Health Technology Assessment 2003; Vol. 7: No. 1

How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study

M Egger
P Jüni
C Bartlett
F Holenstein
J Sterne

Health Technology Assessment
NHS R&D HTA Programme

How to obtain copies of this and other HTA Programme reports.

An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study

M Egger[1,2*]
P Jüni[2,3]
C Bartlett[2]
F Holenstein[1]
J Sterne[2]

[1] Department of Social and Preventive Medicine, University of Berne, Switzerland
[2] Medical Research Council Health Services Research Collaboration, Department of Social Medicine, University of Bristol, UK
[3] Department of Rheumatology and Clinical Immunology, University of Berne, Switzerland

[*] Corresponding author

Declared competing interests of the authors: none

Published January 2003

This report should be referenced as follows:

Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;**7**(1).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA website (see opposite).

Related publications:

Egger M, Smith DG, O'Rourke K. Rationale, potentials and promise of systematic reviews. In: Egger M, Smith DG, Altman, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001.

Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;**31**:115–23.

Egger M, Ebrahim S, Smith DG. Where now for meta-analysis [editorial]. *Int J Epidemiol* 2002;**31**:1–5.

Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;**21**:1513–24.

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

Although the National Coordinating Centre for Health Technology Assessment (NCCHTA) commissions research on behalf of the Methodology Programme, it is the Methodology Group that now considers and advises the Methodology Programme Director on the best research projects to pursue.

The research reported in this monograph was funded as project number 97/18/05.

The views expressed in this publication are those of the authors and not necessarily those of the Methodology Programme, HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

M

Contents

List of abbreviations i

Executive summary iii

1 Introduction 1
The dissemination of research findings 2
Garbage in – garbage out: the importance of study quality 4
Examining for bias: funnel plots 5
Rationale 6
Objectives 6

2 Methods 9
Selection of study sample 9
Definitions 9
Assessment of methodological quality.......... 10
Data extraction and replication of pooled estimates 10
Analysis 10

3 Results 13
Identification and characteristics of eligible meta-analyses 13
The impact of unpublished trials 15
The impact of trials published in languages other than English 21
The impact of trials published in journals not indexed in MEDLINE 25
The impact of trial quality: concealment of allocation.. 32
The impact of trial quality: double-blinding.. 36
Sensitivity analyses using logistic regression 40

4 Discussion 43
Principal findings 43
Strengths and weaknesses............................. 44
The present study in context 45
Implications for research 47
Recommendations for future research 48

Acknowledgements 51

References 53

Appendix 1 Meta-analyses included in one or more of the analyses......................... 57

Appendix 2 Bibliographic references for meta-analyses included in analyses 61

Appendix 3 Reviews meeting inclusion criteria, but lacking trials with characteristics of interest 67

Health Technology Assessment reports published to date 69

Methodology Group 75

HTA Commissioning Board 76

List of abbreviations

CCT	controlled clinical trial
CCTR	Cochrane Controlled Trials Register
CDSR	Cochrane Database of Systematic Reviews
CI	confidence interval
CRD	(NHS) Centre for Reviews and Dissemination
DARE	Database of Abstracts of Reviews of Effectiveness
IPD	individual participant data
LMW	low molecular weight
OR	odds ratio
RCT	randomised controlled trial
SD	standard deviation

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

The inclusion of an unbiased sample of relevant studies is central to the validity of systematic reviews and meta-analyses. Time-consuming and costly literature searches, which cover the grey literature and all relevant languages and databases, are normally recommended to prevent reporting biases. However, the size and direction of these effects is unclear at present. There may be trade-offs between timeliness, cost and the quality of systematic reviews.

Objectives

- To examine the characteristics of clinical trials that are difficult to locate (unpublished trials, trials published in languages other than English, trials published in journals not indexed in the MEDLINE database) and of trials of lower quality (inadequate/unclear concealment of treatment allocation, not double-blind).
- To compare within meta-analyses the treatment effects reported in trials that are difficult to locate with trials that are more accessible, and of trials of lower with trials of higher quality.
- To assess the impact of excluding trials that are difficult to locate and of trials of lower quality on pooled effect estimates, *p*-values and the shape of funnel plots.

Methods

Data sources

The following sources were searched for relevant meta-analyses:

- eight medical journals that regularly publish systematic reviews (handsearch)
- systematic reviews published in the Cochrane Database of Systematic Reviews
- systematic reviews included in the Database of Abstracts of Reviews of Effectiveness
- *Health Technology Assessment* (handsearch).

Study selection

Meta-analyses of therapeutic or preventive interventions that were based on comprehensive literature searches and which combined the binary outcomes of at least five controlled clinical trials were included. Comprehensive literature searches were defined as follows:

- the search was not restricted to the English language literature
- the Cochrane Controlled Trials Register or at least two other electronic databases (such as MEDLINE or EMBASE) had been searched
- at least one indicator of searches for unpublished trials was present (e.g. searches of conference proceedings or contacts with licensing bodies).

Data extraction

Trial reports were classified as published journal articles if they had been published as full or short reports, editorials or letters in a regular or supplementary issue of a journal. Language was assessed using the SERLINE journals database, and published trials were classified according to whether or not they had been published in a MEDLINE-indexed journal. Quality assessment was restricted to trials included in Cochrane reviews.

Data synthesis

Meta-analyses that were able to contribute to the analysis in question were included. For example, only meta-analyses that contained both published and unpublished trials were included in the analyses addressing the impact of publication bias. Within each meta-analysis pooled effect estimates were calculated separately for the trials that are difficult to locate and the remaining trials, applying the same statistical model used by the original authors. For each meta-analysis a ratio of the pooled estimates was derived. A weighted average for all these ratios was calculated using random-effects meta-analysis. The percentage change in the pooled effect estimate which occurred when trials that are difficult to locate were excluded, was also calculated and changes in *p*-values and the impact on the shape of the funnel plot (using a regression method to measure funnel plot asymmetry) were examined.

Results

- A total of 159 systematic reviews met the inclusion criteria but not all included trials that are

difficult to locate. Comparisons of treatment effects were based on the following:
– unpublished versus published (60 meta-analyses)
– other languages versus English (50 meta-analyses)
– non-indexed versus MEDLINE-indexed (66 meta-analyses).

Analyses of trial quality were based on:
– inadequately concealed/unclear versus adequately concealed (39 meta-analyses)
– not double-blind versus double-blind (45 meta-analyses).

- The importance of trials that are difficult to locate appears to vary across medical specialities. For example, unpublished trials are particularly prevalent in oncology whereas trials published in languages other than English and trials published in sources not indexed in MEDLINE are important in psychiatry, rheumatology and orthopaedics. A large proportion of trials of complementary medicine are difficult to locate.
- Unpublished trials show less beneficial effects than published trials whereas non-English language trials and non-indexed trials tend to show larger treatment effects.
- Trials that are difficult to locate tend to be smaller and of lower methodological quality than trials that are easily accessible and published in English.
- Trials with inadequate or unclear concealment of allocation show more beneficial effects than adequately concealed trials. Similarly, open trials tend to be more beneficial than double-blind trials.
- In the majority of meta-analyses exclusion of trials with inadequate or unclear concealment and trials without double-blinding led to a change towards less beneficial treatment effects, which was often substantial.
- Including unpublished trials reduces funnel plot asymmetry whereas the inclusion of trials published in languages other than English and of non-indexed trials increases the degree of asymmetry in the funnel plot. The impact of trials of lower methodological quality on the funnel plot is substantial for trials with inadequate or unclear concealment of allocation.

Conclusions

Systematic reviews that are based on a search of English language literature that is accessible in the major bibliographic databases will often produce results that are close to those obtained from reviews based on more comprehensive searches that are free of language restrictions. We recommend that when planning a review, investigators should consider the type of literature search and the degree of comprehensiveness that are appropriate for the review in question, taking into account budgetary and time constraints.

The finding that trials which are difficult to locate are often of lower quality raises the worrying possibility that rather than preventing bias through extensive literature searches, bias could be introduced by including trials of low methodological quality. We believe that in situations where resources are limited, thorough quality assessments should take precedence over extensive literature searches and translations of articles.

Our results confirm that the funnel plot and the regression method to assess funnel plot asymmetry are useful to detect 'small-study effects', the tendency for smaller studies in a meta-analysis to show larger treatment effects.

Recommendations for future research

- The importance of trials that are difficult to locate appears to vary not only between conventional and complementary medicine but also within conventional medicine. Further research is required to clarify this issue.
- Future studies should prospectively compare the results from rapid reviews that are restricted to the English language with meta-analyses based on extensive searches without language restrictions.
- The inclusion or exclusion of trials of low methodological quality has a substantial impact on results and conclusions from systematic reviews and meta-analyses. Further methodological research into markers of trial quality in different areas of medicine is required.

Chapter 1
Introduction

Systematic, continuously updated reviews of the best evidence that is available on the benefits and risks of medical interventions can valuably inform decision-making in clinical practice and public health medicine, identify areas in which further research is needed and guide allocation of resources.[1] The term 'systematic review' denotes any type of review that has been prepared using strategies to avoid bias and that includes a material and methods section. A systematic review may or may not include meta-analysis, "… a statistical analysis which combines or integrates the results of several independent clinical trials considered by the analyst to be 'combinable'".[2] The best evidence is provided by randomised controlled trials (RCTs), and important additional insights are often gained when results from individual trials are combined.

Meta-analysis is not an infallible tool, however, and several examples exist of meta-analyses where the findings were later contradicted by large randomised trials (*Figure 1*).[3,4] Also, systematic reviews addressing the same issue have reached opposite conclusions.[13] For example, one group reviewing trials comparing low molecular weight (LMW) heparins and standard heparin in the prevention of thrombosis following surgery concluded that "LMW heparins seem to have a higher benefit to risk ratio than unfractionated heparin in preventing perioperative thrombosis",[14] while another group of reviewers considered that "there is at present no convincing evidence that in general surgery patients LMW heparins, compared with standard heparin, generate a clinically important improvement in the benefit to risk ratio".[15] Contrary to one of the central objectives of systematic reviews, to reduce uncertainty, such contradictory reports may contribute to the confusion, a situation that has arisen in other fields, for example when assessing calcium antagonists or cholesterol-lowering interventions in hypertension and coronary heart disease, or mammography for breast cancer screening.[16–18]

Two factors are considered to be central to the validity of meta-analyses and systematic reviews:

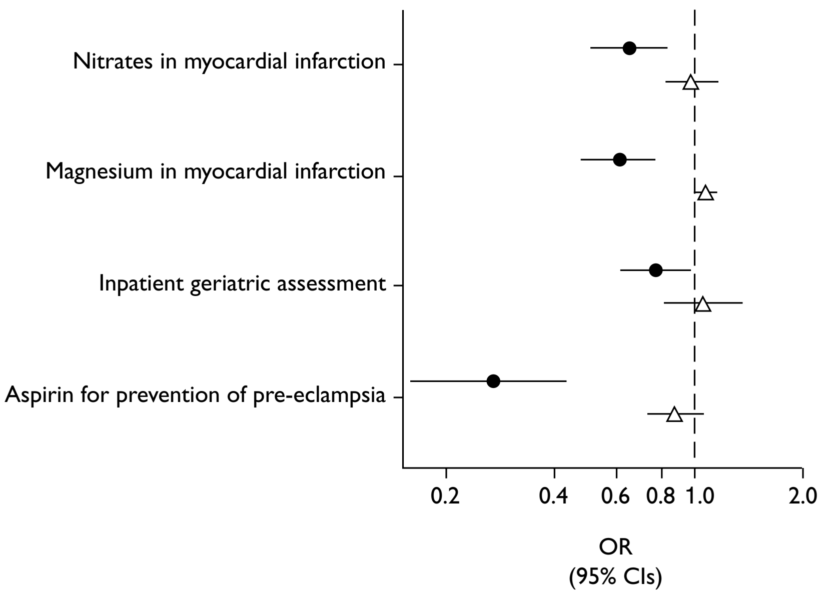

FIGURE 1 *Results from discordant pairs of meta-analyses of small trials and single large trials: Effect of nitrates[5,6] and magnesium[7,8] on mortality in acute myocardial infarction, effect of inpatient geriatric assessment on mortality in the elderly,[9,10] and effect of aspirin on the risk of pre-eclampsia.[11,12] Reproduced from Egger* et al.*[22] by permission of BMJ Books (●, meta-analysis; △, single large trial)*

- the inclusion of all relevant studies or of an unbiased sample of relevant studies
- the methodological quality of component studies.

The dissemination of research findings

The dissemination of research findings is not a dichotomous event but a continuum ranging from the sharing of draft papers among colleagues, presentations at meetings, and published abstracts to papers in journals that are indexed in the major bibliographic databases.[19] It has long been recognised that only a proportion of research projects ultimately reach publication in an indexed journal thus becoming easily identifiable for systematic reviews.[20] Scherer and co-workers[21] showed that only about half of abstracts presented at conferences are later published in full. Dickersin and Meinert examined the fate of doctoral theses from the Department of Epidemiology at Johns Hopkins University School of Hygiene and Public Health and found that one-third of graduates had not published a single article from their thesis.[22] Similar results were found for trainees in public health in the UK.[23] Four separate studies followed up research proposals approved by ethics committees or institutional review boards in Oxford,[24] Sydney,[25] and at the Johns Hopkins School of Medicine[26] and School of Hygiene and Public Health in Baltimore.[26] For each cohort of research proposals the principal investigators were contacted several years later in order to determine the publication status of each completed study. The rates of full publication as journal articles ranged from 49% to 67%. Similarly, 20% of trials funded by the National Institutes of Health and 45% of trials on HIV infection funded by the National Institute of Allergy and Infectious Diseases were still unpublished several years after completion.[27–29]

The fact that a substantial proportion of studies remains unpublished even a decade after the study had been completed and analysed must be of concern as potentially important information remains hidden from reviewers. Worse, the dissemination of research findings is not a random process; rather it is strongly influenced by the nature and direction of results. Statistically significant, ‘positive’ results that indicate that a treatment works are not only more likely to be published, but also more likely to be published rapidly, more likely to be published in English, more likely to be published more than once, and more likely to be cited by others. The different types of reporting biases are defined in *Table 1*. The importance of publication, language and database bias is the focus of the present report and these biases are discussed in more detail below.

Publication bias

In a 1979 article on *The ‘file drawer problem’ and tolerance for null results*, Rosenthal described a gloomy scenario where “the journals are filled with the 5% of the studies that show Type I errors, while the file drawers back at the lab are filled with the 95% of the studies that show non-significant (e.g. $p > 0.05$) results.”[30] The file drawer problem has long been recognised in the social sciences: a review of psychology journals found that of 294 studies published in the 1950s, 97.3% rejected the null hypothesis at the 5% level.[31] The study was recently updated and complemented with three medical journals (*New England Journal of Medicine,*

TABLE 1 *Reporting biases: definitions. Reproduced from Egger* et al.[22] *by permission of BMJ Books*

Type of reporting bias	Definition
Publication bias*	The **publication** or **non-publication** of research findings, depending on the nature and direction of the results
Multiple (duplicate) publication bias	The **multiple** or **singular** publication of research findings, depending on the nature and direction of the results
Language bias*	The publication of research findings **in a particular language**, depending on the nature and direction of the results
Database bias*	The inclusion or exclusion of research findings from widely used bibliographic databases such as MEDLINE, depending on the nature and direction of the results
Citation bias	The **citation** or **non-citation** of research findings, depending on the nature and direction of the results
Outcome reporting bias	The **selective reporting** of some outcomes but not others, depending on the nature and direction of the results

* *Biases examined in the present research project*

American Journal of Epidemiology and the *American Journal of Public Health*).[32] Little had changed in the psychology journals (95.6% reported significant results) and a high proportion of statistically significant results (85.4%) was also found in the general medical and public health journals. Similar results have been reported for emergency medicine[33] and, more recently, in the area of alternative and complementary medicine.[34,35] It is thus possible that studies which suggest a beneficial treatment effect are published, while a mass of data pointing the other way remains unpublished. In this situation, a systematic review of the published trials could identify a spurious beneficial treatment effect, or miss an important adverse effect of a treatment. In the field of cancer chemotherapy such publication bias has been demonstrated by comparing the results from studies identified in a literature search with those contained in an international trials registry.[36,37] In cardiovascular medicine, investigators, who in 1980 found an increased death rate among patients with acute myocardial infarction treated with a class I anti-arrhythmic drug, dismissed it as a chance finding and did not publish their trial at the time.[38] Their findings would have contributed to a more timely detection of the increased mortality that has since become known to be associated with the use of class I anti-arrhythmic agents.[39]

The proportion of all hypotheses tested for which the null hypothesis is truly false is of course unknown and surveys of published results can therefore only provide indirect evidence of publication bias. Convincing, direct evidence is available from the four cohort studies of proposals submitted to ethics committees mentioned earlier,[24–26] from cohorts of trials funded by the National Institutes of Health,[27] trials submitted to licensing authorities,[40] trials conducted by multicentre trial groups in the domain of HIV infection[28] and from analyses of trial registries.[36] In all these studies publication was more likely if effects were large and statistically significant. A meta-analysis of the four ethics committee cohorts is shown in *Figure 2*. The odds of publication were 2.4 times greater if results were statistically significant. Other factors such as the design of the study, its methodological quality, study size and number of study centres, were not consistently associated with the probability of publication.[29]

Studies continued to appear in print many years after approval by the ethics committee. Among proposals submitted to the Royal Prince Alfred Hospital Ethics Committee in Sydney, an estimated 85% of studies with significant results compared with 65% of studies with null results had been published after 10 years.[25] The median time to publication was 4.8 years for studies with significant results and 8.0 years for studies with null results. Similarly, trials conducted by multicentre trial groups in the field of HIV infection in the USA appeared on average 4.2 years after the start of patient enrolment if results were statistically significant but took 6.4 years to be published

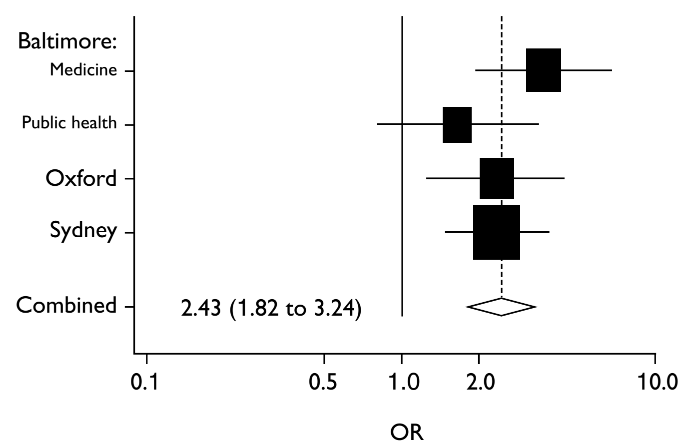

***FIGURE 2** Meta-analysis of six studies examining the association of the statistical significance of results (p < 0.05 versus other) with the probability of publication among research proposals submitted to ethics committees. Meta-analysis was by fixed effects model. Adapted from Egger* et al.[22]

if the results were negative.[28] These findings indicate that time-lag bias,[28] may be introduced in systematic reviews even in situations when most or all trials will eventually be published. Trials with positive results will dominate the literature and introduce bias for several years until the negative, but equally important, results finally appear.

Language bias

Reviews are often exclusively based on trials published in English. For example, among 36 meta-analyses reported in leading English language general medical journals from 1991 to 1993, 26 (72%) had restricted their search to studies reported in English.[41] Investigators working in a non-English speaking country will, however, publish some of their work in local journals.[42] It is conceivable that authors are more likely to report in an international, English language journal if results are positive whereas negative findings are published in a local journal. This has been demonstrated for the German language literature.[43] When comparing pairs of articles published by the same first author, 63% of trials published in English had produced significant ($p < 0.05$) results compared with 35% of trials published in German (*Figure 3*). Bias could thus be introduced in meta-analyses exclusively based on English language reports.[41,44]

MEDLINE bias

A substantial proportion of journals are not indexed in MEDLINE, the most widely used bibliographic database. Studies that are published in non-indexed journals are therefore to some extent hidden from reviewers and meta-analysts. This is of particular importance for the dissemination of the findings from research in less-developed countries. Whereas most of the major west-European journals that are published in languages other than English are indexed, this is not the case for journals published in the less-developed countries. Among the 3000–4000 journals indexed in major databases, only about 2% are from less-developed countries.[45] For example, one survey[46] found that only 30 journals (0.8%) out of a total of 3861 journals indexed in MEDLINE are published in India, despite the fact that India is the developing country with the largest research output and that the medical research is published in English.[47] A minority of trials will be published in indexed local or international journals but it is likely that results and other characteristics differ between these two groups. For example, it is possible that trials with statistically significant results are more likely to be published in an indexed journal whereas trials with null results are published in non-indexed journals.

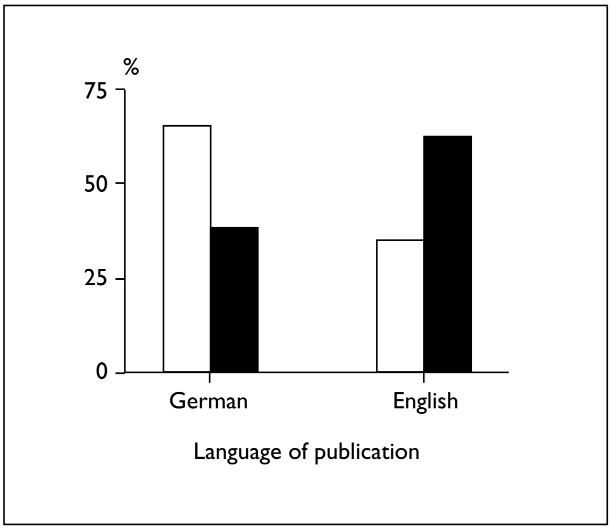

FIGURE 3 *Language bias. The proportion of controlled trial with statistically significant results was higher among reports published in English. Analysis based on 40 pairs of trials published by the same author, with one trial published in English and the other in German. Reproduced from Egger* et al.[22] *by permission of BMJ Books (□, non-significant (p ≥ 0.05); ■, significant (p < 0.05))*

Garbage in – garbage out: the importance of study quality

The quality of component trials is of crucial importance: if the 'raw material' is flawed, then the findings of reviews of this material may also be compromised. Clearly, the trials included in systematic reviews and meta-analyses should ideally be of high methodological quality and free of bias so that any differences in outcomes observed between groups of patients can confidently be attributed to the intervention under investigation. The biases that threaten the validity of clinical trials relate to:

- systematic differences in the patients' characteristics at baseline (**selection bias**)
- unequal provision of care apart from the treatment under evaluation (**performance bias**)
- biased assessment of outcomes (**detection bias**), and
- bias due to exclusion of patients after they have been allocated to treatment groups (**attrition bias**).[48]

Several studies[49–51] have recently attempted to quantify the impact these biases have on the results of controlled clinical trials (CCTs). For example, Schulz and co-workers[49] assessed the methodological quality of 250 trials from 33 meta-analyses from the Cochrane Pregnancy and Childbirth Database and examined the association between dimensions of trial quality and estimated treatment effects. Compared with trials in which authors reported adequately concealed treatment allocation, failure to prevent foreknowledge of treatment allocation or unclear concealment were associated, on average, with an exaggeration of treatment effects by 30–40%. Trials that were not double-blind also yielded larger effects.

The methodological quality of trials could be associated with publication status and language of publication. If unpublished trials or trials published in languages other than English are of lower quality than trials published in English, then their inclusion could in fact introduce bias in systematic reviews and meta-analyses. Moher and co-workers compared the quality of 133 RCTs published in English with 96 trials published in French, German, Italian or Spanish and found no overall difference using a quality score.[44]

Examining for bias: funnel plots

The smaller a study, the larger the treatment effect necessary for the results to be declared statistically significant. In addition, the greater investment of money and time in larger studies means that they are more likely to be of high methodological quality and published even if their results are negative. Bias in a systematic review may therefore be evident in an association between treatment effect and study size, and may be shown graphically in **funnel plots**: scatter plots of the treatment effects estimated from individual studies on the horizontal axis against study size or standard error on the vertical axis.[3,52,53] The name 'funnel plot' is based on the fact that the precision in the estimation of the underlying treatment effect will increase as the sample size of component studies increases. Effect estimates from small studies will therefore scatter more widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias the plot will resemble a symmetrical inverted funnel (see *Figure 4a*).

Bias, for example because smaller studies showing no statistically significant effects (open circles in *Figure 4a*) remain unpublished, will lead to an asymmetrical appearance of the funnel plot with a gap in the right bottom side of the graph (*Figure 4b*). In this situation the combined effect from a meta-analysis will indicate more beneficial

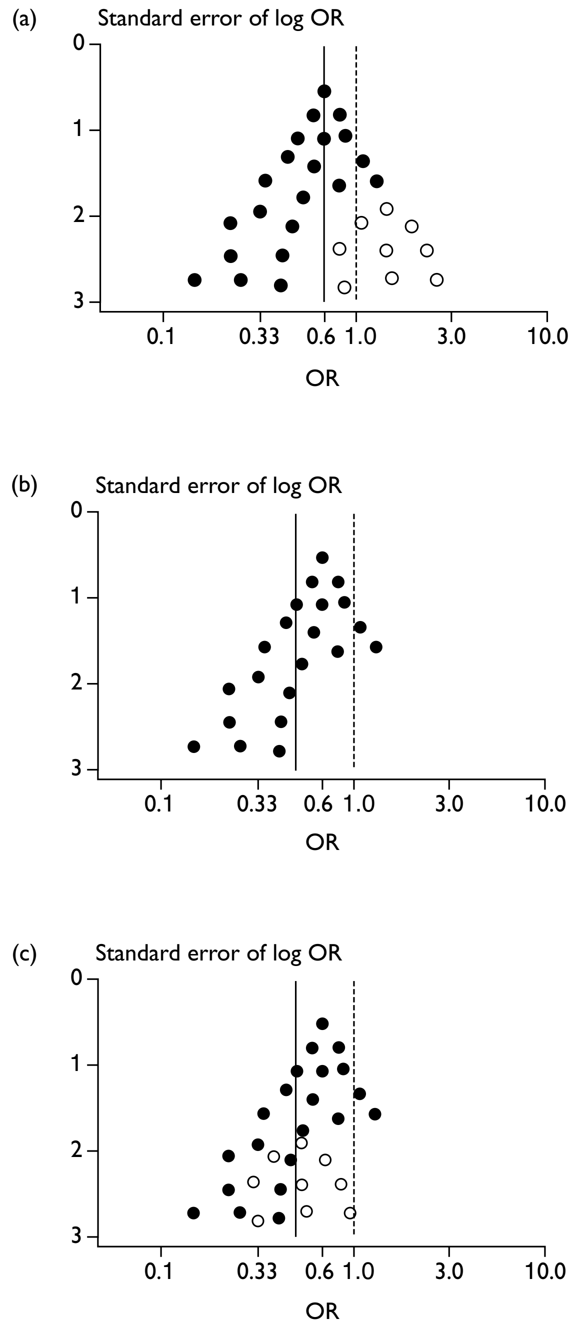

FIGURE 4 *Hypothetical funnel plots: (a) symmetrical plot in the absence of bias (open circles indicate smaller studies showing no beneficial effects); (b) asymmetrical plot in the presence of publication bias (smaller studies showing no beneficial effects are missing); (c) asymmetrical plot in the presence of bias due to low methodological quality of smaller studies (open circles indicate small studies of inadequate quality whose results are biased towards more beneficial effects). The solid line is the pooled ORs, the dotted line is the null effect. The pooled ORs exaggerate treatment effects in the presence of bias. Adapted from* Sterne et al.,[54] *and reproduced by permission of BMJ Books*

treatment effects. Such asymmetry might also result from the tendency of smaller studies of lower methodological quality to show more beneficial effects (*Figure 4c*). We discuss methodological issues relevant to the funnel plot, including the choice of axes, in detail elsewhere.[52,53]

Rationale

The inclusion of an unbiased sample of relevant studies is clearly central to the validity of meta-analytic research. However, the dissemination of medical evidence, including the results from randomised trials, is influenced by a host of factors that affect the probability that a given trial is included in a meta-analysis. Trials with statistically significant (positive) results have been shown to be more likely to be published,[24] and more likely to be published in English[43] than trials with negative results. Such 'positive' trials may also be more likely to be published in MEDLINE-indexed journals. To prevent bias in systematic reviews and meta-analyses, the Cochrane Collaboration,[56] the NHS Centre for Reviews and Dissemination (CRD) (University of York, UK)[57] and other experts in the field[58–60] recommend extensive literature searches that cover the grey literature and all relevant languages and databases. This may involve time-consuming and costly searches and the translation of foreign language articles.

