THE ROYAL SOCIETY

Received 7 April 2001
Accepted 9 August 2001
Published online 4 December 2001

Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution

Michael D. Jennions[1,2*] and Anders P. Møller[3]

[1]*School of Botany and Zoology, Australian National University, Canberra, ACT 0200, Australia*
[2]*Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002-0948, USA*
[3]*Laboratoire d'Ecologie Evolutive Parasitaire, CNRS FRE 2365, Université Pierre et Marie Curie, F-75252 Paris Cedex 5, France*

Both significant positive and negative relationships between the magnitude of research findings (their 'effect size') and their year of publication have been reported in a few areas of biology. These trends have been attributed to Kuhnian paradigm shifts, scientific fads and bias in the choice of study systems. Here we test whether or not these isolated cases reflect a more general trend. We examined the relationship using effect sizes extracted from 44 peer-reviewed meta-analyses covering a wide range of topics in ecological and evolutionary biology. On average, there was a small but significant decline in effect size with year of publication. For the original empirical studies there was also a significant decrease in effect size as sample size increased. However, the effect of year of publication remained even after we controlled for sampling effort. Although these results have several possible explanations, it is suggested that a publication bias against non-significant or weaker findings offers the most parsimonious explanation. As in the medical sciences, non-significant results may take longer to publish and studies with both small sample sizes and non-significant results may be less likely to be published.

Keywords: ecology; evolution; meta-analysis; paradigms; publication bias

1. INTRODUCTION

The basis for scientific investigations is to make generalizations about relationships, and to identify their limits (Hempel & Hempel 1966; Chalmers 1999). Review papers, meta-analyses and handbooks that summarize the available literature are thus invaluable because they serve as sources of generalized knowledge. However, these almost self-evident statements only apply if the literature being reviewed is unbiased. Numerous studies of scientific bias suggest that anything from language to professional seniority may bias what is assessed as the general state of knowledge. More seriously, similar biases may determine what information enters the scientific literature following peer review. Studies, mainly in the social and medical sciences, have shown that publication bias with respect to language, institution, strength of findings, and congruence of research findings with ruling paradigms may bias what is being published (Song *et al.* 2000). A first step towards rectifying such bias is to determine how severe it really is and what form it takes. Subsequently, it might then be possible to adjust generalizations about scientific relationships to correct for any bias.

Recent work shows that the strength of scientific findings has significantly changed with year of publication in at least four specific research areas in evolutionary ecology (Alatalo *et al.* 1997; Gontard-Danek & Møller 1999; Simmons *et al.* 1999; Poulin 2000). No one knows whether these are isolated occurrences or reflect a more widespread trend. The relationship between the strength of research findings and the year of publication requires interpretation. These specific cases have been attributed to changing belief systems (Alatalo *et al.* 1997; Poulin 2000), so-called 'bandwagon' fads (Palmer 2000) or biased study design (Tregenza & Wedell 1997). Alatalo *et al.* (1997) suggested that publication of a theoretical finding that substantiates previous empirical findings might make it easier to publish a certain kind of result. Simmons *et al.* (1999) suggested that, during the initial stages of research in a specific area, it might be easier to publish confirmatory results than later on when a more critical view of the field develops. We should not forget, however, that a relationship between effect size and year is a simple correlation. Determination of causation is difficult, so confounding variables must be considered (Tregenza & Wedell 1997). For example, studies may use smaller sample sizes once a given effect has been determined in an earlier study (Thornhill *et al.* 1999). Many countries require that sample sizes in experiments, but not in observational studies, are kept to a minimum for ethical reasons, but only in certain taxa. Such factors must be taken into account before inferring potential bias, as indicated by a relationship between strength of research findings and year of publication. Might reporting of significant 'year effects' itself be subject to a publication bias exaggerating its prevalence or magnitude?

Here, in what is, to our knowledge, the first systematic inquiry, we examine 44 previously published ecological or evolutionary meta-analyses to test the generality of earlier findings. We find a small, but highly significant, decline in the strength of reported relationships with publication date, while empirical studies with larger sample sizes report more modest findings. The latter finding suggests a publication bias against studies with small sample sizes

*Author for correspondence (michael.jennions@anu.edu.au).

