Testing and adjusting for publication bias

Anders Pape Møller and Michael D. Jennions

How can the scientific literature provide unbiased conclusions if it represents a biased sample of available studies? Publication bias refers to phenomena arising from bias in submitting, reviewing, accepting and publishing scientific results. Direct and indirect methods for investigating publication bias are now readily available, but indirect methods are generally open to alternative interpretations. Publication bias distorts attempts to review a scientific field quantitatively if the likelihood of locating completed studies depends on the strength or direction of the findings of quantitative studies. It is the responsibility of researchers, reviewers and editors to address issues of bias to ensure the existence of an unbiased literature.

PUBLICATION BIAS (PB; see Glossary) occurs when the published literature does not reflect finished research projects in a particular subject area. If the objective of a science, such as ecology or evolutionary biology, is to make generalizations about the organization of the living world and how this has arisen, then any BIAS in our knowledge can distort our view. More generally, DISSEMINATION BIAS affects our ability to retrieve information on completed studies and review a scientific field whenever the time-lag, journal, language, data base indexing, citation rate, public prominence and even likelihood of publication depend on the strength or direction of scientific findings[1–4].

Two aspects of PB with special relevance to biology are choice of research topics and sample sizes. First, most biologists live and work in temperate regions, whereas most species are tropical; most animals are beetles, but field biologists are usually ornithologists. Published studies therefore reflect taxonomic diversity inaccurately. Second, META-ANALYSIS involves weighting EFFECT SIZE by sample size. In ecology and evolutionary biology, however, sample sizes might not be random with respect to the taxonomy or the biology of the species. Large sample sizes are usually associated with naturally common, smaller-bodied species, such as insects. Species abundance varies with latitude and habitat type, rarity and body size are also correlated with many aspects of life history, physiology and behaviour. Although we do not consider these potentially confounding correlations further in this review, they affect both narrative and meta-analysis reviews and do deserve greater attention.

There has been a recent increase in the application of quantitative assessment of scientific publications through meta-analysis[5–7]. Since 1976 (Ref. 8), this has become the main method of assessing a body of literature in the social and medical sciences.

Meta-analysis summarizes a body of literature on a subject by transforming test statistics into a common metric (effect size, which is a standardized measure of the strength of a relationship between two variables of interest), and these quantitative data are subsequently analysed. Meta-analysis has an advantage over simple verbal reviews because unbiased estimates of the overall strength of a hypothetical relationship can be obtained, heterogeneity in research findings identified and new research questions developed. Meta-analysis results in an increase in knowledge. Although only recently introduced to ecology[7], an increase in the use of meta-analysis in this field is clear from bibliographic data bases [~five meta-analyses (1996); six (1997); eight (1998); 19 (1999)].

What is publication bias?

Meta-analysis is based on the assumption that the literature reviewed is unbiased[5,6]. If the total number of studies that have been conducted as the source of all publications is considered, only some studies will be written up and submitted for publication, a fraction of those will be submitted by editors for refereeing, a fraction of those will be recommended by referees for publication, and a fraction of those will be published. Publication bias occurs whenever the strength or direction of the results of published and unpublished studies differ. Although experimental quality, methodological rigour, level of replication (and social factors, such as author or institute prestige) can increase the probability of publication, they do not necessarily give rise to PB. There will be no bias if published articles are a random sample of the original source pool of research with respect to research findings. Given the various potential biases, authors, reviewers and editors could all be responsible for creating a biased scientific literature. Whether that is the case is the topic of this review. Unfortunately there are little specific data on PB in ecological or evolutionary research, and we therefore draw our examples primarily from the medical literature.

Since 1979 (Ref. 9), the meta-analysis literature in medical and social sciences is replete with discussions of bias from the level of choice of research topics, to submission, publication and citation. Indeed, several meta-analyses have addressed these issues and even attempted to assess the potential effect[10]. By contrast, not one narrative review in the medical sciences

Anders Pape Møller*
Laboratoire d'Ecologie Evolutive Parasitaire, CNRS FRE 2365, Université Pierre et Marie Curie, Bât. A, 7ème étage, 7 quai St. Bernard, Case 237, F-75252 Paris Cedex 5, France.
*e-mail: amoller@ hall.snv.jussieu.fr

Michael D. Jennions
Division of Botany and Zoology, Australian National University, Canberra, ACT 0200, Australia.