Although it seems likely that excluding unpublished trials and trials reported in languages other than English will introduce bias and reduce the precision of estimates of treatment effects, the importance and direction of these effects is unclear at present. There may be trade-offs between timeliness, cost and quality of systematic reviews. We examined this issue in rigorously conducted systematic reviews by simulating the effect of less comprehensive literature searches, taking into account the importance of the methodological quality of trials.

Objectives

Our objectives are stated below under headings for the three types of bias that were investigated. All relate to controlled trials, with binary outcomes, included in meta-analyses.

Publication bias

- To examine the characteristics (trial size, results, conventional level of statistical significance, quality) of controlled trials that are unpublished (trials from the grey literature) and compare them with those of controlled trials that have been published.
- To compare, within meta-analyses, the treatment effects reported in grey trials with those reported in published trials.
- To assess the impact of excluding grey trials on pooled effect estimates and associated *p*-values, and on the shape of funnel plots.
- To evaluate whether it is justified for the authors of meta-analyses of healthcare interventions to search only for published trials and thus to exclude grey trials from their syntheses.

Language bias

- To examine the characteristics (trial size, results, conventional level of statistical significance, quality) of controlled trials published in non-English languages and compare them with those of controlled trials published in English.
- To compare, within meta-analyses, the treatment effects reported in non-English language trials with those reported in English language trials.
- To assess the impact of excluding non-English language trials on pooled effect estimates and associated *p*-values, and on the shape of funnel plots.
- To evaluate whether it is justified for the authors of meta-analyses of healthcare interventions to search only for English language trials and thus to exclude non-English language trials from their syntheses.

MEDLINE bias

- To examine the characteristics (trial size, results, conventional level of statistical significance, quality) of controlled trials that are published in journals not indexed in MEDLINE and compare them with those published in MEDLINE-indexed sources.
- To compare, within meta-analyses, the treatment effects reported in trials published in journals not indexed in MEDLINE with those reported in indexed journals.
- To assess the impact of excluding trials that are published in journals not indexed in MEDLINE on pooled effect estimates and associated *p*-values, and on the shape of funnel plots.
- To evaluate whether it is justified for the authors of meta-analyses of healthcare interventions to search only for trials that are published in journals indexed in MEDLINE and thus to exclude reports published in non-indexed sources from their syntheses.

Bias due to inadequate quality of trials

- To examine the characteristics (trial size, results, conventional level of statistical significance) of controlled trials with inadequate/unclear concealment of allocation with trials with adequate concealment and of trials with inadequate blinding with double-blind trials.
- To compare, within meta-analyses, the treatment effects reported in trials of inadequate or unclear concealment or blinding with trials of adequate quality.
- To assess the impact of excluding trials of inadequate quality on pooled effect estimates and associated *p*-values, and on the shape of funnel plots.
- To evaluate whether it is justified for the authors of meta-analyses of healthcare interventions to include trials that appear to be of inadequate quality in their syntheses.

Chapter 2
Methods

Selection of study sample

In a first step we searched four different English language sources for meta-analyses of therapeutic or preventive interventions, which combined the binary outcomes of at least five controlled trials.

1. We conducted a search by hand of all issues of eight high-impact general and specialist medical journals for the period 1994 to 1998 inclusive. These journals were:
   - *American Journal of Cardiology*
   - *Annals of Internal Medicine*
   - *BMJ*
   - *Cancer*
   - *Circulation*
   - *JAMA*
   - *Lancet*
   - *Obstetrics and Gynecology*.

   We chose these journals because an initial MEDLINE search indicated that they publish many meta-analyses. We checked the completeness of the handsearch in an additional MEDLINE search using 'meta-analysis' as medical subject heading term and free-text word and 'systematic review' as free-text word.
2. We searched every systematic review published in issue 1/1998 of the Cochrane Database of Systematic Reviews (CDSR) for relevant meta-analyses.
3. The CRD supplied us with copies of reports of meta-analyses of at least five controlled trials published in any journal for the period 1994 to 1998 inclusive and which had been reviewed by staff at the centre for the Database of Abstracts of Reviews of Effectiveness (DARE).
4. We identified all those *Health Technology Assessment* reports published up to July 1999 by the UK NHS R&D Health Technology Assessment (HTA) Programme that contained systematic reviews and searched them by hand for suitable meta-analyses.

Inclusion criteria

We included meta-analyses that were based on comprehensive literature searches and provided sufficient data and information on techniques used to allow us to replicate the meta-analysis. We defined a comprehensive literature search as follows.

- The search was not restricted to English language literature.
- The Cochrane Controlled Trials Register (CCTR) or at least two other electronic databases (such as MEDLINE or EMBASE) had been searched.
- At least one of the following indicators of searches for unpublished trials:
  – search of conference abstracts
  – search of theses
  – search of a trials register
  – contacts with experts in the field, professional bodies, industry, or licensing bodies to identify unpublished data.

If we identified a report that contained the results of more than one meta-analytic pooling, we used the analysis that included the largest number of trials.

Definitions

Two of the reviewers independently classified all component trials from the eligible meta-analyses, without referring to the trials' results, and resolved any disagreements by consensus. Based on the list of references we classified trial reports as published journal articles if they had been published as full or short reports, editorials or letters in a regular or supplementary issue of a journal. All other reports, including conference abstracts published in proceedings or journals, books and book chapters, unpublished manuscripts and data on file were classified as unpublished (grey) literature. If more than one bibliographic reference was provided for a trial, we gave a reference to a journal article precedence over references to grey literature and so classified the trial as published. We checked references that were unclear using standard databases (MEDLINE, EMBASE and Science Citation Index). If a report could not be satisfactorily classified, we obtained a copy of the report or contacted the authors of the meta-analysis.

We assessed language of publication for published journal articles. Using SERLINE, the journals database produced by the National Library of Medicine (Bethesda, Maryland, USA), we compiled a list of journals that only publish articles in the English language. The language of a journal article was classified as English if the journal publishing a trial report was included in this list. Articles that had a title in a language other than English or were described as being non-English language in the bibliographic details were classified accordingly. For all other articles we checked their respective language fields in MEDLINE or EMBASE. If a report could not be satisfactorily classified in any of these ways, we obtained a copy of the report or contacted the authors of the meta-analysis. If there was more than one reference to a journal article, we gave a reference to an article in English precedence over any references to articles in languages other than English.

Finally, we classified the published trials according to whether or not the latter were published in a MEDLINE-indexed journal. We adopted the rule that for a trial to be classified as published in an indexed source, the journal should have a year of entry into MEDLINE that was prior to the year the article was published. We did not make any direct assessment as to how accurately any particular trial was indexed in MEDLINE in relation to the needs of reviewers undertaking a literature search.

Assessment of methodological quality

Quality assessment was restricted to trials included in meta-analyses published in the CDSR (issue 1/1998) as it was based on information on concealment of allocation of trial participants to treatment groups and blinding of outcome assessment provided in the reviews. All meta-analyses that included five or more trials with binary endpoints were considered for inclusion. Two of the reviewers independently extracted information on trial quality by referring to the text of the review but without considering the trial report or the results of the trial. For concealment of allocation we distinguished between adequately concealed trials (central randomisation, coded drug packs, assignment envelopes, etc.), and inadequately or unclearly concealed trials which either reported an inadequate approach (alternation, open random number tables, etc.) or lacked a statement on concealment. For blinding we distinguished between trials that were described as double-blind or included blinding of the person assessing outcomes (assessor-blind), and those that did not. Inter-observer reliability for this quality assessment procedure was determined using the kappa statistic.

Data extraction and replication of pooled estimates

For each meta-analysis included in our sample, we recorded the outcome, the meta-analytical method used for combining trials, the type of effect measure used (odds ratio (OR), relative risk, or hazard ratio) and the overall pooled effect estimate with its 95% confidence interval (CI), or standard error. One of the reviewers abstracted the raw outcome data for each trial (2 × 2 table) or, if the raw data were unavailable, the point estimate and 95% CI. For the meta-analyses published in the CDSR, Update Software (Oxford, UK) provided the raw data in electronic form.

We checked our data by replicating the meta-analyses, using the original meta-analytical models and compared the pooled results with those of the original meta-analyses. See Deeks and co-workers[61] for a detailed description of the standard statistical methods used for meta-analysis, including the Peto fixed effects model, the Mantel–Haenzel fixed effects model, the inverse variance fixed effects model and the DerSimonian–Laird random effects model. In the case of individual participant data (IPD) reviews we used the results of survival analyses stratified by trial. The log rank expected number of deaths and variance were used to calculate individual and overall pooled hazard ratios by the fixed effects model, in a similar manner to that used in the Peto method for ORs. If the model used was not specified we used the inverse variance approach for fixed effects analyses and the DerSimonian–Laird random effects model for random effects analyses.

To obtain consistency across the meta-analyses in our sample, we re-calculated the pooled effect estimates, where necessary, so that all results were expressed as undesirable results (e.g. mortality, not survival, presence of symptoms, not absence of symptoms). Thus, relative risks and ORs of less than 1.0 indicated treatments providing a beneficial effect whereas values exceeding 1.0 indicated treatments with adverse effects.

Analysis

The same analytic strategy was employed to assess the impact of the different reporting biases (publication, language and MEDLINE bias) and the impact of the quality of component studies. First, we restricted the sample to meta-analyses that were able to contribute to the analysis in question. For example, only meta-analyses that contained both published and unpublished trials were included in the analyses addressing the impact of publication bias. Similarly, only meta-analyses that included trials published in languages other than English and trials published in English were included in the analysis of language bias. Meta-analyses that contained trials published in a journal indexed in MEDLINE and trials published in a non-indexed journal were considered for the analysis of MEDLINE bias. Unpublished trials were excluded from the language and MEDLINE samples because they could not be classified regarding the language of publication and were by definition not indexed in MEDLINE. For the analysis of the importance of the methodological quality of trials we restricted the analysis to the Cochrane sample and to meta-analyses where information on quality was available for at least 80% of trials. As above, only meta-analyses that contained both trials with and without the quality characteristic (blinding or allocation concealment) trials were included in the analyses addressing the influence of methodological quality. We excluded trials published in languages other than English from main analyses to prevent confounding between publication status, language of reporting and trial quality and to make results comparable with a previous study.[49]

The analysis then proceeded in four steps. The description below relates to the analysis of publication status (unpublished versus published) but identical analyses were performed for language (non-English versus English), database (not indexed in MEDLINE versus indexed) and methodological quality (inadequate or unclear concealment of allocation versus adequate concealment; not double-blind versus double-blind).

- We ascertained the characteristics of unpublished trials (year of publication, type of intervention and comparison, sample size, quality and level of statistical significance) and compared these characteristics with those of published trials. If a trial appeared in more than one meta-analysis we counted it only once. We also calculated the percentage weight contributed by unpublished trials to individual meta-analyses.
- Within each meta-analysis we calculated pooled effect estimates separately for the unpublished and published trials, applying the same meta-analytical model used by the authors. We then derived for each meta-analysis a ratio of the pooled estimate from unpublished trials to the pooled estimate from published trials. A ratio below 1.0 would indicate that the unpublished trials showed a more beneficial treatment effect than the published trials. A ratio above 1.0 would indicate the opposite. The log of this ratio is the difference between the log of the treatment effects in published and unpublished studies, and the variance of the log of the ratio was therefore calculated by adding the variance of the log treatment effects in published and unpublished studies. We calculated a weighted average for all these ratios using random effects meta-analysis, also stratifying by clinical area, source (meta-analyses published in the CDSR versus others), type of intervention (drugs versus others), type of control (active control intervention versus others), and complementary versus conventional medicine.
- For each meta-analysis we calculated the percentage change in the pooled effect estimate which occurred when unpublished trials were excluded from the meta-analysis. We also examined the changes in *p*-values and in precision (defined as the inverse of the standard error) which occurred when unpublished trials were excluded from the meta-analyses.
- Finally, we examined the impact of removing unpublished trials on the shape of funnel plots, using the method proposed by Egger and coworkers.[3] The extent of asymmetry is defined by assuming a linear relationship between treatment effect (log OR) and its standard error:[3,52]
  $$\log(\mathrm{OR}) = \text{adjusted treatment effect} + (\text{asymmetry coefficient} \times \text{standard error of log OR})$$
  The 'adjusted' treatment effects refers to the effect in very large trials. In the absence of funnel plot asymmetry the asymmetry coefficient equals 0.0. Negative asymmetry coefficients indicate that treatment effects are more pronounced for smaller trials with larger standard errors. We calculated an asymmetry coefficient separately for each meta-analysis and then combined coefficients using a fixed effects model. We then repeated the analysis excluding unpublished trials and compared combined asymmetry coefficients.

All analyses were performed in Stata version 7.0 (Stata Corporation, College Station, Texas).[62]

Sensitivity analyses using logistic regression

Previous meta-epidemiological studies[49,50,63,64] addressing similar questions, including the landmark study by Schulz and co-workers,[49] used a fixed effects logistic regression approach in which the evidence for an interaction between the effects of trial quality and intervention group is examined, having controlled for the interaction between meta-analysis and treatment group. This assumes that the effect of bias is constant across meta-analyses, which may not be the case in practice. If the effect does vary between meta-analyses then standard errors of estimated differences will be too small.[65] We therefore chose an alternative approach in which we combine estimated effects in a 'meta-meta-analysis', allowing for between-meta-analysis variation. We performed sensitivity analyses comparing results from logistic regression with our meta-analytic approach.

The different biases that affect systematic reviews and meta-analyses are unlikely to operate independently. For example, publication bias may lead to treatment effect estimates being smaller in unpublished trials, but such trials may also tend to be of lower methodological quality and therefore to overestimate treatment effects. The logistic regression model can be used to control for such confounding by including the effects of more than one trial characteristic in the model.

Chapter 3

Results

Identification and characteristics of eligible meta-analyses

We identified 303 meta-analyses with at least five trials and a binary outcome, but had to exclude 144 meta-analyses, mostly because no comprehensive literature search had been performed. A total of 159 meta-analyses, for which comprehensive literature searches had been employed, had usable data (*Figure 5*).

The bulk of meta-analyses (116, 73.0%) were identified from the CDSR, 26 (16.4%) through handsearching of journals, 12 (7.6%) from DARE and 5 (3.1%) were found in *Health Technology Assessment.* Six meta-analyses had been published both electronically in the Cochrane database and as articles published in journals. These were classified as journal articles but included in the analysis of the impact of trial quality which was exclusively based on Cochrane reviews. The number of trials included in the 159 meta-analyses, and the proportion of trials that were difficult to locate, are shown in *Table 2.*

It is clear from *Table 2* that despite comprehensive literature searches only a relatively small number of trials that are difficult to locate were identified for these reviews. The total number of trials included in the 159 meta-analyses was 1635; 153 (9.4%) of these were unpublished, 115 (7.0%) were published in languages other than English and 161 (9.8%) were published in a journal not indexed in MEDLINE. Sixty meta-analyses (37.7%) included at least one unpublished trial, 50 (31.4%) at least one trial published in a language other than English and 66 (41.5%) at least one trial not indexed in MEDLINE. Fifteen meta-analyses (9.4%) included trials from all three categories and 45 reviews (28.3%) did not include any trial that was difficult to locate.

The picture was different for the analyses of the impact of trial quality. There were 122 meta-analyses in the CDSR (issue 1/1998) that included five or more trials with binary endpoints (including the six reviews that were also published in journals). Thirty-nine (32.0%) meta-analyses could be included in the analyses of the impact of concealment of allocation and 45 (36.9%) in the analyses of blinding. The analysis of the impact of concealment was based on 304 trials published in English 186 (61.1%) of which were either inadequately concealed or concealment was

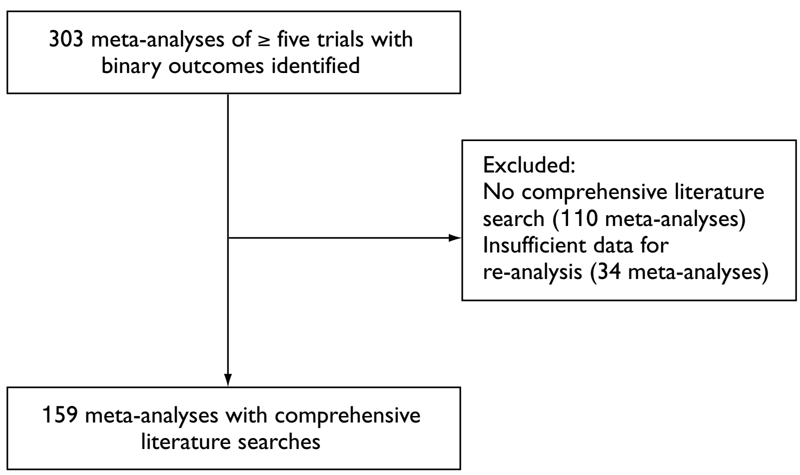

FIGURE 5 *Progress through the stages of identifying eligible meta-analyses*

TABLE 2 The total number of trials included in 159 meta-analyses and the percentage of trials that were difficult to locate. Medians (ranges) are shown

Source of meta-analysis	Total no. of trials	Unpublished (%)	Published in language other than English (%)	Published in journal not indexed in MEDLINE (%)
CDSR (*n* = 116)	7 (5–25)	0 (0–80.0)	0 (0–50.0)	0 (0–80.0)
Journals (*n* = 26)	10 (5–53)	10.4 (0–44.4)	0 (0–61.5)	21.5 (0–69.2)
DARE (*n* = 12)	10 (6–35)	9.3 (0–40.0)	10.0 (0–33.3)	21.1 (0–50.0)
HTA reports (*n* = 5)	5 (5–12)	0 (–)	0 (0–50.0)	0 (0–40.0)
All (*n* = 159)	8 (5–53)	0 (0–80.0)	0 (0–61.5)	0 (4.2–80.0)

unclear. Similarly, the analysis of blinding was based on 399 English language trials, 162 (40.6%) of which were not double-blind. It is clear from these figures that the impact of trials of lower quality may well be greater than the impact of trials that are difficult to locate.

There was some variation in outcome measures and the statistical methods used to combine results from individual trials although most analyses were performed on the OR scale using the Peto fixed effects model (*Table 3*). This reflects the large number of Cochrane reviews in our sample: the Peto method is the default method in the software used by Cochrane reviewers and only works on the OR scale. Among the 116 Cochrane reviews, 93 (80.2%) used the Peto fixed effects model and only four (3.4%) used a random effects model. Random effects models were more popular in reviews published elsewhere (10/43, 23.3%). See Deeks and co-workers[61,66] for a discussion of the different statistical models and effect measures.

All but five meta-analyses were replicated using data from the 2 × 2 table for each trial. There were two meta-analyses for which we used point estimates and 95% CIs, and three IPD meta-analyses for which we used results of survival analyses. All reports specified whether a fixed or random effects model had been used; however, in three instances the exact model used was not described. In these situations we used the inverse variance or DerSimonian–Laird models. As shown in *Figure 6*, we were able to closely reproduce the pooled estimates reported by reviewers for all 159 meta-analyses, using the effect measures chosen by the original reviewers.

A table listing the 133 meta-analyses that were included in one or several analyses reported in the subsequent sections of this report is given in appendix 1. The bibliographic references of these meta-analyses can be found in appendix 2. Finally, the bibliographic references of the 26 meta-analyses that met inclusion criteria, but lacked trials with characteristics of interest, are listed in appendix 3. (NB. Reference numbers in

TABLE 3 Statistical methods and outcome measures used in 159 meta-analyses

Statistical method used in meta-analysis	Outcome measure			
	Hazard ratio	OR	Risk ratio	All
Fixed effects models				
Peto	0	108	0	108
Mantel–Haenzel	0	11	17	28
Inverse variance	0	1	0	1
IPD analysis	3	1	0	4
Not specified	0	2	2	4
Random effects models				
DerSimonian–Laird	0	6	5	11
Not specified	0	1	2	3
All	3	130	26	159

Three meta-analyses that reported relative odds reductions are included in the OR category

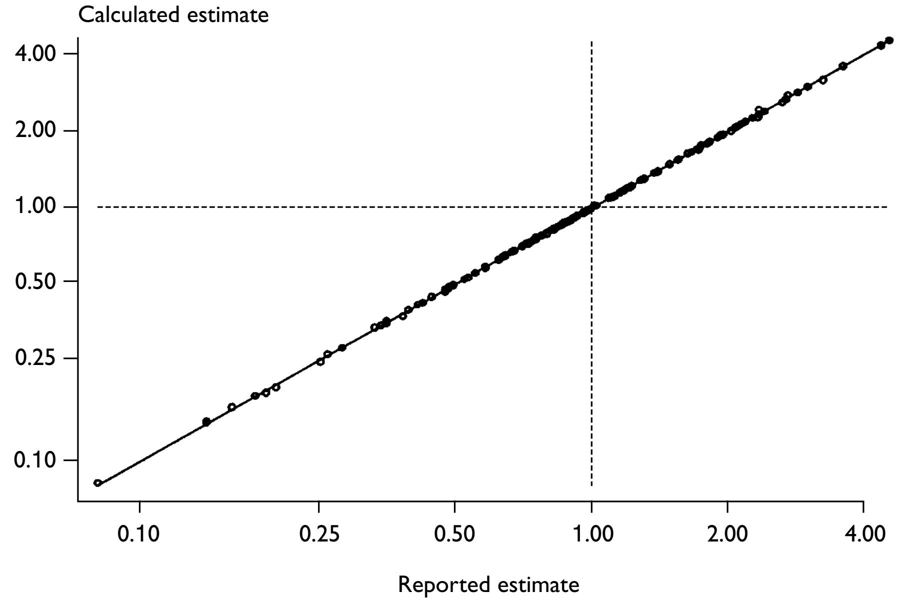

FIGURE 6 *Scatter plot of pooled estimates of treatment effects reported by reviewers against estimates re-calculated in this project*

appendices 2 and 3 do not relate to reference numbers in the main report. The reference list for the main report can be found on page 53.)

The impact of unpublished trials

Inter-observer reliability for the classification of trials according to publication status was excellent (kappa = 0.98). Sixty of 159 meta-analyses with comprehensive literature searches were found to contain at least one grey trial and were therefore included in our analyses (*Figure 7*). Of the 60 meta-analyses, 18 were from journals, 36 were from the CDSR and six were from DARE. The 60 meta-analyses incorporated a total of 783 trials.

Characteristics of trials

Overall 630 trials were published and 153 trials were unpublished grey literature. Of the 153 unpublished trials 69 (45.1%) appeared as abstracts, 22 (14.4%) were reported in books, five (3.3%) appeared in theses, and 57 (37.3%) were other forms of grey literature such as file-drawer data, or material from a trials register. The source of meta-analyses was similar for published and unpublished trials (*Table 4*). Unpublished trials were less frequently concerned with the evaluation of drugs, had smaller sample sizes and were less likely to produce statistically significant results.

The proportion of published and unpublished trials varied according to medical speciality, with oncology and rheumatology/orthopaedics having the highest and lowest proportion of unpublished trials, respectively (*Table 5*).

Assessments by Cochrane Collaboration reviewers relating to concealment of allocation were available for 416 (53.1%) of 783 trials, while assessments relating to double- or assessor-blinding were also available for 416 trials, but not all of these were the same trials. Inter-observer reliability for extraction of these quality assessments by the present researchers was high with kappas of 0.96 for concealment of allocation and 0.94 for blinding. With respect to these two central domains of methodological quality, published trials tended to be of higher quality (*Table 6*).

Estimates of treatment effects from unpublished and published trials

Figure 8 shows the ratios of pooled treatment effects from unpublished trials to those from published trials for all 60 meta-analyses. Pooled estimates from the grey trials were on average 7% less beneficial (average ratio 1.07; 95% CI, 0.98 to 1.18). However, there was notable heterogeneity between meta-analyses ($p = 0.05$) with pooled effect estimates from grey trials, ranging from 97% more to 209% less beneficial than those from the respective published literature.

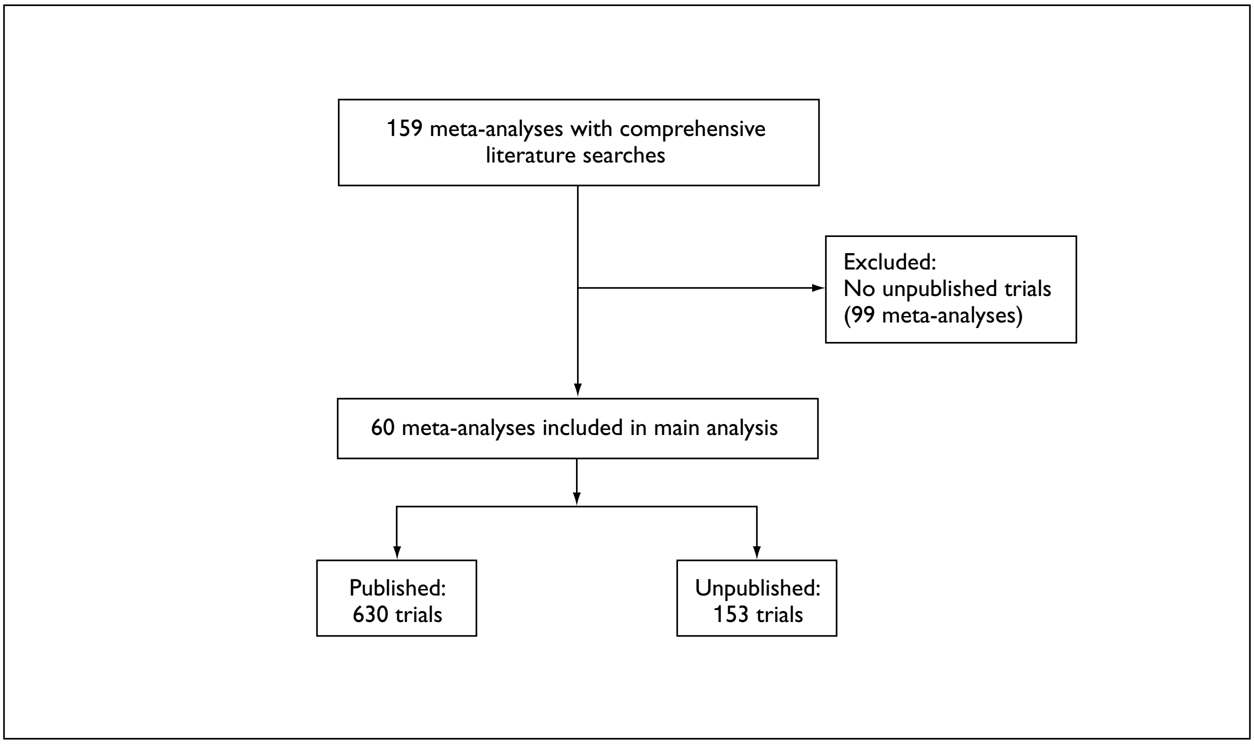

FIGURE 7 *Progress through the stages of identifying eligible meta-analyses which included unpublished trials*

TABLE 4 *Characteristics of published and unpublished trials*

	Published report (n = 630)	**Unpublished report (n = 153)**	***p***
Source of meta-analysis			
CDSR	299 (47.5%)	77 (50.3%)	0.72
General medical journal	257 (40.8%)	61 (39.9%)	
Specialist journal	74 (11.8%)	15 (9.8%)	
Type of intervention and comparison			
Drug intervention	579 (91.9%)	131 (85.6%)	0.016
Complementary medicine	11 (1.8%)	2 (1.3%)	0.70
Active control intervention	157 (24.9%)	36 (23.5%)	0.72
Sample size of trial			
Mean (SD)	232 (442)	141 (151)	0.012
Median (range)	102 (8–5042)	91 (9–1012)	0.073
Statistical significance of trial			
$p < 0.05$	187 (29.7%)	29 (19%)	0.008
$p < 0.01$	100 (15.9%)	18 (11.8%)	0.20

p-values from chi-squared tests, t-tests or Wilcoxon rank sum tests

SD, standard deviation

There was also wide variation between ratios for different medical specialities, with a marked difference (less benefit in unpublished trials) in obstetrics and gynaecology, for example, while in oncology there was little difference between the results of unpublished and published trials.

