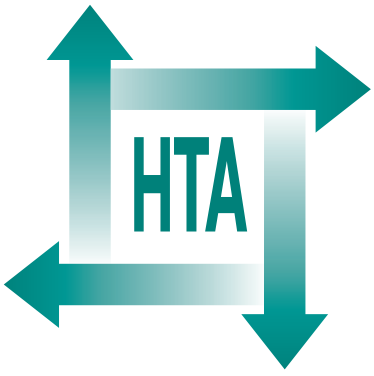
Health Technology Assessment 2003; Vol. 7: No. 41

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

D Moher
B Pham
ML Lawson
TP Klassen

Health Technology Assessment
NHS R&D HTA Programme



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

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

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

You can order HTA monographs from our Despatch Agents:

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

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

Contact details are as follows:

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

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

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

Payment methods

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

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

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

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

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

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

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 Lawson[3]
TP Klassen[5]

[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 Medica*/EMBASE.

NHS R&D HTA Programme

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

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

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.

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

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

M

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 Lawson[3] and TP Klassen[5]

[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

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

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 interventions 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 a 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.[1] 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.[2] In 1999, approximately 1250 publications were identified with the term 'meta-analysis' and 'systematic review' as a medical subject heading (MeSH).[3]

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.[4] 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.[5] 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.[6] 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[7] examined the same question and reported similar results to those of Liberati and colleagues.[6] 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.[8] The relationship between conduct and reporting will be further understood once additional research findings are forthcoming.

A survey[9] 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.[10,11]

Evidence from the spinal manipulation,[12] pain[13] and primary care literature[14] suggests considerable room for improvement in how systematic reviews are conducted and reported. Silagy[14] 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,[24] the inclusion of a German article published in a Swiss journal[25] 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[26] of this systematic review reveals at least two problems not reported by Grégoire and colleagues.

One of the EL reports[27] included in the original systematic review[24] was not an RCT. Brun-Buisson and colleagues[27] 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.[28] Similarly, the German report,[25] 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[22] 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,[29] 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[22] 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[22] 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[18] 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[30] 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[31] 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[32] 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[33] used stratified sampling to select randomly[28] 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[32] 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.[21,34] Studies such as these suggest that 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 fulfilment 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*

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

continued

TABLE I *Search strategies to identify systematic reviews and meta-analyses included in (A) the CRG database and (B) EMBASE (cont'd)*

Search no.	Set	Search terms
(B) EMBASE		
2	7	(integrative research review$ or research integration$).tw.
	8	quantitativ$ synthesis$.tw.
	9	1 or 2 or 3 or 4 or 5 or 7 or 8
	10	randomised controlled trial.de.
	11	clinical trial.de.
	12	controlled study.de.
	13	major clinical study.de.
	14	random$.hw,tw.
	15	"0197".tg.
	16	"0150".tg.
	17	or/10-16
	18	9 and 17
	19	1 or 2
	20	10 or 11 or 14
	21	19 and 20
	22	limit 21 to yr=1985-1999

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]

Quality assessment of the included RCTs

All included trials were quality assessed using a validated scale.[20] 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.[21] 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.[30,34,35]

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.[39] 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 language of publication has 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[21] and used by others.[30,34,35] 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 systematic review. 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.[35] With our modelling conventions, an ROR >1.0 indicates that EL trials tend to report a smaller protective effect compared with LOE trials, on average. RORs and their 95% CIs were derived from the fitted model.

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[40] 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.[41–43] 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.[44,45] The shift in significant funnel plot asymmetry (i.e. detected by Mascakill and colleagues method[45]) with and without language restriction was evaluated using the McNemar test (Z-scores).[46] We plotted the Z-scores testing for funnel plot asymmetry (i.e. the intercept of the

fitted regression line by Macaskill and colleagues[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.[47] The H statistic describes the relative excess in the test for heterogeneity (i.e. in the sense of the DerSimonian–Laird RE model[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.[47] Higgins and Thompson[47] also suggest that values of $I^2 > 56\%$ (or H values > 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

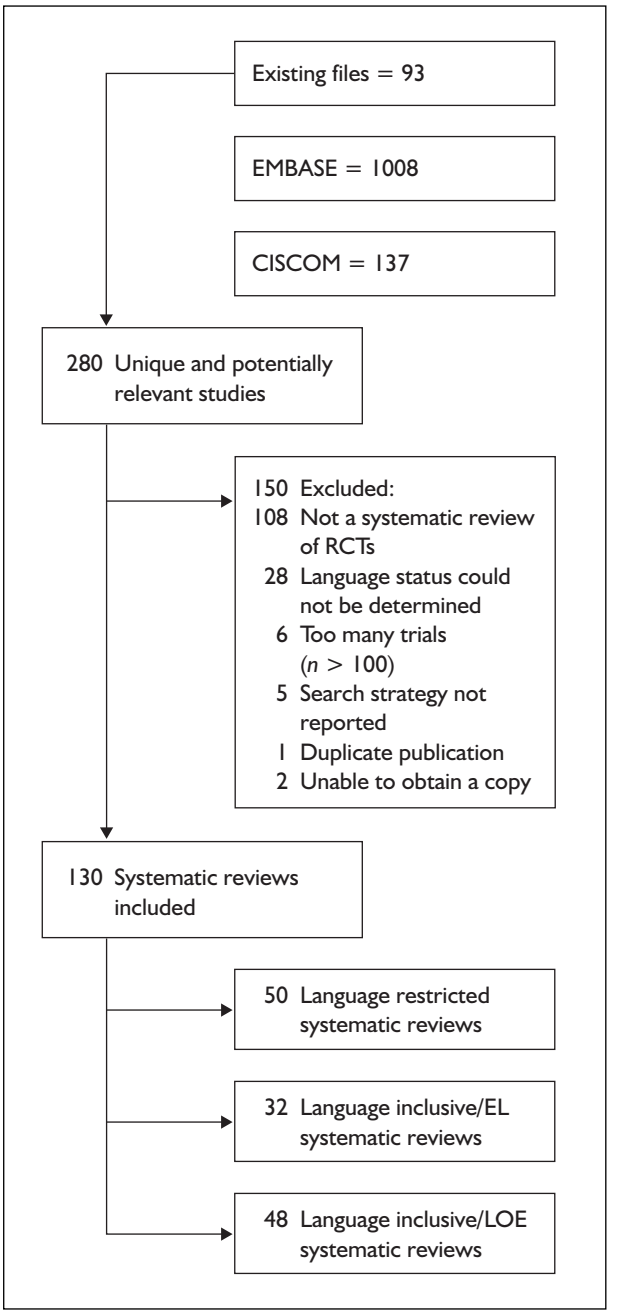


FIGURE 1 *Flow diagram of phases in the inclusion of systematic reviews*

TABLE 2 *General characteristics of language restricted systematic reviews, language inclusive/EL systematic reviews and, language inclusive/LOE systematic reviews*

	Language restricted systematic reviews (*n* = 50): median [1st, 3rd Q[a]] (range)	**Language inclusive/EL systematic reviews (*n* = 32): median [1st, 3rd Q[a]] (range)**	**Language inclusive/LOE systematic reviews (*n* = 48): Median [1st, 3rd Q[a]] (range)**
Year of publication	1994 [1993, 1994] (1989, 1998)	1995 [1994, 1996] (1990, 1999)	1994 [1993, 1996] (1989, 1999)
Number of RCTs	11 [6, 23] (4, 106)	8 [6, 13] (3, 26)	17 [9, 25] (3, 119)
Total number of patients	971 [419, 2641] (112, 52869)	1121 [352, 2385] (150, 361433)	1658 [798, 3604] (112, 40431)
Citation impact[b]	1.67 [0.99, 4.04] (0.08, 15.94)	2.08 [0.71, 4.44] (0.13, 15.94)	1.96 [1.27, 4.34] (0.70, 15.94)
Number of LOE trials			2 [1, 4] (1, 53)
Proportion of LOE trials (%)			14 [11, 36] (5, 67)
			***n* (%)**
Number of languages			9
Danish			3 (0.4)
Dutch			2 (0.3)
English			546 (80)
German			57 (8)
French			52 (8)
Italian			15 (2)
Japanese			1 (0.1)
Portuguese			1 (0.1)
Spanish			6 (0.9)
Total			684 (100)
	***n* (%)**	***n* (%)**	***n* (%)**
Disease area			
Circulatory disease	18 (36)	7 (22)	12 (25)
Infectious disease	2 (4)	7 (22)	1 (2)
Digestive	4 (8)	2 (6)	8 (17)
Pregnancy and childbirth	3 (6)	4 (13)	5 (11)
Genitourinary	2 (4)	2 (6)	6 (13)
Mental health	2 (4)	2 (6)	5 (10)
Nervous system and sense organs	4 (8)	1 (3)	1 (2)
Neoplasms	3 (6)	1 (3)	2 (4)
Respiratory	2 (4)	1 (3)	1 (2)
Other	10 (20)	5 (16)	7 (15)
Type of journal			
General	14 (28)	15 (47)	23 (48)
Specialty	36 (72)	17 (53)	25 (52)
Type of interventions			
CM	39 (78)	28 (88)	38 (79)
CAM	11 (22)	4 (13)	10 (21)
Funding source			
Non-pharmaceutical company	22 (44)	15 (47)	17 (35)
None listed/cannot tell	27 (54)	17 (53)	30 (63)
Single pharmaceutical company	1 (2)		1 (2)

[a] Q = quartile.
[b] SCI Journal Citation Report, 1992.

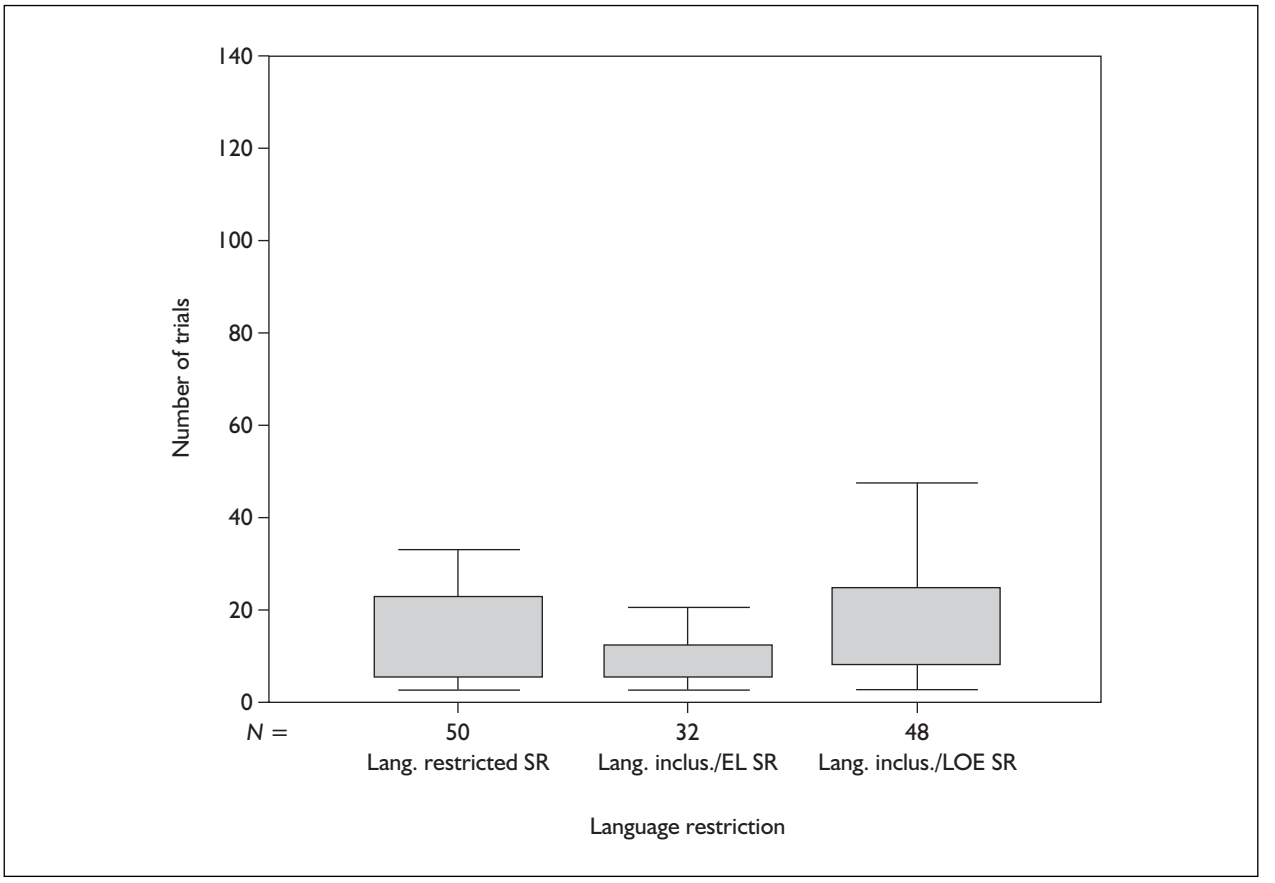


FIGURE 2 *The median (and inter-quartile range) number of reports of RCTs included in language restricted systematic reviews, language inclusive/EL systematic reviews and language inclusive/LOE systematic reviews*

Figure 2). 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/LE, 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%);

TABLE 3 *General characteristics of 105 systematic review examining CM interventions and 25 systematic reviews investigating CAM interventions*

	CM systematic reviews (n = 105): median [1st, 3rd Q[a]] (range)	**CAM systematic reviews (n = 25): median [1st, 3rd Q[a]] (range)**
Year of publication	1994 [1993, 1995] (1989, 1999)	1995 [1994, 1997] (1989, 1999)
Number of RCTs	12 [6, 22] (3, 106)	14 [8, 21] (3, 119)
Total number of patients	1407 [743, 3556] (112, 361433)	793 [184, 1573] (112, 10523)
Citation impact[b]	1.92 [1.22, 4.34] (0.13, 15.94)	1.44 [0.72, 3.02] (0.08, 15.94)
	***n* (%)**	***n* (%)**
Disease area		
Circulatory disease	28 (27)	9 (36)
Infectious disease	10 (10)	
Digestive	12 (11)	2 (8)
Pregnancy and childbirth	11 (11)	1 (4)
Genitourinary	7 (7)	3 (12)
Mental health	7 (7)	2 (8)
Central nervous and sense	4 (4)	2 (8)
Neoplasms	6 (6)	
Respiratory	4 (4)	
Muscular skeletal and connective tissues	2 (2)	
Other	14 (13)	3 (12)
Type of journal		
General	42 (40)	10 (40)
Specialty	63 (60)	15 (60)
Funding source		
Non-pharmaceutical company	41 (39)	13 (52)
None listed/cannot tell	62 (59)	12 (48)
Pharmaceutical company	2 (2)	

[a] Q = quartile.
[b] SCI Journal Citation Report, 1992.

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/LOE systematic reviews, 47% of the language inclusive/EL systematic reviews and 44% 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 between 1989 and 1999.

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*).

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)*

Question	Language restricted systematic reviews (*n* = 50)	Language inclusive/EL systematic reviews (*n* = 32)	Language inclusive/LOE systematic reviews (*n* = 48)	Two-sided *p*-value[a]
1. Were the search methods used to find evidence reported?	30 (60)	14 (44)	24 (50)	0.35
2. Was the search for evidence reasonably comprehensive?	24 (48)	9 (28)	25 (52)	0.09
3. Were the criteria for deciding which studies to include in the overview reported?	40 (80)	27 (84)	36 (75)	0.60
4. Was bias in the selection of studies avoided?	13 (26)	10 (31)	11 (23)	0.69
5. Were the criteria used for assessing the validity of the included studies reported?	18 (36)	11 (34)	33 (69)	0.01
6. Was the validity of all of the studies referred to in the text assessed using appropriate criteria?	14 (28)	11 (34)	30 (63)	0.01
7. Were the methods to combine the findings of the relevant studies reported?	41 (82)	24 (75)	37 (77)	0.76
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addresses?	43 (86)	24 (75)	38 (79)	0.42
9. Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?	40 (80)	25 (78)	39 (81)	0.96
10. How would you rate the scientific quality of this overview?[b]	3 [3, 5.25]	3 [2, 4]	4 [3, 5]	0.25[c]

[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. (%)*

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

continued

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

Characteristics	Language restricted systematic reviews (*n* = 50)	Language inclusive/EL systematic reviews (*n* = 32)	Language inclusive/LOE systematic reviews (*n* = 48)	Two-sided *p*-value
Independent review for study selection				
Single reviewer	0	4 (13)	0	0.01
More than one reviewer	21 (42)	12 (38)	13 (27)	
Inter-reviewer reliability	5 (10)	5 (16)	3 (6)	
Not reported	29 (58)	16 (50)	35 (73)	
Independent data extraction				
One reviewer	2 (4)	2 (6)	1 (2)	0.17
More than one reviewer	18 (36)	16 (50)	27 (56)	
Used a standard form for data extraction	10 (20)	3 (10)	5 (10)	
Not reported	30 (60)	14 (44)	18 (38)	
Assessment of statistical heterogeneity	37 (74)	28 (88)	40 (83)	0.30
Reporting a chi-squared test for heterogeneity	34 (68)	25 (78)	37 (77)	
Using L'Abbé plot	4 (8)	4 (12)	1 (2)	
Investigation of clinical heterogeneity	23 (46)	13 (41)	18 (38)	0.67
Using subgroup analyses	17 (34)	9 (28)	17 (35)	
Using sensitivity analyses	3 (6)	3 (9)	6 (13)	
Using covariates	4 (8)	4 (13)	1 (2)	
Using control rate	1 (2)		1 (2)	
Statistical model				
Fixed effects	31 (62)	15 (47)	25 (52)	0.46
Random effects	9 (18)	12 (38)	12 (25)	
Both fixed and random effects	6 (12)	2 (6)	4 (8)	
Cannot tell	4 (8)	3 (9)	7 (15)	
Inclusion of grey literature	3 (6)	4 (13)	19 (40)	0.01
Number of grey items				
One	2 (4)	1 (3)	6 (13)	
Two			4 (8)	
More than two	1 (2)	2 (6)	9 (19)	
Sources				
Abstracts	1 (2)	1 (3)	11 (23)	
Unpublished		2 (6)	7 (15)	
In press			1 (2)	
Thesis			1 (2)	
Book chapters	2 (4)		2 (4)	
Company reports			4 (8)	
Assessment of publication bias	7 (14)	5 (16)	8 (17)	0.92
Decision to assess publication bias				
A priori	1 (2)		7 (15)	
Posteriori	6 (12)	5 (16)	1 (2)	
Method used				
Fail-safe	5 (10)	5 (16)	6 (13)	
Funnel plot		1 (3)	2 (4)	
Selection model			1 (2)	
Discussed impact of publication bias on results	16 (32)	11 (34)	16 (33)	0.97
Conclusion				
Positive/significant	33 (66)	17 (53)	32 (67)	0.18
Negative/non-significant	10 (20)	8 (25)	13 (27)	
Unclear	7 (14)	7 (22)	2 (4)	

