

- 25 Volpe N, Gesini A, Collini S, Grassano MT, Guzzon D, Marino MR, et al. Single dose ondansetron for prevention of postoperative nausea and vomiting. Results from the Italian Multicentre Ondansetron Study. *Drug Invest* 1994;8:67-72.
- 26 Di Donato A, Collini S, Brauneis S, Pinto G. Impiego clinico dell'ondansetron nella prevenzione in dose singola e nel trattamento del PONV in oftalmochirurgia. *Minerva Anestesiol* 1994;60(suppl 1/1-2):61-5.
- 27 Collini S, Di Donato A, Brauneis S, Turkiewicz AM, Zaccheo AR, Pinto G. Il vomito postoperatorio in oftalmochirurgia: profilassi, trattamento e ritrattamento in uno studio randomizzato in doppio cieco con ondansetron. *Acta Anaesth Ital* 1995;46:49-56.
- 28 Taddei F, Sarli L, Pietra N. L'impiego di ondansetron in chirurgia videolaparoscopica. *Acta Anaesth Ital* 1995;46:43-8.
- 29 Volpe N, Gesini A, Guzzo D, Benini A, Vitale G, Marino MR, et al. Prevenzione della nausea e del vomito postoperatorio con una dose endovenosa singola di ondansetron. *Minerva Anestesiol* 1994;60(suppl 1/1-2):121-3.
- 30 Lundberg GD. Statement by the international committee of medical journal editors on duplicate or redundant publication. *JAMA* 1993;270:2495.
- 31 Frei F. Richtlinien für Reanimationsmassnahmen in der Pädiatrie. *Monatsschr Kinderheilkd* 1996;144:727-36.
- 32 Zideman D, Bingham R, Beattie T, Bland J, Blom C, Bruins-Strassen M, et al. Guidelines for paediatric life support. Paediatric Life Support Working Party of the European Resuscitation Council. *BMJ* 1994;308:1349-55.
- 33 Aspinall RL, Goodman NW. Denial of effective treatment and poor quality of clinical information in placebo controlled trials of ondansetron for postoperative nausea and vomiting: a review of published trials. *BMJ* 1995;311:844-6.
- 34 Angell M, Reiman AS. Redundant publication. *N Engl J Med* 1989;320:1212-4.
- 35 Tramèr M, Moore A, McQuay H. Propofol anaesthesia and postoperative nausea and vomiting: quantitative systematic review of randomized controlled studies. *Br J Anaesth* 1997;78:247-55.
- 36 Leizorovicz A, Haugh MC, Boissel JP. Meta-analysis and multiple publication of clinical reports. *Lancet* 1992;340:1102-3.
- 37 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. *JAMA* 1995;273:408-12.
- 38 Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. *JAMA* 1983;249:1743-5.
- 39 International Steering Committee of Medical Editors. Uniform requirements for manuscripts submitted to biomedical journals. *BMJ* 1979;i:532-5.
- 40 Paxton LD, McKay AC, Mirakhor RK. Prevention of nausea and vomiting after day case gynaecological laparoscopy. A comparison of ondansetron, droperidol, metoclopramide and placebo. *Anaesthesia* 1995;50:403-6.
- 41 Paech MJ, Pavy TJG, Evans SF. Single-dose prophylaxis for postoperative nausea and vomiting after major abdominal surgery: ondansetron versus droperidol. *Anaesth Intens Care* 1995;23:548-54.
- 42 Campbell C, Miller DD. Failure of ondansetron to control postoperative nausea and vomiting in ambulatory surgical patients. *Am J Anesth* 1995;22:81-6.
- 43 Davis A, Krige S, Moyes D. A double-blind randomized prospective study comparing ondansetron with droperidol in the prevention of emesis following strabismus surgery. *Anaesth Intens Care* 1995;23:438-43.
- 44 Watcha MF, White PF. Post-operative nausea and vomiting: do they matter? *Eur J Anaesth* 1995;12(suppl 10):18-23.
- 45 Markham A, Sorkin EM. Ondansetron. An update of its therapeutic use in chemotherapy-induced and postoperative nausea and vomiting. *Drugs* 1993;45:931-52.
- 46 Dershwitz M. Antiemetic drugs. In: White PF, ed. *Ambulatory anesthesia and surgery*. London: Saunders, 1997: 441-53.
- 47 Kovac AL, Pearman MH, Khalil SN, Scuderi PE, Joslyn AF, Prillaman BA, et al. Ondansetron prevents postoperative emesis in male outpatients. *J Clin Anesth* 1996;8:644-51.

