FORUM FORUM FORUM

FORUM is intended for new ideas or new ways of interpreting existing information. It provides a chance for suggesting hypotheses and for challenging current thinking on ecological issues. A lighter prose, designed to attract readers, will be permitted. Formal research reports, albeit short, will not be accepted, and all contributions should be concise with a relatively short list of references. A summary is not required.

Meta-analysis, can it ever fail?

Janne S. Kotiaho, Dept of Biological and Environmental Sciences, Univ. of Jyväskylä, P.O. Box 35, FIN-40351 Jyväskylä, Finland (jkotiaho@cc.jyu.fi). – Joseph L. Tomkins, Dept of Environmental and Evolutionary Biology, Univ. of St Andrews, St Andrews, Fife, UK KY16 9TS.

Conceptual development in science is based on the ability to synthesise bodies of often scattered and heterogeneous research findings. The development of meta-analytical techniques during the last two decades has provided science with powerful tools aiding such synthesis (Light and Pillemer 1984, Hedges and Olkin 1985, Rosenthal 1991, Cooper and Hedges 1994). Compared to the more traditional qualitative and narrative reviews, meta-analytical techniques offer quantitative and objective methods for reviewing literature and are therefore considered to be superior. This superiority is evidenced by the increasing frequency with which meta-analytical approach is adopted across different fields of science (Arnqvist and Wooster 1995, Møller and Jennions 2001).

The main objective of a meta-analysis is to provide an estimate of the true effect based on studies that are available. To obtain this estimate, different test statistics or significance levels are transformed into a common currency called effect size. After necessary transformations, an overall effect size may be calculated and the general significance of the studied phenomenon can be statistically tested (Rosenthal 1991, Cooper and Hedges 1994). The large set of statistical procedures available in meta-analytical techniques makes it a flexible and powerful tool that considerably enhances the quality and interpretative value of any research synthesis.

However, meta-analysis is not without its weaknesses. Already the early advocates of the technique recognised problems associated with it and called for self-critical use of meta-analysis (Leviton and Cook 1981). The major problem, common to all reviews, is that published studies may not be representative of studies that have, in fact, been conducted (Sterling 1959, Greenwald 1975, Coursol and Wagner 1986, Easterbrook et al. 1991, Greenland 1994, Csada et al. 1996). For meta-analysis, there exists an extensive literature on the effects that selective publication or publication bias may have on the estimates of effect sizes, and on how these biases may be reduced or removed (Rosenthal 1979, Light and Pillemer 1984, Begg and Berlin 1988, Iyengar and Greenhouse 1988, Hunter and Schmidt 1990, Dear and Begg 1992, Hedges 1992, Begg and Mazumdar 1994, Cooper and Hedges 1994, Vevea and Hedges 1995, Gleser and Olkin 1996, Cleary and Casella 1997, Egger et al. 1997, Copas 1998, Duvall and Tweedie 2000, Song et al. 2000).

One recommended method for detecting publication bias is to construct a funnel plot (Light and Pillemer 1984, Begg 1994, Vevea and Hedges 1995, Palmer 1999, Duvall and Tweedie 2000, Palmer 2000, Møller and Jennions 2001). The funnel plot is constructed by plotting effect size against the corresponding sample size. The plot is named after the funnel shape of the scatter that is caused by the statistical properties of the variance about the true mean; as sample size decreases, the variance around the true mean increases due to sampling error (Light and Pillemer 1984). The funnel plot has an underlying assumption that effect size is independent of sample size, and any deviation from the funnel shape results from publication bias or some other bias, such as retrieval bias. Bias may be suspected if a correlation exists between the effect size and sample size (Light and Pillemer 1984). In recent ecological and evolutionary literature the use of a funnel plot with meta-analysis has gained increasing popularity (Gurevitch et al. 1992, Arnqvist et al. 1996, Gontard-Danek and Møller 1999, Møller 1999, Palmer 1999, Thornhill et al. 1999, Vøllestad et al. 1999, Palmer 2000).

The objective of our commentary is to highlight one problem associated with significance testing of the overall effect size derived from meta-analysis; we suggest that when publication bias exists, meta-analysis cannot fail to provide an effect size significantly different from zero.

In the ecological literature only 8.6% of published studies ($n = 1201$ papers from 43 biological journals) statistically testing the main hypothesis presented non-significant results (Csada et al. 1996). Almost identical results (9% of studies) have been found in a review of psychological literature (Smart 1964) and several findings of similar magnitude have been made in medical literature (Song et al. 2000). There is also more statistically significant results in the literature than would be expected by chance (Palmer 2000). Moreover, there is an alarming tendency in current and popular evolutionary ideas to go through a significant shift in the magnitude of their effect sizes in relation to year of publication (Alatalo et al. 1997, Gontard-Danek and Møller 1999, Møller and Alatalo 1999, Simmons et al. 1999). One explanation for year effects is that in the early phase of a scientific discovery, the enthusiasm to find supportive results inevitably leads to publication bias; supportive studies get accepted more easily resulting in an overestimate of the true effect size. However, as the discipline matures and scientific rigour settles in, reported effect sizes shift closer to the true effect size (Kuhn 1996, Alatalo et al. 1997, Simmons et al. 1999).