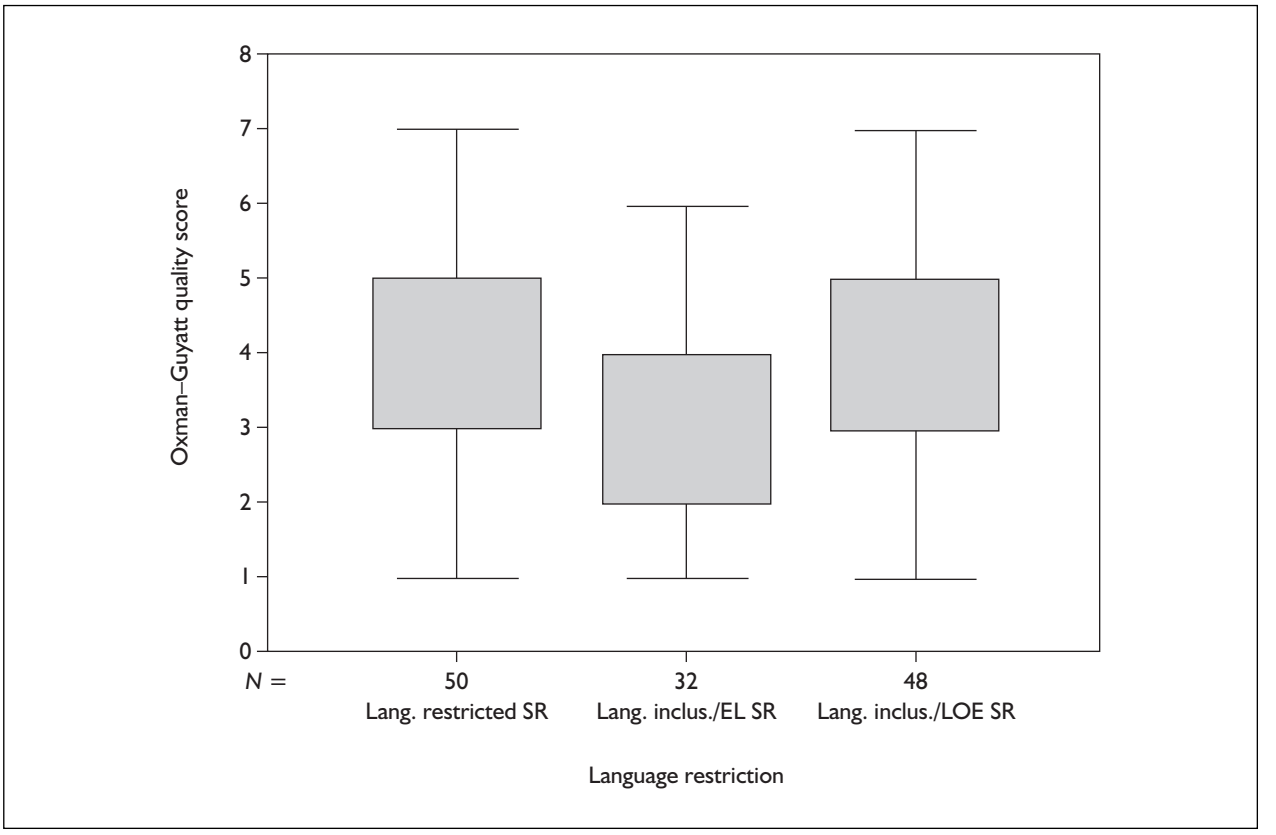


FIGURE 3 *Quality of reporting according to type of systematic review (language restricted systematic reviews, language inclusive/EL systematic reviews and language inclusive/LOE systematic reviews)*

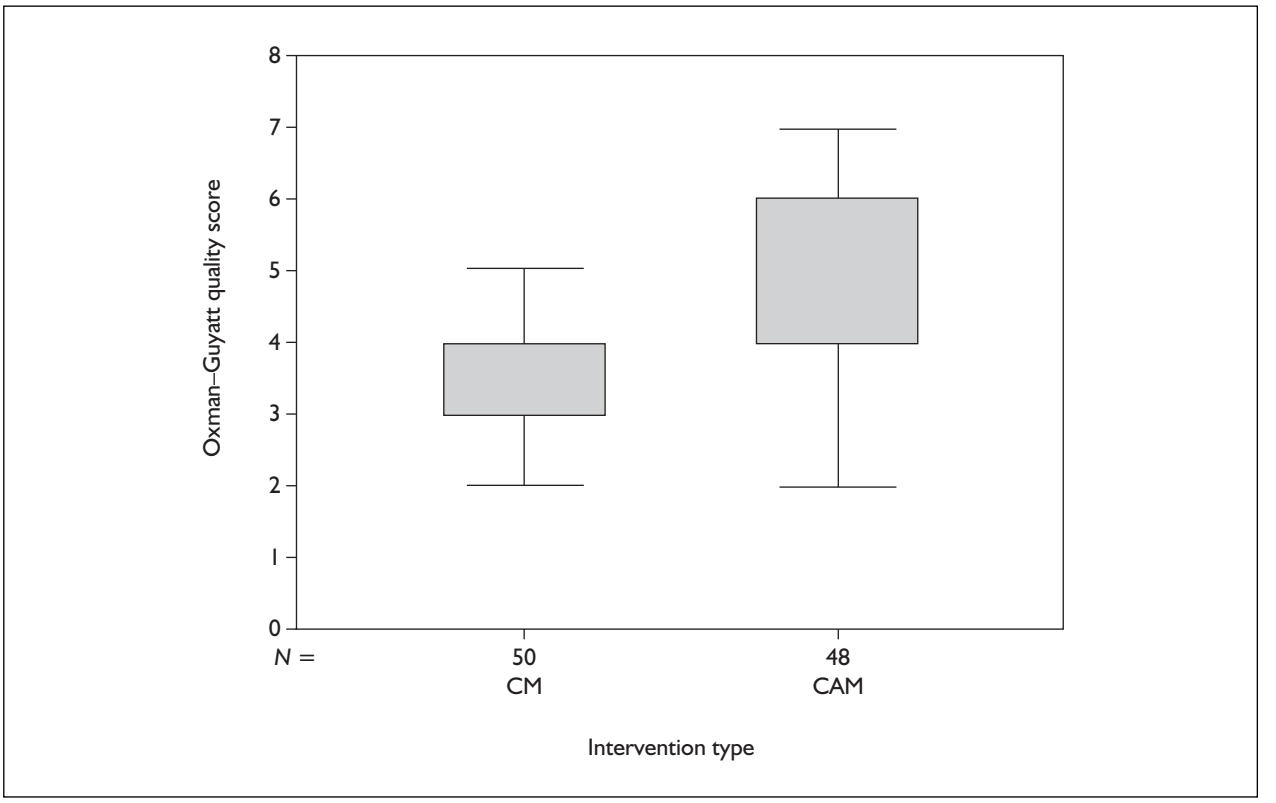


FIGURE 4 *Quality of reporting of systematic reviews and type of interventions*

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[48] 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[49] 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 *a priori* 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[50] 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

TABLE 6 *Quality of reports of CM and CAM systematic reviews: no. (%) (see text for details)*

Question	CM (n = 105)	CAM (n = 25)	Two-sided p-value[a]
1. Were the search methods used to find evidence reported?	51 (49)	17 (68)	0.12
2. Was the search for evidence reasonably comprehensive?	42 (40)	16 (64)	0.04
3. Were the criteria for deciding which studies to include in the overview reported?	78 (74)	25 (100)	0.02
4. Was bias in the selection of studies avoided?	27 (26)	7 (28)	0.80
5. Were the criteria used for assessing the validity of the included studies reported?	42 (40)	20 (80)	0.003
6. Was the validity of all of the studies referred to in the text assessed using appropriate criteria?	38 (36)	17 (68)	0.06
7. Were the methods to combine the findings of the relevant studies reported?	81 (77)	21 (84)	0.59
8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addresses?	83 (79)	22 (88)	0.41
9. Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?	81 (77)	23 (92)	0.16
10. How would you rate the scientific quality of this overview?[b]	3 [3, 4.5]	5 [4, 6.5]	0.01[c]

[a] Fisher's exact test.
[b] Median [inter-quartile range].
[c] Wilcoxon rank sum test.

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[51] 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*

Characteristics	CM (*n* = 105)	CAM (*n* = 25)	Two-sided *p*-value
Search strategy			
MEDLINE	87 (83)	25 (100)	0.02
EMBASE	14 (13)	7 (28)	0.13
Hand search	11 (11)	3 (12)	0.73
Reference lists	60 (57)	16 (64)	0.65
Corresponding authors	20 (19)	10 (40)	0.04
Content experts	13 (12)	4 (16)	0.74
Abstracts	12 (11)	3 (12)	1.00
Conference proceedings	7 (7)	2 (8)	0.68
Other electronic databases	25 (24)	13 (52)	0.08
Cannot tell	8 (8)	1 (4)	1.00
Independent review for study selection			
Single reviewer	4 (4)	0 (0)	1.00
More than one reviewer	37 (36)	9 (36)	
Inter-reviewer reliability	8 (8)	5 (20)	
Not reported	63 (61)	16 (64)	
Independent data extraction			
One reviewer	4 (4)	1 (4)	0.85
More than one reviewer	47 (45)	13 (52)	
Used of a standard form for data extraction	14 (14)	4 (16)	
Not reported	53 (51)	11 (44)	
Assessment of statistical heterogeneity	83 (80)	22 (88)	0.41
Reporting a chi-squared test for heterogeneity	77 (74)	19 (76)	
Using L'Abbé plot	7 (7)	2 (8)	
Investigation of clinical heterogeneity	44 (42)	10 (40)	1.00
Using subgroup analyses	34 (33)	9 (36)	
Using sensitivity analyses	11 (11)	1 (4)	
Using covariates	8 (8)	1 (4)	
Using control rate association	2 (2)		
Statistical model			
Fixed effects	59 (56)	12 (48)	0.51
Random effects	24 (23)	9 (36)	
Both fixed and random effects	11 (11)	1 (4)	
Cannot tell	11 (11)	3 (12)	
Inclusion of grey literature	22 (21)	3 (12)	0.41
Number of grey items			
One	8 (8)	1 (4)	
Two	3 (3)	1 (4)	
More than two	11 (11)	1 (4)	
Sources			
Abstracts	11 (11)	2 (8)	
Unpublished	8 (8)	1 (4)	
In press		1 (4)	
Thesis	1 (1)		
Book chapters	3 (3)	1 (4)	
Company reports	3 (3)	1 (4)	
Assessment of publication bias	16 (15)	4 (16)	1.00
Decision to assess publication bias			
A priori	5 (5)	3 (12)	
Posteriori	11 (11)	1 (4)	
Method used			
Fail-safe	16 (15)		
Funnel plot	1 (1)	2 (8)	
Selection model		1 (4)	

continued

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

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

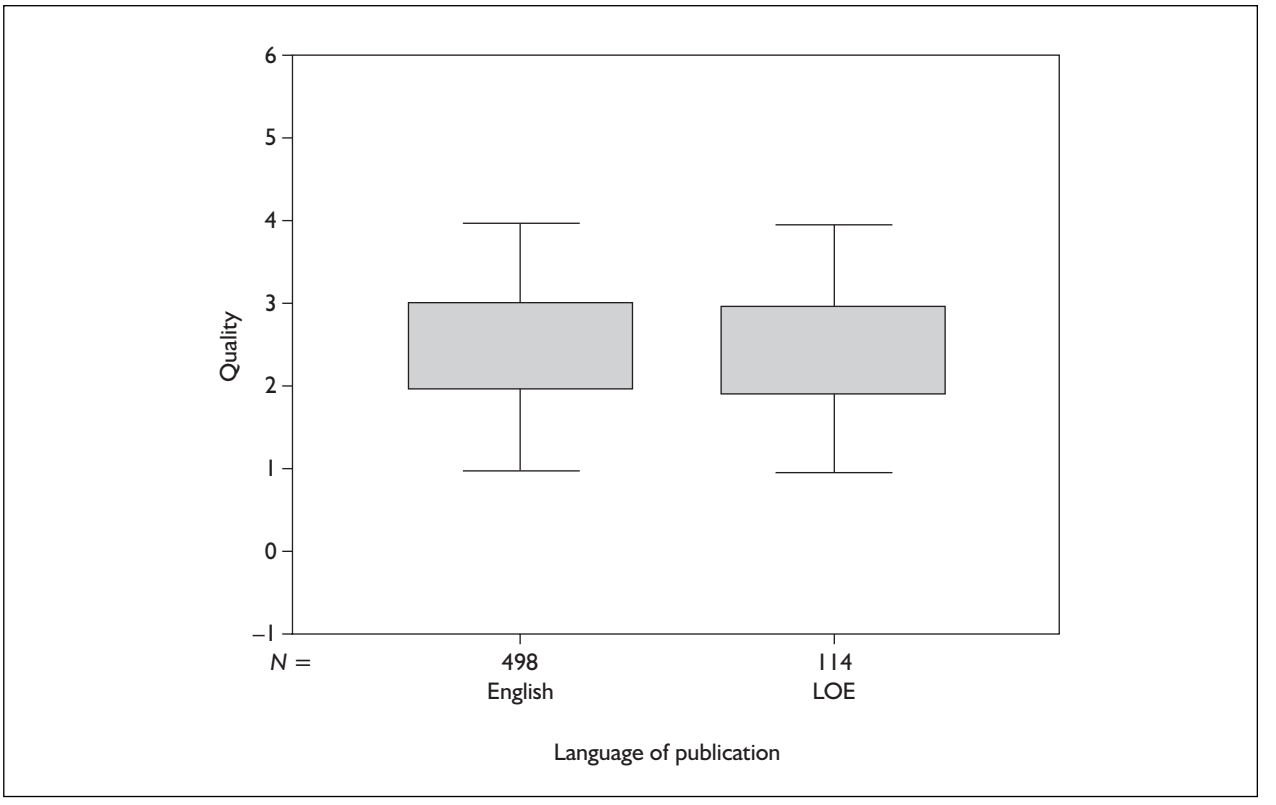


FIGURE 5 *Effect of language of publication on quality of reporting of RCTs*

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

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 8 *Language of publication other than English*

Language	*n* (%)	CM *n* (%)	CAM *n* (%)
Danish	3 (2.6)	1 (2.1)	2 (3)
Dutch	1 (0.9)	0	1 (1.5)
French	46 (40.4)	20 (42.6)	26 (38.8)
German	50 (43.9)	15 (31.9)	35 (52.2)
Italian	8 (7)	5 (10.6)	3 (4.5)
Japanese	1 (0.9)	1 (2.1)	0
Spanish	5 (4.4)	5 (10.6)	0
Total	114 (100)	47	67

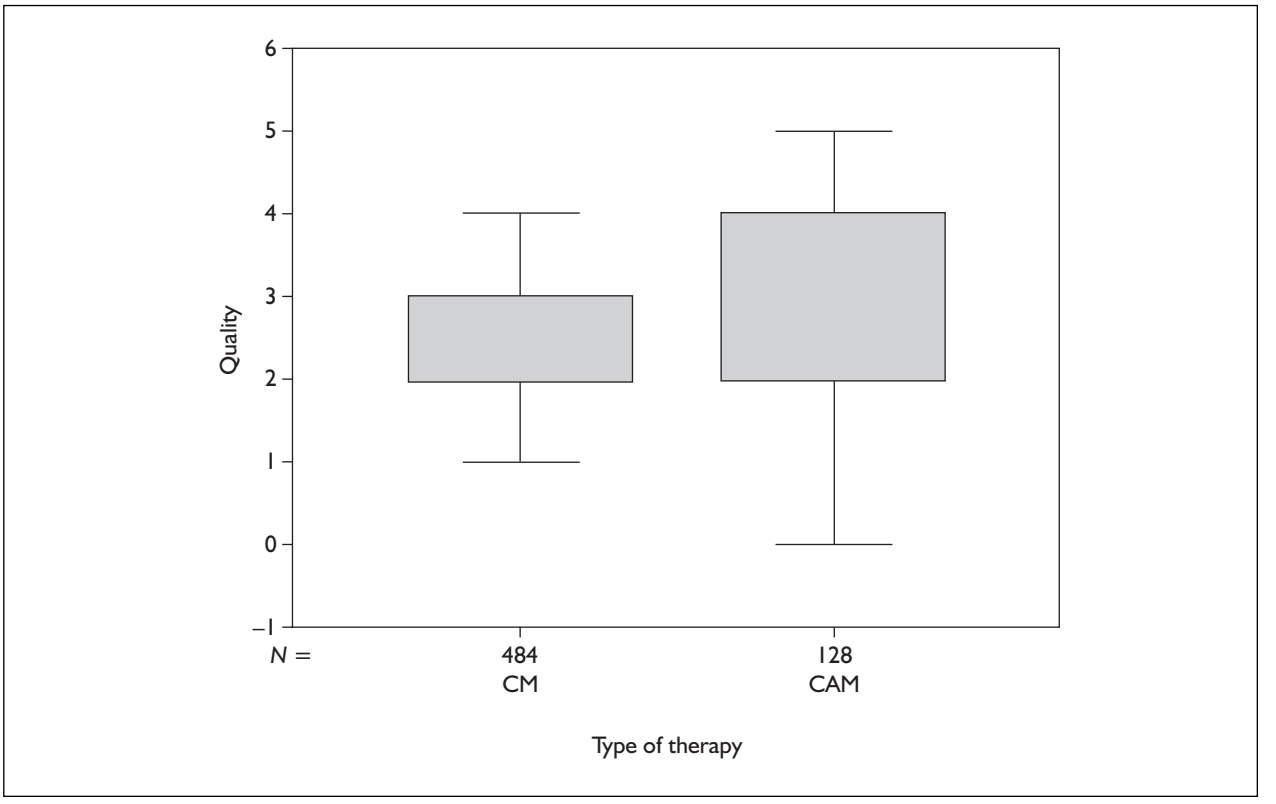


FIGURE 6 *Effect of type of intervention on quality of reporting of RCTs*

less likely to report double blinding (62% versus 71%, $p = 0.07$). Overall, the quality of reporting of EL trials appeared to be slightly better than that of LOE trials (median quality score of 3 versus 2, $p = 0.1$) (*Figure 6*). Allocation concealment was poorly reported in both the English and LOE trials, with 84–85% being scored as inadequate or unclear.