(Accepted 21 July 1997)

Publication bias: evidence of delayed publication in a cohort study of clinical research projects

Jerome M Stern, R John Simes

Abstract

Objectives: To determine the extent to which publication is influenced by study outcome.

Design: A cohort of studies submitted to a hospital ethics committee over 10 years were examined retrospectively by reviewing the protocols and by questionnaire. The primary method of analysis was Cox's proportional hazards model.

Setting: University hospital, Sydney, Australia.

Studies: 748 eligible studies submitted to Royal Prince Alfred Hospital Ethics Committee between 1979 and 1988.

Main outcome measures: Time to publication.

Results: Response to the questionnaire was received for 520 (70%) of the eligible studies. Of the 218 studies analysed with tests of significance, those with positive results ($P < 0.05$) were much more likely to be published than those with negative results ($P \geq 0.10$) (hazard ratio 2.32 (95% confidence interval 1.47 to 3.66), $P = 0.0003$), with a significantly shorter time to publication (median 4.8 v 8.0 years). This finding was even stronger for the group of 130 clinical trials (hazard ratio 3.13 (1.76 to 5.58), $P = 0.0001$), with median times to publication of 4.7 and 8.0 years respectively. These results were not materially changed after adjusting for other significant predictors of publication. Studies with indefinite conclusions ($0.05 \leq P < 0.10$) tended to have an even lower publication rate and longer time to publication than studies with negative results (hazard ratio 0.39

(0.13 to 1.12), $P = 0.08$). For the 103 studies in which outcome was rated qualitatively, there was no clear cut evidence of publication bias, although the number of studies in this group was not large.

Conclusions: This study confirms the evidence of publication bias found in other studies and identifies delay in publication as an additional important factor. The study results support the need for prospective registration of trials to avoid publication bias and also support restricting the selection of trials to those started before a common date in undertaking systematic reviews.

Introduction

In evaluating the effectiveness of treatments, the highest level of evidence is believed to be obtained from a systematic review or meta-analysis of all randomised controlled trials. However, evidence of treatment effectiveness even from randomised trials can still be biased owing to a number of factors: bias within individual trials that are not properly randomised or not analysed according to intention to treat¹; bias in selecting trials for inclusion in a meta-analysis,^{2,3} particularly when only published trials are included⁴; and bias in selecting treatment questions after examining the data.⁵ In particular, if trials with a positive effect of treatment are more likely to be published, a review limited only to published trials would give a more positive effect of treatment than a review based on all trials (published and unpublished).

National Health and Medical Research Council Clinical Trials Centre, Mallet Street Campus, University of Sydney, Sydney NSW 2006, Australia

Jerome M Stern, research associate
R John Simes, director

Correspondence to: Professor Simes john@ctc.trials.su.oz.au

BMJ 1997;315:640-5

The selection of trials for inclusion in a systematic review may be biased if selection is restricted to published trials (publication bias),⁴ to trials published in English language journals (language bias),⁶ to trials published in prestigious journals,⁷ or to trials cited by other authors such as in review articles (reference bias).⁸ Furthermore, even if all relevant trials are eventually published, the selection of trials may still be biased if it is restricted to trials that are published early.

Evidence for publication bias has been shown in many studies,⁹⁻¹⁹ and particularly in three cohort studies of protocols submitted to institutional ethics committees.²⁰⁻²² All three studies showed a significantly greater likelihood for trials with significant results to be published than for those with negative results. These studies did not, however, examine whether there was a delay in publication for trials that were eventually published.

We determined the extent of publication bias for studies submitted to an Australian ethics committee and whether publication was delayed for studies with negative results in comparison with those with positive results.

Methods

Between September 1979 and December 1988, 801 submissions were received by the Royal Prince Alfred Hospital Ethics Committee for proposed medical research. Each was accompanied by a protocol providing a detailed outline of the proposed research, from which was obtained information on approval, details of the research design, the planned sample size, and the nature of any intervention. Research design was classed primarily as observational study, clinical trial, or non-trial experiment according to the definitions of Easterbrook et al.²⁰

In July 1992 the principal investigator for each study was asked to complete a questionnaire providing information on the current status; starting date, closure of recruitment, and finishing date; sample size reached; the nature of funding (none, pharmaceutical, government, other (external), or other (internal)); the rating of scientific importance of the study; the status and date of the most recent analysis; the main research questions posed by the study at the outset; the results for the main research questions; and the publication status and date of initial publication as an article in a peer reviewed journal.