The results of the stratified analyses are presented in *Figure 9*. The differences in the results between unpublished and published trials appeared to be more pronounced in meta-analyses from the CDSR compared with those from other sources, and in meta-analyses with non-active controls compared with those having active controls, but none of these

TABLE 5 *Publication status of trials by medical speciality*

Medical speciality	**Published (*n* = 630)**	**Unpublished (*n* = 153)**
Cardiology & angiology	116 (80.6%)	28 (19.4%)
Gastroenterology	31 (79.5%)	8 (20.5%)
Infectious diseases	57 (73.1%)	21 (26.9%)
Neonatology	31 (79.5%)	8 (21.5%)
Neurology	40 (76.9%)	12 (23.1%)
Obstetrics & gynaecology	97 (77.6%)	28 (18.4%)
Oncology	35 (64.8%)	19 (35.2%)
Psychiatry	70 (88.6%)	9 (11.4%)
Rheumatology & orthopaedics	57 (90.5%)	6 (9.5%)
Other	96 (87.3%)	14 (12.7)

p = *0.007 by chi-squared test*

TABLE 6 *Methodological quality of published and unpublished trials included in Cochrane reviews*

	Published report	**Unpublished report**	***p***
Adequate allocation concealment			0.26
Yes	138/339 (40.7%)	26/77 (33.8%)	
No/unclear	201/339 (59.3%)	51/77 (66.2%)	
Double- or assessor-blinded			0.001
Yes	227/345 (65.8%)	32/71 (45.1%)	
No/unclear	118/345 (34.2%)	39/71 (54.9%)	

Probability by chi-squared tests

differences reached statistical significance in formal tests of interaction ($p > 0.30$). The difference was more pronounced in the numerous drug meta-analyses. Only one meta-analysis of a complementary medical intervention was in the sample, and this showed a substantial effect of unpublished trials.

Impact of unpublished trials on the results of meta-analyses

Unpublished trials contributed a mean of 18.2% (median 14.1%) of the weight in individual meta-analyses, with a range extending from less than 1% to 72.5%. *Figure 10* shows the change in pooled estimates that occurred when grey trials were removed from the sample of meta-analyses. The changes ranged from a 28.1% decrease to a 23.6% increase in benefit. The mean and median changes were –1.40 and –0.84, respectively. In 43 (71.7%) of the 60 meta-analyses the percentage changes were less than 5%. In the 17 meta-analyses in which the change was 5% or more, eight showed increased and ten showed decreased benefit. The average precision of pooled estimates decreased from 9.22 to 8.41 with the removal of grey trials. Three meta-analyses lost and one gained statistical significance at the 5% level.

This analysis is based on the 60 meta-analyses that included unpublished trials but the total number of meta-analyses ($n = 159$) that were based on comprehensive literature searches is arguably a more appropriate denominator for this analysis. Using this denominator the percentage change in pooled estimates would be zero or less than 5% in 142 (89.3%) of 159 meta-analyses.

Impact of unpublished trials on the shape of funnel plots

This analysis was based on 58 meta-analyses and a median of 11 trials (range, 4–53). Two meta-analyses had to be excluded because the number of trials remaining after removal of unpublished trials was too small (less than four) to allow a meaningful funnel plot analysis. The combined asymmetry coefficient from meta-analyses including all trials was –0.44 (95% CI, –0.60 to –0.28). There was thus evidence of funnel plot asymmetry (smaller trials showing larger treatment effects) even when unpublished trials were included in the

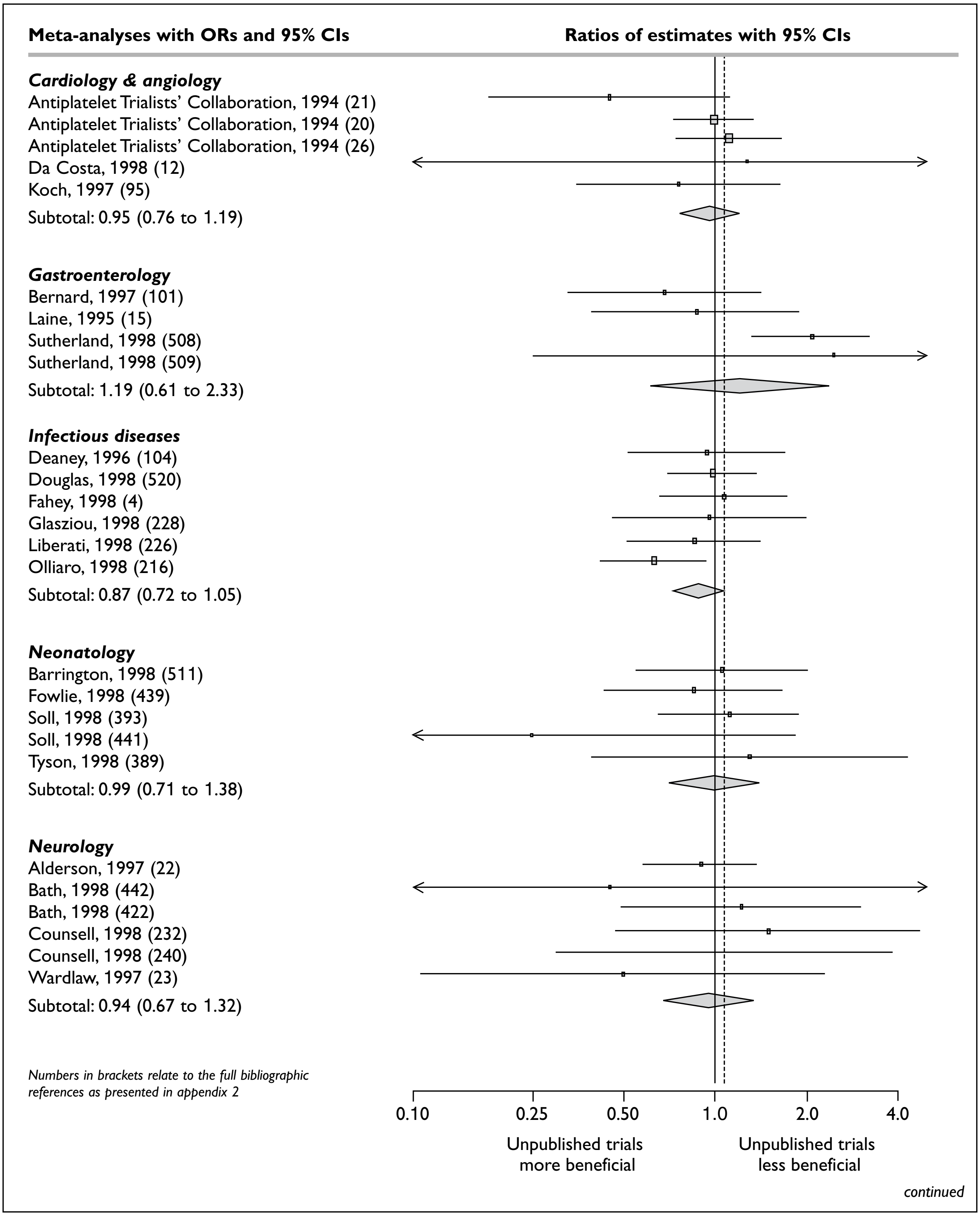

FIGURE 8 *Results from comparisons of treatment effect estimates from unpublished with those from published trials in 60 meta-analyses, calculating ratios of estimates. Ratios of estimates (grey squares) with 95% CIs of individual meta-analyses are shown. The size of the square reflects statistical weight in the overall pooled analysis. The meta-analyses are sub-grouped according to clinical topic, and arranged alphabetically according to the first author. The grey diamonds represent pooled results from clinical sub-groups, the black diamond overall pooled results. Ratio of estimates were pooled using random effects models. A ratio of estimates above 1.0 indicates that grey trials show a less beneficial treatment effect than published trials*

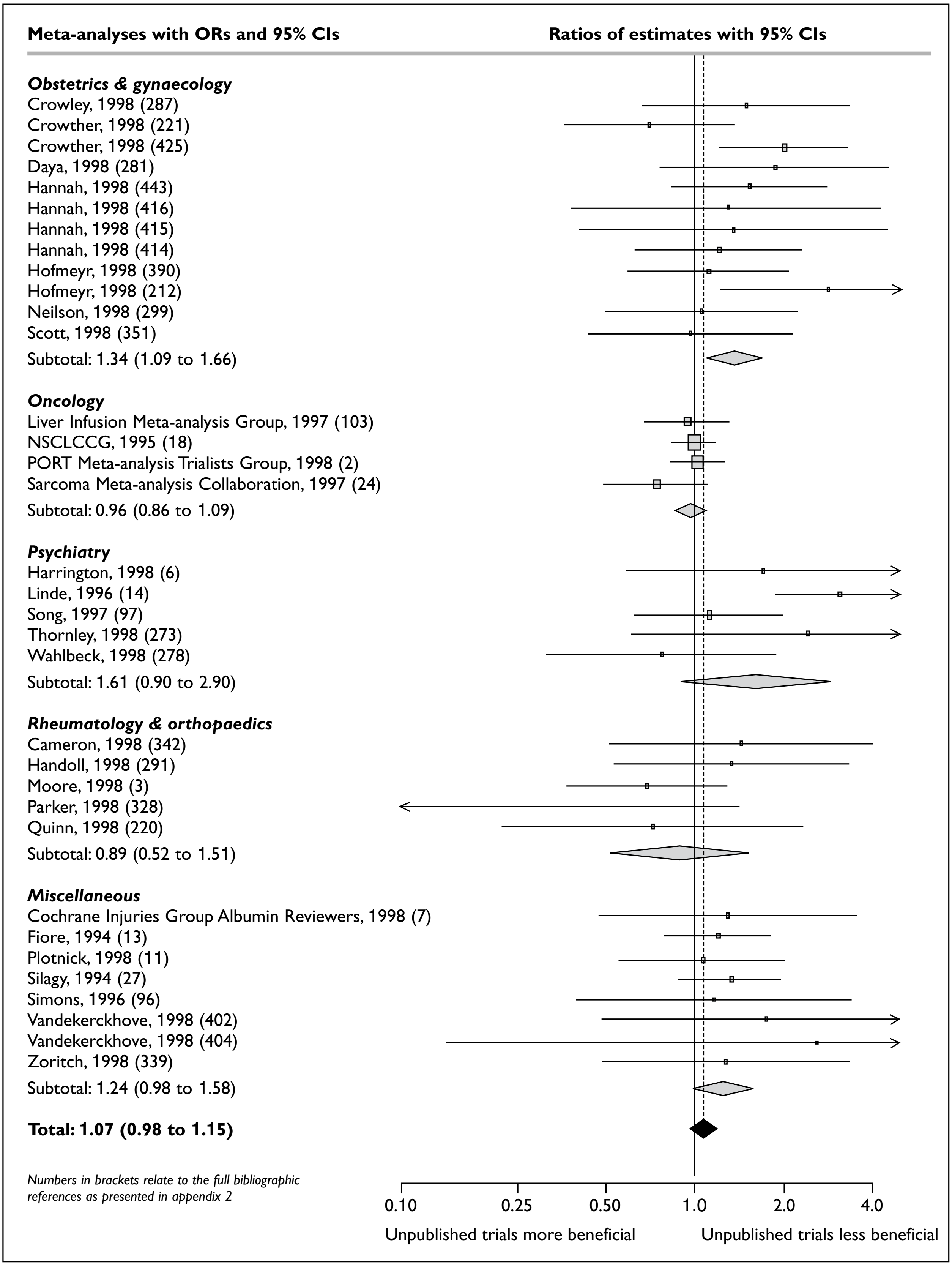

***FIGURE 8* contd** *Results from comparisons of treatment effect estimates from unpublished with those from published trials in 60 meta-analyses, calculating ratios of estimates. Ratios of estimates (grey squares) with 95% CIs of individual meta-analyses are shown. The size of the square reflects statistical weight in the overall pooled analysis. The meta-analyses are sub-grouped according to clinical topic, and arranged alphabetically according to the first author. The grey diamonds represent pooled results from clinical sub-groups, the black diamond overall pooled results. Ratio of estimates were pooled using random effects models. A ratio of estimates above 1.0 indicates that grey trials show a less beneficial treatment effect than published trials*

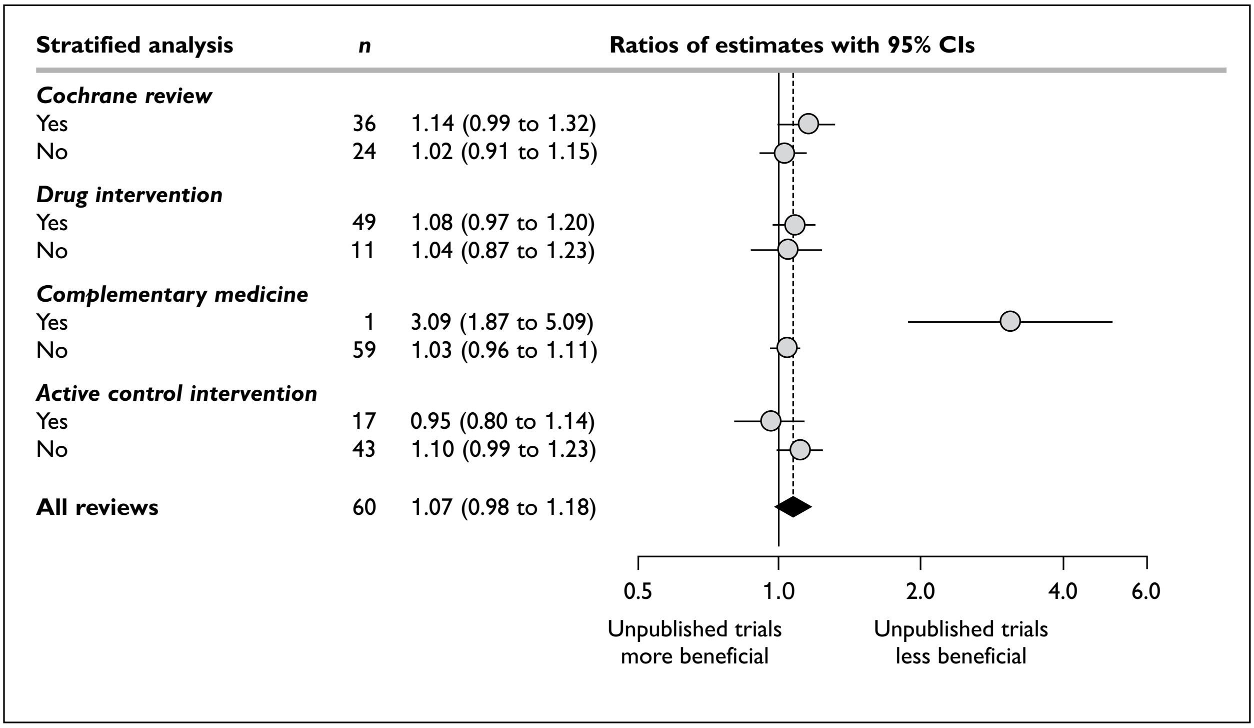

FIGURE 9 Results from stratified analyses comparing treatment effect estimates of unpublished with those of published trials. Ratios of estimates (circles) with 95% CIs of individual strata are shown. The black diamond represents overall pooled results. Estimates were pooled using random effects models. A ratio of estimates above 1.0 indicates that grey trials show a less beneficial treatment effect than published trials

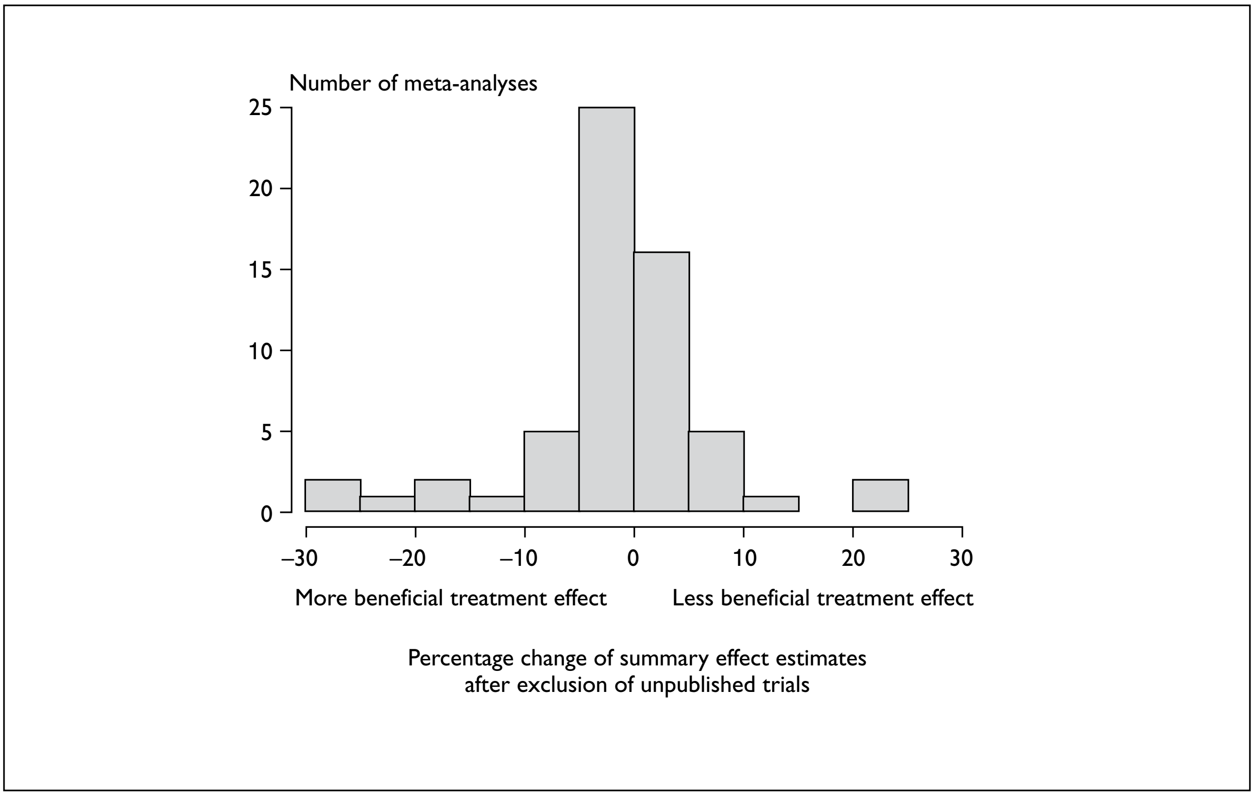

FIGURE 10 Percentage change of treatment effect estimates of individual meta-analyses after exclusion of grey trials. The histogram shows the frequency of percentage changes in pooled estimates that occurred when grey trials were removed from meta-analyses

analysis. As expected, after excluding the unpublished trials funnel plot asymmetry increased and the asymmetry coefficient became more extreme: –0.57 (95% CI, –0.74 to –0.40).

The impact of trials published in languages other than English

Fifty (31.4%) of the 159 meta-analyses with comprehensive literature searches included at least one trial published in a language other than English and were therefore included in the analyses presented in this section. *Figure 11* shows the progress through the stages of identifying eligible meta-analyses which included trials published in languages other than English.

The number of comprehensive meta-analyses including non-English language trials was 29 (25.0%) of 116 meta-analyses published in the CDSR, 12 (46.2%) of 26 meta-analyses published in general medical journals and nine (52.9%) of 17 meta-analyses published in specialist journals. The 50 meta-analyses incorporated 671 trials, but we excluded 71 unpublished trials for the purpose of the language analyses. Six hundred trials that were published in 208 English and 95 non-English language journals thus formed the basis for the analyses reported below.

Characteristics of trials

The language of publication was English in 485 (80.1%) trials. Of the 115 trials published in other languages, 42 (36.5%) were published in German, 29 (25.2%) in French, 12 (10.4%) in Italian, eight (7.0%) in Japanese, seven (6.1%) in Spanish, six (5.2%) in Portuguese, eight (7.0%) in four other European languages and three (2.6%) in Chinese. Characteristics of trials were similar with respect to the year of publication and the type of intervention and comparison. Non-English language trials included fewer participants but they were more likely to show statistically significant results compared with English language trials (*Table 7*). The proportion of trials published in languages other than English varied widely across clinical topics, from 10.1% in tobacco addiction to 35.7% in rheumatology and orthopaedics (*Table 8*). The proportion was notably greater in complementary medicine (41.2%) than in conventional medicine (21.7%). Assessments by Cochrane Collaboration reviewers of concealment of allocation was available for 294 trials (49.0%), while their assessment of blinding was available for 279 (46.5%) trials. Inter-observer reliability for extraction of these assessments by the present researchers was high, with kappas of 0.89 for concealment of allocation and 0.76 for blinding. As shown in *Table 9*, English language trials tended to be of higher methodological quality.

Estimates of treatment effects from trials published in English and trials published in other languages

Figure 12 shows the ratios of estimates of pooled treatment effects from non-English language

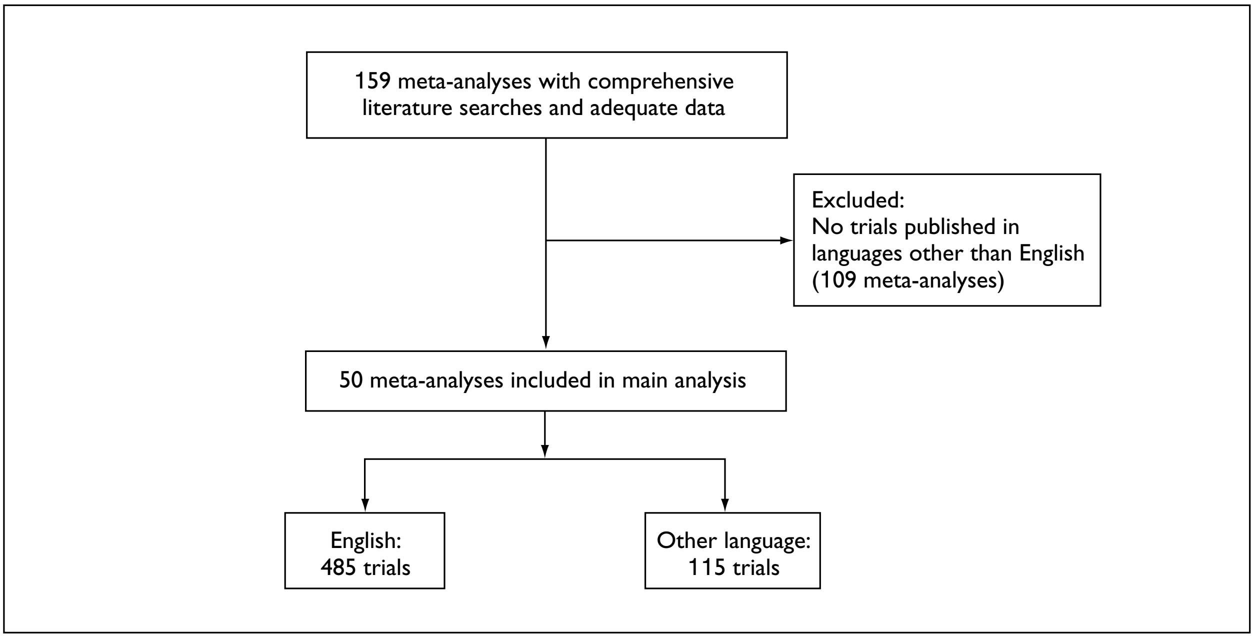

***FIGURE 11** Progress through the stages of identifying eligible meta-analyses, which included trials published in languages other than English*

TABLE 7 *Characteristics of randomised trials published in English and in languages other than English. Reproduced from Jüni* et al.[75] *by permission of Oxford University Press*

	English language report (n = 485)	Non-English language report (n = 115)	p
Source of meta-analysis			0.85
CDSR	232 (47.8%)	52 (45.2%)	
General medical journal	160 (33.0%)	41 (35.7%)	
Specialist journal	93 (19.2%)	22 (19.1%)	
Year of publication of trial			
Mean (SD)	1986 (7)	1986 (6)	0.59
Median (range)	1987 (1955–98)	1987 (1970–96)	0.24
Type of intervention and comparison			
Drug intervention	411 (84.7%)	103 (89.6%)	0.19
Complementary medicine	20 (4.1%)	14 (12.2%)	0.001
Active control intervention	117 (24.1%)	31 (27.0)	0.53
Sample size of trial			
Mean (SD)	269 (487)	147 (195)	0.009
Median (range)	116 (8–4524)	88 (19–1340)	0.006
Statistical significance of trial			
p < 0.05	152 (31.3%)	48 (41.7%)	0.033
p < 0.01	89 (18.4%)	34 (29.6%)	0.007

p-*values from chi-squared tests,* t-*tests or Wilcoxon rank sum tests*

TABLE 8 *Language of publication of trials by medical speciality. Reproduced from Jüni* et al.[75] *by permission of Oxford University Press*

Disease area	English-language report (n = 485)	Non-English language report (n = 115)
Tobacco addiction	62 (89.9%)	7 (10.1%)
Obstetrics & gynaecology	64 (87.7%)	9 (12.3%)
Cardiology & angiology	118 (86.8%)	18 (13.2%)
Infectious diseases	109 (79.6%)	28 (20.4%)
Neurology	42 (77.8%)	12 (22.2%)
Psychiatry	26 (65.0%)	14 (35.0%)
Rheumatology & orthopaedics	36 (64.3%)	20 (35.7%)
Miscellaneous	28 (80.0%)	7 (20.0%)

p *< 0.001 by chi-squared test*

TABLE 9 *Methodological quality of trials published in English and trials published in other languages that were included in Cochrane reviews. Reproduced from Jüni* et al.[75] *by permission of Oxford University Press*

	English-language report	Non-English language report	p
Adequate concealment of allocation			0.15
Yes	88/246 (35.7%)	12/48 (25.0%)	
No/unclear	158/246 (64.3%)	36/48 (75.0%)	
Double- or assessor-blinded			0.016
Yes	153/230 (66.5%)	23/49 (46.9%)	
No/unclear	77/230 (33.5%)	26/49 (53.1%)	

Denominators differ: information on concealment of allocation was provided more frequently than information on blinding. Probability by chi-squared tests