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 CAM 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*

Quality of report	Language of publication		*p*-Value	Type of therapy		*p*-Value
	English (*n* = 498): *n* (%)	LOE (*n* = 114): *n* (%)		CM (*n* = 484): *n* (%)	CAM (*n* = 128): *n* (%)	
Randomisation	447 (90)	90 (79)	0.003	434 (90)	103 (81)	0.001
Double-blind	307 (62)	81 (71)	0.07	273 (56)	115 (90)	<0.01
Withdrawal, lost to follow-up	314 (63)	50 (44)	<0.001	306 (63)	58 (45)	<0.01
Quality score						
Low (0–2)	248 (50)	63 (55)	0.30	252 (52)	59 (46)	0.20
High (3–5)	250 (50)	51 (45)		232 (48)	69 (54)	
Median [IQR] (min., max.)	3 [2, 3] (0, 5)	2 [1.75, 3] (0, 5)	0.10	2 [2, 3] (0, 5)	3 [2, 4] (0, 5)	0.14
	(*n* = 490)	**(*n* = 103)**		**(*n* = 481)**	**(*n* = 112)**	
Allocation concealment						
Adequate	75 (15)	16 (16)	0.59	60 (13)	31 (28)	<0.001
Inadequate	32 (7)	4 (4)		30 (6)	6 (5)	
Unclear	383 (78)	83 (80)		391 (79)	75 (67)	

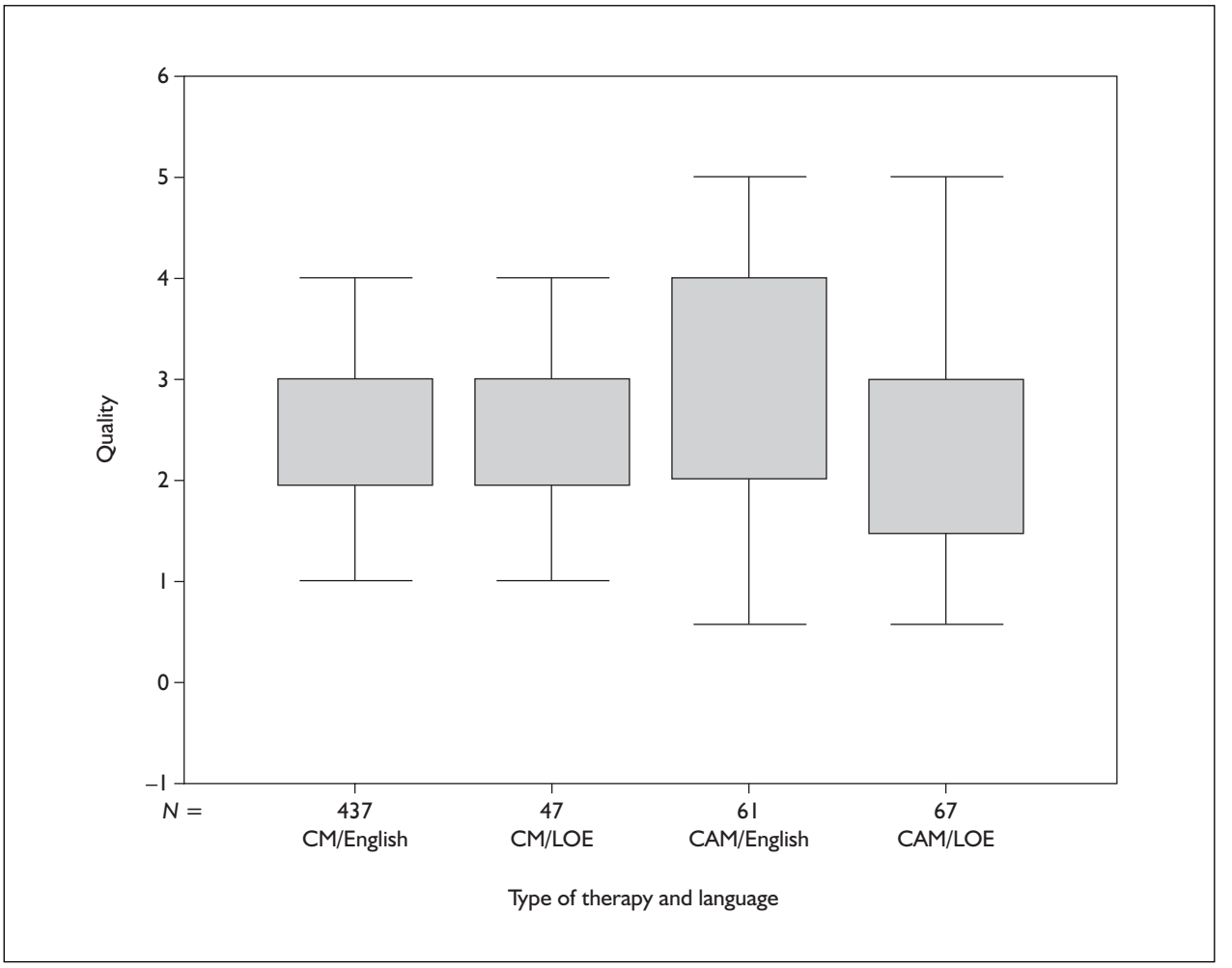


FIGURE 7 *Effect of language of publication on quality of reporting for CM and CAM RCTs*

TABLE 10 *Effect of language of publication on quality of reporting of RCTs in CM and CAM RCTs*

Quality of report	CM		*p*-Value	CAM		*p*-Value
	English (*n* = 437): *n* (%)	LOE (*n* = 47): *n* (%)		English (*n* = 61): *n* (%)	LOE (*n* = 67): *n* (%)	
Randomisation	395 (90)	39 (83)	0.13	52 (85)	51 (76)	0.27
Double-blind	250 (57)	23 (50)	0.29	57 (93)	58 (87)	0.25
Withdrawal, lost to follow-up	279 (64)	27 (57)	0.43	35 (57)	23 (34)	0.01
Quality score						
Low (0–2)	233 (52)	28 (60)	0.23	24 (39)	35 (52)	0.16
High (3–5)	218 (48)	19 (40)		37 (61)	32 (48)	
Median [IQR] (min., max.)	2 [2, 3] (0, 5)	2 [2, 3] (0, 5)	0.12	3 [2, 4] (1, 5)	2 [1, 3] (0, 5)	0.04
	(*n* = 447)	(*n* = 49)		(*n* = 56)	(*n* = 56)	
Allocation concealment						
Adequate	58 (13)	2 (4)	0.07	17 (30)	14 (25)	0.82
Inadequate	29 (7)	1 (2)		3 (5)	3 (5)	
Unclear	347 (80)	44 (94)		36 (64)	39 (70)	

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 11 *Citations and descriptive characteristics of language inclusive/LOE systematic reviews*[a]

ID	Ref. in App. 3	First author	Title	Journal	Year	Intervention	Trials/patients	Outcome
1	16	Hayashi K	Famotindine in the treatment of duodenal ulcer	*Gastroenterol Int*	1993	CM	6/831	Healing rate at 6 weeks
2	24	Malaguarnera M	Interferon-alpha treatment in patients with chronic hepatitis C	*Clin Drug Invest*	1995	CM	26/1490	Complete response
3	29	Pittler MH	Peppermint oil for irritable bowel syndrome	*Am J Gastroenterol*	1998	CAM	5/239	Global improvement
4	6	Carroll D	Transcutaneous electrical nerve stimulation in labour pain	*Br J Obstet Gynaecol*	1997	CAM	4/475	Additional analgesic used
5	41	Wilt TJ	Saw Palmetto extracts for treatment of benign prostatic hyperplasia	*JAMA*	1998	CAM	6/659	Improvement in symptoms self-rating
6	23	Linde K	Are the clinical effects of homeopathy placebo effects?	*Lancet*	1997	CAM	74/8778	Binary efficacy outcome measures
7	39	White AR	Acupuncture for smoking cessation	Cochrane	1998	CAM	12/1947	Early abstinence
8	11	Douglas RM	Vitamin C for the common cold	Cochrane	1998	CM	18/8010	One or more respiratory episodes
9	19	Hofmeyr GJ	External cephalic version at term	Cochrane	1996	CM	6/612	Non-cephalic birth
10	18	Hofmeyr GJ	Cephalic version by postural management	Cochrane	1996	CM	3/192	Non-cephalic birth
11	28	Pace F	Meta-analysis of the effect of placebo on the outcome of medically treated reflux esophagitis	*Scand J Gastroenterol*	1995	CM	22/1224	Healing at 4–6 weeks
12	31	Poynard T	Meta-analysis of smooth muscle relaxants in the treatment of irritable bowel syndrome	*Aliment Pharmacol Ther*	1994	CM	24/1713	Muscle relaxant, global improvement
13	13	Fine MJ	Efficacy of pneumococcal vaccination in adults	*Arch Intern Med*	1994	CM	6/16337	All causes mortality
14	15	Halpern S	Postdural puncture headache and spinal needle design	*Anesthesiology*	1994	CM	9/1720	Headache
15	21	Leizorovicz A	Comparison of efficacy and safety of low molecular weight heparins and unfractionated heparin in initial treatment of deep venous thrombosis	*BMJ*	1994	CM	16/2055	Mortality
16	25	Marino P	Chemotherapy vs supportive care in advance non-small-cell cancer	*Chest*	1994	CM	8/689	Mortality at 6 months
17	14	Glowacki LS	Use of immune globulin to prevent symptomatic cytomegalovirus disease in transplant recipients – a meta-analysis	*Clin Transplant*	1994	CM	17/1016	Symptomatic cytomegalovirus disease
18	33	SDD Trialist Group	Meta-analysis of RCTs of selective decontamination of the digestive tract	*BMJ*	1993	CM	22/3836	Respiratory tract infection

continued

TABLE 11 *Citations and descriptive characteristics of language inclusive/LOE systematic reviews[a] (cont'd)*

ID	Ref. in App. 3	First author	Title	Journal	Year	Intervention	Trials/patients	Outcome
19	40	Wilson AP	A meta-analysis of the use of amoxycillin-clavulanic acid in surgical prophylaxis	*J Hosp Infect*	1992	CM	20/4653	Wound infection
20	7	Cohen HJ	Comparison of two long-term chemotherapy regimens, with or without agents to modify skeletal repair, in multiple myeloma	*Blood*	1984	CM	18/3898	Mortality at 2 years
21	37	Vandekerckhove P	Androgens versus placebo or no treatment for idiopathic oligo/asthenospermia	Cochrane	1999	CM	8/908	Pregnancy
22	8	Covey LS	A meta-analysis of double-blind placebo-controlled trials of clonidine for smoking cessation	*Br J Addict*	1991	CM	9/813	Smoking cessation
23	22	Linde K	St John's wort for depression	*BMJ*	1996	CAM	13/828	Responder (placebo controlled trials)
24	22	Linde K	St John's wort for depression	*BMJ*	1996	CAM	3/317	Responder (active controlled trials)
25	26	Meijer WS	Meta-analysis of RCTs of antibiotics prophylaxis in biliary tract surgery	*Br J Surg*	1990	CM	8/1444	Wound infection
26	38	Vandekerckhove P	The medical treatment of idiopathic oligo/asthenospermia: bromocriptine versus placebo or no treatment	Cochrane	1997	CM	3/102	Pregnancy
27	1	a'Rogvi-Hansen B	Glycerol treatment for acute ischaemic stroke	Cochrane	1996	CM	6/654	Case fatality
28	30	Pouleur H	Effects of dipyridamole in combination with anticoagulant therapy on survival and thromboembolic events in patients with prosthetic heart valves	*J Thorac Cardiovasc Surg*	1995	CM	6/1151	Mortality
29	32	Poynard T	Meta-analysis of hydroxyethylrutosides in the treatment of chronic venous insufficiency	*Vasa*	1994	CM	10/1826	Pain
30	4	Bressa GM	*S*-Adenosyl-1-methionine (SAMe) as antidepressant	*Acta Neurol Scand*	1994	CAM	12/391	Partial to full response
31	17	Heyland DK	Selective decontamination of the digestive tract	*Chest*	1994	CM	24/3405	Mortality
32	35	Silagy C	The effectiveness of nicotine replacement therapies in smoking cessation	*Online J Curr Lin Trials*	1994	CM	48/16921	Smoking cessation
33	5	Brown KH	Use of nonhuman milks in the dietary management of young children with acute diarrhea	*Pediatrics*	1994	CM	14/935	Treatment failure
34	20	Leizorovicz A	Low molecular weight heparin in prevention of perioperative thrombosis	*BMJ*	1992	CM	45/12777	Deep venous thrombosis

continued

TABLE 11 *Citations and descriptive characteristics of language inclusive/LOE systematic reviews[a] (cont'd)*

ID	Ref. in App. 3	First author	Title	Journal	Year	Intervention	Trials/patients	Outcome
35	2	Blondel B	Home visits for pregnancy complications and management of antenatal care	*Br J Obstet Gynaecol*	1992	CM	3/1407	Hospital admission
36	36	Van Ruiswyk J	Efficacy of prophylactic sclerotherapy for prevention of a first variceal hemorrhage	*Gastroenterology*	1992	CM	15/1386	13-month mortality
37	3	Boissel JP	Is it possible to reduce the risk of cardiovascular events in subjects suffering from intermittent claudication of the lower limbs	*Thromb Haemost*	1989	CM	4/618	Thrombotic cardiovascular events
38	10	Daya S	Long vs short gonadotropin releasing hormone agonist protocols for pituitary desensitisation in assisted reproductive cycles	Cochrane	1998	CM	8/988	Clinical pregnancy
39	34	Sikorski J	Support for breastfeeding mothers	Cochrane	1998	CM	11/2934	Stop breastfeeding
40	27	Montgomery SA	Comparison of compliance between serotonin reuptake inhibitors and tricyclic antidepressants	*Int Clin Psychopharmacol*	1995	CM	67/6850	Side-effects
41	9	Crawford F	Tropical treatments for fungal infections of the skin and nails of the foot	Cochrane	1999	CM	5/617	Cure
42	12	Figueredo E	Prophylactic ondansetron for postoperative emesis	*Acta Anaesthesiol Scand*	1999	CM	18/3099	Emesis

[a] See Appendix 3.

TABLE 12 *The effect of language of publication of RCTs on the estimates of intervention effectiveness*

Type of analysis	No. of meta-analyses/ no. of RCTs	Language effect (trials published in English compared with trials published in LOE): ROR (95% CI)	Estimated heterogeneity between trials
Language restricted meta-analyses compared with language inclusive/LOE meta-analyses (overall)	42/662	1.09 (0.99 to 1.21)	2.76 (χ^2 = 1706, 619 df)
	FE	1.07 (0.96 to 1.18)	
	RE	1.11 (0.92 to 1.34)	88.32 (χ^2 41 df), $p < 0.0001$
Language restricted meta-analyses compared with language inclusive/LOE meta-analyses (CM)	34/533	1.01 (0.90 to 1.15)	2.44 (χ^2 =1217, 498 df)
	FE	0.99 (0.88 to 1.12)	
	RE	1.02 (0.83 to 1.26)	66.97 (χ^2 33 df), $p = 0.0004$
Limited to meta-analyses with no. of LOE > 1	17/356	0.98 (0.84 to 1.14)	2.20 (χ^2 = 745, 338 df)
	FE	0.98 (0.84 to 1.14)	
	RE	1.01 (0.77 to 1.32)	37.76 (χ^2 16 df), $p = 0.002$
Limited to meta-analyses with no. of LOE = 1	17/177	1.07 (0.88 to 1.31)	2.96 (χ^2 = 471, 159 df)
	FE	1.01 (0.83 to 1.23)	
	RE	1.06 (0.73 to 1.54)	29.14 (χ^2 16 df), $p = 0.02$
Language restricted meta-analyses compared with language inclusive/LOE meta-analyses (CAM)	8/129	1.37 (1.16 to 1.61)	4.04 (χ^2 = 485, 120 df)
	FE	1.26 (1.05 to 1.52)	
	RE	1.63 (1.03 to 2.60)	16.74 (χ^2 7 df), $p = 0.02$
Limited to meta-analyses with no. of LOE > 1	5/115	1.25 (1.03 to 1.51)	3.81 (χ^2 = 415, 109 df)
	FE	1.22 (1.01 to 1.47)	
	RE	1.34 (0.85 to 2.12)	9.19 (χ^2 4 df), $p = 0.06$
Limited to meta-analyses with no. of LOE = 1	3/14	2.67 (1.21 to 5.86)	6.60 (χ^2 = 66, 10 df)
	FE	2.52 (1.12 to 5.66)	
	RE	3.01 (0.84 to 10.85)	4.62 (χ^2 2 df), $p = 0.10$