If only 9% of studies that were actually performed (rather than reported) find non-significant results (Csada et al. 1996), then the true effect sizes for most hypotheses must be close to one. In reality, however, the reported overall effect sizes for any hypothesis in ecology and evolution rarely exceed 0.3 (measured as Pearson's correlation coefficient r) (Arnqvist et al. 1996, Leung and Forbes 1996, Møller and Thornhill 1997, 1998, Møller 1999, Møller and Alatalo 1999, Palmer 1999, Vøllestad et al. 1999, Møller and Jennions 2001). Although difficult, several studies have obtained direct evidence for publication bias. By conducting surveys of the probability of investigators submitting a study for publication (Greenwald 1975, Coursol and Wagner 1986, Dickersin et al. 1987), following up cohorts of registered studies (Easterbrook et al. 1991, Dickersin et al. 1992, Dickersin 1997) and by comparing unpublished with published results (Smart 1964, Smith 1980, Simes 1986) it has been found that studies finding statistically significant effect sizes are systematically more often published than studies not finding significant effects. Thus, we may be confident that the above facts reveal extensive publication bias.

If we accept that there is extensive publication bias, it seems evident that there is no possibility for a meta-analysis to fail. This is because, if only about 9% of published work reports non-significant results (Csada et al. 1996), it follows that when an overall effect size from this data set is calculated, a meta-analysis cannot fail to find a significant effect. This is evident from a simple exercise. Let us assume that there is a new idea that is not true (i.e. has a true effect size of zero). However, because of publication bias, only 9% of studies report non-significant results. Therefore if the new idea attracted 10 publications, on average 9 of them would be significant (to be conservative we use the just significant two-tailed probability level 0.05 corresponding to $Z = 1.96$) and 1 would be non-significant (two-tailed probability level 1.00 corresponding to $Z = 0.00$). If we assume that all tests had the same methods and sample size we do not need weighting, but the probabilities may be combined straightforward as sum of Z's divided by the square root of the number of studies (Rosenthal 1991). This measure will be distributed as Z and in our hypothetical case the combined probability level of these studies is < 0.001 ($Z = 5.58$). This exercise clearly demonstrates that when publication bias exists, meta-analysis cannot fail to find a significant result.

In conclusion, publication bias should always be tested for as a part of every meta-analysis, and where bias is detected, further analysis and interpretation should only be undertaken with extreme caution. This is because, where bias exists and cannot be accounted for, the overall effect size will be an overestimate of the true effect size. In such cases, testing for the significance of the effect size is pointless and the estimate of the effect size should only be used as an estimate of the upper limit of the true effect size. However, we emphasise that where publication bias has been shown not to exist, meta-analysis remains a useful statistical approach. Meta-analytic techniques may be used as a powerful tool to analyse the results of multiple experiments within a study, and also in other cases, where one can be sure that there is no publication bias.

Acknowledgements – J. S. K. was funded by the Academy of Finland and J. L. T. by a David Phillips Fellowship from the BBSRC.

References

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

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

Arnqvist, G., Rowe, L., Krupa, J. J. and 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: Cooper, H. and Hedges, L. V. (eds), The handbook of research synthesis. Russel Sage Foundation, New York, pp. 399–409.

Begg, C. B. and Berlin, J. A. 1988. Publication bias: a problem in interpreting medical data. – J. R. Statist. Soc. A 151: 419–463.

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

Cleary, R. J. and Casella, G. 1997. An application of Gibbs sampling to estimation in meta-analysis: accounting for publication bias. – J. Educ. Behav. Stat. 22: 141–154.

