The inclusion of reports of randomised trials published in languages other than English in systematic reviews

D Moher
B Pham
ML Lawson
TP Klassen
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.ncchta.org). 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, York Publishing Services.

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, York Publishing Services by:
– 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:
York Publishing Services Email: ncchta@yps-publishing.co.uk
PO Box 642 Tel: 0870 1616662
YORK YO31 7WX Fax: 0870 1616663
UK Fax from outside the UK: +44 1904 430868

NHS and 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 contact York Publishing Services at the address above. 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 York Publishing Distribution 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.ncchta.org/htacd.htm). Or contact York Publishing Services (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.
The inclusion of reports of randomised trials published in languages other than English in systematic reviews

D Moher1–3*
B Pham3,4
ML Lawson3
TP Klassen5

1 Department of Pediatrics, Faculty of Medicine, University of Ottawa, Canada
2 Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Canada
3 Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Canada
4 BioMedical Data Sciences, GlaxoSmithKline, Toronto, Canada
5 Department of Pediatrics, University of Alberta, Edmonton, Canada

* Corresponding author

Declared competing interests of authors: none

Published December 2003

This report should be referenced as follows:

Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technol Assess* 2003;7(41).

*Health Technology Assessment* is indexed in *Index Medicus/MEDLINE* and *Excerpta Medical/EMBASE*. 
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.

The research reported in this monograph was identified as a priority by the HTA Programme’s Methodology Panel and was funded as project number 96/52/99.

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.

Methodology Programme Director: Professor Richard Lilford
HTA Programme Director: Professor Kent Woods
Series Editors: Professor Andrew Stevens, Dr Ken Stein, Professor John Gabbay,
Dr Ruairidh Milne and Dr Rob Riemsma
Managing Editors: Sally Bailey and Sarah Llewellyn Lloyd

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.
Abstract

The inclusion of reports of randomised trials published in languages other than English in systematic reviews

D Moher,1–3* B Pham,3,4 ML Lawson3 and TP Klassen5

1 Department of Pediatrics, Faculty of Medicine, University of Ottawa, Canada
2 Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Canada
3 Chalmers Research Group, Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Canada
4 BioMedical Data Sciences, GlaxoSmithKline, Toronto, Canada
5 Department of Pediatrics, University of Alberta, Edmonton, Canada
* Corresponding author

Objectives: To assemble a large dataset of language restricted and language inclusive systematic reviews, including both conventional medicinal (CM) and complementary and alternative medicine (CAM) interventions. To then assess the quality of these reports by considering and comparing different types of systematic reviews and their associated RCTs; CM and CAM interventions; the effect of language restrictions compared with language inclusions, and whether these results are influenced by other issues, including statistical heterogeneity and publication bias, in the systematic review process.

Data sources: MEDLINE, EMBASE, the Cochrane Database of Systematic Reviews and the Centralised Information Service for Complementary Medicine.

Review methods: Three types of systematic reviews were included: language restricted; language inclusive/English language (EL) reviews that searched RCTs in languages other than English (LOE) but did not find any and, hence, could not include any, in the quantitative data synthesis; and systematic reviews that searched for RCTs in LOE and included them in the quantitative data synthesis. Fisher’s exact test was applied to compare the three different types of systematic reviews with respect to their reporting characteristics and the systematic review quality assessment tool. The odds ratio of LOE trials versus EL trials was computed for each review and this information was pooled across the reviews to examine the influence that language of publication and type of intervention (CM, CAM) have on the estimates of intervention effect. Several sensitivity analyses were performed.

Results: The LOE RCTs were predominantly in French and German. Language inclusive/LOE systematic reviews were of the highest quality compared with the other types of reviews. The CAM reviews were of higher quality compared with the CM reviews. There were only minor differences in the quality of reports of EL RCTs compared with the eight other languages considered. However, there are inconsistent differences in the quality of LOE reports depending on the intervention type. The results, and those reported previously, suggest that excluding reports of RCTs in LOE from the analytical part of a systematic review is reasonable. Because the present research and previous efforts have not included every type of CM RCT and the resulting possibility of the uncertainty as to when bias will be present by excluding LOE, it is always prudent to perform a comprehensive search for all evidence. This result only applies to reviews investigating the benefits of CM interventions. This does not imply that systematic reviewers should neglect reports in LOE. We recommend that systematic reviewers search for reports regardless of the language. There may be merit in including them in some aspects of the review process although this decision is likely to depend on several factors, including fiscal and other resources being available. Language restrictions significantly shift the estimates of an intervention’s effectiveness when the intervention is CAM. Here, excluding trials reported in LOE, compared with their inclusion, resulted in a reduced intervention effect. The present results do not appear to be influenced by statistical heterogeneity and publication bias.
Conclusions: With the exception of CAM systematic reviews, the quality of recently published systematic reviews is less than optimal. Language inclusive/LOE systematic reviews appear to be a marker for a better quality systematic review. Language restrictions do not appear to bias the estimates of a conventional intervention's effectiveness. However, there is substantial bias in the results of a CAM systematic review if LOE reports are excluded from it.
Contents

List of abbreviations ........................................ vii
Executive summary .......................................... ix

1 Introduction: assessing the need to evaluate systematic reviews with language publication restrictions ........................................ 1
Assessing the quality of reporting of systematic reviews ................................. 1
Controlling systematic error ................................ 2
Summary ..................................................... 4
Aim and objectives ......................................... 4

2 The inclusion of non-English language trials in systematic reviews – methodology ............................................. 7
Systematic review eligibility criteria .................. 7
Literature search strategy ................................. 7
Quality assessment strategy ............................. 10
Data extraction strategy .................................. 11
Data analysis ............................................... 11

3 Results .................................................... 15
General characteristics of the included systematic reviews ......................... 15
Quality of reporting of systematic reviews ........................................... 19
Quality of reporting of RCTs ............................. 27
‘Language of publication’ bias and location bias are related to the type of intervention (CM or CAM) ........................................... 30

4 Discussion ............................................ 49
Quality of reporting of systematic reviews ........................................... 50
Quality of reporting of RCTs ............................. 51
‘Language of publication’ bias and location bias are related to the type of intervention (CM or CAM) ........................................... 52
Conclusions ............................................... 56

Acknowledgements ....................................... 61
References ............................................... 63

Appendix 1 Data abstraction form ............. 67
Appendix 2 Jadad and allocation concealment data collection form .......... 71

Appendix 3 Listing and citations of systematic reviews included in research ........................................... 73
Health Technology Assessment reports published to date .......................... 79
Health Technology Assessment Programme ........................................... 87

The impact of language restriction on between-study heterogeneity and publication bias ........................................... 38
**List of abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCG</td>
<td>Bacillus Calmette–Guérin</td>
</tr>
<tr>
<td>CAM</td>
<td>complementary and alternative medicine</td>
</tr>
<tr>
<td>CDSR</td>
<td>Cochrane Database of Systematic Reviews</td>
</tr>
<tr>
<td>CI</td>
<td>confidence interval</td>
</tr>
<tr>
<td>CISCOM</td>
<td>Centralised Information Service for Complementary Medicine</td>
</tr>
<tr>
<td>CM</td>
<td>conventional medicine</td>
</tr>
<tr>
<td>CONSORT</td>
<td>consolidated standards of reporting randomised trials</td>
</tr>
<tr>
<td>CRG</td>
<td>Chalmers Research Group</td>
</tr>
<tr>
<td>DARE</td>
<td>Database of Abstracts of Reviews of Effectiveness</td>
</tr>
<tr>
<td>EL</td>
<td>English language</td>
</tr>
<tr>
<td>FE</td>
<td>fixed effects</td>
</tr>
<tr>
<td>ICC</td>
<td>intra-class correlation coefficient</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Diseases</td>
</tr>
<tr>
<td>IQR</td>
<td>interquartile range</td>
</tr>
<tr>
<td>LOE</td>
<td>languages other than English</td>
</tr>
<tr>
<td>MeSH</td>
<td>medical subject heading</td>
</tr>
<tr>
<td>OG</td>
<td>Oxman and Guyatt</td>
</tr>
<tr>
<td>OR</td>
<td>odds ratio</td>
</tr>
<tr>
<td>QUOROM</td>
<td>quality of reporting of meta-analysis</td>
</tr>
<tr>
<td>RCT</td>
<td>randomised controlled (clinical) trial</td>
</tr>
<tr>
<td>ROR</td>
<td>ratio of odds ratios of intervention</td>
</tr>
<tr>
<td>RE</td>
<td>random effects</td>
</tr>
<tr>
<td>TB</td>
<td>tuberculosis</td>
</tr>
</tbody>
</table>

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**

In an era of evidence-based healthcare, systematic reviews are becoming increasingly important as a source of evidence for decision-making. They afford the reader an opportunity to review quickly the totality of evidence regarding a particular intervention. Ideally, the systematic review process of randomised controlled trials (RCTs) provides the reader with a bias-free estimate of the effects of the intervention under consideration.

There is now evidence regarding the influence that several factors in the review process have on the results of a systematic review. For example, excluding unpublished studies, compared with their inclusion, can exaggerate the estimates of an intervention’s effectiveness by 15%, on average.

The role of including reports of RCTs reported in languages other than English (LOE) (i.e. language restriction) remains uncertain. Such studies are difficult to identify and retrieve. The costs of including these studies can be prohibitive for the average reviewer. Yet excluding them from the systematic review process might introduce substantial bias, make the review process flawed and exaggerate the results of the review. Confounding the decision to exclude these studies is whether they are investigating a conventional medicinal (CM) intervention, such as methylphenidate for attention deficit hyperactivity disorder, or a complementary and alternative medicine (CAM) intervention, such as hypnosis for treating migraines. Traditionally, CAM has been investigated in countries whose first language is not English.

**Objectives**

We set out to assemble a large dataset of language restricted and language inclusive systematic reviews, including both CM and CAM interventions. We also assessed the quality of these different types of systematic reviews and their associated RCTs and compared the quality of systematic reviews investigating a CM intervention with those reviews examining CAM interventions. We also examined whether language restrictions compared with language inclusions exaggerate the estimates of an intervention’s effectiveness. Finally we evaluated whether language restrictions of conventional interventions are similar to those for CAM interventions, and whether these results are influenced by other issues, including statistical heterogeneity and publication bias, in the systematic review process.

**Methods**

**Data sources**

A systematic review was included if the primary data sources were reports of RCTs identified through a collection of systematic reviews assembled by the Chalmers Research Group. This collection was based on searching MEDLINE, EMBASE and the Cochrane Database of Systematic Reviews. We added to this collection for the present work by searching EMBASE and the Centralised Information Service for Complementary Medicine.

**Inclusion criteria**

The methodology section of the systematic review had to state explicitly whether the search was limited to identifying and including English RCTs only, or whether RCTs of all languages were eligible for inclusion. Systematic reviews were excluded if there was no mention of language restriction or inclusion.

Three types of systematic reviews were included: language restricted systematic reviews, meaning that no reports of RCTs reported in LOE were included in the quantitative data synthesis (i.e. ‘language restricted systematic reviews’); language inclusive/English language (EL) systematic reviews that searched for reports of RCTs in LOE but did not find any and, hence, could not include any, in the quantitative data synthesis; and systematic reviews that searched for reports of RCTs in LOE and included them in the quantitative data synthesis (i.e. language inclusive/LOE systematic reviews).

We estimated that 45 language restricted and 45 language inclusive systematic reviews would be required to detect a 25% difference in the ratio of...
odds ratios of intervention (ROR) between trials published in an LOE and reports of English language RCTs, on a logarithmic scale.

**Data extraction**
We assessed the quality of reports of all three types of systematic reviews and the RCTs contained in the language inclusive/LOE reviews. All the assessments were completed using state of the art assessment instruments.

Characteristics (e.g. assessment of publication bias) of each systematic review were abstracted using a standardised data collection form.

Similarly, for each included RCT, we extracted the first author’s name, journal, year of publication, language of publication and whether or not it related to CAM. For the primary outcome, we also extracted the number of events and patients in the control group and the number of events and patients in the experimental group.

**Data synthesis**
We applied Fisher’s exact test to compare the three different types of systematic reviews with respect to their reporting characteristics and the systematic review quality assessment tool. We computed the log ROR of LOE trials versus EL trials for each systematic review and pooled this information across systematic reviews to examine the influence that language of publication and type of intervention (CM, CAM) have on the estimates of intervention effect. Several sensitivity analyses were performed.

**Results**
We included 130 systematic reviews: 50 language restricted, 32 language inclusive/EL and 48 language inclusive/LOE systematic reviews. Approximately 20% of the reviews were investigating CAM. The language inclusive/LOE reviews included the largest number of RCTs and participants. The LOE RCTs were published predominantly in French and German.

Language inclusive/LOE systematic reviews were of the highest quality compared with the other types of reviews, scoring 57% of the maximum possible score. The CAM reviews were of higher quality, averaging 71% of their maximum possible score, compared with the CM reviews. There were only minor differences in the quality of reports of RCTs published in English compared with the eight other languages included in this analysis.

However, there are inconsistent differences in the quality of LOE reports depending upon the type of intervention.

The present results, and those reported previously, suggest that excluding reports of RCTs in LOE from the analytical part of a systematic review is a reasonable way to conduct a review [random effects model (RE) ROR = 1.02; 95% confidence interval (CI): = 0.83 to 1.26]. Because the present research and previous efforts have not included every type of CM RCT and the resulting possibility of the uncertainty as to when bias will be present by excluding LOE, it is always prudent to perform a comprehensive search for all evidence. This result only applies to reviews investigating the benefits of CM interventions. This does not imply that systematic reviewers should neglect reports in LOE. We recommend that systematic reviewers search for reports regardless of the language of their publication. There may be merit in including them in some aspects of the review process although this decision is likely to depend on several factors, including fiscal and other resources being available.

However, language restrictions significantly shift the estimates of an intervention’s effectiveness when the intervention is CAM. Here, excluding trials reported in LOE, compared with their inclusion, resulted in a reduced intervention effect, 63% on average (RE ROR = 1.63; 95% CI: 1.03 to 2.60).

The present results do not appear to be influenced by statistical heterogeneity and publication bias.

**Conclusions**
With the exception of CAM systematic reviews, the quality of recently published systematic reviews is less than optimal. Language inclusive/LOE systematic reviews appear to be a marker for a better quality systematic review. Language restrictions do not appear to bias the estimates of a conventional intervention’s effectiveness. However, there is substantial bias in the results of a CAM systematic review if LOE reports are excluded from it.

**Recommendations for research**
Consideration of the development of a national database of systematic reviews is likely to facilitate meta-epidemiology research undertaken in the UK and elsewhere.
The quality of reporting of systematic reviews of RCTs needs improvement. This is most likely to be achieved if authors and medical journal editors agree to a standardised and evidence-based way of reporting. The quality of reporting of meta-analysis of randomised trials (QUOROM) statement is one option to consider for systematic reviews. Likewise, the consolidated standards of reporting randomised trials (CONSORT) statement is likely to improve the quality of reporting of randomised trials.

To keep QUOROM and CONSORT up to date, regular meetings of these groups should be encouraged.

A more in-depth examination of CAM trials, particularly those conducted in Asian countries, and their influence on the conduct of systematic reviews is required.

Aspects of CAM methodology and content need to be incorporated in critical appraisal skills training programmes.
Chapter 1

Introduction: assessing the need to evaluate systematic reviews with language publication restrictions

Healthcare providers, consumers and others cannot keep up to date with the healthcare literature. For example, healthcare professionals attempting to keep abreast of their field would need to read, on average, 19 original articles each day. Systematic reviews offer the potential to reach that elusive goal of keeping up to date without sacrificing quality and thoroughness.

There has been a striking increase in the number of published systematic reviews, particularly of randomised controlled clinical trials (RCTs). One of the first ‘medical’ systematic reviews was published in the Journal of the American Medical Association in 1955. In 1999, approximately 1250 publications were identified with the term ‘meta-analysis’ and ‘systematic review’ as a medical subject heading (MeSH).

As with any exciting development, the growth of systematic reviews has also had its share of problems. A systematic review evaluating the effectiveness of intravenous magnesium in the treatment of patients with suspected myocardial infarction provided evidence for the effective use of magnesium in clinical practice. However, the results of an international study of infarct survival, ISIS-4, an RCT also examining the effect of magnesium on patients with suspected myocardial infarction, have called into question the results of the systematic review and the effectiveness of magnesium.

Contrary to the results of the systematic review, the ISIS-4 investigators found no statistically significant reduction in mortality for patients receiving magnesium. It is possible that the observed discordance between the results of the systematic review and RCT reflect, in part, how they were conducted and reported.

Assessing the quality of reporting of systematic reviews

We are assuming that the quality of reporting of a systematic review is a reasonable surrogate for how it was conducted. There are, however, sparse data examining the relationship between the conduct of a research study and its report. The evidence that does exist comes from examining reports of breast cancer RCTs and suggests only minimal differences between how a RCT is conducted and reported.

Liberati and colleagues evaluated the internal and external validity of 63 reports of RCTs using a scale with a maximum total score of 100 points. The mean score for all RCTs was 50% [95% confidence interval (CI): 46 to 54]. To elaborate on various aspects of the RCT reports, the authors conducted telephone interviews with 62 (of 63) of the corresponding authors. This resulted in a 7% average improvement in the quality scores.

Hadhazy and colleagues examined the same question and reported similar results to those of Liberati and colleagues. However, a recent study reported by Hill and colleagues questions this relationship and suggests that the written report may not be an accurate reflection of what happens during the conduct of an RCT or there is substantial self-reporting bias. The relationship between conduct and reporting will be further understood once additional research findings are forthcoming.

A survey of 86 reports of English language (EL) systematic reviews assessed each publication on 14 items from six content areas thought to be important in the conduct and reporting of a systematic review: study design, combinability, control of bias, statistical analysis, and problems of applicability. The results suggested that only 24 of the 86 reviews (28%) addressed all six content areas. A more recently updated survey reported similar results.

Evidence from the spinal manipulation, pain and primary care literature suggests considerable room for improvement in how systematic reviews are conducted and reported. Silagyi surveyed 28 review articles published in primary care journals during 1991 using eight methodological criteria thought to be important in the reporting of meta-analyses. Each criterion had a maximum score of two for a total score of 16. Only 25% of the articles
obtained a passing grade with a total score of more than eight. Jadad and McQuay\(^{13}\) reviewed 80 systematic reviews in the literature of pain using the validated Oxman and Guyatt (OG) scale,\(^{15,16}\) including a question about the overall scientific quality of a systematic review (range 1–7, with higher scores indicating superior quality). The median score was four, indicating substantial faults in scientific quality in these reviews. These authors also observed that the lower the score, the more likely the meta-analyses were to have a positive result. Assendelft and colleagues\(^{12}\) used a non-validated tool to assess the validity of 51 systematic reviews of spinal manipulation. A median score of 23 (out of a maximum of 100) was found; in contrast to Jadad and McQuay,\(^{13}\) these authors reported that reviews with ‘positive’ conclusions tended to have higher quality scores.

More recently, Shea and colleagues\(^{17}\) examined 52 systematic reviews published between 1990 and 1995 using the OG validated index, where high scores indicate minimal flaws and low scores extensive flaws. The average OG quality score was 3.35 (out of 7), indicating considerable room for improving the quality of reports of systematic reviews. In summary, these studies suggest that one way to increase the quality of systematic reviews is to improve how they are conducted and reported.

**Controlling systematic error**

**Assessing the quality of reports of RCTs published in other languages**

Systematic reviewers have little control over random errors but can exert some influence over systematic errors (bias). Including only a portion of all available evidence in a systematic review may introduce bias into the review process and threaten its validity. The most comprehensive search strategies would include all relevant literature, regardless of language of publication. However, identifying, obtaining and translating non-English studies significantly increases the time, cost and effort required by the investigators. Grégoire and colleagues\(^{18}\) reported that 78% of identified systematic reviews had language of publication restrictions. The majority (93%) of these restrictions were at the expense of excluding reports of RCTs published in languages other than English (LOE). Perhaps these language restrictions were due to difficulties in identifying reports in LOE or the presumed greater importance and higher quality of English language publications.

One way to evaluate whether language restrictions are a sensible policy for systematic reviewers to abide by is to assess the quality of reports of RCTs published in a variety of languages. Language restrictions might be appropriate if the quality of reporting in LOE differs compared with EL reports. Alternatively, if the quality of reports in LOE and English are similar, this would provide evidence for the inclusion of all trials regardless of their language of publication.

Moher and colleagues\(^{19}\) set out to address whether the quality of reporting in LOE compared with EL reports differs in some meaningful way. The authors compared the quality of reporting design characteristics and analytical approaches of 133 EL RCTs published between 1989 and 1994, with reports in LOE (French, German, Italian and Spanish) over the same time period and type of journal. The assessments were completed under masked conditions, using the Jadad quality assessment scale,\(^{20}\) with a scoring range of zero to five, in which higher scores indicate superior reporting. Each trial report was also assessed for its adequacy of allocation concealment.\(^{21}\)

The authors reported no statistically significant differences between LOE and EL reports with respect to quality of reporting, randomisation, double-blinding, dropouts and withdrawals, or overall total score, despite adequate statistical power. The differences in the quality of reporting in LOE and EL RCTs ranged from 0 to 4% for individual items and 5% in total score. Similarly, there was little difference in the adequacy of allocation concealment between languages. Moher and colleagues, however, did not attempt to address a related issue: do estimates of treatment differences vary across languages?

**Language of publication bias**

Several authors have explored the impact of excluding LOE reports on the results of systematic reviews, with varying conclusions.\(^{18,22,23}\) Grégoire and colleagues\(^{18}\) conducted a MEDLINE search for all systematic reviews published in eight general and internal medicine journals between 1 January 1991 and 1 April 1993. From this search, the investigators identified 28 systematic reviews with language restrictions. The authors repeated four language restricted systematic reviews, adding any LOE report that was originally excluded from the original language restricted data synthesis. In three of the four systematic reviews, the authors reported that the inclusion of RCTs published in all languages did not alter the estimates of the intervention’s effectiveness. However, in a systematic
review of selective decontamination of the digestive tract, the inclusion of a German article published in a Swiss journal changed the results from no statistical effect on mortality [odds ratio (OR) = 0.70; 95% CI: 0.45 to 1.09] to a statistical reduction in mortality (OR = 0.67; 95% CI: 0.47 to 0.95). Unfortunately, closer examination of this systematic review reveals at least two problems not reported by Grégoire and colleagues.

One of the EL reports included in the original systematic review was not an RCT. Brun-Buisson and colleagues report in the abstract of their study that patients participated in an ‘8-week randomised’ trial. The method used to allocate patients to their respective intervention group was described in the Patients and Methods section as ‘odd and even birth year’. Several authorities have noted that this type of alternating does not constitute randomisation. Similarly, the German report, included by Grégoire and colleagues when they repeated the systematic review data synthesis, was not an RCT but a quasi-crossover study. Patients admitted to the intensive care unit during the first 6 months of the year were assigned one therapy that was changed during the latter 6 months of the year. Therefore, the validity of the results reported by Grégoire and colleagues is questionable.

Chalmers and colleagues have also examined the issue of differing treatment effects across languages using a systematic review of intravenous streptokinase in the management of acute myocardial infarction. The investigators included 11 EL RCTs involving 3268 patients and reported a statistically beneficial effect of the intervention (OR = 0.79; 95% CI: 0.66 to 0.94). The authors compared these results with those reported by Yusuf and colleagues, who addressed the same question as Chalmers and colleagues but included 20 RCTs, without language restrictions, involving 5284 patients. Yusuf and colleagues also reported a beneficial effect of the intervention (OR = 0.72; 95% CI: 0.65 to 0.89). The 7% difference in the estimates of the intervention’s effectiveness, reported by the authors, could be explained, in part, by the differences in the language of publications eligibility used by the different investigators.

In a similar examination using beta-blockers, Chalmers and colleagues included 13 English language RCTs, involving 2548 patients, and observed no statistical benefits for beta-blockers (OR = 0.87; 95% CI: 0.75 to 1.24). Again, the investigators compared their results with those of Yusuf and colleagues, who used 21 RCTs without language restrictions, involving 3611 patients, and also reported no statistical benefit of the intervention (OR = 0.95; 95% CI: 0.75 to 1.18). These results are similar to those described in the streptokinase example above and may also be explained, in part, by differences in the inclusion/exclusion criteria of LOE reports used by the investigators.

The results reported by Chalmers and colleagues suggest that including reports in LOE in a systematic review has no effect on the estimates of an intervention’s effectiveness. However, this is based on two systematic reviews. Grégoire and colleagues have shown that including reports in LOE in a systematic review changed the estimates of effectiveness in one case. Unfortunately, this study has methodological limitations. There is a need for much stronger and more reliable evidence based on the results of a larger number of systematic reviews.

Moher and colleagues identified 18 systematic reviews that explicitly stated no language restrictions in their search to identify trials but actually included reports in LOE in their data synthesis (218 RCTs; 83 LOE reports and 128 EL reports). Logistic regression was used to compare the statistical results of each systematic review in which LOE reports were either included or excluded from the analysis. Language restricted systematic reviews, compared with language inclusive reviews, did not exaggerate the benefits of the intervention [ratio of odds ratios of interventions (ROR) = 0.98; 95% CI: 0.79 to 1.15]. These results did not change whether the data synthesis was limited to a single report or multiple reports in LOE or whether reports with small or large sample sizes were included.

This analysis is based on a relatively small number of systematic reviews and associated RCTs. Most (68%) of the systematic reviews included only one report of an RCT in LOE. This may be due to including only systematic reviews identified in MEDLINE and the Cochrane Database of Systematic Reviews (CDSR). It is possible that different sampling frames, such as EMBASE and complementary and alternative medicine (CAM) databases [i.e. Centralised Information Service for Complementary Medicine (CISCOM)], will provide different results.

**Location bias**

Other emerging evidence suggests that the language in which a trial is published may be
related to its statistical results. Egger and colleagues reported that German authors were more likely to report RCTs with statistically positive results in EL journals rather than in German language journals.

Vickers and colleagues examined whether the results of reports of RCTs from certain countries are more likely to be statistically positive. The authors reported that RCTs published in certain countries, such as China, compared with other countries, such as the UK, were more likely to report statistically positive results. Their analysis suggested that some countries, including China, Japan, Hong Kong, Russia/USSR and Taiwan, only reported RCTs with statistically positive results.

**Influence of type of intervention**

**[conventional medicine (CM) or CAM]**

A further factor (bias) to consider in assessing the quality of a systematic review is whether the studies included were investigating a CM intervention or a CAM intervention. We classified a CM intervention as one involving the use of a pharmaceutical intervention or a surgical manoeuvre. An intervention was considered CAM if the intervention dealt with biochemical (e.g. herbs), lifestyle (e.g. mind-body), biomechanical (e.g. chiropractic) or bioenergetics (e.g. acupuncture).

Traditionally, CAM interventions have been investigated in countries whose first language is not English. Hence, for a systematic reviewer evaluating CAM they would have to decide whether they should limit the RCTs to be included in the systematic review to those reported in English only or whether CAM trials published in LOE of sufficient quality should also be included.

In 1999, Tang and colleagues used stratified sampling to select randomly journals from a total of 100 Chinese journals of traditional Chinese medicine. The authors hand searched these journals from the early 1980s to identify RCTs. The 2938 RCTs identified were evaluated for methodological quality. The authors reported that although quality had improved over the years, there continue to be many methodological problems in the Chinese language reports, including inadequate description of method of randomisation, lack of blinding, small sample sizes, inappropriate controls and lack of follow-up. Furthermore, in a subset of 49 trials examining acupuncture in the treatment of stroke, there was evidence of publication bias (assessed by funnel plot). Similarly, Vickers and colleagues found that acupuncture research conducted in certain LOE was uniformly positive, suggesting publication bias. Evidence from CM studies indicates that lower quality studies are more likely to have positive results. Studies such as these suggest that the quality of CAM LOE reports is inadequate, thus justifying their exclusion from systematic reviews. However, investigations comparing the quality of CAM reports with reports of CM are limited.