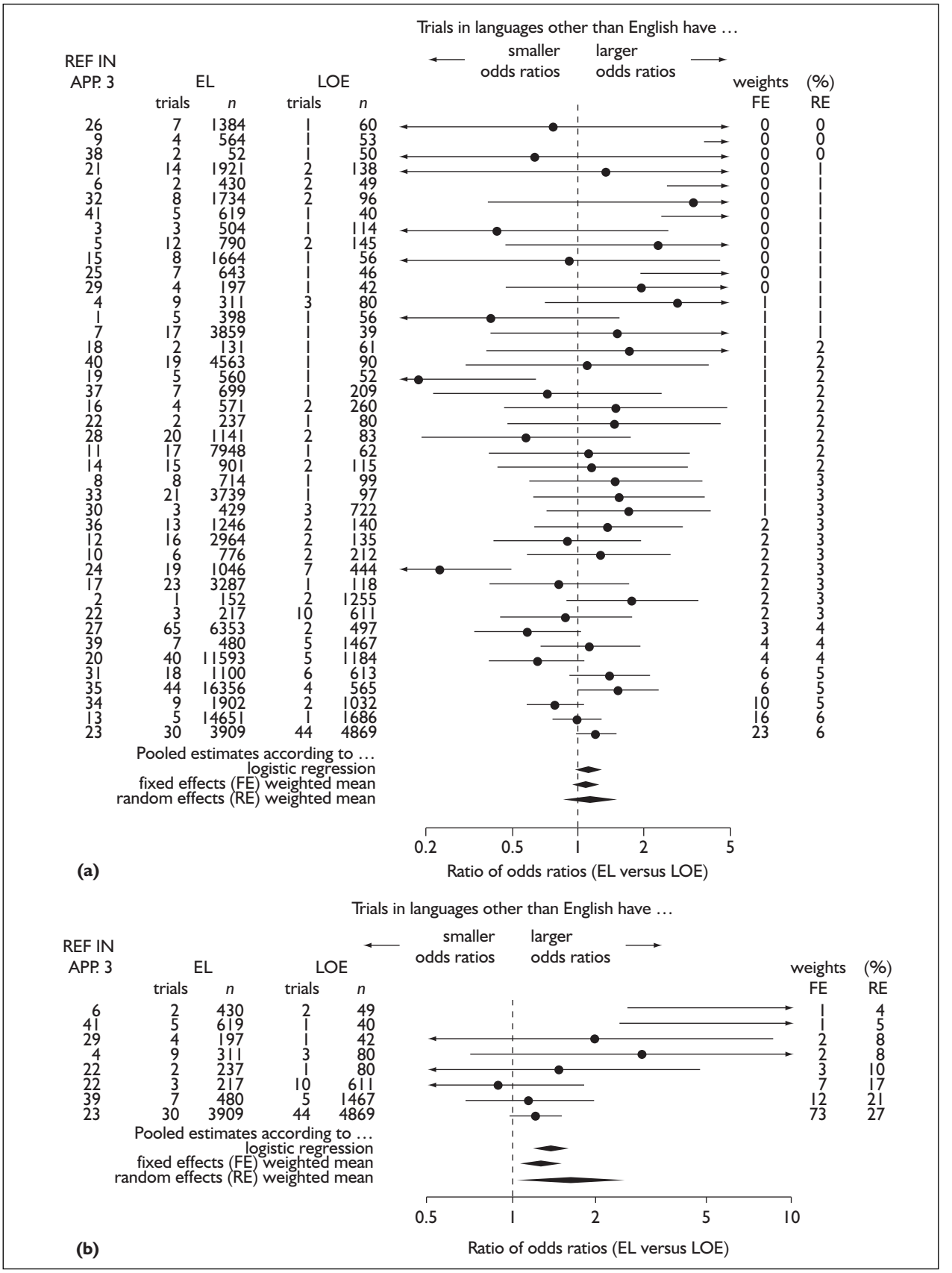


FIGURE 8 *(a) Impact of LOE at the RCT level: RORs with 95% CIs for LOE versus EL for each of 42 systematic reviews, together with pooled estimates from three different approaches (shown as diamonds). (b) Impact of LOE at the RCT level: RORs with 95% CIs for LOE versus EL for each of eight complementary and alternative medicine systematic reviews, together with pooled estimates from three different approaches (shown as diamonds).*

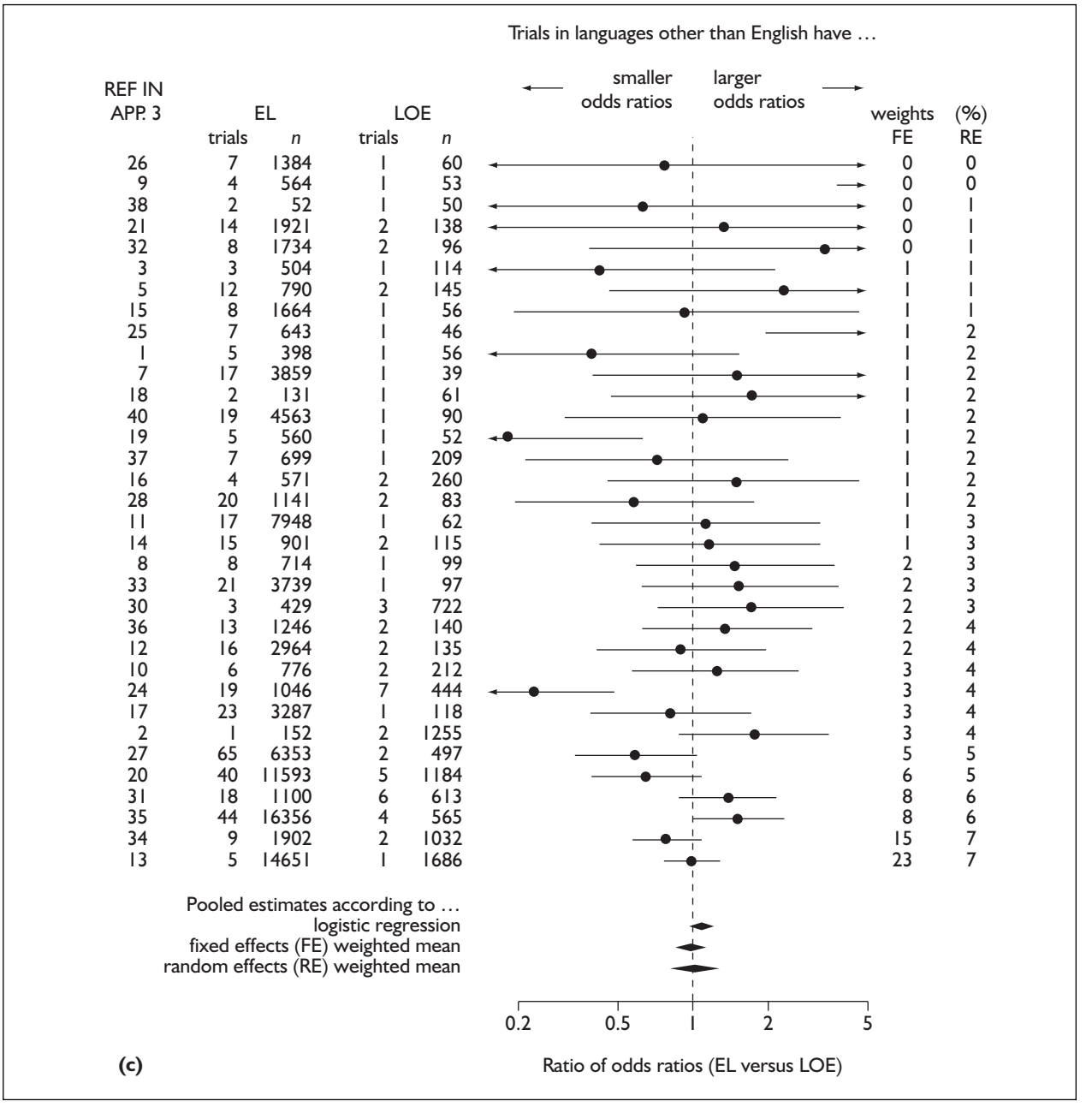


FIGURE 8 *(c) Impact of LOE at the RCT level: RORs with 95% CIs for LOE versus LE for each of 34 conventional medicine systematic reviews, together with pooled estimates from three different approaches (shown as diamonds)*

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[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).

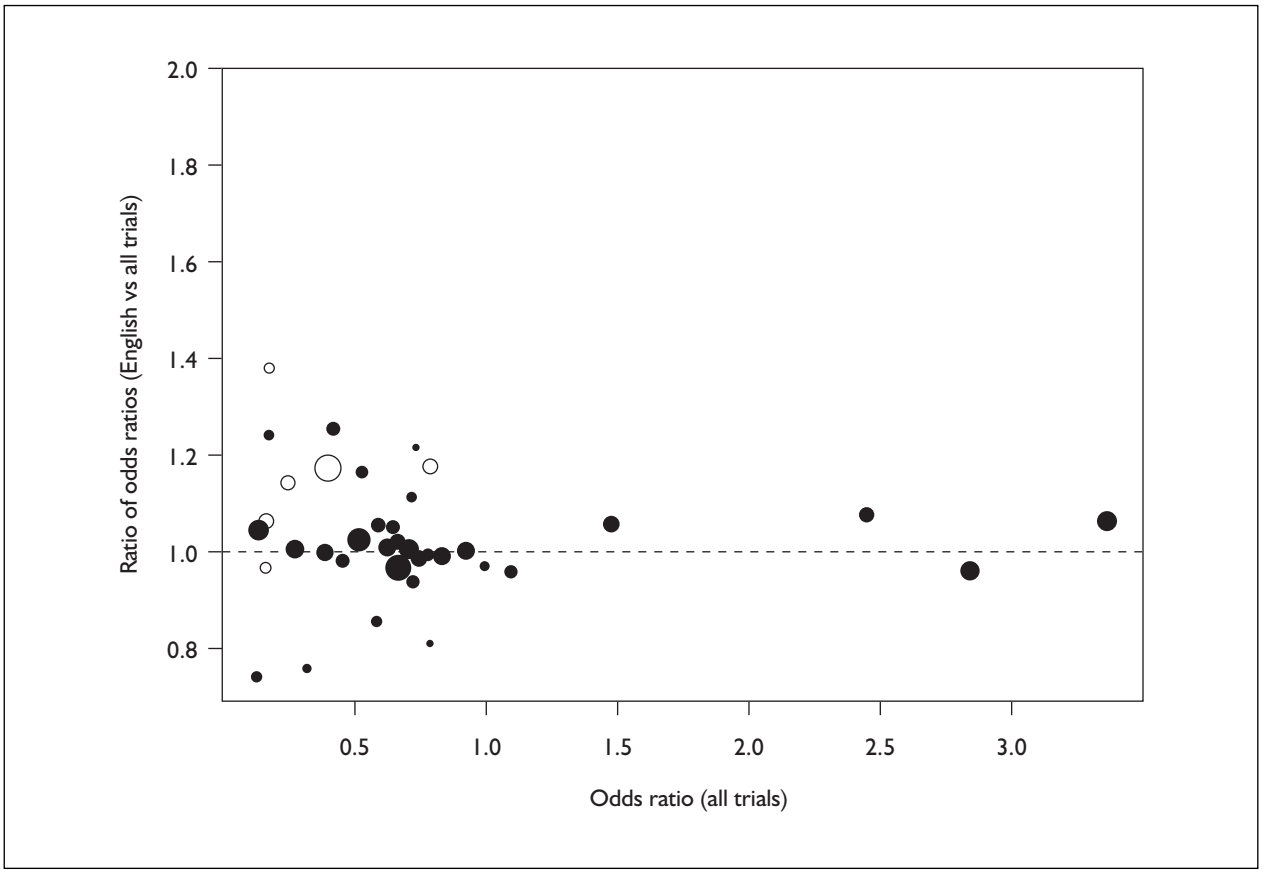


FIGURE 9 *Effect of language restrictions on the estimates of intervention effect. An ROR > 1 estimated from EL trials and all trials indicates that language restricted systematic reviews report smaller intervention effect estimates. Estimates are plotted proportionally to the number of trials in each systematic review with CM (filled circle) and CAM (open circle). For example, a systematic review of homeopathy interventions including both EL and LOE trials reported a protective OR of 0.4. Restricting the data synthesis to EL trials only, the corresponding OR was 0.47 and the ROR between language restricted and language inclusive/LOE systematic review was 1.18.*

Similarly, Linde and colleagues[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.

a'Rogvi-Hansen and Boysen[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).

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

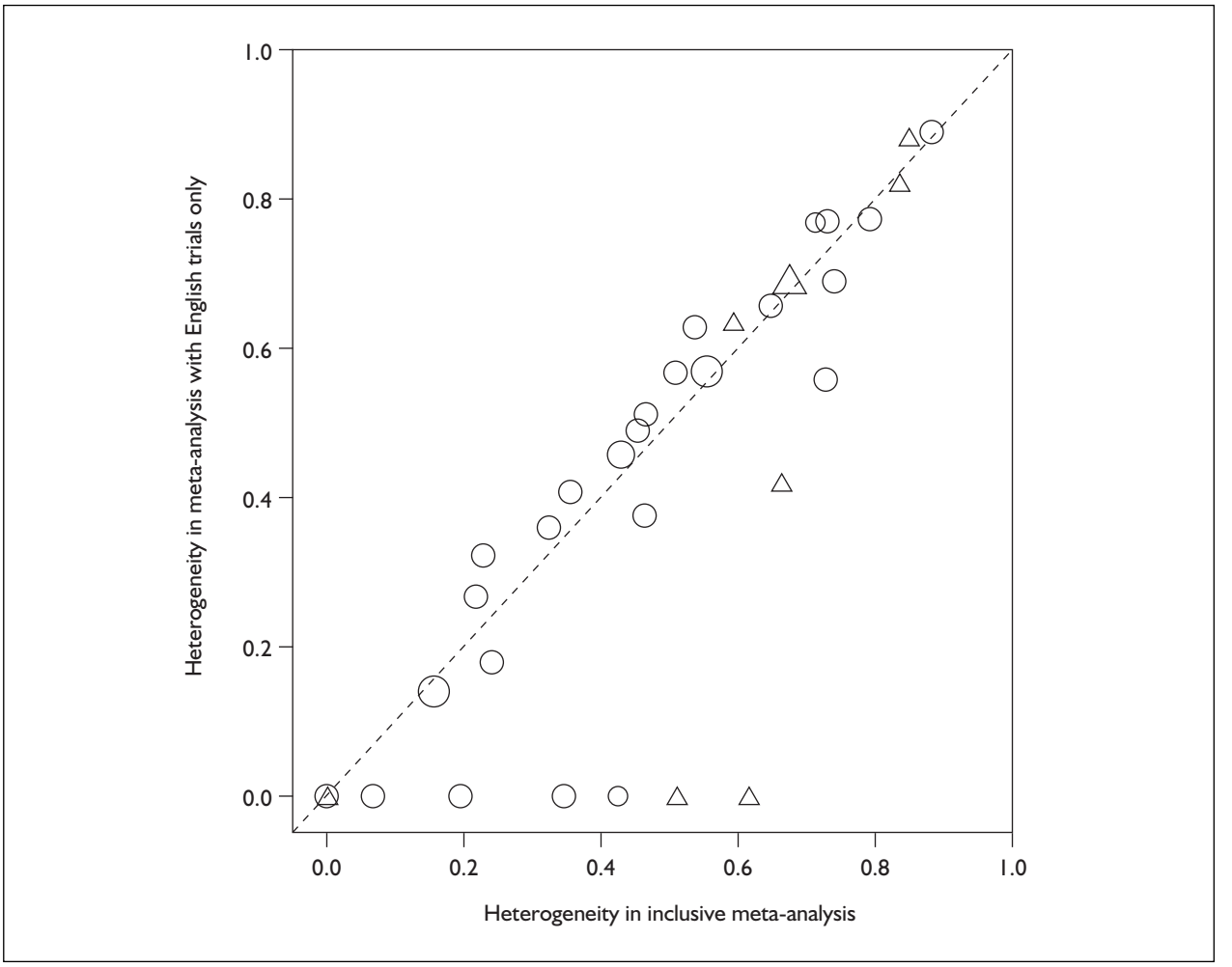


FIGURE 10 *Heterogeneity (i.e. as measured by the I^2 statistic[47]) in language inclusive/LOE systematic reviews versus language restricted systematic reviews. Circles denote systematic reviews of CM and triangles systematic reviews in CAM. Values above 0.56 call for considerable caution in interpreting the intervention effect estimates from the systematic reviews whereas values below 0.30 might cause little concern.*

similar to those reported by Moher and colleagues,[30] 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 13 *Language restriction, statistical heterogeneity and estimates of intervention effect*[a]

