

Prevention is better than cure

Given the above considerations, plagiarism is better dealt with by concentrating on an instructional system of prevention. Many institutions have begun to implement courses on the responsible conduct of research, exploring a range of research integrity issues. Although this is a step in the right direction, such instruction benefits mainly students. Legions of seasoned researchers continue to operate without such instruction. Clearly, research integrity needs to be incorporated into continuing education targeted at this group.

Training on responsible conduct of research must go beyond the "high crimes" of fabrication, falsification, and plagiarism, all three of which are thankfully still relatively rare in science. Zigmond and Fischer have argued that misdemeanours make up the bulk of the ethical transgressions.¹⁰ Instruction on plagiarism should focus on the principles of ethical writing.¹¹ This approach assumes that each of our written works represents an implicit contract between us and our readers in which the reader assumes that, unless otherwise noted, we are the sole authors of the work, the words and ideas are our own, and the ideas, concepts and theories described are accurately and objectively represented to the best of our ability.

Since it is doubtful that we will ever eliminate plagiarism or other forms of intentional misconduct, we need to increase our ability to detect these transgressions and effectively prosecute and punish

offenders. Instruction in ethical writing indirectly touches on several other traditional forms of misconduct. Consequently, I believe that when we internalise and apply its basic principles they will significantly reduce plagiarism and generalise to other areas of scientific research and personal conduct.

I thank Maryellen Reardon for comments on an earlier draft of this article.

Competing interests: MR has written an online instructional resource, with support from the US Office Research Integrity, to help authors avoid plagiarism and other unethical writing practices.

- 1 Chalmers I. Role of systematic reviews in detecting plagiarism: case of Asim Kurjak. *BMJ* 2006;333:594-6.
- 2 Price JH, Dake JA, Islam R. Selected ethical issues in research and publication: perceptions of health education faculty. *Health Educ Behav* 2001; 28:51-64.
- 3 Julliard K. Perceptions of plagiarism in the use of other author's language. *Fam Med* 1994;26:356-60.
- 4 Roig M. When college students' attempts at paraphrasing become instances of potential plagiarism. *Psychol Rep* 1999;84:973-82.
- 5 Roig M. Plagiarism and paraphrasing criteria of college and university professors. *Ethics Behav* 2001;11:307-23.
- 6 Levin JR, Marshall HH. Publishing in the journal of educational psychology: reflections at midstream. *J Educ Psychol* 1993;85:3-6.
- 7 US Department of Health and Human Services, Office of Research Integrity. *ORI policy on plagiarism*. <http://ori.dhhs.gov/policies/plagiarism.shtml> (accessed 1 Jul 2006).
- 8 Hammond KW, Helbig ST, Benson CC, Brathwaite-Sketoe BM. Are electronic records trustworthy? Observations on copying, pasting and duplication. *AMA Annu Symp Proc* 2003;:269-73.
- 9 Bitros MH. *Advice to authors: getting published in Academic Emergency Medicine*, 2000. www.saem.org/inform/aempub.htm (accessed 6 Mar 2003).
- 10 Zigmond MJ, Fischer BA. Beyond fabrication and plagiarism: the little murders of everyday science. Commentary on "Six domains of research ethics." *Sci Eng Ethics* 2002;8:229-34.
- 11 Kolin FC. *Successful writing at work*. 6th ed. Boston: Houghton Mifflin, 2002. doi 10.1136/bmj.38946.501215.68

Evidence based medicine

The case of the misleading funnel plot

Joseph Lau, John P A Ioannidis, Norma Terrin, Christopher H Schmid, Ingram Olkin

Evidence based medicine insists on rigorous standards to appraise clinical interventions. Failure to apply the same rules to its own tools could be equally damaging

The advent of evidence based medicine has generated considerable interest in developing and applying methods that can improve the appraisal and synthesis of data from diverse studies. Some methods have become an integral part of systematic reviews and meta-analyses, with reviewers, editors, instructional handbooks, and guidelines encouraging their routine inclusion. However, the evidence for using these methods is sometimes lacking, as the reliance on funnel plots shows.

What is a funnel plot?

