

# Direction and impact of language bias in meta-analyses of controlled trials: empirical study

Peter Jüni,<sup>a</sup> Franziska Holenstein,<sup>b</sup> Jonathan Sterne,<sup>a</sup> Christopher Bartlett<sup>a</sup> and Matthias Egger<sup>a</sup>

<b>Background</b>	Excluding clinical trials reported in languages other than English from meta-analyses may introduce bias and reduce the precision of combined estimates of treatment effects. We examined the influence of trials published in languages other than English on combined estimates and conclusions of published meta-analyses.
<b>Methods</b>	We searched journals and the <i>Cochrane Database of Systematic Reviews</i> for meta-analyses of at least five trials with binary outcomes that were based on comprehensive literature searches without language restrictions. We compared estimates of treatment effects from trials published in languages other than English to those from trials published in English, and assessed the impact of restricting meta-analyses to trials published in English.
<b>Results</b>	We identified 303 meta-analyses: 159 (52.4%) employed comprehensive literature searches of which 50 included 485 English and 115 non-English language trials. Non-English language trials included fewer participants (median 88 versus 116, $P = 0.006$ ) and were more likely to produce significant results at $P < 0.05$ (41.7% versus 31.3%, $P = 0.033$ ). The methodological quality of non-English language trials tended to be lower than that of trials published in English. Estimates of treatment effects were on average 16% (95% CI: 3–26%) more beneficial in non-English-language trials than in English-language trials. In 29 (58.0%) meta-analyses the change in effect estimates after exclusion of non-English language trials was less than 5%. In the remaining meta-analyses, 5 (10.0%) showed more benefit and 16 (32.0%) less benefit after exclusion of non-English language trials.
<b>Conclusions</b>	This retrospective analysis suggests that excluding trials published in languages other than English has generally little effect on summary treatment effect estimates. The importance of non-English language trials is, however, difficult to predict for individual systematic reviews. Comprehensive literature searches followed by a careful assessment of trial quality are required to assess the contribution of all relevant trials, independent of language of publication.
<b>Keywords</b>	Meta-analysis, randomized-controlled trials, publication bias, language bias, meta-regression
<b>Accepted</b>	2 October 2001

Systematic, continuously updated reviews and meta-analyses of the best evidence that is available on the benefits and risks of medical interventions can inform decision making in clinical practice and public health medicine, identify areas in which further research is needed and guide allocation of resources.<sup>1</sup>

Meta-analysis of randomized clinical trials is not an infallible tool, however, and several examples exist of meta-analyses which were later contradicted by single large randomized controlled trials,<sup>2,3</sup> and of meta-analyses addressing the same issue which have reached opposite conclusions.<sup>4</sup>

The inclusion of an unbiased sample of relevant studies is clearly central to the validity of systematic reviews and meta-analyses. However, the dissemination of medical evidence, including the results from randomized trials, is influenced by a host of factors that affect the probability that a given trial is included in a meta-analysis. Trials with statistically significant ('positive') results have been shown to be more likely to be published,<sup>5</sup> more likely to be published in English,<sup>6</sup> more likely

<sup>a</sup> Medical Research Council Health Services Research Collaboration, Department of Social Medicine, University of Bristol, UK.

<sup>b</sup> Department of Social and Preventive Medicine, University of Berne, Switzerland.

Correspondence: Matthias Egger, Department of Social Medicine, Canynge Hall, Whiteladies Road, University of Bristol, Bristol BS8 2PR, UK. E-mail: m.egger@bristol.ac.uk

to be published more than once<sup>7</sup> and more likely to be cited by other authors.<sup>8</sup> To prevent publication, language and citation biases in meta-analyses, the Cochrane Collaboration,<sup>9</sup> the Centre for Reviews and Dissemination of the British National Health Service<sup>10</sup> and other experts in the field<sup>11–13</sup> recommend extensive literature searches which cover all relevant languages. This may involve time consuming and costly attempts to identify all relevant literature and the translation of foreign language articles.

Although it seems likely that excluding trials reported in languages other than English will introduce bias and reduce the precision of estimates of treatment effects, the importance and direction of these effects is unclear at present. We identified state-of-the-art meta-analyses that were based on comprehensive literature searches and examined the contribution made by trials published in languages other than English and their impact on combined estimates of treatment effects and conclusions.

## Methods

We searched for meta-analyses of therapeutic or preventive interventions that combined the binary outcomes of at least five randomized trials. We manually searched all issues of nine general and specialist medical journals (*American Journal of Cardiology*, *Annals of Internal Medicine*, *British Medical Journal*, *Cancer*, *Circulation*, *Journal of the American Medical Association*, *Lancet*, *New England Journal of Medicine*, and *Obstetrics and Gynecology*) published 1994 through 1998 and all *Health Technology Assessment Reports* published up to July 1999 by the Research and Development Programme of the UK National Health Service.<sup>14</sup> The Centre for Reviews and Dissemination (University of York, UK) supplied us with copies of reports of meta-analyses of at least five trials published in any journal 1994 through 1998 which were reviewed for the *Database of Abstracts of Reviews of Effectiveness (DARE)*.<sup>15</sup> Finally, we checked every review published in issue 1/1998 of the *Cochrane Database of Systematic Reviews*.<sup>16</sup>

### Inclusion criteria

Meta-analyses that were based on comprehensive literature searches and provided sufficient data to allow re-analyses were included in this study. A comprehensive literature search was defined as a search not restricted to the English-language literature, which covered either the *Cochrane Controlled Trials Register* or at least two other electronic databases (such as *Medline* or *Embase*) and at least one other source (for example a search for unpublished material, a search of conference abstracts, theses or other grey literature, or a manual search of journals). If a review included meta-analyses of more than one binary outcome, we included the analysis that was based on the largest number of trials.