MA	n/pat	CAM?	Q	*P*-Homo	OR	Low	High	I^2	*H*	n/pat	Q	*P*-Homo	OR	Low	High	I^2	*H*
1	6/831	0	1.05	0.96	0.72	0.42	1.22	0.00	0.46	4/571	0.40	0.94	0.79	0.43	1.48	0.00	0.36
2	26/1490	0	38.23	0.04	0.15	0.08	0.25	0.35	1.24	19/1046	12.23	0.84	0.15	0.09	0.26	0.00	0.82
3	5/239	1	27.19	0.00	0.17	0.03	0.87	0.85	2.61	4/197	26.04	0.00	0.17	0.02	1.32	0.89	2.95
4	4/475	1	7.84	0.05	0.44	0.14	1.34	0.62	1.62	2/430	0.01	0.91	0.85	0.49	1.49	0.00	0.11
5	6/659	1	31.29	0.00	0.19	0.07	0.50	0.84	2.50	5/619	22.54	0.00	0.26	0.10	0.66	0.82	2.37
6	74/8778	1	226.00	0.00	0.40	0.32	0.50	0.68	1.76	30/3909	94.97	0.00	0.47	0.33	0.68	0.70	1.81
7	12/1947	1	22.56	0.02	0.79	0.55	1.12	0.51	1.43	7/480	0.97	0.99	0.92	0.58	1.49	0.00	0.40
8	18/8010	0	15.53	0.56	0.92	0.83	1.03	0.00	0.96	17/7948	15.48	0.49	0.92	0.83	1.03	0.00	0.98
9	6/612	0	24.55	0.00	0.13	0.05	0.34	0.80	2.22	5/560	17.85	0.00	0.10	0.04	0.26	0.78	2.11
10	3/192	0	1.60	0.45	0.73	0.39	1.38	0.00	0.89	2/131	0.96	0.33	0.89	0.41	1.95	0.00	0.98
11	22/1224	0	38.56	0.01	2.83	1.88	4.26	0.46	1.35	20/1141	37.37	0.01	2.72	1.75	4.22	0.49	1.40
12	24/1713	0	85.88	0.00	3.35	2.12	5.31	0.73	1.93	18/1100	74.91	0.00	3.56	1.92	6.69	0.77	2.10
13	6/16337	0	10.84	0.06	0.99	0.79	1.25	0.54	1.47	5/14651	10.84	0.03	0.97	0.67	1.39	0.63	1.65
14	9/1720	0	10.37	0.24	0.46	0.25	0.86	0.23	1.14	8/1664	10.36	0.17	0.45	0.22	0.93	0.32	1.22
15	16/2055	0	7.58	0.94	0.70	0.45	1.09	0.00	0.71	14/1921	7.36	0.88	0.71	0.45	1.12	0.00	0.75
16	8/689	0	27.21	0.00	0.43	0.21	0.86	0.74	1.97	7/643	19.49	0.00	0.53	0.28	1.03	0.69	1.80
17	17/1016	0	24.86	0.07	0.63	0.42	0.94	0.36	1.25	15/901	23.70	0.05	0.63	0.40	0.99	0.41	1.30
18	22/3836	0	60.12	0.00	0.28	0.20	0.40	0.65	1.69	21/3739	58.80	0.00	0.28	0.20	0.41	0.66	1.71
19	20/4653	0	28.16	0.08	0.75	0.55	1.02	0.33	1.22	19/4563	28.10	0.06	0.74	0.53	1.03	0.36	1.25
20	18/3898	0	149.32	0.00	0.67	0.44	1.02	0.89	2.96	17/3859	148.87	0.00	0.68	0.44	1.06	0.89	3.05
21	8/908	0	5.02	0.66	1.09	0.74	1.62	0.00	0.85	7/699	4.73	0.58	1.05	0.70	1.58	0.00	0.89
22	9/813	0	10.23	0.25	2.44	1.67	3.60	0.22	1.13	8/714	9.57	0.21	2.64	1.70	4.06	0.27	1.17
23	13/828	1	35.86	0.00	0.18	0.09	0.34	0.67	1.73	3/217	3.46	0.18	0.19	0.08	0.42	0.42	1.31
24	3/317	1	1.05	0.59	0.79	0.50	1.26	0.00	0.73	2/237	0.63	0.43	0.85	0.51	1.43	0.00	0.79
25	8/1444	0	5.92	0.55	0.78	0.45	1.35	0.00	0.92	7/1384	5.90	0.44	0.77	0.44	1.35	0.00	0.99
26	3/102	0	0.24	0.89	0.79	0.19	3.19	0.00	0.34	2/52	0.14	0.71	0.64	0.09	4.39	0.00	0.37
27	6/454	0	5.36	0.37	0.59	0.36	0.98	0.07	1.04	5/398	3.63	0.46	0.51	0.30	0.84	0.00	0.95
28	6/1151	0	9.35	0.10	0.53	0.28	1.03	0.47	1.37	3/429	3.22	0.20	0.62	0.25	1.54	0.38	1.27
29	10/1826	0	31.96	0.00	0.59	0.36	0.98	0.72	1.88	8/1734	30.76	0.00	0.63	0.37	1.06	0.77	2.10
30	12/391	1	27.20	0.00	0.25	0.10	0.67	0.60	1.57	9/311	22.04	0.01	0.29	0.09	0.93	0.64	1.66
31	24/3405	0	19.49	0.67	0.84	0.70	0.98	0.00	0.92	23/3287	19.23	0.63	0.83	0.70	0.97	0.00	0.93
32	48/16921	0	105.95	0.00	0.52	0.44	0.61	0.56	1.50	44/16356	100.16	0.00	0.53	0.45	0.63	0.57	1.53
33	14/935	0	8.11	0.84	1.48	0.99	2.20	0.00	0.79	12/790	7.16	0.79	1.57	1.03	2.36	0.00	0.81
34	45/12777	0	77.38	0.00	0.67	0.55	0.82	0.43	1.33	40/11593	72.09	0.00	0.65	0.52	0.80	0.46	1.36
35	3/1407	0	3.48	0.18	0.97	0.69	1.38	0.43	1.32	1/152							
36	15/1386	0	28.55	0.01	0.65	0.44	0.96	0.51	1.43	13/1246	27.84	0.01	0.66	0.42	1.03	0.57	1.52
37	4/618	0	1.03	0.80	0.33	0.15	0.70	0.00	0.58	3/504	0.11	0.95	0.25	0.09	0.64	0.00	0.24
38	8/988	0	8.70	0.28	0.64	0.44	0.95	0.20	1.11	6/776	2.42	0.79	0.68	0.46	0.99	0.00	0.70

continued

TABLE 13 *Language restriction, statistical heterogeneity and estimates of intervention effect[a] (cont'd)*

MA	n/pat	CAM?	*Q*	*P*-Homo	OR	Low	High	I^2	*H*	n/pat	*Q*	*P*-Homo	OR	Low	High	I^2	*H*
39	11/2934	0	18.75	0.04	0.73	0.58	0.92	0.47	1.37	9/1902	16.42	0.04	0.68	0.51	0.92	0.51	1.43
40	67/6850	0	78.20	0.15	0.67	0.57	0.79	0.16	1.09	65/6353	74.60	0.17	0.65	0.55	0.77	0.14	1.08
41	5/617	0	14.80	0.01	0.18	0.09	0.40	0.73	1.92	4/564	6.83	0.08	0.23	0.13	0.40	0.56	1.51
42	18/3099	0	22.42	0.17	0.39	0.32	0.48	0.24	1.15	16/2964	18.31	0.25	0.39	0.32	0.47	0.18	1.10

[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.[43] The I^2 and *H* statistics measure the degree of statistical heterogeneity according to Higgins and Thompson.[47] 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; I^2, 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).[52] 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

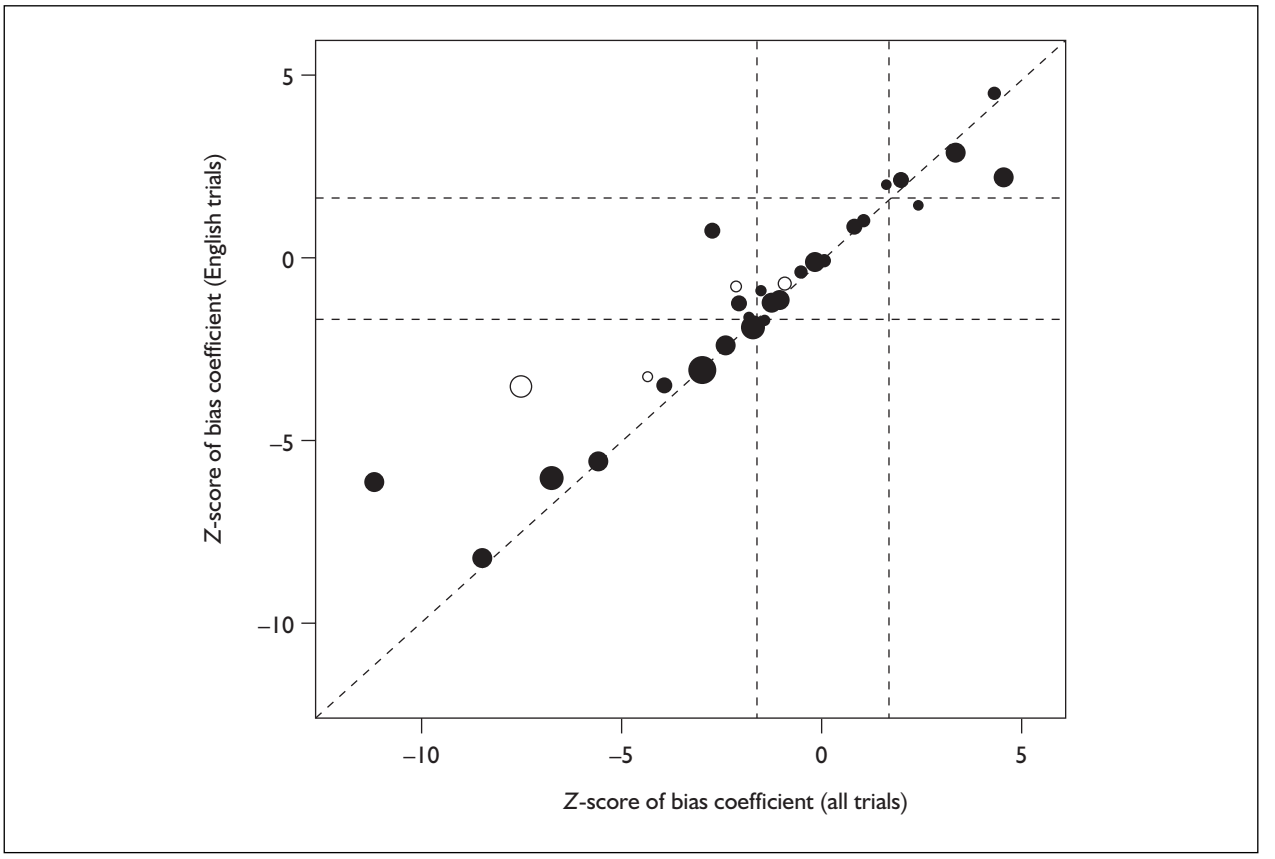


FIGURE 11 *Effect of language restriction on funnel plot asymmetry – a marker for publication bias. The Z-scores for zero bias coefficient from the regression approach to funnel plot[46] from language inclusive/LOE systematic reviews are plotted against their corresponding values from language restricted systematic reviews. A negative Z-score that corresponds to a large protective effect is more likely to be reported in small trials. Estimates from systematic reviews with five or more EL trials are displayed proportionally to the number of trials in each systematic review with CM (filled circle) and CAM (open circle). Dotted lines denote equality, and significance level of 0.1 (i.e. –1.64 and 1.64). The median of the Z-scores was –0.50 [interquartile range (–1.87, 0.93), range (–11.23, 2.29)] for language inclusive/LOE systematic reviews and –0.22 [interquartile range (–1.84, 0.65), range (–4.17, 2.16)] for language restricted systematic reviews.*

TABLE 14 *Language restriction and publication bias*[a]

		Restricted					Inclusive/LOE						
			ME		PM			ME		PM			
ID	**Rx**	**n/pats**	***p*-value**	***Z*-score**	***p*-value**	***Z*-score**	**n/pats**	***p*-value**	***Z*-score**	***p*-value**	***Z*-score**	**Rater I**	**Rater II**
2	CM	19/1046	0	–4.17	0	–6.14	26/1490	0	–11.23	0	–11.15	Y	Y
5	CAM	5/619	0.15	–1.92	0.05	–3.21	6/659	0.06	–2.60	0.01	–4.35		Y
6	CAM	30/3909	0.04	–2.16	0	–3.51	74/8778	0	–4.40	0	–7.50		Y
7	CAM	7/480	0.66	0.47	0.48	–0.77	12/1947	0.14	–1.63	0.06	–2.15	Y	
8	CM	17/7948	0.71	–0.38	0.26	–1.19	18/8010	0.69	–0.41	0.22	–1.28		Y
9	CM	5/560	0.66	–0.48	0.20	–1.66	6/612	0.92	–0.11	0.22	–1.45		
11	CM	20/1141	0.44	0.80	0.01	2.91	22/1224	0.31	1.05	0	3.34		
12	CM	18/1100	0.56	0.60	0.04	2.24	24/1713	0.03	2.29	0	4.51		
13	CM	5/14651	0.81	–0.26	0.13	2.04	6/16337	0.80	–0.27	0.18	1.60		
14	CM	8/1664	1.00	0.01	0.16	–1.66	9/1720	0.97	–0.04	0.12	–1.80		
15	CM	14/1909	0.23	1.28	0.41	0.88	16/2045	0.23	1.26	0.45	0.81		
16	CM	7/643	0.86	–0.18	0.43	–0.87	8/689	0.58	–0.58	0.17	–1.55	Y	
17	CM	15/901	0.84	0.21	0.29	–1.10	17/1016	0.82	0.24	0.31	–1.06		
18	CM	21/3739	0	–3.96	0	–8.23	22/3836	0	–4.28	0	–8.46	Y	Y
19	CM	19/4563	0.05	–2.10	0.03	–2.36	20/4653	0.04	–2.19	0.03	–2.42		
20	CM	17/3859	0.92	–0.11	0.92	–0.10	18/3898	0.90	–0.13	0.86	–0.18		
21	CM	7/699	0.95	–0.07	0.96	–0.05	8/908	0.85	0.20	0.97	0.04		
22	CM	8/714	0.14	1.73	0	4.54	9/813	0.12	1.80	0	4.31		
25	CM	7/1384	0.41	–0.90	0.36	1.04	8/1444	0.41	–0.89	0.36	1.04		
27	CM	5/398	0.32	1.20	0.24	1.48	6/454	0.13	1.88	0.07	2.40		Y
29	CM	8/1734	0.17	1.56	0.73	–0.36	10/1826	0.14	1.65	0.61	–0.54		
30	CAM	9/311	0.28	–1.17	0.52	–0.68	12/391	0.27	–1.17	0.38	–0.92		
31	CM	23/3287	0.02	–2.66	0.01	–2.87	24/3405	0.02	–2.61	0.01	–2.74		
32	CM	44/16356	0.08	–1.82	0	–6.02	48/16921	0.04	–2.09	0	–6.76		
33	CM	12/790	0.06	2.16	0.06	2.17	14/935	0.13	1.61	0.08	1.97		
34	CM	40/11593	0.07	–1.88	0.07	1.85	45/12777	0.08	–1.80	0.09	–1.73		
36	CM	13/1246	0.01	–3.43	0.01	–3.46	15/1386	0	–3.67	0	–3.94	Y	Y
38	CM	6/776	0.68	–0.44	0.33	–1.11	8/988	0.53	–0.66	0.38	–0.95		
39	CM	9/1902	0.84	0.21	0.27	–1.21	11/2934	0.24	–1.26	0.07	–2.08		
40	CM	65/6353	0.43	0.80	0	–3.04	67/6850	0.37	0.90	0	–3.01		
42	CM	16/2964	0.48	–0.72	0	–5.56	18/3099	0.53	–0.64	0	–5.57		

[a] 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[44]; PM, test for funnel plot asymmetry using the method by Macaskill and colleagues[45]; 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,[53] 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*).[55] 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.[56] 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.

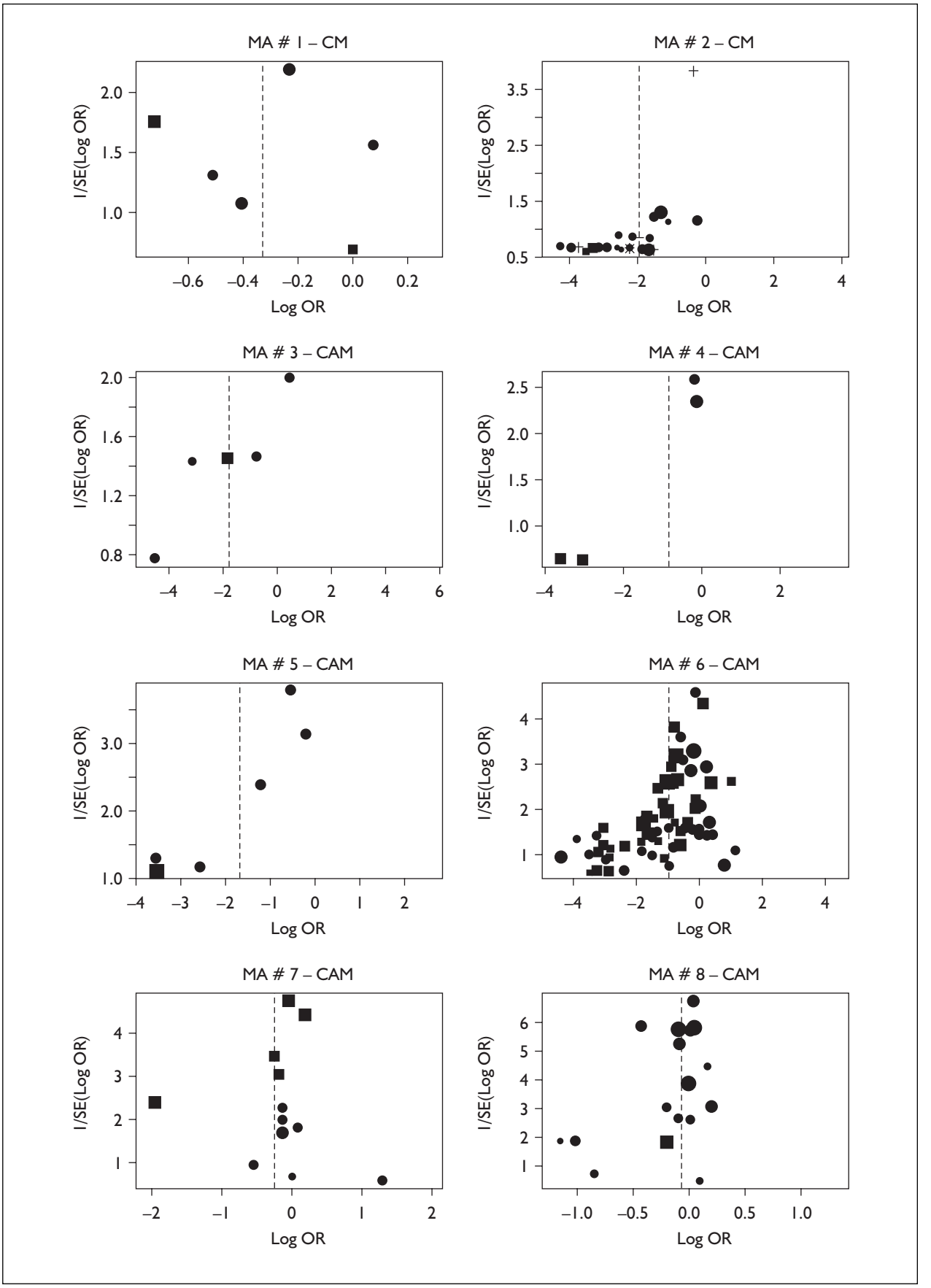


FIGURE 12 *Funnel plots of the 42 language inclusive/LOE systematic reviews. Treatment effect estimates are plotted proportionately to quality using a circle symbol (EL trial), square (LOE trial), * (EL trial with unavailable quality assessment) and + (LOE trial with unavailable quality assessment). 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).*

