors and reviewers in social psychology and found a clear prejudice against the null hypothesis: researchers were more likely to submit, and reviewers more likely to recommend, papers that rejected it. He showed how this prejudice could fill the literature with false findings.

Measuring publication bias directly

The strongest evidence comes from studies that follow a defined set of research projects from approval or registration to publication, so that published and unpublished results can be compared.

Easterbrook et al. (1991) followed studies approved by a research ethics committee in Oxford and found that those with statistically significant results were more than twice as likely to be published as those without. Dickersin (1990) reviewed the early evidence and concluded that publication bias was real and that the main cause was investigators failing to submit null results, rather than journals rejecting them.

Franco et al. (2014) studied social science survey experiments funded through a common programme, where all studies, published or not, could be identified. Strong results were far more likely to be published than null results. Most null results were never written up at all, apparently because researchers expected them to be uninteresting or unpublishable.

The most striking example comes from medicine. Turner et al. (2008) compared antidepressant trials registered with the United States Food and Drug Administration with what was published in journals. According to the regulator's records, about half of the trials had positive results. In the published literature, almost all appeared positive. Many negative trials were not published, and some were published in a way that presented them as positive. As a result, the published literature overstated the effectiveness of the drugs.

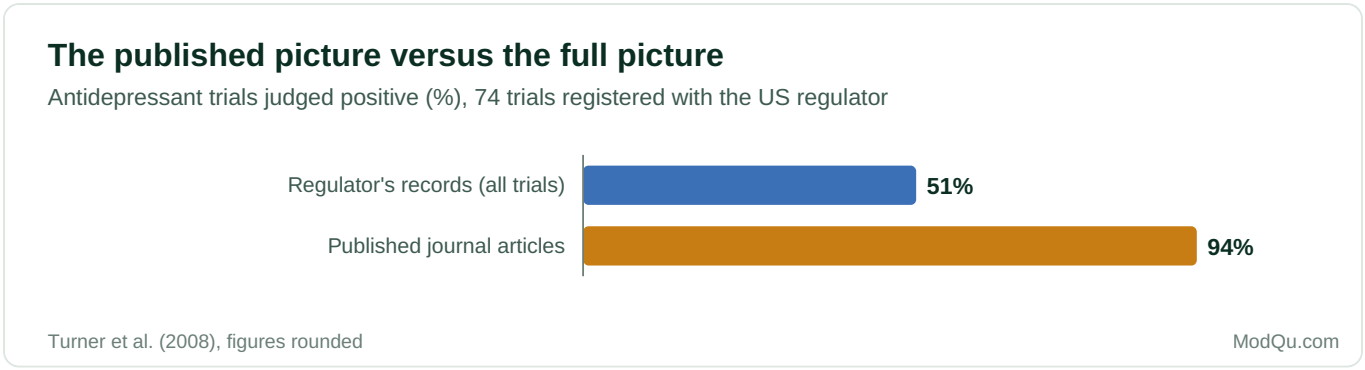


Figure 2. Share of antidepressant trials with positive results in regulatory records and in the published literature. Figures rounded. Data from Turner et al. (2008); chart by ModQu.

Bias within published studies

Publication bias does not only remove whole studies. It also shapes what appears inside published papers.

Chan et al. (2004) compared the protocols of randomised trials with the articles that reported them. In most trials, at least one primary outcome had been changed, added or omitted between protocol and publication, and outcomes with statistically significant results were more likely to be reported in full. Readers of the articles had no way of knowing.

The distribution of reported test statistics shows similar traces. Gerber and Malhotra (2008) found that results in two leading political science journals clustered just above the conventional threshold for significance, more than would be expected by chance. Brodeur et al. (2016) analysed some 50,000 tests published in leading economics journals and found a shortage of results just above the 5 per cent threshold and a surplus just below it, a pattern consistent with researchers adjusting analyses until results crossed the line.

A growing problem?

Fanelli (2010) found that the share of papers reporting support for their hypotheses was higher in the social and behavioural sciences than in the physical sciences. Fanelli (2012) then examined more than 4,600 papers published between 1990 and 2007 across disciplines and countries and found that the share reporting positive results had risen over time. Negative results, he concluded, were disappearing from most disciplines.