DOI 10.1098/rspb.2001.1832

and non-significant findings. Our results mirror those in medicine and the social sciences (Song *et al.* 2000) where it takes longer to publish non-significant results (Stern & Simes 1997; Ioannidis 1998). It is suggested that publication bias based on p-values (Palmer 2000), rather than external factors, may offer a more parsimonious and general explanation for the observed results.

2. METHODS

Ecological and evolutionary literature for meta-analyses up to the end of 2000 were surveyed. The journals *American Naturalist*, *Animal Behaviour*, *Behavioral Ecology*, *Behavioral Ecology and Sociobiology*, *Ecological Monographs*, *Ecology*, *Evolution*, *Evolutionary Biology*, *Journal of Evolutionary Biology* and *Quarterly Review of Biology* were examined. The phrase 'meta-analy*' was entered into the electronic database webSPIRS to locate papers where it occurs in the title or abstract. We then examined the title and place of publication, and directly inspected any papers that seemed to be related to non-human evolutionary or ecological biology (most 'hits' were in the medical or social sciences). Furthermore, a number of colleagues who have used meta-analyses in their research were contacted in order to locate as yet unpublished studies. A total of 44 peer-reviewed meta-analyses were identified for the present study. Thirty-seven meta-analyses that initially appeared suitable were excluded because they: (i) did not present effect sizes for original empirical studies, only mean effect sizes, and the authors were unable to provide the original effect sizes; or (ii) had been updated and overlapped with a more recent meta-analysis that was already included (e.g. Hamilton & Poulin 1997); or (iii) had almost no variation in year of publication (e.g. Arnqvist *et al.* 1996); or (iv) presented too few effect sizes ($n < 5$) (Westlake & Rowe 1999); or (v) were generally unsuitable because they presented complex statistics where effect size depended on a continuous variable (Goldberg *et al.* 1999). Finally, meta-analyses where heritability was treated as an effect size were excluded (e.g. Møller & Thornhill 1997) because it is unclear whether heritability (h^2) or an effect size calculated from the p-value of the associated regression should be used. The meta-analyses used are listed in electronic Appendix A (available on The Royal Society's Publications Web site).

Four relationships were examined: (i) the relationship between effect size and year of publication—the year of publication for unpublished studies was coded as the year after the original meta-analysis was published; (ii) the relationship between effect size and sample size (r_{bias} of Palmer 1999); and (iii) the relationship between standardized effect size and sample size (standardization is necessary, even when using a rank-correlation test, to stabilize the variance in effect size, that increases as sample size decreases (see Begg & Mazumdar 1994, p. 1089)). All three relationships were calculated as Spearman's correlations. Finally, the relationship between effect size and year of publication, weighted for variation in sampling effort, was examined. To estimate this, a random-effects continuous model meta-analysis with year of publication as the independent variable and the inverse of sampling variance as the weighting factor was used. The effect size for the influence of year of publication was calculated by converting the p-value to a standard normal deviate (Z-score) and then using the formula $r = \sqrt{Z^2/n}$ (Rosenthal 1994). The one-tailed p-value for year of publication generated by a randomization method with 999 replicates run in MetaWin 2.0 was used (Rosenberg *et al.* 2000). Finally, for the purpose of performing meta-analyses to determine weighted mean effect size, all correlation coefficients were converted to Fisher's Z. This may legitimately be computed for Spearman's correlations when $n \geq 10$ and $\rho \leq 0.9$ (Zar 1984).

Most of the 44 original meta-analyses asked several related questions and therefore comprised several datasets that examined different relationships (e.g. parasite load versus spleen size; parasite load versus condition). In addition, the original authors often found significantly more heterogeneity in effect size than could be explained by sampling error. They therefore looked for an underlying structure in the data by classifying studies into groups (e.g. temperate versus tropical) and testing for significant among-group variance in effect sizes for each categorical factor using Q_B (variation in effect size explained by differences among rather than within groups) (Rosenberg *et al.* 2000). For each of the 44 meta-analyses we therefore split the data for each response variable using the single categorical factor that generated the greatest differences in effect sizes among groups (but only if $p < 0.05$ for Q_B). This reduces the likelihood that effect size is correlated with sample size and/or year of publication because of confounding variation with study system type. This yielded 232 datasets with 22.8 ± 1.7 (mean ± s.e.) effect sizes per dataset (range = 5–246).