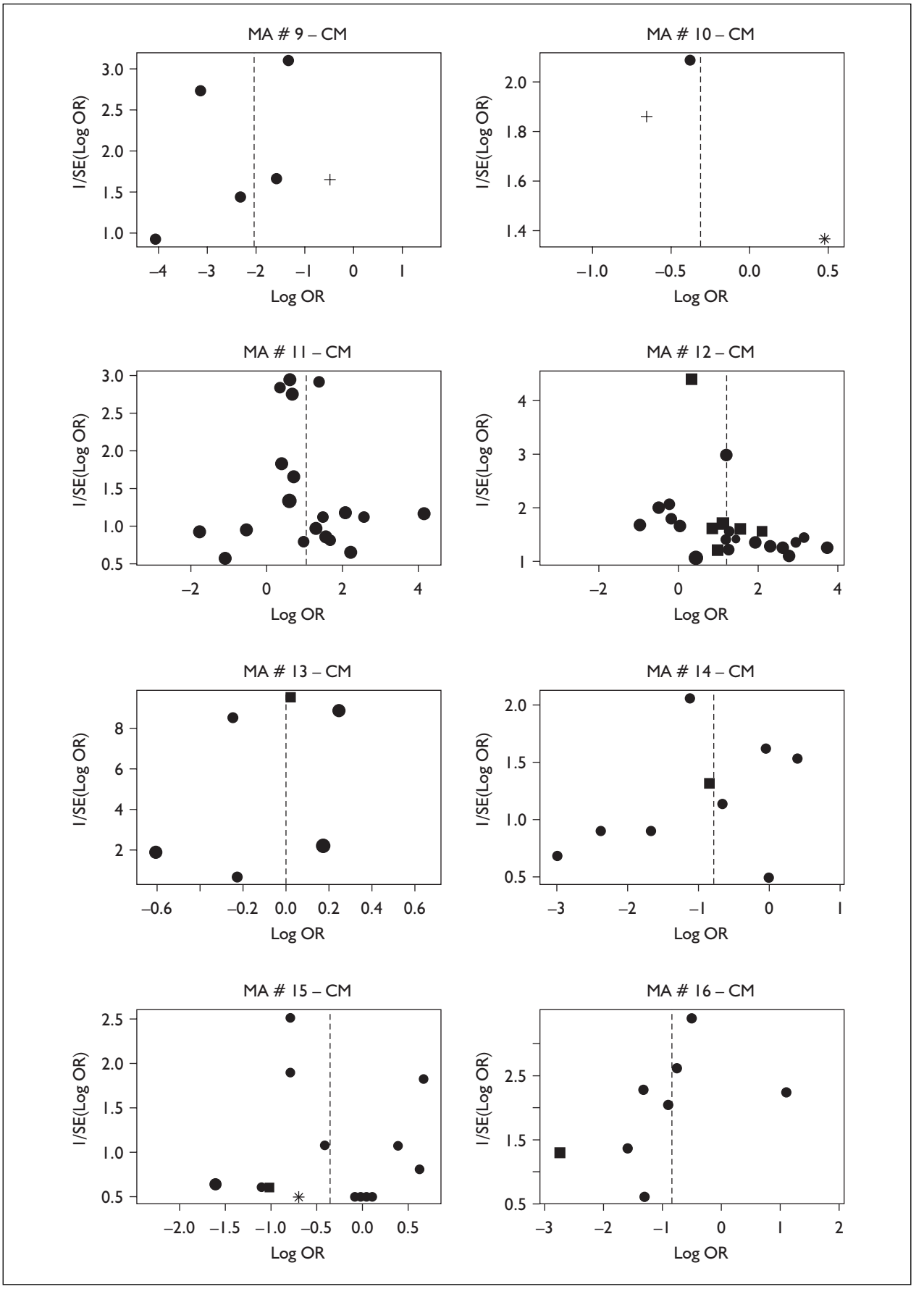


FIGURE 12 *(continued)*

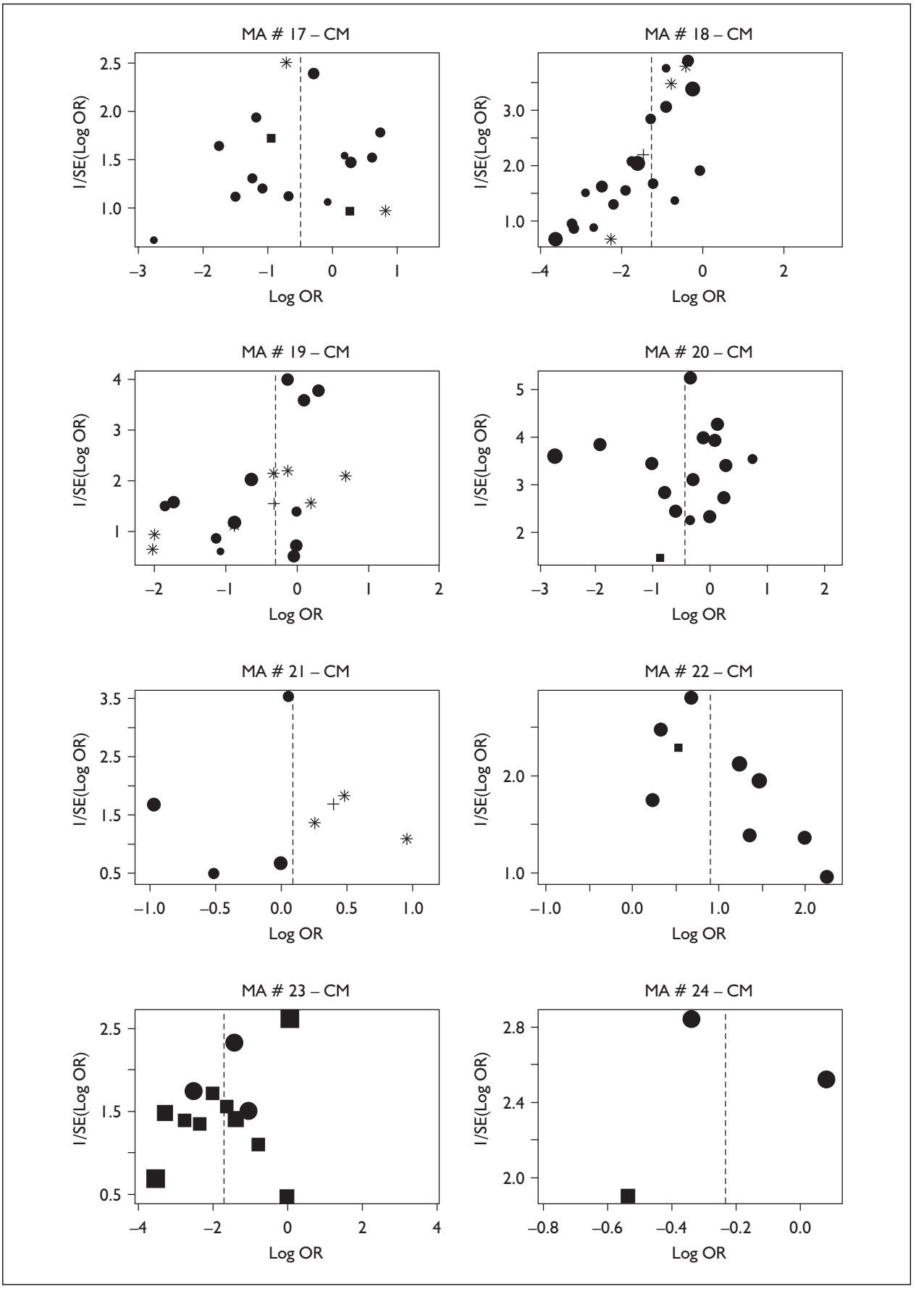


FIGURE 12 *(continued)*

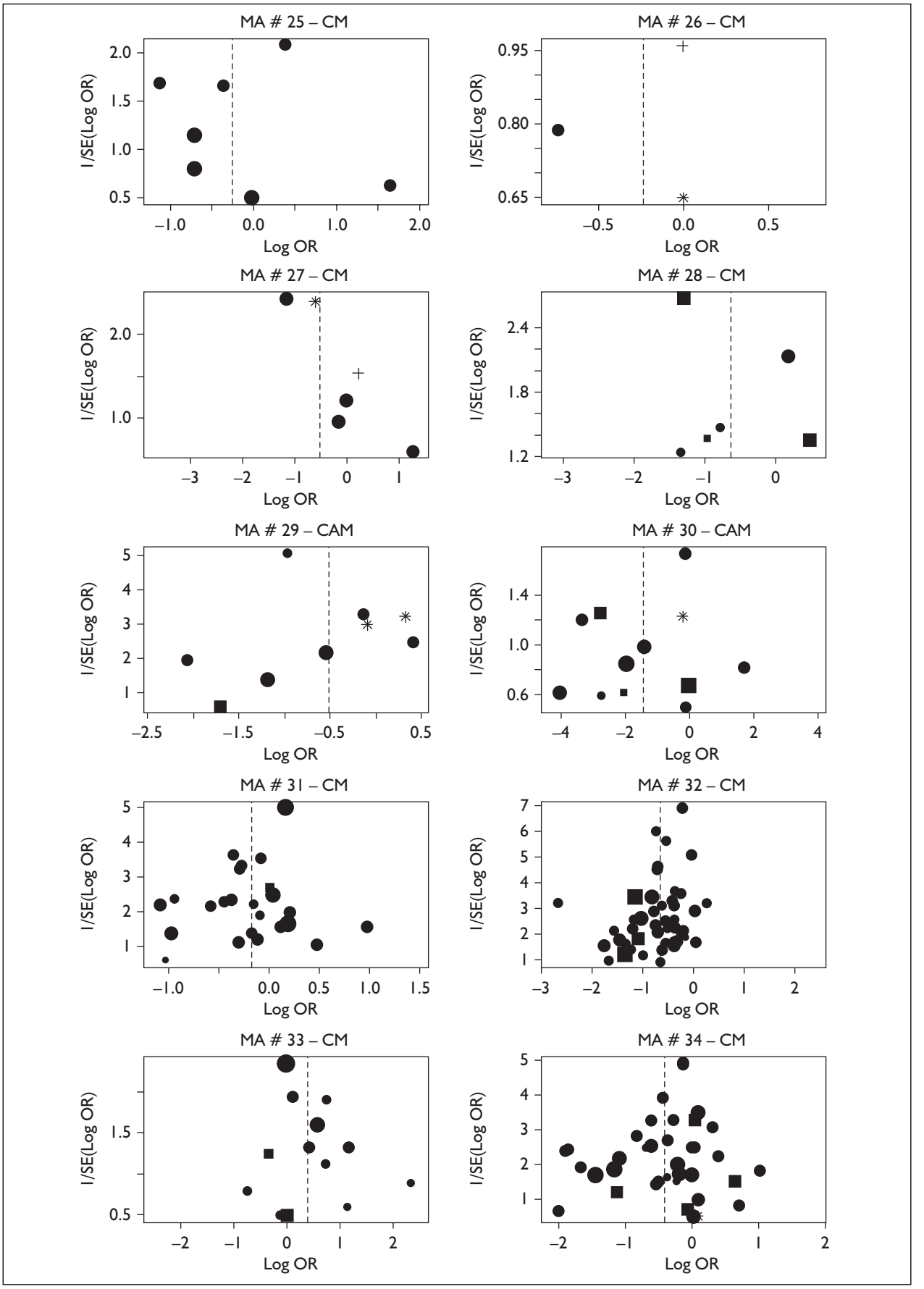


FIGURE 12 *(continued)*

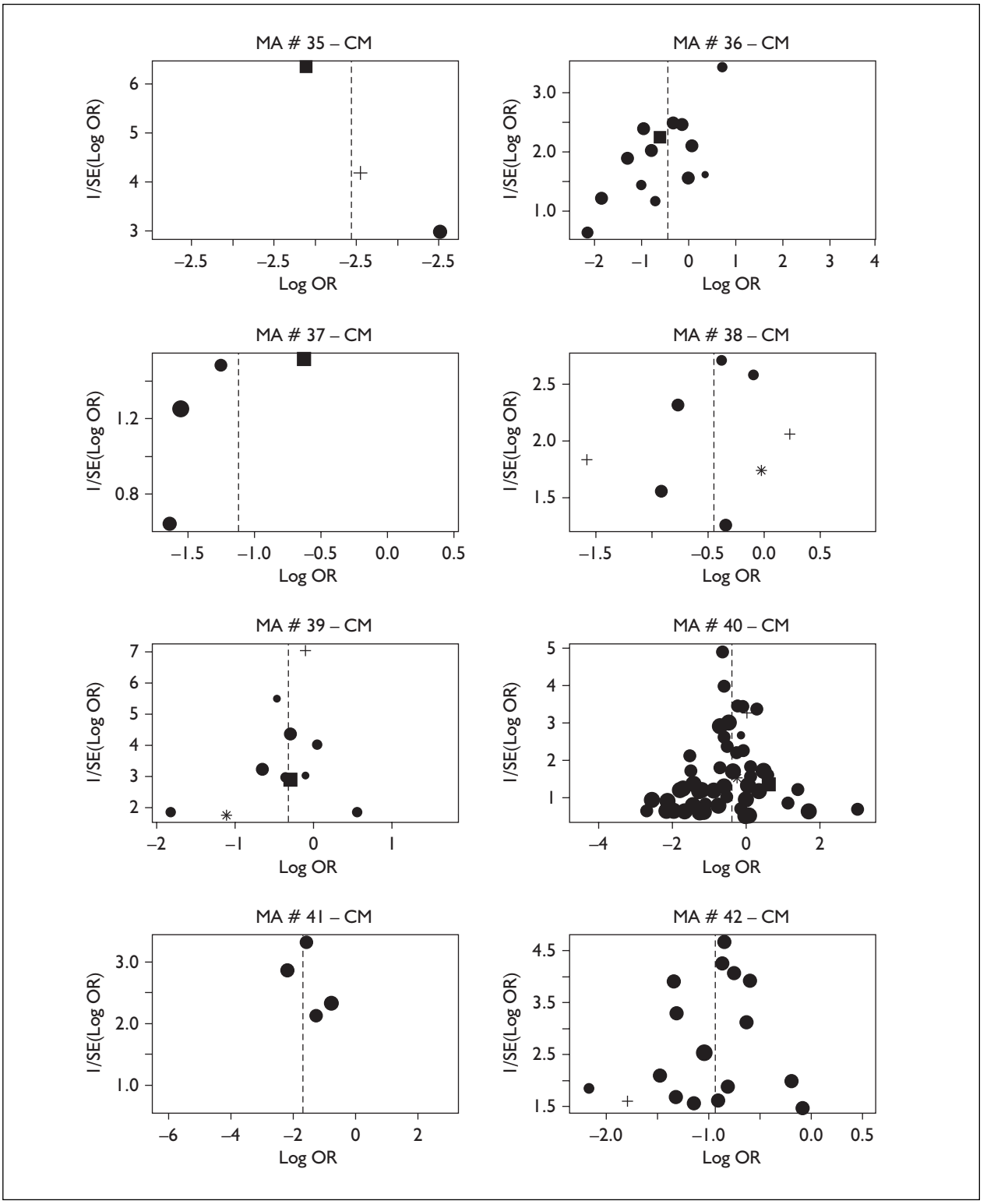


FIGURE 12 *(continued)*

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,[30] 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 reviews. These reviews included almost 700 reports of RCTs, including >100 reports 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. Jüni and colleagues,[23] 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[18] 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,[44] 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[57] 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[58] 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,[30] 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 restrictions on quality of reporting for systematic reviews are consistent with Moher and colleagues' previous findings.[30] 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,[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.[59] 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[13] 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[13] 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 reviews 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[15,16] and has been used extensively since then to assess the quality of reporting of systematic reviews.[13,30,57] 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,[35] effect of language restriction,[30] types of search strategies and publication bias.[60] 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.[61] 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.[62] 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.[18] 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).[19] 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,[19] 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.[63] 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[23] 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 CAM 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[31] 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.[30] This is not surprising given the similar level of quality of reporting that we and others have found for EL and LOE CM trials.[19] 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,[30] although the present results are based on a dataset approximately three times the size of that used in the earlier work. Jüni and colleagues, using a very similar-sized dataset, reported similar results.[23]

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[31] 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[64] 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[41] 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[41] and Sterne and colleagues.[42]

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 associated with the inclusion of LOE trials.

The inclusion of LOE trials in language inclusive/LOE systematic reviews expectedly results in genuine differences in patient populations, interventions and outcome measures. In our sample of systematic reviews, the clinical heterogeneity that we expected did not translate into any consistent patterns of increasing 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.[3,42,44] 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.[3]

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.[31] 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 colleagues[66] 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 colleagues[17] 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 statement[67] 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.[68,69] 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[65] 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,[30] 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,[19] although Egger and colleagues[31] 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.[21] 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 be 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.[70,71]

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[62] and a much larger number of randomised trials.[72] 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.

Acknowledgements

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

1. Davidoff F, Haynes RB, Sackett D, Smith R. Evidence-based medicine: a new journal to help doctors identify the information they need. *BMJ* 1995;**310**:1085–6.
2. Beecher HK. The powerful placebo. *JAMA* 1955;**159**:1602–6.
3. Egger M, Davey-Smith G, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. 2nd ed. London: BMJ Books; 2001.
4. Teo KK, Yusuf S, Collins R, Held PH, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. *BMJ* 1991;**303**:1499–503.
5. ISIS-4 Collaborative Group. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58050 patients with suspected acute myocardial infarction. *Lancet* 1995;**345**:669–85.
6. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomised control trials of primary treatment of breast cancer. *J Clin Oncol* 1986; **4**:942–51.
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.
8. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol* 2002; **55**:783–6.
9. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers T. Meta-analysis of randomised controlled trials. *N Engl J Med* 1987;**316**:450–5.
10. Sacks HS, Berrier J, Reitman D, Pagano D, Chalmers TC. Meta-analyses of randomised controlled trials: an update of the quality and methodology. In: Bailer JC III, Mosteller F, editors. Medical uses of statistics. Boston, MA: NEJM Books, 1992. pp. 427–42.
11. Sacks HS, Reitman D, Pagano D, Kupelnick B. Meta-analysis: an update. *Mt Sinai J Med* 1996; **63**:216–24.
12. Assendelft WJ, Koes BW, Knipschild PG, Bouter LM. The relationship between methodological quality and conclusions in reviews of spinal manipulation. *JAMA* 1995;**274**:1942–8.
13. Jadad AR, McQuay HJ. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J Clin Epidemiol* 1996;**49**:235–43.
14. Silagy CA. An analysis of review articles published in primary care journals. *Fam Pract* 1993;**10**:337–41.
15. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol* 1991; **44**:1271–8.
16. Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA, *et al*. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; **44**:91–8.
17. Shea B, Moher D, Pham B, Graham I, Tugwell P. The quality of reporting of meta-analyses. *Evaluation Health Profess* 2002;**25**:116–29.
18. Grégoire G, Derderian F, Le Lorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol* 1995;**48**:159–63.
19. Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le Lorier J, *et al*. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;**347**:363–6.
20. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, *et al*. Assessing the quality of reports of randomised clinical trials: is blinding necessary? *Control Clin Trials* 1996; **17**:1–12.
21. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**:408–12.
22. Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. II: replicate variability and comparison of studies that agree and disagree. *Stat Med* 1987;**6**:733–44.
23. Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;**31**:115–23
24. Vandenbroucke-Grauls CM, Vandenbroucke JP. Effect of selective decontamination of the digestive tract on respiratory tract infections and mortality in the intensive care unit. *Lancet* 1991;**338**:859–62.
25. Thülig B, Hartenauer U, Diemer W, Lawin P, Fegeler W, Kehrel R, *et al*. Selektive

Florasuppression zur Infektionskontrolle in der operativen Intensivmedizin. *Anästh Intensivther Notfallmed* 1989;**24**:345–54.