Eligible studies

Eligible studies were defined as single studies approved by the Royal Prince Alfred Hospital Ethics Committee between September 1979 and December 1988 with more than one patient and with protocol information available.

Classification of study outcome

For quantitative studies, in which the main study outcome was assessed by using statistical methods with tests of significance, outcome was classed as significant ($P < 0.05$), as showing a non-significant trend ($0.05 \leq P < 0.10$), or as non-significant or null ($P \geq 0.10$). Examples of such studies include those comparing treatments and epidemiological studies examining evidence of association for risk factors.

For qualitative studies, in which the main study outcome was assessed subjectively by the principal investigator, the study was classed as showing striking, important and definite, or unimportant and negative findings. Such studies included uncontrolled phase 2 clinical trials examining response rate to treatment and studies with descriptive statistics of a population.

Analysis

Potential factors predictive of time to publication were examined in a Cox regression analysis, in which time to publication was the time from the date that the study was approved by the ethics committee to the date of first publication. Unpublished studies were censored at the date the questionnaire was completed; studies that were in press were analysed as if published on the date of completion of the questionnaire. Studies for which no analysis had yet been undertaken and two studies whose early findings had first been published before approval from this ethics committee were excluded from the survival analysis. Funding sources were compared in two ways to allow comparison with earlier studies²⁰⁻²¹: pharmaceutical *v* non-pharmaceutical and external *v* internal or none. We made no allowance for multiple comparisons. With the exception of the investigator's rating of scientific importance of the study, which we judged to be largely influenced by study results, all other factors were examined in a multivariate Cox regression to determine the relative importance of study results on time to publication adjusted for any other significant factors. Quantitative and qualitative studies were examined in the same

Table 1 Characteristics of eligible studies

| Characteristic | No (%) of studies |
|---|-------------------|
| All studies (n=748) | |
| Research design: | |
| Observational study | 165 (22) |
| Clinical trial | 418 (56) |
| Non-trial experiment | 165 (22) |
| Pilot study: | |
| Pilot study | 39 (5) |
| Non-pilot study | 709 (95) |
| Study groups: | |
| Uncontrolled | 392 (52) |
| Controlled (non-concurrent, historical) | 5 (1) |
| Controlled (concurrent) | 351 (47) |
| Clinical trials (n=418) | |
| Clinical trial phase: | |
| 1 | 18 (4) |
| 2 | 178 (43) |
| 3 | 221 (53) |
| 4 | 1 (<1) |
| Pilot study: | |
| Pilot study | 22 (5) |
| Non-pilot study | 396 (95) |
| Study groups: | |
| Uncontrolled | 142 (34) |
| Controlled (historical) | 3 (1) |
| Controlled (concurrent), non-randomised | 76 (18) |
| Controlled, randomised | 197 (47) |
| Placebo controlled study: | |
| Placebo controlled | 81 (19) |
| Not placebo controlled | 337 (81) |
| Blinded treatment (single or double blind): | |
| Blinded | 107 (26) |
| Non-blinded | 311 (74) |

model by also including a factor for method of assessment of study outcome. A subsidiary logistic regression analysis was performed to permit a direct comparison of the results of this study with earlier cohort studies of clinical research projects, which were analysed using odds ratios for publication.

The association between the various study characteristics and study status was examined by calculating the χ^2 statistic for an $r \times c$ contingency table.

Results

Inclusion criteria

A total of 801 protocols for 810 separate studies were submitted for approval between September 1979 and December 1988. Of these, 748 separate studies in 741 submissions were approved during this period and included in the study. We excluded studies that had not been approved during this period, 19 duplicate study submissions, four submissions that were not formal studies, five studies on one patient alone, and five studies whose submission protocols or relevant data were not available.

Study characteristics

Table 1 shows the characteristics of the eligible studies and the clinical trials subgroup.

Questionnaires were completed for 520 (70%) of the eligible studies, with a larger proportion of questionnaires completed for the more recently approved studies ($P=0.0007$). Apart from this, the returned questionnaires were representative of the total sample of eligible studies for all factors for which data were available from the study protocols.

Table 2 shows the status of the studies at the time of questionnaire completion. The main reasons for not starting in 100 studies were lack of funding (45), lack of support from colleagues (9), departure of staff (6), and impractical methods (6). Seventy studies (of which 50 were clinical trials) were abandoned after starting, mainly because of difficulties with patient accrual (23), funding problems (6), and technological problems (6) and because the early results that had been obtained were negative (6, all clinical trials).