How selection inflates effects: an illustration

A simple illustrative calculation shows how publication bias distorts what the published record says. Suppose a marketing intervention has a small true effect, raising purchase rates by one percentage point. Twenty independent teams each run a modest study. Because their samples are small, their estimates vary widely around the true value: some teams find no effect or even a negative one, others find effects of three or four points. Only the studies with the largest estimates reach statistical significance.

If only those significant studies are published, a reader of the literature sees a handful of studies reporting effects of three or four points, all significant, and none reporting smaller effects. A meta-analysis of the published studies would estimate an effect several times larger than the true one, with apparently good precision. Nothing in any individual published study is wrong; the distortion comes entirely from which studies were allowed to appear.

The same process explains a common experience in applied research: an effect that looked large in the first published studies shrinks when larger studies are run or when the intervention is rolled out at scale. Part of the shrinkage reflects real differences between research and practice, but part often reflects the selection of early results that happened to be large.

Differences between fields

Publication bias appears in every field that has been studied, but its form and extent vary. In clinical medicine, the existence of regulatory records and trial registries has made it possible to measure missing studies directly, as Turner et al. (2008) did. In psychology and the social sciences, where studies are cheaper and more numerous, the problem shows up mainly as an excess of significant results and a scarcity of null ones (Fanelli, 2010; Franco et al., 2014). In economics, where researchers often analyse existing data in many possible ways, the evidence points to selective reporting within studies as much as to missing studies (Brodeur et al., 2016; Ioannidis et al., 2017).

Fields also differ in how far their questions allow publication bias to matter. Where theories make precise predictions and results are routinely checked by later work, false findings are corrected relatively quickly. Where effects are small, studies are small and replication is rare, distorted findings can persist for years. Business and management research shares many features of the second kind of field: effects are often modest, contexts vary, and exact replications are uncommon. This makes awareness of publication bias especially important for readers of research on markets, consumers and organisations.

Why publication bias happens

Several forces combine.

Researchers' choices. Most of the evidence suggests that the largest losses occur before submission. Researchers who obtain null results often decide not to write them up, because they expect rejection, because the results seem uninformative, or because their time is better spent on new work (Dickersin, 1990; Franco et al., 2014).

Editorial and reviewer preferences. Journals compete for attention and citations, and positive, novel findings attract both. Reviewers may judge a null result as a sign of a poorly designed study (Greenwald, 1975).

Career incentives. Researchers are rewarded for publications, especially in prominent journals, which encourages a focus on publishable results.

Commercial interests. Sponsors may have reasons to delay or withhold unfavourable results, as the antidepressant example showed.

Ambiguity of null results. A non-significant result can mean that there is no effect or that the study lacked the power to detect one. This ambiguity makes null results genuinely harder to interpret, especially from small studies, and so harder to publish.

Consequences

The effects of publication bias go well beyond academic debate.

Exaggerated effects. When only the studies that happened to find larger effects are published, any review of the published literature overstates the true effect. Ioannidis et al. (2017) estimated that in economics, where the typical study had low power, reported effects were commonly exaggerated, often by a factor of two or more.

False findings. Where many teams test the same false hypothesis, publication bias can create an apparent body of supporting evidence where none exists (Ioannidis, 2005).

Misleading reviews. Systematic reviews and meta-analyses are only as good as the studies they include. A meta-analysis of a biased literature produces a precise but wrong answer.

Wasted resources and harm. Researchers may repeat studies that have already failed, unaware that others tried before. In medicine, the result can be treatments that are less effective or more harmful than the evidence suggests.

Detecting publication bias

Several methods have been developed to detect publication bias in a body of studies, each with limitations.

Detecting publication bias

No method is definitive; use several and interpret with care

Method	What it does	Main limitation
Funnel plot	Plots effect against precision	Judged by eye
Egger's test	Tests funnel asymmetry	Asymmetry has other causes
Trim and fill	Imputes missing studies	Strong assumptions
P-curve	Examines significant p-values	Uses significant results only
Fail-safe N	Counts studies needed to overturn	Now seen as unreliable

Drawing on Sterne et al. (2001), Duval & Tweedie (2000), Simonsohn et al. (2014) ModQu.com

Figure 3. Methods for detecting publication bias and their limitations. Table by ModQu, drawing on Sterne et al. (2001), Duval and Tweedie (2000) and Simonsohn et al. (2014).