### Assessment of language of publication

Two of us (PJ/FH or PJ/CB) who were unaware of the results of component trials independently assessed the publication type and language of each trial. Trials were classified as journal reports if they were published as full or short reports, editorials or letters in a regular issue or supplement of a journal. All other reports, including conference abstracts published in journals, were classified as grey literature. We assessed language of publication for journal articles only. Using *Serline*, the journals

database produced by the National Library of Medicine (Bethesda, MA), we compiled a list of journals which only publish in English. The language of a journal article was classified as English if the journal publishing a trial report was included in this list. Articles that had a title in a language other than English or were described as of non-English language in the bibliographic details were classified accordingly. For all other articles we checked the language field in *Medline* or *Embase*. If a report could not be classified we obtained the report or contacted the authors of the meta-analysis.

### Assessment of methodological quality

Quality assessment was restricted to trials included in meta-analyses published in the *Cochrane Database of Systematic Reviews* and was based on information on concealment of allocation and blinding provided in the reviews. Two of us (PJ, FH) independently reviewed this information for each trial while unaware of trial results. For concealment of allocation we distinguished between adequately concealed trials (central randomization, coded drug packs, assignment envelopes, etc), and inadequately or unclearly concealed trials which either reported an inadequate approach (alternation, open random number tables, etc) or lacked a statement on concealment.<sup>17</sup> For blinding we distinguished between trials which were described as double-blind or included blinding of the person assessing outcomes (assessor-blind), and those which did not. Inter-observer reliability was determined using the kappa statistic.<sup>18</sup>

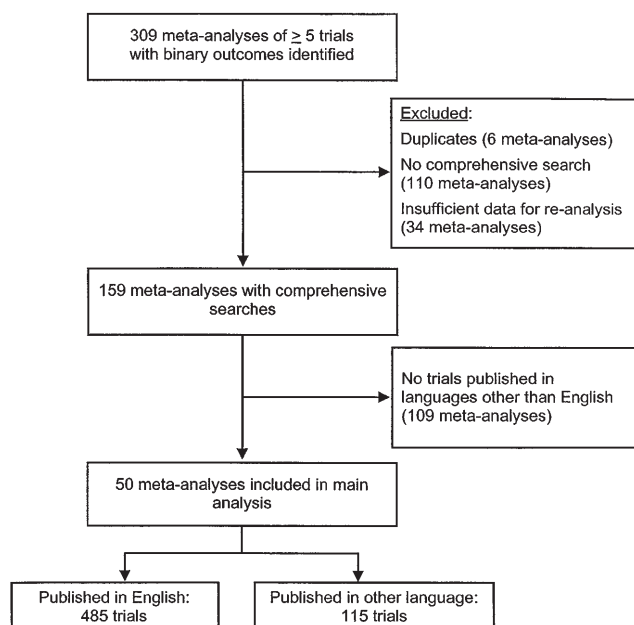
### Data extraction and statistical analysis

For each meta-analysis, we recorded the outcome, the statistical method used for combining trials, the type of effect measure used and the overall pooled estimate with its 95% CI. One of us (CB) abstracted the raw outcome data for each trial or, if raw data were unavailable, the point estimate and CI. For the meta-analyses published in the *Cochrane Database of Systematic Reviews*, Update Software (Oxford, UK) provided raw data in electronic form.

We included meta-analyses that contained at least one trial published in a language other than English, excluding unpublished trials. To obtain consistency across meta-analyses, endpoints were re-coded if necessary, so that odds ratios or relative risks below 1 indicated a beneficial effect of treatment. We calculated the combined effect estimates separately for the non-English and the English language trials, applying the same analytical method used by the original authors. We then derived a ratio of estimates of non-English language to English language trials: a ratio below one indicates that non-English language trials show a more beneficial treatment effect than English language trials. We combined ratios of estimates of treatment effects using random-effects meta-analysis, also stratifying by clinical area, source (meta-analyses published in the *Cochrane Database of Systematic Reviews* versus others), intervention (drugs versus others), type of control (active control intervention versus others), and complementary versus conventional medicine. We calculated the percentage weight contributed by non-English language trials to individual meta-analyses, and the percentage change in the combined estimate of treatment effect that occurred when non-English language trials were excluded from the meta-analysis and examined changes in *P*-values. All analyses were performed in Stata version 6.0 (Stata Corporation, College Station, Texas).

## Results

We identified a total of 309 meta-analyses with at least 5 trials and a binary outcome. After excluding 6 Cochrane reviews also published in journals we identified 159 meta-analyses which employed comprehensive literature searches, of which 50 (31.4%) included at least one trial published in a non-English language and were included in analyses (Figure 1). The number of meta-analyses including non-English language trials were 29 (25.0%) of 116 meta-analyses published in



**Figure 1** Progress through the stages of identifying eligible meta-analyses which included trials published in languages other than English

the *Cochrane Database of Systematic Reviews*, 12 (46.2%) of 26 meta-analyses published in general medicine journals and 9 (52.9%) of 17 meta-analyses published in specialist journals. The 50 meta-analyses included 671 trials; 600 were published in 208 English-language and 95 non-English language journals and analysed further; 71 were unpublished literature and excluded.