Initially, the four correlations of interest were calculated separately for each of the 232 datasets. These analyses are referred to as being at the 'sample level'. Treating these datasets as distinct is a standard practice in meta-analyses because the questions being asked are conceptually different (Rosenthal 1994, p. 241). However, because the different response variables examined were often measured in identical or overlapping sets of original empirical studies, they are not, strictly, statistically independent. So, to be conservative, the weighted mean correlation (Z-transformed) per original meta-analysis for each relationship of interest was also calculated. Subsequent analyses using these means are referred to as being at the 'original meta-analysis level'.

The weighted mean effect sizes for each of the four relationships at both the sample level and the original meta-analysis level were calculated. For the latter, effect size was weighted by the average number of empirical studies per dataset in the original meta-analysis. Mixed-effects models were used. To test for the significance of the mean effect, bias-corrected confidence intervals were calculated using bootstrapping with 999 replications run in MetaWin 2.0 (Rosenberg *et al.* 2000). This approach does not require that effect sizes be parametrically distributed. Finally, unweighted mean effect sizes and confidence intervals were calculated using standard summary statistics and one-sample t-tests. This tests the robustness of the results by dealing with the criticism that research areas where more studies have been conducted have undue influence on estimates of mean relationships. Conversely, of course, calculating unweighted means gives equal importance to relationships based on very small sample sizes. Data are presented as mean ± s.e. or with 95% confidence intervals. Sample sizes vary slightly because, for some studies, variance and sample size were not both available.

3. RESULTS

Results are summarized in table 1. Starting at the sample level of analysis, there was a significant negative relationship between year of publication and effect size ($r = -0.074$, $p < 0.004$; 95% confidence interval (CI): -0.115 to -0.033, $n = 232$). There was, however, also a significant negative relationship between effect size and

Table 1. Relationships (r) between effect size, standardized effect size, year of publication and sample size. (****$p < 0.0001$; ***$p < 0.002$; **$p < 0.01$; *$p < 0.02$.)

method of calculation	year versus effect	n versus effect	n versus standard effect	year versus effect[a]
weighted meta-analysis of datasets[b]	−0.074**** ($n = 232$)	−0.124**** ($n = 210$)	−0.101**** ($n = 201$)	−0.043** ($n = 213$)
weighted meta-analysis of original meta-analyses[c]	−0.133**** ($n = 44$)	−0.188*** ($n = 38$)	−0.121 ($n = 36$)	−0.105*** ($n = 39$)
unweighted mean of original meta-analyses	−0.084* ($n = 44$)	−0.129* ($n = 38$)	−0.113 ($n = 38$)	−0.058 ($n = 39$)

[a] Controlled for variation in sampling effort (see § 2).
[b] Weighted by sample size.
[c] Weighted by the average sample size per dataset.

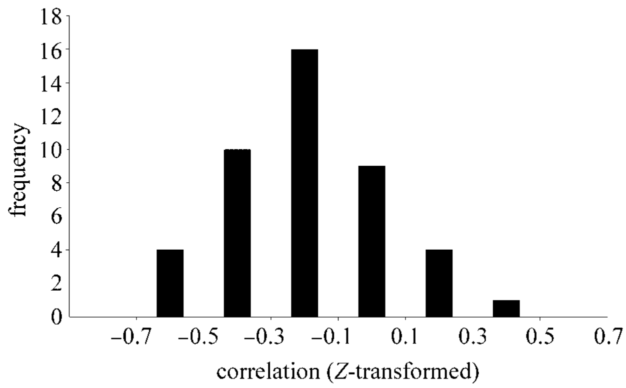

Figure 1. Histogram of the correlation (Z-transformed) between the year of publication and the effect size at the level of original meta-analyses ($n = 44$). Bins are 0.2 units wide.