**Summary**

Taken together, these results may imply that several biases need to be examined when conducting systematic reviews. For example, it is possible that language restricted systematic reviews are reasonable for reviews involving CM interventions but not for those evaluating CAM interventions. Similarly, it is possible that including reports in LOE, especially non-European languages, may introduce additional biases, such as location bias, into the systematic review process. The extent to which the trade-off between including reports in LOE and the introduction of bias influences the results of systematic reviews requires further examination.

Differences in the magnitude, direction or precision of the effectiveness of interventions, depending on whether or not reports in LOE are included in systematic reviews, are also important in terms of efficiency and the costs of conducting systematic reviews. Including reports in LOE will require more effort in identification, retrieval and translation into the language of the systematic reviewer. Before such methodology becomes recommended policy, it is essential to provide as much high-quality evidence as possible to help inform the development of such recommendations.

**Aim and objectives**

We believe that the effect of including reports in LOE on the results of systematic reviews is influenced by several factors requiring further examination. These factors include the number of reports of LOE included in a systematic review, the sample size of these reports, the prevalence of LOE within a given systematic review, the presence of publication bias, the extent of statistically heterogeneity and the type of intervention under investigation.

In addition, in our recent examination of the influence of LOE in a systematic review, we did
not assess the quality of RCTs within a systematic review. This information may be another piece of the puzzle requiring examination.

Therefore, it is the objective of this report to examine the extent to which inclusion of reports in LOE influences the results of systematic reviews. The findings will likely have profound effects on the conduct, results and reporting of all systematic reviews, which will likely have an impact on the therapeutic interventions offered to prospective patients.

To do this, we will first assemble a dataset of language inclusive systematic reviews, including both CM and CAM interventions. Using this dataset we will address the following:

1. Examine the relationship between language restrictions and the quality of systematic reviews for both CM and CAM.
2. Compare the quality of reporting of English and LOE reports of RCTs associated with the systematic reviews and to determine whether there are differences in CM versus CAM reports.
3. Examine whether language restrictions impact the estimates of an intervention’s effectiveness.
4. Evaluate whether language restrictions of CM interventions are similar to those for CAM interventions, and whether these results are influenced by other issues including statistical heterogeneity and publication bias in the systematic review process.
Chapter 2
The inclusion of non-English language trials in systematic reviews – methodology

Systematic review eligibility criteria

A systematic review was included if it was published in English, if the primary data sources were reports of RCTs and if the methodology section of the systematic review explicitly stated whether only English trials were eligible for inclusion or whether trials in other languages were considered. Systematic reviews were excluded if there was no mention of language restriction or inclusion.

Literature search strategy

Selection of electronic databases used for previous research

For the present research we searched two databases, described below (Selection of electronic databases used for present research). Here we provide details of how the Chalmers Research Group (CRG) database of systematic reviews was established. The database is based, in part, on research commissioned by the NHS’s Health Technology Assessment programme. It contains systematic reviews identified from MEDLINE, EMBASE and the CDSR. Systematic reviews from the CRG database were included in the present research.

We began by conducting a MEDLINE search (OVID) from 1 January 1966 to 31 December 1995 to identify systematic reviews. The search strategy included search terms as MeSH, text words and publication types. Abstracts retrieved by the search were reviewed by one member of the team. Determining whether the articles were, in fact, systematic reviews was difficult because the methodology section was insufficiently described in the abstract. Similarly, the citations were not indexed as ‘systematic reviews’. As a result, we had to obtain and read the full reports of the reviews. As an initial step, we randomly retrieved hard copies of 50 reviews only to ascertain fulfillment of eligibility criteria.

To be considered a systematic review, the article had to state (1) the name of database(s) searched (2) the years searched and (3) the search terms included. We found, however, that the majority of the articles failed to report this information in the methodology section. Consequently, we decided to focus on identifying meta-analyses of RCTs.

To identify meta-analysis of RCTs we completed an electronic search of MEDLINE (OVID) from 1 January 1966 to 31 December 1995. The search strategy included 21 search terms such as MeSH, text words and publication types. The MEDLINE search was translated using the appropriate terms to search EMBASE (SILVERPLATTER) from 1 January 1980 to 30 November 1995. Both search strategies aimed to identify meta-analyses published in any language.

The CDSR 1995, Issue 2, was also searched, as was the Database of Abstracts of Reviews of Effectiveness (DARE), both within the Cochrane Library. The 1995 issue of DARE did not provide complete bibliographic information for each reference and we could not retrieve hard copies of the papers. We therefore elected not to include it in our search for meta-analyses. Current versions of DARE now include appropriate sources.

Once the MEDLINE search strategy was refined, we determined its sensitivity (i.e. the number of meta-analyses identified by a search method expressed as a percentage of the total number of relevant articles identified) and precision (i.e. the number of meta-analyses identified by a search method expressed as a percentage of the total number of articles identified by the MEDLINE search strategy) as a quality control check.

Citations identified by the search strategy were compared with established bibliographic lists of meta-analyses. Systematic reviews in the ACP Journal Club were also used as a representative collection of high-quality systematic reviews. Based on the results of the quality control efforts, the search strategy was modified to maximise sensitivity and precision.

The sensitivity and precision of the EMBASE search were not determined because of the small
sample size obtained for this study. We selected few articles because of considerable difficulty in both uploading and mapping the EMBASE CD-ROM discs on to the hard disc (a ‘juke box’ with the ability to accommodate seven CD-ROMs). In addition, there was inadequate technical support and the time and cost restraints involved in retrieving all citations identified in EMBASE were considerable. Instead, we took a random sample of 100 articles that had ‘meta-analysis’ or a variant of the word in the title.

A coding system was developed (a priori) for each article for its potential inclusion in the database. Each article identified by the MEDLINE, EMBASE and CDSR search was evaluated for inclusion based on four criteria: design (did the article refer to meta-analyses?: Yes, no, probably), publication type (was the paper a meta-analysis, editorial or a methodological paper?), primary studies (did the meta-analysis include RCTs, observational or mixed studies?) and type of research question (was the article focused on treatment, diagnostic, prevention, aetiology, association, prognosis and economics). Two members from the research team independently assessed each article. Disagreement was resolved by consensus.

Selection of electronic databases used for present research

The search strategy aimed to identify systematic reviews of RCTs published in English between 1985 and 1999. Eligible reviews were identified from the collection of systematic reviews already assembled by the CRG, and through additional searching of EMBASE and CISCOM. There is an anecdotal view that the EMBASE and CISCOM databases have a greater likelihood of containing non-English trials. For instance, CAM reports may be more likely to be conducted and reported in Europe and East Asia and are less likely to be published in English. MEDLINE and the CDSR were not searched for this study as we have extensively searched them previously.30,34,35

EMBASE is produced by Elsevier and contains more than 8 million records from 1974 to the present from over 4000 journals; approximately 445 000 records are added annually. EMBASE was searched to identify systematic reviews of RCTs using a search strategy based on our published MEDLINE strategy. We translated the MEDLINE MeSH headings to the EMBASE Emtree terminology. Text words were also used to search specific words or phrases in the title and abstract fields of EMBASE citations. EMBASE has no equivalent to the publication type field in MEDLINE, so this term was dropped from the search strategy. We expected this search to retrieve a number of items similar to what would be retrieved from MEDLINE. However, 8129 citations were identified from 1988 to week 43, 1998, almost double the expected number.

After a small sampling, we found that many of the citations were not relevant. Owing to the time and cost associated with downloading references from EMBASE, we decided to modify the search strategy in an attempt to improve precision without sacrificing the sensitivity of the original search. This modification included combining ‘meta-analysis’ as a descriptor with ‘meta-analysis’ used as a text word and then combining this with the study type ‘randomised controlled trials’ as a descriptor with the word ‘random’ searched as a text word. Details of the search strategy are provided in Table 1. Using this strategy, 413 citations were identified for 1995, which were downloaded in order to assess the performance of the strategy. Since the number of citations identified was reduced to a reasonable size with good sensitivity (when compared with a MEDLINE retrieval from 1995 using the original strategy), we decided to use this modified strategy to search the remaining years.

The CISCOM database was searched for systematic reviews of RCTs published since 1985. CISCOM has approximately 60 000 articles spanning the years from 1920 to the present. This database was developed by the Research Council for Complementary Medicine and contains all types of articles on CAM published in the medical literature. The CISCOM database is a result of a systematic search of a number of databases (e.g. MEDLINE, EMBASE, CATS/AMED), hand searching the Cochrane Library, citation tracking and contact with privately held CAM databases. All papers tagged as ‘reviews’ and published in English after 1985 were retrieved and reviewed to see if they met the inclusion criteria.

All identified bibliographic records were imported into Reference Manager (Research Information Systems, Version 8.5). During the import process, a ‘duplicate search’ was conducted. This feature of the software identifies potential duplicate records based on similarity of authors, journal name and title, and gives the option of deleting the duplicate items before they are imported. A second duplicate check was conducted manually, eliminating any missed duplicates.
### TABLE 1  Search strategies to identify systematic reviews and meta-analyses included in (A) the CRG database and (B) EMBASE

<table>
<thead>
<tr>
<th>Search no.</th>
<th>Set</th>
<th>Search terms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(A) CRG database</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>meta-analysis.pt,sh.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>(meta-anal: or metaanal:).tw.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>(quantitativ: review: or quantitativ: overview:).tw.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>(systematic: review: or systematic: overview:).tw.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>(methodologic: review: or methodologic: overview:).tw.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>(integrative research review: or research integration:).tw.</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>review.pt,sh. or review:.tw. or overview:.tw.</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>quantitativ: synthes:.tw.</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>1 or 2 or 3 or 4 or 5 or 6 or 8</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>(medline or medlars).tw,sh. or embase.tw.</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>(scisearch or psychinfo or psycinfo).tw.</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>(psychlit or psyclit).tw.</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>(hand search: or manual search:).tw.</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>(electronic database: or bibliographic database:).tw.</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>(pooling or pooled analys: or mantel haenszel).tw.</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>(peto or der simonian or dersimonian or fixed effect:).tw.</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>10 or 11 or 12 or 13 or 14 or 15 or 16</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>7 and 17</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>9 or 18</td>
</tr>
<tr>
<td><strong>(B) EMBASE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>meta analysis.de.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>(quantitativ$ review$ or quantitativ$ overview$).tw.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>(systematic$ review$ or systematic$ overview$).tw.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>(methodologic$ review$ or methodologic$ overview$).tw.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>review.de. or review$.tw. or overview$.tw.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>(integrative research review$ or research integration$).tw.</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>quantitativ$ synthesis$.tw.</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>1 or 2 or 3 or 4 or 5 or 7 or 8</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>(MEDLINE or medlars or Embase).tw.</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>(scisearch or psychinfo or psychinfo).tw.</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>(hand search$ or manual search$).tw.</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>(electronic database$ or bibliographic database$).tw.</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>(pooling or pooled analys$ or mantel haenszel).tw.</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>(peto or der simonian or dersimonian or fixed effect$).tw.</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>(psychlit or psyclit).tw.</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>or/10-16</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>6 and 17</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>9 or 18</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>randomised controlled trial.de.</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>clinical trial.de.</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>controlled study.de.</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>major clinical study.de.</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>random$.hw,tw.</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>“0197”.tg.</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>“0150”.tg.</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>placebo$.hw,tw.</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>double-blind-procedure.de.</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>blind$.tw,hw.</td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>or/20-29</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>19 and 30</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>limit 31 to yr=1985-1999</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>meta analysis.de.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>(meta-anal$ or metaanal$).tw.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>(quantitativ$ review$ or quantitativ$ overview$).tw.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>(systematic$ review$ or systematic$ overview$).tw.</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>(methodologic$ review$ or methodologic$ overview$).tw.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>review.de. or review$.tw. or overview$.tw.</td>
</tr>
</tbody>
</table>

**continued**
Types of systematic reviews

Hard copies of all the potentially relevant systematic reviews of RCTs were obtained. Each paper was evaluated and categorised according to the inclusion of non-English language trials. Three types of systematic reviews were included: ‘language restricted systematic reviews’, meaning that no reports of RCTs reported in LOE were included in the quantitative data synthesis; ‘language inclusive/EL systematic reviews’, meaning that the systematic reviews searched for reports of RCTs in LOE but did not find any, and hence could not include any in the quantitative data synthesis; and ‘language inclusive/LOE systematic reviews’, meaning that the systematic reviews searched for reports of RCTs in LOE and included them in the quantitative data synthesis. We have used this nomenclature previously.30

Sample size

We estimated that 90 systematic reviews (45 restricted and 45 inclusive) would be required to detect a 25% difference in the ROR between LOE trials and EL reports, on a logarithmic scale. This sample size calculation is based on previous estimates of 18 inclusive systematic reviews including 211 trials (28 LOE and 199 EL, with an average of 12 RCTs per systematic review) from Ref. 30. The median intervention effect OR was 0.5 (i.e. −0.7 with a standard deviation of 1.13 on a log-odds scale) in favour of the intervention. We wished to observe an ROR effect modifier of 0.75 (i.e. a 25% reduction on the log-odds ratio scale) associated with language of publication with a false-positive error of 5% and power of 80% assuming a random effects (RE) model and a two-sided t-test. A total of 484 trials are required under these conditions or 40 systematic reviews, approximately. Given an LOE prevalence of ~14%, we countered the imbalance in language subgroups by a 10% increase in sample size for a total of 45 language inclusive systematic reviews. An optimal 1:1 ratio was used for restricted and inclusive systematic reviews to examine these comparisons.

Quality assessment strategy

Quality assessment of systematic reviews

The quality of reports of each systematic review was assessed using a validated scale based on the OG instrument.15,16,36 The OG instrument includes nine items pertaining to individual aspects in the reporting of a systematic review (e.g. were the search methods used to find evidence on the primary question stated?). Each item is assessed using a three-point scale (i.e. no, partially/cannot tell or yes). A final question elicits an overall scientific quality of the systematic review. The scoring ranges from one to seven, with higher scores indicating superior quality.

We standardised ourselves in using the OG instrument and pretested our methods by completing an inter-observer reliability study, which was assessed using an intra-class correlation coefficient (ICC) obtained with a separate set of 10 systematic reviews.37 Values of ICC >0.61 indicated substantial agreement, based on an a priori decision.38

---

TABLE 1: Search strategies to identify systematic reviews and meta-analyses included in (A) the CRG database and (B) EMBASE (cont’d)

<table>
<thead>
<tr>
<th>Search no.</th>
<th>Set</th>
<th>Search terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B) EMBASE</td>
<td>2</td>
<td>(integrative research review$ or research integration$).tw.</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>quantitativ$ synthesis$.tw.</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1 or 2 or 3 or 4 or 5 or 7 or 8</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>randomised controlled trial.de.</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>clinical trial.de.</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>controlled study.de.</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>major clinical study.de.</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>random$.hw.tw.</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>“0197”.tg.</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>“0150”.tg.</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>or/10-16</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>9 and 17</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>1 or 2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>10 or 11 or 14</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>19 and 20</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>limit 21 to yr=1985-1999</td>
</tr>
</tbody>
</table>
Quality assessment of the included RCTs
All included trials were quality assessed using a validated scale. It includes three items that assess the methods used to generate random assignments, double blinding and a description of dropouts and withdrawals by intervention group. The scoring ranges from one to five, with higher scoring indicating higher quality. In addition, allocation concealment was assessed as adequate, inadequate or unclear. For systematic reviews published in the CDSR, quality assessment of the included trials was obtained directly from the original reviews, if available. Two research assistants and one investigator completed the quality assessments. We used an inter-observer reliability study to standardise ourselves in using the instrument as described in the previous section.

Data extraction strategy

Data extraction of systematic reviews
Prior to data abstraction, the report of each included systematic review was masked to author and any author affiliation, journal, references and other potential identifiers.

Characteristics of each systematic review report were abstracted using a standardised data collection form (see Appendix 1). The characteristics included the source(s) of funding for the preparation of the systematic review; the indication of the intervention(s) or conditions under review using a broad International Classification of Diseases (ICD-9) category; the data sources used to identify the included RCTs; the number of independent reviewers involved in the study selection and whether a reliability test was conducted in case of multiple reviewers; the number of included trials and total number of patients; the assessment of statistical heterogeneity and methods used; the examination of clinical heterogeneity and methods used; the assessment of publication bias and methods used; the inclusion of grey literature and number and type of grey items; and the systematic review conclusion (i.e. positive, negative or unclear). Three members of the research team and one research assistant completed the data abstraction. We did not formally complete an inter-observer reliability study for this exercise as previous data indicated that the same abstractors had adequate reliability.

Validity assessment of the OG scale
The details pertaining to the processes involved in the conduct of the systematic reviews (i.e. listed above) were considered the ‘reference standard’ so as to elucidate the content and construct validity of the item-specific and overall assessment using the OG scale. For example, item one in the OG scale evaluates if a systematic review has reported the search methods used to find evidence. This should be correlated with the number of data sources the authors reported in their search strategy as collected in our Data Collection Form.

Data extraction of the included RCTs
For each language inclusive/LOE systematic review, we identified its primary outcome. If a primary outcome was not stated, a major outcome was selected based on severity (e.g. mortality) or the largest number of trials involved in the comparison between the experimental and control groups. We have previously used this approach to identify primary outcomes. All RCTs included in the primary outcome were identified through the article’s reference list. We then obtained hard copies of the included RCTs. Prior to data extraction, the report of each included RCT was masked to author and any author affiliation, journal, references and other potential identifiers.

For each included RCT, we extracted the first author’s name, journal, year of publication, language of publication and whether or not it related to CAM or CM. For the primary outcome, we also extracted the number of events and participants in the control group and the number of events and participants in the intervention group.

Data analysis

Between meta-analyses comparisons
Fisher’s exact test was used to compare the three types of systematic reviews with respect to their reporting characteristics (e.g. assessment of publication bias) and quality assessment on each of the first nine items of the OG scale. The Kruskal–Wallis test was used to assess the overall scientific quality of the systematic reviews across language restrictions. The Wilcoxon rank sum test was used to verify the association between language of publication and the Jadad quality score. Similar tests were used to evaluate the reporting of the systematic reviews of CM and CAM. We did not attempt to correct for multiple comparisons and have therefore interpreted any significant results, particularly borderline ones, cautiously.
Within language inclusive/LOE systematic review comparisons

Two major questions can be asked about the impact of language of publication on the estimates of an intervention’s effectiveness with a systematic review. First, do reports of LOE trials show systematically different intervention effects than EL trials, adjusting for the effect of the systematic review? This is an RCT-level question. Second, do pooled estimates in systematic reviews change when LOE trials are removed? This is a systematic review-level question.

To assess the impact of LOE at the RCT level, we used a fixed effects (FE) logistic regression model of the type introduced and described by Schulz and colleagues and used by others. Briefly, this model for the log odds of events in an intervention group includes a main effect for an RCT, a main effect for the systematic review, a main effect for the intervention and a main effect for the language of publication. Differences in intervention effect estimates across systematic reviews were accounted for by including an interaction term between intervention and language of publication. Including an interaction term between intervention and language of publication can assess the effect of language of publication on the estimates of intervention effect. The language effect from this logistic regression is reported as an ROR ratio of EL versus LOE trials.

The systematic review-level impact of trial language of publication was assessed as follows. We computed the log ROR of EL trials versus LOE trials for each systematic review. If there was a single LOE trial, the usual estimate of the log OR was computed along with the usual standard error estimate. Otherwise, a pooled OR was computed for the LOE trials using the Mantel–Haenszel procedure (Robins and co-workers’ estimate of the standard error was used). Using the same approach, an estimate of the log OR for the EL trials was computed, along with a standard error estimate. The log ratio of the two OR estimates for each systematic review was then computed, and its standard error was obtained as the square root of the sum of the squared standard errors of the two log ORs.

A weighted mean of these systematic-review specific estimates yields an approximation to the ROR estimate from the logistic model. In the presence of heterogeneity between the systematic-review specific estimates, a DerSimonian–Laird RE version of the weighted mean was used together with a test for heterogeneity of systematic-review specific estimates across the reviews. This approach allows a graphical display of the results to aid in their interpretation and relationship between the overall estimate of the ROR from the logistic regression approach and the individual systematic reviews.

We hypothesised a priori that the effect of language of publication on the estimates of intervention effect may not be similar for CM and CAM interventions. The systematic-review specific estimates were displayed separately for CM and CAM interventions. Subgroup analyses were performed for CM and CAM trials. Other assessments described below also reflected this hypothesis.

We plotted ORs from language inclusive/LOE meta-analyses and the RORs to display visually the effect of language restriction on estimates of intervention effect. We hypothesised a priori that the effect of language of publication on the estimates of intervention effect may not be similar for CM and CAM interventions. Subgroup analyses were performed for CM and CAM trials. Other assessments described below also reflected this hypothesis.

Language restriction and publication bias

To address the question of whether the inclusion of reports of RCTs published in LOE in a systematic review may induce a different likelihood of publication bias, we used visual inspection of funnel plots and statistical regression approaches. For each systematic review, two investigators independently assessed the degree of asymmetry in the funnel plot (log OR versus the inverse of its standard error). We reported the systematic reviews in which funnel plot asymmetry was discerned by at least one assessor and the kappa-coefficient of agreement between the two investigators. For those systematic reviews visually identified with funnel plot asymmetry, we formally tested their degree of asymmetry (at a significance level of 0.10) using two regression approaches to the funnel plot. The shift in significant funnel plot asymmetry (i.e. detected by Mascall and colleagues method) with and without language restriction was evaluated using the McNemar test (Z-scores). We plotted the Z-scores testing for funnel plot asymmetry (i.e. the intercept of the
fitted regression line by Macaskill and colleagues\textsuperscript{46} from language restricted versus language inclusive/LOE) systematic reviews. Deviation from the equal line in this plot indicates a correlation between language restriction and the possibility of publication bias.

**Language restriction and statistical heterogeneity**

For each systematic review, we used two statistics to summarise statistical heterogeneity across trial estimates of intervention effect.\textsuperscript{47} The $H$ statistic describes the relative excess in the test for heterogeneity (i.e. in the sense of the DerSimonian–Laird RE model\textsuperscript{43}) over its degree of freedom. The $I^2$ statistic is a transformation of the $H$ statistic that has an intuitive interpretation. It can be interpreted as the proportion of the between-trial variation in the estimates of intervention effect out of the total variation including both the between- and within-trial variation. Both the $H$ and $I^2$ statistics do not depend on the number of trials included in a systematic review and are recommended as an alternative to the standard statistical test for heterogeneity.\textsuperscript{47} Higgins and Thompson\textsuperscript{47} also suggest that values of $I^2 \geq 56\%$ (or $H$ values $\geq 1.5$) could induce considerable caution about between-study heterogeneity whereas values of $I^2 < 30\%$ (or $H$ values $< 1.2$) might cause little concern.

To gauge the impact of language restriction, we plotted the $I^2$ values derived from language inclusive systematic reviews against corresponding values from language restricted systematic reviews. Specific values of the $H$ and $I^2$ statistics were summarised for each systematic review together with other statistics derived from the DerSimonian–Laird RE model.
Chapter 3

Results

General characteristics of the included systematic reviews

Literature search results
The EMBASE search retrieved 1008 bibliographic records of which 118 articles were potentially relevant and unique to EMBASE. The search of the CISCOM database identified 137 citations. Our own files identified 93 reviews. After initial screening and the removal of duplicate citations, 280 potentially relevant articles remained. After final eligibility determination, the reasons for excluding the studies were as follows: the articles were not systematic reviews of RCTs (108); we could not tell if the article was language restricted or inclusive (28); the study contained too many trials (>100) limited by time and economic restraints (6); the study did not report a search strategy (5); or the study was a duplicate or we were unable to locate a copy of the paper (3); see Figure 1.

As the number of language inclusive systematic reviews that included LOE was small, we decided to include studies where the methodology was not explicitly stated as to whether they were language inclusive or language restrictive. One paper met this criterion.

Type of systematic review
The 130 systematic reviews included in our dataset were divided as follows: 50 language restricted systematic reviews, 32 language inclusive/EL systematic reviews, and 48 language inclusive/LOE systematic reviews.

Publication year
The 130 systematic reviews were published between 1989 and 1999. The median year of publication was 1994 for both language inclusive/LOE systematic reviews and language restricted systematic reviews. Language inclusive/EL systematic reviews that searched for reports of LOE but were not included in the quantitative data synthesis had a median year of publication of 1995 (Table 2).

Number of RCTs
Language inclusive/LOE systematic reviews included substantially more RCTs (median number of RCTs = 17) compared with the number of trial reports in either the language restricted systematic reviews (median number of RCTs = 11) or language inclusive/EL systematic reviews (median number of RCTs = 8) (Table 2 and

FIGURE 1 Flow diagram of phases in the inclusion of systematic reviews
<table>
<thead>
<tr>
<th></th>
<th>Language restricted systematic reviews (n = 50): median [1st, 3rd Q] (range)</th>
<th>Language inclusive/EL systematic reviews (n = 32): median [1st, 3rd Q] (range)</th>
<th>Language inclusive/LOE systematic reviews (n = 48): Median [1st, 3rd Q] (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of RCTs</td>
<td>11 [6, 23] (4, 106)</td>
<td>8 [6, 13] (3, 26)</td>
<td>17 [9, 25] (3, 119)</td>
</tr>
<tr>
<td>Total number of patients</td>
<td>971 [419, 2641] (112, 52869)</td>
<td>1121 [352, 2385] (150, 361433)</td>
<td>1658 [798, 3604] (112, 40431)</td>
</tr>
<tr>
<td>Citation impact(^b)</td>
<td>1.67 [0.99, 4.04] (0.08, 15.94)</td>
<td>2.08 [0.71, 4.44] (0.13, 15.94)</td>
<td>1.96 [1.27, 4.34] (0.70, 15.94)</td>
</tr>
<tr>
<td>Number of LOE trials</td>
<td>2 [1, 4] (1, 4)</td>
<td>14 [11, 36] (5, 67)</td>
<td>14 [11, 36] (5, 67)</td>
</tr>
<tr>
<td>Proportion of LOE trials (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of languages</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danish</td>
<td>3 (0.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>2 (0.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>546 (80)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>57 (8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>52 (8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italian</td>
<td>15 (2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>1 (0.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portuguese</td>
<td>1 (0.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>6 (0.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>684 (100)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circulatory disease</td>
<td>18 (36)</td>
<td>7 (22)</td>
<td>12 (25)</td>
</tr>
<tr>
<td>Infectious disease</td>
<td>2 (4)</td>
<td>7 (22)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Digestive</td>
<td>4 (8)</td>
<td>2 (6)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Pregnancy and childbirth</td>
<td>3 (6)</td>
<td>4 (13)</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Genitourinary</td>
<td>2 (4)</td>
<td>2 (6)</td>
<td>6 (13)</td>
</tr>
<tr>
<td>Mental health</td>
<td>2 (4)</td>
<td>2 (6)</td>
<td>5 (10)</td>
</tr>
<tr>
<td>Nervous system and sense organs</td>
<td>4 (8)</td>
<td>1 (3)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Neoplasms</td>
<td>3 (6)</td>
<td>1 (3)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Respiratory</td>
<td>2 (4)</td>
<td>1 (3)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Other</td>
<td>10 (20)</td>
<td>5 (16)</td>
<td>7 (15)</td>
</tr>
<tr>
<td>Type of journal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>14 (28)</td>
<td>15 (47)</td>
<td>23 (48)</td>
</tr>
<tr>
<td>Specialty</td>
<td>36 (72)</td>
<td>17 (53)</td>
<td>25 (52)</td>
</tr>
<tr>
<td>Type of interventions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>39 (78)</td>
<td>28 (88)</td>
<td>38 (79)</td>
</tr>
<tr>
<td>CAM</td>
<td>11 (22)</td>
<td>4 (13)</td>
<td>10 (21)</td>
</tr>
<tr>
<td>Funding source</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-pharmaceutical company</td>
<td>22 (44)</td>
<td>15 (47)</td>
<td>17 (35)</td>
</tr>
<tr>
<td>None listed/cannot tell</td>
<td>27 (54)</td>
<td>17 (53)</td>
<td>30 (63)</td>
</tr>
<tr>
<td>Single pharmaceutical company</td>
<td>1 (2)</td>
<td></td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

\(^a\) Q = quartile.