Box 1. Fail-safe numbers

Fail-safe or file-drawer methods exploit the fact that an average effect size, adjusted for sample size, has a given associated significance level. A larger number of unpublished null results would be needed to nullify the significance of an analysis with a highly significant mean effect. Rosenthal[a] recommended the use of Eqn I to estimate the fail-safe number of unpublished results (X):

$$X = (\Sigma\, Z_j)^2/2.706 - K \qquad\qquad [I]$$

where $Z_j = Z_{rj}\sqrt{(N_j - 3)}$, Z_{rj} is Fisher's z-transformed correlation coefficient for the relationship between the two variables of interest for sample j, N_j is sample size for sample j and K is the number of analysis units (depending on the level of analysis, this can be the number of samples, studies or species). The value 2.706 is based on a one-tailed P value of 0.05. The fail-safe number estimates the number of studies that are unknown to us as a result of either journals rejecting papers with null results or scientists not writing up null results. A different way of viewing the fail-safe number is that it provides an estimate of the number of future studies needed to change a significant effect to a nonsignificant one. A fail-safe number of $5K + 10$ was considered to provide evidence of a robust average effect size[a].

In a study of survivorship of males in relation to expression of secondary sexual characters, Jennions *et al.*[b] found an average Pearson product-moment correlation coefficient adjusted for sample size of 0.125 based on studies of 40 species. Entering these values in the equation gives a fail-safe number larger than 1200, which by far exceeds $5 \times 40 + 10 = 210$.

References

a Rosenthal, R. (1991) *Meta-analytic Procedures for Social Research*, Sage
b Jennions, M.J. *et al.* (2001) Sexually selected traits and adult survival: a meta-analysis. *Q. Rev. Biol.* 76, 3–36

mentioned PB. Examination of the past decade of *Annual Reviews of Ecology and Systematics, Biological Reviews* and *Quarterly Review of Biology* leaves the same impression: narrative reviews rarely address whether the literature is biased. When they do, the effect of PB is never quantified. Hence, meta-analysis has yet another advantage over simple narrative reviews by quantitatively assessing potential biases.

Recently, several studies in ecology and evolution have suggested that PB might have tainted our view of a given field[11–16]. There are at least three levels in which PB occurs: submission, review and editorial decisions.

Submission bias

SUBMISSION BIAS has been studied little because it requires information on the source population of the research undertaken. A typical method to investigate this bias is to contact prospective authors and members of academic societies to obtain a sample of potential papers. For example, a survey of 1000 psychologists showed that authors were more likely to submit studies that refuted a null hypothesis than those studies that did not (66% versus 22%)[1]. Many other studies have found similar or even higher rates of lack of submission because investigators did not consider their results to be interesting[2] or because the results were 'inexplicable' (i.e. inconsistent with mainstream theory).

Although these studies suggest that submission bias related to the strength of research findings is sometimes important, we cannot be sure. The data are often based on self-reporting, and the investigators surveyed are usually selected nonrandomly. Even so, investigators do appear to be a major source of PB.

Editorial bias

Disappointed authors often blame editors, or explicit or implicit journal policies, as a key cause of PB (Ref. 3). Editors have to present original and interesting findings to the readership, so old, well-known null results rarely get prominent coverage. Authors know this: the presence of the word 'novel' in titles and abstracts has increased exponentially in recent years[17]! A questionnaire study of 429 editors or editorial board members for 19 leading journals in management showed that the likelihood of acceptance was strongly increased by the reputation of the author, successful testing of the author's new theory, and originality. Nonsignificant results, replication and lack of novelty reduced the probability of publication[18]. In biology, replication of a study is so broadly defined that testing for the same relationship in another species or phylum can still be classified as replication. These attitudes strongly discourage authors from replicating studies.

Editorial policies that require greater statistical power when results are not significant can also bias the literature because increasing sample sizes delays publication. It might even discourage researchers from tackling certain topics if they suspect that the null hypothesis is correct and previously published statistically significant findings are simply a result of chance. Finally, editorial bias can also arise when referees of manuscripts are not chosen blindly with respect to author affiliation, institutional prestige or research findings.

Reviewer bias

Reviewer bias based on author prestige, nationality or gender could influence which manuscripts get published and in which journals. Several studies investigating peer-reviewer effects on PB show that inter-reviewer agreement is generally poor and biased by the reviewers' own research findings[19,20]. The prestige of the institution of the authors has been shown to be linked to the probability of recommendation for acceptance of brief reports, but not for full papers, in the *Journal of Pediatrics*[21].