Cooper, H. and Hedges, L. V. 1994. The handbook of research synthesis. – Sage Publ.
Copas, J. 1998. What works? Selectivity models and meta-analysis. – J. R. Statist. Soc. A 162: 95–109.
Coursol, A. and Wagner, E. E. 1986. Effect of positive findings on submission and acceptance rates: a note on meta-analysis bias. – Professional Psychol. 17: 136–137.
Csada, R. D., James, P. C. and Espie, R. H. M. 1996. The "file drawer problem" of non-significant results: does it apply to biological research? – Oikos 76: 591–593.
Dear, K. B. G. and Begg, C. B. 1992. An approach for assessing publication bias prior to performing a meta-analysis. – Stat. Sci. 7: 237–245.
Dickersin, K. 1997. How important is publication bias? A synthesis of available data. – Aids Educ. Prev. 9: 15–21.
Dickersin, K., Chan, S., Chalmers, T. C. et al. 1987. Publication bias and clinical trials. – Control. Clin. Trials 8: 343–353.
Dickersin, K., Min, Y. I. and Meinert, C. L. 1992. Factors influencing publication of research results. Follow up of applications submitted to two institutional review boards. – JAMA 267: 374–378.
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.
Easterbrook, P. J., Berlin, J. A., Gopalan, R. and Matthews, D. R. 1991. Publication bias in clinical research. – Lancet 337: 867–872.
Egger, M., Davey-Smith, G., Schneider, M. and Minder, C. 1997. Bias in meta-analysis detected by a simple, graphical test. – BMJ 315: 629–634.
Gleser, L. J. and Olkin, I. 1996. Models for estimating the number of unpublished studies. – Stat. Med. 15: 2493–2507.
Gontard-Danek, M.-C. and Møller, A. P. 1999. The strength of sexual selection: a meta-analysis of bird studies. – Behav. Ecol. 10: 476–486.
Greenland, S. 1994. A critical look at some popular meta-analytic methods. – Am. J. Epidem. 140: 290–296.
Greenwald, A. G. 1975. Consequences of prejudice against the null hypothesis. – Psychol. Bull. 85: 1–20.
Gurevitch, J., Morrow, L. L., Wallace, A. and Walsh, J. S. 1992. A meta-analysis of competition in field experiments. – Am. Nat. 140: 539–572.
Hedges, L. V. 1992. Modelling publication selection effects in meta-analysis. – Stat. Sci. 7: 246–255.
Hedges, L. V. and Olkin, I. 1985. Statistical methods for meta-analysis. – Academic Press.
Hunter, J. E. and Schmidt, F. L. 1990. Methods of meta-analysis: correcting error and bias in research findings. – Sage Publ.
Iyengar, S. and Greenhouse, J. B. 1988. Selection models and the file drawer problem. – Stat. Sci. 3: 109–135.
Kuhn, T. S. 1996. The structure of scientific revolutions, 3rd ed. – Univ. of Chicago Press.
Leung, B. and Forbes, M. 1996. Fluctuating asymmetry in relation to stress and fitness: effects of trait type as revealed by meta-analysis. – Écoscience 3: 400–413.
Leviton, L. C. and Cook, T. D. 1981. What differentiates meta-analysis from other forms of review. – J. Personality 49: 231–236.
Light, R. J. and Pillemer, D. B. 1984. Summing up. – Harvard Univ. Press.
Møller, A. P. 1999. Asymmetry as a predictor of growth, fecundity, and survival. – Ecol. Lett. 2: 149–156.
Møller, A. P. and Thornhill, R. 1997. A meta-analysis of the heritability of developmental stability. – J. Evol. Biol. 10: 1–16.
Møller, A. P. and Thornhill, R. 1998. Bilateral symmetry and sexual selection: a meta-analysis. – Am. Nat. 151: 174–192.
Møller, A. P. and Alatalo, R. V. 1999. Good-genes effects in sexual selection. – Proc. R. Soc. Lond. B 266: 85–91.
Møller, A. P. and Jennions, M. D. 2001. Testing and adjusting for publication bias. – Trends Ecol. Evol. 16: 580–586.
Palmer, A. R. 1999. Detecting publication bias in meta-analyses: a case study of fluctuating asymmetry and sexual selection. – Am. Nat. 154: 220–233.
Palmer, A. R. 2000. Quasi-replication and the contract of error: lessons from sex rations, heritabilities and fluctuating asymmetry. – Annu. Rev. Ecol. Syst. 31: 441–480.
Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. – Psychol. Bull. 86: 638–641.
Rosenthal, R. 1991. Meta-analytic procedures for social research. – Sage Publ.
Simes, R. J. 1986. Publication bias: the case for an international registry of clinical trials. – J. Clin. Oncol. 4: 1529–1541.
Simmons, L. W., Tomkins, J. L., Kotiaho, J. S. and Hunt, J. 1999. Fluctuating paradigm. – Proc. R. Soc. Lond. B 266: 593–595.
Smart, R. G. 1964. The importance of negative results in psychological research. – Can. Psychol. 5: 225–232.
Smith, M. L. 1980. Publication bias and meta-analysis. – Eval. Educ. 4: 22–24.
Song, F., Eastwood, A. J., Gilbody, S. et al. 2000. Publication and related biases. – Health Technol. Assess. 4: 1–115.
Sterling, T. D. 1959. Publication decisions and their possible effects on inference drawn from tests of significance – or vice versa. – J. Am. Statist. Assoc. 54: 30–34.
Thornhill, R., Møller, A. P. and Gangestad, S. W. 1999. The biological significance of fluctuating asymmetry and sexual selection: a reply to Palmer. – Am. Nat. 154: 234–241.
Vevea, J. L. and Hedges, L. V. 1995. A general linear model for estimation effect size in the presence of publication bias. – Psychometrika 60: 419–435.
Vøllestad, L. A., Hindar, K. and Møller, A. P. 1999. A meta-analysis of fluctuating asymmetry in relation to heterozygosity. – Heredity 83: 206–218.