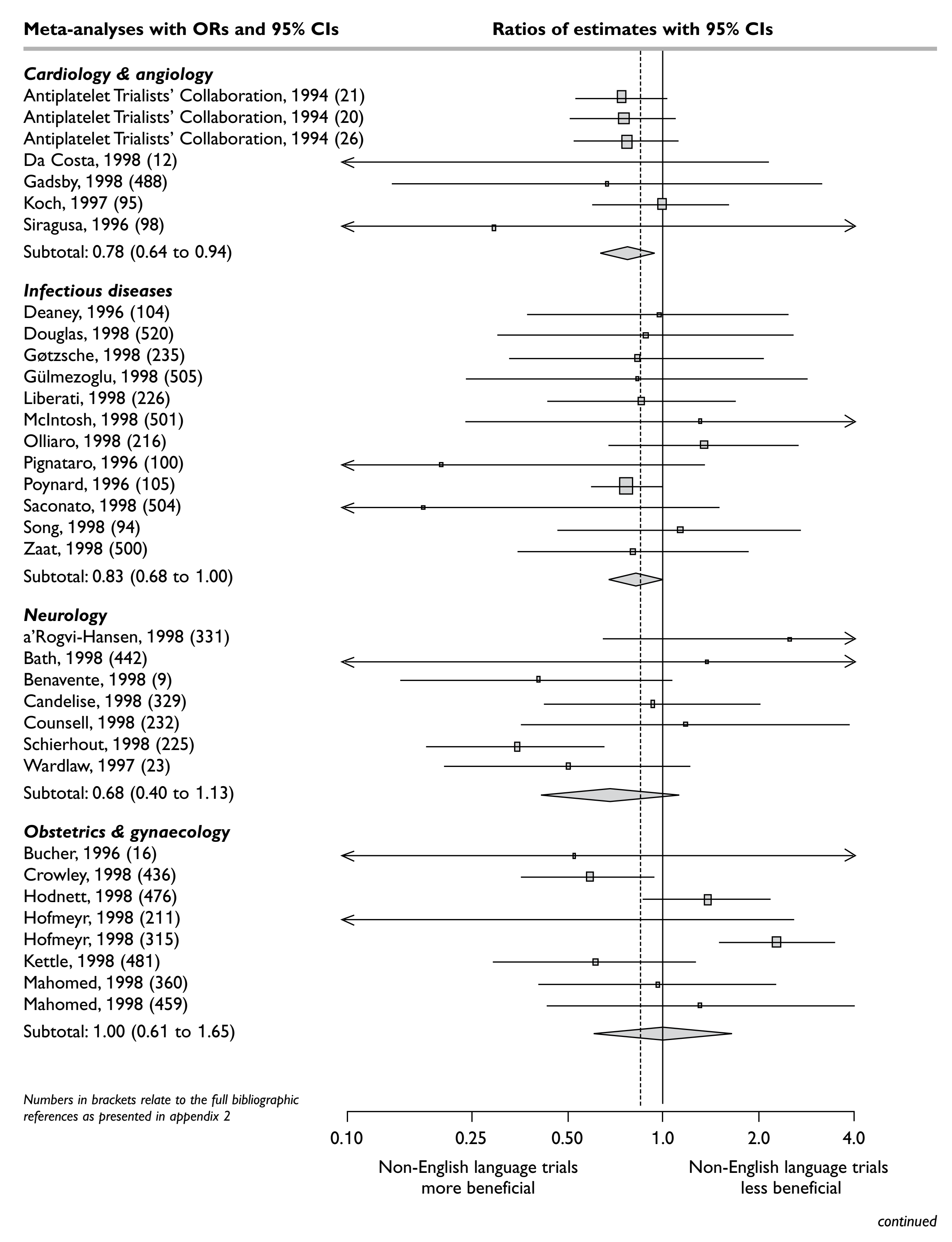

FIGURE 12 *Results from comparisons of treatment effect estimates from trials published in languages other than English with English language trials in 50 meta-analyses. Ratios of estimates (grey squares) with 95% CIs of individual meta-analyses are shown. The size of the square reflects statistical weight in the overall pooled analysis. The meta-analyses are grouped according to clinical topic, and arranged alphabetically according to the first author. The grey diamonds represent pooled results from clinical sub-groups, the black diamond overall pooled results. Ratio of estimates were pooled using random effects models. A ratio of estimates below 1.0 indicates that trials published in languages other than English show a more beneficial treatment effect than trials published in English. Reproduced from Jüni et al.*[75] *by permission of Oxford University Press*

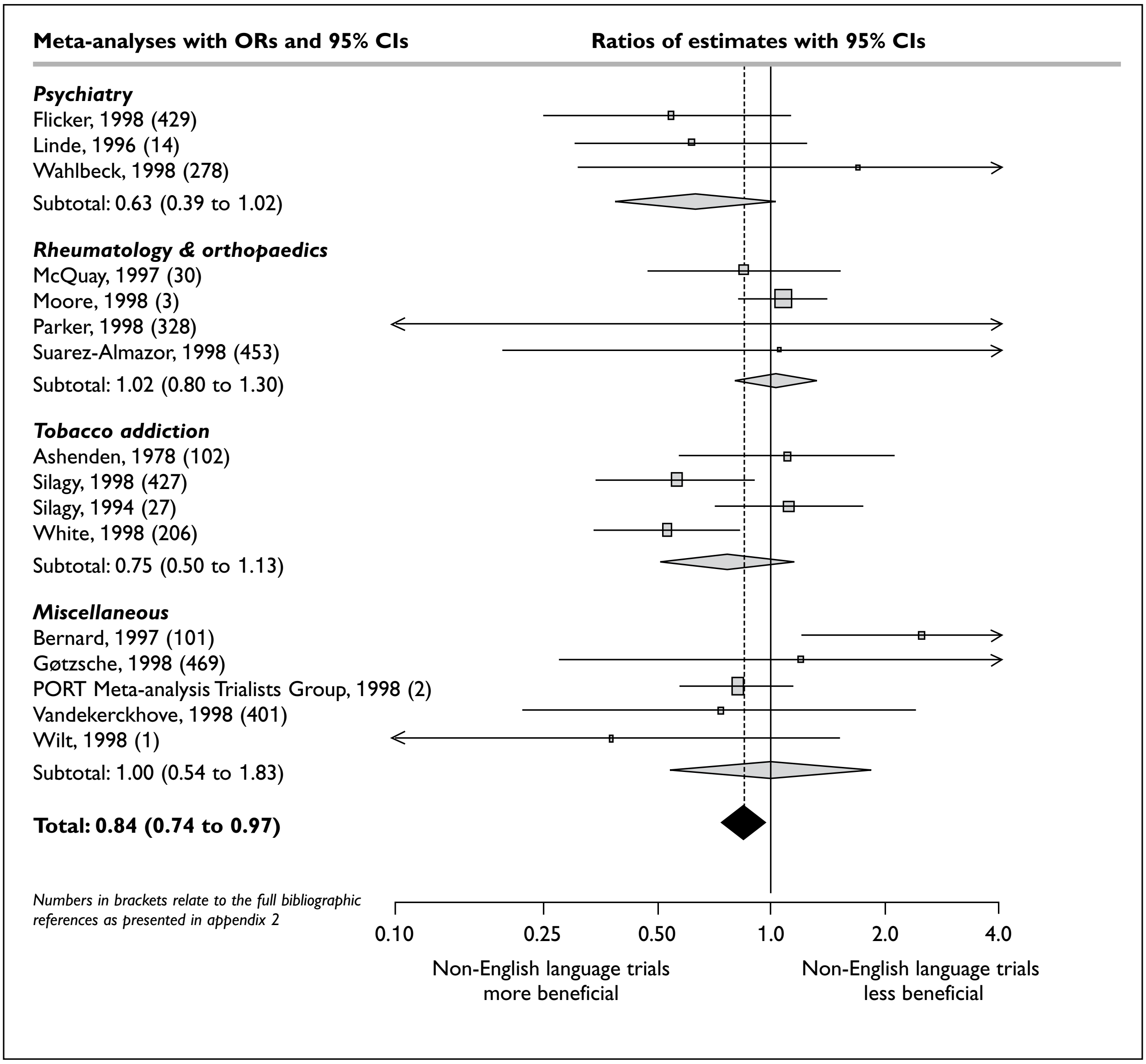

***FIGURE 12* contd** *Results from comparisons of treatment effect estimates from trials published in languages other than English with English language trials in 50 meta-analyses. Ratios of estimates (grey squares) with 95% CIs of individual meta-analyses are shown. The size of the square reflects statistical weight in the overall pooled analysis. The meta-analyses are grouped according to clinical topic, and arranged alphabetically according to the first author. The grey diamonds represent pooled results from clinical sub-groups, the black diamond overall pooled results. Ratio of estimates were pooled using random effects models. A ratio of estimates below 1.0 indicates that trials published in languages other than English show a more beneficial treatment effect than trials published in English. Reproduced from Jüni* et al.[75] *by permission of Oxford University Press*

trials compared with those from English language trials for the 50 meta-analyses. Treatment effect estimates were on average 16% more beneficial in non-English language trials (95% CI, 3% to 26%; $p = 0.011$). However, there was considerable heterogeneity between meta-analyses ($p = 0.003$), with pooled effect estimates of non-English language trials ranging from 90% more to 147% less beneficial compared with English language trials. Results of stratified analyses are presented in *Figure 13*. The effect of language appeared to be more pronounced in complementary medicine, and less pronounced in trials with active control interventions, but none of the differences between strata was statistically significant ($p > 0.20$).

Impact of non-English language trials on the results of meta-analyses

The number of trials published in languages other than English ranged from one to 14 trials and from 4.3% to 72.7% of all trials included. Non-English language trials contributed an average 17.5% of the weight in individual meta-analyses (median 10.2%; range, 1.2–81.1%).

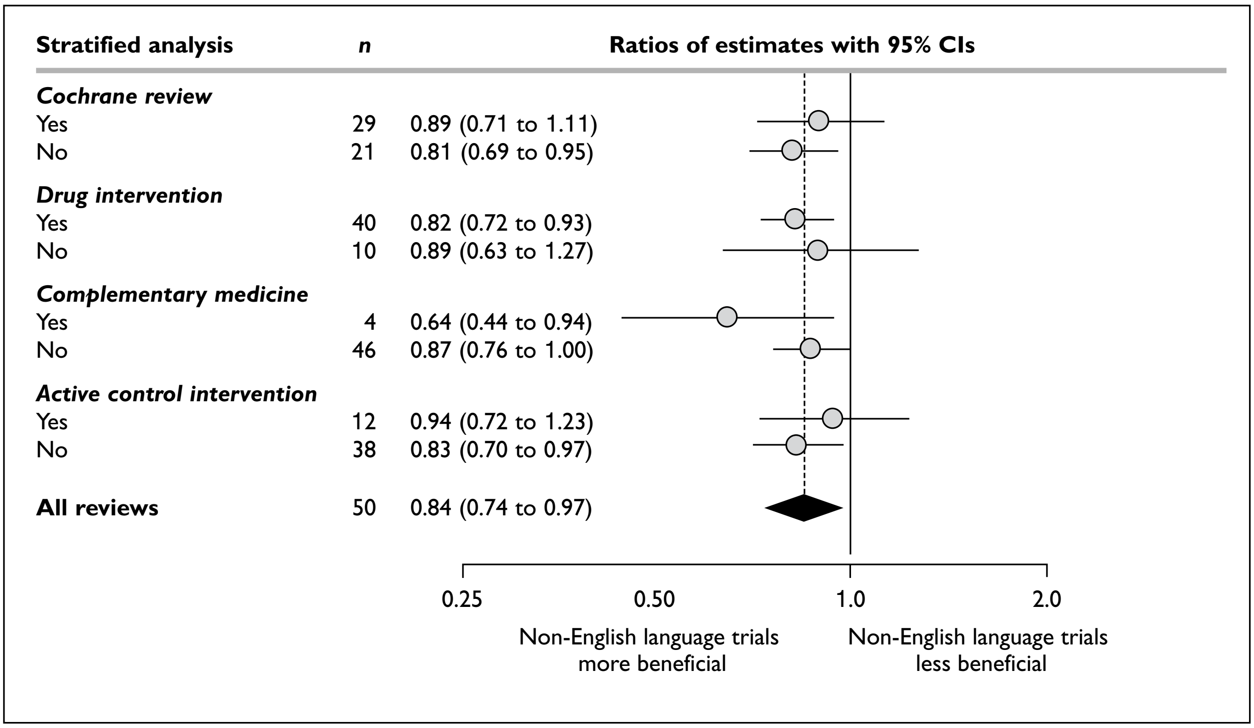

FIGURE 13 *Results from stratified analyses comparing treatment effect estimates of trials published in languages other than English with trials published in English. Ratios of estimates (circles) with 95% CIs of individual strata are shown. The black diamond represents overall pooled results. Estimates were pooled using random effects models. A ratio of estimates below 1.0 indicates that non-English language trials show a more beneficial treatment effect than English language trials. Reproduced from Jüni* et al.[75] *by permission of Oxford University Press*

Figure 14 shows the change in pooled estimates of individual meta-analyses that occurred when non-English language trials were excluded from meta-analyses. The changes ranged from a 42.0% increase (indicating less benefit) to a 22.7% decrease (indicating more benefit) of estimates of treatment effects. However, in 29 (58.0%) meta-analyses the changes were less than 5%. Among the remaining 21 meta-analyses five showed more benefit and 16 less benefit after exclusion of non-English language trials. When the analysis is based on all meta-analyses with comprehensive literature searches (n = 159) then the percentage change in pooled estimates is zero or less than 5% in 138 (86.8%) meta-analyses.

The average precision of pooled effect estimates decreased from 8.34 to 7.68 after exclusion of non-English language trials. Significance levels were affected in nine (18.0%) meta-analyses. In three cases the p-value increased from less than 0.001 to less than 0.01. In a further four cases p increased from less than 0.01 to less than 0.05, and in two instances p decreased from less than 0.05 to less than 0.01. None of the meta-analyses changed statistical significance at the 5% level.

Impact of non-English language trials on the shape of funnel plots

This analysis was based on 49 meta-analyses and a median of ten trials (range, 4–38). One meta-analysis had to be excluded because the number of trials remaining after removal of unpublished trials was too small (less than four) to allow a meaningful funnel plot analysis. The combined asymmetry coefficient for meta-analyses including trials published in any language was –0.49 (95% CI, –0.66 to –0.31). There was thus clear evidence of funnel plot asymmetry when trials published in languages other than English were included in the analysis. Interestingly, after excluding non-English language trials funnel plots became symmetrical, with an asymmetry coefficient of 0.07 (95% CI, –0.09 to 0.24).

The impact of trials published in journals not indexed in MEDLINE

Of the 159 meta-analyses with comprehensive literature searches 66 (41.5%) included at least one trial published in a journal not indexed in MEDLINE. *Figure 15* shows the progress through the stages of identifying eligible meta-analyses.

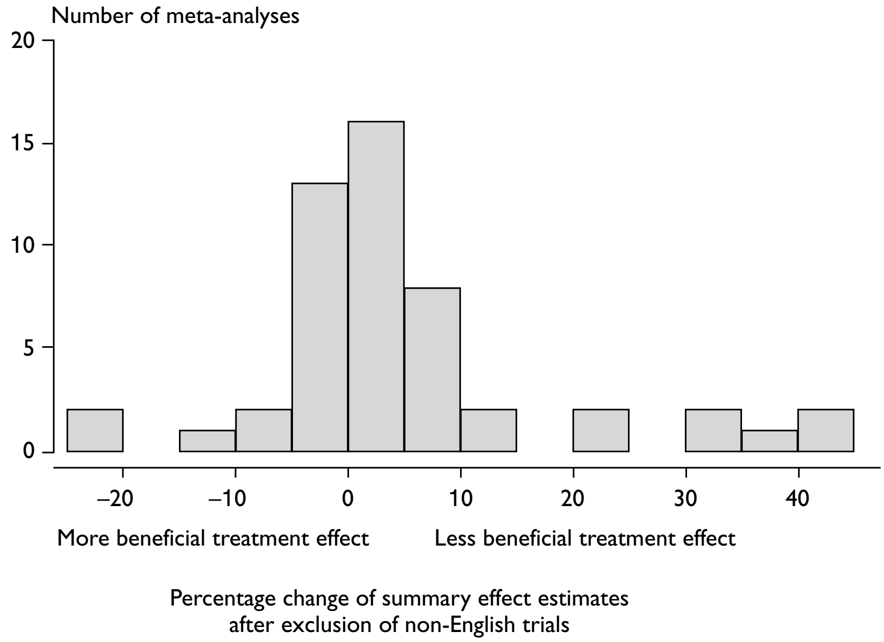

***FIGURE 14** Percentage change of treatment effect estimates of individual meta-analyses after exclusion of trials published in languages other than English. The histogram shows the frequency of percentage changes in pooled estimates that occurred when non-English language trials were removed from meta-analyses. Reproduced from Jüni* et al.[75] *by permission of Oxford University Press*

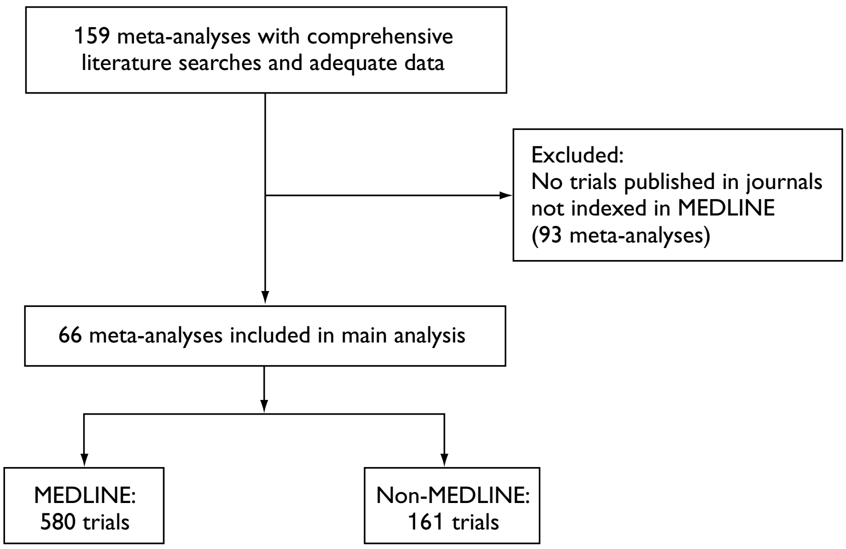

***FIGURE 15** Progress through the stages of identifying eligible meta-analyses, which included trials published in journals not indexed in MEDLINE*

The number of comprehensive meta-analyses including non-indexed trials was 45 (38.8%) of 116 meta-analyses published in the CDSR, 14 (53.8%) of 26 meta-analyses published in general medical journals and seven (41.2%) of 17 meta-analyses published in specialist journals. The 66 meta-analyses incorporated 898 trials, but we excluded 157 unpublished trials for the purpose of the MEDLINE analyses. A total of 741 trials which were published in 222 journals that were indexed in MEDLINE throughout the study period, 86 journals that were never indexed and 22 journals that were indexed at some point during the study period formed the basis for the analyses reported below.

Characteristics of trials

Overall, 580 trials were published in an indexed journal and 161 trials were published in a journal not indexed in MEDLINE. There were clear differences between the two groups (*Table 10*). Non-MEDLINE indexed reports were more likely to be found in Cochrane meta-analyses than in meta-analyses published in journals, more likely to be published in earlier years, and more likely to evaluate complementary medicine. They also tended to be smaller but were more likely to produce a statistically significant result.

The proportion of trials published in journals not indexed in MEDLINE varied across clinical topics, from 7.4% in cardiology and angiology to 28.9% in rheumatology and orthopaedics (*Table 11*). The proportion of non-indexed trials was greater in complementary medicine (40.9%) than in conventional medicine (20.5%; $p = 0.004$).

Assessments by Cochrane Collaboration reviewers of concealment of allocation was available for 339 trials (45.7%), while their assessment of blinding was available for 329 (44.4%) trials. Inter-observer reliability for extraction of these assessments by the present researchers was high, with kappas of 0.91 for concealment of allocation and 0.85 for blinding. As shown in *Table 12*, trials published in journals not indexed in MEDLINE were less likely to conceal allocation adequately although this difference did no reach conventional levels of statistical significance. There was little difference in the frequency of reported double- or assessor-blinding.

Estimates of treatment effects from non-indexed and indexed trials

Figure 16 shows the ratios of estimates of pooled treatment effects from non-MEDLINE trial reports compared with those from trials published in MEDLINE-indexed journals. Treatment effect estimates were on average 6% more beneficial in non-indexed trials (95% CI, 18% more beneficial to 7% less

TABLE 10 *Characteristics of randomised trials indexed in MEDLINE and those not indexed in MEDLINE*

	MEDLINE report (*n* = 580)	**Non-MEDLINE report (*n* = 161)**	***p***
Source of meta-analysis			
CDSR	315 (54.3%)	108 (67.1%)	0.011
General medical journal	194 (33.5%)	42 (26.1%)	
Specialist journal	71 (12.2%)	11 (6.8%)	
Year of publication of trial			
Mean (SD)	1985 (7.91)	1979 (13.27)	< 0.0001
Median (range)	1987 (1953–98)	1984 (1950–97)	< 0.0001
Type of intervention and comparison			
Drug intervention	425 (73.3%)	123 (76.4%)	0.48
Complementary medicine	26 (4.5%)	18 (11.2%)	0.004
Active control intervention	154 (26.6%)	24 (14.9%)	0.002
Sample size of trial			
Mean (SD)	257 (492)	232 (468)	0.55
Median (range)	114 (2–4865)	83 (10–3128)	0.006
Statistical significance of trial			
p < *0.05*	186 (32.1%)	63 (39.1%)	0.11
p < *0.01*	112 (19.3%)	45 (28.0%)	0.022

p-*values from chi-squared tests,* t-*tests or Wilcoxon rank sum tests*

TABLE 11 *Proportion of trials indexed in MEDLINE and those not indexed in MEDLINE by disease area*

Disease area	MEDLINE report (n = 580)	Non-MEDLINE report (n = 161)
Cardiology & angiology	113 (92.6%)	9 (7.4%)
Gastroenterology	18 (81.8%)	4 (18.2%)
Infectious diseases	70 (78.7%)	19 (21.4%)
Neonatology	8 (80%)	2 (20%)
Neurology	61 (76.3%)	19 (23.8%)
Obstetrics & gynaecology	87 (74.4%)	30 (25.6%)
Psychiatry	68 (68.0%)	32 (32.0%)
Rheumatology & orthopaedics	43 (70.5%)	18 (29.5%)
Tobacco addiction	80 (84.2%)	15 (15.8%)
Miscellaneous	32 (71.1%)	13 (28.9%)

p = *0.001 by chi-squared test*

TABLE 12 *Methodological quality of MEDLINE-indexed and non-indexed trials included in Cochrane reviews*

	MEDLINE report	Non-MEDLINE report	p
Adequate concealment of allocation			0.17
Yes	76/252 (30.2%)	19/87 (21.8%)	
No/unclear	176/252 (69.8%)	68/87 (78.2%)	
Double- or assessor-blinded			0.99
Yes	140/249 (56.2%)	45/80 (56.3%)	
No/unclear	109/249 (43.8%)	35/80 (43.8%)	

Denominators differ: information on concealment of allocation was provided more frequently than information on blinding. Probability by chi-squared tests

beneficial; $p = 0.35$). However, there was considerable heterogeneity between meta-analyses ($p < 0.001$), with pooled effect estimates of non-indexed trials ranging from 40% more to 400% less beneficial compared with indexed trials.

Results of stratified analyses are presented in *Figure 17*. The differences appeared to be more pronounced in complementary medicine, and less pronounced in trials with active control interventions, but none of the differences between strata was statistically significant ($p > 0.25$).

Impact of trials not indexed in MEDLINE on the results of meta-analyses

The number of trials per meta-analysis published in journals not indexed in MEDLINE ranged from 1 to 14 trials and from 3.3% to 77.8% of all trials included. Non-indexed trials contributed an average 23.3% of the weight in individual meta-analyses (median 15.6%; range, 0.5–91.2%). *Figure 18* shows the change in pooled estimates of individual meta-analyses that occurred when non-indexed trials were excluded from meta-analyses. The changes ranged from a 59.9% increase (indicating less benefit) to a 52.1% decrease (indicating more benefit) of estimates of treatment effects. However, in 32 (48%) meta-analyses the changes were less than 5%. Among the remaining 34 meta-analyses, 19 showed more benefit and 15 less benefit after exclusion of non-indexed trials. When the analysis is based on all meta-analyses with comprehensive literature searches ($n = 159$) then the percentage change in pooled estimates is zero or less than 5% in 125 (78.6%) meta-analyses.

The average precision of pooled effect estimates decreased from 7.49 to 6.52 after exclusion of trials not indexed in MEDLINE. Significance levels were affected in nine (14%) meta-analyses. In six cases p increased from less than 0.01 to less than 0.05, in one case p increased from less than 0.01 to greater than 0.05, in one instance p decreased from greater than 0.05 to less than 0.05 and in one instance p decreased from less than 0.05 to less than 0.01.