Link[22] found that US reviewers of manuscripts for a US medical journal were more likely to recommend acceptance of US than they did nonUS manuscripts. US reviewers ranked US papers much more favourably than nonUS papers, whereas the difference for nonUS reviewers was nonsignificant. In

Box 2. The funnel graph

Fig. I (a) (b) (c)

TRENDS in Ecology & Evolution
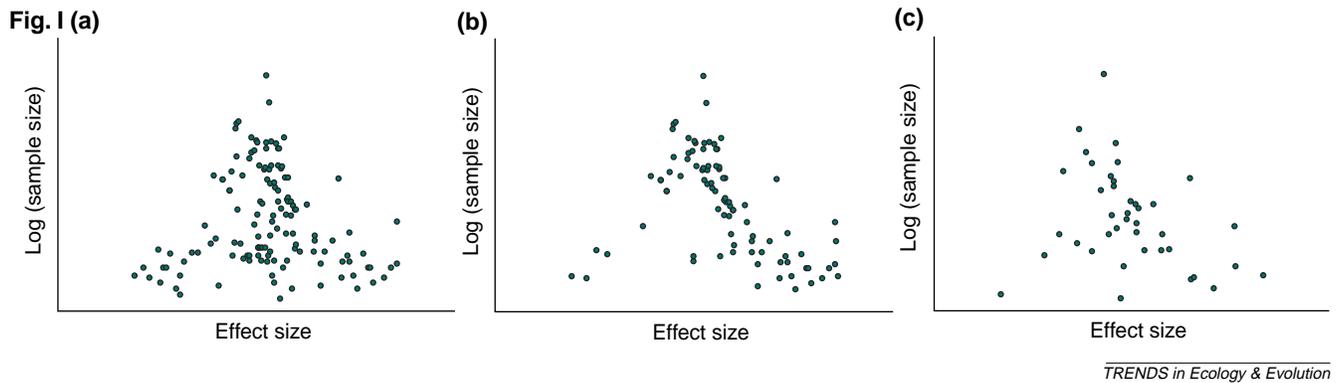
If effect sizes derive from a random sample of studies using similar research methods, a plot of sample size (or the standard error of the effect size) against effect size should reveal a funnel, with larger variance in values at small sample sizes and a decreasing variance with increasing sample size[a] (Fig. Ia). If the true effect differs significantly from zero, but is small to intermediate, and publication is related to statistical significance (unless sample size is large), a decrease in the effect sizes of studies with increasing sample size should be expected (Fig. Ib). This occurs because the smaller the sample size the greater the estimated effect sizes must differ from zero to be statistically significant. So, if the true effect is moderately greater than zero, for smaller sample sizes only those estimates that, owing to sampling error, are larger than the true mean effect are likely to reach significance. Studies to the lower left or centre of Fig. Ib therefore go unpublished, except for the occasional study that is significantly smaller than zero owing to sampling error.

A funnel plot of the study of survivorship of males and expression of secondary sexual characters[b] showed a significant decrease in the variance of effect sizes as sample size increased. Using samples (or species) as the units of analysis did, however, suggest a slight skew towards more positive effect sizes when sample size was low (Fig. Ic).

References
a Light, R.J. and Pillemer, D.B. (1984) *Summing Up: the Science of Reviewing Research*, Harvard University Press
b Jennions, M.J. *et al.* (2001) Sexually selected traits and adult survival: a meta-analysis. *Q. Rev. Biol.* 76, 3–36

a second study, Nylenna *et al.*[23] submitted two fictitious manuscripts for review to Scandinavians: one in English and the other in a Scandinavian language. The English manuscript was awarded consistently higher quality scores than was the nonEnglish manuscript, although the contents were identical. Again, however, PB will only emerge if factors that influence a reviewer's decisions are correlated with the strength of study findings.

Controlled investigation attempting to hide author identity from reviewers provided benefits in terms of paper quality in some[24–26], but not all cases[27,28]. For example, articles from economics journals using double blind peer review were cited more often than were articles from journals that only hide the reviewers' identity[25]. This failure might simply reflect the tendency for reviewers to guess author identity. Furthermore, unless both editors and reviewers are unaware of the authors' names and institutions, the process is not truly anonymous. To avoid reviewer bias, some medical journals require reviewers to reveal their identity. This should reduce bias because reviewers are less likely to express unsubstantiated prejudices[29] (but see Ref. 28). Obviously, revealing your identity as a reviewer could have other consequences, such as affecting the possibility of getting funding or acquiring a job, if a senior author can influence such decisions. The usefulness of different peer-review approaches in ecology and evolutionary biology deserves to be held to referenda by members of academic societies.

The written word: is there direct evidence for publication bias?
Statistically significant results in the literature seem to be more common than would be expected by chance[16]. Csada *et al.*[15] reported that 91% of 1201 papers in 43 biology journals had statistically significant supportive tests. Many similar results have been reported from the medical literature (Ref. 4). Direct comparison of published and unpublished studies is even more informative. For example, Dickersin *et al.*[30] found that 55% of 767 published medical studies showed a new treatment was better than a control, whereas only 14% of 178 unpublished studies found this effect. However, in two biological meta-analyses concerning sexual selection, neither Thornhill *et al.*[31] nor Jennions *et al.*[32] found a significant difference in effect size adjusted for sample size between published and unpublished studies. In both these studies, extensive attempts to locate unpublished studies were made through contacts with scientists in the field in question. One caveat is that studies might be unpublished simply because they have only recently been completed.