Funnel plots. Light and Pillemer (1984) proposed plotting each study's effect estimate against a measure of its size or precision. Without bias, the plot should resemble an inverted funnel: small studies scatter widely around the true effect and large studies cluster closely, symmetrically on both sides. If small studies with unfavourable results are missing, the plot will be asymmetric.

Tests of asymmetry. Egger et al. (1997) proposed a regression test to assess funnel plot asymmetry statistically. Sterne et al. (2001) cautioned that asymmetry can have causes other than publication bias, such as genuine differences between small and large studies or poorer methods in small studies, and that such tests have low power when there are few studies.

Trim and fill. Duval and Tweedie (2000) developed a method that estimates how many studies are missing from the asymmetric side of the funnel, imputes them and recalculates the pooled effect. The method gives an idea of how much the estimate might change, but its assumptions are strong, and it can mislead when asymmetry has other causes.

P-curve. Simonsohn et al. (2014) proposed examining the distribution of statistically significant p-values in a set of studies. If an effect is real, very small p-values should be more common than p-values just below 0.05. If significant results arise mainly from selective reporting, they will cluster near the threshold.

Precision-effect tests. Stanley (2005) developed regression-based methods, widely used in economics, that estimate what the effect would be in a hypothetical study of perfect precision, on the reasoning that the most precise studies are least affected by selection.

Fail-safe N. Rosenthal (1979) also suggested calculating how many unpublished null studies would be needed to overturn a significant pooled result. The method is simple but is now generally regarded as unreliable, because it assumes that missing studies average zero effect and ignores the size of the effect.

No method is definitive. The best practice is to use several, interpret them cautiously and, above all, prevent the bias at source by finding unpublished studies.

Correcting at the review stage

Reviewers can reduce the influence of publication bias by searching beyond published journals. The Cochrane handbook recommends searching trial registers, conference abstracts, theses and other unpublished sources, and contacting researchers in the field (Higgins et al., 2019). Hopewell et al. (2007) found that meta-analyses that excluded grey literature tended to show larger effects than those that included it. Rothstein et al. (2005) set out a range of methods for preventing, assessing and adjusting for publication bias in meta-analysis, emphasising that thorough searching is the first line of defence.

Reforms

The most effective responses change how research is registered and published.

Trial registration. In 2004 the International Committee of Medical Journal Editors announced that its journals would publish clinical trials only if they had been registered in a public registry before patients were enrolled (De Angelis et al., 2004). Registration makes it possible to see which trials were started and to compare their planned outcomes with what was reported. Kaplan and Irvin (2015) found that, among large trials funded by one United States institute, the share reporting a significant benefit fell sharply after registration of outcomes in advance became mandatory, which suggests that earlier results had been inflated by flexible reporting.

Registered Reports. Under this format, journals review a study's question and methods before data are collected and commit to publishing the results regardless of outcome (Chambers, 2013). Scheel et al. (2021) found that Registered Reports in psychology supported their first hypothesis far less often than standard articles, a difference consistent with reduced publication bias.

Journals for null results and replications. Some journals and article types now explicitly welcome null results, replications and negative findings.

Results reporting requirements. Some regulators and funders require that the results of registered trials be posted within a fixed period, whether or not they are published in a journal.

Preprints and repositories. Posting results publicly, even when a journal does not accept them, ensures that they remain available to reviewers.

What individual researchers can do

Reforms to journals and registries matter most, but individual researchers can also reduce the problem.

- **Write up every study.** A short report deposited in a repository or posted as a preprint keeps a null result available to future reviewers.
- **Register plans in advance.** Recording hypotheses and outcomes before data collection makes selective reporting visible and protects the researcher's own credibility.
- **Report all outcomes.** Include every measure and analysis that was planned, and label additional analyses as exploratory.
- **Design studies to be informative whatever the result.** Studies with adequate power give null results that can be interpreted, which makes them easier to publish and more useful.
- **Value null results as a reviewer or editor.** Judge studies on their question and methods, not on whether the result is positive.