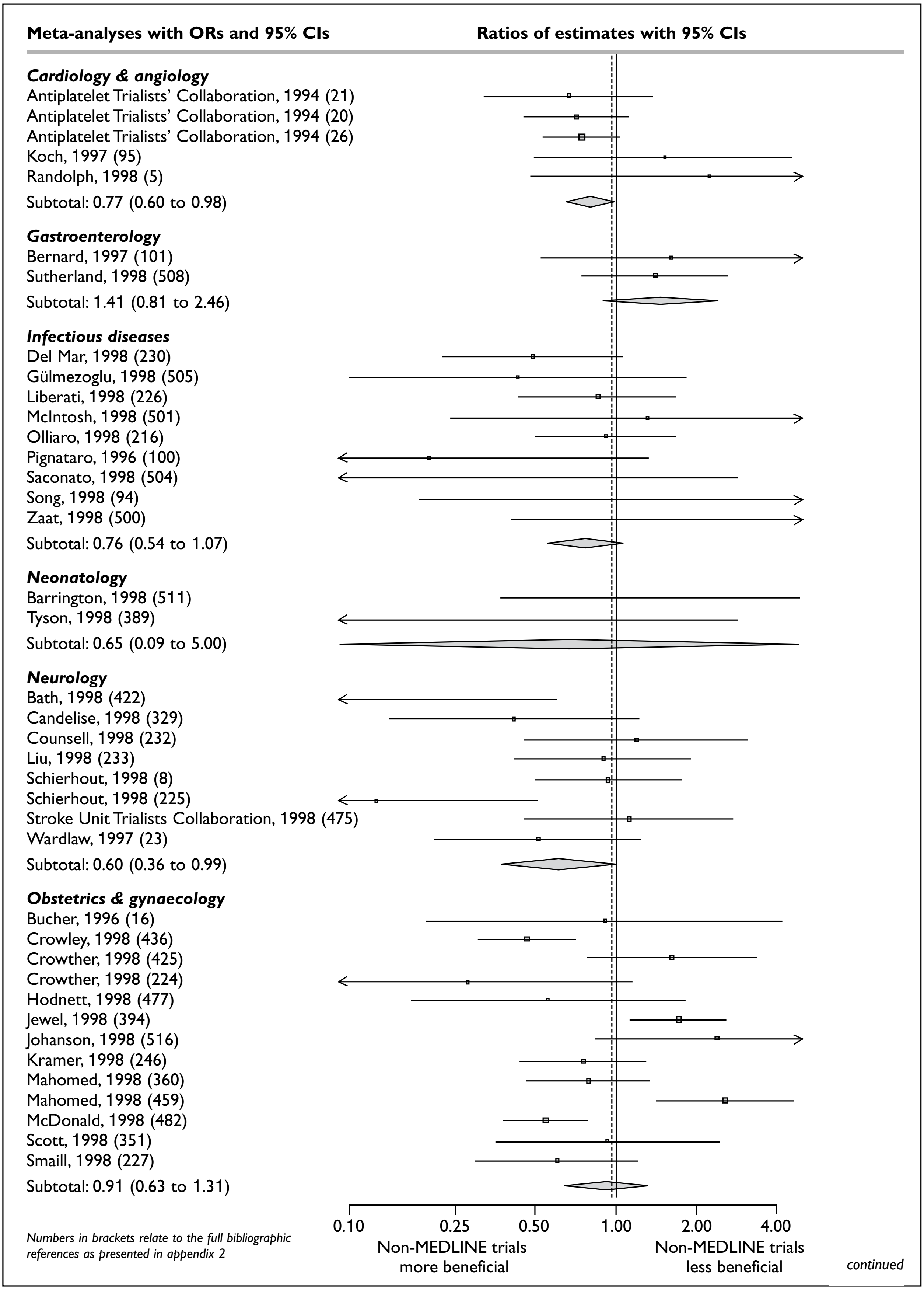

FIGURE 16 *Results from comparisons of treatment effect estimates from trials published in journals not indexed in MEDLINE with trials published in indexed journals in 66 meta-analyses. A ratio of estimates below 1.0 indicates that trials published in journals not indexed in MEDLINE show a more beneficial treatment effect than trials published in indexed journals*

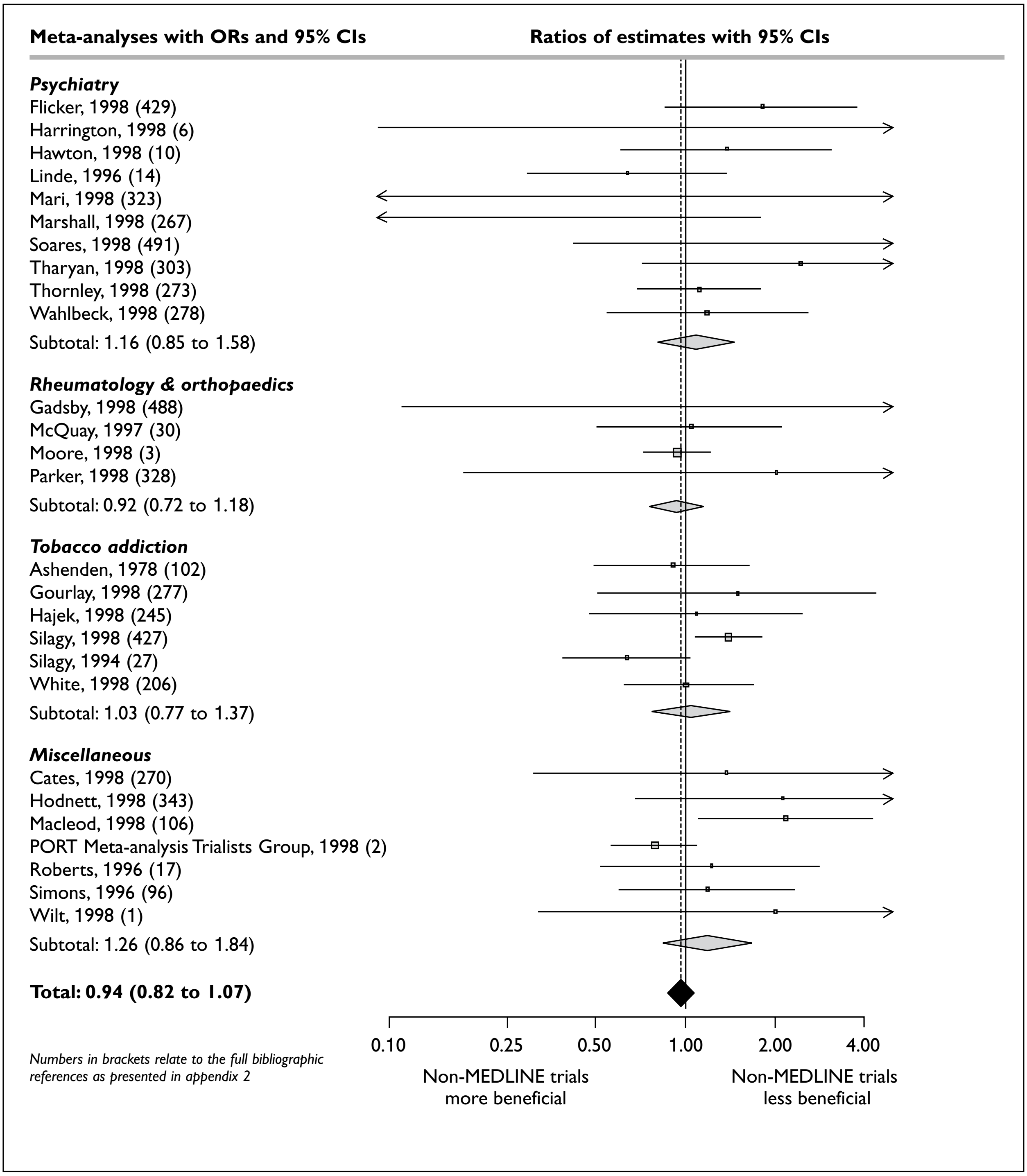

***FIGURE 16* contd** *Results from comparisons of treatment effect estimates from trials published in journals not indexed in MEDLINE with trials published in indexed journals in 66 meta-analyses. A ratio of estimates below 1.0 indicates that trials published in journals not indexed in MEDLINE show a more beneficial treatment effect than trials published in indexed journals*

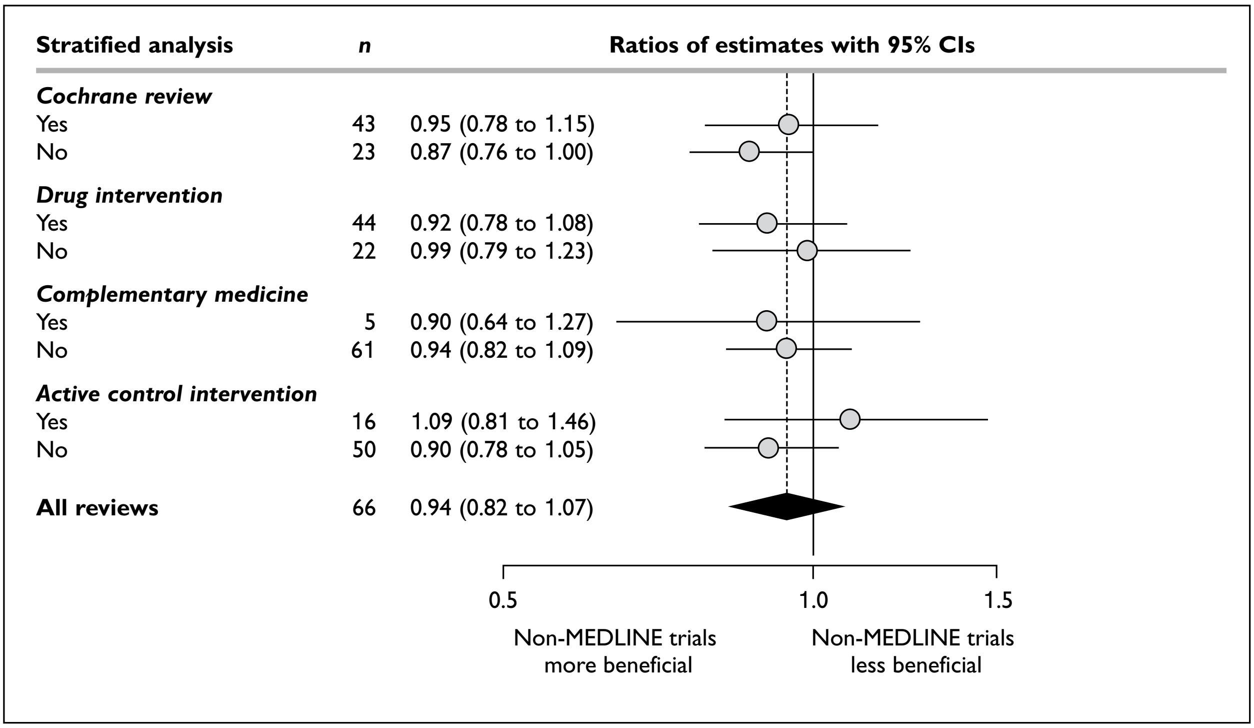

FIGURE 17 *Results from stratified analyses comparing treatment effect estimates of trials published in journals not indexed in MEDLINE with trials published in indexed journals in 66 meta-analyses. Ratios of estimates (circles) with 95% CIs of individual strata are shown. The black diamond represents overall pooled results. Estimates were pooled using random effects models. A ratio of estimates below 1.0 indicates that trials published in journals not indexed in MEDLINE show a more beneficial treatment effect than indexed trials*

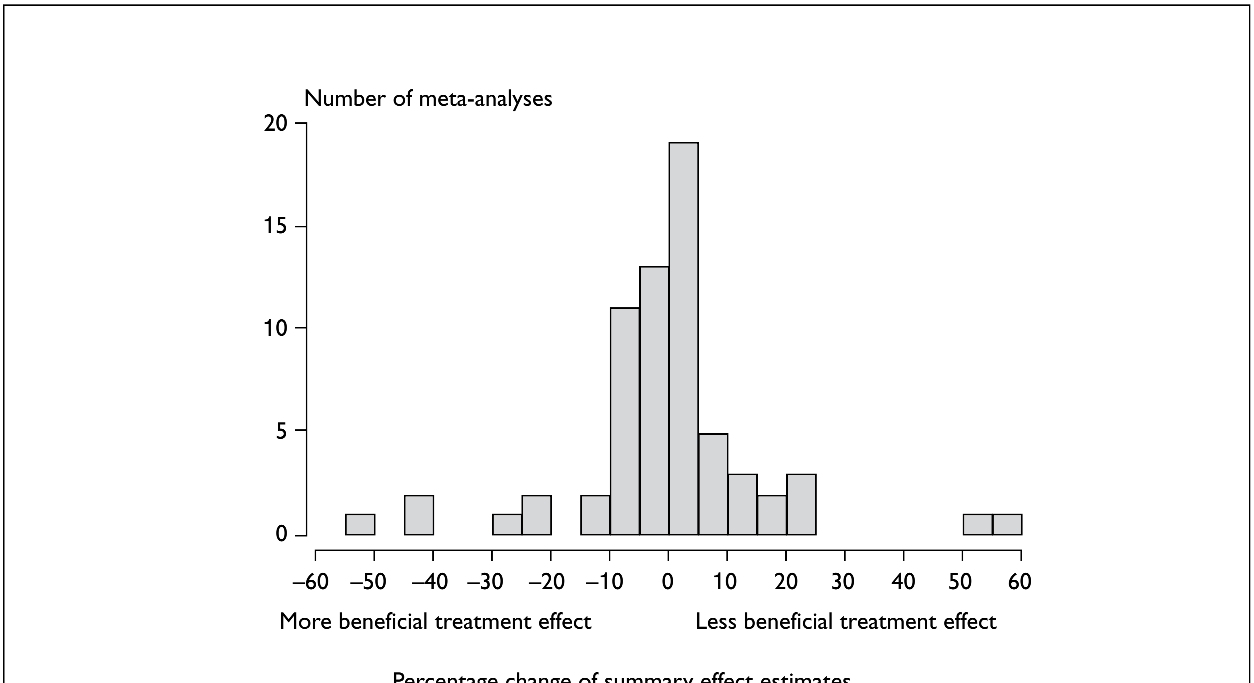

FIGURE 18 *Percentage change of treatment effect estimates of individual meta-analyses after exclusion of trials published in journals not indexed in MEDLINE. The histogram shows the frequency of percentage changes in pooled estimates that occurred when non-indexed trials were removed from meta-analyses*

Impact of trials not indexed in MEDLINE on the shape of funnel plots

This analysis was based on 62 meta-analyses and a median of 9 trials (range, 5–38). Four meta-analyses had to be excluded because the number of trials remaining after removal of unpublished trials was too small (less than four) to allow a meaningful funnel plot analysis. The combined asymmetry coefficient from meta-analyses including trials published in journals that are indexed in MEDLINE was –0.58 (95% CI, –0.75 to –0.41). There was thus clear evidence of funnel plot asymmetry when MEDLINE-indexed trials were included in the analysis. After excluding non-indexed trials funnel plots became slightly more symmetrical, with a combined asymmetry coefficient of –0.49 (95% CI, –0.63 to –0.36).

The impact of trial quality: concealment of allocation

This analysis was based on the 122 meta-analyses included in the CDSR (issue 1/1998) that included five or more trials with binary outcomes. We had to exclude 83 meta-analyses either because an assessment of concealment of allocation was not available in at least 80% of trials or because no differences were noted in the way allocation was concealed. Only trials published in English language journals were considered in this analysis. *Figure 19* shows the progress through the stages of identifying eligible meta-analyses and trials: 39 meta-analyses including 118 trials with adequate concealment and 186 trials with inadequate or unclear concealment were analysed.

Characteristics of trials

There were clear differences between the trials with adequate concealment and the other trials (*Table 13*). Trials that reported adequate concealment of allocation were published more recently and enrolled more participants than trials with inadequate or unclear concealment of allocation. Interestingly, there was no difference in the distribution of *p*-values, despite the clear difference in sample sizes.

There was also some variation across topics: trials with adequate concealment were more

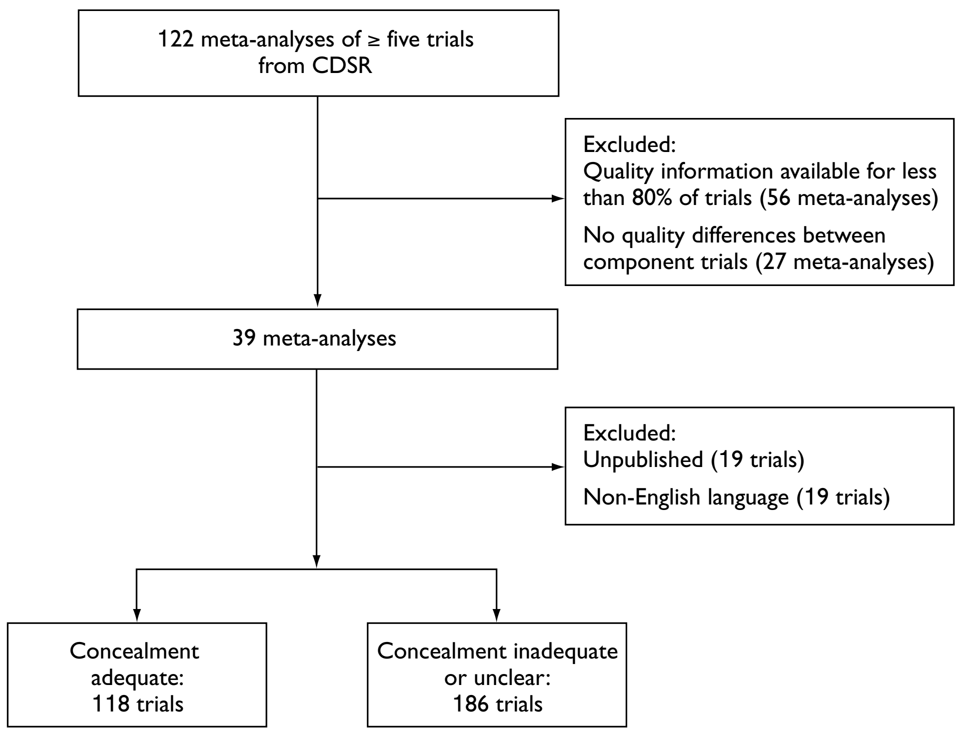

FIGURE 19 *Progress through the stages of identifying eligible meta-analyses for analyses of the impact of inadequate concealment of allocation*

TABLE 13 *Characteristics of randomised trials with adequate concealment of allocation and trials with inadequate or unclear concealment*

	Concealment adequate (n = 118)	Concealment inadequate or unclear (n = 186)	p
Year of publication of trial			
Mean (SD)	1987 (9)	1983 (11)	0.0004
Median (range)	1990 (1951–97)	1986 (1950–97)	0.0002
Type of intervention and comparison			
Drug intervention	97 (82.2%)	158 (84.9%)	0.53
Complementary medicine	0	0	–
Active control intervention	30 (25.4%)	70 (37.6%)	0.027
Sample size of trial			
Mean (SD)	382 (638)	200 (361)	0.002
Median (range)	154 (15–3510)	97 (2–2844)	< 0.0001
Statistical significance of trial			
p < 0.05	31 (26.3%)	52 (28.0%)	0.75
p < 0.01	16 (13.6%)	30 (16.1%)	0.54

p*-values from chi-squared tests,* t*-tests or Wilcoxon rank sum tests*

TABLE 14 *Adequacy of allocation concealment by medical speciality*

Disease area	Concealment adequate (n = 118)	Concealment inadequate or unclear (n = 186)
Infectious diseases	30 (54.5%)	25 (45.5%)
Neurology	18 (52.9%)	16 (47.1%)
Obstetrics & gynaecology	46 (37.7%)	76 (62.3%)
Other	24 (25.8%)	69 (74.2%)

p *= 0.002 by chi-squared test*

likely to be concerned with infectious diseases and neurological conditions than trials with inadequate or unclear concealment (*Table 14*).

Estimates of treatment effects from trials with inadequate/unclear concealment and trials with adequate concealment of allocation

Figure 20 shows the ratios of estimates of pooled treatment effects from adequately concealed trials compared with those from trials with inadequate or unclear concealment of treatment allocation. Treatment effect estimates were on average 21% more beneficial in the trials of lower methodological quality (95% CI, 11% to 30% more beneficial; $p < 0.001$). There was some evidence for heterogeneity between meta-analyses ($p = 0.01$).

Results of stratified analyses are presented in *Figure 21*. The differences were somewhat more pronounced when active control intervention were used but this difference may well have been produced by chance ($p = 0.37$).

Impact of trials with inadequate/unclear concealment of allocation on the results of meta-analyses

The proportion of trials with inadequate or unclear concealment of allocation in individual meta-analyses ranged from 8.3% to 88.9% with a median of 66.7%. The weight contributed by inadequately/unclearly concealed trials ranged from 1% to 97.7%, with a median of 50.9%. Based on these figures a considerable impact on pooled estimates is expected. *Figure 22* shows the change in pooled estimates of individual meta-analyses that occurred when trials of lower methodological quality were excluded from meta-analyses. The changes ranged from a 515% increase (indicating less benefit) to a 86% decrease (indicating more benefit) of estimates of treatment effects. In the majority of meta-analyses (29, 74%) exclusion of trials with inadequate/unclear concealment led to a change towards less beneficial treatment effects, which was often substantial (more than 10% in 21 or 54% of meta-analyses).

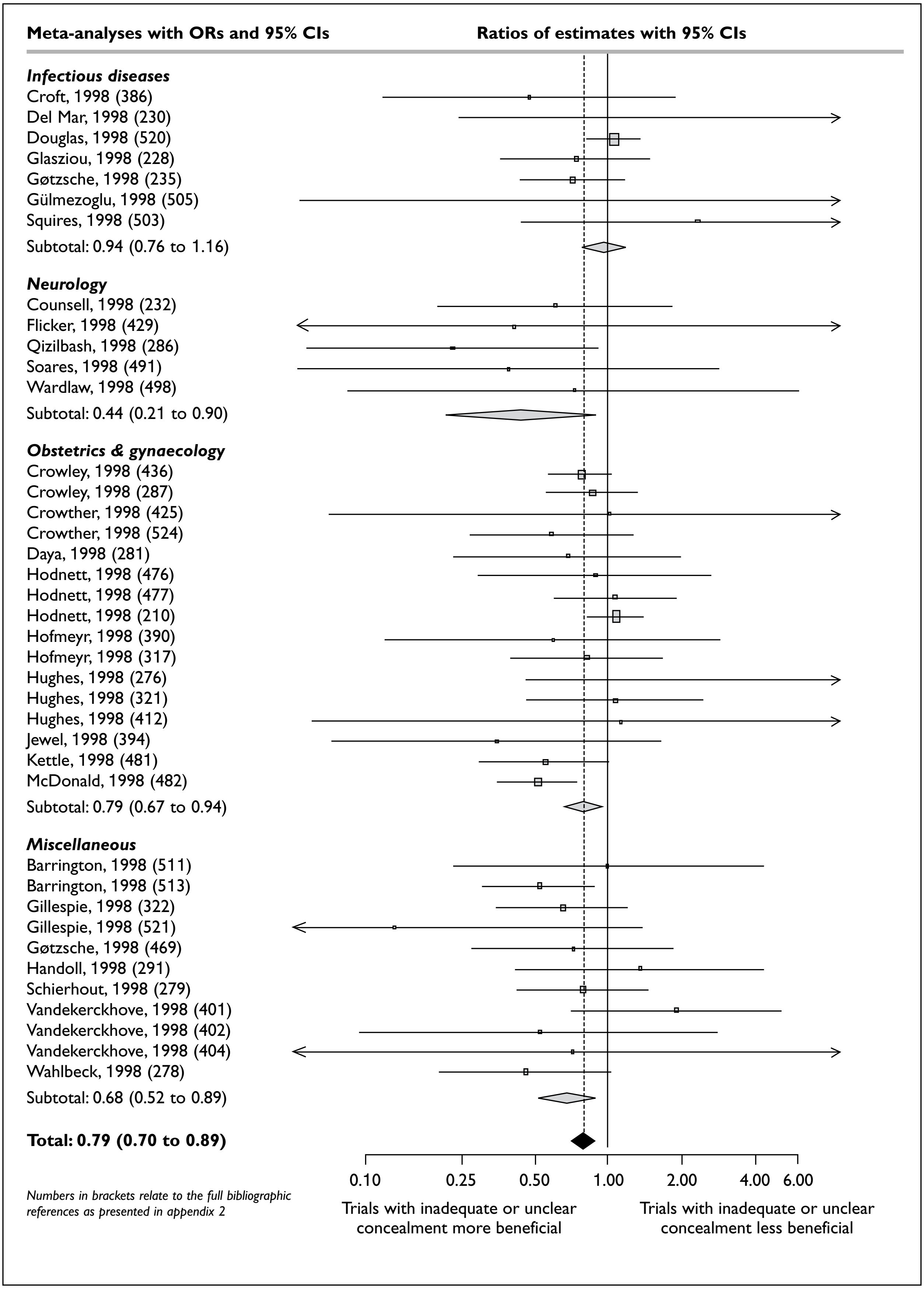

FIGURE 20 *Results from comparisons of treatment effect estimates from trials with inadequate or unclear allocation concealment with adequately concealed trials in 39 meta-analyses, calculating ratios of estimates. A ratio of estimates below 1.0 indicates that trials with inadequate or unclear allocation concealment show a more beneficial treatment effect than adequately concealed trials*

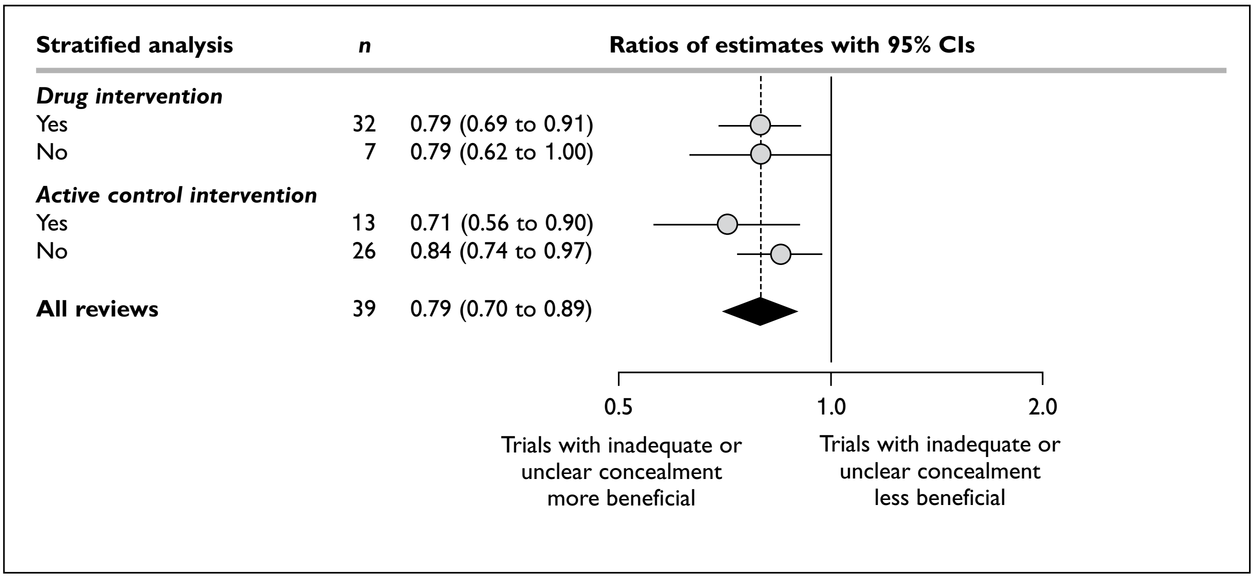

FIGURE 21 *Results from stratified analyses comparing treatment effect estimates of trials with inadequate or unclear concealment with adequately concealed trials. A ratio of estimates below 1.0 indicates that inadequately concealed trials show a more beneficial treatment effect than adequately concealed trials*

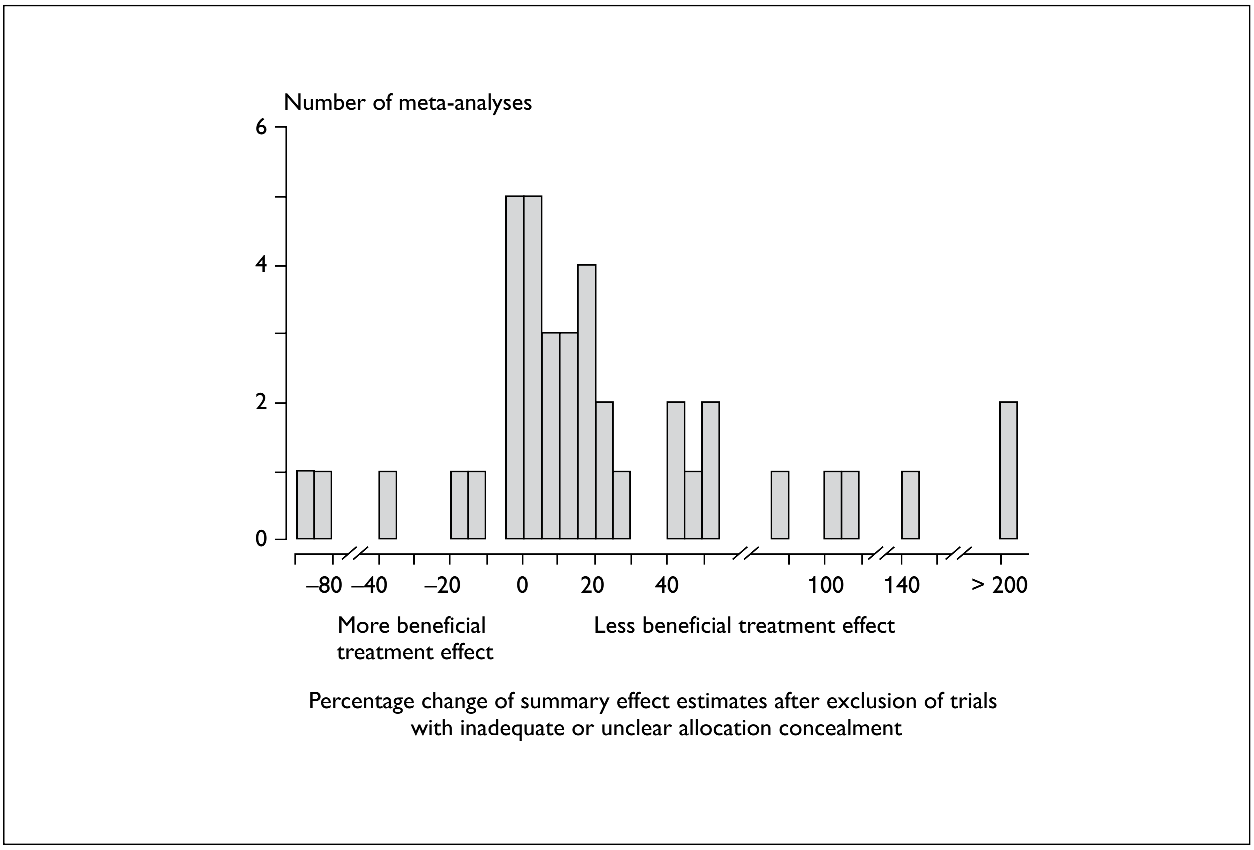

FIGURE 22 *Percentage change of treatment effect estimates of individual meta-analyses after exclusion of trials with inadequate or unclear allocation concealment. The histogram shows the frequency of percentage changes in pooled estimates that occurred when inadequately concealed trials were removed from meta-analyses*

The average precision of pooled effect estimates decreased from 7.09 to 4.97 after exclusion of trials with inadequate/unclear concealment. Statistical significance at the 5% level was affected in 16 meta-analyses (41%). In 15 cases *p* increased from less than 0.05 to greater than 0.05, in one case *p* decreased from greater than 0.05 to less than 0.05. At the 1% level significance was affected in 13 meta-analyses (33%): in all cases *p* increased from less than 0.01 to greater than 0.01.