<div style="background:yellow">

Box 3. Rank correlation analysis of Begg and Mazumdar

The rank correlation test of Begg and Mazumdar[a] uses the Spearman rank correlation to investigate the relationship between standardized effect size and sample size or variance in effect size. The significance of the test can be found in standard statistical tables. In the study[b] described in Box 2, there was no obvious heterogeneity in the funnel plot, using a graphical inspection. Applying the rank correlation test to the data revealed, however, that effect size decreased as sample size increased at both the study and species level of analysis (both: Spearman's $r = -0.317$, $P < 0.05$). There were fewer than expected studies with negative effects at low sample sizes. Most biological meta-analyses appear to report the simple correlation between sample size and effect size without first standardizing effect size as recommended by Begg and Mazumdar. This could change conclusions. For the study[b] described in Box 2 the nonstandardized correlations initially reported were Spearman's $r = -0.35$.

A linear regression approach to address the same question was reported by Allison *et al.*[c] This test is statistically more powerful, but requires several assumptions about the distribution of the data that might be unwarranted, for example, homoscedacity.

References
a Begg, C.B. and Mazumdar, M. (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50, 1088–1101
b Jennions, M.J. *et al.* (2001) Sexually selected traits and adult survival: a meta-analysis. *Q. Rev. Biol.* 76, 3–36
c Allison, D.B. *et al.* (1996) Publication bias in obesity treatment trials. *Int. J. Obes.* 20, 445–447

</div>

Although a high proportion of significant results in published studies might indicate bias, this is not necessarily so if the hypotheses under test are not randomly formulated[33]. Indeed, biologists often perform empirical tests to 'confirm' established theoretical models. Even so, in evolutionary and ecological studies, effect sizes are usually small (for 44 published meta-analyses, the mean variance explained is ±5%; A.P. Møller and M.D. Jennions, unpublished). Power analysis shows that, given small effect sizes ($r < 0.2$) and the characteristically small sample sizes of field biology, a far greater proportion of published studies should report nonsignificant results if they represent an unbiased sample.

The strongest case for PB is made when the publication fate of a group of preregistered studies is followed. In medicine, all available analyses show that studies producing significant results ($P < 0.05$) are more likely to be published than are those with nonsignificant results[4]. Interestingly, studies with a weak but nonsignificant trend are even less likely to be published[34]. Positive results confirming a hypothesis with a significant result often get priority in publication[2,34]. In one study, there was a median of 4.8 years to publication for significant positive results, but eight years for negative results[34]. This should lead to a decline in effect size over time, as seen in three recent biological meta-analyses[12–14] (although alternative explanations also exist, e.g. studies with weaker effects take longer to be published).

Dissemination bias: not all publications are equal

Electronic searches are biased when the strength of the findings of located and unlocated publications differs. For example, meta-analyses often rely on searching with language restrictions. Gregoire *et al.*[35] showed than 28 out of 36 meta-analyses in medical journals had language restrictions, and at least two conclusions would have been altered by inclusion of nonEnglish literature. This is potentially a problem because nonEnglish-speaking authors might be more likely to publish nonsignificant results in their native language and significant findings in English[3]. More generally, nonsignificant results are published in journals with smaller circulation because of the previously described submission, review and editorial bias. These journals are less likely to be indexed in electronic data bases.

When compiling data for a meta-analysis, examination of reference lists of relevant papers is a standard practice. This could also produce a bias because biased use of available references is common. There is, for example, a bizarre finding that the alphabetical position of surnames is correlated with citation rate[36]. More importantly, a study of a randomly selected citations from 283 research articles in psychology journals revealed that authors most commonly used citations to support their own argument, whereas study quality was only important for 2% of the citations[37]. A study of cholesterol-lowering medical trials revealed that supportive trials had a mean annual number of citations of 40, whereas other trials only had 7.4 citations[38], independent of journal of publication and sample size of trials. Nationality of review authors might seriously bias citations, as shown in a study of reviews on chronic fatigue syndrome[39]. Among papers cited by authors from the USA, 72% were from US journals, whereas 55% of citations by British authors were from British journals.

Editors generally favour original and novel findings, and this decision might bias our view of how science normally functions. When last was a null result the cover story in *Nature* or in *Science*? A similar hunger for reporting positive and novel studies is common in the news media, including general review journals, such as *TREE*. Newspaper coverage of a positive and a negative study on risk of cancer in relation to radiation published together in *JAMA* in 1991 showed that nine of 19 articles only covered the positive result. In the other ten articles, an average of 354 words was used for the positive study and 192 for the negative study[40]. Similarly, conference talks usually focus on significant findings; therefore, scientists are more aware of these studies.

