Opinion is intended to facilitate communication between reader and author and reader and reader. Comments, viewpoints or suggestions arising from published papers are welcome. Discussion and debate about important issues in ecology, e.g. theory or terminology, may also be included. Contributions should be as precise as possible and references should be kept to a minimum. A summary is not required.

Non-significant results in ecology: a burden or a blessing in disguise?

Julia Koricheva, Section of Ecology, Dept of Biology, Univ. of Turku, FIN-20014 Turku, Finland (julkoric@utu.fi).

Null hypothesis significance testing remains a common practice in ecology, despite criticism by statisticians (Yoccoz 1991, Cohen 1994) and numerous suggested alternatives (Jones and Matloff 1986, Fernandez-Duque 1996, Parkhurst 2001). The preoccupation of scientists with the statistical significance of tests as a criterion of study value may lead to the under-reporting of non-significant ($P > 0.05$) results in the published literature (the 'file drawer problem', Rosenthal 1979). This situation may arise either because non-significant results are not submitted for publication or because they are rejected in the review process. Less severe forms of bias against non-significant results (sometimes referred to as 'dissemination bias', Song et al. 2000) include time-lag bias (delayed publication) and place of publication bias (publication in low-circulation journals or in the form of technical reports, conference abstracts or dissertations). As a result, even when published, studies reporting non-significant results may be less accessible to researchers, undercited, and less likely to be indexed in major reference databases – and therefore more likely to go unnoticed.

The under-reporting of non-significant results has long been suspected to occur in ecology (Csada et al. 1996), and potential problems caused by publication and dissemination bias against non-significant results for research synthesis, design of ecological experiments and the presentation of the results have been discussed repeatedly in Oikos (Csada et al. 1996, Bauchau 1997, Lortie and Dyer 1999, Kotiaho and Tomkins 2002). However, a recent review by Møller and Jennions (2001) demonstrated that the existing evidence of bias against non-significant results in ecology is based largely on indirect methods (e.g. analysis of the funnel plots or calculation of the fail-safe numbers) which are open to alternative interpretations. Direct evidence of publication bias comes from comparisons of results between unpublished and published studies and from follow-ups of the publication fate of a group of pre-registered studies. Several surveys of this kind conducted in medicine have demonstrated that published studies often report a higher effectiveness of treatments than unpublished trials, and that the publication rate of trials providing significant results is considerably higher than those that found no significant differences among treatments (reviewed in Song et al. 2000). Similar analyses of ecological studies have not been undertaken so far because of the logistic problems in obtaining an unbiased sample of unpublished studies. Although some ecological meta-analyses included considerable number of unpublished studies and were able to compare the magnitude of effect size for published and unpublished studies (Thornhill et al. 1999, Jennions et al. 2001), most of the unpublished studies included in the above reviews were very recent, and might therefore be unpublished simply because they had only recently been conducted (Møller and Jennions 2001). Direct evidence of publication bias in ecology is therefore still lacking.

In order to find out whether the statistical significance of the results affects the publication of ecological studies, I have followed up the fate of manuscripts from Finnish and Swedish doctoral dissertations on ecological topics. A typical PhD thesis in Finland and Sweden consists of several articles (usually 4–6) and a summary. Some of the papers included in the thesis may already have been published or accepted for publication, while others are in manuscript; these are often submitted and published after the PhD student obtains a doctoral degree. Given that PhD students usually have limited time for completion of their studies and little experience of conducting and writing up research, I assumed that all or at least most of the research conducted during the PhD studies is included in the thesis and, thus, Finnish and Swedish doctoral dissertations provide an unbiased source of completed but as

yet unpublished studies. I have examined all doctoral dissertations on ecological topics produced at University of Turku, Finland, between 1982 and 1998, together with other Finnish and Swedish ecological doctoral dissertations available in the library of University of Turku and completed during the same period. Dissertations completed after 1998 were not included in the survey because the researcher might not have had enough time to publish the manuscripts. Only the manuscripts which were neither published nor accepted for publication by the time of dissertation publication were considered. Altogether I examined 187 manuscripts from 93 dissertations defended in 12 different universities (6 Finnish and 6 Swedish) and covering different ecological subdisciplines. I addressed the following questions: (1) are studies reporting non-significant findings less likely to be submitted and/or accepted for publication as compared to studies with significant results (publication bias)?; (2) do studies with significant results get published in higher quality journals as compared to studies with non-significant results (place of publication bias)?; (3) do studies with significant results get published earlier than those with non-significant results (time-lag bias)?