Impact of trials with inadequate/unclear allocation concealment on the shape of funnel plots

This analysis was based on 18 meta-analyses only. Twenty-one meta-analyses had to be excluded because the number of trials remaining after removal of trials with inadequate or unclear concealment was too small (less than four) to allow a meaningful funnel plot analysis. The median number of trials in the remaining 18 meta-analyses was 4 (range, 4–11). The combined asymmetry coefficient including all trials was 0.069 (95% CI, –0.27 to 0.41). There was thus little evidence of asymmetry. After excluding the trials with inadequate or unclear concealment the plot became asymmetrical with a positive asymmetry coefficient of 0.97 (95% CI, 0.43 to 1.52), indicating that asymmetry was introduced by the removal of smaller trials showing relatively large treatment effects.

The impact of trial quality: double-blinding

This analysis was again based on the 122 meta-analyses included in the CDSR (issue 1/1998) that included five or more trials. We had to exclude 77 meta-analyses either because an assessment of blinding was not available in at least 80% of trials or because no differences in blinding were noted. Only trials published in English language journals were considered. *Figure 23* shows the progress through the stages of identifying eligible meta-analyses and trials: 45 meta-analyses including 237 trials described as double-blind and 162 trials that were not described as double-blind were analysed.

Characteristics of trials

There were few differences between double-blind trials and other trials (*Table 15*). Double-blind trials

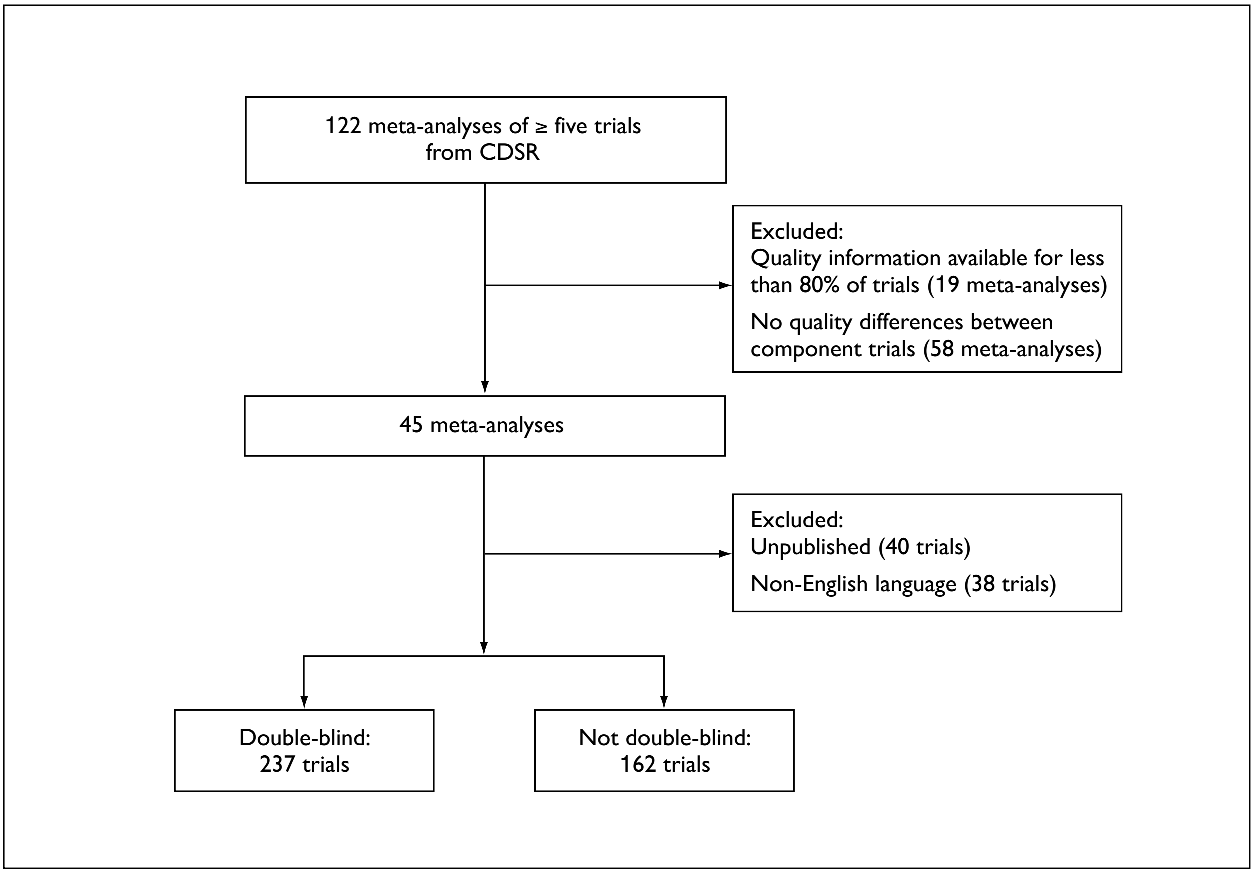

FIGURE 23 *Progress through the stages of identifying eligible meta-analyses for analyses of the impact of double-blinding*

TABLE 15 *Characteristics of double-blind trials and other trials*

	Double-blind (n = 237)	Not double-blind (n = 162)	p
Year of publication of trial			
Mean (SD)	1986 (10)	1983 (12)	0.005
Median (range)	1989 (1955–98)	1987 (1950–97)	0.005
Type of intervention and comparison			
Drug intervention	223 (94.1%)	156 (96.3%)	0.32
Complementary medicine	4 (1.7%)	2 (1.2%)	0.72
Active control intervention	41 (17.3%)	28 (17.3%)	0.99
Sample size of trial			
Mean (SD)	217 (545)	273 (539)	0.32
Median (range)	88 (8–4736)	101 (9–5042)	0.06
Statistical significance of trial			
$p < 0.05$	60 (25.3%)	47 (29.0%)	0.41
$p < 0.01$	37 (15.6%)	30 (18.5%)	0.45

p-values from chi-squared tests, t-tests or Wilcoxon rank sum tests

TABLE 16 *Blinding of trials by disease area*

Disease area	Double-blind (n = 237)	Not double-blind (n = 162)
Infectious diseases	37 (59.7%)	25 (40.3%)
Neonatology	18 (52.9%)	16 (47.1%)
Neurology	33 (57.9%)	24 (42.1%)
Obstetrics & gynaecology	23 (35.4%)	42 (64.6%)
Psychiatry	44 (84.6%)	8 (15.4%)
Other	82 (63.6%)	47 (36.4%)

$p < 0.0001$ *by chi-squared test*

were published more recently but there were no clear differences in sample sizes or the distribution of p-values.

There was also some variation across topics: double-blind trials were less likely in obstetrics and gynaecology but more likely in psychiatry (*Table 16*).

Estimates of treatment effects from double-blind trials and other trials

Figure 24 shows the ratios of estimates of pooled treatment effects from double-blind trials compared with other trials. Treatment effect estimates were on average 12% more beneficial in the trials of lower methodological quality (95% CI, 25% more beneficial to 4% less beneficial; $p = 0.13$). There was some evidence of heterogeneity between meta-analyses ($p = 0.051$), with pooled effect estimates of open trials ranging from 100% more to 493% less beneficial compared with double-blind trials.

Results of stratified analyses are presented in *Figure 25*. Interestingly, open trials showed larger effects for drug trials but produced less beneficial results for other types of interventions. Discordant results were also observed for complementary medicine. These analyses were based on only one meta-analysis from complementary medicine and three meta-analyses for non-drug intervention trials. The p-values from tests of interaction were 0.16 for complementary medicine versus other and 0.37 for non-drug interventions versus drug interventions.

Impact on the results of meta-analyses

The proportion of open trials in individual meta-analyses ranged from 5.9% to 85.7% with a median of 33.3%. The weight contributed to the meta-analysis ranged from 3.2% to 98.8%, with a median of 40.8%. *Figure 26* shows the change in pooled estimates of individual meta-analyses that occurred when trials without double-blinding

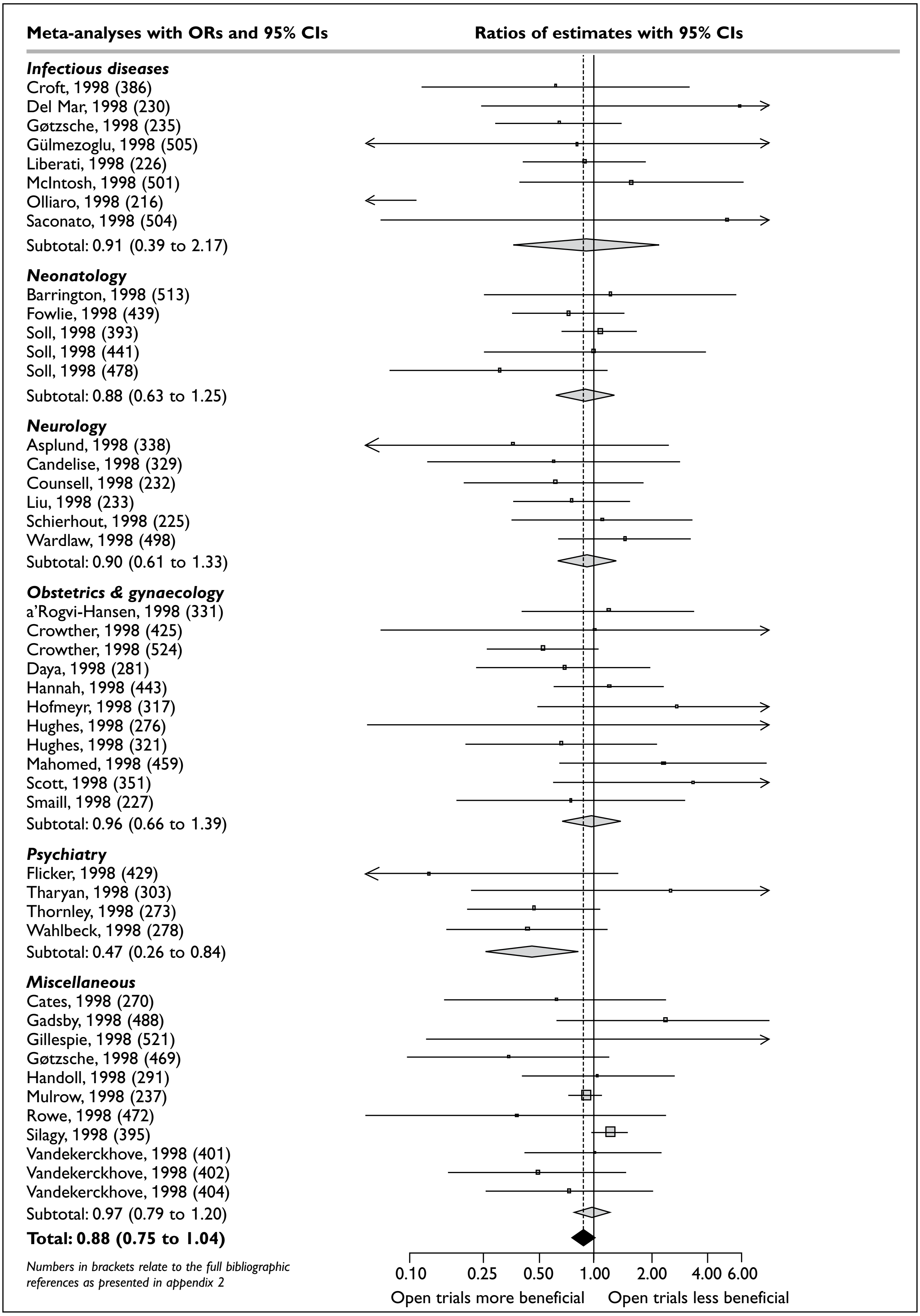

FIGURE 24 Results from comparisons of treatment effect estimates from double-blind trials with those from other trials in 45 meta-analyses, calculating ratios of estimates. A ratio of estimates below 1.0 indicates that open trials show a more beneficial treatment effect than double-blind trials

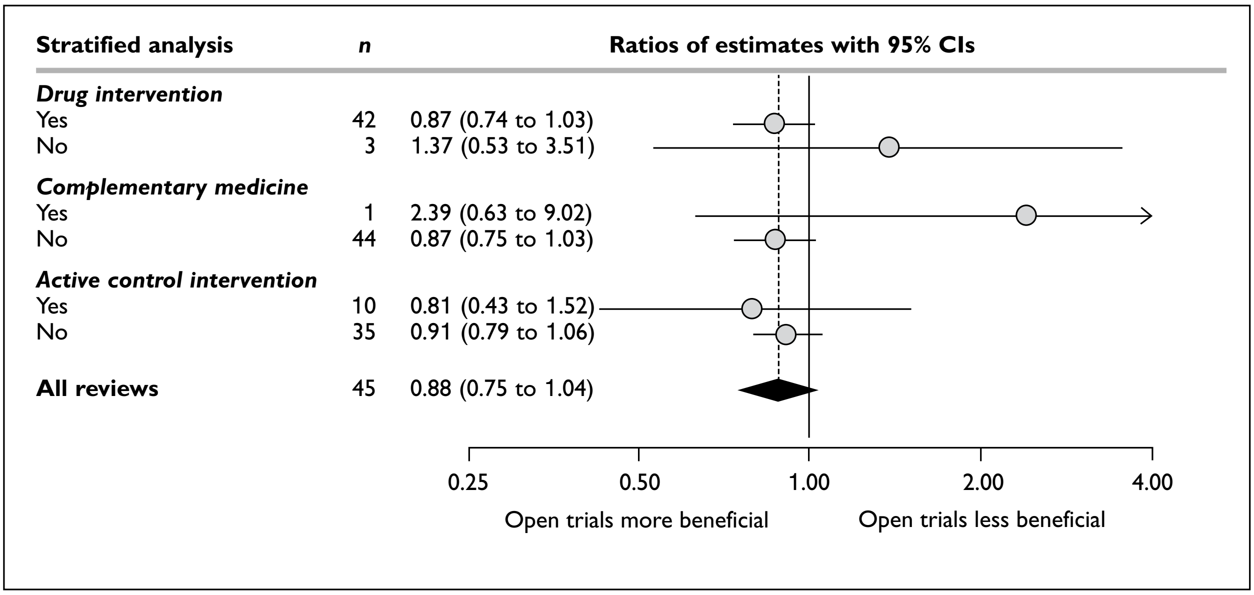

FIGURE 25 *Results from stratified analyses comparing treatment effect estimates of trials that were not double-blind with double-blind trials. Ratios of estimates with 95% CIs of individual strata are shown. A ratio of estimates below 1.0 indicates that open trials show a more beneficial treatment effect than double-blind trials*

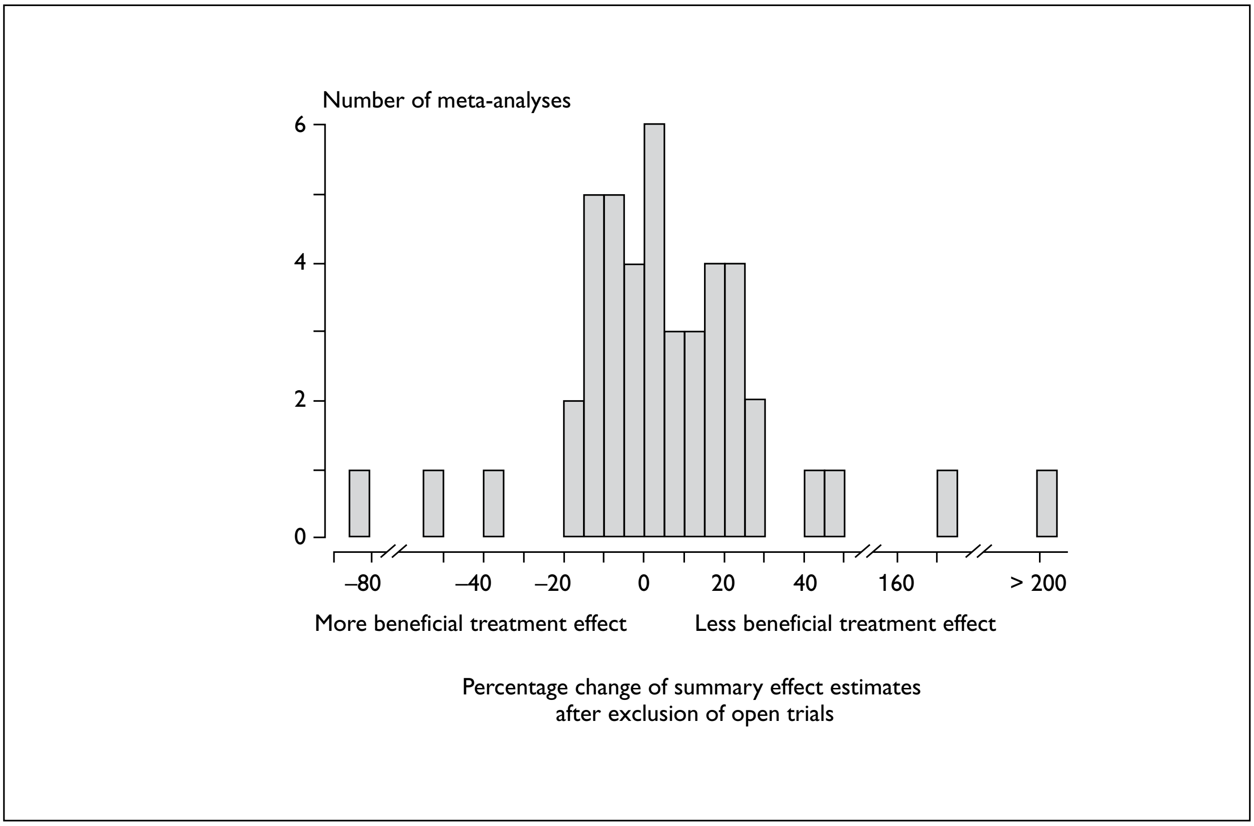

FIGURE 26 *Percentage change of treatment effect estimates of individual meta-analyses after exclusion of open trials. The histogram shows the frequency of percentage changes in pooled estimates that occurred when open trials were removed from meta-analyses*

were excluded from meta-analyses. The changes ranged from substantial increases (indicating less benefit) to decreases (indicating more benefit) in estimates of treatment effects. In the majority of meta-analyses (26, 58%) exclusion of trials without double-blinding led to a change towards less beneficial treatment effects, which was often substantial (more than 10% in 17 or 38% of meta-analyses). The average precision of pooled effect estimates decreased from 6.25 to 4.44 after exclusion of trials without double-blinding. Statistical significance at the 5% level was affected in six meta-analyses (13%). In all cases *p* increased from less than 0.05 to greater than 0.05. At the 1% level significance was affected in 12 meta-analyses (27%); in all cases *p* increased from less than 0.01 to greater than 0.01.

Impact of trials that were not double-blind on the shape of funnel plots

This analysis was based on 30 meta-analyses only. Fifteen meta-analyses had to be excluded because the number of trials remaining after removal of trials that were not double-blind was too small (less than four) to allow a meaningful funnel plot analysis. The median number of trials in the remaining 30 meta-analysis was 8.5 (range, 4–62). The combined asymmetry coefficient including all trials was negative: –0.16 (95% CI, –0.40 to 0.07). There was thus weak evidence of asymmetry, indicating that smaller trials produced somewhat larger treatment effects. After excluding trials that were not double-blind asymmetry was reduced: the asymmetry coefficient became –0.11 (95% CI, –0.32 to 0.10), indicating that asymmetry was reduced by the removal of smaller trials showing relative large treatment effects.

Sensitivity analyses using logistic regression

We repeated analyses using logistic regression to allow comparison with previous meta-epidemiological studies,[49,50,63,64] and to control analyses of reporting biases for confounding by methodological quality.

To use logistic regression the numbers of patients and events in each group must be tabulated in the report of the meta-analysis. Meta-analyses were only included in the sensitivity analyses if, in addition, the component trials reported at least one event in both the treatment and control groups in trials with and without the characteristic. The data set was restricted to meta-analyses that employed comprehensive literature searches and were published either in the CDSR, or in a medical journal. We omitted other meta-analyses, because of limitations in the number of covariates which can be included in logistic regression models. Multivariate analyses were restricted to meta-analyses from the CDSR that included information on the methodological quality of component trials.

Table 17 compares the effects of publication status, language and indexing of trials on treatment effect estimates, using fixed effects logistic regression and the meta-analytical approach used in the main analysis.

Table 18 presents crude estimates of effects of publication status, language and indexing of trials on treatment effects and estimates adjusted for methodological quality, using fixed effects logistic regression.

TABLE 17 *Comparison of effects of publication status, language and indexing of trials on treatment effect estimates, using fixed effects logistic regression and the meta-analytical approach used in the main analysis*

	Fixed effects logistic regression		Random effects meta-analysis	
	Ratio of ORs (95% CI)	***p***	**Ratio of effect estimates (95% CI)**	***p***
Unpublished vs published	1.11 (1.01 to 1.23)	0.031	1.13 (1.01 to 1.27)	0.035
Non-English vs English	0.81 (0.72 to 0.90)	< 0.001	0.82 (0.70 to 0.96)	0.016
Non-indexed vs MEDLINE-indexed	0.94 (0.86 to 1.04)	0.23	0.93 (0.81 to 1.08)	0.35

Analyses based on 49 meta-analyses and 673 trials (unpublished vs published), 36 meta-analyses and 461 trials (non-English vs English) and 50 meta-analyses and 591 trials (non-indexed vs MEDLINE indexed)

TABLE 18 *Effects of publication status, language and indexing of trials on treatment effect estimates, with and without controlling for trial quality*

	Crude		Controlled for trial quality*	
	Ratio of ORs (95% CI)	***p***	**Ratio of ORs (95% CI)**	***p***
Unpublished vs published	1.12 (0.99 to 1.26)	0.068	1.14 (1.01 to 1.28)	0.035
Non-English vs English	0.78 (0.67 to 0.91)	0.001	0.81 (0.70 to 0.95)	0.010
Non-indexed vs MEDLINE-indexed	0.98 (0.88 to 1.10)	0.75	1.00 (0.89 to1.12)	0.95

* *Controlling for concealment of allocation and blinding*

Analyses based on 39 meta-analyses and 482 trials (unpublished vs published), 29 meta-analyses and 349 trials (non-English vs English) and 42 meta-analyses and 476 trials (non-indexed vs MEDLINE-indexed)

Chapter 4
Discussion

It has long been understood that the results of systematic reviews and meta-analyses may be undermined by reporting biases and the poor methodological quality of trials,[67,68] but knowledge of the extent, nature and relative importance of these biases is still limited. In an ideal world reviews of medical research would always include all relevant studies of acceptable quality, independent of publication status or language of publication. However, in the real world unpublished trials, trials published in languages other than English and trials published in journals that are not indexed in MEDLINE are difficult to locate, and may require translation, which will increase costs and delay the conclusion of a review. Although performing reviews that produce misleading results is never justified, there may be trade-offs between the timeliness, costs and quality of systematic reviews.

The effects of bias cannot be estimated precisely in individual meta-analyses, which typically contain only small numbers of trials.[52] The imprecision of estimates from individual meta-analyses, again illustrated in the present study (see for example *Figure 8*) means that it is necessary to combine evidence from many meta-analyses in order to estimate the effects of factors such as publication status or language of publication on treatment effect estimates. We identified a large number of state-of-the-art meta-analyses that were based on comprehensive literature searches and examined the contribution made by trials that are difficult to locate, as well as their methodological quality.

Principal findings

The main findings relating to reporting biases can be summarised as follows.

- A substantial proportion of state-of-the-art systematic reviews do not actually include trials that are difficult to locate, despite comprehensive literature searches.
- The importance of trials that are difficult to locate appears to vary across medical specialities. For example, unpublished trials are particularly prevalent in oncology, whereas trials published in languages other than English and trials not indexed in MEDLINE are important in psychiatry, rheumatology and orthopaedics. Trials in complementary medicine are frequently difficult to locate.
- Unpublished trials are smaller and less likely to produce statistically significant results than published trials. Conversely, non-English language trials and non-indexed trials are more likely to produce statistically significant results, despite smaller sample sizes.
- Similarly, unpublished trials tend to show less beneficial effects than published trials, whereas non-English language trials and non-indexed trials show larger treatment effects.
- Trials that are difficult to locate tend to be of lower methodological quality than trials that are easily accessible and published in English.
- Including unpublished trials reduces funnel plot asymmetry whereas the inclusion of trials published in languages other than English and of non-indexed trials increases the degree of asymmetry in the funnel plot.
- In the majority of meta-analyses excluding trials that are unpublished, not indexed in MEDLINE or published in languages other than English has only relatively small effects on estimates of treatment effects and the precision of these estimates, although more substantial changes were observed in some instances.

Our findings regarding the methodological quality of trials were as follows.

- In only about 40% of trials was it clear that allocation of participants to treatment groups was adequately concealed and only about 60% of trials were double-blind.
- Adequately concealed and double-blind trials were published more recently than open trials and trials with inadequate or unclear concealment of allocation, indicating that the quality of trials has improved in recent years.
- Trials with inadequate or unclear concealment are smaller than adequately concealed trials but there was no difference in the proportion of trials with statistically significant results.
- Trials with inadequate or unclear concealment of allocation show more beneficial effects than adequately concealed trials. Similarly,

open trials tend to be more beneficial than double-blind trials.

- In the majority of meta-analyses exclusion of trials of lower methodological quality led to a change towards less beneficial treatment effects, which, unsurprisingly, was often substantial. The precision of estimates was also reduced substantially, reflecting the relatively large number of trials of doubtful quality included in these reviews.
- The impact on the funnel plot was also substantial, particularly when including or excluding trials with inadequate or unclear concealment of allocation.

Strengths and weaknesses

To our knowledge this is the largest and most comprehensive study to date into the reporting and dissemination biases that distort the evidence from RCTs. To increase the generalisability of results we examined a wide range of sources of meta-analyses, including the journals that are known to publish many meta-analyses, the CDSR, DARE and *Health Technology Assessment*. Cochrane reviews dominated analyses, which reflects the fact that they employ comprehensive literature searches and therefore met our inclusion criteria whereas many meta-analyses published in journals had to be excluded. However, stratified analyses showed that among included reviews effects were similar for Cochrane reviews and reviews identified from other sources. For practical reasons the analyses of the impact of trial quality were restricted to the Cochrane sample and based on assessments made by Cochrane reviewers. This made it possible to consider both reporting biases and bias associated with the often inadequate methodological quality of trials, and to gain an understanding of the interrelations between different sources of bias.

We stress that our results are applicable only to meta-analyses where five or more trials have been located. As recently pointed out by Clarke,[69] such meta-analyses will be more robust to the effects of removing one or two trials than meta-analyses with fewer trials. Meta-analyses of few trials are not uncommon and it might be expected that the exclusion of, for example, one unpublished trial from a total of three could have a larger effect.

We were interested in the effect of bias on the results of meta-analyses as performed by the original reviewers and restricted our analysis to systematic reviews with comprehensive literature searches. In other words, we asked what would have happened to the results of a meta-analysis, had the literature search been less comprehensive, keeping everything else constant. We performed a 'meta-analysis of meta-analyses' for this purpose, which allows the inclusion of all meta-analyses, independently of whether or not the numbers of patients and events in each group were reported, using the same summary statistic chosen by the original reviewers. The exact replication of the analyses performed by the original reviewers is important because it allows an unbiased assessment of the impact of the reviewers' decision to perform a comprehensive literature search. Changing the statistical methods or summary statistics may introduce bias: reviews that used random effects models may have done so because of unexplained between-trial heterogeneity and it would be inappropriate in this situation to analyse the data using a fixed effects model. Similarly, outcome measures will generally have been selected because they are most appropriate in that particular context, for example in the case of the hazard ratio in IPD meta-analyses.