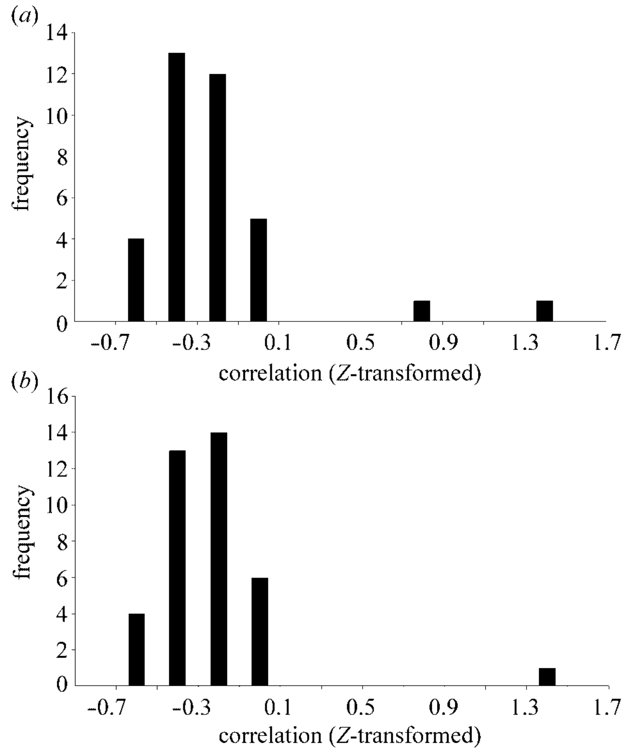

Figure 2. Histogram of the correlation (Z-transformed) between the sample size and (a) standardized effect size ($n = 38$) and (b) effect size ($n = 36$) at the level of original meta-analyses. Bins are 0.2 units wide.

the sample size used to estimate the effect ($r = -0.124$, $p < 0.0001$; 95% CI: -0.172 to -0.080, $n = 210$; using standardized effect size: $r = -0.101$, $p < 0.0001$; 95% CI: -0.147 to -0.049, $n = 201$). Given this, the influence of year of publication was reassessed, but controlled for variation in sampling effort. The relationship was still significant ($r = -0.043$, $p < 0.01$; 95% CI: -0.088 to -0.006, $n = 213$).

Next, these analyses were repeated at the original meta-analyses level. The average number of datasets per original meta-analysis was 5.3 ± 1.0, and the average number of samples per dataset was 31.2 ± 4.9 (range = 5–199). Again, there was a significant negative relationship between year of publication and effect size ($r = -0.133$, $p < 0.01$; 95% CI: -0.189 to -0.062, $n = 44$). There was also a significant negative relationship between effect size and sample size ($r = -0.188$, $p < 0.01$; 95% CI: -0.246 to -0.125, $n = 38$), but not for standardized effect size ($r = -0.121$, $p > 0.10$; 95% CI: -0.218 to 0.044, $n = 36$). The significant negative relationship between year of publication and effect size remained even after controlling for sampling effort ($r = -0.105$, $p < 0.01$; 95% CI: -0.151 to -0.055, $n = 39$).

Finally, we just calculated unweighted means at the original meta-analysis level (see § 2). There was a negative relationship between year of publication and effect size ($r = -0.084$, $t_{43} = 2.43$, $p = 0.019$; 95% CI: -0.153 to -0.014) (figure 1), and between effect size and sample size ($r = -0.129$, $t_{37} = 2.48$, $p = 0.018$; 95% CI: -0.234 to -0.024; standardized effect size: $r = -0.113$, $t_{35} = 1.90$, $p = 0.066$; 95% CI: -0.234 to 0.008). If one or two extreme outliers, respectively, are removed (figure 2), both relationships are highly significant ($r = -0.172$, $t_{36} = 5.85$, $p < 0.0001$; 95% CI: -0.232 to -0.112; standardized effect size: $r = -0.188$, $t_{33} = 7.03$, $p < 0.001$; 95% CI: -0.243 to -0.134). There was no significant relationship between year of publication and effect size after controlling for sampling effort ($r = -0.058$, $t_{38} = 1.59$, $p = 0.119$; 95% CI: -0.132 to 0.016) (figure 3). However, if less reliable estimates from meta-analyses where the average number of effect sizes per dataset was ≤ 10 are excluded ($n = 6$), the effect of publication date remains significant ($r = -0.095$, $t_{32} = 2.65$, $p = 0.012$; 95% CI: -0.168 to -0.022).