Ultimately, dissemination bias is only a dilemma when literature reviewers correlate the strength of research findings with the rate of encounter. The available evidence suggests that studies with nonsignificant effects are encountered less often than are studies with significant effects. They are more

---

Box 4. The 'trim and fill' method of testing and adjusting for publication bias

The 'trim and fill' method of Duval and Tweedie[a] is a nonparametric method testing and adjusting for publication bias (PB). The only assumption is a reliance on the symmetric distribution of effect sizes around the 'true' effect size in a funnel plot. The method assumes that it is the most extreme negative results that go unpublished. To find these studies, asymmetry in the funnel plot is detected. The extreme positive studies responsible for the asymmetry are then 'trimmed' off. An initial estimate of the mean effect size is then based on the symmetrical remainder. The original full data set is then re-examined to detect asymmetry around the new estimate of the mean effect size. Additional iterations are made until the result converges on the 'true' effect size. The missing, extreme negative results are then added by using the mirror image of the trimmed positive results, and this new distribution is used to derive final estimates of the mean effect size and its variance. Duval and Tweedie[a] used three different estimators for the number of missing studies.

In the study of survivorship of males and expression of secondary sexual characters by Jennions *et al.*[b] the number of missing studies can be obtained based on the method of Duval and Tweedie[a]. At the species level of analysis, the R and L estimators of the method indicated two and 11 missing studies for fixed-effects meta-analysis, but none for a random-effects meta-analysis. The mean effect size after trimming and filling for two or 11 missing studies respectively was $r = 0.118$ (95% CI 0.090–0.146) and $r = 0.072$ (95% CI 0.046–0.099) (both $P < 0.05$) compared with an original estimate of 0.125 (95% CI 0.098–0.152). The conclusion is therefore robust even if it is assumed that 11 studies are missing owing to PB.

Several other methods (e.g. weighted distribution theory, general linear models and Bayesian modelling) have been derived to model PB (Ref. c). Unfortunately, these are not implemented in available software, and require considerable skills in statistical modelling. At present, 'trim and fill' appears to be the most parsimonious (and easiest) method with which to estimate the number of missing studies.

References
a Duvall, S. and Tweedie, R. (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56, 455–463
b Jennions, M.J. *et al.* (2001) Sexually selected traits and adult survival: a meta-analysis. *Q. Rev. Biol.* 76, 3–36
c Song, F. *et al.* (2000) Publication and related biases. *Health Technol. Assess.* 4, 1–115

---

likely to be unpublished, published in low-circulation journals and published in languages other than English; and are less likely to be cited and talked or written about.

How to assess publication bias? The use of indirect methods

PB can be assessed directly in a comparison of published and unpublished studies, or by following a cohort of studies from their inception. This is, however, a logistical challenge. Consequently, at least four types of indirect methods have been proposed to quantify PB. Unfortunately, the findings from these methods might have plausible alternative explanations that are unrelated to PB. Conclusions should thus be made cautiously[31,32,41]. There are currently no comparative analyses of the efficiency of different methods to identify PB and it is good practice to use all of them. Consistency in findings provides greater confidence.

The fail-safe number of unpublished studies

Among several available methods, calculating Rosenthal's FAIL-SAFE NUMBER is the best known[42]. It estimates the number of unpublished studies, with a mean effect of zero, required to eliminate a significant overall effect size (Box 1). If the number of such studies is extremely large, it is unlikely that PB can alter the main conclusion of a meta-analysis regarding the significance of an effect[43,44].

The use of the fail-safe method is problematic because of its emphasis on statistical rather than on biological significance. More importantly, if studies go unlocated because they reported results contrary to those in located studies, the 'true' fail-safe number is smaller than that based on an average effect of zero in unlocated studies.

The funnel graph

As a result of sampling error, small studies generate a greater range of findings than do larger studies, resulting in a funnel-shaped relationship between effect size and sample size[45]. If the probability of publication is greater for studies with statistically significant results, this skews or 'hollows out' the funnel shape of the graph. If the true effect is small or moderate, but non-zero, and nonsignificant results tend not to be published, this creates a decrease in effect size as sample size increases (Box 2). However, a skewed FUNNEL PLOT can be caused by factors other than PB because previous knowledge of effect sizes from pilot studies, reduced sample sizes for certain species, choice of effect measures, chance and many other factors might cause asymmetric plots[31,46].