Previous studies[49,50,63,64] addressing similar questions used a fixed effects logistic regression model. This approach requires the raw data from each trial, expresses results on the OR scale only, and ignores heterogeneity between-trials and between meta-analyses. Furthermore, the need to include an indicator variable for each trial and each meta-analysis in the data set means that estimation is slow and that the number of variables required may reach the limits permitted in standard statistical packages. In sensitivity analyses we compared the effects of publication status, language and indexing of trials on treatment effect estimates, using fixed effects logistic regression and the meta-meta-analytical approach used in the main analysis (*Table 17*). It is clear from these comparisons that the logistic regression analysis will tend to underestimate standard errors because of the presence of between-meta-analysis heterogeneity. These analyses thus support the notion that between-meta-analysis heterogeneity should be allowed for in the analysis of meta-epidemiological studies. In a methodological paper[65] based on data from this and another study[49] we discuss these statistical issues in more detail and argue that too little consideration has so far been given to appropriate statistical methods for this type of meta-epidemiological research. An approach similar to ours has recently been developed by David Moher's group (Children's Hospital of Eastern Ontario Research Institute: personal communication, July 2002).

The different biases that affect systematic reviews and meta-analyses are unlikely to operate independently. For example, our findings indicate that trials published in languages other than English tend to show more beneficial treatment effects but such trials also tend to be of lower methodological quality, which may explain the larger treatment effects. On the other hand, the smaller effects observed in unpublished trials may not be an accurate reflection of the effect of publication bias considering that unpublished trials also tend to be of lower methodological quality. We made an attempt to control for such confounding by controlling for the effects of trial quality, using the logistic model described above (*Table 18*). We found little differences between crude and adjusted estimates, possibly because our assessment of trial quality relied on information derived by many different Cochrane reviewers. Such assessments are unlikely to be consistent, or consistently reliable, across reviews, despite the standardised guidelines specified in the *Cochrane Reviewers' Handbook*.[56] Also, we only selected Cochrane reviews that reported on the quality of trials, which may have introduced selection bias. Despite these shortcomings, the impact of trial quality on estimates of treatment effects was in line with previous studies[49,50,70] in which quality had consistently been assessed by the same observers (see also *Figure 27* below). Finally, reporting on important methodological detail is often incomplete in trial reports. For example, in many trials it remained unclear whether concealment of allocation was adequate or inadequate. We considered these trials in the same category as trials with clearly inadequate concealment, which, based on Schulz and co-workers' results,[49] is justified. Classification regarding blinding relied on description of trials as 'double-blind'. This term implies that neither the caregiver nor the patient knows which treatment was received; however, it is ambiguous with regard to blinding of other persons, including those assessing patient outcome.[71] Misclassification bias may thus have been introduced in our analysis. Authors should clearly state who exactly was blinded (participants, care providers, evaluators, or data analysts) and the methods used to achieve blinding.[72]

A weakness of our study relates to its retrospective nature and its reliance on what authors described as comprehensive literature searches. We did not assess whether the sample of trials identified by these authors was in fact complete and whether searches were truly comprehensive. If searches were inadequate, so that many unpublished trials, or published trials that were difficult to locate were omitted then our results might underestimate the potential impact of reporting bias. Our sample was, however, large and our inclusion criteria well defined and stringent, reflecting current recommendations for comprehensive, state-of-the-art searches. The results reported here should thus reflect what is gained or lost by attempts to identify unpublished trials, trials published in languages other than English and trials published in journals not indexed in MEDLINE.

To our knowledge this is the first study examining the effects of publication status, language and indexing of trials on the shape of funnel plots. Our results show that the funnel plot is affected in a manner that is entirely predictable considering the differences observed in the size and results of trials. Asymmetry was reduced upon inclusion of unpublished trials but increased when non-English language trials or non-indexed trials were added to the plot. Asymmetry coefficients are, however, not directly comparable between analyses because samples differed. For example, unpublished trials were excluded from the language and MEDLINE samples because they could not be classified regarding the language of publication and were by definition not indexed in MEDLINE.

The present study in context

As mentioned earlier, there are several previous studies, published and unpublished, that have examined the influence of unpublished trials, trials published in languages other than English, and of trial quality on the results of systematic reviews and meta-analyses of RCTs. We performed a meta-analysis of all studies we are aware of in order to put our results in context with the existing evidence. We may have missed some studies: the conduct of a formal systematic review was outside the scope of this project. Furthermore, there may be some overlap in the meta-analyses included in these studies. The authors of two studies[49,70] kindly provided us with unpublished data, which made consistent definitions and coding across studies possible. There were two studies assessing publication bias (McAuley and co-workers[63] and our study), two studies on language bias (Moher and co-workers[64] and our study), the present study on MEDLINE bias, and four studies each on the importance of concealment of allocation and double-blinding (Schulz and co-workers,[49] Moher and co-workers,[50] Kjaergard and co-workers[70] and the present study). As shown in *Figure 27* results were fairly homogeneous (despite the differences in statistical methodology discussed above) and

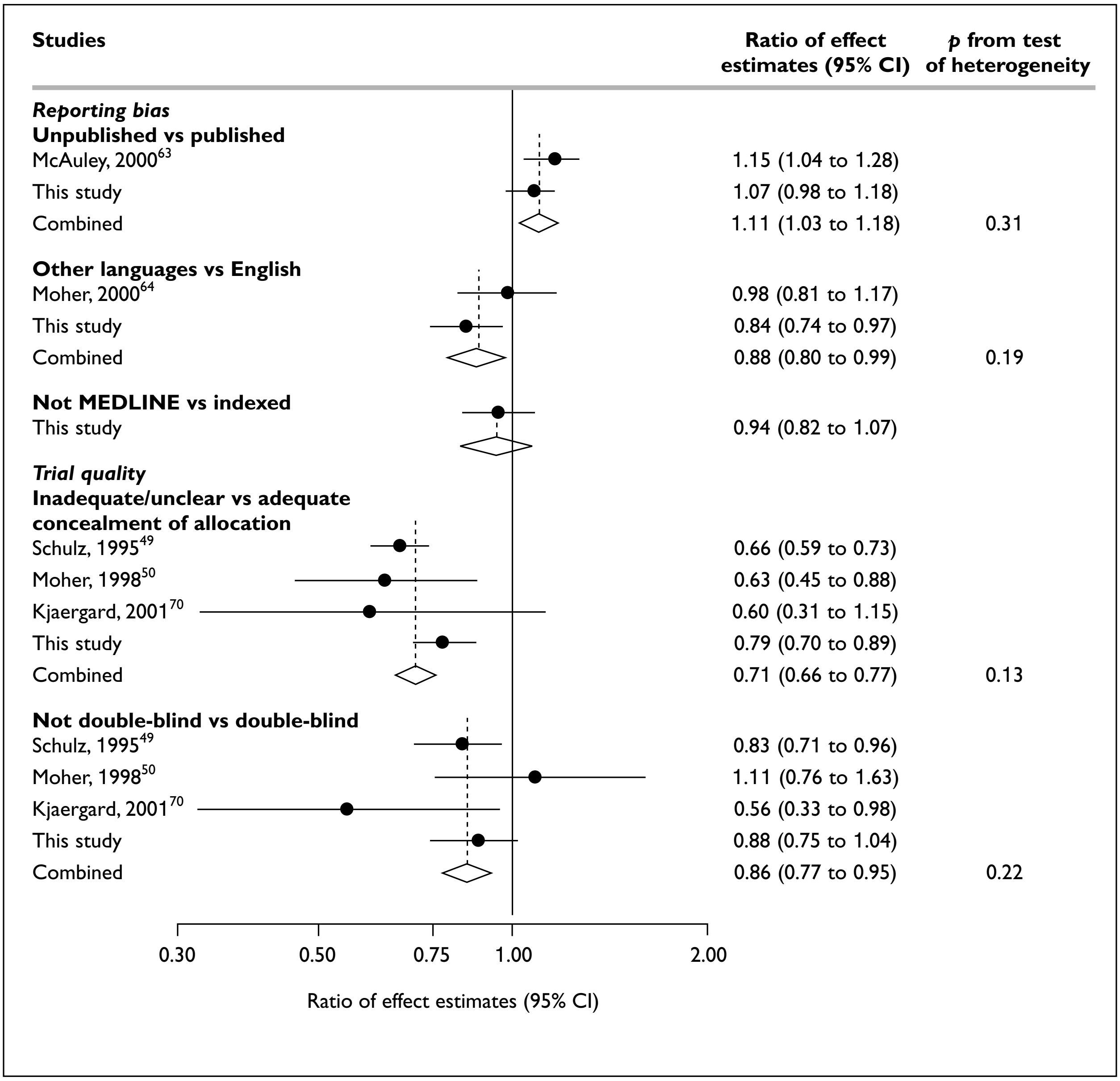

***FIGURE 27** Meta-analysis of empirical studies of reporting bias and trial quality. All studies compared estimates of treatment effects within a large number of meta-analyses and calculated ratios of effect estimates for this purpose. A ratio of estimates below 1.0 indicates that trials with the characteristic (e.g published in a language other than English) showed a more beneficial treatment effect. Adapted from Egger* et al.[80] *by permission of Oxford University Press*

formal tests of heterogeneity were non-significant ($p > 0.10$). Combined results indicate that, on average, unpublished trials will produce 11% less beneficial treatment effects, trials published in languages other than English will show 12% more beneficial effects and trials not indexed in MEDLINE 6% more beneficial effects, although the latter result did not reach conventional levels of statistical significance. Effects of trial quality were more pronounced; trials with inadequate or unclear concealment of allocation and trials that are not double-blind produce treatment effects that on average are 29% and 14% more beneficial (*Figure 27*).

In contrast to Moher and co-workers[44] we found that trials published in languages other than English tend to be of lower quality than studies published in English language journals. Moher compared 133 trials published in English with 98 trials published in other languages during 1992 to 1994 and found little differences in reporting and overall quality. Their study was based on 13 selected journals of relatively high impact whereas our sample included a much wider range of journals (208 journals published in English and 95 journals published in other languages). Moher[44] used the scale developed by Jadad and co-workers[73] to gauge quality.

This scale gives more weight to the quality of reporting, that is the extent to which a report of a clinical trial provides adequate information about the design, conduct, and analysis of the trial than to actual methodological quality. Furthermore, the Jadad scale addresses the generation of allocation sequences, a domain not consistently related to bias,[74] but it does not assess allocation concealment, which has repeatedly been shown to be associated with exaggerated treatment effects (*Figure 27*). It thus seems likely that the discrepant findings are explained by differences in the samples examined and quality features assessed.[75]

In an earlier investigation we examined factors predicting the language of publication for pairs of reports of RCTs, with one report published by the same author in German and the other in English.[43] A statistically significant result was the only characteristic that predicted publication in an English language journal. Based on these findings we hypothesised that significant findings are over-represented in the English language literature whereas more non-significant results would be found in journals published in other European languages. The present study not only failed to confirm this prediction but showed that articles published in languages other than English were more likely to report statistically significant findings. Trialists in German-speaking Europe who publish both in English and German may thus not be representative of the majority of authors publishing clinical trials in languages other than English.

Clarke pointed out that in the present study the trials published in languages other than English were all identified for inclusion in meta-analyses and it is therefore possible that 'positive' trials published in other languages were more likely to be identified, for example because these trials were more widely cited in the English language literature than 'negative' trials.[69] Such differential publication bias could explain the tendency of trials published in languages other than English to report statistically significant findings more often than trials published in English. As discussed elsewhere[75] an alternative, and perhaps more likely, explanation lies in the lower methodological quality of trials published in other languages, which would be expected, on average, to lead to more beneficial treatment effects in these trials.

The proportion of published trials showing superior efficacy of the experimental treatment has been shown to vary from country to country. Vickers and co-workers examined 252 abstracts of clinical trials of acupuncture and 405 abstracts from trials of other interventions.[34] They found unusually high proportions of trials favouring experimental treatments in some countries, for example China, Russia and Taiwan. Our sample included only few reports published in these countries but our results indicate that journals published in western Europe may also contain a relatively high proportion of 'positive' trials.

Implications for research

Our findings have important implications for the conduct of systematic reviews. First, systematic reviews that are based on a search of the English language literature that is accessible in the major bibliographic databases will often produce results that are close to those obtained from reviews based on more comprehensive searches that are free of language restrictions. This is certainly the case for specialities where most relevant trials appear to be published in English, for example cardiology or obstetrics and gynaecology. On the other hand, it is clear that in some areas of medicine it is essential to broaden the search to include the grey literature and material published in languages other than English. The importance of unpublished trials and trials published in languages other than English is well known in complementary medicine, for example homoeopathy[76,77] or phytotherapy.[78] Our results indicate that unpublished trials are also important in oncology, whereas non-English language trials are particularly prevalent in psychiatry, rheumatology and orthopaedics. We stress that in any review it is possible that a study which would affect the conclusions is missed if the search is not comprehensive and free of language restrictions, although this appears to be a relatively rare situation in many areas of conventional medicine.

We recommend that when planning a review, investigators should consider the type of literature search and the degree of comprehensiveness that are likely to be appropriate for the review in question, taking into account budgetary and time constraints. Assessments should be based on preliminary searches, existing reviews and advice from people with expertise in the area that is reviewed. Whenever possible, reviewers should attempt to include all relevant trials of acceptable quality. The inclusion of unpublished trials and trials published in different languages will increase precision, generalisability and applicability of findings. The exclusion of trials on the grounds that they are difficult to locate discriminates against some investigators and countries and will always introduce an element of doubt. Reviewers should take into account that thanks to the CCTR which contains over a quarter of a million reports of controlled trials, and registers of unpublished trials, the identification of all relevant trials has become an easier task.[79] If literature searches have to be restricted then reviewers should discuss the possible implications. The results from the present study should inform these considerations and facilitate sensitivity and scenario analyses.

Second, the low quality of trials that are difficult to locate is an important finding, which raises the worrying possibility that rather than preventing bias through extensive literature searches, bias could be introduced by including trials of low methodological quality. If all resources are spent on extensive literature searches and no careful assessment of the methodological quality of candidate trials is performed, then bias may well be introduced. This happened in a review[14] of trials comparing LMW heparin with standard heparin for the prevention of postoperative deep vein thrombosis which erroneously concluded that LMW heparins have higher benefit to risk ratio in preventing perioperative thrombosis.[22] The results from the present and other studies (see *Figure 27*) underscores the crucial importance of a sound assessment of trial quality. We believe that in situations where resources are limited, thorough quality assessments should take precedence over extensive literature searches and translations of articles.[80] Two bibliographic studies[81,82] have recently shown that only a minority of meta-analyses published in medical journals assessed trial quality whereas Cochrane reviews always included some form of quality assessment.

There is debate on how trial quality should best be assessed.[74,83–86] Quality scales combine information on several features in a single numerical value whereas the component approach examines key dimensions individually, without calculation of a summary score. Published scales vary considerably in terms of dimensions covered, size and complexity and many scales include items that are not in fact related to the internal validity of a trial.[87] Even when based on relevant items the interpretation of summary scores is difficult. In the absence of an association between treatment effects and the summary score, associations with one or several components may still exist, but these components will often contribute little weight to the summary score. Also, associations between two or more components may cancel out. In the presence of an association investigators will always have to identify the component or components that are responsible for this association in order to interpret this finding. For these reasons we recommend an assessment based on individual components of study quality, which is transparent, avoids the problem of weighting individual items and takes into account that the importance of individual quality domains varies between the contexts in which trials are performed.

Independent of the method used, the assessment of trial quality is hampered by the fact that reports frequently omit important methodological detail.[44,88–92] This situation has been improving somewhat in recent years with the adoption of the CONSORT guidelines.[93–97] Special efforts may be needed to improve reporting of clinical trials in journals published in languages other than English.

Third, our results confirm that the funnel plot and the regression method to assess funnel plot asymmetry are potentially useful to detect bias in systematic reviews and meta-analyses.[3] As mentioned above, the asymmetry coefficients and changes in coefficients calculated in this study cannot directly be compared with each other. It is nevertheless noteworthy that the most pronounced change in the shape of the funnel plot was not observed when adding unpublished trials but when including trials with inadequate or unclear concealment of allocation. These results lend empirical support to the notion that the funnel plot should be seen as a generic means of examining 'small-study effects', the tendency for smaller studies in a meta-analysis to show larger treatment effects, rather than as a tool to diagnose specific types of bias.[3,52,54]

Recommendations for future research

Our study was designed to examine the overall impact of unpublished trials, trials published in languages other than English and trials not indexed in MEDLINE. We found that the importance of trials that are difficult to locate varies not only between conventional and complementary medicine but also within conventional medicine. Further research is required to clarify this issue: in what medical specialities and conditions are trials predominantly published in accessible

English language journals? In what specialities are trials more difficult to locate? For example, a manual search of Chinese journals recently yielded 166 randomised trials in neurology, the majority of which were in stroke.[98]

To overcome the limitations associated with the retrospective design of the present investigation, future studies should prospectively compare the results from rapid reviews that are restricted to the English language with meta-analyses based on extensive searches without language restrictions. To what extent has the CCTR made it possible to perform searches that are both comprehensive and rapid?

Concealment of the process of treatment allocation has consistently been shown to be the domain of methodological quality that is most strongly associated with bias in clinical trials. The inclusion or exclusion of trials of low methodological quality has a substantial impact on results and conclusions from systematic reviews and meta-analyses. We believe that further methodological research into different markers of trial quality is required. To what extent could the effects of quality be explained by other factors, for example differences in the proportion of high-risk patients, in control treatments or the thoroughness of the implementation of the experimental treatment. What are the important dimensions of quality in placebo-controlled trials, trials with active control intervention and trials with 'hard' and 'soft' endpoints? Are the differences in trial quality across specialities observed in the present study real? What are the problems and mechanisms leading to erosion of trial quality as seen from the perspective of trialists, including investigators, treating clinicians, trial nurses and other research staff?

Acknowledgements

We thank Jos Kleijnen and the staff of the NHS CRD, University of York for the supply of articles from the DARE database, Mark Starr from Update Software for kindly providing raw data from the CDSR, Carol Lefebvre of the Oxford Cochrane Centre for bibliographical advice, and Guido Schwarzer and Deborah Tallon for preliminary work, which was partly drawn upon for this study. We also thank Dr Joanna Wardlaw and Dr Lesley Stewart for guidance on issues arising in specific meta-analyses. Finally, we are grateful to Tim Peters, George Davey Smith and Tom Fahey for valuable input in the early stages of this project.

The authors are also grateful to the HTA referees for their comments on the report. The views expressed in this report are those of the authors who are also responsible for any errors.

The Department of Social Medicine is the lead centre of the MRC Health Services Research Collaboration. Peter Jüni is supported by the Swiss National Science Foundation.

Contribution of authors

Matthias Egger conceived the study and wrote the original grant proposal; Peter Jüni and Jonathan Sterne made substantial contributions to the final study design. Christopher Bartlett performed electronic and manual literature searches and extracted and managed data. Peter Jüni and Franziska Holenstein assessed the methodological quality of trials included in Cochrane reviews. Jonathan Sterne, Peter Jüni and Matthias Egger performed statistical analyses. Matthias Egger wrote the first draft of the report; all authors contributed to the final text.

References

1. Egger M, Davey Smith G, O'Rourke K. Rationale, potentials and promise of systematic reviews. In: Egger M, Smith DG, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 23–42.

2. Huque MF. Experiences with meta-analysis in NDA submissions. *Proc Biopharmaceutical Section of the Am Stat Assoc* 1988;**2**:28–33.

3. Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34.

4. LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;**337**:536–42.

5. Yusuf S, Collins R, MacMahon S, Peto R. Effect of intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomised trials. *Lancet* 1988;**i**:1088–92.

6. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI). GISSI-3: effects of lisinopril and transdermal glyceryl trinitrate singly and together on 6-week mortality and ventricular function after acute myocardial infarction. *Lancet* 1994;**343**:1115–22.

7. Teo KK, Yusuf S. Role of magnesium in reducing mortality in acute myocardial infarction. A review of the evidence. *Drugs* 1993;**46**:347–59.

8. Collaborative Group. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet* 1995;**345**:669–87.

9. Stuck AE, Siu AL, Wieland GD, Adams J, Rubenstein LZ. Comprehensive geriatric assessment: a meta-analysis of controlled trials. *Lancet* 1993;**342**:1032–6.

10. Reuben DB, Borok G, Wolde-Tsadik G, Ershoff DH, Fishman LK, Ambrosini VL, *et al*. Randomized trial of comprehensive geriatric assessment in the care of hospitalized patients. *N Engl J Med* 1995;**332**:1345–50.

11. Imperiale TF, Stollenwerk Petrullis A. A meta-analysis of low-dose aspirin for the prevention of pregnancy-induced hypertensive disease. *JAMA* 1991;**266**:261–5.

12. CLASP Collaborative Group. CLASP: a randomized trial of low-dose aspirin for the prevention and treatment of pre-eclampsia among 9364 pregnant women. *Lancet* 1994;**343**:619–29.

13. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *Can Med Assoc J* 1997;**156**:1411–16.

14. Leizorovicz A, Haugh MC, Chapuis FR, Samama MM, Boissel JP. Low molecular weight heparin in prevention of perioperative thrombosis. *BMJ* 1992;**305**:913–20.

15. Nurmohamed MT, Rosendaal FR, Bueller HR, Dekker E, Hommes DW, Vandenbroucke JP, *et al.* Low-molecular-weight heparin versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet* 1992;**340**:152–6.

16. Yusuf S. Calcium antagonists in coronary heart disease and hypertension. Time for reevaluation? *Circulation* 1995;**92**:1079–82.

17. Dunnigan MG. The problem with cholesterol. No light at the end of this tunnel? *BMJ* 1993;**306**:1355–6.

18. de Koning HJ. Assessment of nationwide cancer-screening programmes. *Lancet* 2000;**355**:80–1.

19. Smith R. What is publication? A continuum. *BMJ* 1999;**318**:142.

20. Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;**263**:1385–9.

21. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994;**272**:158–62.

22. Egger M, Dickersin K, Davey Smith G. Problems and limitations in conducting systematic reviews. In: Egger M, Smith DG, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 43–68.

23. Donaldson LJ, Cresswell PA. Dissemination of the work of public health medicine trainees in peer-reviewed publications: an unfulfilled potential. *Public Health* 1996;**110**:61–3.

24. Easterbrook PJ, Berlin J, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;**337**:867–72.

25. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;**315**:640–5.

26. Dickersin K, Min YL, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;**267**:374–8.

27. Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993; Doc No 50:4967.

28. Ioannidis JPA. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;**279**:281–6.

29. Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;**9**:15–21.

30. Rosenthal R. The 'file drawer problem' and tolerance for null results. *Psychol Bull* 1979;**86**:638–41.

31. 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.

32. Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *Am Stat* 1995;**49**:108–12.

33. Moscati R, Jehle D, Ellis D, Fiorello A, Landi M. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. *Acad Emerg Med* 1994;**1**:267–71.

34. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Controlled Clin Trials* 1998;**19**:159–66.

35. Pittler MH, Abbot NC, Harkness EF, Ernst E. Location bias in controlled clinical trials of complementary/alternative therapies. *J Clin Epidemiol* 2000;**53**:485–9.

36. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987;**6**:11–29.

37. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;**4**:1529–41.

38. Cowley AJ, Skene A, Stainer K, Hampton JR. The effect of lorcainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias. *Int J Cardiol* 1993;**40**:161–6.

39. Chalmers I. Using systematic reviews and registers of ongoing trials for scientific and ethical trial design, monitoring, and reporting. In: Egger M, Smith DG, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 429–43

40. Bardy AH. Bias in reporting clinical trials. *Br J Clin Pharmacol* 1998;**46**:147–50.

41. Grégoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol* 1995;**48**:159–63.

42. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;**309**:1286–91.

43. Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;**350**:326–9.

44. Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le Lorier J, *et al.* Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;**347**:363–6.

45. Zielinski C. New equities of information in an electronic age. *BMJ* 1995;**310**:1480–1.

46. Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. *BMJ* 1998; **316**:61–6.

47. Singh R, Singh S. Research and doctors. *Lancet* 1994;**344**:546.

48. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 87–108.

49. Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;**273**:408–12.

50. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, *et al.* Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609–13.

51. Jüni P, Tallon D, Egger M. 'Garbage in – garbage out'? Assessment of the quality of controlled trials in meta-analyses published in leading journals. 3rd Symposium on Systematic Reviews: Beyond the Basics. 2000 July 3–5; Oxford, UK.

52. Sterne JAC, Gavaghan DJ, 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.

53. Sterne JAC, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;**54**:1046–55.

54. Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases. In: Egger M, Smith DG, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 189–208.

55. Sterne JA, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;**323**:101–5.

56. Cochrane Reviewers' Handbook 4.0 [updated July 1999]. In: Clarke M, Oxman AD, editors. The Cochrane Library. The Cochrane Collaboration. Oxford: Update Software;1999.

57. NHS Centre for Reviews and Dissemination. Undertaking systematic reviews of research or effectiveness. York: Publications Office, CRD, University of York; 1996.

58. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 1995;**48**:167–71.

59. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet* 1998;**351**:47–52.

60. 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. *Lancet* 1999;**354**:1896–900.

61. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 285–312.

62. Sterne JAC, Bradburn MJ, Egger M. Meta-analysis in stata. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 347–69.

63. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in m eta-analyses? *Lancet* 2000;**356**:1228–31.

64. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, *et al.* What contributions do languages other than English make on the results of meta-analyses. *J Clin Epidemiol* 2000;**53**:964–72.

65. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;**21**:1513–24.

66. Deeks JJ, Altman DG. Effect measures for meta-analysis of trials with binary outcomes. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 313–35.

67. Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;**106**:485–8.

68. Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988; **138**:697–703.

69. Clarke M. Searching for trials for systematic reviews: what difference does it make? [Commentary] *Int J Epidemiol* 2002;**31**:123–4.

70. Kjaergard LL, Villumsen J, Gluud C. Reported methodological quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;**135**:982–9.

71. Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, *et al.* Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;**285**:2000–3.

72. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;**134**:663–94.

73. Jadad AR, Moore RA, Carrol D, Jenkinson C, Reynolds DJM, Gavaghan DJ, *et al.* Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clin Trials* 1996;**17**:1–12.

74. Jüni P, Altman DG, Egger M. Assessing the quality of controlled clinical trials. *BMJ* 2001;**323**:42–6.

75. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;**31**:115–23.

76. Kleijnen J, Knipschild P, ter Riet G. Clinical trials of homoeopathy. *BMJ* 1991;**302**:316–23.

77. Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV, *et al.* Are the clinical effects of homoeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997;**350**:834–43.

78. Kleijnen J, Knipschild P. Ginkgo bilboa. *Lancet* 1992;**340**:1136–9.

79. Lefèbvre C, Clarke M. Identifying randomised trials. In: Egger M, Davey Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. London: BMJ Books; 2001. p. 69–86.

80. Egger M, Ebrahim S, Davey Smith G. Where now for meta-analysis? *Int J Epidemiol* 2002;**31**:1–5.

81. Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, *et al.* Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses. *Health Technol Assess* 1999;**3**(12).

82. Schwarzer G, Antes G, Tallon D, Egger M. Review publication bias? Matched comparative study of Cochrane and journal meta-analyses [abstract]. BMC Meeting Abstracts: 9th International Cochrane Colloquium 2001;1:pc142.

83. Greenland S. Quality scores are useless and potentially misleading. *Am J Epidemiol* 1994;**140**:300–2.

84. Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care* 1996;**12**:195–208.

85. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trial for meta-analysis. *JAMA* 1999;**282**:1054–60.

86. Klassen TP. Bias against quality scales. Electronic letter. eBMJ http://www.bmj.com/cgi/eletters/323/7303/42#EL1 (Accessed 9-7-2001).

87. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clin Trials* 1995;**16**:62–73.

88. Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *BMJ* 1996;**312**:742–4.

89. DerSimonian R, Charette LJ, McPeek B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;**306**:1332–7.

90. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999;**319**:670–4.