### Characteristics of trials

The language of publication was English in 485 (80.1%) trials. Of the 115 trials published in non-English languages, 42 (36.5%) were published in German, 29 (25.2%) in French, 12 (10.4%) in Italian, 8 (7.0%) in Japanese, 7 (6.1%) in Spanish, 6 (5.2%) in Portuguese, 8 (7.0%) in four other European languages and 3 (2.6%) in Chinese. Characteristics of trials were similar with respect to the year of publication and the type of intervention and comparison. Non-English language trials included fewer participants but were more likely to show statistically significant results (Table 1). The proportion of trials published in languages other than English varied widely across clinical topics, from 10.1% in tobacco addiction to 35.7% in rheumatology and orthopaedics (Table 2). It was higher in complementary medicine (41.2%) than in conventional medicine (21.7%). Cochrane reviewers' assessment of concealment of allocation was available for 294 trials (49.0%), of blinding for 279 trials (46.5%). Inter-observer reliability was high with kappas of 0.89 (95% CI: 0.80–0.98) for concealment of allocation and 0.76 for blinding (95% CI: 0.67–0.84). As shown in Table 3, English-language trials were of higher methodological quality.

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

Figure 2 shows the ratios of estimates of treatment effects from non-English language trials compared to English language trials for the 50 meta-analyses. Treatment effect estimates were on average 16% more beneficial in non-English language trials

**Table 1** Characteristics of randomized trials published in English and languages other than English

	English language report (n = 485)	Non-English language report (n = 115)	P
<b>Source of meta-analysis</b>			0.85
Cochrane Database of Systematic Reviews	232 (47.8%)	52 (45.2%)	
General medicine journal	160 (33.0%)	41 (35.7%)	
Specialist journal	93 (19.2%)	22 (19.1%)	
<b>Year of publication of trial</b>			
Mean (SD)	1986 (7)	1986 (6)	0.59
Median (Range)	1987 (1955–1998)	1987 (1970–1996)	0.24
<b>Type of intervention and comparison</b>			
Drug intervention	411 (84.7%)	103 (89.6%)	0.19
Complementary medicine	20 (4.1%)	14 (12.2%)	0.001
Active control intervention	117 (24.1%)	31 (27.0%)	0.53
<b>Sample size of trial</b>			
Mean (SD)	269 (487)	147 (195)	0.009
Median (Range)	116 (8–4524)	88 (19–1340)	0.0063
<b>Statistical significance of trial</b>			
P < 0.05	152 (31.3%)	48 (41.7%)	0.033
P < 0.01	89 (18.4%)	34 (29.6%)	0.007

P-values from  $\chi^2$  tests, t-tests or Wilcoxon rank sum tests.

**Table 2** Proportion of trials published in English and in languages other than English in different disease areas

Disease area	English-language report	Non-English language report	All trials
Tobacco addiction	62 (89.9%)	7 (10.1%)	69 (100%)
Obstetrics and gynaecology	64 (87.7%)	9 (12.3%)	73 (100%)
Cardiology and angiology	118 (86.8%)	18 (13.2%)	136 (100%)
Infectious disease	109 (79.6%)	28 (20.4%)	137 (100%)
Neurology	42 (77.8%)	12 (22.2%)	54 (100%)
Psychiatry	26 (65.0%)	14 (35.0%)	40 (100%)
Rheumatology and orthopaedics	36 (64.3%)	20 (35.7%)	56 (100%)
Miscellaneous	28 (80.0%)	7 (20.0%)	35 (100%)

$P < 0.001$  by  $\chi^2$  test.

**Table 3** Methodological quality of trials included in Cochrane reviews

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

Denominators differ: information on concealment of allocation was provided more frequently than information on blinding.

Probability values by  $\chi^2$  tests.

(ratio of estimates 0.84, 95% CI : 0.74–0.97,  $P = 0.011$ ). However, there was considerable heterogeneity between meta-analyses ( $P = 0.003$ ). Results of stratified analyses are presented in Figure 3. The effect of language appeared to be more pronounced in complementary medicine and less pronounced in trials with active control interventions, but none of the differences between strata was statistically significant ( $P > 0.20$ ).

