Finding evidence to inform the conduct and reporting of systematic reviews of randomized trials

Moher, D.

Publication date
2004

Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews

David Moher, Paul Fortin, Alejandro R Jadad, Peter Jüni, Terry Klassen, Jacques Le Lorier, Alessandro Liberati, Klaus Linde, Angelo Penna

Lancet 1996; 347:363-6

Summary

Background Lately, the number of systematic reviews published has increased substantially. Many systematic reviews exclude trials published in languages other than English. However, there is little empirical evidence to support this action. We looked for differences in the completeness of reporting between trials published in other languages and those published in English, to see whether the exclusion of trials published in other languages is justified.

Methods We compared completeness of reporting, design characteristics, and analytical approaches of 133 randomised controlled trials (RCTs) published in English between 1989 and 1994 and 96 published in French, German, Italian, or Spanish during the same time. RCTs were identified by hand searching of journals (seven in English and six in the other languages).

Findings We found no significant differences between trials published in English and other-language trials for any single item in the completeness of reporting scale (□moxicillin□a, double-blinding, withdrawals), or for the overall score (percentage of maximum possible score 51.0% for trials in English, 46.2% for trials in other languages; 95% CI for difference–1.1 to 10.5). Other-language trials were more likely than English-language trials to have adult participants, to use two or more interventions, and to compare two or more active treatments without an untreated control group. Trials in other languages were less likely to report a clearly prespecified primary outcome or any rationale for sample size estimation.

Interpretation These results provide evidence for inclusion of all trial reports, irrespective of the language in which they are published, in systematic reviews. Their inclusion is likely to increase precision and may reduce systematic errors. We hope that our findings will prove useful to those developing guidelines and policies for the conduct of reporting of systematic reviews.

Introduction

Evidence-based health care ideally involves the systematic collection, synthesis, and application of scientific evidence to guide clinical practice and policy-making. Systematic reviews are a key component of evidence-based health care. Over the past few years the number of systematic reviews published has increased substantially.[1]

Systematic performance of a review has the potential advantage of keeping biases to a minimum and improving the precision of its result. Systematic reviewers have little control over random errors but much control over systematic ones. Inclusion of only a selection of all possible evidence is likely to introduce systematic errors (biases), thus threatening the validity of the systematic review-the extent to which its conduct has guarded against these biases. There is evidence that most systematic reviews do not include all potential evidence. Grégoire and colleagues reported that 78% of the meta-analyses they identified had language restrictions.[2] Most (93%) of these restrictions led to the exclusion of □moxicilli controlled trials (RCTs) reported in languages other than English. Perhaps these language restrictions were applied because of difficulties in identifying trials published in languages other then English or the presumed greater importance and quality of English-language publications.

One way to assess whether language restrictions are a sensible policy for systematic reviewers is to assess the completeness of reporting of RCTs. Language restrictions might be appropriate if trials published in other languages are reported less completely than those published in English. On the other hand, if the completeness of reporting of English-language and other-language trials is similar, there would be empirical evidence for the non-a-priori language-based exclusion of RCTs from systematic reviews and for those developing recommendations and policies on how they are conducted. We set out to assess whether the completeness of reporting of trials differs significantly between those published in English and those published in other languages.

Methods

Before the start of the project we defined the completeness of a trial report[3] as providing information about the design, conduct, and analysis of the trial that should avoid biases in its treatment comparisons. In this study we limited ourselves to the assessment of completeness of reporting of □moxicillin□a, double-blinding, and dropouts and withdrawals. This was the a-priori primary outcome measure in this study. Our secondary outcome variable was the reporting of several design characteristics and analytical approaches.

Thirteen journals classified by the Science Citation Index[4] as general and internal medicine were selected and searched by hand. Four English-language journals were ranked in the top ten according to their citation impact factor, three other English-language journals were ranked within the top 40, and the six journals published in other languages were ranked first or second according to language (all within the top 90).

Table 1. Completeness-of-reporting scores of trials

	% of maximum possible score* for trials in:		
	Other languages	**English**	**% difference (95% CI)**
All trials	(n=96)	(n=133)	
Randomization	23.8	23.9	0.1 (-2.6 to 2.7)
Double-blinding	12.1	15.6	3.5 (-0.8 to 7.9)
Withdrawals	10.4	11.4	1.0 (-1.6 to 3.6)
Total score	46.2	51.0	4.8 (-1.1 to 10.5)
Inadequately reported trials	(n=61)	(n=71)	
Randomization	22.3	20.6	1.7 (-1.1 to 4.6)
Double-blinding	2.6	4.2	1.6 (-1.0 to 4.2)
Withdrawals	7.2	8.7	1.5 (-1.9 to 4.9)
Total score	32.1	33.5	1.4 (-2.2 to 5.0)
Adequately reported trials (score>2)	(n=35)	(n=62)	
Randomization	26.3	27.7	1.4 (-2.8 to 5.7)
Double-blinding	28.6	28.7	0.1 (-5.7 to 6.0)
Withdrawals	16.0	14.5	1.5 (-2.2 to 5.1)
Total score	70.9	71.0	0.1 (-5.3 to 5.6)

