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.

The "file drawer problem" of non-significant results: does it apply to biological research?

Ryan D. Csada, 82 Straub Cres., Regina, SK, Canada S4T 656. – Paul C. James, Wildlife Branch, Saskatchewan Environment and Resource Management, 3211 Albert St., Regina, SK, Canada S4S 5W6. – Richard H. M. Espie, Dept of Veterinary Anatomy, Western College of Veterinary Medicine, Univ. of Saskatchewan, Saskatoon, SK, Canada S7N 0W0.

We show that there appears to be a publication bias against non-significant results in the biological literature. We suggest reasons why non-significant results are not published, the implications of not publishing non-significant results, and why we need to correct the problem.

Scientists in a variety of disciplines have long noted that studies published in peer reviewed journals are a biased sample of the research actually conducted (Sterling 1959, McNemar 1960, Smart 1964, Bakan 1967, Rosenthal 1979, Vandenbroucke 1988, Greenland 1994). In paper after paper one finds that the author(s) has formulated and tested hypotheses and almost invariably has found statistical support (a significant result) for them. An extreme view of this is Rosenthal's (1979) "file drawer problem". He argues that journals are filled with the 5% of studies that show significant results while the file drawers in the lab contain the non-significant results (those that lack statistical support for the hypothesis) from the other 95% of studies done.

In a review of the psychological literature, Smart (1964) found that studies with non-significant results constituted only 9% of the total number of published papers. Based on data from unpublished sources (e.g., conference abstracts and Ph.D. theses), he showed that non-significant results are less likely to be published. For example, 30% of the theses he examined lacked significant results to support the main hypothesis.

To our knowledge, no one has quantitatively examined whether the file drawer problem occurs in biological research, although it is our impression, and the impression of many practising biologists, that it does. The purpose of our study was to determine if the proportion of non-significant results found in a range of biological journals is the same as that found in the psychological literature. Assuming that there is evidence the file drawer problem exists in biology, we will suggest reasons why non-significant results are not published as often as significant ones, the implications of not publishing non-significant results, and reasons why we need to correct the problem.

We reviewed 1812 papers from 43 biological journals (Table 1) that cover all biological subdisciplines. Only issues in the years 1989–1995 were considered and the number of issues considered varied from journal to journal. We reviewed a maximum of five randomly selected papers per issue. We defined a paper as having non-significant results if the main hypothesis was not statistically supported. Once one paper in an issue was found to have non-significant results, no further papers in that issue were reviewed.

Of the 1812 papers that we reviewed, 1201 statistically tested the main hypothesis. Of these papers, only 8.6% (103) presented non-significant results. This percentage is not statistically different from Smart's (1964) percentage ($\chi^2 = 0.02$, $p > 0.05$) for the psychological literature. We conclude from this that the file drawer problem occurs in biological research.

Possible reasons why non-significant results are not published

There are two likely reasons for the low proportion of non-significant results published in the biological literature. One is that these papers are not submitted for publication and a second is that they are submitted but rejected by the review process. There are many examples (e.g., Colwell and Hurtt 1994, Evans and Cain 1995) of references to unpublished non-significant results in the literature. It is likely that researchers con-

Table 1. List of journals examined and numbers of papers reviewed per journal (Total), number of those papers that statistically tested a hypothesis (Stat.), and number of these papers that lacked statistical support for the main hypothesis (Non-sig.).

	Total	Stat.	Non-sig.
American Journal of Botany	55	34	3
American Naturalist	30	26	0
Annals of Botany	28	19	2
Bulletin of Entomological Research	34	28	3
Canadian Entomologist	51	24	6
Canadian Field-Naturalist	21	9	3
Canadian Journal of Botany	89	60	1
Canadian Journal of Zoology	162	107	9
Cell	30	20	0
Condor	52	37	5
Copeia	32	17	4
Developmental Biology	30	17	0
Ecology	22	18	3
Environmental and Experimental Botany	52	36	3
Evolution	56	34	4
Fish Physiology and Biochemistry	27	20	2
Freshwater Biology	25	16	3
General and Comparative Endocrinology	37	30	2
Genome	30	12	0
Heredity	24	17	3
Ibis	30	26	3
Journal of Ecology	26	20	2
Journal of Endocrinology	26	22	3
Journal of Experimental Botany	85	54	3
Journal of Herpetology	58	36	7
Journal of Insect Physiology	57	48	2
Journal of Mammalogy	72	44	4
Journal of Reproduction and Fertility	21	15	2
Journal of Zoology	56	37	1
Molecular Biology and Evolution	19	14	1
Nature	20	14	0
Oecologia	69	63	6
Phycologia	30	2	0
Physiologia Plantarum	35	25	3
Physiological Zoology	57	42	2
Plant and Cell Physiology	30	11	0
Plant Cell and Environment	19	16	1
Planta	30	16	0
Science	20	14	0
The Auk	52	38	4
The Journal of General Physiology	30	14	0
The Journal of Parasitology	53	27	3
The New Phytologist	30	22	0
	1812	1201	103

sciously or unconsciously decide that non-significant results have a lower priority than significant ones.