The funnel plot is a scatter plot of the component studies in a meta-analysis, with the treatment effect on the horizontal axis and some measure of weight, such as the inverse variance, the standard error, or the sample size, on the vertical axis. Light and Pillemer proposed in 1984: "If all studies come from a single underlying population, this graph should look like a funnel, with the effect sizes homing in on the true

underlying value as n increases. [If there is publication bias] there should be a bite out of the funnel."¹ Many meta-analyses show funnel plots or perform various tests that examine whether there is asymmetry in the funnel plot and directly interpret the results as showing evidence for or against the presence of publication bias.

The plot's wide popularity followed an article published in the *BMJ* in 1997.² That pivotal article has already received over 800 citations (as of December 2005) in the Web of Science. With two exceptions, this is more citations than for any other paper published by the *BMJ* in the past decade. The authors were careful to state many reasons why funnel plot asymmetry may not necessarily reflect publication bias. However, apparently many readers did not go beyond the title of "Bias in meta-analysis detected by a simple, graphical test."


Institute for Clinical Research and Health Policy Studies, Tufts-New England Medical Center, Boston, MA 02111, USA

Joseph Lau
professor
Norma Terrin
associate professor
Christopher H Schmid
professor of medicine

Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece

John P A Ioannidis
professor

continued over

 Details of *BMJ* systematic reviews mentioning funnel plots are on bmj.com

BMJ 2006;333:597-600

Statistics
Department,
Stanford University,
Palo Alto, USA
Ingram Olkin
professor
Correspondence to:
J Lau
jlau1@tufts-nemc.org

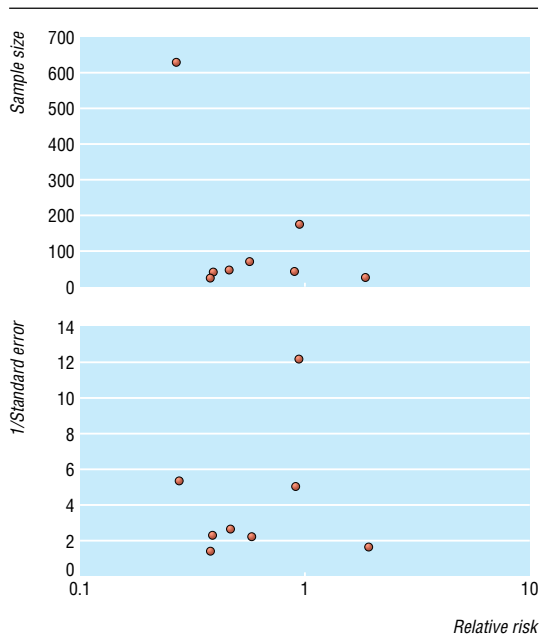


Fig 1 Effect of measure of precision (plotted on y axis) on appearance of funnel plots. Redrawn from Tang and Liu¹¹

The influential *Cochrane Handbook* adopts a relatively conservative view and acknowledges that there are problems with the concept.³ Yet it devotes more than four pages to this subject, far more than for any other test of bias and heterogeneity in meta-analysis. Whereas the widely accepted quality of reporting of meta-analysis (QUOROM) statement simply requires in its proposed checklist a description of “any assessment for publication bias,”¹⁴ its equally accepted counterpart for meta-analyses of observational studies in epidemiology (MOOSE) states that “methods should be used to aid in the detection of publication bias, eg, fail safe methods or funnel plots.”⁵ In an article on quantitative synthesis in systematic reviews commissioned by the American College of Physicians, even we advocated funnel plots and devoted a figure and considerable text to them.⁶

Use and abuse

Hand searching of the *BMJ* from July 2003 to June 2005 shows that funnel plots were mentioned in 20 of the 47 systematic reviews that included some quantitative data synthesis (see bmj.com for full details). In all 20 cases, the plots were mentioned specifically as tests for evaluating publication bias. Four of the 20 systematic reviews eventually did not perform these tests because they felt too few studies were available (maximum 3 to 10 per meta-analysis), one made no further mention besides the methods, and only one performed the tests and acknowledged that “the funnel plot may not detect publication bias when the number of studies is small.” The other 14 systematic reviews did not question the inferences from these tests and typically made categorical statements about conclusively finding or excluding publication bias with these methods.