*5 points=100%; 40% each for randomization and double-blinding, and 20% for withdrawals

To identify trials published in languages other than English (French by PF, J LeL, PJ; German by PJ and KL; Italian by AL and TK), hand searching was done from Dec 31, 1993, backwards until 20 RCTs from each journal were identified or until the 1989 publication year. There is evidence that electronic searching is not sensitive enough to identify all relevant RCTs.[5] We chose 1989 as the cut-off year because the quality of reporting of clinical trials has improved over time.[6] To be classified as an RCT the report had to state that the participants were randomly assigned to their respective intervention groups.

Completeness of reporting was assessed on a scale developed with appropriate rigorous standards.[7] Briefly, the report describes how the items were

initially selected, how and why the final items were included, how the scale discriminates between trials of differing quality, and what ranges of scores were identified during its development. The scale consists of three items pertaining to descriptions of □moxicillin□a, double-blinding, and dropouts and withdrawals as described in the report of an RCT. The scale ranges from zero to five (two points each for □moxicillin□a and double-blinding, and one point for withdrawals) with higher scores indicating better reporting. In addition, we recorded whether the adequacy of allocation concealment was described.

Once RCTs had been identified, information about design characteristics and analytical approaches was extracted from each trial report: number of treatment groups (two or more), design (parallel group *vs* other), sex and age range of participants (adult *vs* other), type of interventions (pharmacological vs other), control group comparisons (placebo or active) and outcomes, sample size and rationale for its estimation, whether a primary outcome was specified (yes/no), type of primary outcome assessed (mortality and/or morbidity *vs* other specific), number and handling of withdrawals, and whether the trial was reported as statistically positive or negative.

Completeness of reporting and design characteristics and analytical approaches were assessed, in the language of the trial report. Information on author, author affiliation, all journal identifications, references, acknowledgements, and locations in which the trial had been conducted was concealed from the assessors by means of a black marker. Chalmers et al[8] suggested more than ten years ago that the quality of clinical trial reports should be assessed under blind conditions. Empirical evidence to support this recommendation has been produced.[7]

Before the completeness of reporting of the trials included in this study was assessed, a separate set of ten trials published in English (language common to all investigators) was sent to each of us for evaluation. Inter-observer reliability was assessed with the intraclass correlation coefficient (values above 0.65 indicate high reliability, a priori).[9] We also established criteria for adequate (>2 points) and inadequate (≤2 points) completeness of reporting.

Table 2. Design characteristics of RCTs

	% of maximum possible score* for trials in:		
	Other languages (n=96)	**English (n=133)%**	**% difference (95% CI)**
Demographic			
Female	50.8	52.6	2.3 (-7.5 to 12.1)
Adult	57.9	46.4	11.5 (1.3 to 21.8)
Descriptive			
Assessed change in specific outcome*	36.2	59.8	23.6 (11.0 to 36.3)
Intervention)			
Parallel groups	87.5	85.6	1.9 (-7.0 to 10.8)
Two intervention groups	88.5	77.7	10.8 (1.3 to 20.4)
Comparison group receiving active treatment	59.4	39.1	20.3 (7.4 to 33.1)

*For example: pain-free walking distance in patients with claudication

To estimate (a priori) the appropriate sample size, for the primary outcome, we used previously reported date.[7] Trials with adequate and inadequate completeness of reporting had mean scores of 3.4 (SD 0.26) and 0.7 (0.24), respectively. We calculated that with the sample size we had we would have more than 95% statistical power to detect mean differences between adequate and inadequate completeness of reporting.[10]

To assess mean differences between trials published in English and those published in other languages in their completeness of reporting we used independent *t* tests (two-sided). Unless otherwise stated, all other outcomes were assessed by t tests of $_x^2$ analysis. We made no adjustments for multiple comparisons. All outcomes are reported as absolute differences between the two types of trials with 95% CI.

Results

We established substantial agreement among ourselves in assessing completeness of reporting of the same set of English-language trials (intraclass coefficient 0.7 [95% CI 0.5-0.8]). We identified 133 RCTs from seven English-language journals-*British Medical Journal* (20), *Canadian Medical Association Journal* (20), *The Lancet* (20), *Medical Journal of Australia* (20), *New England Journal of*

Medicine (20), and *New Zealand Medical Journal* (13)-and 96 RCTs from six journals published in other languages-*Deutsche Medizinische Wochenschrift* (20, German), *La Presse Médicale* (20, French), *Revista Medica de Chile* (9, Spanish), *Schweizerische Medizinische Wochenschrift* (20, French/German), *Minerva Medica* (20, Italian: searched from March 31, 1994), and *Revista Clinica Española* (7, Spanish). 73.8% of the RCTs identified were published between 1992 and 1994 (82.7% English, 61.5% other languages); 41.4% of the trials involved infectious diseases or diseases of the circulatory or endocrine systems (36.1% English, 46.9% other languages).