The journals widely considered to be the most prestigious (e.g., Science, Nature, Cell, American Naturalist, Ecology, Evolution) tend to publish a lower proportion of non-significant papers than the whole (see Table 1). This is important in that one of the criteria that granting agencies use to assess the quality of research, and hence the amount of money awarded to investigators, is the prestige of the journals published in. Therefore, due to the lower likelihood of non-significant papers being published in the more prestigious journals, researchers may file non-significant results and concentrate on significant ones.

Many papers with non-significant results likely fail to be published because of editorial policies that increase the likelihood of outright rejection or subject them to more critical review. Although we could find no unambiguous statements regarding editorial policy towards non-significant results in biological journals, Smith (1956) during his editorship of the Journal of Abnormal Psychology wrote that:

> "Too often the failure to obtain anticipated results occurs in a context in which the procedures employed are of questionable adequacy and reliability. In such instances, interpretation is inherently so ambiguous that the report is properly judged to fall below the threshold of acceptable significance..."

This policy clearly states that non-significant results were to be reviewed more critically.

The argument that studies with significant results, as opposed to those with non-significant results, tend to support the validity of the procedures employed is

based on the assumption that only false null hypotheses are being tested. Significant results, however, far from confirming the validity of the procedures used, might merely confirm the researcher's expected outcome. Rosenthal and Fode (1963) have shown a researcher bias in that the researcher usually obtains the results expected or wanted. This raises the question as to how many studies reported to have significant results are nothing more than a researcher's expectations of such results. It also raises the question of how much this contributes to the publishing of fraudulent results.

Implications of not publishing non-significant results

Not publishing non-significant results can lead to false decisions as to the warrantability of hypotheses. If the reader is aware only of studies supporting the hypothesis while others failing to support it are unpublished, then an evaluation about its validity may be wrong. Ostensibly, the purpose of any science is to describe hypotheses relating the variables within its scope. This also implies that something is known of the variables that are unimportant and of the instances in which a particular hypothesis does not hold.

An example of the effects of not being aware of all studies done can be found in the recent increase in interest in meta-analysis (see Arnqvist and Wooster 1995); summarizing bodies of literature systematically and quantitatively with respect to significance levels and effect-size estimation. Rosenthal (1979) showed that small numbers of studies that are not very significant, even when the combined p is significant, may be misleading in that a few studies filed away instead of being published could change the combined significant result to a non-significant one.

There is also the trend in science to overlook the fact that in many studies, slight procedural variations, a different study population, or varied statistical analysis might lead to different results. This problem is compounded by the fact that too often a significant result is accepted and no subsequent effort is made to replicate the study. In his review of the psychological literature, Sterling (1959) found that none of the 294 papers he reviewed were replicates of a previous study. Although we did not collect similar data in our study, it is our impression that there is a similar lack of published replicates in the biological literature.

Potential solutions to the file drawer problem

To produce a journal of non-significant results would not be economical or practical. The easiest solution to the problem is for editors to right the ratio of non-significant to significant papers that are published. This is unlikely to happen, however, given the scarcity of journal space. A more feasible solution is that biologists submit non-significant results that are of biological significance and that each journal publish the abstracts of these papers or their titles with the authors' names and addresses. Either way, it is a situation that needs to be addressed. We believe that withholding non-significant results from publication introduces a serious bias in the biological literature and hence has a retrogressive effect on scientific development.

The idea that there is an apparent publication bias against non-significant results is not new, as many practising biologists have suspected this for years. Our study is, however, the first one to quantify support for this suspicion. We seriously hope that attitudes and practices can be changed to rectify the problem.

Acknowledgements – We would like to thank R. M. Brigham, D. J. Lieske and J. Lawton, for making valuable comments on earlier versions of the manuscript.

References

Arnqvist, G. and Wooster, D. 1995. Meta-analysis: synthesizing research findings in ecology and evolution. – Trends Ecol. Evol. 10: 236–240.
Bakan, D. 1967. On method. – Jossey-Bass, San Francisco.
Colwell, R. K. and Hurtt, G. C. 1994. Non-biological gradients in species richness and a spurious Rapoport effect. – Am. Nat. 144: 570–595.
Evans, J. P. and Cain, M. L. 1995. A spatially explicit test of foraging behavior in a clonal plant. – Ecology 76: 1147–1155.
Greenland, S. 1994. A critical look at some popular meta-analytic methods. – Am. J. Epidem. 140: 290–296.
Mcnemar, Q. 1960. At random: sense and nonsense. – Am. Psychol. 15: 295–300.
Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. – Psychol. Bull. 86: 638–641.
– and Fode, K. L. 1963. The psychology of a scientist: three experiments in experimenter bias. – Psychol. Rep. 12: 491–511.
Smart, R. G. 1964. The importance of negative results in psychological research. – Can. Psychol. 5: 225–232.
Smith, M. B. 1956. Editorial. – J. Abnorm. Soc. Psychol. 52: 1–4.
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.
Vandenbroucke, J. P. 1988. Passive smoking and lung cancer: a publication bias? – Brit. Med. J. 296: 391–392.