\(^b\) SCI Journal Citation Report, 1992.
The 48 language inclusive/LOE systematic reviews included 684 reports of RCTs.  

**Number of participants**

Language inclusive/LOE systematic reviews included substantially larger numbers of participants (median number of participants = 1658) compared with either the language restricted systematic reviews (median number of participants = 971) or language inclusive/EL systematic reviews (median number of participants = 1121) (Table 2).

**Disease classification**

The 130 systematic reviews reported investigating a broad spectrum of diseases. For all three types of systematic reviews, diseases of the circulatory system were most commonly investigated [language restricted, n = 18 (36%); language inclusive/EL, n = 7 (22%); language inclusive/LOE, n = 12 (25%) (Table 2)].

**Journal type**

Slightly more than half (53%) of the language inclusive/EL systematic reviews and language inclusive/LOE systematic reviews (52%) were published in specialty journals as compared with general medical journals (Table 2). This contrasts with almost three-quarters (72%) of the language restricted systematic reviews that were retrieved from specialty journals as compared with general medical journals (Table 2). Language restricted systematic reviews had the lowest citation impact factor (median = 1.67), whereas language inclusive/EL systematic reviews had the highest citation impact factor (median = 2.08) (Table 2).

**Language of publication**

Reports of English language RCTs dominated those reports included in the language inclusive/LOE systematic reviews (n = 546, 80%) (Table 2). Of the remaining 138 reports of RCTs, 57 (8% of the total) were reported in German and 52 (8% of the total) in French. In total, there were reports of RCTs in eight LOE.

**Interventions**

Approximately three-quarters of the interventions examined in the 130 systematic review reports were in CM [language restricted, n = 39 (78%);
language inclusive/LE, n = 28 (88%); and
language inclusive/LOE, n = 38 (79%).

Funding source
A single pharmaceutical company funded a
minority of the 130 systematic reviews.
Approximately one-third (35%) of the language
inclusive/EL systematic reviews, 47% of the
language restricted systematic reviews had
no pharmaceutical funding. For more than half of
all the systematic reviews, no funding source was
specified or we could not ascertain a funding
source (Table 2).

Type of intervention: comparison of
CM versus CAM
Of the 130 systematic reviews, 105 examined CM
interventions with the remaining 25 reviews
examining CAM interventions. The CAM
interventions included examining the effectiveness
of acupuncture for smoking cessation, garlic for
blood pressure control, acupuncture for chronic
pain, fish oil for the management of rheumatoid
arthritis and cognitive behaviour therapy for
hypertension. As specified a priori, we set out to
evaluate the effect language restrictions have on
the results of both CM and CAM systematic
reviews. As such, we qualitatively compared the
CM systematic reviews with the CAM reviews.
These results are reported below and in Table 3.

Publication year
The CAM reviews were published more recently
(median year of publication = 1995) than the CM
reviews (median year of publication = 1994)
(Table 3). Both types of systematic reviews (i.e. CM
and CAM systematic reviews) were published

Number of RCTs
Systematic reviews examining CAM interventions
included more reports of RCTs (median number
of RCTs = 14) compared with their CM
counterparts (median number of RCTs = 12)
(Table 3).

<table>
<thead>
<tr>
<th>TABLE 3 General characteristics of 105 systematic review examining CM interventions and 25 systematic reviews investigating CAM interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CM systematic reviews</strong>&lt;br&gt;<strong>(n = 105):</strong>&lt;br&gt;median [1st, 3rd Q] (range)</td>
</tr>
<tr>
<td><strong>CAM systematic reviews</strong>&lt;br&gt;<strong>(n = 25):</strong>&lt;br&gt;median [1st, 3rd Q] (range)</td>
</tr>
<tr>
<td><strong>Number of RCTs</strong></td>
</tr>
<tr>
<td><strong>Total number of patients</strong></td>
</tr>
<tr>
<td><strong>Citation impact</strong>&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Disease area</strong></td>
</tr>
<tr>
<td><strong>Circulatory disease</strong></td>
</tr>
<tr>
<td><strong>Infectious disease</strong></td>
</tr>
<tr>
<td><strong>Digestive</strong></td>
</tr>
<tr>
<td><strong>Pregnancy and childbirth</strong></td>
</tr>
<tr>
<td><strong>Genitourinary</strong></td>
</tr>
<tr>
<td><strong>Mental health</strong></td>
</tr>
<tr>
<td><strong>Central nervous and sense</strong></td>
</tr>
<tr>
<td><strong>Neoplasms</strong></td>
</tr>
<tr>
<td><strong>Respiratory</strong></td>
</tr>
<tr>
<td><strong>Muscular skeletal and connective tissues</strong></td>
</tr>
<tr>
<td><strong>Other</strong></td>
</tr>
<tr>
<td><strong>Type of journal</strong></td>
</tr>
<tr>
<td><strong>General</strong></td>
</tr>
<tr>
<td><strong>Specialty</strong></td>
</tr>
<tr>
<td><strong>Funding source</strong></td>
</tr>
<tr>
<td><strong>Non-pharmaceutical company</strong></td>
</tr>
<tr>
<td><strong>None listed/cannot tell</strong></td>
</tr>
<tr>
<td><strong>Pharmaceutical company</strong></td>
</tr>
</tbody>
</table>

<sup>a</sup> Q = quartile.<br><sup>b</sup> SCI Journal Citation Report, 1992.
Number of participants
Systematic reviews examining CAM interventions included fewer participants (median number of participants = 793) compared with the systematic reviews of CM interventions (median number of participants = 1407).

Disease classification
The largest percentage (36%) of systematic reviews in CAM evaluated interventions in circulatory disease. A smaller percentage (27%), but also the largest percentage of systematic reviews in CM, evaluated interventions for people diagnosed with circulatory disease.

The 105 systematic reviews in CM evaluated interventions within certain disease areas not examined by systematic reviews in CAM reports included in our sample (i.e. infectious disease, neoplasms, respiratory disease, muscular, skeletal and connective tissues). Our sample did not include any systematic reviews in CAM that examined interventions for disease categories not already covered by those systematic reviews in CM.

Journal type
Sixty per cent of systematic reviews in CM and CAM were published in specialty journals, although the citation impact factor was higher for systematic reviews in CM (median impact factor = 1.92) compared with CAM (median impact factor = 1.44) (Table 3).

Funding source
Approximately half of the systematic reviews in both CM (59%) and CAM (48%) did not report any funding source or we could not locate any information about funding source from the report.

Quality of reporting of systematic reviews (Tables 4 and 5, Figure 3)

Quality assessment of systematic reviews
We used the OG instrument and data derived from our Data Collection Form to evaluate the quality of reporting of the 50 language restricted systematic reviews, the 32 language inclusive/EL systematic reviews and the 48 language inclusive/LOE systematic reviews. The quality assessment (OG and Jadad) reliability results (ICC) were 0.66. For allocation concealment the corresponding result was 0.67.

Overall, language inclusive/LOE systematic reviews seemed to have better reporting compared with the other two groups; language inclusive/LOE systematic reviews had a median overall OG score of four out of seven, compared with a median OG score of three out of seven for language restricted and language inclusive/EL reviews (p = 0.25, Figure 4). However, each of the systematic reviews from all three groups had some major flaws in their reporting. For example, only 44–60% of systematic reviews reported the search methods; search strategies were reasonably comprehensive in 28–52% of the systematic reviews, and 23–31% of the systematic reviews avoided bias in their selection of studies.

Search strategy
According to item one of the OG instrument, 60% of language restricted, 44% of language inclusive/EL and 50% of language inclusive/LOE systematic reviews reported the search methods used to find evidence. Both language restricted and language inclusive/EL systematic reviews reported a range of search sources, most notably MEDLINE, reference lists of potentially relevant reports and other electronic databases (e.g. CDSR). Language inclusive/LOE systematic reviews also reported a wide range of search sources, including MEDLINE, reference lists, corresponding authors, other electronic databases and EMBASE. Overall, 84–90% of the systematic reviews searched MEDLINE. However, language inclusive/LOE systematic reviews were more likely to search EMBASE (30%) compared with 6% for the language restricted and 13% for the language inclusive/EL reviews (p = 0.07). The language inclusive/LOE systematic reviews also reported corresponding with authors of potentially relevant studies more frequently (34%) compared with 17 and 18% for the language restricted and the language inclusive/EL systematic reviews, respectively. This composition of literature sources appeared to be more comprehensive than the other two systematic review groups. However, this observation was not reflected in the scoring of the OG instrument (item two), which indicated that 48–52% of the language restricted and language inclusive/LOE systematic reviews employed a reasonably comprehensive search. The corresponding number for language inclusive/EL reviews was 28%.

Study selection
According to item three of the OG instrument, 75–84% of the systematic reviews reported the criteria for deciding which study to include, with no significant differences between the three types of systematic reviews. However, only 23–31% avoided bias in the selection of studies (OG item four). This poor rating was supported by the data.
TABLE 4  Quality of reports of language inclusive systematic reviews and language restricted systematic reviews: no. (%) (see text for details)

<table>
<thead>
<tr>
<th>Question</th>
<th>Language restricted systematic reviews (n = 50)</th>
<th>Language inclusive/EL systematic reviews (n = 32)</th>
<th>Language inclusive/LOE systematic reviews (n = 48)</th>
<th>Two-sided p-valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Were the search methods used to find evidence reported?</td>
<td>30 (60)</td>
<td>14 (44)</td>
<td>24 (50)</td>
<td>0.35</td>
</tr>
<tr>
<td>2. Was the search for evidence reasonably comprehensive?</td>
<td>24 (48)</td>
<td>9 (28)</td>
<td>25 (52)</td>
<td>0.09</td>
</tr>
<tr>
<td>3. Were the criteria for deciding which studies to include in the overview reported?</td>
<td>40 (80)</td>
<td>27 (84)</td>
<td>36 (75)</td>
<td>0.60</td>
</tr>
<tr>
<td>4. Was bias in the selection of studies avoided?</td>
<td>13 (26)</td>
<td>10 (31)</td>
<td>11 (23)</td>
<td>0.69</td>
</tr>
<tr>
<td>5. Were the criteria used for assessing the validity of the included studies reported?</td>
<td>18 (36)</td>
<td>11 (34)</td>
<td>33 (69)</td>
<td>0.01</td>
</tr>
<tr>
<td>6. Was the validity of all of the studies referred to in the text assessed using appropriate criteria?</td>
<td>14 (28)</td>
<td>11 (34)</td>
<td>30 (63)</td>
<td>0.01</td>
</tr>
<tr>
<td>7. Were the methods to combine the findings of the relevant studies reported?</td>
<td>41 (82)</td>
<td>24 (75)</td>
<td>37 (77)</td>
<td>0.76</td>
</tr>
<tr>
<td>8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addresses?</td>
<td>43 (86)</td>
<td>24 (75)</td>
<td>38 (79)</td>
<td>0.42</td>
</tr>
<tr>
<td>9. Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?</td>
<td>40 (80)</td>
<td>25 (78)</td>
<td>39 (81)</td>
<td>0.96</td>
</tr>
</tbody>
</table>

a Fisher’s exact test.  
b Median [inter-quartile range].  
c Kruskal–Wallis test.

TABLE 5  Reporting characteristics of language restrictive systematic reviews and language inclusive systematic reviews: no. (%) 

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Language restricted systematic reviews (n = 50)</th>
<th>Language inclusive/EL systematic reviews (n = 32)</th>
<th>Language inclusive/LOE systematic reviews (n = 48)</th>
<th>Two-sided p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEDLINE</td>
<td>45 (90)</td>
<td>27 (84)</td>
<td>40 (83)</td>
<td>0.62</td>
</tr>
<tr>
<td>EMBASE</td>
<td>3 (6)</td>
<td>4 (13)</td>
<td>14 (29)</td>
<td>0.07</td>
</tr>
<tr>
<td>Hand search</td>
<td>7 (14)</td>
<td>3 (9)</td>
<td>4 (8)</td>
<td>0.68</td>
</tr>
<tr>
<td>Reference lists</td>
<td>31 (62)</td>
<td>18 (56)</td>
<td>27 (56)</td>
<td>0.83</td>
</tr>
<tr>
<td>Corresponding authors</td>
<td>9 (18)</td>
<td>5 (17)</td>
<td>16 (33)</td>
<td>0.11</td>
</tr>
<tr>
<td>Content experts</td>
<td>6 (12)</td>
<td>4 (13)</td>
<td>7 (15)</td>
<td>0.95</td>
</tr>
<tr>
<td>Abstracts</td>
<td>4 (8)</td>
<td>4 (13)</td>
<td>7 (15)</td>
<td>0.61</td>
</tr>
<tr>
<td>Conference proceedings</td>
<td>1 (2)</td>
<td>3 (9)</td>
<td>5 (10)</td>
<td>0.23</td>
</tr>
<tr>
<td>Other electronic databases</td>
<td>12 (24)</td>
<td>9 (28)</td>
<td>16 (33)</td>
<td>0.60</td>
</tr>
<tr>
<td>Cannot tell</td>
<td>4 (8)</td>
<td>1 (3)</td>
<td>4 (8)</td>
<td>0.75</td>
</tr>
</tbody>
</table>

continued
## TABLE 5 Reporting characteristics of language restrictive systematic reviews and language inclusive systematic reviews: no. (%) (cont’d)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Language restricted systematic reviews</th>
<th>Language inclusive/EL systematic reviews</th>
<th>Language inclusive/LOE systematic reviews</th>
<th>Two-sided p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent review for study selection</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single reviewer</td>
<td>0</td>
<td>4 (13)</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>More than one reviewer</td>
<td>21 (42)</td>
<td>12 (38)</td>
<td>13 (27)</td>
<td></td>
</tr>
<tr>
<td>Inter-reviewer reliability</td>
<td>5 (10)</td>
<td>5 (16)</td>
<td>3 (6)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>29 (58)</td>
<td>16 (50)</td>
<td>35 (73)</td>
<td></td>
</tr>
<tr>
<td>Independent data extraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One reviewer</td>
<td>2 (4)</td>
<td>2 (6)</td>
<td>1 (2)</td>
<td>0.17</td>
</tr>
<tr>
<td>More than one reviewer</td>
<td>18 (36)</td>
<td>16 (50)</td>
<td>27 (56)</td>
<td></td>
</tr>
<tr>
<td>Used a standard form for data extraction</td>
<td>10 (20)</td>
<td>3 (10)</td>
<td>5 (10)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>30 (60)</td>
<td>14 (44)</td>
<td>18 (38)</td>
<td></td>
</tr>
<tr>
<td>Assessment of statistical heterogeneity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting a chi-squared test for heterogeneity</td>
<td>37 (74)</td>
<td>28 (88)</td>
<td>40 (83)</td>
<td>0.30</td>
</tr>
<tr>
<td>Using L’Abbé plot</td>
<td>4 (8)</td>
<td>4 (12)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Investigation of clinical heterogeneity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using subgroup analyses</td>
<td>23 (46)</td>
<td>13 (41)</td>
<td>18 (38)</td>
<td>0.67</td>
</tr>
<tr>
<td>Using sensitivity analyses</td>
<td>17 (34)</td>
<td>9 (28)</td>
<td>17 (35)</td>
<td></td>
</tr>
<tr>
<td>Using covariates</td>
<td>3 (6)</td>
<td>3 (9)</td>
<td>6 (13)</td>
<td></td>
</tr>
<tr>
<td>Using control rate</td>
<td>4 (8)</td>
<td>4 (13)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Statistical model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed effects</td>
<td>31 (62)</td>
<td>15 (47)</td>
<td>25 (52)</td>
<td>0.46</td>
</tr>
<tr>
<td>Random effects</td>
<td>9 (18)</td>
<td>12 (38)</td>
<td>12 (25)</td>
<td></td>
</tr>
<tr>
<td>Both fixed and random effects</td>
<td>6 (12)</td>
<td>2 (6)</td>
<td>4 (8)</td>
<td></td>
</tr>
<tr>
<td>Cannot tell</td>
<td>4 (8)</td>
<td>3 (9)</td>
<td>7 (15)</td>
<td></td>
</tr>
<tr>
<td>Inclusion of grey literature</td>
<td>3 (6)</td>
<td>4 (13)</td>
<td>19 (40)</td>
<td>0.01</td>
</tr>
<tr>
<td>Number of grey items</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One</td>
<td>2 (4)</td>
<td>1 (3)</td>
<td>6 (13)</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>4 (8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than two</td>
<td>1 (2)</td>
<td>2 (6)</td>
<td>9 (19)</td>
<td></td>
</tr>
<tr>
<td>Sources</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abstracts</td>
<td>1 (2)</td>
<td>1 (3)</td>
<td>11 (23)</td>
<td></td>
</tr>
<tr>
<td>Unpublished</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In press</td>
<td>2 (6)</td>
<td>7 (15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thesis</td>
<td>1 (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Book chapters</td>
<td>2 (4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Company reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment of publication bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision to assess publication bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A priori</td>
<td>1 (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posteriori</td>
<td>6 (12)</td>
<td>5 (16)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Method used</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fail-safe</td>
<td>5 (10)</td>
<td>5 (16)</td>
<td>6 (13)</td>
<td></td>
</tr>
<tr>
<td>Funnel plot</td>
<td>1 (3)</td>
<td>2 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selection model</td>
<td>1 (2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussed impact of publication bias on results</td>
<td>16 (32)</td>
<td>11 (34)</td>
<td>16 (33)</td>
<td>0.97</td>
</tr>
<tr>
<td>Conclusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive/significant</td>
<td>33 (66)</td>
<td>17 (53)</td>
<td>32 (67)</td>
<td>0.18</td>
</tr>
<tr>
<td>Negative/non-significant</td>
<td>10 (20)</td>
<td>8 (25)</td>
<td>13 (27)</td>
<td></td>
</tr>
<tr>
<td>Unclear</td>
<td>7 (14)</td>
<td>7 (22)</td>
<td>2 (4)</td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 3 Quality of reporting according to type of systematic review (language restricted systematic reviews, language inclusive/EL systematic reviews and language inclusive/LOE systematic reviews).

FIGURE 4 Quality of reporting of systematic reviews and type of interventions.
derived from our Data Collection Forms, which indicated that the majority of systematic reviews (50–73%) did not report the number of independent reviewers for study selection. Of the language inclusive/LOE systematic reviews, 27% were conducted with multiple reviewers for study selection compared with 42% of the language restricted systematic reviews. None of the language restricted or the language inclusive/LOE reviews reported using only a single reviewer for study selection, although 13% of the language inclusive/EL systematic reviews reported using only a single reviewer (p = 0.01). Inter-reviewer reliability was assessed by only 6% of the language inclusive/LOE reviews compared with 10% of the language restricted and 16% of the language inclusive/EL systematic reviews.

Inclusion of grey literature
Consistent with their broader search strategy, language inclusive/LOE systematic reviews included more grey literature (40%) compared with the language restricted (6%) and the language inclusive/EL reviews (13%). Of the LOE reviews, 19% included more than two types of grey literature, with 23% including abstracts and 15% including unpublished material.

Trial size and number of patients
The broader search strategy of the language inclusive/LOE systematic reviews is supported by the fact that these reviews included significantly more trials in their reviews than the language restricted or language inclusive/EL systematic reviews (median of 17 trials per review compared with 11 and 8, respectively) (Figure 3). In addition, the language inclusive/LOE systematic reviews included substantially more participants (median number of participants = 1658) than the language restricted (median number = 971) and language inclusive/EL systematic reviews (median number = 1121).

Validity assessment
Language inclusive/LOE systematic reviews were more likely to report the criteria that they used for assessing the validity of included trials (69%) than were the language restricted (36%) and language inclusive/EL systematic reviews (34%) (OG item five). The language inclusive/LOE systematic reviews were also more likely to assess appropriately the validity of included trials (63%) than were the language restricted (28%) and language inclusive/EL reviews (34%) (OG item six).

Data extraction
Of the systematic reviews, 44–60% did not identify how data extraction was done. More than one reviewer was used by 36% of the language inclusive/LOE systematic reviews compared with 50% of the language inclusive/EL reviews and only 36% of the language restricted reviews. Only 10–20% of the reviews reported using a standard form for data extraction.

Quantitative synthesis
Of the systematic reviews, 75–82% reported the method used to combine the relevant studies (OG item seven), with 75–86% using an appropriate method (OG item eight). Both FE (47–67%) and RE models (18–38%) were used in these systematic reviews, with only 6–12% of authors reporting the use of both models in the same systematic review.

Statistical heterogeneity was assessed by ~77–88% of the systematic reviews, with 68–78% reporting the chi-squared test for heterogeneity, and only 2–12% reporting the use of the L’Abbé plot for this assessment. Fewer than half of the systematic reviews investigated for clinical heterogeneity, with subgroup analyses being the methods most frequently used (28–34%), followed by sensitivity analyses (6–13%). Two examples in which the investigation for clinical heterogeneity was relatively well conducted are briefly described below.

In a systematic review of the efficacy of Bacillus Calmette–Guérin (BCG) vaccination of newborns and infants in the prevention of tuberculosis (TB), Colditz and colleagues investigated BCG efficacy using studies conducted over a period of more than 50 years, reflecting decades of changes in medical practice, reporting techniques and the design and conduct of studies. In order partially to account for these sources of heterogeneity, they developed a validity scale to assess the potential for bias and ascertainment of TB diagnosis. They specified methods dealing with variation in outcomes (e.g. TB cases, laboratory-confirmed cases, TB deaths), TB strain and duration of BCG protection. They reported BCG efficacy under a range of clinical heterogeneity and estimated that the validity score could explain up to 15% of the among-study heterogeneity in BCG effect.

Rowe and colleagues performed a systematic review of seven RCTs examining the effectiveness of steroid therapy in acute exacerbations of asthma. At the outset of the review, they qualitatively assessed characteristics of the included trials (e.g. study design, patient population, intervention and outcome measurement) that could potentially result in statistical heterogeneity. Interventions
and outcome measures were thought to be similar in the various trial categories. Clinical heterogeneity was then thought to arise as a result of differences in populations (adults versus children) and/or study design. When statistical heterogeneity was encountered, subgroup analysis of these two factors was performed in an attempt to explain the findings.

The assessment of publication bias was not well conducted among any of the three types of systematic reviews. Publication bias was assessed for only 14–17% of the reviews. The fail-safe number (i.e. the number of unpublished studies finding no treatment difference that would be necessary to refute a statistically significant treatment difference) was reportedly the method most frequently used (10–16%). The funnel plot (i.e. the plot of effect size versus its precision or sample size) used to detect for potential publication bias was displayed in no more than 4% of the reviews. A passing mention of publication bias in the discussion section of the report was more frequent (32%) than the actual assessment of the bias (15%), with one notable exception. Linde and colleagues\(^5\) examined whether the clinical effects of homoeopathy were placebo effects in a meta-analysis of double-blind and/or randomised placebo-controlled trials. The authors assumed that publication bias occurred in the data despite extensive efforts to collect all relevant studies. They assessed for publication bias using funnel plots and evaluated the robustness of homoeopathy effect estimate using the fail-safe method. In addition, a statistical test for publication bias and a correction for its effects on the estimate of homoeopathy effect were performed. For these, they used a random effects model for treatment effect estimates and a selection model in which the likelihood that a study was reported depended on the significance level of treatment comparison derived from the study.

**Systematic review conclusions**

The likelihood of reaching a positive conclusion did not seem to differ among the three language restriction groups; 66% of the language restricted reviews had a positive conclusion compared with 53% of the language inclusive/EL reviews and 67% of the language inclusive/LOE reviews. However, the conclusions were unclear in 14–22% of the language restricted and language inclusive/EL reviews, whereas only 4% of the language inclusive/LOE reviews had unclear conclusions. The conclusions appeared to be supported by the data and/or conclusions in 78–81% of the three types of systematic reviews (OG item nine).

**Comparison of quality of reporting of CM versus CAM systematic reviews (Tables 6 and 7, Figure 4)**

Our database included 105 systematic reviews involving CM interventions and 25 systematic reviews of CAM interventions. The quality of
### TABLE 7 Reporting characteristics of CM and CAM systematic reviews

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CM</th>
<th>CAM</th>
<th>Two-sided p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search strategy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEDLINE</td>
<td>87 (83)</td>
<td>25 (100)</td>
<td>0.02</td>
</tr>
<tr>
<td>EMBASE</td>
<td>14 (13)</td>
<td>7 (28)</td>
<td>0.13</td>
</tr>
<tr>
<td>Hand search</td>
<td>11 (11)</td>
<td>3 (12)</td>
<td>0.73</td>
</tr>
<tr>
<td>Reference lists</td>
<td>60 (57)</td>
<td>16 (64)</td>
<td>0.65</td>
</tr>
<tr>
<td>Corresponding authors</td>
<td>20 (19)</td>
<td>10 (40)</td>
<td>0.04</td>
</tr>
<tr>
<td>Content experts</td>
<td>13 (12)</td>
<td>4 (16)</td>
<td>0.74</td>
</tr>
<tr>
<td>Abstracts</td>
<td>12 (11)</td>
<td>3 (12)</td>
<td>1.00</td>
</tr>
<tr>
<td>Conference proceedings</td>
<td>7 (7)</td>
<td>2 (8)</td>
<td>0.68</td>
</tr>
<tr>
<td>Other electronic databases</td>
<td>25 (24)</td>
<td>13 (52)</td>
<td>0.08</td>
</tr>
<tr>
<td>Cannot tell</td>
<td>8 (8)</td>
<td>1 (4)</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Independent review for study selection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single reviewer</td>
<td>4 (4)</td>
<td>0 (0)</td>
<td>1.00</td>
</tr>
<tr>
<td>More than one reviewer</td>
<td>37 (36)</td>
<td>9 (36)</td>
<td></td>
</tr>
<tr>
<td>Inter-reviewer reliability</td>
<td>8 (8)</td>
<td>5 (20)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>63 (61)</td>
<td>16 (64)</td>
<td></td>
</tr>
<tr>
<td><strong>Independent data extraction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One reviewer</td>
<td>4 (4)</td>
<td>1 (4)</td>
<td>0.85</td>
</tr>
<tr>
<td>More than one reviewer</td>
<td>47 (45)</td>
<td>13 (52)</td>
<td></td>
</tr>
<tr>
<td>Used of a standard form for data extraction</td>
<td>14 (14)</td>
<td>4 (16)</td>
<td></td>
</tr>
<tr>
<td>Not reported</td>
<td>53 (51)</td>
<td>11 (44)</td>
<td></td>
</tr>
<tr>
<td><strong>Assessment of statistical heterogeneity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting a chi-squared test for heterogeneity</td>
<td>83 (80)</td>
<td>22 (88)</td>
<td>0.41</td>
</tr>
<tr>
<td>Using L’Abbé plot</td>
<td>7 (7)</td>
<td>2 (8)</td>
<td></td>
</tr>
<tr>
<td><strong>Investigation of clinical heterogeneity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using subgroup analyses</td>
<td>44 (42)</td>
<td>10 (40)</td>
<td>1.00</td>
</tr>
<tr>
<td>Using sensitivity analyses</td>
<td>34 (33)</td>
<td>9 (36)</td>
<td></td>
</tr>
<tr>
<td>Using covariates</td>
<td>11 (11)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Using control rate association</td>
<td>8 (8)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td><strong>Statistical model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed effects</td>
<td>59 (56)</td>
<td>12 (48)</td>
<td>0.51</td>
</tr>
<tr>
<td>Random effects</td>
<td>24 (23)</td>
<td>9 (36)</td>
<td></td>
</tr>
<tr>
<td>Both fixed and random effects</td>
<td>11 (11)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Cannot tell</td>
<td>11 (11)</td>
<td>3 (12)</td>
<td></td>
</tr>
<tr>
<td><strong>Inclusion of grey literature</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of grey items</td>
<td>22 (21)</td>
<td>3 (12)</td>
<td>0.41</td>
</tr>
<tr>
<td>One</td>
<td>8 (8)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Two</td>
<td>3 (3)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>More than two</td>
<td>11 (11)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td><strong>Sources</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abstracts</td>
<td>11 (11)</td>
<td>2 (8)</td>
<td></td>
</tr>
<tr>
<td>Unpublished</td>
<td>8 (8)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>In press</td>
<td></td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Thesis</td>
<td>1 (1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Book chapters</td>
<td>3 (3)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Company reports</td>
<td>3 (3)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td><strong>Assessment of publication bias</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision to assess publication bias</td>
<td>16 (15)</td>
<td>4 (16)</td>
<td>1.00</td>
</tr>
<tr>
<td>A priori</td>
<td>5 (5)</td>
<td>3 (12)</td>
<td></td>
</tr>
<tr>
<td>Posteriori</td>
<td>11 (11)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td><strong>Method used</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fail-safe</td>
<td>16 (15)</td>
<td>2 (8)</td>
<td></td>
</tr>
<tr>
<td>Funnel plot</td>
<td>1 (1)</td>
<td>1 (4)</td>
<td></td>
</tr>
</tbody>
</table>

*continued*
reporting of the CM systematic reviews was compared with that of the CAM systematic reviews using the OG scale and the quality of reporting data from our Data Collection Forms. Overall, the CAM reviews were rated as being of higher quality with a median OG score of five out of seven compared with an OG score of three out of seven for the CM systematic reviews (Figure 5 and Table 7).