Effect size and sample size

Several biological meta-analyses have shown significant relationships between sample size and effect size[12,16,32]. If there is PB towards significant results, and the true effect size is small, but non-zero, effect size is expected to weaken as sample size increases. Begg and Mazumdar[47] proposed a distribution-free, rank correlation test to investigate this bias. Simulations show that the test is fairly powerful for large meta-analyses with 75 studies, but only moderately so with 25 studies, requiring caution of interpretation (Box 3).

Glossary

Bias: systematic deviation of results from the truth, or processes leading to such deviation.

Dissemination bias: bias in the performance of research, publication, interpretation and review of scientific findings[a]. The accessibility of research findings to potential users depends on the direction or strength of these findings[b].

Effect size: a statistical measure of the magnitude of a relationship between two variables of interest. Typical measures of effect size in meta-analyses are the Pearson correlation coefficient, Hedges' *d* (a measure of effect size in terms of normal standard deviates) or the odds ratio[a,d].

Fail-safe number: The number of null results with a mean effect size of zero needed to change a given mean effect size adjusted for sample size into a nonsignificant result[d–g].

Funnel plot: A graph illustrating the relationship between effect size and its variance. Traditionally, sample size or the standard error of the effect size is plotted against effect size.

Meta-analysis: A group of methodological techniques developed to synthesize a body of literature in a standardized way: (1) to obtain an estimate of the overall strength of a relationship; (2) to test for robustness of a common research finding; (3) to determine heterogeneity and its causes in a body of literature; and (4) to test for and investigate biases[c,d,h].

Publication bias: bias owing to the influence of research findings on submission, review and editorial decisions[a,b,i].

Submission bias: bias owing to the influence of research findings on the probability of submission[b].

References
a Chalmers, I. (1990) Under-reporting research is scientific misconduct. *JAMA* 263, 1405–1408
b Song, F. *et al.* (2000) Publication and related biases. *Health Technol. Assess.* 4, 1–115
c Hedges, L.V. and Olkin, I. (1985) *Statistical Methods for Meta-analysis*, Academic Press
d Rosenthal, R. (1991) *Meta-analytic Procedures for Social Research*, Sage
e Rosenthal, R. (1979) The 'file drawer problem' and tolerance for null results. *Psychol. Bull.* 86, 638–641
f Gleser, L.J. and Olkin, I. (1996) Models for estimating the number of unpublished studies. *Stat. Med.* 15, 2493–2507
g Gurevitch, J. and Hedges, L.V. (1999) Statistical issues in ecological meta-analyses. *Ecology* 80, 1142–1149
h Arnqvist, G. and Wooster, D. (1995) Meta-analysis: synthesizing research findings in ecology and evolution. *Trends Ecol. Evol.* 10, 236–240
i Dickersin, K. (1990) The existence of publication bias and risk factors for its occurrence. *JAMA* 263, 1385–1389

The trim and fill method

The features of the funnel plot have been used to derive a nonparametric method of testing and adjusting for PB in meta-analysis[48]. This provides an estimate of the number of 'missing' studies owing to PB. Furthermore, the funnel plot is then used to obtain an estimate of effect size once these 'missing' studies are added (Box 4). This simple iterative technique is based on estimating the number of studies in the asymmetrical, outlying part of the plot and then 'trimming' them off. The remaining studies are then used to estimate more accurately the 'true' mean effect size. The full data set is then used to re-estimate the number of 'missing' studies based on asymmetry around the new estimate of mean effect size.

After just a few iterations, a robust estimate of the number of 'missing' studies and the mean effect size is obtained. Finally, the variance in effect size is estimated from all available studies and the 'filled in' missing studies. The 'filled in' missing studies are simply the mirror-image counterparts of the trimmed-off studies with respect to the final estimated true mean effect.

How can publication bias be avoided?

What can be done to avoid the problem of PB? First, studies of PB might be subject to the same problems of publication bias as are other studies. Hence, indirect evidence of PB should be interpreted cautiously. As Johnston and Breimer[49] suggest: would studies of publication bias be published if no significant difference was found?

Meta-analysis might provide a powerful tool for identifying PB. Although methods exist to correct estimates of effect size if PB has been suggested, most assessments of potential bias are indirect. Alternative interpretations prevail so these methods are not particularly good 'remedies for publication bias'[50]. Instead, attempts to identify or adjust for PB should be used to investigate the robustness of conclusions, as is done in sensitivity analysis[10,47].

Current knowledge of the causes of PB suggests that investigators are the main cause of bias because they fail to write and submit (or re-submit) nonsignificant findings. Hopefully, with the increase in online journals, reduced competition for space and lower costs of publication will ensure the eventual publication of all methodologically sound studies regardless of their actual research findings[51,52].