Association between study characteristics and completion of study

Analysis of the association between the various study characteristics and completion of the 520 studies for which questionnaires were completed showed a higher rate of not starting for non-trial experiments (31/114; 27%) than for observational studies (23/129; 18%) or for clinical trials (43/277; 16%), and a higher completion rate for observational studies (85/129; 66%) than for clinical trials (156/277; 56%) or for non-trial experiments (46/114; 40%) ($P<0.0001$). There was also a far higher completion rate for studies that were undertaken as part of a degree compared with those that were not (85/107; 79% *v* 202/396; 51%) ($P<0.0001$). Single centre studies were more likely never to have started or to have been abandoned than were multicentre studies (123/337; 36% *v* 28/163; 17%) ($P<0.0001$). Studies with a small sample size reached had a higher rate of abandonment (<100 (40/231; 17%) *v* ≥ 100 (0/141)) ($P<0.0001$), which

Table 2 Stage of studies for 520 questionnaires with completed questionnaires

| Stage | No of studies |
|-------------------|---------------|
| Not started: | |
| Abandoned | 97 |
| Yet to be started | 3 |
| Started: | |
| Abandoned | 70 |
| Still in progress | 63 |
| Completed | 287 |

Table 3 Outcome (results) of quantitative and qualitative studies

| | All analysed studies | Clinical trials |
|--------------------------|----------------------|-----------------|
| Quantitative studies | | |
| Total No | 218 | 130 |
| Results: | | |
| Significant | 146 (67) | 76 (58) |
| Non-significant trend | 20 (9) | 15 (12) |
| Non-significant | 52 (24) | 39 (30) |
| Qualitative studies | | |
| Total No | 103 | 37 |
| Results: | | |
| Striking | 27 (26) | 6 (16) |
| Important and definite | 59 (57) | 18 (49) |
| Unimportant and negative | 17 (17) | 13 (35) |

may simply reflect being abandoned because of poor accrual. Clinical trials with a more rigorous study design were far more likely to be completed (randomised (87/130; 67%) *v* non-randomised (69/147; 47%), $P=0.002$; placebo controlled (31/44; 70%) *v* non-placebo controlled (125/233; 54%), $P=0.06$; and blinded (43/58; 74%) *v* non-blinded (113/219; 52%), $P=0.004$).

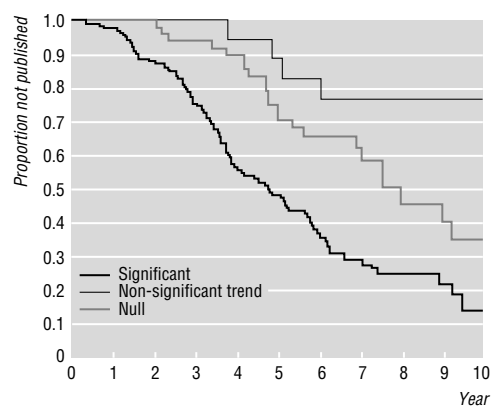
Study outcome and method of analysis

Of the 520 studies with completed questionnaires, 321 had had analysis undertaken with results available and were included in further analysis of the association between study outcome and time to publication.

Table 3 summarises study outcome and method of analysis. 67% of quantitative studies, had a significant result for the primary research question, whereas only 26% of qualitative studies were in the most positive category. Accordingly, quantitative and qualitative studies were considered separately in subsequent analyses.

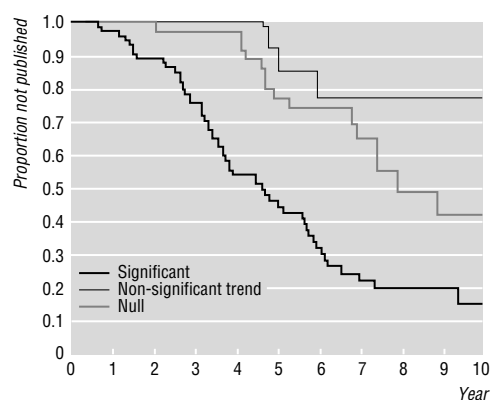
Study characteristics and publication

Table 4 shows the hazard ratios for publication for selected variables. Factors predictive of publication were significant results; research design using non-trial experiments; a high scientific importance rating of the study by the investigator; external funding; and studies with non-comparative study groups and clinical trials that were non-randomised. Other variables tested and found not to be significant were whether the study was undertaken as part of a degree; the number of data collection sites (single *v* multicentre); pharmaceutical funding; the research department undertaking the study; the year of study approval; the classification of study outcome (qualitative *v* quantitative); and, for clinical trials, placebo control and blinding. For the 218 quantitative studies, positive studies were much more likely to be published than negative studies (hazard ratio 2.32 (95% confidence interval 1.47 to 3.66),