**Search strategy**

Of the CAM systematic reviews, 68% reported the search methods used to find evidence compared with 49% of the CM reviews (OG item one). The CAM systematic reviews also had more comprehensive search strategies than the CM reviews (64% vs 40%, \( p = 0.04 \), OG item two). This rating was supported by the detailed quality of reporting we examined which indicated that the CAM systematic reviews were significantly more likely to search MEDLINE (100% versus 83%, \( p = 0.02 \)), contact corresponding authors (40% versus 19%, \( p = 0.04 \)) and use other electronic databases (52% versus 24%, \( p = 0.08 \)).

**Study selection**

The CAM systematic reviews were more likely than the CM reviews to report the criteria that they used for deciding which studies to include (100% versus 74%, \( p = 0.02 \), OG item three). Only 26–28% of the reviews avoided bias in the selection of studies (OG item four). None of the CAM reviews reported using a single reviewer compared with 4% of the CM reviews. However, 61–64% did not report how many reviewers were

---

**TABLE 7 Reporting characteristics of CM and CAM systematic reviews (cont’d)**

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>CM (n = 105)</th>
<th>CAM (n = 25)</th>
<th>Two-sided p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussed impact of publication bias on results</td>
<td>33 (32)</td>
<td>10 (40)</td>
<td>0.48</td>
</tr>
<tr>
<td>Conclusion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive/significant</td>
<td>68 (65)</td>
<td>14 (56)</td>
<td>0.07</td>
</tr>
<tr>
<td>Negative/non-significant</td>
<td>21 (20)</td>
<td>10 (40)</td>
<td></td>
</tr>
<tr>
<td>Unclear</td>
<td>15 (14)</td>
<td>1 (4)</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 5 Effect of language of publication on quality of reporting of RCTs**

Language of publication

English

Language of publication

LOE

N = 498

N = 114

LOE

LOE

CM

CAM

Results
used. About 36% of the language restricted systematic reviews reported using more than one reviewer for study selection, but only 8% of the CM reviews reported that they had assessed inter-reviewer reliability compared with 20% of the CAM reviews.

Inclusion of grey literature
About 21% of the CM systematic reviews included grey literature compared with 12% of the CAM reviews ($p = 0.41$), with abstracts and unpublished material being the most frequently included types of grey literature.

Trial size and number of patients
Consistent with their broader search strategy, the number of trials included in the CAM reviews was likely to be larger (median number of trials = 14) compared with the CM reviews (median number of trials = 12). However, the CAM trials on average were smaller than the CM trials, resulting in CAM systematic reviews including fewer participants (median number of participants = 793) than the CM reviews (median number = 1407).

Validity assessment
About 80% of the CAM systematic reviews reported the criteria that they used to assess validity of the included studies (OG item five) compared with 40% of the CM reviews ($p = 0.003$), and these criteria were more likely to be appropriate than those used for the CM reviews (68% versus 36%, $p = 0.06$, OG item six).

Data extraction
There were no significant differences in the methods of data extraction between CM and CAM systematic reviews. Some 44–51% did not report how many reviewers were used and only 14–16% reported using a standard form for data extraction.

Quantitative synthesis
Of the CM reviews, 77% reported the methods that they used to combine the studies compared with 88% of the CAM reviews (OG item seven). The method was deemed appropriate by 79% of the CM reviews and 88% of the CAM reviews (OG item eight). The FE model was reported by 56% of the CM and 48% of the CAM reviews. Statistical heterogeneity was assessed by 80% of the CM and 88% of the CAM reviews, most commonly using a chi-squared test. There was no difference in the reporting of investigation for clinical heterogeneity with 40% and 42% reporting that this was done, most often with subgroup analyses (33–36%). Publication bias was rarely assessed in both types of reviews (15–16%), but the decision to do so was more likely to be made a priori in the CAM reviews (12% versus 5%). About 40% of the CAM reviews discussed the impact of publication bias on the results compared with 32% of the CM reviews.

Systematic review conclusions
Of the CAM reviews, 40% had a negative conclusion compared with 20% of the CM reviews ($p = 0.07$). The conclusions of the CM reviews were more likely to be unclear (14% versus 4%). According to the OG scale (item 9), 92% of the conclusions of the CAM reviews were supported by the data and/or analysis compared with only 77% of the CM reviews.

Quality of reporting of RCTs
Quality assessment of RCTs
According to our selection criteria, a total of 42 systematic reviews were selected from the 48 language inclusive/LOE systematic reviews. We excluded six systematic reviews that did not report at least one binary outcome. The 42 systematic reviews included 622 RCTs, although we could only obtain information on allocation concealment for 593 trials and quality assessment using the Jadad scale could only be completed for 612 trials.

Comparison of RCTs published in English versus those published in an LOE (Table 8, Figure 5)
Of the 612 trials with quality assessment using the Jadad scale, 498 trials were published in English and 114 trials in LOE. Of these 114 trials, 50 (44%) were published in German and 45 (40%) in French (Table 8). The English trials were more likely to report a valid method of randomisation than the LOE trials (90% versus 79%, $p = 0.003$). They were also more likely to account for withdrawals and losses to follow-up (63% versus 44%, $p < 0.001$). However, they appeared to be

<table>
<thead>
<tr>
<th>Language</th>
<th>CM n (%)</th>
<th>CAM n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Danish</td>
<td>3 (2.6)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Dutch</td>
<td>1 (0.9)</td>
<td>1 (1.5)</td>
</tr>
<tr>
<td>French</td>
<td>46 (40.4)</td>
<td>26 (38.8)</td>
</tr>
<tr>
<td>German</td>
<td>50 (43.9)</td>
<td>35 (52.2)</td>
</tr>
<tr>
<td>Italian</td>
<td>8 (7)</td>
<td>3 (4.5)</td>
</tr>
<tr>
<td>Japanese</td>
<td>1 (0.9)</td>
<td>0</td>
</tr>
<tr>
<td>Spanish</td>
<td>5 (4.4)</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>114 (100)</td>
<td>67</td>
</tr>
</tbody>
</table>

© Queen’s Printer and Controller of HMSO 2003. All rights reserved.
Results

Comparison of RCTs for CM versus CAM interventions (Table 9, Figure 6)

Of the 638 trials, 500 trials were for CM interventions and 128 trials were for CAM interventions. Quality assessment using the Jadad scale was done on 484 CM trials and 128 CAM trials. CM trials were more likely to report adequate randomisation than the CAM trials (90% compared to 81%, $p = 0.001$). They were also more likely to adequately report losses to follow-up and withdrawals (58% compared to 45%, $p < 0.01$). However, the CAM trials were significantly more likely to report double blinding than the CM trials (90% versus 56%, $p < 0.01$). The total quality score was higher for the CM trials (median score of 3 versus 2, $p = 0.14$, Figure 7). In addition, adequate reporting of allocation concealment was more frequent amongst the 112 CAM trials than amongst the 496 CM trials (28% versus 13%, $p < 0.001$), although it was unclear in 67–81% of the trials.

Effect of language of publication on quality of reporting in CM versus CAM trials (Table 10)

There was no effect of language of publication (English versus LOE) on quality of reporting in the 484 CM trials (Figure 7). Reporting of randomisation, double-blind status, withdrawals and losses to follow-up and allocation concealment was similar between the CM trials published in English and those published in LOE. In contrast, amongst the 128 CAM trials, English trials were more likely to report adequately losses to follow-up and withdrawals than the LOE trials (57% versus 34%, $p = 0.01$). Overall, quality scores were also higher in the EL CAM trials than the LOE CAM trials (median score of 3 versus 2, $p = 0.04$) (Figure 7).

Summary

Our results indicate that the overall quality of reporting of RCTs of CAM interventions is as good as, or better than, that for CM interventions. In contrast, the quality of EL reports of CAM interventions was higher than that of the LOE CAM reports, and also both the EL and LOE
TABLE 9  Effect of language of publication and type of therapy on quality of reporting of RCTs

<table>
<thead>
<tr>
<th>Quality of report</th>
<th>Language of publication</th>
<th>p-Value</th>
<th>Type of therapy</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English (n = 498): n (%)</td>
<td></td>
<td>CM (n = 484): n (%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LOE (n = 114): n (%)</td>
<td></td>
<td>CAM (n = 128): n (%)</td>
<td></td>
</tr>
<tr>
<td>Randomisation</td>
<td>447 (90)</td>
<td>&lt;0.001</td>
<td>434 (90)</td>
<td>0.001</td>
</tr>
<tr>
<td>Double-blind</td>
<td>307 (62)</td>
<td>0.07</td>
<td>273 (56)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Withdrawal, lost to</td>
<td>314 (63)</td>
<td>&lt;0.001</td>
<td>306 (63)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (0–2)</td>
<td>248 (50)</td>
<td>0.30</td>
<td>252 (52)</td>
<td>0.20</td>
</tr>
<tr>
<td>High (3–5)</td>
<td>250 (50)</td>
<td>232 (48)</td>
<td>69 (54)</td>
<td></td>
</tr>
<tr>
<td>Median [IQR] (min., max.)</td>
<td>3 [2, 3] (0, 5)</td>
<td>0.10</td>
<td>2 [2, 3] (0, 5)</td>
<td>0.14</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate</td>
<td>75 (15)</td>
<td>0.59</td>
<td>60 (13)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Inadequate</td>
<td>32 (7)</td>
<td>0.30</td>
<td>30 (6)</td>
<td>0.5</td>
</tr>
<tr>
<td>Unclear</td>
<td>383 (78)</td>
<td>391 (79)</td>
<td>75 (67)</td>
<td></td>
</tr>
</tbody>
</table>

FIGURE 7  Effect of language of publication on quality of reporting for CM and CAM RCTs
CM reports. These results may mean that there are implications to excluding LOE trials from systematic reviews of CAM interventions. This issue is explored below.

‘Language of publication’ bias and location bias are related to the type of intervention (CM or CAM) ([Tables 11 and 12, Figures 8a, 8b and 9])

Characteristics of identified language inclusive/LOE systematic reviews
We identified 48 language inclusive/LOE systematic reviews. Of these, 42 independent systematic reviews from 41 separate publications were included for further data analysis. The remaining seven systematic reviews were excluded because the primary outcome data were recorded using continuous data.

The 42 systematic reviews are presented in [Table 11]. The reviews were published in a variety of paper-based peer-reviewed journals and the CDSR between 1984 and 1999. These systematic reviews included 662 RCTs (120 545 participants), of which 133 were trials in LOE (17 810 participants). The trials investigated a broad spectrum of CM interventions, such as pneumococcal vaccination in adults, and CAM interventions, including the effectiveness of St John’s wort. Likewise, the outcomes examined varied from being objective (e.g. pregnancy), to softer and more subjective outcomes such as respondent’s perception of the severity of a headache.

Excluding trials reported in LOE, compared with their inclusion, did not provide biased results in terms of estimates of the effectiveness of an intervention ([Table 12]) (RE ROR = 1.11; 95% CI: 0.92 to 1.34) ([Figure 8a]).

The effect was more pronounced and statistically significant when the analysis was repeated separately for CAM interventions ([Table 12]). Here, excluding trials reported in LOE, compared with their inclusion, resulted in a 63% smaller intervention effect to prevent an unwanted outcome, on average (RE ROR = 1.63; 95%CI: 1.03 to 2.60) ([Figure 8b]). This movement indicates that when reports of LOE are excluded from the meta-analytical calculations, the treatment estimates are smaller (i.e. less pronounced). However, when the analysis was limited to CM interventions, no such effect was observed ([Table 12]). Here, excluding reports of LOE, compared with their inclusion, did not bias the estimates of an intervention’s effectiveness (RE ROR = 1.02; 95% CI: 0.83 to 1.26) ([Figure 8c]).

Sensitivity analysis
When the data analysis was limited to CM interventions, the results did not change whether the systematic review included one or more than one report of a trial in LOE (see [Table 12]). That is,
<table>
<thead>
<tr>
<th>Ref. in ID</th>
<th>App. 3 First author</th>
<th>Title</th>
<th>Journal</th>
<th>Year</th>
<th>Intervention</th>
<th>Trials/patients</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 16</td>
<td>Hayashi K</td>
<td>Famotidine in the treatment of duodenal ulcer</td>
<td>Gastroenterol Int</td>
<td>1993</td>
<td>CM</td>
<td>6/831</td>
<td>Healing rate at 6 weeks</td>
</tr>
<tr>
<td>3 29</td>
<td>Pittler MH</td>
<td>Peppermint oil for irritable bowel syndrome</td>
<td>Am J Gastroenterol</td>
<td>1998</td>
<td>CAM</td>
<td>5/239</td>
<td>Global improvement</td>
</tr>
<tr>
<td>4 6</td>
<td>Carroll D</td>
<td>Transcutaneous electrical nerve stimulation in labour pain</td>
<td>Br J Obstet Gynaecol</td>
<td>1997</td>
<td>CAM</td>
<td>4/475</td>
<td>Additional analgesic used</td>
</tr>
<tr>
<td>6 23</td>
<td>Linde K</td>
<td>Are the clinical effects of homeopathy placebo effects?</td>
<td>Lancet</td>
<td>1997</td>
<td>CAM</td>
<td>74/8778</td>
<td>Binary efficacy outcome measures</td>
</tr>
<tr>
<td>7 39</td>
<td>White AR</td>
<td>Acupuncture for smoking cessation</td>
<td>Cochrane</td>
<td>1998</td>
<td>CAM</td>
<td>12/1947</td>
<td>Early abstinence</td>
</tr>
<tr>
<td>8 11</td>
<td>Douglas RM</td>
<td>Vitamin C for the common cold</td>
<td>Cochrane</td>
<td>1998</td>
<td>CM</td>
<td>18/8010</td>
<td>One or more respiratory episodes</td>
</tr>
<tr>
<td>9 19</td>
<td>Hofmeyr GJ</td>
<td>External cephalic version at term</td>
<td>Cochrane</td>
<td>1996</td>
<td>CM</td>
<td>6/612</td>
<td>Non-cephalic birth</td>
</tr>
<tr>
<td>10 18</td>
<td>Hofmeyr GJ</td>
<td>Cephalic version by postural management</td>
<td>Cochrane</td>
<td>1996</td>
<td>CM</td>
<td>3/192</td>
<td>Non-cephalic birth</td>
</tr>
<tr>
<td>11 28</td>
<td>Pace F</td>
<td>Meta-analysis of the effect of placebo on the outcome of medically treated reflux esophagitis</td>
<td>Scand J Gastroenterol</td>
<td>1995</td>
<td>CM</td>
<td>22/1224</td>
<td>Healing at 4–6 weeks</td>
</tr>
<tr>
<td>13 13</td>
<td>Fine MJ</td>
<td>Efficacy of pneumococcal vaccination in adults</td>
<td>Arch Intern Med</td>
<td>1994</td>
<td>CM</td>
<td>6/16337</td>
<td>All causes mortality</td>
</tr>
<tr>
<td>14 15</td>
<td>Halpern S</td>
<td>Postdural puncture headache and spinal needle design</td>
<td>Anesthesiology</td>
<td>1994</td>
<td>CM</td>
<td>9/1720</td>
<td>Headache</td>
</tr>
<tr>
<td>15 21</td>
<td>Leizorovicz A</td>
<td>Comparison of efficacy and safety of low molecular weight heparins and unfractionated heparin in initial treatment of deep venous thrombosis</td>
<td>BMJ</td>
<td>1994</td>
<td>CM</td>
<td>16/2055</td>
<td>Mortality</td>
</tr>
<tr>
<td>16 25</td>
<td>Marino P</td>
<td>Chemotherapy vs supportive care in advance non-small-cell cancer</td>
<td>Chest</td>
<td>1994</td>
<td>CM</td>
<td>8/689</td>
<td>Mortality at 6 months</td>
</tr>
<tr>
<td>17 14</td>
<td>Glowacki LS</td>
<td>Use of immune globulin to prevent symptomatic cytomegalovirus disease in transplant recipients – a meta-analysis</td>
<td>Clin Transplant</td>
<td>1994</td>
<td>CM</td>
<td>17/1016</td>
<td>Symptomatic cytomegalovirus disease</td>
</tr>
<tr>
<td>18 33</td>
<td>SDD Trialist Group</td>
<td>Meta-analysis of RCTs of selective decontamination of the digestive tract</td>
<td>BMJ</td>
<td>1993</td>
<td>CM</td>
<td>22/3836</td>
<td>Respiratory tract infection</td>
</tr>
</tbody>
</table>

continued


<table>
<thead>
<tr>
<th>ID</th>
<th>Ref. in App. 3</th>
<th>First author</th>
<th>Title</th>
<th>Journal</th>
<th>Year</th>
<th>Intervention</th>
<th>Trials/patients</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>40</td>
<td>Wilson AP</td>
<td>A meta-analysis of the use of amoxycillin-clavulanic acid in surgical prophylaxis</td>
<td>J Hosp Infect</td>
<td>1992</td>
<td>CM</td>
<td>20/4653</td>
<td>Wound infection</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>Cohen HJ</td>
<td>Comparison of two long-term chemotherapy regimens, with or without agents to modify skeletal repair, in multiple myeloma</td>
<td>Blood</td>
<td>1984</td>
<td>CM</td>
<td>18/3898</td>
<td>Mortality at 2 years</td>
</tr>
<tr>
<td>21</td>
<td>37</td>
<td>Vandekerckhove P</td>
<td>Androgens versus placebo or no treatment for idiopathic oligo/asthenospermia</td>
<td>Cochrane</td>
<td>1999</td>
<td>CM</td>
<td>8/908</td>
<td>Pregnancy</td>
</tr>
<tr>
<td>22</td>
<td>8</td>
<td>Covey LS</td>
<td>A meta-analysis of double-blind placebo-controlled trials of clonidine for smoking cessation</td>
<td>Br J Addict</td>
<td>1991</td>
<td>CM</td>
<td>9/813</td>
<td>Smoking cessation</td>
</tr>
<tr>
<td>23</td>
<td>22</td>
<td>Linde K</td>
<td>St John’s wort for depression</td>
<td>BMJ</td>
<td>1996</td>
<td>CAM</td>
<td>13/828</td>
<td>Responder (placebo controlled trials)</td>
</tr>
<tr>
<td>24</td>
<td>22</td>
<td>Linde K</td>
<td>St John’s wort for depression</td>
<td>BMJ</td>
<td>1996</td>
<td>CAM</td>
<td>3/317</td>
<td>Responder (active controlled trials)</td>
</tr>
<tr>
<td>26</td>
<td>38</td>
<td>Vandekerckhove P</td>
<td>The medical treatment of idiopathic oligo/asthenospermia: bromocriptine versus placebo or no treatment</td>
<td>Cochrane</td>
<td>1997</td>
<td>CM</td>
<td>3/102</td>
<td>Pregnancy</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>a’Rogvi-Hansen B</td>
<td>Glycerol treatment for acute ischaemic stroke</td>
<td>Cochrane</td>
<td>1996</td>
<td>CM</td>
<td>6/654</td>
<td>Case fatality</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>Bressa GM</td>
<td>S-Adenosyl-l-methionine (SAMe) as antidepressant</td>
<td>Acta Neurol Scand</td>
<td>1994</td>
<td>CAM</td>
<td>12/391</td>
<td>Partial to full response</td>
</tr>
<tr>
<td>31</td>
<td>17</td>
<td>Heyland DK</td>
<td>Selective decontamination of the digestive tract</td>
<td>Chest</td>
<td>1994</td>
<td>CM</td>
<td>24/3405</td>
<td>Mortality</td>
</tr>
<tr>
<td>32</td>
<td>35</td>
<td>Silagy C</td>
<td>The effectiveness of nicotine replacement therapies in smoking cessation</td>
<td>Online J Curr Lin Trials</td>
<td>1994</td>
<td>CM</td>
<td>48/16921</td>
<td>Smoking cessation</td>
</tr>
<tr>
<td>33</td>
<td>5</td>
<td>Brown KH</td>
<td>Use of nonhuman milks in the dietary management of young children with acute diarrhea</td>
<td>Pediatrics</td>
<td>1994</td>
<td>CM</td>
<td>14/935</td>
<td>Treatment failure</td>
</tr>
<tr>
<td>34</td>
<td>20</td>
<td>Leizorovich A</td>
<td>Low molecular weight heparin in prevention of perioperative thrombosis</td>
<td>BMJ</td>
<td>1992</td>
<td>CM</td>
<td>45/12777</td>
<td>Deep venous thrombosis</td>
</tr>
</tbody>
</table>

continued
<table>
<thead>
<tr>
<th>ID</th>
<th>Ref. in App. 3</th>
<th>First author</th>
<th>Title</th>
<th>Journal</th>
<th>Year</th>
<th>Intervention</th>
<th>Trials/patients</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>3</td>
<td>Boissel JP</td>
<td>Is it possible to reduce the risk of cardiovascular events in subjects suffering from intermittent claudication of the lower limbs</td>
<td>Thromb Haemost</td>
<td>1989</td>
<td>CM</td>
<td>4/618</td>
<td>Thrombotic cardiovascular events</td>
</tr>
<tr>
<td>38</td>
<td>10</td>
<td>Daya S</td>
<td>Long vs short gonadotropin releasing hormone agonist protocols for pituitary desensitisation in assisted reproductive cycles</td>
<td>Cochrane</td>
<td>1998</td>
<td>CM</td>
<td>8/988</td>
<td>Clinical pregnancy</td>
</tr>
<tr>
<td>39</td>
<td>34</td>
<td>Sikorski J</td>
<td>Support for breastfeeding mothers</td>
<td>Cochrane</td>
<td>1998</td>
<td>CM</td>
<td>11/2934</td>
<td>Stop breastfeeding</td>
</tr>
<tr>
<td>40</td>
<td>27</td>
<td>Montgomery SA</td>
<td>Comparison of compliance between serotonin reuptake inhibitors and tricyclic antidepressants</td>
<td>Int Clin Psychopharmacol</td>
<td>1995</td>
<td>CM</td>
<td>67/6850</td>
<td>Side-effects</td>
</tr>
<tr>
<td>41</td>
<td>9</td>
<td>Crawford F</td>
<td>Tropical treatments for fungal infections of the skin and nails of the foot</td>
<td>Cochrane</td>
<td>1999</td>
<td>CM</td>
<td>5/617</td>
<td>Cure</td>
</tr>
<tr>
<td>42</td>
<td>12</td>
<td>Figueredo E</td>
<td>Prophylactic ondansetron for postoperative emesis</td>
<td>Acta Anaesthesiol Scand</td>
<td>1999</td>
<td>CM</td>
<td>18/3099</td>
<td>Emesis</td>
</tr>
</tbody>
</table>

*See Appendix 3.*
<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>No. of meta-analyses/ no. of RCTs</th>
<th>Language effect (trials published in English compared with trials published in LOE): ROR (95% CI)</th>
<th>Estimated heterogeneity between trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language restricted meta-analyses compared with language inclusive/LOE meta-analyses (overall)</td>
<td>42/662</td>
<td>1.09 (0.99 to 1.21)</td>
<td>2.76 ($\chi^2 = 1706$, 619 df)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>1.07 (0.96 to 1.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE</td>
<td>1.11 (0.92 to 1.34)</td>
</tr>
<tr>
<td>Language restricted meta-analyses compared with language inclusive/LOE meta-analyses (CM)</td>
<td>34/533</td>
<td>1.01 (0.90 to 1.15)</td>
<td>2.44 ($\chi^2 = 1217$, 498 df)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>0.99 (0.88 to 1.12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE</td>
<td>1.02 (0.83 to 1.26)</td>
</tr>
<tr>
<td>Limited to meta-analyses with no. of LOE &gt; 1</td>
<td>17/356</td>
<td>0.98 (0.84 to 1.14)</td>
<td>2.20 ($\chi^2 = 745$, 338 df)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>0.98 (0.84 to 1.14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE</td>
<td>1.01 (0.77 to 1.32)</td>
</tr>
<tr>
<td>Limited to meta-analyses with no. of LOE = 1</td>
<td>17/177</td>
<td>1.07 (0.88 to 1.31)</td>
<td>2.96 ($\chi^2 = 471$, 159 df)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>1.01 (0.83 to 1.23)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE</td>
<td>1.06 (0.73 to 1.54)</td>
</tr>
<tr>
<td>Language restricted meta-analyses compared with language inclusive/LOE meta-analyses (CAM)</td>
<td>8/129</td>
<td>1.37 (1.16 to 1.61)</td>
<td>4.04 ($\chi^2 = 485$, 120 df)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>1.26 (1.05 to 1.52)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE</td>
<td>1.63 (1.03 to 2.60)</td>
</tr>
<tr>
<td>Limited to meta-analyses with no. of LOE &gt; 1</td>
<td>5/115</td>
<td>1.25 (1.03 to 1.51)</td>
<td>3.81 ($\chi^2 = 415$, 109 df)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>1.22 (1.01 to 1.47)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE</td>
<td>1.34 (0.85 to 2.12)</td>
</tr>
<tr>
<td>Limited to meta-analyses with no. of LOE = 1</td>
<td>3/14</td>
<td>2.67 (1.21 to 5.86)</td>
<td>6.60 ($\chi^2 = 66$, 10 df)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FE</td>
<td>2.52 (1.12 to 5.66)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RE</td>
<td>3.01 (0.84 to 10.85)</td>
</tr>
</tbody>
</table>
FIGURE 8 (a) Impact of LOE at the RCT level: RORs with 95% CIs for LOE versus EL for each of 42 systematic reviews, together with pooled estimates from three different approaches (shown as diamonds). (b) Impact of LOE at the RCT level: RORs with 95% CIs for LOE versus EL for each of eight complementary and alternative medicine systematic reviews, together with pooled estimates from three different approaches (shown as diamonds).

© Queen’s Printer and Controller of HMSO 2003. All rights reserved.
excluding reports of trials in LOE did not exaggerate the estimates of the intervention’s effectiveness. However, excluding even a single report of an RCT published in an LOE from a systematic review examining a CAM intervention results in a significant exaggeration of the estimates of its effectiveness.

For example, White and colleagues\(^\text{52}\) conducted a systematic review to evaluate the effectiveness of acupuncture, compared with sham acupuncture, for those wanting to quit smoking. These authors included 12 RCTs in their quantitative data synthesis, of which five were published in LOE. Acupuncture was 21\% (OR = 0.79; 95\% CI: 0.55 to 1.11) more effective (although not statistically significant) in producing early abstinence, compared with sham acupuncture (Figure 12, number 7). Restricting the data synthesis to EL reports resulted in a 7\% (OR = 0.93; 95\% CI: 0.58 to 1.49) acupuncture effect. In contrast, when the analysis was limited to reports in LOE, the acupuncture effect was substantially more pronounced at 31\% (OR = 0.69; 95\% CI: 0.39 to 1.21).
Similarly, Linde and colleagues\textsuperscript{51} reported that homeopathy was 60\% (OR = 0.40; 95\% CI: 0.32 to 0.50) more effective than placebo. When we re-synthesised the data, limiting them to EL reports, the estimate of the homeopathy’s effectiveness was similar at 53\% (OR = 0.47; 95\% CI: 0.30 to 0.68). The corresponding ROR between EL only and language inclusive meta-analyses was 1.18 (Figure 9; 0.47/0.40). We restricted our replication of Linde and colleagues’ review to RCTs reporting binary outcomes.