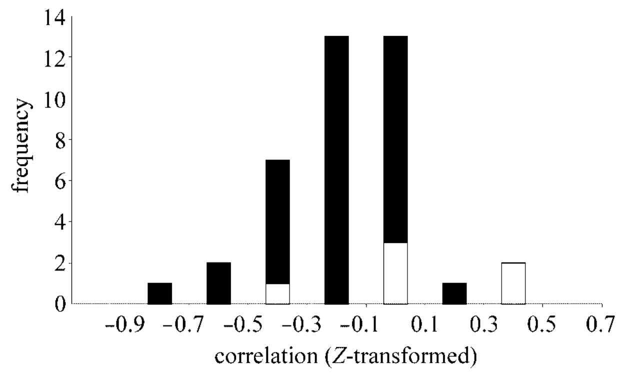

Figure 3. Histogram of the correlation (Z-transformed) between the year of publication and the effect size at the level of original meta-analyses after controlling for sampling effort ($n = 39$). Meta-analyses with fewer than 10 studies per dataset are indicated by the white bar area. Bins are 0.2 units wide.

It has been suggested that some authors may currently contribute disproportionately to meta-analyses and this might distort results (Palmer 1999). The strength of the relationships of interest between meta-analyses were compared where we were the authors ($n = 11$) and in the remaining meta-analyses ($n = 33$). We investigated this at the original meta-analysis level. There was no significant difference in the weighted mean relationship between effect size and year of publication ($Q_B = 2.54$, $p = 0.147$). When corrected for sampling effort, the meta-analyses where we were authors had a marginally *weaker* mean relationship ($Q_B = 2.54$, $p = 0.058$; $r = -0.142$ versus -0.040). There was also no difference for the relationship between sample size and effect size ($Q_B = 0.46$, $p = 0.506$; $r = -0.175$ versus -0.218). For sample size versus standardized effect size the difference was non-significant, but stronger for the meta-analyses that we authored ($Q_B = 2.29$, $p = 0.122$; $r = -0.064$ versus -0.255). Examining only those meta-analyses of other authors, the relationship between year of publication and effect size remains significant both before ($r = -0.166$ $p < 0.01$, $n = 33$) and after correcting for sampling effort ($r = -0.142$, $p < 0.01$, $n = 28$). The relationship between sample size and effect size was significant ($r = -0.175$, $p < 0.01$, $n = 27$), but not that between sample size and standardized effect size ($r = -0.053$, $p > 0.25$, $n = 25$).

4. DISCUSSION

Previous studies have reported either a significant increase (e.g. Alatalo *et al.* 1997) or decrease (e.g. Møller & Alatalo 1999) in the strength of research findings over time. Here it was found that, for all three analytical approaches, on average, there is a significant negative relationship between effect size and year of publication. Even after controlling for variation in sampling effort, the relationship between year of publication and effect size remained significant in weighted meta-analyses. In unweighted analysis the trend was towards a significant negative relationship, but this only became significant if a few less-reliable studies, based on small samples, were excluded. In general, it is concluded that more recently published studies in ecology and evolution report weaker findings (have smaller 'effect sizes'). The effect of year of publication is, however, fairly weak. It explains, at most, 1.8% of the variance in effect sizes ($r^2 = -0.133^2 = 0.18$). An effect where $r^2 = 1\%$ and 9% are defined as small and medium, respectively, by Cohen (1988). To place this in perspective, however, in a recent review it was found that the mean effect size for biological relationships in published ecology and evolution meta-analyses is such that r^2 is only 4.4–7.3% (A. P. Møller and M. D. Jennions, unpublished data).