There were no statistically significant differences in completeness of reporting of trials published in languages other than English and those published in English as regards □moxicillin□at, double-blinding, dropouts and withdrawals, or overall total score (table 1). The differences in the completeness of reporting between English-language and other-language trials ranged from 0 to 4% for individual items and 5% in total score. Similarly, we found no differences between the two types of trials for any of the items used to assess completeness of reporting when we compared adequately or inadequately reported trials only (Table 1). Fewer than 7% of trials reported on allocation concealment (4.2% of trials published in languages other than English *vs* 6.8% of trials published in English; difference 2.6% [95% CI –1.4 to 7.7]).

Table 3. Analytical approaches

	% of trials using approach:			Number of trials for which approach could not be ascertained
	Other languages (n=96)	English (n=133)%	% difference (95% CI)	
Primary outcome specified	38.5	56.4	17.9 (4.8 to 30.8)	7
Rationale for sample size estimation stated	3.2	33.3	30.1 (21.4 to 39.0)	2
Trials in which there were withdrawals	74.6	86.5	11.9 (-0.9 to 24.7)	59
Participants randomized and included in analysis	92.8	93.7	0.9 (-5.5 to 7.2)	22
Statistically positive result	60.0	51.9	8.1 (-5.0 to 21.1)	3

There were differences between trials published in English and those published in other languages in demographic and descriptive characteristics (Table 2). A significantly lower proportion of trials published in languages other than English assessed specific outcomes, and significantly higher proportions had adult participants, used two or more intervention groups, and compared two or more active therapies without an untreated control group.

We also found some differences in the analytical approaches (Table 3). Trials published in languages other than English were significantly less likely than those published in English to report a clearly prespecified primary outcome or any rationale for their sample size estimation. Moreover, we could not ascertain whether there had been withdrawals in a quarter of all trials (25.8%; English 16.5%, other languages 38.5%).

A subgroup analysis revealed that amount trials with poor completeness of reporting (≤40% of possible maximum) statistically positive trial results were significantly more likely (odds ration 2.4 [95% CI 1.1-5.2]) to be reported in trials published in languages other than English than in those published in English. There was no such difference for trials with good completeness of reporting (0.8 [0.3-1.9]).

Discussion

The results of this study suggest that completeness of reporting about important features of study design and conduct does not differ between trials published in English and those published in other languages. We did observe some differences between the two types of trial in design characteristics and analytical approaches. More trials reported in English used a placebo group as the control comparison, had a primary outcome clearly specified, and had an estimate of the sample size needed to ascertain a given treatment difference. This difference may reflect more strict regulatory standards particularly in the USA and Canada, where 39% of such trials were done. It is also possible, however, that trials published in English are more methodologically sound and explore questions of greater clinical relevance.[11] This difference needs further investigation.

Systematic reviews are more likely to reach valid conclusions if their results are based on all available evidence, not only English-language evidence. Sole reliance on evidence published in English is likely to result in reduced precision and may be a subsequent loss in validity.

There is a common perception that systematic reviews can reasonably be limited to trials published in English, because those published in other languages represent small numbers of trials, have weaker methods, and report fewer significant results. Dickersin et al[5] used Medline to identify clinical trials on vision. They reported that 20% of relevant trials were published in languages other than English, but noted that this percentage was likely to be an underestimate of relevant trials because they did not search Embase, which includes several journals published in languages other than English that are not indexed by Medline.

We identified only very small differences in the completeness of reporting between trials published in English and in other languages. There is growing empicical evidence that the items included in our assessment of completeness of reporting are important predicators for biased estimates of treatment effect.[12]

Ottenbacher and Diffabio[13] observed that studies reported in journals published in the USA had larger treatment differences and more positive results than similar studies published in journals in other □moxicilli countries. Our results show that of trials with poor completeness of reporting, those published in languages other than English are more likely to report statistically positive results than trials published in English. These findings suggest that estimates of intervention effects may differ between the two types of trial and that trials published in languages other than English should be considered for inclusion in any systematic review. This result also provides further empirical evidence[8,13,14] of the need to do sensitivity analysis as part of a systematic review.