91. Schulz KF, Chalmers I, Grimes DA, Altman D. Assessing the quality of randomization from reports of controlled trials published on obstetrics and gynecology journals. *JAMA* 1994;**272**:125–8.

92. Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998;**317**:1181–4.

93. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, *et al.* Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;**276**:637–9.

94. Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Lancet* 2001;**357**:1191–4.

95. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, *et al.* The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;**134**:663–94.

96. Moher D, Jones A, LePage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;**285**:1992–5.

97. Egger M, Jüni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;**285**:1996–9.

98. He L, Liu M. A report of handsearching Chinese neurological journals. 5th Annual Cochrane Colloquium. Amsterdam, October 1997.

Appendix 1

Meta-analyses included in one or more of the analyses (*n* = 133)

Meta-analyses are shown alphabetically by first author, by source, and according to the analyses for which they qualified. Where an author contributed more than one meta-analysis, the meta-analyses are listed in the order in which they appear in the forest plots (see main report). The reference numbers in the second column link to the full bibliographical references in appendix 2.

Author	**Reference no.***	**Source**	**Publication analysis**	**Language analysis**	**MEDLINE analysis**	**Concealment analysis**	**Blinding analysis**
Alderson P	22	Journal	Yes				
a'Rogvi-Hansen B	331	CDSR		Yes			Yes
Antiplatelet Trialists[a]	21	Journal	Yes	Yes	Yes		
Antiplatelet Trialist[a]	20	Journal	Yes	Yes	Yes		
Antiplatelet Trialists[a]	26	Journal	Yes	Yes	Yes		
Ashenden R	102	DARE		Yes	Yes		
Asplund K	338	CDSR					Yes
Barrington K	511	CDSR	Yes		Yes	Yes	
Barrington K	513	CDSR				Yes	Yes
Bath P	442	CDSR	Yes	Yes			
Bath P	422	CDSR	Yes		Yes		
Benavente O	9	Journal		Yes			
Bernard B	101	CDSR	Yes	Yes	Yes		
Bucher H	16	Journal		Yes	Yes		
Cameron I	342	CDSR	Yes				
Candelise L	329	CDSR		Yes	Yes		Yes
Cates C	270	CDSR			Yes		Yes
Cochrane Albumin[b]	7	Journal	Yes				
Counsell C	232	CDSR	Yes	Yes	Yes	Yes	Yes
Counsell C	240	CDSR	Yes				
Croft A	386	CDSR				Yes	Yes
Crowley P	436	CDSR		Yes	Yes	Yes	
Crowley P	287	CDSR	Yes			Yes	
Crowther C	221	CDSR	Yes				
Crowther C	425	CDSR	Yes		Yes	Yes	Yes
Crowther C	224	CDSR			Yes		
Crowther C	524	CDSR				Yes	Yes
Da Costa A	12	Journal	Yes	Yes			
Daya S	281	CDSR	Yes			Yes	Yes
Deaney N	104	DARE	Yes	Yes			
Del Mar C	230	CDSR			Yes	Yes	Yes
Douglas R	520	CDSR	Yes	Yes		Yes	
Fahey T	4	Journal	Yes				
Fiore M	13	Journal	Yes				
Flicker L	429	CDSR		Yes	Yes	Yes	Yes
Fowlie P	439	CDSR	Yes				Yes

continued

Author	Reference no.*	Source	Publication analysis	Language analysis	MEDLINE analysis	Concealment analysis	Blinding analysis
Gadsby J	488	CDSR			Yes		Yes
Gillespie L	322	CDSR				Yes	
Gillespie W	521	CDSR				Yes	Yes
Glasziou P	228	CDSR	Yes			Yes	
Gøtzsche P	235	CDSR		Yes		Yes	Yes
Gøtzsche P	469	CDSR		Yes		Yes	Yes
Gourlay S	277	CDSR			Yes		
Gülmezoglu A	505	CDSR		Yes	Yes	Yes	Yes
Handoll H	291	CDSR	Yes	Yes		Yes	Yes
Hannah M	443	CDSR	Yes				Yes
Hannah M	416	CDSR	Yes				
Hannah M	415	CDSR	Yes				
Hannah M	414	CDSR	Yes				
Hajek P	245	CDSR			Yes		
Harrington R	6	Journal	Yes		Yes		
Hawton K	10	Journal			Yes		
Hodnett E	476	CDSR		Yes		Yes	
Hodnett E	477	CDSR			Yes	Yes	
Hodnett E	343	CDSR			Yes		
Hodnett E	210	CDSR				Yes	
Hofmeyr G	211	CDSR		Yes			
Hofmeyr G	315	CDSR		Yes			
Hofmeyr G	390	CDSR	Yes			Yes	
Hofmeyr G	212	CDSR	Yes				
Hofmeyr G	317	CDSR				Yes	Yes
Hughes E	276	CDSR				Yes	Yes
Hughes E	321	CDSR				Yes	Yes
Hughes E	412	CDSR				Yes	
Jewel D	394	CDSR			Yes	Yes	
Johanson R	516	CDSR			Yes		
Kettle C	481	CDSR		Yes		Yes	
Koch A	95	DARE	Yes	Yes	Yes		
Kramer M	246	CDSR			Yes		
Laine L	15	Journal	Yes				
Liberati A	226	CDSR	Yes	Yes	Yes		Yes
Linde K	14	Journal	Yes	Yes	Yes		
Liu M	233	CDSR			Yes		Yes
Liver Infusion[c]	103	DARE	Yes				
Macleod A	106	HTA			Yes		
McQuay H	30	HTA		Yes	Yes		
Mahomed K	360	CDSR		Yes	Yes		
Mahomed K	459	CDSR		Yes	Yes		Yes
Mari J	323	CDSR			Yes		
Marshall M	267	CDSR			Yes		
McDonald S	482	CDSR			Yes	Yes	
McIntosh H	501	CDSR		Yes	Yes		Yes
Moore R	3	Journal	Yes	Yes	Yes		
Mulrow C	237	CDSR					Yes
Neilson J	299	CDSR	Yes				
NSCLCCG[d]	18	Journal	Yes				

continued

Author	Reference no.*	Source	Publication analysis	Language analysis	MEDLINE analysis	Concealment analysis	Blinding analysis
Olliaro P	216	CDSR	Yes	Yes	Yes		Yes
Parker M	328	CDSR	Yes	Yes	Yes		
Pignataro O	100	DARE		Yes	Yes		
Plotnick L	11	Journal	Yes				
PORT[e]	2	Journal	Yes	Yes	Yes		
Poynard T	105	DARE		Yes			
Qizilbash N	286	CDSR				Yes	
Quinn K	220	CDSR	Yes				
Randolph A	5	Journal			Yes		
Roberts I	17	Journal			Yes		
Rowe B	472	CDSR					Yes
Saconato H	504	CDSR		Yes	Yes		Yes
Sarcoma[f]	24	Journal	Yes				
Schierhout G	8	CDSR			Yes		
Schierhout G	225	Journal		Yes	Yes		Yes
Schierhout G	279	CDSR				Yes	
Scott J	351	CDSR	Yes		Yes		Yes
Silagy C	427	CDSR		Yes	Yes		
Silagy C	27	Journal	Yes	Yes	Yes		
Silagy C	395	CDSR					Yes
Simons M	96	DARE	Yes		Yes		
Siragusa S	98	DARE		Yes			
Smaill F	227	CDSR			Yes		Yes
Soares K	491	CDSR			Yes	Yes	
Soll R	393	CDSR	Yes				Yes
Soll R	441	CDSR	Yes				Yes
Soll R	478	CDSR					Yes
Song F	94	DARE		Yes	Yes		
Song F	97	DARE	Yes				
Squires N	503	CDSR				Yes	
Stroke Trialists[g]	475	CDSR			Yes		
Suarez-Almazor M	453	CDSR		Yes			
Sutherland L	508	CDSR	Yes		Yes		
Sutherland L	509	CDSR	Yes				
Tharyan F	303	CDSR			Yes		Yes
Thornley B	273	CDSR	Yes		Yes		Yes
Tyson J	389	CDSR	Yes		Yes		
Vandekerckhove P	401	CDSR		Yes		Yes	Yes
Vandekerckhove P	402	CDSR	Yes			Yes	Yes
Vanderkerckhove P	404	CDSR	Yes			Yes	Yes
Wahlbeck K	278	CDSR	Yes	Yes	Yes	Yes	Yes
Wardlaw J	23	Journal	Yes	Yes	Yes		
Wardlaw J	498	CDSR				Yes	Yes
White A	206	CDSR		Yes	Yes		
Wilt T	1	Journal		Yes	Yes		
Zaat J	500	CDSR		Yes	Yes		
Zoritch B	339	CDSR	Yes				

[*] See appendix 2

[a] Antiplatelet Trialists' Collaboration
[b] Cochrane Injuries Group Albumin Reviewers
[c] Liver Infusion Meta-analysis Group
[d] Non-Small Cell Lung Cancer Collaborative Group
[e] PORT Meta-analysis Trialists' Group
[f] Sarcoma Meta-analysis Collaboration
[g] Stroke Unit Trialists' Collaboration

Appendix 2

Bibliographic references for meta-analyses included in analyses (n = 133)

(NB. Reference numbers relate to table in appendix 1 and not the main report.)

1. Wilt T. Saw palmetto extracts for treatment of benign prostatic hyperplasia. A systematic review. *JAMA* 1998;**280**:1604–9.
2. PORT Meta-analysis Trialists Group. Postoperative radiotherapy on non-small-cell lung cancer: systematic review and meta-analysis of individual patient data from nine RCTs. *Lancet* 1998;**352**:257–63.
3. Moore R, Tramèr M, Carroll D, Wiffen P, McQuay H. Quantitative systematic review of topically applied non-steroidal anti-inflammatory drugs. *BMJ* 1998;**316**:333–8.
4. Fahey T, Stocks N, Thomas T. Quantitative systematic review of RCT comparing antibiotic with placebo for acute cough in adults. *BMJ* 1998;**316**:906–10.
5. Randolph A, Cook D, Gonzales C, Andrew M. Benefit of heparin in peripheral venous and arterial catheters: systematic review and meta-analysis of randomised controlled trials. *BMJ* 1998;**316**:969–75.
6. Harrington R, Whittaker, Shoebridge P, Campbell F. Systematic review of efficacy of cognitive behaviour therapies in childhood and adolescent depressive disorder. *BMJ* 1998;**316**:1559–63.
7. Cochrane Injuries Group Albumin Reviewers, Human albumin administration in critically ill patients: systematic review of randomised controlled trials. *BMJ* 1998;**317**:235–40.
8. Schierhout G, Roberts I. Fluid resuscitation with colloid or crystalloid solutions in critically ill patients. *BMJ* 1998;**316**:961–4.
9. Benavente O. Carotid endarterectomy for asymptomatic carotid stenosis: a meta-analysis. *BMJ* 1998;**317**:1477–80.
10. Hawton K, Arensman E, Townsend E, Bremner S, Feldman E, Goldney R, *et al.* Deliberate self harm: systematic review of efficacy of psycho-social and pharmacological treatments in preventing repetition. *BMJ* 1998;**317**:441–7.
11. Plotnick L, Ducharme F. Should inhaled anticholinergics be added to Beta2 agonists for treating acute childhood and adolescent asthma? A systematic review. *BMJ* 1998;**317**:971–7.
12. Da Costa A. Antibiotic prophylaxis for permanent pacemaker implantation. A meta-analysis. *Circulation* 1998;**97**:1796–801.
13. Fiore M, Smith S, Jorenby D, Baker T. The effectiveness of the nicotine patch for smoking cessation: a meta-analysis. *JAMA* 1994;**271**:1940–7.
14. Linde K. St John's wort for depression – an overview and meta-analysis of RCTs. *BMJ* 1996;**313**:253–8.
15. Laine L, Cook D. Endoscopic ligation compared with sclerotherapy for treatment of esophageal variceal bleeding: a meta-analysis. *Ann Int Med* 1995;**123**:280–7.
16. Bucher H, Guyatt G, Cook R, Hatala R, Cook D, Lang J, *et al.* Effect of calcium supplementation on pregnancy-induced hypertension and pre-eclampsia: a meta-analysis of RCTs. *JAMA* 1996;**275**:1113–17.
17. Roberts I, Kramer M, Suissa S. Does home visiting prevent childhood injury? A systematic review of RCTs. *BMJ* 1996;**312**:29–33.
18. NSCLCCG. Chemotherapy in non small cell lung cancer: a meta-analysis using updated data on individual patients from 52 RCTs. *BMJ* 1995;**311**:899–909.

20. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy – I. Prevention of death, myocardial infarction and stroke by prolonged antiplatelet therapy in various categories of patients. *BMJ* 1994;**308**:81–106.
21. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy – II. maintenance of vascular graft or arterial patency by antiplatelet therapy. *BMJ* 1994;**308**:159–68.
22. Alderson P, Roberts I. Corticosteroids in acute traumatic brain injury: systematic review of RCTs. *BMJ* 1997;**314**:1855–9.
23. Wardlaw J, Warlow C, Counsell C. Systematic review of evidence on thrombolytic therapy for acute ischaemic stroke. *Lancet* 1997;**350**:607–14.

24. Sarcoma Meta-analysis Collaboration. Adjuvant chemotherapy for localised resectable soft tissue sarcoma of adults: meta-analysis of individual data. *Lancet* 1997;**350**:1647–54.

26. Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy – III: reduction in venous thrombosis and pulmonary embolism by antiplatelet prophylaxis among surgical and medical patients. *BMJ* 1994;**308**:235–46.

27. Silagy C, Mant D, Fowler G, Lodge M, Meta-analysis on efficacy of nicotine replacement therapies in smoking cessation. *Lancet* 1994;**343**:139–42.

30. McQuay H, Moore R, Eccleston C, Morley S, de C Williams A. Postoperative analgesia and vomiting with special reference to daycase surgery: a systematic review. *Health Technol Assess* 1997;**1**(6).

94. Song F, Glenny A. Antimicrobial prophylaxis in colorectal surgery: a systematic review of randomized controlled trials. *Br J Surg* 1998;**85**:1232–41.

95. Koch A, Bouges S, Ziegler S, Dinkel H, Daures J, Victor N. Low molecular weight heparin and unfractionated heparin in thrombosis prophylaxis after major surgical intervention: update of previous meta-analyses. *Br J Surg* 1997;**84**:750–9.

96. Simons M, Kleijnen J, van Geldere, Hoitsma H, Obertop H. Role of the Shouldice technique in inguinal hernia repair: a systematic review of controlled trials and a meta-analysis. *Br J Surg* 1996;**83**:734–8.

97. Song F. Risperidone in the treatment of schizophrenia: a meta-analysis of randomized controlled trials. *J Psychopharmacol* 1997;**11**:65–71.

98. Siragusa S, Cosmi B, Piovella F, Hirsh J, Ginsberg J. Low molecular weight heparin and unfractionated heparin in the treatment of patients with acute venous thrombo-embolism: results of a meta-analysis. *Am J Med* 1996;**100**:269–77.

100. Pignataro O, Pignataro L, Gallus G, Calori G, Cordaro C. Otitis media with effusion and S-carboxymethylcysteine and/or its lyseine salt: a critical overview. *Int J Pediatr Otorhinolaryngol* 1996;**35**:231–41.

101. Bernard B. Beta-adrenergic antagonists in the prevention of gastro-intestinal rebleeding in patients with cirrhosis: a meta-analysis. *Hepatology* 1997;**25**:63–70.

102. Ashenden R, Silagy C, Lodge M, Fowler G. A meta-analysis of the effectiveness of acupuncture in smoking cessation. *Drug Alcohol Rev* 1978;**16**:33–40.

103. Liver Infusion Meta-analysis Group. Portal vein chemotherapy for colorectal cancer: a meta-analysis of 4,000 patients in 10 studies. *J Nat Cancer Inst* 1997;**89**:497–505.

104. Deaney N, Tate H. A meta-analysis of clinical studies of imipenem-cilastatin for empirically treating febrile neutropenic patients. *J Antimicrob Chemother* 1996;**37**:975–86.

105. Poynard T, Leroy V, Cohard M, Thevenot T, Mathurin P, Opolan P, *et al.* Meta-analysis of interferon randomized trials in the treatment of viral hepatitis C: effects of dose and duration. *Hepatology* 1996;**24**:778–89.

106. Macleod A, Grant A, Donaldson C, Khan I, Campbell M, Daly C, *et al.* Effectiveness and efficiency of methods of dialysis therapy for end-stage renal disease: systematic reviews. *Health Technol Assess* 1998;**2**(5).

206. White A, Rampes H. Acupuncture in smoking cessation (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

210. Hodnett E. Alternative versus conventional delivery settings (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

211. Hofmeyr G. Amnioinfusion for cord compression (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

212. Hofmeyr G. Amnioinfusion for meconium liquor in labour (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

216. Olliaro P, Mussano P. Amodiaquine treatment in malaria (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

220. Quinn K, Parker P, de Bie R, Rowe B, Handoll H. Ankle ligament injuries – prevention (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

221. Crowther C, Alfirevic Z, Haslam R. Antenatal TRH prior to preterm delivery (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

224. Crowther C, Middleton P. Anti Rh-D prophylaxis post-partum (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

225. Schierhout G, Roberts I. Anti-epileptics following brain injury (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

226. Liberati A. Antibiotic prophylaxis in ICU (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

227. Smaill F. Antibiotics for asymptomatic bacteriuria in pregnancy (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

228. Glasziou P, Hayem M, Del Mar C. Antibiotic versus placebo for acute otitis media (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

230. Del Mar C, Glasziou P. Antibiotics for sore throat (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

232. Counsell C, Sandercock P. Anticoagulants in acute stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

233. Liu M, Counsell C, Sandercock P. Anticoagulation following non-embolic stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

235. Gøtzsche P, Johansen H. Antifungal therapy in cancer patients (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

237. Mulrow C, Lau J, Cornell J, Brand M. Antihypertensive drug therapy in the elderly (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

240. Counsell C, Sandercock P. Antiplatelet therapy in acute stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

245. Hajek P, Stead L. Aversive smoking for smoking cessation (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

246. Kramer M. Balanced protein/energy supplementation (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

267. Marshall M. Case management for severe mental disorders (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

270. Cates C. Chambers/nebulisers – acute asthma (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

273. Thornley B, Adams C, Awad G. Chlorpromazine versus placebo for schizophrenia (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

276. Hughes E, Collins J, Vandekerckhove P. Clomiphene, unexplained subfertility (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

277. Gourlay S, Stead L, Benowitz N. Clonidine for smoking cessation (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

278. Wahlbeck K, Cheine M, Essali M, Rezk E. Clozapine for schizophrenia (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

279. Schierhout G, Roberts I. Colloids vs crystalloids in fluid resuscitation (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

281. Daya S. Comparison of FSH and HMG in IVF (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

286. Qizilbash N, Lewington S, Lopez-Arrieta J. Corticosteroids in acute ischaemic stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

287. Crowley P. Corticosteroids prior to pre-term delivery (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

291. Handoll H, Farrar M, McBirnie J, Tytherleigh-Strong G, Awal K, Milne A. DVT prevention heparin plus physical for hip fracture (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

299. Neilson J, Alfirevic Z. Doppler in high risk pregnancies (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

303. Tharyan F. ECT for schizophrenia (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

315. Hofmeyr G. External cephalic version at term (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

317. Hofmeyr G. External cephalic version facilitation at term (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

321. Hughes E, Collins J, Vandekerckhove P. FSH vs hMG for PCOS (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

322. Gillespie L, Gillespie W, Cumming R, Lamb S, Rowe B. Fall prevention in the elderly (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

323. Mari J. Family Intervention for schizophrenia (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

328. Parker M, Handoll H, Robinson C. Gamma nail versus sliding hip screw for hip fractures (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

329. Candelise L, Ciccone A. Gangliosides in acute stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

331. a'Rogvi-Hansen B, Boysen G. Glycerol for acute stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

338. Asplund K, Israelsson K, Schampi I. Haemodilution in acute stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

339. Zoritch B, Roberts I, The health and welfare effects of pre-school daycare (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

342. Cameron I, Finnegan T, Langhorne P, Handoll H. Hip fracture: inpatient rehabilitation (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

343. Hodnett E, Roberts I. Home-based maternal support (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

351. Scott J. Immunotherapy for recurrent miscarriage (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

360. Mahomed K. Iron supplementation in pregnancy (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

386. Croft A. Mefloquine malaria prohylaxis (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

389. Tyson J, Kennedy K. Minimal enteral nutrition (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

390. Hofmeyr G. Misoprostol vaginally for labour induction (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

393. Soll R. Natural versus synthetic surfactant (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

394. Jewel D. Nausea in early pregnancy (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

395. Silagy C, Mant D, Fowler G, Lancaster T. Nicotine replacement therapy for smoking cessation (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

401. Vandekerckhove P, Lilford R, Vail A, Hughes E. Oligospermia: treatment with androgens (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

402. Vandekerckhove P, Lilford R, Vail A, Hughes, Oligospermia: treatment with anti-oestrogens (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

404. Vandekerckhove P, Lilford R, Vail A, Hughes E. Oligospermia: treatment with kinin enhancing drugs (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

412. Hughes E, Fedorkow D, Collins J, Vandekerckhove P. Ovulation suppression, endometriosis (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

414. Hannah M, Tan B. Oxytocin for PROM at or near term (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

415. Hannah M, Tan B. PGs versus oxytocin for PROM (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

416. Hannah M, Tan B. PGs versus oxytocin for PROM at term (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

422. Bath P, Bath F, Asplund K. Pentoxifylline in acute stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

425. Crowther C, Henderson-Smart D. Phenobarbital prior to preterm birth (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

427. Silagy C, Ketteridge S. Physician advice for smoking cessation (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

429. Flicker L, Grimley Evans J. Piracetam in dementia (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

436. Crowley P. Prevention/optimising outcome in post-term pregnancy (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

439. Fowlie P. Prophylactic indomethacin (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

441. Soll R. Prophylactic natural surfactant (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

442. Bath P, Bath F. Prostacyclin in acute stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

443. Hannah M, Tan B. Prostaglandins for PROM at/near term (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

453. Suarez-Almazor M, Belseck E, Shea B, Wells G, Tugwell P. Rheumatoid Arthritis: Methotrexate versus Placebo (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

459. Mahomed K. Routine folate supplementation in pregnancy (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

469. Gøtzsche P. Somatostatin/octreotide in acute bleeding varices (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

472. Rowe BH, Spooner CH, Ducharme FM, Bretzlaff JA, Bota GW. Sterioids and asthma relapse (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

475. Stroke Unit Trialists' Collaboration, Stroke Units (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

476. Hodnett E. Support during at-risk pregnancy (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

477. Hodnett E. Support during childbirth (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

478. Soll RF, Morley, CJ. Surfactant: Prophylaxis vs treatment (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

481. Kettle C, Johanson R. Suturing materials for perineal repair (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

482. McDonald S. Syntometrine versus oxytocin (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

488. Gadsby J, Flowerdew M. TENS/ALTENS effectiveness in chronic low back pain (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

491. Soares K, McGrath J, Deeks J. Tardive dyskinesia: GABA agonist drugs (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

498. Wardlaw JM, Yamaguchi T, del Zoppo G. Thrombolysis vs control in acute ischaemic stroke (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

500. Zaat J. Treating giardiasis (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

501. McIntosh H. Treating malaria with CQ/AQ and SP (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

503. Squires N. Treating schistosomiasis haematobium (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

504. Saconato H, Atallah A, Treating schistosomiasis mansoni (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

505. Gülmezoglu A. Treating trichomoniasis in pregnancy (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

508. Sutherland L, Roth D, Beck P, May G, Makiyama K. Ulcerative colitis: maintenance remission 5-ASA (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

509. Sutherland L, Roth D, Beck P, May G, Makiyama K. Ulcerative colitis: induction remission 5-ASA (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

511. Barrington K. Umbilical artery catheters: catheter position (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

513. Barrington K, Umbilical artery catheters: heparin usage (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

516. Johanson R, Vacuum extraction versus forceps delivery (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

520. Douglas R, Chalker E, Treacy B. Vitamin C for the common cold (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

521. Gillespie W, Henry D, O'Connell D, Robertson J, Vitamin D in fracture prevention (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

524. Crowther C, Henderson-Smart D, Vitamin K prior to pre-term birth (Cochrane Reviews), In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

Appendix 3

Reviews meeting inclusion criteria, but lacking trials with characteristics of interest (*n* = 26)

(NB. Reference numbers do not relate to the main report.)

19. Hazell P, O'Connell D, Heathcote D, Robertson J, Henry D. Efficacy of tricyclic drugs in treating child and adolescent depression: a meta-analysis. *BMJ* 1995;**310**:897–901.

28. Johnson PWM, Simnett SJ, Sweetenham JW, Morgan GJ, Stewart LA. Bone marrow and peripheral blood stem cell transplantation for malignancy. *Health Technol Assess* 1998;**2**(8).

29. MacLeod A, Grant A, Donaldson C, Khan I, Campbell M, Daly C, *et al.* Effectiveness and efficiency of methods of dialysis therapy for end stage renal disease: systematic reviews. *Health Technol Assess* 1998;**2**(5).

99. Rinaldi M, Bardelli F, Rampazzo R, Lusuriello, Messori A. Effectiveness of immunoglobulins for the prevention of systemic infections. A meta-analysis of 8 clinical studies in premature infants. *Clin Drug Invest* 1995;**10**:328–36.

207. Ahonen J, Cheine M, Wahlbeck K. Adding beta-blockers to standard medication. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

215. Fraser WD, Krauss I, Brisson-Carrol G, Thornton J, Breart G. Amniotomy to shorten spontaneous labour. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

229. Becker L, Glazier R, McIsaac W, Smucny J. Antibiotics for Acute Bronchitis. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

231. King J, Flenady V. Antibiotics in preterm labour (intact membranes). (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

238. Fuller J, Stevens LK, Chaturvedi N, Holloway JF. Antihypertensive therapy in diabetes mellitus. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

269. Martin-Hirsch P, Jarvis G, Kitchener H, Lilford R. Cervical smear collection devices. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

284. Thacker SB, Stroup DF. Continuous Electronic Fetal Monitoring. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

289. Sandborn W, Sutherland L, Pearson D, May G, Schoenfeld P, Modigliani R, *et al.* Crohn's: induction remission, AZA or 6-MP. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

290. Pearson DC, May GR, Fick G, Sutherland LR. Crohn's: maintenance, azathioprine. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

311. Carroli G, Belizan J, Stamp G. Episiotomy policies in vaginal births. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

348. Crowther C. Hospitalisation for bed rest in multiple pregnancy. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

375. Graves P. Malaria vaccines. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

392. Davis PG, Henderson-Smart DJ. Nasal CPAP after extubation in preterm infants. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

418. Counsell C, Salinas R, Warlow C, Naylor R. Patch vs no patch in carotid endarterectomy. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

435. Wilkinson D. Preventing TB in HIV. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

452. Ortiz Z, Shea B, Suarez Almazor M, Moher D, Wells G, Tugwell P. Rheumatoid Arthritis: Folic Acid and Folinic Acid. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

458. Neilson JP. Routine early pregnancy ultrasound. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

474. Boulvain M, Irion O. Stripping/sweeping of the membranes. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

499. Lancaster T, Silagy C, Fowler G, Spiers I. Training health professionals in smoking cessation. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

506. Villar J, Lydon-Rochelle MT, Gülmezoglu AM. Treatment duration for bacteriuria in pregnancy. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

525. Mahomed K. Zinc supplementation in pregnancy. (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

600. Gross A, Aker P, Goldsmith C, Peloso P. Physical medicine modalities for neck pain (Cochrane Review). In: The Cochrane Library. Issue 1. Oxford: Update Software; 1998.

Methodology Group

Current and past membership details of all HTA 'committees' are available from the HTA website (see inside front cover for details)

continued

HTA Commissioning Board

Current and past membership details of all HTA ‘committees’ are available from the HTA website (see inside front cover for details)

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278