Conclusions

Although problems of PB have been investigated in the medical and social sciences for more than two decades, preliminary studies have only recently appeared in ecology and evolution[11–16]. To assess the literature, more information about potential biases is needed, in particular, more direct tests of PB comparing what has actually been published with all studies completed. Calls in society journals for members to register studies online would be a useful starting point. The fate of studies funded by major granting bodies could also be examined.

The problem of PB puts responsibility on the shoulders of editors, peer reviewers and authors to allow fair publication of high-quality research, independent of research findings. Publication ethics deal with proper methods for analysing and presenting information in the scientific literature[53]. Based on the literature in medical sciences, it is potential authors in particular who should recognize their responsibility to publish their findings. Lack of publication by potential authors is even considered to be scientific misconduct by some scientists[49,54].

Finally, when making reviews, authors should not only search electronic data bases, but also check references and contact experts in the field. Although controversial in some scientific circles, unpublished studies should be included to assess any PB. Although they can be unreliable guides to the existence of genuine PB, indirect methods do test the robustness of conclusions from meta-analyses and should always be used to test for PB.

References

1 Coursol, A. and Wagner, E.E. (1986) Effect of positive findings on submission and acceptance rates: a note on meta-analysis bias. *Prof. Psychol.* 17, 136–137

2 Dickersin, K. and Min, Y.I. (1993) Publication bias: the problem that won't go away. *Ann. New York Acad. Sci.* 703, 135–146

3 Sharp, D.W. (1990) What can and should be done to reduce publication bias? The perspective of an editor. *JAMA* 263, 1390–1391

4 Song, F. *et al.* (2000) Publication and related biases. *Health Technol. Assess.* 4, 1–115

5 Hedges, L.V. and Olkin, I. (1985) *Statistical Methods for Meta-analysis*, Academic Press

6 Rosenthal, R. (1991) *Meta-analytic Procedures for Social Research*, Sage

7 Arnqvist, G. and Wooster, D. (1995) Meta-analysis: synthesizing research findings in ecology and evolution. *Trends Ecol. Evol.* 10, 236–240

8 Glass, G.V. (1976) Primary, secondary, and meta-analysis of research. *Educat. Res.* 5, 3–8

9 Boisen, E. (1979) Testicular size and shape of 47,XYY and 47,XXY men in a double-blind, double-matched population survey. *Am. J. Hum. Genet.* 31, 697–703

10 Song, F. and Gilbody, S. (1998) Increase in studies of publication bias coincided with increasing use of meta-analysis. *Br. Med. J.* 316, 471

11 Alatalo, R.V. *et al.* (1997) Heritabilities and paradigm shifts. *Nature* 385, 402–403

12 Gontard-Danek, M-C. and Møller, A.P. (1999) The strength of sexual selection: meta-analysis of bird studies. *Behav. Ecol.* 10, 476–486

13 Simmons, L.W. *et al.* (1999) Fluctuating paradigm. *Proc. R. Soc. London B Biol. Sci.* 266, 593–595

14 Poulin, R. (2000) Manipulation of host behaviour by parasites: a weakening paradigm? *Proc. R. Soc. London B Biol. Sci.* 267, 787–792

15 Csada, R.D. *et al.* (1996) The 'file drawer problem' of non-significant results: does it apply to biological research? *Oikos* 76, 591–593

16 Palmer, A.R. (2000) Quasireplication and the contract of error: lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annu. Rev. Ecol. Syst.* 31, 441–480

17 Friedmann, S.H. and Karlsson, J.O.M. (1997) A novel paradigm. *Nature* 385, 480

18 Kerr, S. *et al.* (1977) Manuscript characteristics which influence acceptance for management and social science journals. *Acad. Manage. J.* 20, 132–141

19 Mahoney, M.J. (1977) Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cognit. Theor. Res.* 1, 161–175

20 Ernst, R. and Resch, K.L. (1994) Reviewer bias – a blinded experimental study. *J. Lab. Clin. Med.* 124, 178–182

21 Garfunkel, J.M. *et al.* (1994) Effect of institutional prestige on reviewers' recommendations and editorial decisions. *JAMA* 272, 137–138

22 Link, A.M. (1998) US and non-US submissions: an analysis of reviewer bias. *JAMA* 280, 246–247

23 Nylenna, M. *et al.* (1994) Multiple blinded reviews of the same two manuscripts: effects of referee characteristics and publication language. *JAMA* 272, 149–151

24 Fisher, M. *et al.* (1994) The effects of blinding on acceptance of research papers by peer review. *JAMA* 272, 143–146

25 Laband, D.N. and Piette, M.J. (1994) A citation analysis of the impact of blinded peer review. *JAMA* 272, 147–149

26 McNutt, R.A. *et al.* (1990) The effect of blinding on the quality of peer-review. A randomised trial. *JAMA* 263, 1371–1376