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

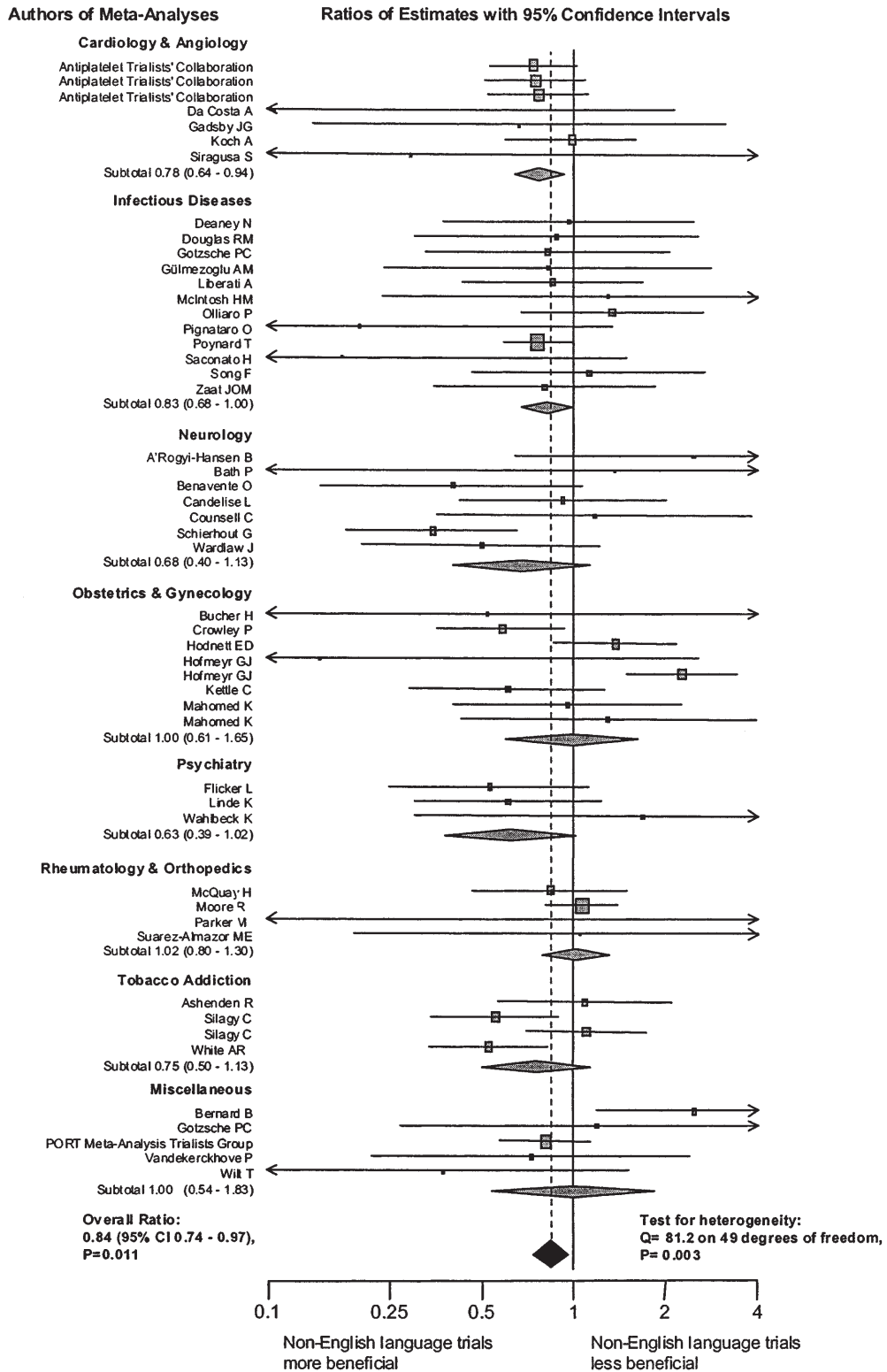
The number of trials published in languages other than English ranged from one to 14 trials and from 4.3% to 72.7% of all trials included. Non-English language trials contributed an average 17.5% of the weight in individual meta-analyses (median 10.2%; range 1.2–81.1%). The average precision (the inverse of the standard error) of treatment effect estimates decreased from 8.34 to 7.68 after exclusion of non-English language trials. Figure 4 shows the change in pooled estimates of individual meta-analyses that occurred when non-English language trials were excluded from meta-analyses. The changes ranged from a 42.0% increase (indicating less benefit) to a 22.7% decrease (indicating more benefit) of estimates of treatment effects. In 29 (58.0%) meta-analyses changes were less than 5%. Among the remaining 21 meta-analyses 5 showed more benefit and 16 less benefit after exclusion of non-English language trials. Significance levels were affected in 9 (18.0%) meta-analyses. In three cases  $P$  increased from  $P < 0.001$  to  $P < 0.01$ , in a further four cases  $P$  increased from  $P < 0.01$  to  $P < 0.05$  and in two instances  $P$  decreased from  $P < 0.05$  to  $P < 0.01$ .

## Discussion

In an ideal world reviews of medical research would always include all relevant studies, independent of the language of

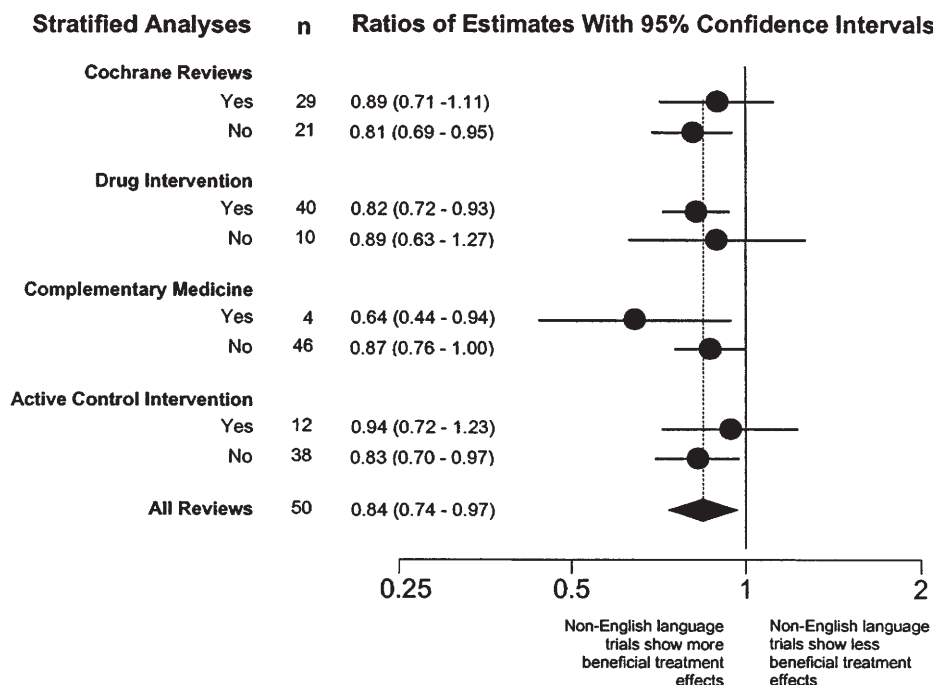
publication. The inclusion of studies published in languages other than English could avoid bias<sup>6,19</sup> and may often add valuable additional information. However, trials published in other languages can be more difficult to locate, and may require translation, which will increase costs and delay the conclusion of a review. Although performing reviews that produce misleading results is never justified, there may be trade-offs between the timeliness, costs and quality of systematic reviews. We examined the importance of including trials published in languages other than English in rigorously conducted systematic reviews by examining the effect of excluding these trials on overall estimates of treatment effects and the conclusion of the reviews.