The publication fate of the manuscripts was followed up by searching the Science Citation Index database. When I was unable to find a manuscript from a thesis in the Science Citation Index, I contacted the author by e-mail and inquired about the fate of the paper. If the study had been published, the authors were asked to provide a reference. For unpublished research, authors were asked whether they had submitted a manuscript to a journal; if not, they were asked to indicate the reasons for non-submission. Finally, if the manuscript had been submitted for publication but was rejected in the reviewing process, authors were asked to indicate the name of the journal and the main reason(s) for rejection.

As a measure of study outcome, I calculated for each manuscript the proportion of non-significant results (denoted by $P > 0.05$) in tests of the main hypotheses (as defined in the introduction). I also recorded the impact factor of the journal to which the manuscript was submitted and, for published studies, the time in months from manuscript submission to acceptance and from acceptance to publication. The above dates are usually indicated in the printed versions of the articles, although some journals (including Oikos) indicate only the date of manuscript acceptance.

Fifty-seven (30.5%) of the 187 manuscripts examined were unpublished as of the time of this survey. About half (53.7%) of the unpublished manuscripts (10.1% of total number of examined manuscripts) had not been submitted for publication. The remaining part of the unpublished manuscripts (46.3%) were submitted to journals, but rejected during the review process.

The responses of 35 researchers to the questionnaires indicated that the statistical significance of the results was not an important determinant of the publication fate of ecological studies. The most common reason given by researchers for not submitting a study for publication was lack of time (68% of reasons given), often due to a change in research topic after completion of the PhD. The authors of six of the 22 non-submitted manuscripts are still planning to make modifications (e.g. reanalysing the existing data or adding new data) and to submit the manuscripts in the near future. A third reason for non-submission, indicated by the two Swedish researchers, was the low priority of publication of manuscripts from doctoral dissertations, since according to Swedish regulations such publications are ranked lower than new research conducted after the completion of the PhD. To my knowledge no such difference in the ranking of publications exists in Finland. The above factor may account for the lower submission rate of manuscripts from doctoral dissertations in Sweden as compared to Finland (81% vs 97%, respectively). None of the researchers who replied to my questionnaire indicated the non-significance of the results as a reason for not submitting a study for publication. This is in sharp contrast with a study by Rotton et al. (1995), who reported that ''non-significant results'' was the reason most frequently given for not publishing psychological studies (59.9%).

The most frequent reasons for manuscript rejection appear to be problems with study design (19% of reasons given) and data analysis (14%), lack of novelty in the findings (14%), and unsuitability of the study topic for the scope of the journal (14%). Some manuscripts were rejected because they were poorly written (9.5%) or based on a poor theoretical background (9.5%). Only in one case did the statistical significance of the results play some role in manuscript rejection.

The proportion of non-significant results varied from 0 to 100% in both published and unpublished studies. Contrary to predictions by some ecologists (Csada et al. 1996, Jennions and Møller 2002) and evidence coming from medicine (reviewed in Song et al. 2000), the mean proportion of non-significant results was significantly higher in published (49%) than in unpublished (41%) manuscripts (Wilcoxon's two-sample test, $S = 4237.5$, $N_{published} = 130$, $N_{unpublished} = 54$, $P = 0.021$), indicating that null results do not reduce the probability of publication. When the comparison was conducted for Finland and Sweden separately, the trend remained the same although the difference between the proportion of non-significant results in published and unpublished manuscripts was no longer statistically significant (Finland: 52% vs 43%, $S = 340.5$, $N_{published} = 66$, $N_{unpublished} = 11$, $P = 0.200$; Sweden: 46% vs 41%, $S = 2156.5$, $N_{published} = 62$, $N_{unpublished} = 44$, $P = 0.206$).