27 Godlee, F. *et al.* (1998) Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA* 280, 237–240

28 Rooyen, S.V. *et al.* (1998) Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA* 280, 234–237

29 Walsh, E. *et al.* (2000) Open peer review: a randomised control trial. *Br. J. Psychiatry* 176, 47–51

30 Dickersin, K. *et al.* (1987) Publication bias and clinical trials. *Cont. Clin. Trials* 8, 343–353

31 Thornhill, R. *et al.* (1999) The biological significance of fluctuating asymmetry and sexual selection: A reply to Palmer. *Am. Nat.* 154, 234–241

32 Jennions, M.J. *et al.* (2001) Sexually selected traits and adult survival: a meta-analysis. *Q. Rev. Biol.* 76, 3–36

33 Bauchau, V. (1997) Is there a 'file drawer problem' in biological research? *Oikos* 79, 407–409

34 Stern, J.M. and Simes, R.J. (1997) Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *Br. Med. J.* 315, 640–645

35 Gregoire, G. *et al.* (1995) Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J. Clin. Epidemiol.* 48, 159–163

36 Trengenza, T. (1997) Darwin a better name than Wallace? *Nature* 385, 480

37 Shadish, W.R. *et al.* (1995) Author judgements about works they cite: three studies from psychology journals. *Soc. Stud. Sci.* 25, 477–498

38 Ravnskov, U. (1992) Cholesterol lowering trials in coronary heart disease: frequency of citation and outcome. *Br. Med. J.* 305, 15–19

39 Joyce, J. *et al.* (1998) Reviewing the reviews. The example of chronic fatigue syndrome. *JAMA* 280, 264–266

40 Koren, G. and Klein, N. (1991) Bias against negative studies in newspaper reports of medical research. *JAMA* 266, 1824–1826

41 Egger, M. *et al.* (1997) Language bias in randomised controlled trials published in English and German. *Lancet* 350, 326–329

42 Rosenthal, R. (1979) The 'file drawer problem' and tolerance for null results. *Psychol. Bull.* 86, 638–641

43 Gleser, L.J. and Olkin, I. (1996) Models for estimating the number of unpublished studies. *Stat. Med.* 15, 2493–2507

44 Gurevitch, J. and Hedges, L.V. (1999) Statistical issues in ecological meta-analyses. *Ecology* 80, 1142–1149

45 Light, R.J. and Pillemer, D.B. (1984) *Summing Up: The Science of Reviewing Research*, Harvard University Press

46 Egger, M. *et al.* (1997) Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.* 315, 629–634

47 Begg, C.B. and Mazumdar, M. (1994) Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50, 1088–1101

48 Duvall, S. and Tweedie, R. (2000) Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56, 455–463

49 Johnston, J.D. and Breimer, L.H. (1991) Publication bias. *Br. Med. J.* 337, 1102

50 Dickersin, K. (1990) The existence of publication bias and risk factors for its occurrence. *JAMA* 263, 1385–1389

51 Berlin, J.A. (1992) Will publication bias vanish in the age of online journals? *Online J. Curr. Clin. Trials* 1992 Jul. 8, Doc. No. 12 http://govt.ucsd.edu/newjour/o/msg01600.html

52 Song, F. *et al.* (1999) The role of electronic journals in reducing publication bias. *Med. Inform. Internet Med.* 24, 223–229

53 Meslin, E.M. (1994) Toward an ethic in dissemination of new knowledge in primary care research. In *Research Methods for Primary Care* (Dunn, E.V., ed.), pp. 32–44, Sage

54 Chalmers, I. (1990) Under-reporting research is scientific misconduct. *JAMA* 263, 1405–1408

Articles of interest

Articles of ecological or evolutionary interest in recent issues of other *Trends* journals

Memory consolidation of Pavlovian fear conditioning: a cellular and molecular perspective, *G.E. Schafe, K. Nader, H.T. Blair and J.E. LeDoux* **Trends in Neurosciences** 24, 540–546 (September 2001)

Mechanisms of plant desiccation tolerance, *F.A. Hoekstra, E.A. Golovina and J. Buitink* **Trends in Plant Science** 6, 431–438 (September 2001)

Intracellular protozoan parasites and apoptosis: diverse strategies to modulate host–parasite interaction, *C.G.K. Luder, U. Gross and M.F. Lopes* **Trends in Parasitology** 17, 480–486 (October 2001)

Aspartyl proteinase genes from apicomplexan pararsites: evidence for the evolution of the gene structure, *L. Jean, M. Long, J. Young, P. Péry and F. Tomley* **Trends in Parasitology** 16, 491–498 (October 2001)