Of 309 meta-analyses identified by our search around half employed comprehensive literature searches that were free of language restriction. Moher *et al.* found that 41 (52%) out of 79 meta-analyses were 'language inclusive', i.e. authors did not report any restrictions.<sup>20</sup> Conversely, a 1995 survey found that 26 (72%) out of 36 meta-analyses had restricted their search to studies published in English.<sup>19</sup> Search strategies may thus have become more comprehensive in recent years. In our study only 50 (31.4%) of the 159 meta-analyses which reported comprehensive searches had in fact located reports published in languages other than English. Our study nevertheless included 485 English-language trials and 115 trials published in 11 other languages. We found that treatment-effect estimates from trials published in non-English languages were on average 16% more beneficial but the heterogeneity we observed between different meta-analyses means that both the size and the direction of this difference are unpredictable. Although trials published in languages other than English were smaller, they were more likely to report statistically significant results than trials published in English. However, in the majority of meta-analyses excluding reports published in other languages did not change estimates of treatment effects substantially although the



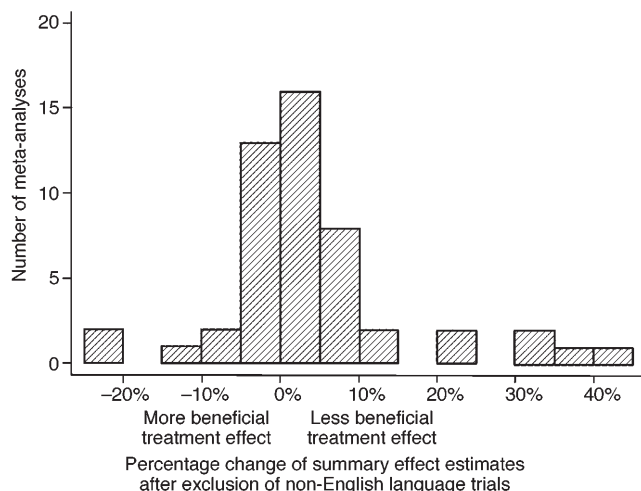
**Figure 2** Ratios of estimates of treatment effects from non-English language trials compared to English language trials for 50 meta-analyses. Ratios of estimates (grey squares) with 95% CI of individual meta-analyses are shown. The size of the square reflects statistical weight in the overall pooled analysis. Meta-analyses are grouped according to clinical topic, and arranged alphabetically according to the first author. The grey diamonds represent pooled results from clinical subgroups, the black diamond overall pooled results. Ratios of estimates were pooled using random-effects models. A ratio of estimates below one indicates that trials published in languages other than English show a more beneficial treatment effect than trials published in English.





**Figure 3** Ratios of estimates of treatment effects from non-English language trials compared to English language trials: stratified analyses

Ratios of estimates (circles) with 95% CI of individual strata are shown. The black diamond represents overall pooled results. Estimates were pooled using random-effects models. There was little evidence that ratios differed between strata ( $P > 0.20$ ).



**Figure 4** Percentage change of treatment effect estimates of individual meta-analyses after exclusion of non-English language trials

A negative change indicates that the ratio became smaller after excluding non-English trials, thus indicating a more beneficial effect. A positive change indicates the opposite.

precision of estimates was somewhat reduced. More substantial changes did occur in some instances; usually this meant that estimates of treatment effects were more conservative after excluding trials published in languages other than English.

Our study relied on the accuracy of meta-analysts' description of their literature searches: we did not assess whether the

sample of trials identified by these authors was in fact complete. If searches were inadequate, so that many trials published in languages other than English were omitted, then our results might underestimate the contribution of this literature. Our sample was, however, large and our inclusion criteria well defined and stringent, reflecting current recommendations for comprehensive searches. The results reported here should thus reflect what is gained or lost by attempts to identify trials published in languages other than English for meta-analyses. Future studies could prospectively compare the results from rapid reviews that are restricted to the English language with subsequent meta-analyses based on extensive searches without language restrictions. We also relied on the information on study quality provided by many different Cochrane reviewers. However, the *Cochrane Reviewers' Handbook* specifies a standardized method for assessment of trial quality.<sup>9</sup>

Our methods differ in two respects from previous studies which combined results from many meta-analyses.<sup>17,21</sup> First, we used the statistical methods of the original meta-analysis. For example if the authors used a random-effects model because of the presence of between-trial heterogeneity in their review then so did we. We were thus able to focus on the impact of omission of non-English literature on the meta-analysis as actually published. Second, we allowed for differences in the disparity between treatment effects in English and non-English trials between different meta-analyses, and found clear evidence of such differences. Previous studies<sup>17,21</sup> have statistically combined different meta-analyses assuming no differences between meta-analyses, an approach which may exaggerate statistical precision.