One reason why more recently published ecological and evolutionary studies document weaker scientific findings is that clear-cut results reduce the time-lag from study completion to publication due to bias in submission, reviewing and editorial decisions. A larger effect size leads to earlier publication (e.g. Ioannides 1998). The role of statistical significance in publication decisions is well known in the medical sciences (Song *et al.* 2000). The same biases also seem likely to arise in biology (Palmer 2000), although convincing data based on direct comparison of published and unpublished studies, or tracking the fate of a known set of studies, is not yet, to our knowledge, available (Møller & Jennions 2001). Although ecological and evolutionary research may be influenced by broader social trends (Gowaty 1996), sudden paradigm shifts (Kuhn 1996), or even idiosyncratic changes in what topics, findings or explanations are of interest (Feyerabend 1993), an average temporal decline in the strength of published findings is not predicted by these phenomena. Fads, social trends and paradigm shifts would seem as likely to lead to an increase as a decrease in effect sizes. For example, Kuhnian paradigm shifts are usually about how an established phenomenon is explained and what research questions are seen to be interesting: when Einstein replaced Newton apples still fell from the tree.

A slightly different perspective on these matters has been taken by Simmons *et al.* (1999), Poulin (2000) and Palmer (2000), who argue that a new field of research (or paradigm) may generate a so-called 'bandwagon' effect where corroborative studies are readily published. Later, as skepticism grows and methodology improves, results that refute or question earlier findings may find a more receptive audience (Simmons *et al.* 1999). This would also explain a decline in effect size with time. Similarly, one could argue that in biology, replication of an influential study identifying a novel relationship is likely to be conducted sooner by those working on related species. Researchers are usually more alert to trends in their own narrow field of research because of who they interact with and which journals they read. Later, as findings are more widely disseminated, those working on more distant taxa may also 'replicate' the original study when looking for a similar relationship. However, for biological reasons, a mating pattern detected in long-tailed widow birds may be less likely to be seen in house finches and unlikely to be seen in crickets or frogs. This, too, could result in a decline in effect size with year of publication (e.g. Poulin 2000, p. 791). Alternatively, one could argue that an influential new theoretical paper suddenly makes it possible to publish a result that was previously considered implausible because it is now 'explicable' (Alatalo *et al.* 1997).

The problem with all these speculative scenarios, however plausible, is that they have little independent support. They all stem from the same two patterns: the relationships between effect size and year of publication or sample size. In our view it is premature to attribute temporal trends to the publication of specific papers, as did Alatalo *et al.* (1997) and Poulin (2000). For example, while Alatalo *et al.* (1997) showed that heritability estimates increased with year of publication and attributed this to Pomiankowski's 1988 model showing that 'good gene' models were theoretically workable, Møller & Alatalo (1999) found a significant temporal decline in paternal effects on offspring viability. In general, explanations for temporal trends based on Kuhnian paradigm shifts or socially driven fads may have been readily accepted simply because fluctuating asymmetry and good genes are, for whatever reason, controversial topics. The potential for human prejudice to influence the literature therefore seems greater. In contrast, it has been shown here that, on average, 'year effects' are a widespread phenomenon in ecology and evolution (at least for the currently available datasets). A more general explanation must therefore be sought.

It is suggested that the most parsimonious explanation for our finding of a year effect is the preoccupation of scientists with statistical significance as criteria for publication (see also Palmer 2000). This leads to studies with stronger effects being published sooner. This explanation has the additional advantage that it is consistent with trends in the social and medical sciences, where the role of statistical significance in publication has been directly demonstrated (Song *et al.* 2000). Of course, other factors may be at play in specific cases. The temporal increase in the reported heritability of sexual characters is opposite to that predicted by a bias towards more strongly significant results (Alatalo *et al.* 1997). Likewise, Poulin noted that he observed a temporal decline even though all the individual studies were statistically significant (see fig. 4 of Poulin (2000)).