\textsuperscript{a}Rogvi-Hansen and Boysen\textsuperscript{53} assessed the effectiveness of glycerol treatment compared with a control for patients in the acute period following a stroke. In a systematic review of six RCTs, including one LOE report, these authors reported that glycerol provided a strong mortality protective effect of 41\% at 14 days post-stroke (OR = 0.59; 95\% CI: 0.36 to 0.98). The corresponding protective effect after excluding the LOE report was 49\% (OR 0.51; 95\% CI: 0.30 to 0.85).

\textbf{Summary}

These results provide new evidence concerning what effects language of publication restrictions can have on the results of systematic reviews. These analyses suggest that language of publication restrictions, in the quantitative data synthesis of a systematic review, depend, in part, on the type of intervention under investigation. Specifically, the present analyses suggest that limiting a CAM systematic review to EL reports will produce smaller (less beneficial) treatment effects on average. These analyses suggest that individual trial results in LOE are important and need to be included, along with EL ones, in any CAM systematic review.

In contrast, the present results suggest that limiting the language of publication of trial reports to English does not appear to result in any measurable effect on the estimates of an intervention’s effectiveness, when the intervention under investigation is CM. These results are very
similar to those reported by Moher and colleagues, although the present results are based on a dataset approximately three times the size as that used in the earlier publication. The present results, and those previously reported by Moher and colleagues, remain unchanged whether one or more LOE report is included in the systematic review. Despite the low LOE prevalence, we observed a substantial language of publication bias effect for CAM reports. The present results suggest both language of publication bias and possible location bias. As such, it is probable that the results are influenced by broader selective publication biases. Publication bias is an important selective publication bias that may be contributing to the observed results reported in this chapter. We set out to explore the impact of publication bias on these results and we report our findings below.

The impact of language restriction on between-study heterogeneity and publication bias

Language restriction and statistical heterogeneity

Figure 10 displays the degree of statistical heterogeneity of language restricted versus language inclusive/LOE systematic reviews. Of the 42 systematic reviews we studied, language of publication reduced a high degree of statistical heterogeneity (i.e. the $I^2$ statistic above 0.30) to no
<table>
<thead>
<tr>
<th>MA</th>
<th>n/pat</th>
<th>CAM?</th>
<th>Q</th>
<th>P-Homo OR</th>
<th>Low</th>
<th>High</th>
<th>I²</th>
<th>( H )</th>
<th>n/pat</th>
<th>Q</th>
<th>P-Homo OR</th>
<th>Low</th>
<th>High</th>
<th>I²</th>
<th>( H )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6/831</td>
<td>1</td>
<td>0.05</td>
<td>0.96</td>
<td>0.72</td>
<td>0.42</td>
<td>1.22</td>
<td>0.00</td>
<td>0.46</td>
<td>4/571</td>
<td>0.40</td>
<td>0.94</td>
<td>0.79</td>
<td>0.43</td>
<td>1.48</td>
</tr>
<tr>
<td>2</td>
<td>26/1490</td>
<td>1</td>
<td>0.38</td>
<td>0.23</td>
<td>0.15</td>
<td>0.08</td>
<td>0.25</td>
<td>0.35</td>
<td>1.24</td>
<td>19/1046</td>
<td>12.23</td>
<td>0.84</td>
<td>0.15</td>
<td>0.09</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>5/239</td>
<td>1</td>
<td>0.27</td>
<td>0.19</td>
<td>0.00</td>
<td>0.17</td>
<td>0.03</td>
<td>0.87</td>
<td>0.85</td>
<td>2.61</td>
<td>4/197</td>
<td>26.04</td>
<td>0.00</td>
<td>0.17</td>
<td>0.02</td>
</tr>
<tr>
<td>4</td>
<td>4/475</td>
<td>0</td>
<td>0.78</td>
<td>0.05</td>
<td>0.44</td>
<td>0.14</td>
<td>1.34</td>
<td>0.62</td>
<td>1.62</td>
<td>2/430</td>
<td>0.01</td>
<td>0.91</td>
<td>0.85</td>
<td>0.49</td>
<td>1.49</td>
</tr>
<tr>
<td>5</td>
<td>6/659</td>
<td>0</td>
<td>0.31</td>
<td>0.29</td>
<td>0.00</td>
<td>0.19</td>
<td>0.07</td>
<td>0.50</td>
<td>0.84</td>
<td>2.50</td>
<td>5/619</td>
<td>22.54</td>
<td>0.00</td>
<td>0.26</td>
<td>0.10</td>
</tr>
<tr>
<td>6</td>
<td>14/378</td>
<td>0</td>
<td>1</td>
<td>1.07</td>
<td>0.00</td>
<td>1.11</td>
<td>0.52</td>
<td>1.54</td>
<td>0.71</td>
<td>4/571</td>
<td>0.40</td>
<td>0.94</td>
<td>0.79</td>
<td>0.43</td>
<td>1.48</td>
</tr>
<tr>
<td>7</td>
<td>12/1947</td>
<td>0</td>
<td>2.36</td>
<td>0.00</td>
<td>0.50</td>
<td>0.32</td>
<td>0.58</td>
<td>0.69</td>
<td>1.76</td>
<td>30/3909</td>
<td>94.97</td>
<td>0.00</td>
<td>0.47</td>
<td>0.33</td>
<td>0.68</td>
</tr>
<tr>
<td>8</td>
<td>18/8010</td>
<td>0</td>
<td>1.53</td>
<td>0.56</td>
<td>0.92</td>
<td>0.83</td>
<td>1.03</td>
<td>0.00</td>
<td>0.96</td>
<td>17/7948</td>
<td>15.48</td>
<td>0.49</td>
<td>0.92</td>
<td>0.83</td>
<td>1.03</td>
</tr>
<tr>
<td>9</td>
<td>6/612</td>
<td>0</td>
<td>0.56</td>
<td>0.92</td>
<td>0.83</td>
<td>1.03</td>
<td>0.00</td>
<td>0.96</td>
<td>17/7948</td>
<td>15.48</td>
<td>0.49</td>
<td>0.92</td>
<td>0.83</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>24/1713</td>
<td>0</td>
<td>8.58</td>
<td>0.00</td>
<td>0.52</td>
<td>0.31</td>
<td>2.12</td>
<td>0.53</td>
<td>1.93</td>
<td>18/1100</td>
<td>74.91</td>
<td>0.00</td>
<td>3.56</td>
<td>1.92</td>
<td>6.69</td>
</tr>
<tr>
<td>11</td>
<td>16/337</td>
<td>0</td>
<td>10.84</td>
<td>0.06</td>
<td>0.99</td>
<td>0.79</td>
<td>1.25</td>
<td>0.54</td>
<td>1.47</td>
<td>5/14651</td>
<td>10.84</td>
<td>0.03</td>
<td>0.97</td>
<td>0.67</td>
<td>1.39</td>
</tr>
<tr>
<td>12</td>
<td>10/1720</td>
<td>0</td>
<td>10.37</td>
<td>0.24</td>
<td>0.46</td>
<td>0.25</td>
<td>0.86</td>
<td>0.23</td>
<td>1.14</td>
<td>8/1664</td>
<td>10.36</td>
<td>0.17</td>
<td>0.45</td>
<td>0.22</td>
<td>0.93</td>
</tr>
<tr>
<td>13</td>
<td>17/1016</td>
<td>0</td>
<td>7.58</td>
<td>0.94</td>
<td>0.70</td>
<td>0.45</td>
<td>1.09</td>
<td>0.00</td>
<td>0.71</td>
<td>14/1921</td>
<td>7.36</td>
<td>0.88</td>
<td>0.71</td>
<td>0.45</td>
<td>1.12</td>
</tr>
<tr>
<td>14</td>
<td>8/689</td>
<td>0</td>
<td>27.21</td>
<td>0.00</td>
<td>0.63</td>
<td>0.42</td>
<td>0.94</td>
<td>0.36</td>
<td>1.25</td>
<td>15/901</td>
<td>23.70</td>
<td>0.05</td>
<td>0.63</td>
<td>0.40</td>
<td>0.99</td>
</tr>
<tr>
<td>15</td>
<td>3/192</td>
<td>0</td>
<td>1.60</td>
<td>0.45</td>
<td>0.73</td>
<td>0.39</td>
<td>1.38</td>
<td>0.00</td>
<td>0.89</td>
<td>2/131</td>
<td>0.96</td>
<td>0.33</td>
<td>0.89</td>
<td>0.41</td>
<td>1.95</td>
</tr>
<tr>
<td>16</td>
<td>22/1224</td>
<td>0</td>
<td>38.56</td>
<td>0.01</td>
<td>2.83</td>
<td>1.88</td>
<td>4.26</td>
<td>0.46</td>
<td>1.35</td>
<td>20/1141</td>
<td>37.37</td>
<td>0.01</td>
<td>2.72</td>
<td>1.75</td>
<td>4.22</td>
</tr>
<tr>
<td>17</td>
<td>24/1713</td>
<td>0</td>
<td>85.88</td>
<td>0.00</td>
<td>3.32</td>
<td>2.12</td>
<td>5.31</td>
<td>0.73</td>
<td>2.50</td>
<td>30/3909</td>
<td>94.97</td>
<td>0.00</td>
<td>3.56</td>
<td>1.92</td>
<td>6.69</td>
</tr>
<tr>
<td>18</td>
<td>16/2055</td>
<td>0</td>
<td>7.58</td>
<td>0.94</td>
<td>0.70</td>
<td>0.45</td>
<td>1.09</td>
<td>0.00</td>
<td>0.71</td>
<td>14/1921</td>
<td>7.36</td>
<td>0.88</td>
<td>0.71</td>
<td>0.45</td>
<td>1.12</td>
</tr>
<tr>
<td>19</td>
<td>8/689</td>
<td>0</td>
<td>27.21</td>
<td>0.00</td>
<td>0.63</td>
<td>0.42</td>
<td>0.94</td>
<td>0.36</td>
<td>1.25</td>
<td>15/901</td>
<td>23.70</td>
<td>0.05</td>
<td>0.63</td>
<td>0.40</td>
<td>0.99</td>
</tr>
<tr>
<td>20</td>
<td>3/192</td>
<td>0</td>
<td>1.60</td>
<td>0.45</td>
<td>0.73</td>
<td>0.39</td>
<td>1.38</td>
<td>0.00</td>
<td>0.89</td>
<td>2/131</td>
<td>0.96</td>
<td>0.33</td>
<td>0.89</td>
<td>0.41</td>
<td>1.95</td>
</tr>
<tr>
<td>21</td>
<td>22/1224</td>
<td>0</td>
<td>38.56</td>
<td>0.01</td>
<td>2.83</td>
<td>1.88</td>
<td>4.26</td>
<td>0.46</td>
<td>1.35</td>
<td>20/1141</td>
<td>37.37</td>
<td>0.01</td>
<td>2.72</td>
<td>1.75</td>
<td>4.22</td>
</tr>
<tr>
<td>22</td>
<td>24/1713</td>
<td>0</td>
<td>85.88</td>
<td>0.00</td>
<td>3.32</td>
<td>2.12</td>
<td>5.31</td>
<td>0.73</td>
<td>2.50</td>
<td>30/3909</td>
<td>94.97</td>
<td>0.00</td>
<td>3.56</td>
<td>1.92</td>
<td>6.69</td>
</tr>
<tr>
<td>23</td>
<td>14/935</td>
<td>0</td>
<td>8.11</td>
<td>0.84</td>
<td>1.48</td>
<td>0.99</td>
<td>2.20</td>
<td>0.00</td>
<td>0.79</td>
<td>12/790</td>
<td>8.11</td>
<td>0.84</td>
<td>1.48</td>
<td>0.99</td>
<td>1.30</td>
</tr>
<tr>
<td>24</td>
<td>3/1407</td>
<td>0</td>
<td>3.48</td>
<td>0.18</td>
<td>0.97</td>
<td>0.69</td>
<td>1.38</td>
<td>0.43</td>
<td>1.32</td>
<td>1/152</td>
<td>3.48</td>
<td>0.18</td>
<td>0.97</td>
<td>0.69</td>
<td>1.38</td>
</tr>
<tr>
<td>25</td>
<td>15/1386</td>
<td>0</td>
<td>28.55</td>
<td>0.01</td>
<td>0.65</td>
<td>0.44</td>
<td>0.56</td>
<td>0.95</td>
<td>0.43</td>
<td>13/1246</td>
<td>27.74</td>
<td>0.01</td>
<td>0.66</td>
<td>0.42</td>
<td>1.03</td>
</tr>
<tr>
<td>26</td>
<td>4/618</td>
<td>0</td>
<td>1.03</td>
<td>0.80</td>
<td>0.33</td>
<td>0.15</td>
<td>0.70</td>
<td>0.00</td>
<td>0.58</td>
<td>3/504</td>
<td>0.11</td>
<td>0.95</td>
<td>0.25</td>
<td>0.09</td>
<td>0.64</td>
</tr>
<tr>
<td>27</td>
<td>8/988</td>
<td>0</td>
<td>8.70</td>
<td>0.28</td>
<td>0.64</td>
<td>0.44</td>
<td>0.95</td>
<td>0.20</td>
<td>1.11</td>
<td>6/776</td>
<td>2.42</td>
<td>0.79</td>
<td>0.68</td>
<td>0.46</td>
<td>0.99</td>
</tr>
</tbody>
</table>

continued
TABLE 13  Language restriction, statistical heterogeneity and estimates of intervention effecta (cont’d)

<table>
<thead>
<tr>
<th>MA</th>
<th>n/pat</th>
<th>CAM?</th>
<th>Q</th>
<th>P-Homo</th>
<th>OR</th>
<th>Low</th>
<th>High</th>
<th>I²</th>
<th>H</th>
<th>n/pat</th>
<th>Q</th>
<th>P-Homo</th>
<th>OR</th>
<th>Low</th>
<th>High</th>
<th>I²</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>11/2934</td>
<td>0</td>
<td>18.75</td>
<td>0.04</td>
<td>0.73</td>
<td>0.58</td>
<td>0.92</td>
<td>0.47</td>
<td>1.37</td>
<td>9/1902</td>
<td>16.42</td>
<td>0.04</td>
<td>0.68</td>
<td>0.51</td>
<td>0.92</td>
<td>0.47</td>
<td>1.37</td>
</tr>
<tr>
<td>40</td>
<td>67/6850</td>
<td>0</td>
<td>78.20</td>
<td>0.15</td>
<td>0.67</td>
<td>0.57</td>
<td>0.79</td>
<td>0.16</td>
<td>1.09</td>
<td>65/6353</td>
<td>74.60</td>
<td>0.17</td>
<td>0.65</td>
<td>0.55</td>
<td>0.77</td>
<td>0.14</td>
<td>1.08</td>
</tr>
<tr>
<td>41</td>
<td>5/617</td>
<td>0</td>
<td>14.80</td>
<td>0.01</td>
<td>0.18</td>
<td>0.09</td>
<td>0.40</td>
<td>0.73</td>
<td>1.92</td>
<td>4/564</td>
<td>6.83</td>
<td>0.08</td>
<td>0.23</td>
<td>0.13</td>
<td>0.40</td>
<td>0.56</td>
<td>1.51</td>
</tr>
<tr>
<td>42</td>
<td>18/3099</td>
<td>0</td>
<td>22.42</td>
<td>0.17</td>
<td>0.39</td>
<td>0.32</td>
<td>0.48</td>
<td>0.24</td>
<td>1.15</td>
<td>16/2964</td>
<td>18.31</td>
<td>0.25</td>
<td>0.39</td>
<td>0.32</td>
<td>0.47</td>
<td>0.18</td>
<td>1.10</td>
</tr>
</tbody>
</table>

a The Q-statistic, test for heterogeneity (p-value), the intervention effect estimate (i.e. odds ratio and 95% CI, log-scale) were derived from the DerSimonian–Laird random effects model. The I² and H statistics measure the degree of statistical heterogeneity according to Higgins and Thompson. The section on the left displays statistics for language inclusive/LOE systematic review and the section on the right displays corresponding statistics for language restricted systematic reviews.

MA, Language inclusive/LOE systematic review number, which corresponds to the ID number and its related citation in Table 11 and Appendix 3; n/pat, number of RCTs/number of participants; CAM®, complementary and alternative medicine topic (0 = no, 1 = yes); Q, Cochran’s q test, P-homo, probability of homogeneity; OR, odds ratio, Low, lower end of the 95% confidence interval; High, higher end of 95% confidence interval; I², proportion of between-trial variation in the estimates of an intervention effect; H, statistical heterogeneity across RCTs.
heterogeneity in one CM and two CAM systematic reviews (#2, 4, and 7). In all three instances, the restriction eliminated a relatively large number of LOE trials. Otherwise, we did not observe any noticeable trend in the correlation between language of publication restrictions and statistical heterogeneity. Details of statistical heterogeneity and language of publication restriction are provided in Table 13.

For example, a systematic review evaluating the effect of acupuncture for smoking cessation included seven EL trials ($n = 480$) and five trials published in LOE ($n = 1467$). When all 12 trials were included, the degree of statistical heterogeneity was of borderline cause for concern ($I^2 = 0.51$) regarding the impact of the between-trial variation on the pooled estimate of acupuncture effect. When the analysis was limited to EL trial reports only, no significant statistical heterogeneity was observed ($I^2 = 0$). The effect of acupuncture on smoking cessation, however, was consistent despite language of publication restriction. Both language restricted and language inclusive/LOE systematic reviews reported non-significant effects on failure to quit smoking with acupuncture (OR = 0.92; 95% CI: 0.58 to 1.49 and OR = 0.79; 95% CI: 0.55 to 1.22, respectively).

### Language restriction and publication bias

A total of 31 language restricted systematic reviews had five or more trials. The median $Z$-score testing for funnel plot asymmetry from language inclusive/LOE systematic reviews was $-0.50$ [IQR (interquartile range): $-1.87$ to 0.93], the negative value indicating a large degree of asymmetry. The corresponding median from language restricted systematic reviews was $-0.22$ (IQR: $-1.84$ to 0.65). The slight decrease in the median $Z$-score for funnel plot asymmetry could be inspected using Figure 11. We did not observe any significant

![Figure 11](image-url)
## Results

**TABLE 14** Language restriction and publication bias

<table>
<thead>
<tr>
<th>ID</th>
<th>Restricted ME</th>
<th>Restricted PM</th>
<th>Inclusive/LOE ME</th>
<th>Inclusive/LOE PM</th>
<th>Rater I</th>
<th>Rater II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n/pats</td>
<td>p-value</td>
<td>Z-score</td>
<td>p-value</td>
<td>Z-score</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19/1046</td>
<td>0</td>
<td>-4.17</td>
<td>0</td>
<td>-6.14</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5/619</td>
<td>0.15</td>
<td>-1.92</td>
<td>0.05</td>
<td>-3.21</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>30/3009</td>
<td>0.04</td>
<td>-2.16</td>
<td>0</td>
<td>-3.51</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7/480</td>
<td>0.66</td>
<td>0.47</td>
<td>0.48</td>
<td>-0.77</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>17/7948</td>
<td>0.71</td>
<td>-0.38</td>
<td>0.26</td>
<td>-1.19</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5/560</td>
<td>0.66</td>
<td>-0.48</td>
<td>0.20</td>
<td>-1.66</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>20/1141</td>
<td>0.44</td>
<td>0.80</td>
<td>0.01</td>
<td>2.91</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>18/1100</td>
<td>0.56</td>
<td>0.60</td>
<td>0.04</td>
<td>2.24</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>5/1465</td>
<td>0.81</td>
<td>-0.26</td>
<td>0.13</td>
<td>2.04</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>8/1664</td>
<td>1.00</td>
<td>0.01</td>
<td>0.16</td>
<td>-1.66</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>14/1909</td>
<td>0.23</td>
<td>1.28</td>
<td>0.41</td>
<td>0.88</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>7/643</td>
<td>0.86</td>
<td>-0.18</td>
<td>0.43</td>
<td>-0.87</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>15/901</td>
<td>0.84</td>
<td>0.21</td>
<td>0.29</td>
<td>-1.10</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>21/3739</td>
<td>0</td>
<td>-3.96</td>
<td>0</td>
<td>-8.23</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>19/4563</td>
<td>0.05</td>
<td>-2.10</td>
<td>0.03</td>
<td>-2.36</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>17/3859</td>
<td>0.92</td>
<td>-0.11</td>
<td>0.92</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>7/699</td>
<td>0.95</td>
<td>-0.07</td>
<td>0.96</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>8/714</td>
<td>0.14</td>
<td>1.73</td>
<td>0</td>
<td>4.54</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>7/1384</td>
<td>0.41</td>
<td>-0.90</td>
<td>0.36</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>5/398</td>
<td>0.32</td>
<td>1.20</td>
<td>0.24</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>8/1734</td>
<td>0.17</td>
<td>1.56</td>
<td>0.73</td>
<td>-0.36</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>9/311</td>
<td>0.28</td>
<td>-1.17</td>
<td>0.52</td>
<td>-0.68</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>23/3287</td>
<td>0.02</td>
<td>-2.66</td>
<td>0.01</td>
<td>-2.87</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>44/16356</td>
<td>0.08</td>
<td>-1.82</td>
<td>0.02</td>
<td>-6.02</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>12/790</td>
<td>0.06</td>
<td>2.16</td>
<td>0.06</td>
<td>2.17</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>40/11593</td>
<td>0.07</td>
<td>-1.88</td>
<td>0.07</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>13/1246</td>
<td>0.01</td>
<td>-3.43</td>
<td>0.01</td>
<td>-3.46</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>6/776</td>
<td>0.68</td>
<td>-0.44</td>
<td>0.33</td>
<td>-1.11</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>9/1902</td>
<td>0.84</td>
<td>0.21</td>
<td>0.27</td>
<td>-1.21</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>65/6353</td>
<td>0.43</td>
<td>0.80</td>
<td>0</td>
<td>-3.04</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>16/2964</td>
<td>0.48</td>
<td>-0.72</td>
<td>0</td>
<td>-5.56</td>
<td></td>
</tr>
</tbody>
</table>

*All estimates of treatment effect were derived using the DerSimonian–Laird random effects model.43 The p-value was derived from a t-test for zero intercept of the regression line. The Z-score was the ratio between the estimate coefficient of the regression line and its standard error. A negative Z-score indicates small trials with large estimate of intervention effect. The ID refers to the systematic review number (see Table 11 and Appendix 3). Rx, type of intervention; n/pats, number of RCTs/no of participants; ME, test for funnel plot asymmetry using the method by Egger and colleagues; PM, test for funnel plot asymmetry using the method by Macaskill and colleagues; Rater I and Rater II, visual inspection of funnel plot asymmetry by two assessors.
increase in funnel plot asymmetry associated with language inclusive/LOE systematic reviews compared with language restricted reviews (Figure 11). In the quadrant of Figure 11 in which funnel plot asymmetry is likely to be related to publication bias (i.e. with a Z-score <0), restricting the systematic reviews to reports of EL trials might reduce funnel plot asymmetry in about three systematic reviews. Table 14 displays the results of two regression approaches for detecting funnel plot asymmetry for both language inclusive/LOE and restricted systematic reviews. Table 14 also displays the results from the visual inspection for asymmetry of individual funnel plots illustrated in Figure 12.

Nine (of 31) systematic reviews were with funnel plot asymmetry as assessed by one or both assessors. The visual inspection was completed for language inclusive/LOE systematic reviews (Figure 12). There were three systematic reviews in which both assessors agreed (numbers 2, 18 and 36; Figure 12 and Table 14). The agreement between the two visual inspections was mild, with a kappa of 0.38 (95% CI: –0.01 to 0.78).

Significant funnel plot asymmetry was detected in 18 (of 31) language inclusive/LOE systematic reviews. The corresponding number of language restricted systematic reviews was 15 (of 31). Language inclusive/LOE systematic reviews were not associated with an increase in funnel plot asymmetry (i.e. McNemar test, \( p = 0.25 \)). Three language inclusive/LOE systematic reviews (one CAM and two CM) were with significant funnel plot asymmetry. The test for heterogeneity became non-significant when the three systematic reviews were language restricted (i.e. plots numbered 7, 27 and 39 in Figure 12; Table 14). The first systematic review (i.e. plot number 7 in Figure 12) compared the effect of acupuncture with sham acupuncture. Its findings were mentioned previously. We discuss the other two systematic reviews below.

Assessing the effect of glycerol treatment compared with a control group in the acute period after stroke, a language inclusive/LOE systematic review of six studies reported a 41% protective effect (14 days post-stroke mortality OR 0.59; 95% CI: 0.36 to 0.98; number 27, Figure 12). The language-restricted systematic review reported an intervention effect of 49% (OR 0.51; 95% CI: 0.30 to 0.85).

A systematic review evaluating the effects of support for breastfeeding mothers included nine EL trials and two reports of trials published in LOE (number 39, Figure 12). The language inclusive/LOE systematic review reported a 27% favourable effect for support services (OR for early stopping of breastfeeding 0.73; 95% CI: 0.58 to 0.92). The language-restricted systematic review reported a corresponding effect of 32% (OR 0.68; 95% CI: 0.51 to 0.92).

Given the total number of trials included in the language inclusive/LOE systematic reviews, the number of LOE trials was relatively large for systematic reviews of CAM interventions. It was, however, relatively small for systematic reviews of CM interventions (Table 13). For example, a systematic review evaluating the effect of St John’s wort for depression included 10 German and three EL RCTs. There was substantial funnel plot asymmetry in the systematic review (\( p = 0.01 \)); this was also the case when the data was restricted to German trials (\( p = 0.02 \)). The proportion of non-response to St John’s Wort treatment seemed to be reduced substantially according to both the 10 German trials (OR 0.17; 95% CI: 0.07 to 0.40) and the three EL (OR 0.19; 95% CI: 0.08 to 0.42).

**Summary**

In this study, the association between language of publication restriction and statistical heterogeneity should be considered separately for systematic reviews of CM and CAM interventions. The inclusion of LOE trials in the 34 CM systematic reviews resulted in an absolute increase of 2.4% in the percentage of between-trial variation expressed as a portion of the total between- and within-trial variation. This increase did not produce any meaningful impact in the estimates of intervention effect across the 34 CM systematic review.

Issues related to clinical heterogeneity become the main considerations for the inclusion of LOE trials in CAM systematic reviews. Often, such inclusion increases the number of included trials substantially (median 37%, range 17–77%). As such, careful evaluation of potential variation in patient population, intervention and outcome ascertainment among LOE and EL trials is required.
Results

FIGURE 12 Funnel plots of the 42 language inclusive/LOE systematic reviews. Treatment effect estimates are plotted proportionately to quality using a circle symbol (EL trial), square (LOE trial), * (EL trial with unavailable quality assessment) and + (LOE trial with unavailable quality assessment). The systematic review number (MA #) corresponds to ID number and its related citation in Tables 11 and 13. A negative logarithm OR indicates a protective intervention (i.e. preventing an undesirable outcome, relative to control).
FIGURE 12 (continued)
FIGURE 12 (continued)
FIGURE 12 (continued)
Chapter 4
Discussion

Examining our sample of 130 recently published systematic reviews revealed some noteworthy points regarding language inclusive/LOE reviews. These reviews included a larger number of RCTs and participants compared with either the language restricted or the language inclusive/EL systematic reviews. In an average language inclusive/LOE systematic review included in our sample, 14% of the reports were in LOE.

To investigate the effects of excluding reports of RCTs in LOE, we identified systematic reviews that included LOE in the quantitative data, and then repeated the systematic review in two ways: first, we simply replicated the data synthesis, and second, we repeated the data synthesis for each language inclusive/LOE systematic review by a pooled OR being computed separately for LOE and EL trials. The ROR of the two OR estimates was then computed. Both approaches provide similar results.

In our earlier research, the results were limited in that we could only identify and include 18 language inclusive/LOE systematic reviews, including 211 reports of RCTs. Included in that sample were 30 reports of RCTs in LOE. Nevertheless, the 2% observed difference between language restricted and language inclusive/LOE systematic reviews, along with the narrow CIs and several sensitivity analyses gave some comfort to the findings. Recognising that this result was based on a single study, we felt a need to repeat and expand on it. Such replication is a cornerstone of the scientific method.