In an earlier investigation we examined factors predicting the language of publication for pairs of reports of randomized controlled trials, with one report published by the same author in German and the other in English.<sup>6</sup> A statistically significant result was the only characteristic that predicted publication in an English-language journal. Based on these findings we hypothesized that significant findings are over-represented in the English-language literature whereas more non-significant results would be found in journals published in other European languages. The present study not only failed to confirm this prediction but showed that articles published in languages other than English were more likely to report statistically significant findings. Trialists in German-speaking Europe who publish both in English and German may thus not be representative of the majority of authors publishing clinical trials in languages other than English. The proportion of published trials showing superior efficacy of the experimental treatment has been shown to vary from country to country. Vickers and colleagues examined 252 abstracts of clinical trials of acupuncture and 405 abstracts from trials of other interventions.<sup>22</sup> They found unusually high proportions of trials favouring experimental treatments in some countries, for example China, Russia and Taiwan. Our sample included only few reports published in these countries but our results indicate that journals published in Western Europe may also contain a relatively high proportion of 'positive' trials.

Assessments by Cochrane reviewers found non-English language trials to be of lower methodological quality than English language trials. Two recent studies examined to what extent estimates of treatment effects from clinical trials are affected by dimensions of methodological quality.<sup>17,21</sup> In both studies inadequate concealment of treatment allocation was, on average, associated with an exaggeration of treatment effects by around 40%. We also found inadequate methodological quality to be associated with larger effects (data available on request). The lower quality of trials may therefore partly explain the more beneficial treatment effects observed in trials published in languages other than English. This must be of concern: bias could thus be introduced by including trials published in languages other than English, leading to overoptimistic assessments of treatment effects. The methodological quality and quality of reporting was fairly poor in both language groups, however, and our findings underscore the importance of a sound assessment of trial quality in meta-analyses.<sup>23</sup> At present, trial reports frequently omit important methodological detail,<sup>24-29</sup> a situation which will hopefully improve in the future with a more widespread adoption of the CONSORT guidelines.<sup>30,31</sup> Special efforts may be needed to improve reporting of clinical trials in journals published in languages other than English.

Our findings on study quality contrast with the results from an earlier study by Moher *et al.*<sup>24</sup> Moher *et al.* compared 133 trials published in English with 98 trials published in other languages during 1992 to 1994 and found little differences in reporting and overall quality. Their study was based on 13 selected journals of relatively high impact whereas our sample included a much wider range of journals (208 journals published in English and 95 journals published in other languages). Moher *et al.*<sup>24</sup> used the scale developed by Jadad *et al.*<sup>32</sup> to gauge quality. This scale gives more weight to the quality of reporting, that is the extent to which a report of a clinical trial provides adequate information about the design, conduct, and analysis of the trial than to actual

methodological quality. Furthermore, the Jadad scale addresses the generation of allocation sequences, a domain not consistently related to bias,<sup>17,21</sup> but it does not assess allocation concealment, which has been shown to be associated with exaggerated treatment effects.<sup>17,21</sup> It thus seems likely that the discrepant findings are explained by differences in the samples examined and quality features assessed. It could be argued that the different time period covered in our study might explain the discrepant findings, however, we found that differences between language groups in trial quality were in fact more pronounced in the 1990s.

What are the implications for the conduct of future reviews? Could reviews that are performed in a short period of time but ignore the non-English language literature still produce valid and reasonably precise results? In many situations the answer may be yes, particularly in specialties where most relevant trials appear to be published in English, for example in cardiology or obstetrics and gynaecology. The importance of trials published in non-English language journals is well known in complementary medicine, for example homoeopathy<sup>33,34</sup> or phytotherapy.<sup>35</sup> Within a specialty the situation may depend on the disease: about 80% of trials included in meta-analyses in neurology were published in English, however, a recent manual search of Chinese journals yielded 166 randomized trials in neurology the majority of whom (70%) were in stroke.<sup>36</sup> We emphasize that our study was designed to examine the overall impact of the non-English language literature. Further studies should clarify its importance in different specialties and conditions.

Important considerations support the inclusion of all relevant trials of acceptable methodological quality in systematic reviews. The inclusion of trials published in many different languages will often increase the precision, generalizability and applicability of findings. The effect of excluding non-English language trials on summary estimates is unpredictable and the exclusion of trials on the grounds of language alone runs against the principles and spirit of systematic reviews, discriminates against some investigators and countries and will always introduce an element of doubt. However, our findings indicate that in many cases exclusion of non-English literature makes little practical difference and if anything will lead to more conservative estimates of treatment effects.

## Acknowledgements

We thank Jos Kleijnen and the staff of the NHS Centre for Reviews and Dissemination for the supply of articles from the DARE database, Mark Starr from Update Software for kindly providing raw data from the Cochrane Database of Systematic Reviews, Carol Lefebvre of the UK Cochrane Centre for bibliographical advice, and Guido Schwarzer and Deborah Tallon for preliminary work which was partly drawn upon for this study. We also thank Joanna Wardlaw and Lesley Stewart for guidance on issues arising in specific meta-analyses. We are grateful to Doug Altman, Gerd Antes, Iain Chalmers, Mike Clarke, Philippa Middleton and David Moher for helpful comments on a previous draft of the manuscript. Peter Jüni was supported by the Swiss National Science Foundation. The views and opinions expressed are those of the authors and do not necessarily reflect those of the Department of Health. The project was funded by the UK National Health Service Health Technology Assessment Programme (Project No: 97/18/05).