The observed difference between the proportion of non-significant results in published and unpublished manuscripts may arise due to differential rates of submission of manuscripts with significant and non-significant results and/or due to their differential acceptance by journals. In order to distinguish between these explanations, I compared the proportion of non-significant results in submitted vs not submitted manuscripts and in accepted vs rejected manuscripts. The proportion of non-significant results was similar in submitted (48%) and not submitted (43%) manuscripts (Wilcoxon's two-sample test, $S = 1425$, $N_{submitted} = 151$, $N_{not\ submitted} = 18$, $P = 0.594$), indicating that manuscripts with non-significant results are not less likely to be submitted for publication. Among submitted manuscripts, however, the proportion of non-significant results was significantly higher in accepted (49%) than in rejected (39%) manuscripts (Wilcoxon's two-sample test, $S = 1210.5$, $N_{accepted} = 128$, $N_{rejected} = 22$, $P = 0.017$). At first sight this result appears paradoxical: why would journals preferentially accept manuscripts with higher proportion of non-significant results? The explanation may lie in the fact that manuscripts with significant and non-significant results are submitted to different journals that differ in rejection rates. In medicine, for instance, Easterbrook et al. (1991) demonstrated that studies with significant results were more likely to be published in journals with a high impact factor; such journals also tend to have higher rejection rates as compared to lower quality journals.

I found a similar tendency in ecological studies. The proportion of non-significant results in a study tended to be negatively associated with the impact factor (IF) of the journal in which the study was published ($r = -0.167$, $N = 126$, $P = 0.063$). Studies reporting mainly significant results (proportion of non-significant results 0–30%) were published in journals with a mean IF of 2.06, those reporting mainly non-significant results (proportion of non-significant results 70–100%) in journals with an average IF of 1.45. Therefore, while I did not find evidence of publication bias against non-significant results, I found some evidence of place of publication bias. This finding agrees with observation by Csada et al. (1996) that the most prestigious ecological journals tend to publish a lower proportion of non-significant papers than journals taken as a whole. Similarly, Murtaugh (2002) has recently reported a positive correlation in two out of four ecological data sets analysed between the magnitude of the effect and the impact factor of the journal in which the study was published. Taken together, the results of the present study and those of Murtaugh (2002) indicate that both statistical significance and magnitude of effect influence the place of publication.

The observed tendency for non-significant studies to be published in journals with a lower impact factor than statistically significant ones may be due either to the higher rejection rates of manuscripts with non-significant results by high-impact journals or to the tendency of researchers to submit non-significant results to less prestigious journals. I found no evidence of the higher rejection rates of manuscripts with non-significant results by high-impact journals since the proportion of non-significant results tended to be lower in rejected studies than in accepted ones both in high- and low-impact journals (31% vs 48% in journals with IF > 2, Wilcoxon's two-sample test, $S = 170.5$, $N_{rejected} = 8$, $N_{accepted} = 59$, $P = 0.051$; 39% vs 50% in journals with IF < 2, Wilcoxon's two-sample test, $S = 307$, $N_{rejected} = 11$, $N_{accepted} = 68$, $P = 0.061$). It appears, therefore, that researchers deliberately submit non-significant results to less prestigious journals, either because they consider those results to be less interesting and important or because they believe that studies reporting non-significant results will be rejected by high-impact journals.

The above trend has some implications for research synthesis in ecology. Papers published in low-impact journals may be less accessible to researchers as compared to high impact journals. Furthermore, in some ecological meta-analyses the search for relevant studies is restricted to the relatively small number of journals which are considered to publish most of the literature on the topic (Gurevitch et al. 1992, Tonhasca and Byrne 1994). If this approach is taken, it is important that both high and low impact journals be selected for the survey to avoid possible dissemination bias.