In the present study, we set out to increase these numbers and to include systematic reviews that examined CAM interventions. The current sample of 130 systematic reviews included 48 language inclusive/LOE systematic reviews, including 700 reports of RCTs. Included in that sample were 115 reports of RCTs in LOE. Although these numbers are a substantial increase compared with some previous research, the distribution of reports of RCTs in LOE was disappointing; those reported in German and French dominated the included RCTs in LOE. Although we completed extensive searches in both CM databases and the more specialised CAM databases, and contacted researchers who had published systematic reviews in CAM, it is possible that we did not identify all of the relevant systematic reviews. Juni and colleagues, in a similar investigation, reported that 62% of their 115 LOE reports were in German or French.

Although reports of RCTs appear in many languages, it is possible that they cover a very broad range of topics and interventions and that there are insufficient numbers in any one area, making them appear scarce in any particular systematic review. Alternatively, it is possible that many trial reports in CAM appear in English language only. One issue (explored above) is whether systematic reviews in CAM are more likely to be language restricted compared with systematic reviews in CM.

One way to increase the pool of language inclusive/LOE systematic reviews, and thus reports of RCTs in LOE, would be to identify language restricted systematic reviews and subsequently identify reports of RCTs in LOE that meet the eligibility criteria of the systematic review in question, and could therefore be included in the systematic review. A similar approach has been tried before with disappointing results. Grégoire and colleagues identified 36 language restricted systematic reviews and were able to identify four additional reports of RCTs in LOE, whereby four language inclusive/LOE systematic reviews were further examined. However, on closer inspection only two (of the four) systematic reviews met the criteria of including RCTs. Therefore, extrapolating this approach to our group of 50 language restricted systematic reviews might have meant the addition of only 2.5 language inclusive/LOE systematic reviews.

We observed that systematic reviews evaluating a CAM intervention included more RCTs compared with systematic reviews evaluating a CM intervention. However, even though they contained more trials, they included far fewer participants. Therefore, systematic reviews in CAM include smaller RCTs compared with their CM counterparts. If smaller RCTs are more prone to bias, as has been suggested, our findings have important implications for those conducting systematic reviews. For example, we might expect...
smaller CAM RCTs to be of lower quality, such as a
greater proportion of reports with unclear
allocation concealment. Similarly, this bias might
influence the results of systematic reviews.

**Quality of reporting of systematic reviews**

The early reports on the quality of reporting for
systematic reviews indicated that many reviews had
serious flaws. More recently, Jadad and
colleagues\(^5\) reported on the quality of reporting
in 50 systematic reviews (38 paper-based and 12
Cochrane Reviews) that examined the treatment of
asthma. Of these reviews, 58\% were published in
1997 or 1998. The authors found that 80\% had
serious or extensive flaws. However, they reported
that the Cochrane Reviews were more rigorous
and better reported than the paper-based
publications. In contrast, Choi and colleagues\(^5\)
found only minor or minimal flaws in the quality
of reporting in nearly half of 82 systematic reviews
of perioperative medicine. This study suggests
that there may be an association between quality
of reporting and the content area of the systematic
review. However, neither of these studies examined
the effect of language of publication or CAM
interventions.

In the current study, we aimed to examine both
the effect of language restriction on quality of
reporting and reporting characteristics of systematic
reviews and whether there are differences between
quality of reporting for CM interventions and
CAM interventions. We employed a much larger
sample size and sampling frame than the previous
studies, with particular emphasis on systematic
reviews involving CAM. To do so, we searched
EMBASE and a CAM Database and combined this
with the Medline and Cochrane database
identified in a previous study,\(^3\) thus increasing
our sample size to 130 and more than doubling
the number of language inclusive/LOE systematic
reviews found previously (from 18 to 48). In
addition, 25\% of the systematic reviews identified
in our current study involved CAM interventions.

The results of our examination of the effect of
language restriction on quality of reporting for
systematic reviews are consistent with Moher and
colleagues’ previous findings.\(^3\) Our data indicated
that the overall quality of reporting of language
inclusive/LOE reviews is only slightly better than
that of language restricted or language
inclusive/EL reviews. Authors who do not impose
language restrictions appear to be more thorough
and comprehensive in their search for and
inclusion of potentially relevant literature. It is not
surprising, therefore, that the language
inclusive/LOE systematic reviews identified and
included a larger number of trials and more grey
literature than the language inclusive/EL and
language restricted reviews.

Unlike the previous studies examining quality of
systematic reviews,\(^1,11–13\) a significant proportion of
our reviews (25\%) involved CAM interventions. We
found the CAM systematic reviews to be of higher
quality than those that focused on CM
interventions. The CAM systematic reviews used
more comprehensive search strategies than the
CM systematic reviews. The CAM systematic
reviewers may search more broadly because it may
be more difficult to identify CAM RCTs in certain
journals. For example, Moher and colleagues
found that trials published in English are likely to
report positive findings of CM interventions but
significantly less likely to do so for CAM
interventions.\(^5\) The CAM systematic reviews were
also more likely to report the criteria used to select
studies for inclusion, in addition to the criteria
used for assessing the validity of the studies that
they included. Furthermore, the validity criteria
were more likely to be appropriate than those
reported by the CM systematic reviews. In
addition, the CM systematic reviews were more
likely to report positive or unclear conclusions
compared with the CAM systematic reviews, and to
have their conclusions supported by their data
and/or analysis.

Ours is the first study to compare systematically
the quality of reporting of CAM systematic reviews
with those involving CM interventions. Jadad and
McQuay\(^1\) reviewed the quality of reporting of 74
systematic reviews of pain interventions, including
some CAM interventions. Only three of these
reports were in LOE (Italian, Portuguese and
German). Two of the 74 reports were specifically
described as involving a complementary
intervention. However, the 74 reports included 52
unclassified psychological interventions, 10
physical interventions (four manipulation studies)
and two acupuncture studies whose classification
was unclear. Two-thirds of the reports were
published after 1990.

Jadad and McQuay\(^1\) reported that 90\% of these
systematic reviews had methodological flaws that
could limit their validity, with the main
deficiencies being insufficient information on their
search strategy, validity assessment of the included
studies and design of the primary studies. Only
10% of the reviews satisfied all of the OG criteria: 16% were given the lowest possible score; 26% did not describe their search strategy; more than 60% failed to describe the methods used, if any, for validity assessment of the included studies; and 21% did not even describe the design features of the primary studies. The two CAM systematic reviews identified by these authors received quality ratings above the median, as did both acupuncture and two of the four manipulation systematic reviews. There were also two systematic reviews assessing laser therapy for musculoskeletal pain, and both were given a quality rating of six out of seven. Of concern, these authors found that systematic reviews of low quality (based on the OG scale) produced significantly more positive conclusions than those rated as being of high quality (total score over four out of seven). The authors did not comment on the quality assessment of the three LOE reports.

The OG scale was developed and validated in 1991 and has been used extensively since then to assess the quality of reporting of systematic reviews. Its brief nature and ease of administration facilitate its use. However, since this scale’s development, there has been considerable research examining the methodology of systematic reviews, such as the inclusion of grey literature, effect of language restriction, types of search strategies and publication bias. We questioned whether the OG scale still provides a valid assessment of quality in the reporting of systematic reviews. To examine this, we informally compared the quality assessment based on the OG scale with the data on quality of reporting collected through our Data Collection Forms. Overall, the OG scale performed well, providing overall quality ratings that were consistent with the general impressions from the information collected through our Data Collection Forms. The only exception to this was the OG rating for comprehensiveness of the search strategy; the OG scale did not identify any significant differences based on language restriction, whereas our Data Collection Forms indicated that the language inclusive/LOE reviews were much more comprehensive in their search strategies. In all other respects, the quality assessments based on the OG scale and Data Collection Forms were consistent. However, as expected, the Data Collection Forms provided significantly more information on specific aspects of quality of reporting (e.g. inclusion of grey literature, assessment of publication bias, methods for avoiding bias in study selection). If these specific components have sufficient impact on the overall quality of a systematic review, it may be justified to incorporate them into an assessment instrument such as the OG scale. Further research in this area is warranted.

The use of CAM interventions by the general population has increased dramatically over the last decade. This has been paralleled, albeit more slowly, by an increase in the number of trials evaluating CAM interventions. Many of these trials are small and/or have conflicting results, thus uncertainty remains about the effectiveness of many CAM interventions. It is under these circumstances that systematic reviews are most useful in quantitatively summarising the available evidence. It is surprising, therefore, that we are only able to identify 25 systematic reviews of CAM interventions that met our eligibility criteria. That said, these numbers are in keeping with what has been reported in the literature recently. Our findings regarding the quality and reporting of CAM systematic reviews are limited by the number of CAM reviews that we were able to evaluate. However, this report provides data on the largest and most systematic comparison of CAM and CM systematic reviews to date. As a result, our study presents new information that demonstrates that CAM systematic reviews have quality of reporting similar to or better than those of CM interventions. This suggests that consumers and health professionals evaluating CAM interventions can be reassured that CAM systematic reviews are a reliable source of information.

Our results also indicate that the quality of reporting and reporting characteristics of systematic reviews are not affected by the inclusion or exclusion of LOE. However, a systematic review’s conclusions and estimate of effectiveness may be influenced by language restrictions, if there is an association between the quality of the included trials and the language of publication.

Quality of reporting of RCTs

Systematic reviews must include all the available and relevant evidence pertaining to the question at hand. Exclusion of certain studies is likely to introduce systematic error or biases that threaten the validity of the findings. There is evidence that most systematic reviews do not include all the available evidence, with language of publication being the most common reason for exclusion. Whether this is justified depends on the quality of the individual studies being excluded; it would seem reasonable to exclude studies published in
LOE if their quality of reporting was inferior to that of EL trials. Several authors have previously examined this question.

Moher and colleagues compared the quality of reporting of 133 RCTs published in English with 96 trials published in French, German, Italian or Spanish during the same time period (1989–96). They did not find any significant differences in the quality of report (using the Jadad scale) between the EL and LOE language trials. The present results are in agreement with these findings, namely that the overall quality of reports published in LOE is similar to that of EL reports of RCTs. These results are strengthened by the much larger size of our database, which included 498 EL trials and 114 trials published in LOE. In addition, our study included four additional non-English languages: Danish, Dutch, Japanese and Portuguese, although French, German and Italian continued to predominate.

When we examined the effect of language of publication and type of intervention, we found similar results in the 484 trials of CM interventions. As previously recommended by Moher and colleagues, our findings support the inclusion of all trials, regardless of language of publication, in systematic reviews of CM interventions. However, do these conclusions apply to trials of CAM interventions?

Other authors have questioned the quality of CAM trials, regardless of language of publication. Linde and colleagues reviewed the quality of 207 RCTs published in five systematic reviews on homeopathy, herbal medicine and acupuncture. They found significant methodological problems in the majority of the trials with overall mean Jadad scores of 2.2–3.2. Most trials did not adequately describe the randomisation procedure, allocation concealment or the number and reasons for withdrawals and dropouts. In addition, they found that larger trials published in MEDLINE-indexed journals and in English were of higher quality. The 61 trials published in English had a mean Jadad score of 2.88 compared with 2.36 for the 59 LOE trials (p = 0.027). Furthermore, there was a significant association between language of publication and adequate allocation concealment (OR 4.54, 95% CI: 1.47 to 14.02). However, Linde and colleagues’ study only included trials of CAM interventions.

Jüni and colleagues reported that LOE reports were of lower quality (i.e. lower frequency of adequate allocation concealment) than EL reports. Although our results differ, there are more similarities when our results were stratified by type of intervention: CM or CAM. LOE CM reports were of lower quality (i.e. lower frequency of adequate allocation concealment) than CAM LOE reports. We found no differences in quality of reporting of LOE and EL CAM trials.

Our results indicate that the overall quality of reporting of CAM RCTs is as good as or better than that for CM interventions. In contrast, the quality of EL reports of CM interventions was higher than that of the LOE CAM reports and also both the EL and LOE CM reports. This suggests that a higher level of quality for CAM RCTs is required for these trials to be published in EL journals, raising the possibility of selective bias by editors against CAM trials. Alternatively, it may be that authors of CAM trials submit their highest quality work to EL journals. Egger and colleagues have previously shown that German authors of CM trials preferentially publish their positive RCTs in EL journals.

Previous authors have shown that exclusion of LOE reports from systematic reviews of CM interventions does not affect the estimates of the interventions’ effectiveness. This is not surprising given the similar level of quality of reporting that we and others have found for EL and LOE CM trials. However, our results from the current study suggest that there are differences in the quality of reporting between EL and LOE CAM reports. These results might also help explain the bias found in CAM language restricted systematic reviews (i.e. exclusion LOE reports).

‘Language of publication’ bias and location bias are related to the type of intervention (CM or CAM)

These results, from a dataset of 42 language inclusive/LOE systematic reviews including 662 RCTs, provide new evidence concerning what effects language of publication restrictions can have on the results of systematic reviews. As indicated in our methods, we set out to include 45 language restricted systematic reviews and 45 language inclusive reviews. This sample size enabled us to observe a 25% difference in the ROR if it existed. We met this target by including 50 language restricted systematic reviews and 48 language inclusive/LOE reviews.
Our analyses suggest that language of publication restrictions, in the quantitative data synthesis of a systematic review, depend on the type of intervention under investigation. Systematic reviews focused on evaluating the effectiveness of CAM interventions appear to produce different results when the quantitative data synthesis is limited to trials reported in English only. Specifically, the present analysis suggests that limiting a CAM systematic review to EL reports will produce exaggerated estimates of effectiveness of 63%, on average. This analysis suggests that individual trial results in LOE are important and need to be included, along with EL results, in any CAM systematic review.

In contrast, the present results suggest that limiting the language of publication of trial reports to English does not look as if it will result in any measurable effect on the estimates of an intervention’s effectiveness, when the intervention under investigation is CM. These results are very similar to those reported by us previously, although the present results are based on a dataset approximately three times the size of that used in the earlier work. Juni and colleagues, using a very similar-sized dataset, reported similar results.

Moher and colleagues identified 18 systematic reviews that included 178 EL trial reports and 33 publications in seven LOE. Excluding these reports, compared with including them, resulted in a 2% shift of the estimates of the intervention effect at the RCT level. Given the low prevalence for trials published in LOE, it is likely that this very small non-significant trial effect will be even smaller at the level of an individual systematic review. The present results, from a database of 42 systematic reviews including 129 reports of RCTs in eight LOE, found very similar results. The potential shift in estimating the effectiveness of an intervention’s effectiveness was only 2%, on average.

The present results, and those previously reported by us, remain unchanged whether one or more report in LOE is included in the systematic review. Taken together, they strongly suggest that excluding reports of RCTs in LOE from the meta-analytical part of a systematic review is a reasonable way to conduct a review. This recommendation only applies to reviews investigating the benefits of CM interventions. This does not imply that systematic reviewers should neglect reports in LOE. We recommend that systematic reviewers search for reports regardless of the language of their publication. There may be merit in including them in some aspects of the review process although this decision is likely to depend on several factors, including fiscal and other resources being available.

Egger and colleagues examined a similar ‘language’ issue, albeit they asked a different question. Here the investigators identified ‘pairs’ of reports from investigators who published in both German and English, matching for first author and time of publication, with one report appearing in English while the other report was reported in German. Of the 40 pairs of reports analysed, those published in English, compared with German, were nearly four times more likely to be statistically positive. This result might reflect German author submission bias with authors concluding that statistically positive results are of more appeal to the wider ‘international’ English-reading audience. Alternatively, the results might reflect a journal bias to publishing statistically positive results.

Our findings could be interpreted as being the result of a different set of biases. The exaggeration due to CAM interventions might reflect author and/or journal bias. It is possible that authors reporting statistically positive trial results would elect to submit their findings to CAM journals appealing to their practitioner audience. It is also possible that some journals, particularly those with higher citation impact factors, are less likely to publish results of CAM trials with statistically positive results. This observation has recently been reported.

Pittler and colleagues set out to investigate whether there was a relationship between the statistical directions of CAM RCT results and where they were likely to be published. These authors classified 352 controlled trials, included in 19 systematic reviews from the CDSR, as either positive or negative. They also categorised journals as either mainstream medical or CAM, and each journal’s citation impact factor (i.e. no impact factor, impact factor <1 and impact factor ≥ 1). Among their findings, the authors observed that mainstream medical journals with an impact factor ≥ 1 published an equal number of CAM trials with positive and negative results. This finding is in contrast with more CAM positive findings in similar journals with a lower citation impact factor.

Journals may need to take a more explicit stance with regard to their view of publishing CAM RCTs. This is already starting to happen. Recently, the
British Medical Journal and the Journal of the American Medical Association have published CAM theme issues. Likewise, authors will need to review their preconceived notions regarding journals’ views of CAM.

Despite the low LOE prevalence, we observed a substantial language of publication bias effect for CAM reports. However, this is a ‘trial’-level result owing to the logistic approach that we used. It is possible that because of the low prevalence, the effects of this bias will be diluted at the level of the systematic review.

Sampson and colleagues\(^4\) examined the potential for biasing the results of a systematic review, if the search to identify relevant RCTs excluded the EMBASE electronic database. These authors observed a large effect of 29% when EMBASE-unique trials were excluded from the data synthesis compared with when these trials were included. However, this trial level result was less pronounced at the systematic review level (i.e. 6%), due in part to the low prevalence of EMBASE unique trials and the data synthesis approach proposed by Sampson and colleagues\(^4\) and Sterne and colleagues.\(^4\)

The results reported here have implications for those interpreting reports of systematic reviews. There is considerable room for improvement in how systematic reviews are reported. As such, close scrutiny of systematic reviews using critical appraisal skills is required. We encourage readers to use one of several available tools, such as the OG index and the quality of reporting of meta-analysis (QUOROM) checklist, alongside the review of any systematic review. Interestingly, closer inspection of CM reviews, compared with CAM reviews, may be warranted because their quality (of reporting) was lower in our sample.

Systematic reviewers who have examined the benefits of CM interventions and have limited their meta-analytical analysis to only reports of randomised trials in English are not likely to have introduced bias in the reported pooled estimate. That said, there may be merit in knowing what trials were excluded. As such, we recommend caution when interpreting systematic reviews that have completely neglected reports in LOE.

Systematic reviewers should report on whether they used a language filter to locate reports in any language. Likewise, reviewers should report on whether any such reports were identified, retrieved, translated and included in any part of the systematic review process, such as quality assessment.

Of course, it is possible that the systematic reviewers searched to identify trials reported in any language but for whatever reason were unable to identify any. Readers will have to judge the ‘face validity’ of such reports in the context of a broader picture.

For reviews that report on the benefits of CAM interventions, readers should be very cautious in how they are interpreted for all the reasons discussed above. In addition, readers should interpret with considerable suspicion the purported benefits of any CAM intervention if the systematic reviewers have used any language restrictions in the meta-analytical part of their review. Such exclusions are likely to introduce bias into the review process and exaggerate the reported pooled estimates of benefit.

**Limitations**

This data analysis included 42 systematic reviews including 662 RCTs. Of these, there were eight systematic reviews of CAM interventions including 129 RCTs. Despite our best intentions and comprehensive search to identify systematic reviews meeting our eligibility criteria, we were not able to increase the number of CAM systematic reviews, and therefore RCTs, in our evaluations of CAM findings. Nevertheless, these results appear to be robust and generalisable. The findings of no bias for CM are in keeping with those already reported in the literature, although the database upon which these results are based is considerably larger than any used in previous studies. The CM and CAM findings also appear to be robust, based on the reported sensitivity analysis.

The present analysis is based on eight LOE. Only one of these languages was Asian, where CAM has had a long and respected history. German and French reports accounted for the majority of LOE.

Our data analyses provide results that are limited to the influence at the level of the trial. To what extent they apply at the level of systematic review is unknown. It is not likely to be a problem for CM interventions because no bias effect was observed. However, the substantive CAM effect may be smaller at the level of the systematic review.

Despite the potential limitation that this data analysis approach has, it is in line with current thinking as to how to address these methodological issues within systematic reviews. The new approach will have to be further assessed before it can be used more widely.
In completing our analysis, we categorised interventions as either CM or CAM. This is easier to do for certain interventions. For example, the use of antibiotics for the management of children with acute otitis media is clearly classifiable as a CM intervention. Likewise, the use of chiropractic spinal manipulation for reducing symptoms of neonates with colic is clearly outside the realm of what conventional healthcare providers would suggest or recommend for managing neonates with colic. However, such an intervention falls well within that offered by CAM practitioners.

Between these 'extreme' examples of CM and CAM interventions lies a grey zone making classification of interventions more difficult and a moving target. Traditional healthcare practitioners may well offer mind–body interventions, such as cognitive therapy, as part of the armamentarium of interventions available for a wide variety of problems, such as anxiety disorders. In categorising interventions as either CM or CAM we relied on a typology proposed by the Cochrane Collaboration. As such, CM interventions included surgical and/or pharmaceutical products. Other interventions fell under the CAM umbrella. It is possible that using a different set of definitions would lead to a different series of results other than those which we observed. We are confident that the observed results are robust for at least two reasons. Our classification, although perhaps conservative, is in keeping with other typologies within the literature. The strong effect that we observed could be reduced by using a different classification system, but it is very unlikely that the observed 63% effect would disappear completely.

Taken together, these results suggest that for systematic reviewers evaluating the merits of CM interventions, it is reasonable to limit their quantitative data synthesis to English language reports of RCTs.

However, when the evaluation is focused on assessment of CAM interventions, it is important to include trials published in any language in the quantitative data synthesis. Limiting the data analysis to EL reports is likely to result in a substantial bias in the result.

For the larger reading audience of healthcare providers, consumers and health policy analysts, caution is required during critical appraisal and interpretation of a systematic review, depending upon the type of intervention being considered. Decision-making based on the results of CM systematic reviews limited to reports of EL RCTs is reasonable. It is unlikely that the results obtained from such a systematic review will be biased. At most, there may be a slight decrease in the precision of the estimate of the intervention's effectiveness.

If decision-making is required from a systematic review of a CAM intervention, it is essential that the systematic reviewers have included reports of RCTs in all languages as part of their quantitative data synthesis. If not, the decision-making is likely to result in a biased estimate of the intervention’s effectiveness, something to be avoided whenever possible.

Regardless of the type of intervention, caution in the interpretation of the results is needed if the systematic reviewers have not searched to identify reports in all languages.

The data analysis in this research has focused on the synthesis of quantitative information from RCTs in English compared with those in other languages. These results have little bearing on the efforts required to identify and retrieve potentially relevant articles. Systematic reviewers and those needing critically to appraise and/or make decisions based, in part, on the results of systematic reviews will need to develop search strategies without language of publication restrictions or location bias restrictions.

Impact of language restriction on between-study heterogeneity and publication bias

In this study, the inclusion of LOE trials in the 42 systematic reviews resulted in an ~12% increase in the proportion of systematic reviews with significant statistical heterogeneity. This increase, however, did not result in any significant changes in the estimates of intervention effect. In addition, we did not observe any consistent patterns of increasing statistical heterogeneity in the intervention effect estimates. Our results suggest that the increase in generalisability of the review findings from language inclusive/LOE systematic reviews is associated with only a small increase in statistical heterogeneity.
We noted a difference in the number of LOE trials contributing to language inclusive/LOE systematic reviews evaluating interventions in CM versus CAM. With respect to CAM interventions, the number of LOE trials is fairly large relative to the total number of included trials, on average. With respect to CM interventions, however, the number of LOE trials is relatively small. The decision regarding the inclusion of LOE trials entails different considerations for systematic reviews of CM and CAM interventions.

The likelihood of publication bias associated with the inclusion of LOE trials in language inclusive/LOE systematic reviews is one of these considerations. Compared with language restricted systematic reviews, our results did not indicate any significant change in the likelihood of publication bias associated with language inclusive/LOE reviews. However, ~30% of language inclusive/LOE systematic reviews in our sample displayed significant asymmetry in their funnel plots. In ~10% of these cases, the significant funnel plot asymmetry was reduced substantially with language restriction.

We used a combination of approaches for the assessment of publication or related biases in systematic reviews. These included visual inspection of, and two regression approaches to, funnel plots. Although none of these approaches perform consistently well, their limitations are relatively well studied. The observation that funnel plot asymmetry is common, as reported above, has been similarly concluded elsewhere using a different method to assess for publication and related biases.

Asymmetry in a funnel plot can be explained by a number of sources, most notably genuine trial differences, and publication and related biases, among others. The inclusion of LOE trials in language inclusive/LOE systematic reviews may result in heterogeneity in the patient populations, interventions and outcome measures pertaining to the clinical question under investigation. As such, true heterogeneity is one likely cause of the funnel plot asymmetry reported here.

Publication or related biases are also possible, especially with language inclusive/LOE systematic reviews evaluating CAM interventions. In our sample, the LOE trials that were included in these systematic reviews were dominantly published in German and French. These were trials with a relatively high quality of reporting. On average, they appeared to report larger intervention effect estimates compared with trials reporting in English and assessing the same interventions. Their inclusion led to overestimation of the benefits of CAM interventions. This is not congruent to at least one other study examining the language of publication issue. With respect to CM, it was reported that German authors often published trials conducted in German-speaking settings with statistically significant findings in EL journals. Further investigations may be needed here to sort out the direction of these potential selection biases.

In summary, we have examined two issues that are increasingly considered as part of the regular conduct of a systematic review. We observed no significant association between statistical heterogeneity and the inclusion of LOE trials in language inclusive/LOE systematic reviews. This is also true for the association between publication and related biases and language inclusion. As such, the results reported in the previous chapters were not affected by either statistical heterogeneity or publication or related biases.

**Conclusions**

Using a database of 130 recently published systematic reviews with varying degrees of language restrictions, we observed that the quality of reporting of systematic reviews is low, with room for considerable improvement. The median quality of reporting was in the range of 48% of the maximum possible score for a systematic review. Language inclusive/LOE systematic reviews were of higher quality than language restricted reviews. Language inclusive/LOE reviews searched more comprehensively than language restricted reviews to identify relevant literature. For example, language inclusive/LOE systematic reviews were more likely than language restricted reviews to include grey literature in the review process. Language inclusive/LOE systematic reviews included trials published in eight LOE but were dominated by those reported in German and French. Language inclusive/LOE reviews included more reports of RCTs and more participants than language restricted systematic reviews. CAM systematic reviews were of higher quality than CM reviews.

Using the Jadad quality assessment tool, we observed that the quality of reporting of randomisation, and dropouts and withdrawals, was significantly higher in EL RCTs than LOE RCTs. These results appear to be driven by the type of
intervention under consideration. The quality of reporting of CM RCTs is similar whether they are reported in English or another language. However, the quality of reporting of randomisation, double blinding and dropouts and withdrawals was higher for EL CAM RCTs than LOE CAM trials. This finding has been observed elsewhere.65 There are several possible explanations for this finding. The most obvious one is that the higher quality CAM trials are published in English. Alternatively, editors of EL publications might have a differential quality threshold for accepting reports of CAM trials. Whatever the reasons for this finding, we believe that the relationship between CAM and quality requires further investigation.

Language restrictions do not bias the results of CM systematic reviews but substantially bias the results of CAM systematic reviews. These results are robust even after sensitivity analyses, and do not appear to be influenced by statistical heterogeneity and publication bias.

Our group is experienced in assembling large datasets of systematic reviews. For example, we identified 73 systematic reviews to examine the influence of excluding grey literature from systematic reviews. The development and assembly of a large database of systematic reviews is a labour-intensive and costly exercise. We experienced all of these problems and more when assembling the 130 systematic reviews included in the present study. For example, despite considerable efforts, we were only able to include 25 CAM systematic reviews that met our eligibility criteria. Although we think this is a reasonable number of CAM reviews to address our objectives, French and German predominated as the languages of publication of the RCTs. We were not able to identify a large number of trials in other languages, such as Mandarin, that were included in a systematic review.