A total of 34 meta-analyses had been evaluated with these methods: 14 of them had nine or fewer

studies and 18 of them had significant between-study heterogeneity; only five of the 34 meta-analyses had 10 or more studies and no significant between-study heterogeneity. Although 10 studies is not an adequate number for the funnel plot,⁷ we chose it as a cut-off to show that systematic reviewers did not meet even this liberal criterion.

Inconsistent interpretations were notable between different tests in the same meta-analysis. For example, in a meta-analysis of breastfeeding and blood pressure in later life,¹³ the results said: “evidence of such [publication bias] was provided by a funnel plot. The Egger test was significant ($P=0.033$), but not the Begg test ($P=0.186$)” and a figure shows “Begg’s funnel plot (pseudo 95% confidence limits).” Inconsistent interpretations were notable even for the same test between results and discussion. For example, in a meta-analysis of metformin for polycystic ovary syndrome¹⁴ the results stated that, “the funnel plot implies publication bias” whereas the discussion concluded that, “these data seem robust with no evidence of major publication bias.”

Accuracy of test

The evaluation of a methodological test is directly analogous to the evaluation of a clinical diagnostic test. Fryback and Thornbury have proposed a six level model for evaluating a diagnostic test.⁸ This provides a good discussion framework. The six expectations of a clinical diagnostic test are technical feasibility, diagnostic accuracy, diagnostic effect, treatment effect, effect on patient outcome, and societal effect. If the conclusions of evidence based medicine are based on poor tests, the negative effect eventually may be considerable. So we must examine closely at least the technical feasibility and diagnostic accuracy of these methods.

An evaluation of the technical feasibility of the funnel plot shows many problems that are difficult to solve. Strong empirical evidence exists that the appearance of the plot may be affected by the choice of the coding of the outcome (binary versus continuous),⁹ the choice of the metric (risk ratio, odds ratio, or logarithms thereof), and the choice of the weight on the vertical axis (inverse variance, inverse standard error, sample size, etc).^{10, 11} Figure 1 gives an example of how these choices can make a difference.

Even in the unlikely event that agreement is reached on what metric and what expression of weight to use on the axes, enormous uncertainty and subjectivity remains in the visual interpretation of the same plot by different researchers. Our team recently designed a survey to examine this question using simulated plots with or without publication bias.¹² The ability of researchers to identify publication bias using a funnel plot was practically identical to chance (53% accuracy).

Formal statistical tests may eliminate the subjectivity in visual inspection of asymmetry. Investigators commonly use the rank correlation test¹³ or one of many tests based on regression.^{2, 7, 10, 11, 14} The validity of these tests depends on assumptions often unmet in practice, however, and the choice of test introduces further subjectivity into the procedure. The methods theoretically require a considerable number of available studies, generally at least 30 for sufficient

power. But the number needed depends on the size of the studies and on the true treatment effect—for example, for an odds ratio of 0.67, even 60 studies are not adequate.⁷ Most meta-analyses of clinical trials, however, have far fewer studies. For instance, the average Cochrane meta-analysis has fewer than 10.¹⁵ Thus the tests typically have low power¹⁶ and may be inappropriate.

Even ignoring statistical concerns of power and choice of metric and weights, it is still unclear if funnel plots really diagnose publication bias. Strictly speaking, funnel plots probe whether studies with little precision (small studies) give different results from studies with greater precision (larger studies). Asymmetry in the funnel plot may therefore result not from a systematic under-reporting of negative trials but from an essential difference between smaller and larger studies that arises from inherent between-study heterogeneity. For example, small studies may focus on high risk patients, for whom the treatment is more effective because such patients have more events that could potentially be prevented¹⁷; or studies with small weight may generally have shorter follow-up and differ because the treatment effect decreases with time.¹⁸ Early studies may target different populations (with different effect sizes) than subsequent studies,¹⁹ and subsequent studies may be much larger, trying to test the concept on less selected patients. Variation in quality can affect