Several studies in medicine have indicated that clinical trials with statistically significant results are published on average 2–3 years earlier than non-significant ones (Stern and Simes 1997, Ioannidis 1998, Misakian and Bero 1998). Jennions and Møller (2002) have suggested that a similar time-lag in the publication of non-significant results occurs in ecological studies, and may account for the observed decrease in the magnitude of effect size with publication year in ecological meta-analyses. Contrary to this prediction, I found no significant relationship between the proportion of non-significant results and time from manuscript submission to acceptance ($r = -0.010$, $N = 80$, $P = 0.932$), indicating that statistical significance of the results does not affect the duration of the review process. The length of time interval from manuscript submission to acceptance did not depend on the impact factor of the journal ($r = -0.080$, $N = 80$, $P = 0.483$). However, the time from manuscript acceptance to publication was positively correlated with the impact factor ($r = 0.271$, $N = 79$, $P = 0.015$), indicating that time to publication tends to be longer in more prestigious journals (with the exception of journals like Nature and Science which are specifically aimed at rapid publication). Given that studies with non-significant results tend to be submitted to low-impact journals, which typically have lower rejection rates and shorter time from acceptance to publication, non-significant results in ecology may actually

get published earlier than studies reporting significant results. I cannot exclude the possibility, however, that there might be a time-lag in submission of non-significant results to journals.

To summarize, I found no indication of publication bias and time-lag bias and a limited evidence of place of publication bias against non-significant results for manuscripts from Finnish and Swedish doctoral dissertations on ecological topics. The results of this survey thus differ from those for medicine, where strong evidence of publication and dissemination bias against non-significant results has been found (reviewed in Song et al. 2000), but this is hardly surprising. The aim of clinical medicine is to find treatments which have a significant effect on a particular medical condition, and there is enormous pressure from society, from competing institutions and from funding agencies to obtain results as quickly as possible. From this perspective, treatments which have no effect are of no interest, and it is understandable (although unethical) that medical researchers often choose to start a new trial with different treatments rather than write up and publish the results of the trial that failed to produce an effect. The situation in ecology is somewhat different. Although statistically significant results may be more highly valued by ecologists, some null results also have considerable scientific interest. Few ecologists would argue, for instance, that studies demonstrating the absence of trophic cascades in certain ecosystems or of trade-offs between allocation to certain functions under some conditions are not worth publishing. Ecologists may thus be less inclined than their medical colleagues to dump non-significant results into the file drawer. Rather, ecologists tend to submit non-significant findings to journals with a lower impact factor, which have lower rejection rates. As a result, the chance of non-significant results being published may be higher in ecology than in medicine. On the other hand, the publication fate of studies with statistically significant results may suffer from the more ambitious attitude of ecologists towards such findings; these studies are usually submitted to top ecological journals, which have higher standards, more severe competition for space and higher rejection rates. This gives rise to a paradox: non-significant results may actually have a higher chance being published than statistically significant ones in ecology, i.e. non-significant results may be a blessing in disguise.

The current study represents the first attempt to obtain direct evidence of publication and dissemination bias in ecology by following up the publication fate of manuscripts. My findings indicate that publication bias against non-significant results is likely to be less pronounced in ecology than in medicine, and suggest possible reasons for such a difference. In other words, publication bias should not be automatically assumed to occur in one discipline because it has been found in another. The question remains, however, to what extent the publication fate of manuscripts from Finnish and Swedish doctoral dissertations reflects the situation in ecology in general. The present study should therefore be complemented by similar analyses of other sources, e.g. follow-ups of the publication fate of studies presented as abstracts at ecological conferences (e.g. ESA annual meetings) or of projects funded by different foundations.