One option to avoid this effort could be to use an established database in which systematic reviews are already collected. The CDSR is an excellent example, and several groups have used it for methodological research. There are approximately 1500 completed systematic reviews in the CDSR which include all the relevant data required for meta-epidemiological investigations such as ours. Although the CDSR may well have good internal validity, there are questions regarding its generalisability. There is now growing evidence that the quality of reporting differs between Cochrane systematic reviews and paper-based reviews. Jadad and colleagues66 observed that in 36 Cochrane Reviews, compared with 39 paper-based ones, Cochrane Reviews were more likely to include a description of their inclusion and exclusion criteria and to assess trial quality. Similar data can be found elsewhere and we recognise that these findings are not universal. For example, Shea and colleagues17 assessed the quality of 50 Cochrane Reviews and compared their quality with that of 49 paper-based reviews; they found no difference in the quality of reporting when the reviews were evaluated using the OG index. As such, it is unclear to what extent Cochrane Reviews reflect how systematic reviewers handle specific methodological issues in the conduct of a systematic review. The findings of Cochrane Reviews may not be generalisable to paper-based reviews. The latter are published with substantially greater frequency.

One option would be to fund the development and maintenance of a large database of systematic reviews (Cochrane and paper-based). This database could be shared with the growing number of investigators conducting meta-epidemiological research. In the short term, this move is likely to reduce the cost of conducting important methodological research. This move is also likely to make such investigations more time efficient.

Although the sample of systematic reviews included in our analysis were recently published (the median year of publication was 1994 and 1995 depending on the type of review), the quality of reporting was less than optimum. However, two exceptions were noted: language inclusive/LOE reviews were of higher quality than language restricted reviews; and CAM systematic reviews were of significantly higher quality than CM reviews. It is possible that reviewers who are sensitive to the possibility of including reports of LOE are also more comprehensive in the entire conduct of the systematic review process. As such, language inclusive/LOE reviews are possibly a general ‘marker’ for better systematic review conduct. A similar explanation is likely warranted as to why CAM reviews are of higher quality. CAM systematic reviewers may also be motivated to do a better job because of the anecdotal view that CAM research is not scientific.

One way to improve the quality of reporting of systematic reviews is to introduce a reporting standard. The QUOROM of randomised trials statement67 was developed with this objective in mind. It is an evidence-based approach to help
improve the quality of reporting of meta-analyses. QUOROM, which consists of a 20-item checklist and flow diagram, follows similar efforts undertaken to help improve the quality of reporting of RCTs. QUOROM was recently published, and the sample of reviews included in this report were published before its introduction. Although the QUOROM group could only identify a small number of evidence-based items to guide the conduct of a systematic review, this list is growing as new evidence is published. The broader research community need to redouble their efforts to help improve the quality of reporting of systematic reviews. More healthcare journals, editorial groups and granting agencies need to promote QUOROM.

We observed that reports of English language CAM RCTs were of higher quality than those published in LOE. The higher quality EL CAM trials also tended to be more statistically positive, helping to explain, in part, the consequences of conducting a language restricted systematic review when evaluating the merits of a CAM intervention. This new form of selective publication requires further investigation before we can be confident that our observation is robust and valid. However, there is already some data supporting this position. Linde and colleagues reported similar results using the same assessment tool as ours.

To examine the relationship between the exclusion of reports of RCTs in LOE from the conduct of a systematic review, and its effect on producing biased results, we identified a large number of systematic reviews, 48 of which included LOE in the quantitative data synthesis. This allowed us to repeat the analyses with and without the LOE. Such an analysis permits the exploration of whether the systematic exclusion of reports of LOE from systematic reviews biases the estimates of the point estimate. We also explored whether language of publication exclusions and the type of interventions under consideration (i.e. CM and CAM) produced differential results.

We found that language restrictions (i.e. excluding reports of RCTs published in LOE) will not bias the results of systematic reviews provided that the intervention under consideration is a CM one. This finding holds true even after several planned sensitivity analyses. This finding supports an earlier one reported by us, although the database used in the present investigation was nearly three times larger than the earlier one. Taken together, we think that this finding is robust and recommend that a policy of language restrictions (i.e. limiting the quantitative data synthesis to reports of EL trials) is a reasonable approach for the conduct of CM systematic reviews.

However, when investigating the benefits of CAM interventions within the context of a systematic review, language restrictions are not recommended. Such restrictions will result in a substantial exaggeration of the estimates of the effects of the intervention. This finding did not appreciably change after several sensitivity analyses. We are not aware of this finding being reported previously.

We are confident that this is a robust result and can be generalised to systematic reviews in which the dominant languages are European (e.g. German, French). However, to what extent the results can be generalised to language restrictions in Asian languages remains uncertain. Even though we used an experienced information specialist to help develop the database, we were unable to include an appreciable number of reviews in which the associated trials were in Asian languages.

We investigated whether our language restriction results were influenced by statistical heterogeneity and whether publication bias played an important role in explaining the results. Traditional publication bias (i.e. the exclusion of negative and neutral results from the published literature) was not evident in our analysis. We used a variety of methods to explore this possibility. However, a new type of selective publication, not previously reported, may be evident. We observed that editors are likely to publish reports of CM RCTs in any language with similar quality, even if it is low. Similar findings have been reported previously, although Egger and colleagues have noted that German authors of conventional trials are more likely to report statistically positive results in English than German.

In conclusion, when conducting a systematic review, a decision needs to be made regarding whether the data analysis can be language restricted or not. There is now solid evidence that language restrictions are reasonable when conducting a systematic review that investigates the benefits of a CM intervention. However, such language restrictions are inappropriate, and can lead to substantial bias, when conducting a systematic review examining the merits of a CAM intervention. In this case, all trials regardless of the language in which they are reported should be included in the data synthesis.
Research recommendations

Priority 1: creating a meta-epidemiology database

Over the last 10 years, we have seen the emergence of the use of systematic reviews to examine the impact that specific biases can have on the results of randomised trials. Schulz and colleagues reported a landmark study in 1995. These investigators used trials with binary outcomes included in systematic reviews from the Cochrane Pregnancy and Childbirth Database. They showed that inadequately concealed trials, compared with those with adequate (allocation) concealment, yielded exaggerated estimates of treatment effect. Several other investigations have used this general approach to examine the impact of other biases on randomised trials and systematic reviews, including the impact of excluding grey literature, reports in LOE and EMBASE.

Typically these studies are initiated by assembling a database of systematic reviews and associated primary studies. In each case this usually necessitates developing the database de novo. There are considerable resource implications here. Additionally, the database becomes outdated quickly unless additional resources are brought to bear to keep it up to date. Similarly, these databases are usually developed in ‘isolation’, whereby a group of investigators assemble and extract specific data points unique to their specific needs. Although it is possible to share such a database, this is not without further requirements. The upshot is that this approach to conducting methods research (meta-epidemiology) is probably less than optimum from several perspectives.

Creating a single meta-epidemiology database that is kept up to date has several advantages. It will allow for sharing of a common database across different investigators. It will be even more useful if there is an agreed upon set of common data elements; such an approach will make for easier replication, examining the influence of new biases and collaboration.

We recommend that a committee be established to develop the meta-epidemiology database, ensuring that sufficient resources are allocated to keeping it up to date. This committee could also be tasked with developing an agreed upon common set of generic data elements. Beyond this core list of variables, individual groups of investigators could add project specific data elements. The committee could also produce a resource and procedure manual for ease of use. The database could be established by any group with previous experience in developing such databases. Additionally, it could be posted, along with the appropriate user manuals (discussed above), on a dedicated website.

Systematic reviews of randomised trials have been the focus of the present report and that of many others. However, the majority of healthcare literature consists of observational type designs. We recommend that the creation of any database should include reviews and associated studies that use observational design.

The inception cohort of systematic reviews and associated randomised trials could come from the NHS commissioned research through their Health Technology Assessment programme.

Priority 2: improving the quality of reporting of randomised trials

This report provides data suggestive of the need to improve the quality of reporting of systematic reviews and randomised trials. The latter are essential to informing the former and, importantly, investigating the impact of bias as already described. Individual meta-epidemiology databases have been used to show that low-quality reports of randomised trials, compared with higher quality ones, can introduce bias into the purported benefits of interventions. Clinical researchers and medical journal editors have come together and developed an approach to help improve the quality of reporting of randomised trials.

The consolidated standards of reporting randomised trials (CONSORT) statement, consisting of a flow diagram and a 22-item checklist, is an evidence-based approach to reporting two group parallel randomised trials. It has been endorsed by leading medical journals, including the premier British general medical journals (i.e. BMJ and The Lancet) and international editorial groups, such as the World Association of Medical Editors, as a way to improve the quality of reports of randomised trials. New data suggest that authors using CONSORT, compared with those not doing so, have higher quality reports of randomised trials.

To keep CONSORT evidence-based and up to date with emerging literature requires resources to spearhead research, the development of CONSORT for other clinical trial designs, including multi-arm, crossover and equivalence, and regular meetings of the group. We recommend using the CONSORT template and evidence-based process
to develop reporting ‘statements’ for other randomised trial designs, such as multi-arm, crossover, within participant (n of 1) and equivalence. We also recommend that the NHS Health Technology Assessment programme continues its support for methods research, the results of which can be used to help inform the continual development and refinement of the CONSORT statement.

Priority 3: Improving the quality of reporting of systematic reviews
Although randomised trials are the primary unit of a systematic review, we should not lose sight of the review process itself. The present results suggest considerable room for improving the quality of reporting of systematic reviews. The QUOROM statement, consisting of a flow diagram and a 20-item checklist, is an evidence-based approach to reporting systematic reviews of randomised trials. Although the QUOROM statement is only 3 years old, it is fast becoming updated owing, in part, to the fast pace of methods development. For example, some of the data reported in this report can be used to help inform the refinement of the QUOROM checklist. We recommend resources to use to bring together the QUOROM group with the specific purpose of updating the checklist using the most recent methods evidence.

Priority 4: Exploring the bias introduced by language of publication restrictions when conducting systematic reviews
The data presented in this report suggest that language of publication restrictions may introduce bias during the systematic review process, particularly if the intervention under evaluation is classified as CAM. To gain more confidence in this result, we recommend that a similar study be conducted using a separate set of randomised trials and systematic reviews, with a specific focus on those published in Asian languages. Such an examination could describe and quantify the impact of language, culture and quality of reports and their multiplicative effect that could help assessing without bias the effectiveness of CAM interventions amongst populations.

Similarly, we recommend exploring whether different cultures, who use a similar language, such as French and French Canadians, have different reporting standards.

Priority 5: A more in-depth evaluation of complementary and alternative medicine randomised trials and systematic reviews
CAM is becoming an increasingly common ‘treatment’ option for those needing to access the healthcare system. In a recent survey conducted by the WHO, 70% of Canadians reported using CAM at least once, which is comparable to 75% in the French population. As a consequence, the health literature has seen the number of trials evaluating CAM becoming more prevalent. A recent review of the paediatric CAM literature identified a small number of systematic reviews and a much larger number of randomised trials. The present report has shed some light on the possible biases these studies can exert. The data presented in this report suggests that LOE CAM trial reports are different compared with their CM comparators. As such, we recommend a more in-depth examination of CAM trials, particularly those reported in LOE.

Priority 6: Incorporating aspects of CAM methodology and content into critical appraisal skills training programmes
The data presented in this report suggest that reports of randomised trials and systematic reviews of CAM are different from those of CM. We recommend that critical appraisal programmes be developed that include specific attention to methods issues relating to CAM trials and systematic reviews.
We thank Nick Barrowman, Kaitryn Campbell, Manchun Fang, Alison Jones, Jessie McGowan, Leah Lepage, Margaret Sampson and Jesse Berlin for helping to prepare this report.
References


7. Hadhazy V, Ezzo J, Berman B. How valuable is effort to contact authors to obtain missing data in systematic reviews. Presented at the VII Cochrane Colloquium, Rome, 5–9 October 1999.


26. Selective Decontamination of the Digestive Tract
   Trials’ Collaborative Group. Meta-analysis of randomised controlled trials of selective

decontamination for control of nosocomial multi resistant gram-negative bacilli. *Ann Intern Med*


## Appendix 1

### Data abstraction form

<table>
<thead>
<tr>
<th>ID. NO. _____________________</th>
<th>REVIEWER _____________________</th>
</tr>
</thead>
</table>

### MA Characteristics

1. What year was the MARCT published? ________________

What country was the MARCT conducted in? (Use country of corresponding author) ________________

3a. What journal was the MARCT published in?

3b. What is the journal citation impact factor? ________________

4. What is the funding source for the MARCT?

<table>
<thead>
<tr>
<th>Option</th>
<th>selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single pharmaceutical company</td>
<td>☐</td>
</tr>
<tr>
<td>Multiple pharmaceutical companies</td>
<td>☐</td>
</tr>
<tr>
<td>Pharmaceutical company and a non-drug sponsor</td>
<td>☐</td>
</tr>
<tr>
<td>[one or more pharmaceutical companies and a non-pharmaceutical organisation]</td>
<td>☐</td>
</tr>
<tr>
<td>Non-drug company</td>
<td>☐</td>
</tr>
<tr>
<td>[a non-pharmaceutical company or sponsor including universities, medical societies, government and foundations]</td>
<td>☐</td>
</tr>
<tr>
<td>None listed/can’t tell</td>
<td>☐</td>
</tr>
</tbody>
</table>

### Data abstraction form

<table>
<thead>
<tr>
<th>ID. NO. _____________________</th>
<th>REVIEWER _____________________</th>
</tr>
</thead>
</table>

## 1.

What is the broad ICD-10 category investigated in this MARCT?

- ☐ 1. Intestinal infectious diseases
- ☐ 2. Neoplasms
- ☐ 3. Endocrine, nutritional and metabolic diseases and immunity disorders
- ☐ 4. Diseases of the blood and blood-forming organs
- ☐ 5. Mental disorders
- ☐ 6. Diseases of the nervous system and sense organs
- ☐ 7. Diseases of the circulatory system
- ☐ 8. Diseases of the respiratory system
- ☐ 9. Diseases of the digestive system
- ☐ 10. Diseases of the genitourinary system
- ☐ 11. Complications of pregnancy, childbirth and the puerperium
- ☐ 12. Diseases of the skin and subcutaneous tissue
- ☐ 13. Diseases of the musculoskeletal system and connective tissue
- ☐ 14. Congenital anomalies
- ☐ 15. Certain conditions originating in the perinatal period
- ☐ 16. Symptoms, signs and ill-defined conditions

© Queen’s Printer and Controller of HMSO 2003. All rights reserved.
17. Injury and poisoning
18. Supplemental classification of external causes of injury and poisoning
19. Supplementary classification of factors influencing health status and contact with health services

2. How would you classify the MARCT?
- Language restricted
- Language inclusive – with TLPOE
- Language inclusive – without TPLOE

If the MARCT contains TPLOE, please answer the next question. If not, please go to Question 4.

3. Which non-English languages are included in the MARCT? Please indicate the language and how many trials in each language are included.

Search, Study Selection and Abstraction

4. How were the trials identified?
- Medline
- Embase
- Other Electronic Databases
- Hand Searching
- Reference Lists
- Corresponding Authors
- Content Experts
- Abstracts
- Conference Proceedings
- Can’t Tell

5a. How many independent reviewers assisted in the study selection process?
- One
- Two
- Three
- More than three
- Can’t Tell

5b. If more than one reviewer, was there a reliability test done?
- Yes
- No
- Can’t Tell

6a. How many independent reviewers completed the data extraction?
- One
- Two
- Three
- More than three
- Can’t Tell

6b. If more than one reviewer, was a pre-defined form used?
- Yes
- No
- Can’t Tell

Data Synthesis

7. How many trials were considered for inclusion?

8. What is the total number of participants randomised?

9a. Was there any formal evaluation of statistical heterogeneity reported?
- Yes
- No
- Can’t Tell

9b. If yes, which methods were used?
- Visual Inspection of Forest Plot
- L’Abbé Plot
- Statistical test of heterogeneity
- Can’t Tell
- Other, please specify

Appendix 1
10a. Was there any formal evaluation of clinical heterogeneity reported?
   - Yes  □ No  □ Can’t Tell

10b. If yes, which methods were used?
   - Association with Control Rate
   - Include a covariate analysis
   - Can’t tell
   - Other, please specify____________________________

11. What model was used for the analyses?
   - Fixed effects model
   - Random effects model
   - Can’t tell

---

**PUBLICATION BIAS**

12. Was publication bias assessed in the MARCT?
   - Yes  □ No  □ Can’t Tell

13a. Did the authors include grey literature?
   - Yes  □ No  □ Can’t Tell

13b. If yes, how many “grey” items were included?
   - One  □ Two  □ Three  □ More than three  □ Can’t Tell

13c. If yes, what sources of grey literature (Reports that are unpublished, have limited distribution and/or are not included in bibliographic retrieval systems).
   - Abstracts  □ Unpublished studies  □ Thesis  □ Papers in press
   - Book chapters  □ Company reports  □ Book chapters
   - Drug company reports

14. Was the decision to assess publication bias:
   - A priori  □ Posteriori  □ Can’t tell

15. If publication was assessed which method did the authors explicitly report using?
   - Visual inspection  □ Regression methods  □ Rank correlation
   - Trim and fill  □ Subgroup analyses by sample size  □ Funnel plot
   - Fail-safe method  □ Selection model
   - Can’t tell  □ Other __________________________

16. Did the authors report combined results (e.g. pooled odds ratio, pooled effect size) adjusting for publication bias?
   - Yes  □ No  □ Can’t Tell

17. Did the authors discuss the results of their trials in relation to publication bias?
   - Yes  □ No  □ Can’t Tell
Appendix 2

Jadad and allocation concealment data collection form

Study Quality

Trial: _______________________

Randomisation: Total Points: □□ 0 □□ 1 □□ 2
A trial reporting that it is “randomised” is to receive one point. Trials describing an appropriate method of randomisation (table of random numbers, computer generated) receive an additional point. However, if the report describes the trial as randomised and uses an inappropriate method of randomisation (date of birth, hospital numbers) a point is deducted.

Double-blinding: Total Points: □□ 0 □□ 1 □□ 2
A trial reporting that is “double blind”, it is to receive one point. Trials that describe an appropriate method of double blinding (identical placebo, active placebo) are to receive an additional point. However, if the report describes the trial as double blind and uses an inappropriate method (comparison of tablets versus injection with no double dummy), a point is deducted.

Withdrawals and dropouts: Total Points: □□ 0 □□ 1 □□ 2
A trial reporting the number and reasons for withdrawals is to receive one point. If there is no statement, no point is given.

TOTAL Score: □ Low (0-2 pts) □ Moderate (3-4 pts) □ High (5 pts)

Allocation concealment: □□ Adequate □□ Inadequate □□ Unclear

Adequate: Central randomisation; numbered or coded bottles or containers; drugs prepared by a pharmacy, serially numbered, opaque, sealed envelopes, etc.

Inadequate: alternation; reference to case record # or date of birth, etc.

Unclear: allocation concealment approach is not reported or fits neither above category.
Appendix 3
Listing and citations of systematic reviews included in research

Language inclusive/LOE systematic reviews


Language restricted systematic reviews


Language inclusive/EL systematic reviews


Health Technology Assessment reports published to date

Volume 1, 1997

No. 1
Home parenteral nutrition: a systematic review.
By Richards DM, Deeks JJ, Sheldon TA, Shaffer JL.

No. 2
Diagnosis, management and screening of early localised prostate cancer.
A review by Selley S, Donovan J, Faulkner A, Coast J, Gillatt D.

No. 3
The diagnosis, management, treatment and costs of prostate cancer in England and Wales.
A review by Chamberlain J, Melia J, Moss S, Brown J.

No. 4
Screening for fragile X syndrome.
A review by Murray J, Cuckle H, Taylor G, Hewison J.

No. 5
A review of near patient testing in primary care.

No. 6
Systematic review of outpatient services for chronic pain control.
By McQuay HJ, Moore RA, Eccleston C, Morley S, de C Williams AC.

No. 7
Neonatal screening for inborn errors of metabolism: cost, yield and outcome.

No. 8
Preschool vision screening.
A review by Snowdon SK, Stewart-Brown SL.

No. 9
Implications of socio-cultural contexts for the ethics of clinical trials.
A review by Ashcroft RE, Chadwick DW, Clark SRL, Edwards RHT, Frith L, Hutton JL.

No. 10
A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment.
By Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S.

Volume 2, 1998

No. 1
Antenatal screening for Down’s syndrome.
A review by Wald NJ, Kennard A, Hackshaw A, McGuire A.

No. 2
Screening for ovarian cancer: a systematic review.
By Bell R, Petticrew M, Luengo S, Sheldon TA.

No. 3
Consensus development methods, and their use in clinical guideline development.

No. 4
A cost-utility analysis of interferon beta for multiple sclerosis.

No. 5
Effectiveness and efficiency of methods of dialysis therapy for end-stage renal disease: systematic reviews.
By MacLeod A, Grant A, Donaldson C, Khan I, Campbell M, Daly C, et al.

No. 6
Effectiveness of hip prostheses in primary total hip replacement: a critical review of evidence and an economic model.

No. 7
Antimicrobial prophylaxis in colorectal surgery: a systematic review of randomised controlled trials.
By Song F, Glenny AM.

No. 8
Bone marrow and peripheral blood stem cell transplantation for malignancy.
A review by Johnson PWM, Simnett SJ, Sweetenham JW, Morgan GJ, Stewart LA.

No. 9
Screening for speech and language delay: a systematic review of the literature.
By Law J, Boyle J, Harris F, Harkness A, Nye C.

No. 10
By Sculpher MJ, Petticrew M, Kelland JL, Elliott RA, Holdright DR, Buxton MJ.

No. 11
Detection, adherence and control of hypertension for the prevention of stroke: a systematic review.
By Ebrahim S.

No. 12
Postoperative analgesia and vomiting, with special reference to day-case surgery: a systematic review.
By McQuay HJ, Moore RA.

No. 13
Choosing between randomised and nonrandomised studies: a systematic review.
By Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C.

No. 14
Evaluating patient-based outcome measures for use in clinical trials.
A review by Fitzpatrick R, Davey C, Buxton MJ, Jones DR.
Health Technology Assessment reports published to date

No. 1
Informed decision making: an annotated bibliography and systematic review.

No. 2
Handling uncertainty when performing economic evaluation of healthcare interventions.
A review by Briggs AH, Gray AM.

No. 3
The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review.

No. 4

No. 5
Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review.
By Ukoumunne OC, Guilford MC, Chinn S, Sterne JAC, Burney PGJ.

No. 6
Assessing the costs of healthcare technologies in clinical trials. A review by Johnston K, Buxton MJ, Jones DR, Fitzpatrick R.

No. 7
Cooperatives and their primary care emergency centres: organisation and impact.
By Hallam L, Henthorne K.

No. 8
Screening for cystic fibrosis. A review by Murray J, Cuckle H, Taylor G, Littlewood J, Hewison J.

No. 9
A review of the use of health status measures in economic evaluation.
By Brazier J, Deverill M, Green C, Harper R, Booth A.

No. 10

No. 11
Antenatal and neonatal haemoglobinopathy screening in the UK: review and economic analysis.
By Zeuner D, Ades AE, Karnon J, Brown J, Dezateux C, Anionwu EN.

No. 12

No. 13
‘Early warning systems’ for identifying new healthcare technologies.
By Robert G, Stevens A, Gabbay J.

No. 14
A systematic review of the role of human papillomavirus testing within a cervical screening programme.

No. 15
Near patient testing in diabetes clinics: appraising the costs and outcomes. By Grieve R, Beech R, Vincent J, Mazurkiewicz J.

No. 16
Positron emission tomography: establishing priorities for health technology assessment. A review by Robert G, Milne R.

No. 17 (Pt 1)
The debridement of chronic wounds: a systematic review.
By Bradley M, Cullum N, Sheldon T.

No. 17 (Pt 2)
Systematic reviews of wound care management: (2) Dressings and topical agents used in the healing of chronic wounds.
By Bradley M, Cullum N, Nelson EA, Petticrew M, Sheldon T, Torgerson D.

No. 18
A systematic literature review of spiral and electron beam computed tomography: with particular reference to clinical applications in hepatic lesions, pulmonary embolus and coronary artery disease.

No. 19
What role for statins? A review and economic model.

No. 20
Factors that limit the quality, number and progress of randomised controlled trials.
A review by Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiuaka S, et al.

No. 21
Antimicrobial prophylaxis in total hip replacement: a systematic review.
By Glenn AM, Song F.

No. 22
Health promoting schools and health promotion in schools: two systematic reviews.
By Lister-Sharp D, Chapman S, Stewart-Brown S, Sowden A.

No. 23
Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee.

Volume 4, 2000

No. 1
The estimation of marginal time preference in a UK-wide sample (TEMPUS) project. A review by Cairns JA, van der Pol MM.

No. 2
Geriatric rehabilitation following fractures in older people: a systematic review.
No. 3  Screening for sickle cell disease and thalassaemia: a systematic review with supplementary research.
By Davies SC, Cronin E, Gill M, Greengross P, Hickman M, Normand C.

No. 4  Community provision of hearing aids and related audiology services.
A review by Reeves DJ, Alborz A, Hickson FS, Bamford JM.

No. 5  False-negative results in screening programmes: systematic review of impact and implications.
By Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K.

No. 6  Costs and benefits of community postnatal support workers: a randomised controlled trial.
By Morrell CJ, Spiby H, Stewart P, Walters S, Morgan A.

No. 7  Implantable contraceptives (subdermal implants and hormonally impregnated intrauterine systems) versus other forms of reversible contraceptives: two systematic reviews to assess relative effectiveness, acceptability, tolerability and cost-effectiveness.

No. 8  An introduction to statistical methods for health technology assessment.
A review by White SJ, Ashby D, Brown PJ.

No. 9  Disease-modifying drugs for multiple sclerosis: a rapid and systematic review.
By Clegg A, Bryant J, Milne R.

No. 10  Publication and related biases.
A review by Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ.

No. 11  Cost and outcome implications of the organisation of vascular services.
By Michaels J, Brazier J, Palfreyman S, Shackley P, Slack R.

No. 12  Monitoring blood glucose control in diabetes mellitus: a systematic review.
By Coster S, Gulliford MC, Seed PT, Powrie JK, Swanninathan R.

No. 13  The effectiveness of domiciliary health visiting: a systematic review of international studies and a selective review of the British literature.

No. 14  The determinants of screening uptake and interventions for increasing uptake: a systematic review.

No. 15  The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth.
A rapid review by Song F, O'Meara S, Wilson P, Golding S, Kleijnen J.

No. 16  Ultrasound screening in pregnancy: a systematic review of the clinical effectiveness, cost-effectiveness and women’s views.

No. 17  A rapid and systematic review of the effectiveness and cost-effectiveness of the taxanes used in the treatment of advanced breast and ovarian cancer.
By Lister-Sharp D, McDonagh MS, Khan KS, Kleijnen J.

No. 18  Liquid-based cytology in cervical screening: a rapid and systematic review.
By Payne N, Chilcott J, McGoogan E.

No. 19  Randomised controlled trial of non-directive counselling, cognitive–behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care.

No. 20  Routine referral for radiography of patients presenting with low back pain: is patients’ outcome influenced by GPs’ referral for plain radiography?
By Kerry S, Hilton S, Patel S, Dundas D, Rink E, Lord J.

No. 21  Systematic reviews of wound care management: (3) antimicrobial agents for chronic wounds; (4) diabetic foot ulceration.
By O’Meara S, Cullum N, Majid M, Sheldon T.

No. 22  Using routine data to complement and enhance the results of randomised controlled trials.
By Lewsey JD, Leyland AH, Murray GD, Boddy FA.

No. 23  Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review.
By Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C.

No. 24  Outcome measures for adult critical care: a systematic review.
By Hayes JA, Black NA, Jenkinson C, Young JD, Rowan KM, Daly K, et al.

No. 25  A systematic review to evaluate the effectiveness of interventions to promote the initiation of breastfeeding.
By Fairbank L, O’Meara S, Renfrew MJ, Woolridge M, Sowden AJ, Lister-Sharp D.

No. 26  Implantable cardioverter defibrillators: arrhythmias. A rapid and systematic review.
By Parkes J, Bryant J, Milne R.