Our study focused on the completeness of reporting of trials. It is possible that investigators carried out their trials appropriately but did not report the relevant information. In one study[15] the internal and external validity of 63 reports of RCTs was assessed on a scale with a maximum total score of 100 points. The mean score for all trials was 50% (95% CI:46-54). To elaborate on various features of the trial reports the investigators conducted telephone interviews with 62 (of 63) corresponding authors. This resulted in a 7% (mean) improvement in the quality scores. Systematic reviewers, however, have to rely on trial reports especially if a large number of trials must be considered for inclusion. Telephone costs and other logistical factors could preclude further contact with corresponding authors.

We □moxicill that many systematic reviewers may have avoided including trials published in languages other than English in their reviews because of anglocentricity (a minority, we hope), a search strategy that excludes the identification of such trials, lack of awareness of the evidence about the completeness of reports of trials published in other languages, and the logistical and cost barriers of translation. There are several approaches to examine for assistance in translation and clarification of issues in trials published in languages other than English; many health-care facilities have registers of individuals fluent in other languages, students associated with university languages departments, and immigrant community centres. Initial experience with the latter group has proved position. Similarly, international cooperation is likely to break down perceived language barriers. The Cochrane Collaboration is trying to involve individuals world wide with the objective of performing, maintaining, updating, and disseminating systematic reviews in all areas of health care.[16]

We found few trials reporting on methods of allocation concealment. This feature was reported in much higher frequency in obstetrics and gynaecology trials.[17] This difference may be due partly to the editorial efforts of at least one journal, which reported more information than any other journal.

Our results also provide further evidence that the overall completeness of reporting of RCTs is poor,[17-19] irrespective of the language of publication. The Standards of Reporting Trials Group (SORT)[20] and the Asilomar group[21] have made recommendations about the items, including the primary outcome measures in this study, that should be included in the reporting of RCTs. The SORT group also provided details about why these items should be included (i.e., evidence that their exclusion leads to biased estimates of intervention differences) and a format for including them. The aim is that such efforts will ultimately improve the reporting of RCTs and thus help in the conduct of systematic review.

Our results suggest that trials published in languages other than English, as well as those published in English, should be included in the conduct of all systematic reviews. We hope these results will be useful to groups developing guidelines and policy for the conduct of systematic review. If trials published in other languages are excluded from systematic reviews, this fact and a justification for the action should be given in the paper.

Acknowledgments

We thank Brenda Malanka for help throughout the study and Phil Wells for reviewing an earlier version of the paper.

References

1. Chalmers TC, Lau J. Meta-analytic stimulus for changes in clinical trials. Stat Methods Med Res 1993; 2: 161-72.
2. 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.
3. Moher D, Jadad AR, Nichol G, Penman M. Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995; 16: 62-73.
4. Science Citation Index. SCI journal citation reports: journals ranked by impact factor within category. Philadelphia: Institute for Scientific Information, 1991.
5. Dickersin K., Scherer R, Lefebvre C. Identifying relevant studies for systematic review. BMJ 1994; 309: 1286-91.
6. Andrew E, Eide H, Fuglerud P, et al. Publications on clinical trials with x-ray contrast media: differences in quality between journals and decades. Eur J Radiol 1990; 10: 92-7.
7. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials (in press).
8. Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled clinical trials. N Eng J Med 1983; 309: 1358-61.
9. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979; 86: 420-8.
10. Dupont WD, Plommer WD. Power and sample size calculations: a review and computer program. Control Clin Trials 1990; 11: 116-28.
11. Rossetti L, Marchetti I, Orzalesi N, Scorpiglione N, Liberati A. Is proper methodology associated with the use of a clinically relevant outcome measure? The case of randomized clinical trials on medical treatment of open angle glaucoma. Online J Curr Clin Trials 1993; (Doc No. 100).
12. 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.
13. Ottenbacher K, Difabio RP. Efficacy of spinal manipulation/mobilization therapy: a meta-anaylsis. Spine 1985; 10: 833-7.

14. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy I: medical. Stat Med 1989; 8: 441-54.

15. Liberati A, Himel HN, Chalmers TC. A quality assessment of randomized control trials of primary treatment of breast cancer. J Clin Oncol 1986; 4: 942-51.

16. Chalmers I. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. Ann NY Acad Sci 1993, 703: 156-65.

17. Schulz KF, Chalmers I, Grimes DA, Altman DG. Assessing the quality of randomization from reports of controlled trials published in obstetrics and Gynecology. JAMA 1994; 272: 125-8.

18. Gotzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidak anti-inflammatory drugs in rheumatoid arthritis. Control Clin Trials 1989; 10: 31-56 (correction1989; 10: 356)

19. Altman DG, Doré CJ. Randomized and baseline comparisons in clinical trials. Lancet 1990; 335: 149-53.

20. The Standards of Reporting Trials Group. A proposal for structured reporting of randomized controlled trials. JAMA 1994; 272: 1926-31.

21. Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature. Call for comments on a proposal to improve reporting of clinical trials in the biomedical literature. Ann Intern Med 1994; 121: 894-5.