Finally, the under-reporting of non-significant results is just one form of publication bias. Another common type of selective reporting is the bias against results which contradict the current theory on the subject. For instance, statistically significant results which indicate a different direction of the effect than that predicted by the favoured hypothesis may be published less frequently than studies supporting the theory. Several recent studies have demonstrated that the magnitude and direction of research findings may change with time following a shift in a common ecological paradigm (Alatalo et al. 1997, Simmons et al. 1999, Poulin 2000, Nykänen and Koricheva, in press). This type of publication bias may be more widespread in ecology than bias against non-significant results, and deserves further attention.

Acknowledgements – I am very grateful to the numerous Finnish and Swedish researchers who replied to my questionnaire and supplied me with information on the publication status of the manuscripts included in their dissertations. I apologise for inadvertently awakening painful memories of rejected or never submitted manuscripts. The manuscript benefited greatly from critical comments by Erkki Haukioja, Janne Kotiaho and Esa Lehikoinen and from linguistic corrections by Ellen Valle. The study was supported financially by the Academy of Finland.

References

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

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

Cohen, J. 1994. The Earth is round ($p < .05$). – Am. Psychol. 49: 997–1003.

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.

Easterbrook, P. J., Berlin, J. A., Gopalan, R. and Matthews, D. R. 1991. Publication bias in clinical research. – Lancet 337: 867–872.

Fernandez-Duque, E. 1996. Comparing and combining data across studies: alternatives to significance testing. – Oikos 79: 616–618.

Gurevitch, J., Morrow, L. L., Wallace, A. and Walsh, J. S. 1992. A meta-analysis of competition in field experiments. – Am. Nat. 155: 435–453.

Ioannidis, J. P. A. 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. and Petrie, M. 2001. Sexually selected traits and adult survival: a meta-analysis. – Q. Rev. Biol. 76: 3–36.

Jennions, M. D. and Møller, A. P. 2002. Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. – Proc. R. Soc. Lond. B 269: 43–48.
Jones, D. and Matloff, N. 1986. Statistical hypothesis testing in biology: a contradiction in terms. – J. Econ. Entomol. 79: 1156–1160.
Kotiaho, J. S. and Tomkins, J. L. 2002. Meta-analysis, can it ever fail? – Oikos 96: 551–553.
Lortie, C. J. and Dyer, A. R. 1999. Over-interpretation: avoiding the stigma of non-significant results. – Oikos 87: 183–184.
Misakian, A. L. and Bero, L. A. 1998. Publication bias and research on passive smoking: comparison of published and unpublished studies. – J. Am. Med. Assoc. 280: 250–253.
Møller, A. P. and Jennions, M. D. 2001. Testing and adjusting for publication bias. – Trends Ecol. Evol. 16: 580–586.
Murtaugh, P. A. 2002. Journal quality, effect size, and publication bias in meta-analysis. – Ecology 83: 1162–1166.
Nykänen, H. and Koricheva, J. Damage-induced changes in woody plants and their effects on insect herbivore performance: a meta-analysis. – Oikos in press.
Parkhurst, D. F. 2001. Statistical significance tests: equivalence and reverse tests should reduce misinterpretation. – BioScience 51: 1051–1057.
Poulin, R. 2000. Manipulation of host behaviour by parasites: a weakening paradigm? – Proc. R. Soc. Lond. B 267: 787–792.
Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. – Psychol. Bull. 86: 638–641.
Rotton, J., Foos, P. W., Vanmeek, L. and Levitt, M. 1995. Publication practices and the file drawer problem – a survey of published authors. – J. Soc. Behav. Pers. 10: 1–13.
Simmons, L.W., Tomkins, J. L., Kotiaho, J. S. and Hunt, J. 1999. Fluctuating paradigm. – Proc. R. Soc. Lond. B 266: 593–595.
Song, F., Eastwood, A. J., Gilbody, S. et al. 2000. Publication and related biases. – Health Technol. Assess. 4: 1–115.
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.
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.
Tonhasca, A. and Byrne, D. N. 1994. The effects of crop diversification on herbivorous insects: a meta-analysis approach. – Ecol. Entomol. 19: 239–244.
Yoccoz, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. – Bull. Ecol. Soc. Am. 72: 106–111.