26. Selective Decontamination of the Digestive Tract Trialists' Collaborative Group. Meta-analysis of randomised controlled trials of selective decontamination of the digestive tract. *BMJ* 1993; **307**:525–32.
27. Brun-Buisson C, Legrand P, Rauss A, Richard C, Montravers F, Besbes M, *et al*. Intestinal decontamination for control of nosocomial multi resistant gram-negative bacilli. *Ann Intern Med* 1989;**110**:873–81.
28. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals. *JAMA* 1994;**272**:125–8.
29. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomised clinical trials. *JAMA* 1991;**266**:93–8.
30. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, *et al*. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000;**53**:964–72.
31. Egger M, Zellweger-Zahner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;**350**:326–9.
32. Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? a systematic review of controlled trials. *Control Clin Trials* 1998; **19**:159–66.
33. Tang JL, Zhan SY, Ernst E. Review of randomised controlled trials of traditional Chinese medicine. *BMJ* 1999;**319**:160–1.
34. Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, *et al*. Assessing the quality of randomised controlled trials: implications for the conduct of meta-analyses. *Health Technol Assess* 1999;**3**:1–98.
35. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence the estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000;**356**:1228–31.
36. Shea B, Dube C, Moher D. Assessing the quality of reports of systematic reviews: the QUOROM statement compared to other tools. In: Egger M, Davey-Smith G, Altman DG, editors. Systematic reviews in health care. London: BMJ Books, 2001. pp. 122–39.
37. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–8.
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**:159–74.
39. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomised controlled trials. *JAMA* 1994;**272**:122–4.
40. Robins J, Greenland S, Breslow NE. A general estimator for the variance of the Mantel–Haenszel odds ratio. *Am J Epidemiol* 1986;**124**:719–23.
41. Sampson M, Barrowman NJ, Moher D, Klassen TP, Pham B, Platt R, *et al*. Should systematic reviewers search Embase in addition to Medline? *J Clin Epidemiol* 2003;**56**:943–55.
42. Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;**21**:1513–24.
43. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88.
44. Egger M, Davey-Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34.
45. Macaskill P, Walter S, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001;**20**:641–54.
46. Eliasziw M, Donner A. Application of the McNemar test to non-independent matched pair data. *Stat Med* 1991;**12**:1981–91.
47. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**:1539–58.
48. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987;**107**:224–33.
49. Colditz GA, Berkey CS, Mosteller F, Brewer TF, Wilson ME, Burdick E, *et al*. The efficacy of bacillus Calmette–Guérin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature. *Pediatrics* 1995; **96**:29–35.
50. Rowe BH, Spooner CH, Ducharme FM, Bretzlaff JA, Bota GW. Corticosteroids for preventing relapse following acute exacerbations of asthma. *Cochrane Database Syst Rev* 2001;Issue 1.
51. Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997;**350**:834–43.
52. White AR, Rampes H, Ernst E. Acupuncture for smoking cessation. *Cochrane Database Syst Rev* 1997; Issue 2.
53. a'Rogvi-Hansen B, Boysen G. Glycerol for acute ischaemic stroke. *Cochrane Database Syst Rev* 1996; Issue 2.
54. Malaguarnera M, Restuccia S, Trovato G, Siciliano R, Motta M, Trovato BA, *et al*. Interferon-alpha treatment in patients with chronic hepatitis C:

a meta-analytic evaluation. *Clin Drug Invest* 1995; **9**:141–9.

55. Sikorski J, Renfrew MJ. Support for breastfeeding mothers. *Cochrane Database Syst Rev* 1998; Issue 2.

56. Linde K, Ramirez G, Mulrow CD, Pauls A, Weidenhammer W, Melchart D. St John's wort for depression – an overview and meta-analysis of randomised clinical trials. *BMJ* 1996;**313**:253–8.

57. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, *et al*. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. *BMJ* 2000;**320**:537–40.

58. Choi PT, Halpern SH, Malik N, Jadad AR, Tramer MR, Walder B. Examining the evidence in anesthesia literature: a critical appraisal of systematic reviews. *Anesth Analg* 2001;**92**:700–9.

59. Moher D, Klassen TP, Pham B, Lawson ML. Evidence of journal bias against complementary and alternative medicine. Presented at the fourth International Congress on Peer Review in Biomedical Publication, Barcelona, 2001.

60. Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998;**316**:61–6.

61. Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompay M, *et al*. Trends in alternative medicine use in the United States, 1990–1997: results of a follow-up national survey. *JAMA* 1998;**280**:1569–75.

62. Moher D, Soeken K, Sampson M, Campbell K, Ben Perot L, Berman B. Assessing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. *BMC (BioMedCentral) Pediatrics* 2002;**2**:3.

63. Linde K, Jonas WB, Melchart D, Willich S. The methodological quality of randomised controlled trials of homeopathy, herbal medicines and acupuncture. *Int J Epidemiol* 2001;**30**:526–31.

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

65. Linde K, Scholz M, Ramirez G, Clausius N, Melchart D, Jonas WB. Impact of study quality on outcome in placebo-controlled trials of homeopathy. *J Clin Epidemiol* 1999;**52**:631–6.

66. Jadad AR, Cook DJ, Jones A, Klassen T, Moher M, Tugwell P, Moher D. Methodology and reports of systematic reviews and meta-analyses: a comparison of Cochrane reviews with articles published in paper-based journals. *JAMA* 1998;**280**:178–80.

67. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, *et al*. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of reporting of meta-analyses. *Lancet* 1999;**354**:1896–900.

68. Moher D, Schulz KF, Altman DG, for the CONSORT group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Ann Intern Med* 2001;**134**:657–62.

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

70. Moher D, Jones A, Lepage L, for the CONSORT group. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before and after evaluation? *JAMA* 2001; **285**:1992–5.

71. P. J. Devereaux, B. J. Manns, W. A. Ghali, H. Quan, G. H. Guyatt. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. *Control Clin Trials* 2002;**23**:380–8.

72. Moher D, Sampson M, Campbell K, Beckner W, Lepage L, Gaboury I, Berman B. Assessing the quality of reports of randomized controlled trials in pediatric complementary and alternative medicine. *BMC (BioMedCentral) Pediatrics* 2002;**2**:2.

Appendix 1

Data abstraction form

DATA ABSTRACTION FORM
Meta-analysis and the Language of Publication of Included Randomised Controlled Trials

ID. NO. _____ REVIEWER _____

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?

Single pharmaceutical company ❒

Multiple pharmaceutical companies ❒

Pharmaceutical company and a non-drug sponsor ❒ **[one or more pharmaceutical companies and a non-pharmaceutical organisation]**

Non-drug company ❒ **[a non-pharmaceutical company or sponsor including universities, medical societies, government and foundations]**

None listed/can't tell ❒

DATA ABSTRACTION FORM
Meta-analysis and the Language of Publication of Included Randomised Controlled Trials

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

❒ 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 ❒ Other _____

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? _____ ❒ Can't Tell

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_____

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_____
❒ Subgroup analysis
❒ Sensitivity analysis

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

1. a'Rogvi-Hansen B, Boysen G. Glycerol for acute ischaemic stroke. *Cochrane Database Syst Rev* 2000;**77**(2):*CD000096*.
2. Blondel B, Breart G. Home visits for pregnancy complications and management of antenatal care: an overview of three randomized controlled trials. *Br J Obstet Gynaecol* 1992;**99**:283–6.
3. Boissel JP, Peyrieux JC, Destors JM. Is it possible to reduce the risk of cardiovascular events in subjects suffering from intermittent claudication of the lower limbs? *Thromb Haemost* 1989;**62**:681–5.
4. Bressa GM. *S*-Adenosyl-l-methionine (SAMe) as antidepressant: meta-analysis of clinical studies. *Acta Neurol. Scand. Suppl.* 1994;**154**:7–14.
5. Brown KH, Peerson JM, Fontaine O. Use of nonhuman milks in the dietary management of young children with acute diarrhea: a meta-analysis of clinical trials. *Pediatrics* 1994;**93**:17–27.
6. Carroll D, Tramer M, McQuay H, Nye B, Moore A. Transcutaneous electrical nerve stimulation in labour pain: a systematic review. *Br J Obstet Gynaecol* 1997;**104**:169–75.
7. Cohen HJ, Silberman HR, Tornyos K, Bartolucci AA. Comparison of two long-term chemotherapy regimens, with or without agents to modify skeletal repair, in multiple myeloma. *Blood* 1984;**63**:639–48.
8. Covey LS, Glassman AH. A meta-analysis of double-blind placebo-controlled trials of clonidine for smoking cessation. *Br J Addict* 1991;**86**:991–8.
9. Crawford F, Hart R, Bell-Syer S, Torgerson D, Young P, Russell I. Topical treatments for fungal infections of the skin and nails of the foot. *Cochrane Database Syst Rev* 2000;**9** Suppl.4(2):*CD001434*.
10. Daya S. Gonadotropin releasing hormone agonist protocols for pituitary desensitization in *in vitro* fertilization and gamete intrafallopian transfer cycles. *Cochrane Database Syst Rev* 2000;(2):*CD001299*.
11. Douglas RM, Chalker EB, Treacy B. Vitamin C for preventing and treating the common cold. *Cochrane Database Syst Rev* 2000;**350**(2):*CD000980*.
12. Figueredo E, Canosa L. Prophylactic ondansetron for postoperative emesis. Meta-analysis of its effectiveness in patients with previous history of postoperative nausea and vomiting. *Acta Anaesthesiol Scand* 1999;**43**:637–44.
13. Fine MJ, Smith MA, Carson CA, Meffe F, Sankey SS, Weissfeld LA, *et al*. Efficacy of pneumococcal vaccination in adults. A meta-analysis of randomized controlled trials. *Arch Intern Med* 1994;**154**:2666–77.
14. Glowacki LS, Smaill FM. Use of immune globulin to prevent symptomatic cytomegalovirus disease in transplant recipients – a meta-analysis. *Clin Transplant* 1994;**8**:10–18.
15. Halpern S, Preston R. Postdural puncture headache and spinal needle design. Metaanalyses. *Anesthesiology* 1994;**81**:1376–83.
16. Hayashi K, Chalmers TC. Famotidine in the treatment of duodenal ulcer. A meta-analysis of randomized control trials. *Gastroenterol Int* 1993; **6**:19–25.
17. Heyland DK, Cook DJ, Jaeschke R, Griffith L, Lee HN, Guyatt GH. Selective decontamination of the digestive tract. An overview. *Chest* 1994; **105**:1221–9.
18. Hofmeyr GJ, Kulier R. Cephalic version by postural management for breech presentation. Cochrane *Database Syst Rev* 2000;**350**(3):*CD000051*.
19. Hofmeyr GJ. External cephalic version facilitation for breech presentation at term. *Cochrane Database Syst Rev* 2000;**350**(2):*CD000184*.
20. Leizorovicz A, Haugh MC, Chapuis FR, Samama MM, Boissel JP. Low molecular weight heparin in prevention of perioperative thrombosis. *BMJ* 1992;**305**:913–20.
21. Leizorovicz A, Simonneau G, Decousus H, Boissel JP. Comparison of efficacy and safety of low molecular weight heparins and unfractionated heparin in initial treatment of deep venous thrombosis: a meta-analysis. *BMJ* 1994;**309**:299–304.
22. Linde K, Ramirez G, Mulrow CD, Pauls A, Weidenhammer W, Melchart D. St John's wort for depression – an overview and meta-analysis of randomised clinical trials. *BMJ* 1996;**313**:253–8.
23. Linde K, Clausius N, Ramirez G, Melchart D, Eitel F, Hedges LV, *et al*. Are the clinical effects of homeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997;**350**:834–43.

24. Malaguarnera M, Restuccia S, Trovato G, Siciliano R, Motta M, Trovato BA, *et al*. Interferon-alpha treatment in patients with chronic hepatitis C: a meta-analytic evaluation. *Clin Drug Invest* 1995;**9**:141–9.

25. Marino P, Pampallona S, Preatoni A, Cantoni A, Invernizzi F. Chemotherapy vs supportive care in advanced non-small-cell lung cancer. Results of a meta-analysis of the literature. *Chest* 1994;**106**:861–5.

26. Meijer WS, Schmitz PI, Jeekel J. Meta-analysis of randomized, controlled clinical trials of antibiotic prophylaxis in biliary tract surgery. *Br J Surg* 1990; **77**:283–90.

27. Montgomery SA, Kasper S. Comparison of compliance between serotonin reuptake inhibitors and tricyclic antidepressants: a meta-analysis. *Int Clin Psychopharmacol* 1995;**9** Suppl 4:33–40.

28. Pace F, Maconi G, Molteni P, Minguzzi M, Bianchi PG. Meta-analysis of the effect of placebo on the outcome of medically treated reflux esophagitis. *Scand J Gastroenterol* 1995;**30**:101–5.

29. Pittler MH,.Ernst E. Peppermint oil for irritable bowel syndrome: a critical review and metaanalysis. *Am J Gastroenterol* 1998;**93**:1131–5.

30. Pouleur H,.Buyse M. Effects of dipyridamole in combination with anticoagulant therapy on survival and thromboembolic events in patients with prosthetic heart valves. A meta-analysis of the randomized trials. *J Thorac Cardiovasc Surg* 1995; **110**:463–72.

31. Poynard T, Naveau S, Mory B, Chaput JC. Meta-analysis of smooth muscle relaxants in the treatment of irritable bowel syndrome. *Aliment Pharmacol Ther* 1994;**8**:499–510.

32. Poynard T, Valterio C. Meta-analysis of hydroxyethylrutosides in the treatment of chronic venous insufficiency. *Vasa* 1994;**23**:244–50.

33. SDD Trialist Group. Meta-analysis of randomised controlled trials of selective decontamination of the digestive tract. Selective Decontamination of the Digestive Tract Trialists' Collaborative Group. *BMJ* 1993;**307**:525–32.

34. Sikorski J, Renfrew MJ. Support for breastfeeding mothers. *Cochrane Database Syst Rev* 2000; (2):*CD001141*.

35. Silagy C, Mant D, Fowler G, Lodge M. The effectiveness of nicotine replacement therapies in smoking cessation. *Online J Curr Clin Trials* 1994; Doc No 113:7906.

36. Van Ruiswyk J, Byrd JC. Efficacy of prophylactic sclerotherapy for prevention of a first variceal hemorrhage. *Gastroenterology* 1992;**102**:587–97.

37. Vandekerckhove P, Lilford R, Vail A, Hughes E. Androgens versus placebo or no treatment for idiopathic oligo/asthenospermia. *Cochrane Database Syst Rev* 2000;**63**(2):*CD000150*.

38. Vandekerckhove P, Lilford R, Vail A, Hughes E. The medical treatment of idiopathic oligo/ asthenospermia: bromocriptine versus placebo or no treatment. *Cochrane Database Syst Rev* 2000; **77**(2):*CD000152*.

39. White AR, Rampes H, Ernst E. Acupuncture for smoking cessation. *Cochrane Database Syst Rev* 2000;**350**(2):*CD000009*.

40. Wilson AP, Shrimpton S, Jaderberg M. A meta-analysis of the use of amoxycillin-clavulanic acid in surgical prophylaxis. *J Hosp Infect* 1992;**22** Suppl A:9–21.

41. Wilt TJ, Ishani A, Stark G, MacDonald R, Lau J, Mulrow C. Saw palmetto extracts for treatment of benign prostatic hyperplasia: a systematic review. *JAMA* 1998;**280**:1604–9.

Language restricted systematic reviews

1. Abramson MJ, Puy RM, Weiner JM. Is allergen immunotherapy effective in asthma? A meta-analysis of randomized controlled trials. *Am J Respir Crit Care Med* 1995;**151**:969–74.

2. Anderson R, Meeker WC, Wirick BE, Mootz RD, Kirk DH, Adams A. A meta-analysis of clinical trials of spinal manipulation. *J Manipulative Physiol Ther* 1992;**15**:181–94.

3. Appel LJ, Miller ER, Seidler AJ, Whelton PK. Does supplementation of diet with 'fish oil' reduce blood pressure? A meta-analysis of controlled clinical trials. *Arch Intern Med* 1993;**153**:1429–38.

4. Avgerinos A, Armonis A, Raptis S. Somatostatin or octreotide versus endoscopic sclerotherapy in acute variceal haemorrhage: a meta-analysis study. *J Hepatol* 1995;**22**:247–8.

5. Barbui T, Finazzi G, Grassi A, Marchioli R. Thrombosis in cancer patients treated with hematopoietic growth factors – a meta-analysis. On behalf of the Subcommittee on Haemostasis and Malignancy of the Scientific and Standardization Committee of the ISTH. *Thromb Haemost* 1996; **75**:368–71.

6. Browman GP. Evidence-based recommendations against neoadjuvant chemotherapy for routine management of patients with squamous cell head and neck cancer. *Cancer Invest* 1994;**12**:662–70.