No. 27  Treatments for fatigue in multiple sclerosis: a rapid and systematic review.
By Bráñas P, Jordan R, Fry-Smith A, Burls A, Hyde C.

No. 28  Early asthma prophylaxis, natural history, skeletal development and economy (EASE): a pilot randomised controlled trial.

No. 29  Screening for hypercholesterolaemia versus case finding for familial hypercholesterolaemia: a systematic review and cost-effectiveness analysis.
By Marks D, Wonderling D, Thorogood M, Lambert H, Humphries SE, Neil HAW.

No. 30  A rapid and systematic review of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists in the medical management of unstable angina.
By McDonagh MS, Bachmann LM, Golder S, Kleijnen J, ter Riet G.

No. 31  A randomised controlled trial of prehospital intravenous fluid replacement therapy in serious trauma.
By Turner J, Nicholl J, Webber L, Cox H, Dixon S, Yates D.

No. 32  Intrathecal pumps for giving opioids in chronic pain: a systematic review.
By Williams JE, Louw G, Towlerton G.

No. 33  Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C: a rapid and systematic review.
By Shepherd J, Waugh N, Hewitson P.
No. 34  A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies.
By MacLhose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS.

No. 35  Intravascular ultrasound-guided interventions in coronary artery disease: a systematic literature review, with decision-analytic modelling, of outcomes and cost-effectiveness.
By Berry E, Kelly S, Hutton J, Lindsay HSJ, Blaxill JM, Evans JA, et al.

No. 36  A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression.
By Simpson S, Corney R, Fitzgerald P, Beecham J.

No. 37  Systematic review of treatments for atopic eczema.
By Hoare C, Li Wan Po A, Williams H.

No. 38  Bayesian methods in health technology assessment: a review.
By Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR.

No. 39  The management of dyspepsia: a systematic review.

No. 40  A systematic review of treatments for severe psoriasis.
By Griffiths CEM, Clark CM, Chalmers R, Li Wan Po A, Williams HC.

Volume 5, 2001

No. 1  Clinical and cost-effectiveness of donepezil, rivastigmine and galantamine for Alzheimer’s disease: a rapid and systematic review.

No. 2  The clinical effectiveness and cost-effectiveness of riluzole for motor neurone disease: a rapid and systematic review.

No. 3  Equity and the economic evaluation of healthcare.
By Sassi F, Archard L, Le Grand J.

No. 4  Quality-of-life measures in chronic diseases of childhood.
By Eiser C, Morse R.

No. 5  Eliciting public preferences for healthcare: a systematic review of techniques.

No. 6  General health status measures for people with cognitive impairment: learning disability and acquired brain injury.
By Riemsma RP, Forbes CA, Glanville JM, Eastwood AJ, Kleijnen J.

No. 7  An assessment of screening strategies for fragile X syndrome in the UK.
By Pembrey ME, Barnicoat AJ, Carmichael B, Bobrow M, Turner G.

No. 8  Issues in methodological research: perspectives from researchers and commissioners.
By Lilford RJ, Richardson A, Stevens K, Sheldon T.

No. 9  Systematic reviews of wound care management: (5) beds; (6) compression; (7) laser therapy, therapeutic ultrasound, electrotherapy and electromagnetic therapy.
By Callum N, Nelson EA, Flemming K, Sheldon T.

No. 10  Effects of educational and psychosocial interventions for adolescents with diabetes mellitus: a systematic review.
By Hampson SE, Skinner TC, Hart J, Storey L, Gage H, Foxcroft D, et al.

No. 11  Effectiveness of autologous chondrocyte transplantation for hyaline cartilage defects in knees: a rapid and systematic review.
By Johanputra P, Parry D, Fry-Smith A, Burls A.

No. 12  Statistical assessment of the learning curves of health technologies.
By Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT.

No. 13  The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma: a rapid and systematic review.
By Dinnes J, Cave C, Huang S, Major K, Milne R.

No. 14  A rapid and systematic review of the clinical effectiveness and cost-effectiveness of debriding agents in treating surgical wounds healing by secondary intention.
By Lewis R, Whiting P, ter Riet G, O’Meara S, Glanville J.

No. 15  Home treatment for mental health problems: a systematic review.

No. 16  How to develop cost-conscious guidelines.
By Eccles M, Mason J.

No. 17  The role of specialist nurses in multiple sclerosis: a rapid and systematic review.
By De Broe S, Christopher F, Waugh N.

No. 18  A rapid and systematic review of the clinical effectiveness and cost-effectiveness of orlistat in the management of obesity.
By O’Meara S, Riemsma R, Sharran L, Mathet L, ter Riet G.

No. 19  The clinical effectiveness and cost-effectiveness of pioglitazone for type 2 diabetes mellitus: a rapid and systematic review.
By Chilcott J, Wight J, Lloyd Jones M, Tappenden P.

No. 20  Extended scope of nursing practice: a multicentre randomised controlled trial of appropriately trained nurses and preregistration house officers in pre-operative assessment in elective general surgery.

No. 21  Systematic reviews of the effectiveness of day care for people with severe mental disorders: (1) Acute day hospital versus admission; (2) Vocational rehabilitation; (3) Day hospital versus outpatient care.

No. 22  The measurement and monitoring of surgical adverse events.
By Bruce J, Russell EM, Mollison J, Krukowski ZH.

No. 23  Action research: a systematic review and guidance for assessment.
By Waterman H, Tillen D, Dickson R, de Koning K.

No. 24  A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer.
No. 25  A rapid and systematic review of the evidence for the clinical effectiveness and cost-effectiveness of ziprasidone, olanzapine, and quetiapine for the treatment of schizophrenia.
   By Lloyd Jones M, Hummel S, Bansback N, Orr B, Seymour M.

No. 26  Comparison of the effectiveness of depot antipsychotic medication in the treatment of patients with schizophrenia:

No. 27  The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint.

No. 28  A rapid and systematic review of the clinical effectiveness and cost-effectiveness of topotecan for ovarian cancer.
   By Forbes C, Shirran L, Bagnall A-M, Duffy S, ter Riet G.

No. 29  Superseded by a report published in a later volume.

No. 30  The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial.
   By Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M.

No. 31  Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients.

No. 32  A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer.

No. 33  Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives.
   By Brooks ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G.

No. 34  Depot antipsychotic medication in the treatment of patients with schizophrenia: (1) Meta-review; (2) Patient and nurse attitudes.
   By David AS, Adams C.

No. 35  A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression.

No. 36  Cost analysis of child health surveillance.
   By Sanderson D, Wright D, Acton C, Duree D.

Volume 6, 2002

No. 1  A study of the methods used to select review criteria for clinical audit.
   By Hearshaw H, Harker R, Cheater F, Baker R, Grimshaw G.

No. 2  Fluorarabine as second-line therapy for B cell chronic lymphocytic leukaemia: a technology assessment.

No. 3  Rituximab as third-line treatment for refractory or recurrent Stage III or IV follicular non-Hodgkin’s lymphoma: a systematic review and economic evaluation.

No. 4  A systematic review of discharge arrangements for older people.

No. 5  The clinical effectiveness and cost-effectiveness of inhaler devices used in the routine management of chronic asthma in older children: a systematic review and economic evaluation.
   By Peters J, Stevenson M, Beverley C, Lim J, Smith S.

No. 6  The clinical effectiveness and cost-effectiveness of sibutramine in the management of obesity: a technology assessment.
   By O’Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 7  The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review.

No. 8  Promoting physical activity in South Asian Muslim women through ‘exercise on prescription’.
   By Carroll B, Ali N, Azam N.

No. 9  Zanamivir for the treatment of influenza in adults: a systematic review and economic evaluation.

No. 10  A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models.
   By Richards RG, Sampson FC, Beard SM, Tappenden P.

No. 11  Screening for gestational diabetes: a systematic review and economic evaluation.
   By Scott DA, Loveman E, McIntyre L, Waugh N.

No. 12  The clinical effectiveness and cost-effectiveness of surgery for people with morbid obesity: a systematic review and economic evaluation.

No. 13  The clinical effectiveness of trastuzumab for breast cancer: a systematic review.

No. 14  The clinical effectiveness and cost-effectiveness of vinorelbine for breast cancer: a systematic review and economic evaluation.

No. 15  A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease.
   By Vale L, Wynn L, McCormack K, McKenzie L, Brazzelli M, Stearns SC.

No. 16  The clinical effectiveness and cost-effectiveness of bupropion and nicotine replacement therapy for smoking cessation: a systematic review and economic evaluation.
   By Woolacott NF, Jones L, Forbes CA, Mather LC, Sowden AJ, Song FJ, et al.

No. 17  A systematic review of the effectiveness and economic evaluation of new drug treatments for juvenile idiopathic arthritis: etanercept.
   By Cummins C, Connoock M, Fry-Smith A, Burds A.

No. 18  Clinical effectiveness and cost-effectiveness of growth hormone in children: a systematic review and economic evaluation.
By Bryant J, Loveman E, Chase D, Mihaylova B, Cave C, Gerad K, et al.

No. 20 Clinical medication review by a pharmacist of patients on repeat prescriptions in general practice: a randomised controlled trial.
By Zemansky AG, Petty DR, Raynor DK, Lowe CJ, Freemantle N, Vail A.

No. 21 The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation.
By Jobanputra P, Barton P, Bryan S, Burls A.

No. 22 A systematic review and economic evaluation of computerised cognitive behaviour therapy for depression and anxiety.
By Kaltenhalter E, Shackley P, Stevens K, Beverley C, Parry G, Chilcott J.

No. 23 A systematic review and economic evaluation of pegylated liposomal doxorubicin hydrochloride for ovarian cancer.
By Forbes C, Wilby J, Richardson G, Sculpher M, Mather L, Reimsmra R.

No. 24 A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behaviour change.

No. 25 A systematic review update of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists.

No. 26 A systematic review of the effectiveness, cost-effectiveness and barriers to implementation of thrombolytic and neuroprotective therapy for acute ischaemic stroke in the NHS.

No. 27 A randomised controlled crossover trial of nurse practitioner versus doctor-led outpatient care in a bronchiectasis clinic.

No. 28 Clinical effectiveness and cost–consequences of selective serotonin reuptake inhibitors in the treatment of sex offenders.
By Adi Y, Ashcroft D, Browne K, Beech A, Fry-Smith A, Hyde C.

No. 29 Treatment of established osteoporosis: a systematic review and cost–utility analysis.
By Kears JA, Brazier JE, Stevenson M, Calvert NW, Lloyd Jones M.

No. 30 Which anaesthetic agents are cost-effective in day surgery? Literature review, national survey of practice and randomised controlled trial.

No. 31 Screening for hepatitis C among injecting drug users and in genitourinary medicine clinics: systematic reviews of effectiveness, modelling study and national survey of current practice.

No. 32 The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature.

No. 33 The effectiveness and cost-effectiveness of imatinib in chronic myeloid leukaemia: a systematic review.
By Garside R, Round A, Dalziel K, Stein K, Royle R.

No. 34 A comparative study of hypertonic saline, daily and alternate-day rhDNase in children with cystic fibrosis.

No. 35 A systematic review of the costs and effectiveness of different models of paediatric home care.

**Volume 7, 2003**

No. 1 How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study.
By Egger M, Juni P, Bartlett C, Holenstein F, Sterne J.

No. 2 Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of home versus hospital or satellite unit haemodialysis for people with end-stage renal failure.

No. 3 Systematic review and economic evaluation of the effectiveness of infliximab for the treatment of Crohn’s disease.
By Clark W, Raftery J, Barton P, Song F, Fry-Smith A, Burls A.

No. 4 A review of the clinical effectiveness and cost-effectiveness of routine anti-D prophylaxis for pregnant women who are rhesus negative.

No. 5 Systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing’s sarcoma and neuroblastoma.

No. 6 The cost-effectiveness of screening for *Helicobacter pylori* to reduce mortality and morbidity from gastric cancer and peptic ulcer disease: a discrete-event simulation model.

No. 7 The clinical effectiveness and cost-effectiveness of routine dental checks: a systematic review and economic evaluation.

No. 8 A multicentre randomised controlled trial assessing the costs and benefits of using structured information and analysis of women’s preferences in the management of menorrhagia.

No. 9 Clinical effectiveness and cost–utility of photodynamic therapy for wet age-related macular degeneration: a systematic review and economic evaluation.
By Meads C, Salas C, Roberts T, Moore D, Fry-Smith A, Hyde C.

No. 10 Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities.
No. 11  
First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS).  
By Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM.

No. 12  
The effectiveness and cost-effectiveness of ultrasound locating devices for central venous access: a systematic review and economic evaluation.  
By Calvert N, Hind D, McWilliams RG, Thomas SM, Beverley C, Davidson A.

No. 13  
A systematic review of atypical antipsychotics in schizophrenia.  

No. 14  
Prostate Testing for Cancer and Treatment (ProtecT) feasibility study.  
By Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L, et al.

No. 15  
Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation.  

No. 16  
Screening for fragile X syndrome: a literature review and modelling.  
By Song FJ, Barton P, Sleightholme V, Yao GL, Fry-Smith A.

No. 17  
Systematic review of endoscopic sinus surgery for nasal polyps.  
By Dalziel K, Stein K, Round A, By Song FJ, Barton P, Sleightholme V, Yao GL, Fry-Smith A.

No. 18  
Towards efficient guidelines: how to monitor guideline use in primary care.  
By Hutchinson A, McIntosh A, Cox S, Gilbert C.

No. 19  
Effectiveness and cost-effectiveness of acute hospital-based spinal cord injuries services: systematic review.  
By Bagall A-M, Jones L, Richardson G, Duffy S, Riemsma R.

No. 20  
Prioritisation of health technology assessment. The PATHS model: methods and case studies.  
By Townsend J, Buxton M, Harper G.

No. 21  

No. 22  
By Loveman E, Cave C, Green C, Royle P, Dunn N, Waugh N.

No. 23  
The role of modelling in prioritising and planning clinical trials.  
By Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P.

No. 24  
Cost-benefit evaluation of routine influenza immunisation in people 65–74 years of age.  
By Allsup S, Gosney M, Haycox A, Regin M.

No. 25  
The clinical and cost-effectiveness of pulsatile machine perfusion versus cold storage of kidneys for transplantation retrieved from heart-beating and non-heart-beating donors.  
By Wight J, Chilcott J, Holmes M, Brewer N.

No. 26  
Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment.  
By Williams JG, Cheung WY, Cohen DR, Hutchings HA, Longo MF, Russell IT.

No. 27  
Evaluating non-randomised intervention studies.  

No. 28  
A randomised controlled trial to assess the impact of a package comprising a patient-oriented, evidence-based self-help guidebook and patient-centred consultations on disease management and satisfaction in inflammatory bowel disease.  

No. 29  
The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review.  
By Dinnes J, Loveman E, McIntyre L, Waugh N.

No. 30  
The value of digital imaging in diabetic retinopathy.  

No. 31  
Lowering blood pressure to prevent myocardial infarction and stroke: a new preventive strategy.  
By Law M, Wald N, Morris J.

No. 32  
Clinical and cost-effectiveness of capecitabine and tegafur with uracil for the treatment of metastatic colorectal cancer: systematic review and economic evaluation.  
By Ward S, Kaltenthaler E, Cowan J, Brewer N.

No. 33  
By Hummel S, Paisley S, Morgan A, Currie E, Brewer N.

No. 34  
Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system.  
By Royle P, Waugh N.

No. 35  
Systematic review and economic decision modelling for the prevention and treatment of influenza A and B.  

No. 36  
A randomised controlled trial to evaluate the clinical and cost-effectiveness of Hickman line insertions in adult cancer patients by nurses.  
By Boland A, Haycox A, Bagust A, Fitzsimmons L.

No. 37  
Redesigning postnatal care: a randomised controlled trial of protocol-based midwifery-led care focused on individual women’s physical and psychological health needs.  

No. 38  
Estimating implied rates of discount in healthcare decision-making.  
By West RR, McNabb R, Thompson AGH, Sheldon TA, Grimley Evans J.
No. 39
Systematic review of isolation policies in the hospital management of methicillin-resistant *Staphylococcus aureus*: a review of the literature with epidemiological and economic modelling.
By Cooper BS, Stone SP, Kibbler CC, Cookson BD, Roberts JA, Medley GF, et al.

No. 40
Treatments for spasticity and pain in multiple sclerosis: a systematic review.
By Beard S, Hunn A, Wight J.

No. 41
The inclusion of reports of randomised trials published in languages other than English in systematic reviews.
By Moher D, Pham B, Lawson ML, Klassen TP.
**Health Technology Assessment Programme**

**Prioritisation Strategy Group**

<table>
<thead>
<tr>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chair,</strong> Professor Kent Woods, Director, NHS HTA Programme &amp; Professor of Therapeutics, University of Leicester</td>
</tr>
<tr>
<td>Professor Bruce Campbell, Consultant Vascular &amp; General Surgeon, Royal Devon &amp; Exeter Hospital</td>
</tr>
<tr>
<td>Professor Shah Ebrahim, Professor in Epidemiology of Ageing, University of Bristol</td>
</tr>
<tr>
<td>Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Radcliffe Hospital, Oxford</td>
</tr>
<tr>
<td>Dr Ron Zimmern, Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Programme Director,</strong> Professor Kent Woods, Director, NHS HTA Programme, Department of Medicine and Therapeutics, Leicester Royal Infirmary, Robert Kilpatrick Clinical Sciences Building, Leicester</td>
</tr>
<tr>
<td>Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health &amp; Related Research, University of Sheffield, ScHARR Regent Court, Sheffield</td>
</tr>
<tr>
<td>Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Oxford</td>
</tr>
<tr>
<td>Dr Christine Clark, Medical Writer &amp; Consultant Pharmacist, Cloudside, Rossendale, Lancashire and Principal Research Fellow, Clinical Therapeutics in the School of Pharmacy, Bradford University, Bradford</td>
</tr>
<tr>
<td>Professor Nicky Callum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York, Research Section, Sebohm Roomtree Building, Heslington, York</td>
</tr>
<tr>
<td>Dr Andrew Farmer, Senior Lecturer in General Practice, Department of Primary Health Care, University of Oxford, Institute of Health Sciences, Headington, Oxford</td>
</tr>
<tr>
<td>Professor Fiona J Gilbert, Professor of Radiology, Department of Radiology, University of Aberdeen, Lillian Sutton Building, Foresterhill, Aberdeen</td>
</tr>
<tr>
<td>Professor Adrian Grant, Director, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen</td>
</tr>
<tr>
<td>Professor F D Richard Hobbs, Professor of Primary Care &amp; General Practice, Department of Primary Care &amp; General Practice, University of Birmingham, Primary Care and Clinical Sciences Building, Edgbaston, Birmingham</td>
</tr>
<tr>
<td>Professor Peter Jones, Head of Department, University Department of Psychiatry, University of Cambridge, Addenbrooke’s Hospital, Cambridge</td>
</tr>
<tr>
<td>Professor Sallie Lamb, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health, Coventry University, Coventry</td>
</tr>
<tr>
<td>Dr Donna Lamping, Senior Lecturer, Health Services Research Unit, Public Health and Policy, London School of Hygiene and Tropical Medicine, London</td>
</tr>
<tr>
<td>Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Park House, Birmingham</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Professor Douglas Altman,</strong> Professor of Statistics in Medicine, Centre for Statistics in Medicine, Oxford University, Institute of Health Sciences, Cancer Research UK Medical Statistics Group, Headington, Oxford</td>
</tr>
<tr>
<td>Professor John Bond, Professor of Health Services Research, Centre for Health Services Research, University of Newcastle, School of Health Sciences, Newcastle upon Tyne</td>
</tr>
<tr>
<td>Professor Alastair Gray, Director, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen</td>
</tr>
<tr>
<td>Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health &amp; Related Research, University of Sheffield, ScHARR Regent Court, Sheffield</td>
</tr>
<tr>
<td>Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine, Health &amp; Social Care, St George’s Building, Portsmouth</td>
</tr>
<tr>
<td>Professor David Neal, Professor of Surgical Oncology, Oncology Centre, Addenbrooke’s Hospital, Cambridge</td>
</tr>
<tr>
<td>Professor Ian Roberts, Professor of Epidemiology &amp; Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine, London</td>
</tr>
<tr>
<td>Professor Peter Sandercoc, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital NHS Trust, Bramwell Dott Building, Edinburgh</td>
</tr>
<tr>
<td>Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine, Health &amp; Social Care, St George’s Building, Portsmouth</td>
</tr>
<tr>
<td>Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Park House, Birmingham</td>
</tr>
</tbody>
</table>

Current and past membership details of all HTA ‘committees’ are available from the HTA website (www.nchta.org)
## Diagnostic Technologies & Screening Panel

<table>
<thead>
<tr>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chair,</strong> Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge</td>
</tr>
<tr>
<td>Dr David Elliman, Consultant in Community Child Health, London</td>
</tr>
<tr>
<td>Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London</td>
</tr>
<tr>
<td>Dr Andrew Farmer, Senior Lecturer in General Practice, Institute of Health Sciences, University of Oxford</td>
</tr>
<tr>
<td>Dr Karen N Roster, Clinical Lecturer, Dept of General Practice &amp; Primary Care, University of Aberdeen</td>
</tr>
<tr>
<td>Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton</td>
</tr>
<tr>
<td>Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff</td>
</tr>
<tr>
<td>Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London</td>
</tr>
<tr>
<td>Professor Jane Franklyn, Professor of Medicine, University of Birmingham</td>
</tr>
<tr>
<td>Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust</td>
</tr>
<tr>
<td>Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust, Devon</td>
</tr>
<tr>
<td>Mr Tony Tester, Chief Officer, South Bedfordshire Community Health Council, Luton</td>
</tr>
<tr>
<td>Dr Andrew Walker, Senior Lecturer in Health Economics, University of Gloucester</td>
</tr>
<tr>
<td>Professor Martin J Whittle, Head of Division of Reproductive &amp; Child Health, University of Birmingham</td>
</tr>
<tr>
<td>Dr Dennis Wright, Consultant Biochemist &amp; Clinical Director, Pathology &amp; The Kennedy Galton Centre, Northwick Park &amp; St Mark’s Hospitals, Harrow</td>
</tr>
</tbody>
</table>

## Pharmaceuticals Panel

<table>
<thead>
<tr>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chair,</strong> Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Oxford Radcliffe Hospital</td>
</tr>
<tr>
<td>Dr Christopher Cates, GP and Cochrane Editor, Bashey Health Centre, Bushey, Herts.</td>
</tr>
<tr>
<td>Mr Charles Dobson, Special Projects Adviser, Department of Health</td>
</tr>
<tr>
<td>Dr Robin Ferrier, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham</td>
</tr>
<tr>
<td>Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff</td>
</tr>
<tr>
<td>Dr Christine Hine, Consultant in Public Health Medicine, Bristol South &amp; West Primary Care Trust</td>
</tr>
<tr>
<td>Dr Frances Rothlat, CPMP Delegate, Medicines Control Agency, London</td>
</tr>
<tr>
<td>Mrs Sharon Hart, Managing Editor, Drug &amp; Therapeutics Bulletin, London</td>
</tr>
<tr>
<td>Professor Robert Pevler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton</td>
</tr>
<tr>
<td>Mrs Katrina Sinister, New Products Manager, National Prescribing Centre, Liverpool</td>
</tr>
<tr>
<td>Dr Ken Stein, Senior Lecturer in Public Health, University of Exeter</td>
</tr>
<tr>
<td>Professor Terence Stephenson, Professor of Child Health, University of Nottingham</td>
</tr>
<tr>
<td>Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry, London</td>
</tr>
<tr>
<td>Professor Dame Jennifer Wilson-Barnett, Head of Florence Nightingale School of Nursing &amp; Midwifery, King’s College, London</td>
</tr>
</tbody>
</table>

Current and past membership details of all HTA ‘committees’ are available from the HTA website (www.ncchta.org)
Therapeutic Procedures Panel

Members

Chair, 
Professor Bruce Campbell, 
Consultant Vascular and 
General Surgeon, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Head of 
Clinical Support & Health 
Protection, Directorate of 
Health and Social Care (North), 
Department of Health, 
Manchester

Professor John Bond, Head of 
Centre for Health Services 
Research, University of 
Newcastle upon Tyne

Mr Michael Clancy, Consultant 
in A & E Medicine, 
Southampton General Hospital

Dr Carl E Counsell, Senior 
Lecturer in Neurology, 
University of Aberdeen

Dr Keith Dodd, Consultant 
Paediatrician, Derbyshire 
Children's Hospital, Derby

Professor Gene Feder, Professor 
of Primary Care R&D, Barts & 
the London, Queen Mary’s 
School of Medicine and 
Dentistry, University of London

Ms Bec Hanley, Freelance 
Consumer Advocate, 
Hurstpierpoint, West Sussex

Professor Alan Horwich, 
Director of Clinical R&D, The 
Institute of Cancer Research, 
London

Dr Phillip Leech, Principal 
Medical Officer for Primary 
Care, Department of Health, 
London

Mr George Levy, Chief 
Executive, Motor Neurone 
Disease Association, 
Northampton

Professor James Lindesay, 
Professor of Psychiatry for the 
Elderly, University of Leicester

Dr Mike McGovern, Senior 
Medical Officer, Heart Team, 
Department of Health, London

Dr John C Pounsford, 
Consultant Physician, North 
Bristol NHS Trust

Professor Mark Sculpher, 
Professor of Health Economics, 
Institute for Research in the 
Social Sciences, University of 
York

Dr L David Smith, Consultant 
Cardiologist, Royal Devon & 
Exeter Hospital

Professor Norman Waugh, 
Professor of Public Health, 
University of Aberdeen

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)
Expert Advisory Network

Members

Mr Gordon Aylward,
Chief Executive,
Association of British Health-Care Industries, London
Ms Judith Brodie,
Head of Cancer Support Service, Cancer BACUP, London
Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury, Bucks
Ms Tracy Bury,
Project Manager, World Confederation for Physical Therapy, London
Mr John A Cairns,
Professor of Health Economics, Health Economics Research Unit, University of Aberdeen
Professor Howard Stephen Cackle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds
Professor Nicky Cullum,
Director of Centre for Evidence Based Nursing, University of York
Dr Katherine Darton,
Information Unit, MIND – The Mental Health Charity, London
Professor Carol Dezateaux,
Professor of Paediatric Epidemiology, London
Professor Martin Eccles,
Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne
Professor Pam Enderby,
Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield
Mr Leonard R Emwick,
Chief Executive, Newcastle upon Tyne Hospitals NHS Trust
Professor David Field,
Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust
Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield, West Sussex
Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Services, West Middlesex University Hospital, Isleworth, Middlesex
Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol
Professor Robert E Hawkins,
CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester
Professor F D Richard Hobbs,
Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham
Professor Allen Hutchinson,
Director of Public Health & Deputy Dean of SchARR, Department of Public Health, University of Sheffield
Professor Rajan Madhok,
Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York
Professor David Mant,
Professor of General Practice, Department of Primary Care, University of Oxford
Professor Alexander Markham,
Director, Molecular Medicine Unit, St James’s University Hospital, Leeds
Dr Chris McColl,
General Practitioner, The Hadleigh Practice, Castle Mullen, Dorset
Professor Alistair McGuire,
Professor of Health Economics, London School of Economics
Dr Peter Moore,
Freelance Science Writer, Ashtead, Surrey
Dr Andrew Mortimore,
Consultant in Public Health Medicine, Southampton City Primary Care Trust
Dr Sue Moss,
Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton, Surrey
Professor Jon Nicholl,
Director of Medical Care Research Unit, School of Health and Related Research, University of Sheffield
Mrs Julietta Patnick,
National Co-ordinator, NHS Cancer Screening Programmes, Sheffield
Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester
Ms Marianne Rigge,
Director, College of Health, London
Professor Sarah Stewart-Brown,
Director HSUR/Honorary Consultant in PH Medicine, Department of Public Health, University of Oxford
Professor Ala Szczepura,
Professor of Health Service Research, Centre for Health Services Studies, University of Warwick
Dr Ross Taylor,
Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen
Mrs Joan Webster,
Consumer member, HTA – Expert Advisory Network

Current and past membership details of all HTA ‘committees’ are available from the HTA website (www.ncchta.org)
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.*