| No at risk | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|-----|-----|----|----|----|---|---|---|---|---|----|
| Significant | 144 | 127 | 77 | 36 | 15 | 2 | | | | | |
| Non-significant trend | 20 | 20 | 19 | 14 | 4 | 3 | | | | | |
| Null | 52 | 52 | 46 | 24 | 10 | 7 | | | | | |

Fig 1 Proportion of quantitative studies not published, according to type of results



| No at risk | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|----|----|----|----|---|---|---|---|---|---|----|
| Significant | 75 | 67 | 38 | 19 | 8 | 2 | | | | | |
| Non-significant trend | 15 | 15 | 15 | 11 | 3 | 2 | | | | | |
| Null | 39 | 39 | 37 | 20 | 8 | 6 | | | | | |

Fig 2 Proportion of quantitative clinical trials not published, according to type of results

$P=0.0003$) This finding was even stronger for the subgroup of 130 quantitative clinical trials (hazard ratio 3.13 (1.76 to 5.58), $P=0.0001$).

Figure 1 shows the survival curve plots for time to publication for quantitative studies and figure 2 those for quantitative clinical trials. The median time to publication was 4.82 (3.87 to 5.72) years for studies with significant results, 7.99 (6.91 to ∞) years for studies with null results, and not reached for studies with indefinite conclusions. The median time to publication was 4.69 (3.75 to 5.72) years for clinical trials with significant results and 7.99 (7.02 to ∞) years for clinical trials with null results.

The survival curve plots for sample size (not shown) show that studies with a larger sample size (≥ 100) continued to be published right until the end of the study period, at which time about 90% of studies and clinical trials had been published, compared with only about 64% of studies and clinical trials with a sample size < 100 .

The odds ratio for publication quantitative studies was 2.66 (1.32 to 5.35) ($P=0.003$) for significant compared with null studies. For quantitative clinical trials it was 4.19 (1.71 to 10.32) ($P=0.0004$).

Figure 3 shows the results of multivariate analysis. For quantitative studies the adjusted hazard ratio for publication for positive compared with negative studies was 2.34 (1.47 to 3.73) ($P=0.0004$); for studies with intermediate results it was 0.43 (0.15 to 1.24) ($P=0.12$). For quantitative clinical trials the adjusted hazard ratio for publication for positive compared with negative trials was 3.29 (1.84 to 5.90) ($P=0.0001$); for trials with intermediate results it was 0.50 (0.14 to 1.74) ($P=0.27$). For quantitative studies the adjusted odds ratio for publication for positive compared with negative studies was 2.93 (1.49 to 5.74); for studies with intermediate results it was 0.34 (0.17 to 0.67). For quantitative clinical trials the adjusted odds ratio for publication for positive compared with negative trials was 4.57 (1.96 to 10.63); for trials with intermediate results it was 0.44 (0.10 to 1.91).

Table 4 Risk factors for time to publication using univariate Cox regression analysis