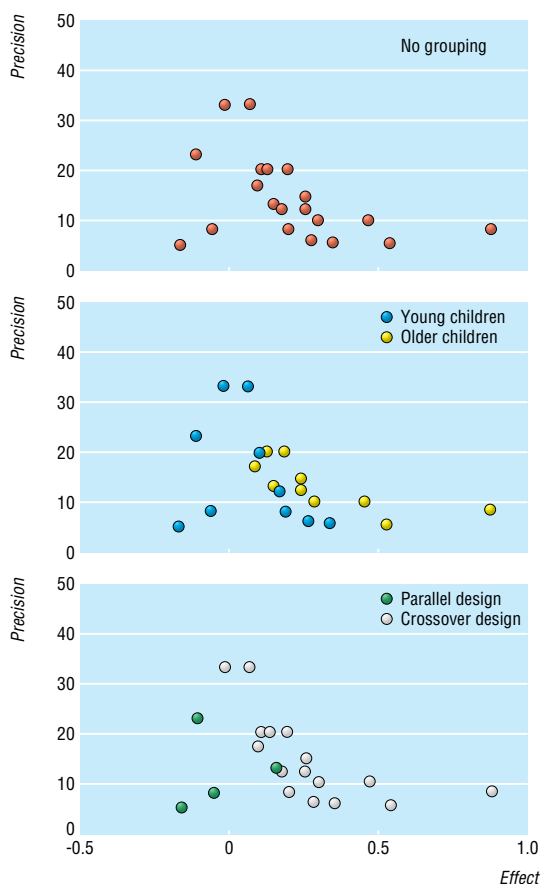


Fig 2 Effect of grouping of data on impression of funnel plot for meta-analysis of inhaled disodium cromoglicate that found significant heterogeneity²⁴

Summary points

Methods used by evidence based medicine should be evaluated with rigorous standards

The funnel plot is widely used in systematic reviews and meta-analyses as a test for publication bias

Asymmetry of the funnel plot, either visually interpreted or statistically tested, does not accurately predict publication bias

Inappropriate or misleading use of funnel plot tests may do more harm than good

the shape of the funnel plot, with smaller, lower quality studies showing greater benefit of treatment.²⁰

Heterogeneity may sometimes be both statistically and clinically obvious—that is, studies may be examining different questions.²¹ Yet the authors of a meta-analysis, such as the one investigating the relation between garlic consumption and cancer,²¹ may still pool all studies together when it comes to the funnel plot, even though they have analysed them separately for the main analysis. In other cases, it may not be possible to identify a source for the existing heterogeneity.²² Simulation studies of funnel plots have found that bias may be incorrectly inferred if studies are heterogeneous.^{21 23}

For example, figure 2 shows the funnel plot for a meta-analysis of inhaled disodium cromoglicate as maintenance therapy in children with asthma.²⁴ The authors found both statistical and clinical heterogeneity, yet they published a funnel plot (fig 2, top), stating: “Studies with low precision and negative outcome are under-represented, indicating publication bias.” Grouping the studies according to age of participants (middle) and study design (bottom) creates a different impression.

Finally, we have no gold standard against which to compare the results of funnel plot tests. A true standard measure of publication bias would require prospective registries of trials with detailed knowledge of which studies have been published and which are unpublished. It would then be feasible to test whether tests of publication bias capture accurately the presence of unpublished studies and whether one variant performs better than others. Given that efforts for study registration have only recently started,²⁵ this evaluation is currently difficult. Although a large number of alternative tests for publication bias exist,²⁶ none has been validated against a standard.

Prevention of bias

In conclusion, evidence based methods, including the funnel plot, should be evidence based. If treatment decisions are made on the basis of misleading methodological tests, the costs to patients and society could be high. Decisions guided by the easy assurance of a symmetrical funnel plot may overlook serious bias. Equally, it may be misleading to discredit and

abandon valid evidence simply because of an asymmetrical funnel plot. The prevention of publication bias is much more desirable than any diagnostic or corrective analysis.

Contributors and sources: The authors have worked for a long time on methodological research in systematic reviews and meta-analyses. The idea was generated by JL and expanded by the other authors. The manuscript was written by JPAI and JL and commented on critically by the other authors. All authors approved the final version. JL is the guarantor.

Funding: Supported in part by AHRQ grant R01 HS10254.

Competing interests: None declared.