7. Cappelleri JC, Fiore LD, Brophy MT, Deykin D, Lau J. Efficacy and safety of combined anticoagulant and antiplatelet therapy versus anticoagulant monotherapy after mechanical heart-valve replacement: a metaanalysis. *Am Heart J* 1995; 130:547-52.

8. Cummings P, Psaty BM. The association between cholesterol and death from injury. *Ann Intern Med* 1994;**120**:848–55.

9. Egarter C, Leitich H, Husslein P, Kaider A, Schemper M. Adjunctive antibiotic treatment in

preterm labor and neonatal morbidity: a meta-analysis. *Obstet Gynecol* 1996;**88**:303–9.

10. Eisenberg DM, Delbanco TL, Berkey CS, Kaptchuk TJ, Kupelnick B, Kuhl J, *et al*. Cognitive behavioral techniques for hypertension: are they effective? *Ann Intern Med* 1993;**118**:964–72.

11. Eisenberg E, Berkey CS, Carr DB, Mosteller F, Chalmers TC. Efficacy and safety of nonsteroidal antiinflammatory drugs for cancer pain: a meta-analysis. *J Clin Oncol* 1994;**12**:2756–65.

12. Fortin PR, Lew RA, Liang MH, Wright EA, Beckett LA, Chalmers TC, *et al*. Validation of a meta-analysis: the effects of fish oil in rheumatoid arthritis. *J Clin Epidemiol* 1995;**48**:1379–90.

13. Fraser EJ, Grimes DA, Schulz KF. Immunization as therapy for recurrent spontaneous abortion: a review and meta-analysis. *Obstet Gynecol* 1993;**82**:854–9.

14. Fremes SE, Wong BI, Lee E, Mai R, Christakis GT, McLean RF, *et al*. Metaanalysis of prophylactic drug treatment in the prevention of postoperative bleeding. *Ann Thorac Surg* 1994;**58**:1580–8.

15. Hansen JF. Review of postinfarct treatment with verapamil: combined experience of early and late intervention studies with verapamil in patients with acute myocardial infarction. Danish Study Group on Verapamil in Myocardial Infarction. *Cardiovasc Drugs Ther* 1994;**8** Suppl 3:543–7.

16. Hauth JC, Goldenberg RL, Parker CR, Cutter GR, Cliver SP. Low-dose aspirin: lack of association with an increase in abruptio placentae or perinatal mortality. *Obstet Gynecol* 1995;**85**:1055–8.

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

18. Hillegass WB, Ohman EM, Leimberger JD, Califf RM. A meta-analysis of randomized trials of calcium antagonists to reduce restenosis after coronary angioplasty. *Am J Cardiol* 1994;**73**:835–9.

19. Hooker KD, DiPiro JT, Wynn JJ. Aminoglycoside combinations versus beta-lactams alone for penetrating abdominal trauma: a meta-analysis. *J Trauma* 1991;**31**:1155–60.

20. Hricik DE, O'Toole MA, Schulak JA, Herson J. Steroid-free immunosuppression in cyclosporine-treated renal transplant recipients: a meta-analysis. *J Am Soc Nephrol* 1993;**4**:1300–5.

21. Kreter B, Woods M. Antibiotic prophylaxis for cardiothoracic operations. Meta-analysis of thirty years of clinical trials. *J Thorac Cardiovasc Surg* 1992;**104**:590–9.

22. Labrecque M, Dostaler LP, Rousselle R, Nguyen T, Poirier S. Efficacy of nonsteroidal anti-inflammatory drugs in the treatment of acute renal colic. A meta-analysis. *Arch Intern Med* 1994;**154**:1381–7.

23. Lensing AW, Prins MH, Davidson BL, Hirsh J. Treatment of deep venous thrombosis with low-molecular-weight heparins. A meta-analysis. *Arch Intern Med* 1995;**155**:601–7.

24. Lu-Yao GL, Keller RB, Littenberg B, Wennberg JE. Outcomes after displaced fractures of the femoral neck. A meta-analysis of one hundred and six published reports. *J Bone Joint Surg Am* 1994; **76**:15–25.

25. Macharia WM, Leon G, Rowe BH, Stephenson BJ, Haynes RB. An overview of interventions to improve compliance with appointment keeping for medical services. *JAMA* 1992;**267**:1813–17.

26. May GR, Sutherland LR, Shaffer EA. Efficacy of bile acid therapy for gallstone dissolution: a meta-analysis of randomized trials. *Aliment Pharmacol Ther* 1993;**7**:139–48.

27. Meunier F, Paesmans M, Autier P. Value of antifungal prophylaxis with antifungal drugs against oropharyngeal candidiasis in cancer patients. *Eur J Cancer* 1994;**30B**:196–9.

28. Midgette AS, O'Connor GT, Baron JA, Bell J. Effect of intravenous streptokinase on early mortality in patients with suspected acute myocardial infarction. A meta-analysis by anatomic location of infarction. *Ann Intern Med* 1990;**113**:961–8.

29. Midgette AS, Wong JB, Beshansky JR, Porath A, Fleming C, Pauker SG. Cost-effectiveness of streptokinase for acute myocardial infarction: a combined meta-analysis and decision analysis of the effects of infarct location and of likelihood of infarction. *Med Decis Making* 1994;**14**:108–17.

30. Moreland J, Thomson MA. Efficacy of electromyographic biofeedback compared with conventional physical therapy for upper-extremity function in patients following stroke: a research overview and meta-analysis. *Phys Ther* 1994; **74**:534–43.

31. Moreland JD, Thomson MA, Fuoco AR. Electromyographic biofeedback to improve lower extremity function after stroke: a meta-analysis. *Arch Phys Med Rehabil* 1998;**79**:134–40.

32. Morganroth J, Goin JE. Quinidine-related mortality in the short-to-medium-term treatment of ventricular arrhythmias. A meta-analysis. *Circulation* 1991;**84**:1977–83.

33. Ng CM, Rivera JO. Meta-analysis of streptokinase and heparin in deep vein thrombosis. *Am J Health Syst Pharm* 1998;**55**:1995–2001.

34. O'Connor GT, Malenka DJ, Olmstead EM, Johnson PS, Hennekens CH. A meta-analysis of randomized trials of fish oil in prevention of restenosis following coronary angioplasty. *Am J Prev Med* 1992;**8**:186–92.

35. Pichichero ME, Margolis PA. A comparison of cephalosporins and penicillins in the treatment of group A beta-hemolytic streptococcal pharyngitis: a

meta-analysis supporting the concept of microbial copathogenicity. *Pediatr Infect Dis J* 1991;**10**:275–81.

36. Pyorala S, Huttunen NP, Uhari M. A review and meta-analysis of hormonal treatment of cryptorchidism. *J Clin Endocrinol Metab* 1995; **80**:2795–9.

37. Rossetti L, Marchetti I, Orzalesi N, Scorpiglione N, Torri V, Liberati A. Randomized clinical trials on medical treatment of glaucoma. Are they appropriate to guide clinical practice? *Arch Ophthalmol* 1993;**111**:96–103.

38. Rowe BH, Keller JL, Oxman AD. Effectiveness of steroid therapy in acute exacerbations of asthma: a meta-analysis. *Am J Emerg Med* 1992;**10**:301–10.

39. Sacks HS, Chalmers TC, Blum AL, Berrier J, Pagano D. Endoscopic hemostasis. An effective therapy for bleeding peptic ulcers. *JAMA* 1990; **264**:494–9.

40. Schleenbaker RE, Mainous AG. Electromyographic biofeedback for neuromuscular reeducation in the hemiplegic stroke patient: a meta-analysis. *Arch Phys Med Rehabil* 1993;**74**:1301–4.

41. Thomas JA, McIntosh JM. Are incentive spirometry, intermittent positive pressure breathing, and deep breathing exercises effective in the prevention of postoperative pulmonary complications after upper abdominal surgery? A systematic overview and meta-analysis. *Phys Ther* 1994;**74**:3–10.

42. Wang PH, Lau J, Chalmers TC. Meta-analysis of effects of intensive blood-glucose control on late complications of type I diabetes. *Lancet* 1993; **341**:1306–9.

43. Yurkowski PJ, Plaisance KI. Prevention of auditory sequelae in pediatric bacterial meningitis: a meta-analysis. *Pharmacotherapy* 1993;**13**:494–9.

44. Zhang WY, Li WP. The effectiveness of topically applied capsaicin. A meta-analysis. *Eur J Clin Pharmacol* 1994;**46**:517–22.

Language inclusive/EL systematic reviews

1. Andrews TC, Reimold SC, Berlin JA, Antman EM. Prevention of supraventricular arrhythmias after coronary artery bypass surgery. A meta-analysis of randomized control trials. *Circulation* 1991; **84**:III236–44.

2. Barker FG. Efficacy of prophylactic antibiotics for craniotomy: a meta-analysis. *Neurosurgery* 1994; **35**:484–90.

3. Cohard M, Poynard T, Mathurin P, Zarski JP. Prednisone–interferon combination in the treatment of chronic hepatitis B: direct and indirect metanalysis. *Hepatology* 1994;**20**:1390–8.

4. Colditz GA, Brewer TF, Berkey CS, Wilson ME, Burdick E, Fineberg HV, *et al*. Efficacy of BCG vaccine in the prevention of tuberculosis. Meta-analysis of the published literature. *JAMA* 1994; **271**:698–702.

5. Colditz GA, Berkey CS, Mosteller F, Brewer TF, Wilson ME, Burdick E, *et al*. The efficacy of bacillus Calmette–Guérin vaccination of newborns and infants in the prevention of tuberculosis: meta-analyses of the published literature. *Pediatrics* 1995;**96**:29–35.

6. Coplen SE, Antman EM, Berlin JA, Hewitt P, Chalmers TC. Efficacy and safety of quinidine therapy for maintenance of sinus rhythm after cardioversion. A meta-analysis of randomized control trials. *Circulation* 1990;**82**:1106–16.

7. Cummings P, Del BM. Antibiotics to prevent infection of simple wounds: a meta-analysis of randomized studies. *Am J Emerg Med* 1995; **13**:396–400.

8. Di MR, Marchioli R, Tognoni G. From pharmacological promises to controlled clinical trials to meta-analysis and back: the case of nimodipine in cerebrovascular disorders. *Clin Trials Metaanal* 1994;**29**:57–79.

9. Fardy JM, Laupacis A. A meta-analysis of prophylactic endoscopic sclerotherapy for esophageal varices. *Am J Gastroenterol* 1994;**89**:1938–48.

10. Fontaine O, Gore SM, Pierce NF. Rice-based oral rehydration solution for treating diarrhoea. *Cochrane Database Syst Rev* 2000;**28**(2):*CD001264*.

11. Glanz M, Klawansky S, Stason W, Berkey C, Shah N, Phan H, *et al*. Biofeedback therapy in poststroke rehabilitation: a meta-analysis of the randomized controlled trials. *Arch Phys Med Rehabil* 1995; **76**:508–15.

12. Hemila H. Vitamin C intake and susceptibility to the common cold. *Br J Nutr* 1997;**77**:59–72.

13. Lancaster T, Silagy C, Gray S. Primary care management of acute herpes zoster: systematic review of evidence from randomized controlled trials. *Br J Gen Pract* 1995;**45**:39–45.

14. Langhorne P, Williams BO, Gilchrist W, Howie K. Do stroke units save lives? *Lancet* 1993;**342**:395–8.

15. Lee A, Done ML. The use of nonpharmacologic techniques to prevent postoperative nausea and vomiting: a meta-analysis. *Anesth Analg* 1999; **88**:1362–9.

16. Lugo-Miro VI, Green M, Mazur L. Comparison of different metronidazole therapeutic regimens for bacterial vaginosis. A meta-analysis. *JAMA* 1992; **268**:92–5.

17. Mari JJ, Streiner DL. An overview of family interventions and relapse on schizophrenia: meta-analysis of research findings. *Psychol Med* 1994; **24**:565–78.

18. Martin-Hirsch PL, Lilford RJ, Jarvis GJ. Adjuvant progestagen therapy for the treatment of endometrial cancer: review and meta-analyses of published randomised controlled trials. *Eur J Obstet Gynecol Reprod Biol* 1996;**65**:201–7.

19. Ortiz Z, Shea B, Suarez-Almazor ME, Moher D, Wells GA, Tugwell P. The efficacy of folic acid and folinic acid in reducing methotrexate gastrointestinal toxicity in rheumatoid arthritis. A metaanalysis of randomized controlled trials. *J Rheumatol* 1998; **25**:36–43.

20. Roberts L, Ahmed I, Hall S, Sargent C, Adams C. Intercessory prayer for ill health: a systematic review. *Forsch Komplementarmed* 1998;**5** Suppl S1:82–6.

21. Sikorski J, Renfrew MJ. Support for breastfeeding mothers. *Cochrane Database Syst Rev* 2000; (2):*CD001141*.

22. Silagy C, Neil A. Garlic as a lipid lowering agent – a meta-analysis. *J R Coll Physicians Lond* 1994; **28**:39–45.

23. Silagy CA, Neil HA. A meta-analysis of the effect of garlic on blood pressure. *J Hypertens* 1994;**12**:463–8.

24. Spina GP, Henderson JM, Rikkers LF, Teres J, Burroughs AK, Conn HO, *et al*. Distal spleno-renal shunt versus endoscopic sclerotherapy in the prevention of variceal rebleeding. A meta-analysis of 4 randomized clinical trials. *J Hepatol* 1992; **16**:338–45.

25. Tholl DA, Miller TQ, Henderson WG, Myers TF. Meta-analysis of phenobarbital usage for prevention of intraventricular hemorrhage in premature infants: factors related to variation in outcome. *Clin Trials Metaanal* 1994;**29**:177–90.

26. Tine F, Magrin S, Craxi A, Pagliaro L. Interferon for non-A, non-B chronic hepatitis. A meta-analysis of randomised clinical trials. *J Hepatol* 1991; **13**:192–9.

27. Tramer MR, Fuchs-Buder T. Omitting antagonism of neuromuscular block: effect on postoperative nausea and vomiting and risk of residual paralysis. A systematic review. *Br J Anaesth* 1999;**82**:379–86.

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Kent Woods,
Director, NHS HTA Programme & Professor of Therapeutics, University of Leicester

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

Professor Shah Ebrahim, Professor in Epidemiology of Ageing, University of Bristol

Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Radcliffe Hospital, Oxford

Dr Ron Zimmern, Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Kent Woods, Director, NHS HTA Programme, Department of Medicine and Therapeutics, Leicester Royal Infirmary, Robert Kilpatrick Clinical Sciences Building, Leicester

Chair,
Professor Shah Ebrahim, Professor in Epidemiology of Ageing, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol

Deputy Chair,
Professor Jenny Hewison, Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine, Leeds

Professor Douglas Altman, Professor of Statistics in Medicine, Centre for Statistics in Medicine, Oxford University, Institute of Health Sciences, Cancer Research UK Medical Statistics Group, Headington, Oxford

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

Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health & Related Research, University of Sheffield, ScHARR Regent Court, Sheffield

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Oxford

Dr Christine Clark, Medical Writer & Consultant Pharmacist, Cloudside, Rossendale, Lancs
and
Principal Research Fellow, Clinical Therapeutics in the School of Pharmacy, Bradford University, Bradford

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York, Research Section, Seebohm Rowntree Building, Heslington, York

Dr Andrew Farmer, Senior Lecturer in General Practice, Department of Primary Health Care, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Fiona J Gilbert, Professor of Radiology, Department of Radiology, University of Aberdeen, Lilian Sutton Building, Foresterhill, Aberdeen

Professor Adrian Grant, Director, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen

Professor Alastair Gray, Director, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Mark Haggard, Director, MRC ESS Team, CBU Elsworth House, Addenbrooke's Hospital, Cambridge

Professor F D Richard Hobbs, Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham, Primary Care and Clinical Sciences Building, Edgbaston, Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Cambridge

Professor Sallie Lamb, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health, Coventry University, Coventry

Dr Donna Lamping, Senior Lecturer, Health Services Research Unit, Public Health and Policy, London School of Hygiene and Tropical Medicine, London

Professor David Neal, Professor of Surgical Oncology, Oncology Centre, Addenbrooke's Hospital, Cambridge

Professor Tim Peters, Professor of Primary Care Health Services Research, Division of Primary Health Care, University of Bristol, Cotham House, Cotham Hill, Bristol

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine, London

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital NHS Trust, Bramwell Dott Building, Edinburgh

Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine, Health & Social Care, St George's Building, Portsmouth

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Park House, Birmingham

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

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Dr Paul Cockcroft, Consultant Medical Microbiologist/ Laboratory Director, Public Health Laboratory, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, Addenbrooke's Hospital, Cambridge

Dr David Elliman, Consultant in Community Child Health, London

Dr Andrew Farmer, Senior Lecturer in General Practice, Institute of Health Sciences, University of Oxford

Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen

Professor Jane Franklyn, Professor of Medicine, University of Birmingham

Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London

Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton

Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust

Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust, Devon

Mr Tony Tester, Chief Officer, South Bedfordshire Community Health Council, Luton

Dr Andrew Walker, Senior Lecturer in Health Economics, University of Glasgow

Professor Martin J Whittle, Head of Division of Reproductive & Child Health, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Oxford Radcliffe Hospital

Professor Tony Avery, Professor of Primary Health Care, University of Nottingham

Professor Iain T Cameron, Professor of Obstetrics & Gynaecology, University of Southampton

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Dr Christopher Cates, GP and Cochrane Editor, Bushey Health Centre, Bushey, Herts.

Mr Charles Dobson, Special Projects Adviser, Department of Health

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff

Professor Alastair Gray, Professor of Health Economics, Institute of Health Sciences, University of Oxford

Mrs Sharon Hart, Managing Editor, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, Bristol South & West Primary Care Trust

Professor Robert Peveler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton

Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London

Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool

Dr Ken Stein, Senior Lecturer in Public Health, University of Exeter

Professor Terence Stephenson, Professor of Child Health, University of Nottingham

Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry, London

Professor Dame Jenifer Wilson-Barnett, Head of Florence Nightingale School of Nursing & Midwifery, King's College, London

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 Levvy, 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 Services, University of York

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

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

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 Cuckle,
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 Dezateux,
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 Fenwick,
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 Servs., 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 McCall,
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 HSRU/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

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.

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

ISSN 1366-5278