| Characteristic | No not published | No published | Hazard ratio (95% CI) | P value |
|--------------------------------|------------------|--------------|-----------------------|---------|
| All studies (n=321) | | | | |
| Outcome: | | | | |
| Quantitative studies: | | | | |
| Null | 29 | 23 | 1.00 | <0.0001 |
| Non-significant trend | 16 | 4 | 0.39 (0.13 to 1.12) | |
| Significant | 47 | 99 | 2.32 (1.47 to 3.66) | |
| Qualitative studies: | | | | |
| Negative and unimportant | 8 | 9 | 1.00 | 0.25 |
| Important and definite | 24 | 35 | 1.22 (0.59 to 2.55) | |
| Striking | 8 | 19 | 1.56 (0.70 to 3.45) | |
| Research design: | | | | |
| Observational study | 42 | 48 | 1.00 | 0.008 |
| Clinical trial | 71 | 96 | 1.03 (0.73 to 1.46) | |
| Non-trial experiment | 19 | 45 | 1.74 (1.15 to 2.61) | |
| Funding: | | | | |
| Internal or none | 59 | 58 | 1.00 | 0.02 |
| External | 73 | 133 | 1.45 (1.06 to 1.99) | |
| Pilot study: | | | | |
| Non-pilot study | 120 | 183 | 1.00 | 0.07 |
| Pilot study | 12 | 6 | 0.51 (0.23 to 1.15) | |
| Study groups: | | | | |
| Non-comparative | 62 | 100 | 1.00 | 0.02 |
| Comparative | 70 | 89 | 0.72 (0.54 to 0.96) | |
| Scientific importance: | | | | |
| Low | 74 | 51 | 1.00 | <0.0001 |
| High | 58 | 138 | 2.35 (1.69 to 3.26) | |
| Sample size: | | | | |
| <100 | 80 | 100 | 1.00 | 0.44 |
| ≥ 100 | 52 | 82 | 1.12 (0.84 to 1.51) | |
| Clinical trials (n=167) | | | | |
| Outcome: | | | | |
| Quantitative trials: | | | | |
| Null | 24 | 15 | 1.00 | <0.0001 |
| Non-significant trend | 12 | 3 | 0.46 (0.13 to 1.59) | |
| Significant | 21 | 55 | 3.13 (1.76 to 5.58) | |
| Qualitative trials: | | | | |
| Negative and unimportant | 4 | 9 | 1.00 | 0.39 |
| Important and definite | 7 | 11 | 0.97 (0.40 to 2.36) | |
| Striking | 3 | 3 | 0.53 (0.14 to 1.95) | |
| Clinical trial phase: | | | | |
| 1 | 3 | 5 | 1.15 (0.45 to 2.91) | 0.13 |
| 2 | 22 | 40 | 1.00 | |
| 3 | 46 | 51 | 0.67 (0.44 to 1.02) | |
| 4 | 0 | 0 | | |
| Study groups: | | | | |
| Non-comparative | 19 | 33 | 1.00 | 0.02 |
| Comparative (non-randomised) | 10 | 17 | 1.07 (0.60 to 1.93) | |
| Randomised | 42 | 46 | 0.61 (0.39 to 0.96) | |

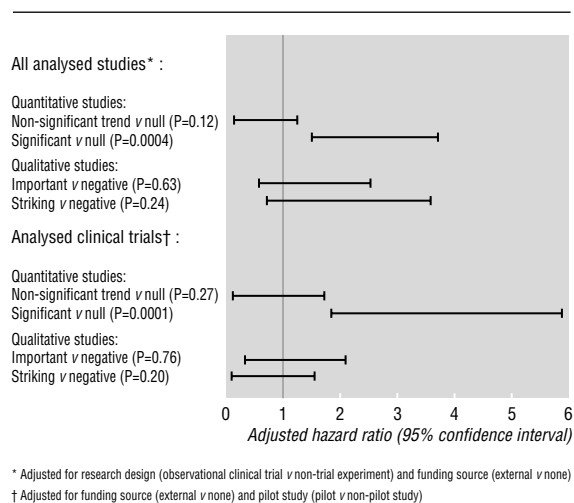


Fig 3 Multivariate analysis of outcome in studies with analysed results, according to type of study and results

Quantitative studies with significant study outcomes were much more likely to be submitted for publication than studies with a null study outcome (114/146; 78% v 28/52; 54%) ($P = 0.0009$). By contrast, 130 of the 148 quantitative studies submitted had been published or were in press, and the publication rate for those studies with significant outcomes (99/114; 87%) were similar to those with null results (23/28; 82%) ($P = 0.54$).

Discussion

We found publication bias, after allowing for confounding factors, in a cohort of studies that were approved by an Australian ethics committee, and our results confirm those of American and British studies.²⁰⁻²² To our knowledge, ours is the first study to show the delay in publication of studies with negative results. The evidence was definite for quantitative studies, with a publication rate 2.3 times greater for studies with significant compared with null results. Within the subgroup of clinical trials, the evidence of publication bias was even stronger, with a publication rate 3.3 times greater. These figures represent a higher estimated publication rate at 5 years of 52% v 27% respectively for quantitative studies, and 54% v 23% respectively for clinical trials. Although the trend for qualitative studies was not significant ($P = 0.21$), this may relate to the smaller number of qualitative studies. Other factors in a multivariate analysis that were associated with publication were research design (observational studies being more likely to be published than clinical trials and non-trial experiments) and source of funding (externally funded studies being more likely to be published). The impact of study results on publication was essentially the same after adjusting for these factors. The impact of other factors on publication rate was examined only for the studies with analysed results. Interestingly, clinical trials that used better study designs (randomisation, placebo control, or blinding) were not published more often than other studies but were more likely to be completed and hence more likely to be analysed than other studies. Consequently, these factors may still positively influence the publication rate when all studies, whether analysed or not, are considered.