- 1 Light RJ, Pillemer DB. *Summing up. The science of reviewing research*. Cambridge, MA: Harvard University Press, 1984.
- 2 Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
- 3 Alderson P, Green S, Higgins JPT, eds. *Cochrane reviewers' handbook 4.2.2* [updated March 2004]. In: *Cochrane Library*, Issue 1. Chichester: Wiley, 2004.
- 4 Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999;354:1896-900.
- 5 Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008-12.
- 6 Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997;127:820-6.
- 7 Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001;20:641-54.
- 8 Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making* 1991;11:88-94.
- 9 Hrobjartsson A, Gotzsche PC. Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. *J Intern Med* 2004;256:91-100.
- 10 Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;54:1046-55.
- 11 Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 2000;53:477-84.

- 12 Terrin N, Schmid CH, Lau J. In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. *J Clin Epidemiol* 2005;58:894-901.
- 13 Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088-99.
- 14 Peters JL, Sutton AJ, Jones DR, Abram KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006;295:676-80.
- 15 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
- 16 Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;53:1119-29.
- 17 Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med* 1998;17:1923-42.
- 18 Ioannidis JP, Cappelleri JC, Lau J, Skolnik PR, Melville B, Chalmers TC, et al. Early or deferred zidovudine therapy in HIV-infected patients without an AIDS-defining illness. *Ann Intern Med* 1995;122:856-66.
- 19 Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative meta-analyses. *Proc Natl Acad Sci USA* 2001;98:831-6.
- 20 Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*. London: BMJ Books, 2001.
- 21 Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Stat Med* 2003;22:2113-26.
- 22 Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol* 2004;57:683-97.
- 23 Schwartz G, Antes G, Schumacher M. Inflation of type I error rate in two statistical tests for the detection of publication bias in meta-analyses with binary outcomes. *Stat Med* 2002;21:2465-77.
- 24 Tasche MJ, Uijen JH, Bernsen RM, van der Woude JC. Inhaled disodium cromoglycate as maintenance therapy in children with asthma: a systematic review. *Thorax* 2000;55:913-20.
- 25 DeAngelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *JAMA* 2004;292:1363-4.
- 26 Rothstein HR, Sutton AJ, Borenstein M, eds. *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Sussex: John Wiley and Sons, 2005.

(Accepted 16 June 2006)



bmjlearning.com

Everything bad is good for you

There is no shortage of complaints from doctors these days about dumbing down. Consultants complain that they can't trust their registrars any more, and registrars say that foundation year doctors don't get enough experience and don't know what they're doing. And grumpy old men everywhere complain that exam results don't mean what they used to. But what objective evidence do we have that medicine is being dumbed down? The truth is that there is little such evidence that doctors are better or worse than they were a generation ago. Mowa and Alcolado have shown a large increase in the number of final year medical students being awarded honours, but this could reflect either smarter students or lower standards.¹

The reason that we have such little evidence is that the context in which doctors practise medicine has changed. Early medical curriculums concentrated on doctors' retention of knowledge, but more recent ones emphasise the importance of skills, attitudes, and behaviours. But if we look beyond the world of medicine then we see strong evidence that IQ scores of the population at large are rising, a phenomenon dubbed the Flynn effect. Because of the Flynn effect, IQ tests are periodically readjusted, making them harder.² Why is this happening? Steven Johnson, author of *Everything Bad Is Good for You*, attributes it to

the rise of the popular media and computer games. Could it be that the hours we spend on the internet and playing games are teaching us vital skills such as understanding complex systems and spatial awareness? Are more sophisticated games enabling us to learn strategic thinking?

The interactive case histories presented by BMJ Learning are designed to help you learn knowledge and skills in a safe environment—without anyone looking over your shoulder. One of the latest is on acne: it starts by explaining the basics of diagnosis and treatment and goes on to deal with more complex clinical problems such as how drug induced acne differs from acne vulgaris. It also looks at other issues such as how to advise a patient with mild acne but who is very distressed by it. To have a look, go to www.bmjlearning.com.

Kieran Walsh *clinical editor, BMJ Learning*
(bmjlearning@bmjgroup.com)

- 1 Mowa R, Alcolado J. Medical degrees with honours; the 'dumbing down' of undergraduate examinations? *Clin Med* 2004;4:381-3.
- 2 Kanaya T, Scullin MH, Ceci SJ. Rising IQ scores in British children: recent evidence. *J Child Psychol Psychiatry* 1992;33:1241-7.