Our claim that publication bias based on statistical significance explains the 'year effect' is further supported by the significant negative correlation between effect size and sample size in all three analyses. This explained, at most, 3.5% of the variation in effect sizes ($r = -0.188$). If there is no bias, a scatter plot of effect size against sample size should generate a funnel shape around the 'true' effect size, with effect sizes based on larger samples being closer to the 'true' effect size (Light & Pillemer 1984). This occurs because estimates of effect size based on small sample sizes are subject to greater sampling error, yielding a wider range of estimated effects. However, at any given sample size the reported effect sizes should be normally distributed around the mean effect and there should be no relationship between sample size and effect size (Palmer 1999). Any deviations from this symmetrical funnel-shaped plot can be used to infer publication bias. Specifically, if there is a moderate 'true' effect size and researchers, reviewers and editors respond more favourably towards statistically significant results (Song *et al.* 2000; Møller & Jennions 2001), this should generate a skewed funnel plot in which effect size decreases as sample size increases (Begg 1994; Begg & Mazumdar 1994). Studies with small sample sizes will only be published if the results are significant. Conversely, because of low statistical power, non-significant results will probably only be published if they are based on large samples (Cohen 1988). Given a moderate 'true' effect, therefore, when sample size is small only those findings in the direction of the true effect are likely to reach statistical significance. Thus, studies with small samples and small effect sizes are systematically under-reported (Begg 1994).

Earlier studies reported a negative relationship between effect size and sample size for several isolated topics in ecology or evolution (e.g. Palmer (1999) and Jennions *et al.* (2001) and references therein). These reports, however, appear to have used the simple correlation between effect size and sample size, while Begg & Mazumdar (1994) and Begg (1994) recommend that effect size be standardized. On average, correlations using standardized and non-standardized effect size are similar ($r = 0.89$, $p < 0.0001$, $n = 201$), but they can differ considerably. For example, Palmer (2000) reported $r_s = -0.39$ for asymmetry-fitness effect sizes, but using standardized effect size the value is $r_s = -0.16$. The certainty with which it can be concluded that there is funnel plot asymmetry depends on which correlation is used. There is a significant negative correlation for all three analyses based on simple effect size; but for standardized effect size, the correlation is significant at the sample level, non-significant at the original meta-analysis level, and marginally significant using an unweighted approach (although highly significant if two clear outliers are removed). The trend, however, is always a decrease in effect size as sample size increases, as predicted by a publication bias against non-significant results.

It has previously been argued that this relationship could arise due to heterogeneity in the data. For example, if an effect has been shown to be small, those working on the same topic may adjust sample sizes upward (Thornhill *et al.* 1999); experimental studies may generate larger effect sizes with smaller sample sizes because confounding variables are controlled; and taxa where sample sizes tend to be smaller may fortuitously be those where the actual effect is stronger. Although these alternative explanations cannot be completely excluded, the fact that the data were initially partitioned to reduce within-group heterogeneity makes them less plausible.

In conclusion, publication bias, whatever the underlying cause, appears to be a problem in biology because both year of publication and sample size are correlated with effect size. This raises questions about the validity of drawing general conclusions from the biological literature, using formal meta-analysis or traditional narrative reviews. Because the year of publication and the sample size each explained less than 4% of the variation in effect size the problem may, at first glance, appear negligible. Elsewhere, we have estimated the number of studies 'missing' due to publication bias (Jennions & Møller 2002). In about 15% of cases robust, positive estimates of mean effect size no longer differ significantly from zero if 'missing' unpublished studies are taken into consideration. Publication bias is therefore a general problem, which is apparently not unique to strongly hypothesis-driven science (cf. Poulin 2000).

The authors thank Göran Arnqvist, Michael Brett, Isabelle Côté, Peter Curtis, Mark Forbes, Peter Hamback, Nick Jonsson, Julia Koricheva, Dean McCurdy, Fiorenza Micheli, Iago

Mosqueira, Robert Poulin, Howie Riessen, Michael Rosenberg, Gina Schalk, Xianzhong Wang, Peter van Zandt and many others for generously providing unpublished information. J. Shykoff, D. Pope, P. Backwell and J.H. Christy provided instructive comments on this work.

REFERENCES

Alatalo, R. V., Mappes, J. & Elgar, M. 1997 Heritabilities and paradigm shifts. *Nature* **385**, 402–403.

Arnqvist, G., Rowe, L., Krupa, J. J. & Sih, A. 1996 Assortative mating by size: a meta-analysis of mating patterns in water striders. *Evol. Ecol.* **10**, 265–284.

Begg, C. B. 1994 Publication bias. In *The handbook of research synthesis* (ed. H. Cooper & L. V. Hedges), pp. 399–409. New York: Russell Sage Foundation.