Other studies reporting on publication bias have analysed quantitative and qualitative studies as a single

group,²⁰⁻²² but we considered them separately. This was because the categories were not comparable in terms of outcomes and because there was evidence of interaction between these subgroups and the degree of publication bias ($P = 0.02$).

The validity of our study was not affected by including studies with only interim findings because the findings of publication bias were not materially altered by excluding these studies. Consequently, the bias relates to delay in publication of studies with negative results rather than just the premature publication of positive results.

Quantitative studies and clinical trials with an indefinite study outcome ($0.05 \leq P < 0.10$) were less likely to be published than studies and clinical trials with a non-significant study outcome ($P \geq 0.10$). A similar trend was seen in the study of Easterbrook et al, who found an adjusted odds ratio of 0.61 (0.26 to 1.59) for non-significant trend studies compared with null studies.²⁰ Thus an indefinite study outcome may be even more likely than a definitely negative outcome to deter researchers from submitting their studies for publication, or editors from accepting them.

The data also imply that publication bias is primarily due to the failure of researchers to submit negative studies for publication rather than to the failure of journal editors to publish them after they have been submitted. However, we cannot exclude that the failure of investigators to submit negative studies for publication may partly relate to editorial bias; previous experience in the submission of negative studies may have conditioned them to expect rejection of such studies for publication.

Sample size

Berlin et al have argued that large scale studies are eventually published and that a meta-analysis restricted to large studies may be free of publication bias.²³ They also argue that small studies should be excluded from meta-analyses on the grounds that they have greater random fluctuation in their estimates and that they may be more subject to bias than large studies.²³ We found that studies with a sample size ≥ 100 continue to be published, with more than 90% published at the end of the study period. However, at any time point, since some studies are still in progress, there is always publication bias due to a delay in the publication of negative studies, provided that the results from large studies are not systemically different from other studies and that large negative studies are subject to the same delay in publication as other studies. We did not find significant interaction between study results and sample size; in addition, publication bias was still evident when analysis was restricted to the 103 quantitative studies with a sample size ≥ 100 , with a hazard ratio for publication of positive compared with negative studies of 2.00 (1.09 to 3.66) ($P = 0.02$). Hence, a strategy that excludes small trials is not a solution to the problem of publication bias.

Implications for the conduct of meta-analysis

The use of meta-analysis in medical research is becoming more common; its results seem to be precise and convincing, and it is beginning to have an impact on clinical practice and on the planning of future research. Consequently, it is important that modest dif-

ferences in outcome can be validly attributed to the effect of treatment rather than bias. Although an increasing number of meta-analyses are based on controlled clinical trials, a computerised literature search for 1982-9 by Easterbrook et al²⁰ found that about two thirds of meta-analyses are based on observational studies, which are more prone to bias than controlled clinical trials.

Our results have four important implications for the conduct of meta-analysis.

Firstly, the studies used in any meta-analysis should not comprise solely published studies. This is not an unbiased sample, and meta-analysis based on published studies may result in bias in favour of positive results.

Secondly, meta-analyses should be restricted to studies that have started before a certain date. This avoids the problem of the delayed but eventual publication of studies with negative results. To include all studies, regardless of starting date, is likely to result in a selection that is biased in favour of studies with positive results.

Thirdly, retrospective collection of data is difficult—the lower questionnaire completion rate for studies that were undertaken in the early years confirms the findings of Hetherington et al.²⁴ Meta-analyses that are based on retrospective data collection are likely to be subject to significant selection bias.

Finally, restricting meta-analyses to large scale studies, even if they are all eventually published, does not by itself solve the problem of publication bias, since it does not take into account the delay in publishing negative studies.