Publication bias in business and commercial research

Publication bias is not only an academic problem. Organisations face their own versions of the file drawer. Market tests that fail are rarely written up, case studies of successful initiatives are widely shared while failures are quietly forgotten, and consultancy reports tend to feature the projects that went well. Over time, an organisation's internal evidence base can come to overstate what works, in the same way as an academic literature.

The remedies are similar. Record every test or study at the start, including what it is expected to show. Report the results of all of them, including those that found nothing. When reviewing evidence, whether published research or internal reports, ask what might be missing, and look for it.

Questions to ask of any body of evidence

When reading a review, a literature or a set of internal reports, a few questions help to judge whether publication bias may be distorting the picture:

- **Is there a registry or complete list of studies?** If so, compare it with what was published.
- **Are the studies mostly small, with large effects?** This pattern is a warning sign.
- **Do almost all studies report positive results?** In most fields, a literature with no null results is implausible.
- **Did the review search for unpublished work?** Reviews that searched only published journals are more likely to overstate effects.
- **Did the review test for bias?** Funnel plots and related methods are imperfect but informative.
- **Who funded the studies?** Sponsors with an interest in a particular result may influence what is reported.

Conclusion

Publication bias means that the published record is not a fair sample of the research that has been done. Studies with positive results are more likely to be written up, submitted, published, published quickly and cited. The result is a literature that overstates effects and contains more false findings than it should. Methods such as funnel plots, p-curves and trim and fill can help to detect the problem, but none is definitive. The lasting solutions lie in registering studies in advance, publishing results regardless of outcome and searching thoroughly for what has not been published. For anyone relying on evidence, the essential question is not only what the published studies say, but what the unpublished ones might have said.

References

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star Wars: The empirics strike back. *American Economic Journal: Applied Economics*, *8*(1), 1–32. https://doi.org/10.1257/app.20150044

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at *Cortex*. *Cortex*, *49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Chan, A.-W., Hróbjartsson, A., Haahr, M. T., Gøtzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles. *JAMA*, *291*(20), 2457–2465. https://doi.org/10.1001/jama.291.20.2457

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., Kotzin, S., Laine, C., Marusic, A., Overbeke, A. J. P. M., Schroeder, T. V., Sox, H. C., & Van Der Weyden, M. B. (2004). Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, *351*(12), 1250–1251. https://doi.org/10.1056/NEJMe048225

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *JAMA, 263*(10), 1385–1389. https://doi.org/10.1001/jama.1990.03440100097014

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet, 337*(8746), 867–872. https://doi.org/10.1016/0140-6736(91)90201-Y

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE, 5*(4), Article e10068. https://doi.org/10.1371/journal.pone.0010068

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Gerber, A. S., & Malhotra, N. (2008). Do statistical reporting standards affect what is published? Publication bias in two leading political science journals. *Quarterly Journal of Political Science, 3*(3), 313–326. https://doi.org/10.1561/100.00008024

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82*(1), 1–20. https://doi.org/10.1037/h0076157

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Wiley. https://doi.org/10.1002/9781119536604

Hopewell, S., McDonald, S., Clarke, M. J., & Egger, M. (2007). Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews*, 2007(2), Article MR000010. https://doi.org/10.1002/14651858.MR000010.pub3

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

Ioannidis, J. P. A., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal, 127*(605), F236–F265. https://doi.org/10.1111/ecoj.12461

Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLOS ONE, 10*(8), Article e0132382. https://doi.org/10.1371/journal.pone.0132382

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. https://doi.org/10.1002/0470870168

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science, 4*(2). https://doi.org/10.1177/25152459211007467

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547. https://doi.org/10.1037/a0033242

Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J., Hing, C., Kwok, C. S., Pang, C., & Harvey, I. (2010). Dissemination and publication of research findings: An updated review of related biases. *Health Technology Assessment, 14*(8). https://doi.org/10.3310/hta14080

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys, 19*(3), 309–345. https://doi.org/10.1111/j.0950-0804.2005.00250.x

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

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician, 49*(1), 108–112. https://doi.org/10.1080/00031305.1995.10476125

Sterne, J. A. C., Egger, M., & Davey Smith, G. (2001). Investigating and dealing with publication and other biases in meta-analysis. *BMJ, 323*(7304), 101–105. https://doi.org/10.1136/bmj.323.7304.101

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