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

Chalmers, A. F. 1999 *What is this thing called science?* 3rd edn. Indianapolis, IN: Hackett Publishing.

Cohen, J. 1988 *Statistical power analysis for the behavioral sciences*, 2nd edn. Mahwah, NJ: Lawrence Erlbaum.

Feyerabend, P. 1993 *Against method*, 3rd edn. New York: Verso Books.

Goldberg, D. E., Rajaniemi, T., Gurevitch, J. & Stewart, O. A. 1999 Empirical approaches to quantifying interaction intensity: competition and facilitation along productivity gradients. *Ecology* **80**, 1118–1131.

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

Gowaty, P. A. (ed.) 1996 *Feminism and evolutionary biology: boundaries, intersections and frontiers*. London: Chapman & Hall.

Hamilton, W. J. & Poulin, R. 1997 The Hamilton and Zuk hypothesis: a meta analytic approach. *Behaviour* **134**, 299–320.

Hempel, G. G. & Hempel, C. G. 1966 *Philosophy of natural science*. Englewood Cliffs, NJ: Prentice-Hall.

Ioannidis, J. 1998 Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *J. Am. Med. Assoc.* **279**, 281–286.

Jennions, M. D. & Møller, A. P. 2002 Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. *Biol. Rev.* (In the press.)

Jennions, M. D., Møller, A. P. & Petrie, M. 2001 Sexually selected traits and adult survival: a meta-analysis. *Q. Rev. Biol.* **76**, 3–36.

Kuhn, T. S. 1996 *The structure of scientific revolutions*, 3rd edn. University of Chicago Press.

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

Møller, A. P. & Alatalo, R. V. 1999 Good genes effects in sexual selection. *Proc. R. Soc. Lond.* B **266**, 85–91.

Møller, A. P. & Jennions, M. D. 2001 Testing and adjusting for publication bias. *Trends Ecol. Evol.* **16**, 580–586.

Møller, A. P. & Thornhill, R. 1997 A meta-analysis of the heritability of developmental stability. *J. Evol. Biol.* **10**, 1–16.

Palmer, A. R. 1999 Detecting publication bias in meta-analysis: a case study of fluctuating asymmetry and sexual selection. *Am. Naturalist* **154**, 220–233.

Palmer, A. R. 2000 Quasireplication and the contract of error: lessons from sex ratios, heritabilities and fluctuating asymmetry. *Ann. Rev. Ecol. Sys.* **31**, 441–480.

Pomiankowski, A. 1988 The evolution of female mating preferences for male genetic quality. *Oxf. Surv. Evol. Biol.* **5**, 136–184.

Poulin, R. 2000 Manipulation of host behaviour by parasites: a weakening paradigm?. *Proc. R. Soc. Lond.* B **267**, 787–792.

Rosenberg, M. S., Adams, D. C. & Gurevitch, J. 2000 *MetaWin: statistical software for meta-analysis*, v. 2.0. Sunderland, MA: Sinauer.

Rosenthal, R. 1994 Parametric measures of effect size. In *The handbook of research synthesis* (ed. H. Cooper & L. V. Hedges), pp. 231–244. New York: Russell Sage Foundation.

Simmons, L. W., Tomkins, J. L., Kotiaho, J. S. & Hunt, J. 1999 Fluctuating paradigm. *Proc. R. Soc. Lond.* B **266**, 593–595.

Song, F., Eastwood, A. J., Gilbody, S., Duley, L. & Sutton, A. J. 2000 Publication and related biases. *Hlth Technol. Assess.* **4**, 1–115.

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

Thornhill, R., Møller, A. P. & Gangestad, S. 1999 The biological significance of fluctuating asymmetry and sexual selection: a reply to Palmer. *Am. Naturalist* **154**, 234–241.

Trengenza, T. & Wedell, N. 1997 Natural selection bias. *Nature* **386**, 234.

Westlake, K. P. & Rowe, L. 1999 Developmental costs of male sexual traits in the water strider *Rheumatobates rileyi*. *Can. J. Zool.* **77**, 917–922.

Zar, J. R. 1984 *Biostatistical analysis*, 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.

Visit http://www.pubs.royalsoc.ac.uk to see an electronic appendix to this paper.