Prospective registration—a solution to the problem of publication bias

A strategy that provides a solution to publication bias is to register prospectively all trials before their results are known and to select trials from such a registry when undertaking any systematic review.⁴ This approach has been supported by the Cochrane Collaboration and has led to the statement on the need for prospective registration of all controlled trials developed at the second annual Cochrane colloquium in 1994. For evidence based medicine, the support of government bodies to ensure that mechanisms and funding are provided to ensure universal registration is essential and can be argued to be in the interests of developing the highest level of evidence. There are also strong ethical arguments on behalf of participating patients to ensure that the results of all studies are eventually published.^{25 26} The mechanisms for the establishment of universal registration already exist—it has been mandatory since 1985 in Australia, as in many other countries, for all research projects on human subjects to be approved by an institutional ethics committee.²⁷ The identification of studies for national registration at the time of initial approval by each institutional ethics committee is therefore a fairly simple matter. The need for universal prospective registration was recognised over 10 years ago and is being actively pursued by the Cochrane Collaboration, but it is yet to be implemented. With the increasing recognition of the importance of evidence based medicine, the establishment of universal prospective registration is an important and urgent priority.

Key messages

- This retrospective cohort study of clinical research projects confirms the findings of publication bias found in previous studies
- Delay in the publication of studies with negative results has been identified as an additional important factor in publication bias
- With the recognised importance of evidence based medicine, these results have important implications for the selection of studies included in systematic reviews
- Prospective registration of clinical research projects will avoid many of the problems associated with publication bias
- However, it is also important to restrict inclusion in systematic reviews to studies started before a certain date to allow for the delay in completing studies with negative results

We thank the ethics committee and investigators from the Royal Prince Alfred Hospital for their support, and Davina Ghersi for insightful comments on the manuscript.

Funding: This study was supported by a unit grant from the National Health and Medical Research Council, Australia.

Conflict of interest: None.

- 1 Newell DJ. Intention to treat analysis: implications for quantitative and qualitative research. *Int J Epidemiol* 1992;21:837-41.
- 2 Chalmers TC, Levin H, Sacks HS, Reitman D, Berrier J, Nagalingham R. Meta-analysis of clinical trials as a scientific discipline: control of bias and comparison with large co-operative trials. *Stat Med* 1987;6:315-25.
- 3 Counsell CE, Clarke MJ, Slattery J, Sandercock PA. The miracle of DICE therapy for acute stroke: fact or fictional product of subgroup analysis? *BMJ* 1994;309:1677-81.
- 4 Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;4:1529-41.
- 5 Simes RJ. Prospective meta-analysis of cholesterol-lowering studies: the prospective pravastatin pooling (PPP) project and the Cholesterol Treatment Trialists (CTT) Collaboration. *Am J Cardiol* 1995;76:122-6C.
- 6 Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;347:363-6.
- 7 Frank E. Authors' criteria for selecting journals. *JAMA* 1994;272:163-4.
- 8 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;295:654-6.
- 9 Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *J Am Stat Assoc* 1959;54:30-4.
- 10 Juhl E, Christensen E, Tygstrup N. The epidemiology of the gastrointestinal randomised clinical trial. *N Engl J Med* 1977;296:20-2.
- 11 Begg CB, Pocock LJ, Freedman L, Zelen M. State of the art in comparative cancer clinical trials. *Cancer* 1987;60:2811-5.
- 12 Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986;1:55-8.
- 13 Gøtzsche PC. Methodology and overt and hidden bias in reports in 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Controlled Clin Trials* 1989;10:31-56.
- 14 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;295:654-6.
- 15 Mahoney MJ. Publication prejudices: an experimental study of confirmatory bias in the peer review system. *Cog Ther Res* 1977;1:161-75.
- 16 Koren G, Klein N. Bias against newspaper reports of medical research. *JAMA* 1991;266:1824-6.
- 17 Kotelchuk D. Asbestos research: winning the battle but losing the war. *Health/PAC Bulletin* 1974;61:1-27.
- 18 Sommer B. The file drawer effect and publication rates in menstrual cycle research. *Psychol Women Q* 1987;11:233-42.
- 19 Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H. Publication bias and clinical trials. *Controlled Clin Trials* 1987;8:343-53.
- 20 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
- 21 Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results. *JAMA* 1992;267:374-8.
- 22 Dickersin K, Min Y-I. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993;Document No 50.
- 23 Berlin JA, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989;84:381-92.
- 24 Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Paediatrics* 1989;84:374-80.
- 25 Chalmers I, Gray M, Sheldon T. Handling scientific fraud: prospective registration of health care research would help. *BMJ* 1995;311:262.
- 26 Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996;313:1390-3.
- 27 National Health and Medical Research Council. *NHMRC Statement on human experimentation and supplementary notes*. Canberra: NHMRC, 1985. (Accepted 26 August 1997)