## KEY MESSAGES

- Studies published in languages other than English are often *a priori* excluded from systematic reviews and meta-analyses, which may introduce bias.
- This study shows that trials published in languages other than English tend to be of lower quality and produce more favourable treatment effects than trials published in English.
- Excluding non-English language trials has generally only modest effects on summary treatment effect estimates but the effect is difficult to predict for individual meta-analyses.
- Comprehensive literature searches followed by a careful assessment of trial quality are required to assess the contribution of all relevant trials, independent of language of publication.

## References

- Egger M, Davey Smith G, O'Rourke K. Rationale, potentials and promise of systematic reviews. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-Analysis in Context*. London: BMJ Books, 2001, pp.23–42.
- Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997;**315**: 629–34.
- LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;**337**:536–42.
- Egger M, Davey Smith G. Meta-analysis: bias in location and selection of studies. *Br Med J* 1998;**316**:61–66.
- Easterbrook PJ, Berlin J, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;**337**:867–72.
- Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;**350**:326–29.
- Tramèr MR, Reynolds DJM, Moore RA, McQuay HJ. Impact of covert duplicate publication on meta-analysis: a case study. *Br Med J* 1997;**315**:635–40.
- Gøtzsche PC. Reference bias in reports of drug trials. *Br Med J* 1987; **295**:654–56.
- Clarke M, Oxman AD (eds). *Cochrane Reviewers' Handbook 4.1* [updated June 2000]. In: *The Cochrane Library* [database on CDROM]. The Cochrane Collaboration, Oxford: Update Software, 2001, issue 2.
- NHS Centre for Reviews and Dissemination. *Undertaking Systematic Reviews of Research or Effectiveness. 2nd Edn*. York: Publications Office, CRD, University of York, 2001.
- Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 1995; **48**:167–71.
- Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet* 1998;**351**:47–52.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999;**354**:1896–900.
- Health Technology Assessment. A national R&D programme for the NHS. <http://www.hta.nhsweb.nhs.uk/index.htm> (accessed September 2001).
- Database of Abstracts of Reviews of Effectiveness (DARE). In: *The Cochrane Library* [database on CDROM]. The Cochrane Collaboration, Oxford: Update Software, 2001, issue 2.
- The Cochrane Database of Systematic Reviews. In: *The Cochrane Library* [database on CDROM]. The Cochrane Collaboration, Oxford: Update Software, 2001, issue 2.
- Schulz KF, Chalmers I, Hayes RJ, Altman D. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; **273**:408–12.
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;**86**:420–28.
- Grégoire G, Derderian F, LeLorier J. Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *J Clin Epidemiol* 1995;**48**:159–63.
- Moher D, Pham B, Klassen TP *et al*. What contributions do languages other than English make on the results of meta-analyses. *J Clin Epidemiol* 2000;**53**:964–72.
- Moher D, Pham B, Jones A *et al*. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;**352**:609–13.
- Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Contr Clin Trials* 1998;**19**:159–66.
- Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trial for meta-analysis. *JAMA* 1999;**282**:1054–60.
- Moher D, Fortin P, Jadad AR *et al*. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;**347**: 363–66.
- Schulz KF, Grimes DA, Altman DG, Hayes RJ. Blinding and exclusions after allocation in randomised controlled trials: survey of published parallel group trials in obstetrics and gynaecology. *Br Med J* 1996;**312**: 742–44.
- DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;**306**:1332–37.
- Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *Br Med J* 1999;**319**: 670–74.
- Schulz KF, Chalmers I, Grimes DA, Altman D. Assessing the quality of randomization from reports of controlled trials published on obstetrics and gynecology journals. *JAMA* 1994;**272**:125–28.
- Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *Br Med J* 1998;**317**:1181–84.
- Moher D, Schulz KF, Altman DG, for the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Lancet* 2001;**357**: 1191–94.
- Altman DG, Schulz KF, Moher D *et al*. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;**134**:663–94.
- Jadad AR, Moore RA, Carrol D *et al*. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Contr Clin Trials* 1996;**17**:1–12.



<sup>33</sup> Kleijnen J, Knipschild P, ter Riet G. Clinical trials of homoeopathy. *Br Med J* 1991;**302**:316–23.

<sup>34</sup> Linde K, Clausius N, Ramirez G *et al.* Are the clinical effects of homoeopathy placebo effects? A meta-analysis of placebo-controlled trials. *Lancet* 1997;**350**:834–43.

<sup>35</sup> Kleijnen J, Knipschild P. *Ginkgo biloba*. *Lancet* 1992;**340**:1136–39.

<sup>36</sup> He L, Liu M. *A Report of Handsearching Chinese Neurological Journals*. 5th Annual Cochrane Colloquium, October 1997, Amsterdam.

## Commentary: Searching for trials for systematic reviews: what difference does it make?

Mike Clarke

Systematic reviews are increasingly common as a means of assessing the relative effectiveness of health care interventions. This is not least because of the formation of the Cochrane Collaboration with its aim to prepare, maintain and promote the accessibility of reviews in all areas of health care.<sup>1</sup> Just as systematic reviews contribute to the practice of evidence-based health care, so should the conduct of reviews be based on good evidence, ideally from empirical research. The paper by Juni *et al.*<sup>2</sup> adds to this evidence base.

For all but the last century, decisions on how to treat patients were almost always based on personal experience, anecdotal case histories and highly selective retrospective comparisons; patients who had been treated in one way compared with others who had been treated differently. These processes are subject to many biases. The adoption of the randomized trial provided a more reliable way to estimate the relative effects of different interventions. However, for many conditions, the results of a single trial will rarely be sufficient. Most trials are too small, or too focused on a particular type of patient to provide a result that is either easily or reliably applicable to future patients.

To overcome this, a number of trials need to be brought together and this needs to be done reliably. It should minimize the possibility that chance will dominate the results (by combining the maximum amount of relevant information), minimize the possibility that bias will dominate the results (by combining an unbiased set of trials) and maximize the ability of users of the review to judge its applicability for specific health care decisions (by combining as wide a variety of trials as sensible and possible). Ideally, this would require that all relevant randomized trials are identified and included in the review.<sup>3</sup> Practically, though, this might rarely be possible given the time and resources needed to identify all trials and to obtain complete information and data from them.<sup>4</sup>

Trial identification is one of the most important steps in the conduct of a systematic review. It needs to ensure that as many as possible of the relevant trials are found and, in particular, that a biased set of trials, where the bias stems from the results of the trials, is not identified. Publication bias, in which trials are more likely to be published if they have statistically significant positive results, makes this especially important. Coupling this with the fact that trials with more positive results are more likely to be published in English,<sup>5</sup> it implies that searching must encompass publications in other languages.

Juni *et al.* investigated a sample of meta-analyses published in the mid-1990s to assess the effect of including trials published in languages other than English in these. They found that, on average, the exclusion of such trials would have had little effect on the summary effect estimates.<sup>2</sup> However, as they point out, the size or direction of the effect was difficult to predict for individual reviews and they conclude that comprehensive searching is still required. I agree with this conclusion but, is it possible that they have actually underestimated the effect of trials published in languages other than English?

The meta-analyses chosen for the research had to include a minimum of five trials for sound statistical reasons. These are, as a consequence, more robust to the effects of removing one or two trials than meta-analyses with fewer trials. Thus it is difficult to know how applicable the results of this research might be for meta-analyses that include fewer than five trials. Such meta-analyses are not uncommon and it might be expected that the exclusion of, for example, one trial from a total of three could have a much larger effect.

The included trials that were published in languages other than English were typically smaller and had more positive results than the trials published in English. This might be an indication that the former were more subject to publication bias than the latter. This could result from the fact that, in general, trials reported in English were more readily available in the period when the reviews under investigation were done,

because for example of the searching of electronic databases such as MEDLINE, which have a preponderance for articles published in English. This means that the trials published in languages other than English that were most likely to be found by reviewers—given the time and resource difficulties of searching extensively in the non-English-language literature—would be those that were published and cited with greatest prominence. Namely, the trials that were particularly positive and, most likely, particularly small.

This raises a very challenging question. Was it worthwhile for the reviewers in the sample studied by Juni *et al.* to have done what was probably a limited amount of searching for trials published in languages other than English, if this identified a biased set, albeit containing a relatively small amount of randomized evidence? Fortunately, recent developments have made this question less pertinent but it still needs to be borne in mind by the people who do reviews and the people who use them.

The two most important developments are the continuing growth in the number of Cochrane reviews and the increased ease of trial identification in recent years. Cochrane reviews are published electronically in full and can therefore be revised and updated as additional evidence comes to light. This evidence might be from new research or it might be newly found evidence from older research. Trial finding has become easier with the development of *The Cochrane Controlled Trials Register* in *The Cochrane Library*. This incorporates records for reports of randomized trials regardless of language restrictions and one of its main sources is the hand searching of thousands of journals and other sources around the world. This searching, which would be impossible within the context of an individual review, is done in such a way that each source is searched for all reports of studies that might be randomized trials and the results of this searching are then incorporated into *The Cochrane Controlled Trials Register*.<sup>6</sup>

Therefore, in agreeing with Juni *et al.* that comprehensive searching should remain an important component of systematic reviews, I also hope that reviewers of the mid-2000s will discover that it is now much easier to do this. The remaining challenge for reviewers will then be how to extract or obtain the necessary information and data from trials that have been published in languages other than their own.

## Acknowledgements

Mike Clarke is Associate Director (Research) at the UK Cochrane Centre, which is part of the NHS Research and Development Programme. The views expressed in this editorial are his own and are not necessarily the views or the official policy of the Cochrane Collaboration.

## References

- <sup>1</sup> Clarke M, Langhorne P. Revisiting the Cochrane Collaboration. *Br Med J* 2001;**323**:821.
- <sup>2</sup> Juni 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.
- <sup>3</sup> Clarke MJ, Stewart LA. Obtaining data from randomised controlled trials: how much do we need to perform reliable and informative meta-analyses? *Br Med J* 1994;**309**:1007–10.
- <sup>4</sup> Egger M, Davey Smith G. Bias in location and selection of studies. *Br Med J* 1998;**316**:61–66.
- <sup>5</sup> 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**:326–29.
- <sup>6</sup> Lefebvre C, Clarke MJ. Identifying randomised trials. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in Context*. London: BMJ Books, 2001, pp.69–86.