

Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality study cross sectional study

Lorenzo P Moja, Elena Telaro, Roberto D'Amico, Ivan Moschetti, Laura Coe, Alessandro Liberati on behalf of the Metaquality Study Group

Abstract

Objectives To describe how the methodological quality of primary studies is assessed in systematic reviews and whether the quality assessment is taken into account in the interpretation of results.

Data sources Cochrane systematic reviews and systematic reviews in paper based journals.

Study selection 965 systematic reviews (809 Cochrane reviews and 156 paper based reviews) published between 1995 and 2002.

Data synthesis The methodological quality of primary studies was assessed in 854 of the 965 systematic reviews (88.5%). This occurred more often in Cochrane reviews than in paper based reviews (93.9% v 60.3%, $P < 0.0001$). Overall, only 496 (51.4%) used the quality assessment in the analysis and interpretation of the results or in their discussion, with no significant differences between Cochrane reviews and paper based reviews (52% v 49%, $P = 0.58$). The tools and methods used for quality assessment varied widely.

Conclusions Cochrane reviews fared better than systematic reviews published in paper based journals in terms of assessment of methodological quality of primary studies, although they both largely failed to take it into account in the interpretation of results. Methods for assessment of methodological quality by systematic reviews are still in their infancy and there is substantial room for improvement.

Introduction

Critical appraisal of the methodological quality of primary studies is an essential feature of systematic reviews.¹⁻³ Lack of adherence to a priori defined validity criteria may help explain why primary studies on the same topic provide different results.⁴ Some key issues still remain unresolved: which checklists and scales are the ideal approaches⁵ and how the results of quality assessment in a systematic review should be handled in the analysis and interpretation of results.⁶⁻¹⁰

We compared the approaches used for quality assessment of primary studies by Cochrane systematic reviews with systematic reviews published in paper based journals. We determined how quality assessment is used and whether systematic reviews consider quality assessment in their results.

Methods

We sampled systematic reviews from two databases: the Cochrane Database of Systematic Reviews and the Database of Abstracts of Reviews of Effectiveness (DARE), which selects systematic reviews published in peer reviewed journals on the basis of their adherence to a few methodological requirements.¹¹

We selected a stratified sample of Cochrane systematic reviews published in issue 1, 2002, of the Cochrane Library (see bmj.com). Our final sample represented 62.4% ($n = 809$) of the Cochrane reviews. The paper based systematic reviews were extracted from DARE, including all systematic reviews published in 2001 registered up to November 2002.

Data extraction form

We assessed the systematic reviews by using an ad hoc data extraction form. We developed this form by taking into account published reports on the quality assessment of trials included in systematic reviews.

We accepted at face value what was reported by the authors of individual studies. As a common taxonomy for quality assessment does not exist, we used a large number of descriptive quality components to capture as many of the different definitions as possible.

For each systematic review we evaluated what authors reported in the methods section of their review for quality assessment. In particular, we tried to ascertain whether authors stated they would have assessed the quality and how (scale or checklist, components studied, composite score) and in what way they planned to use the quality assessment (for example, for sensitivity analysis). See bmj.com for a summary version of the data extraction form.

We then evaluated how authors assessed quality. We recorded if trials were combined in a quantitative meta-analysis; if the quality was evaluated; if scales, checklists, and scores were used; and how the quality was formally incorporated. Assessors judged whether an attempt had been made to incorporate the quality assessment in the results, either qualitatively or quantitatively. We purposely did not make our operationalised definition of qualitative too stringent. Our definition of quantitative was more stringent and included the carrying out of a sensitivity or subgroup analysis (with quality as a stratifying factor) and use of a quality score as a weight or factor for cumulative meta-analysis or metaregression.

Data extraction

We drafted and piloted an extraction checklist. Inter-rater agreement, based on a random sample of the Cochrane reviews and paper based reviews, was high. Inter-rater reliability was moderate to perfect (percentage mean agreement 94, range 71.1-100; prevalence and bias adjusted κ statistic mean 0.80, range 0.40-1.00).

Centro Cochrane Italiano, Istituto Mario Negri, Via Eritrea 62, 20157 Milan, Italy

Lorenzo P Moja
researcher

Elena Telaro
researcher

Ivan Moschetti
researcher

Laura Coe
researcher

Alessandro Liberati
director

Università degli Studi di Modena e Reggio Emilia, Modena, Italy

Roberto D'Amico
researcher

Correspondence to: A Liberati, Università degli Studi di Modena e Reggio Emilia, Modena, Italy
alesslib@tin.it

BMJ 2005;330:1053-5



Summary version of data extraction form is on bmj.com



This is the abridged version of an article that was posted on bmj.com on 7 April 2005: <http://bmj.com/cgi/doi/10.1136/bmj.38414.515938.8F>

Table 1 Distribution of three main quality assessment related variables investigated in study according to Cochrane systematic reviews and systematic reviews published in paper based journals. Values are numbers (percentages) unless stated otherwise

Variable	Cochrane systematic reviews (n=809)	Paper based systematic reviews (n=156)	Difference between proportions (95% CI)	P value (χ^2 test)
Authors would carry out quality assessment	758 (93.7)	99 (63.5)	30.2 (22.5 to 38.0)	<0.0001
Quality assessment carried out	760 (93.9)	94 (60.3)	33.6 (25.8 to 41.5)	<0.0001
Quality assessment linked to results	419 (51.8)	77 (49.4)	2.4 (-6.1 to 11.0)	0.5777

Twelve pairs of investigators independently extracted the data. Disagreements were resolved through discussion and, when necessary, centrally reviewed.

Results

We analysed 965 systematic reviews: 809 Cochrane reviews and 156 reviews published in paper based journals (see bmj.com). Quality assessment was assessed in 854 (88.5%) of the reviews and was more often carried out in Cochrane reviews than in paper based reviews for both formal and informal comparisons (tables 1 and 2).

The quality components most frequently assessed were, in decreasing order, allocation concealment, blinding, and losses to follow-up. The difference between Cochrane reviews and paper based reviews for these and intention to treat analysis significantly favoured Cochrane reviews.

The most commonly used quality scale was the Jadad scale (n=113; 12.0%). Cochrane reviews used the scale less often than paper based reviews. In 65.0% (n=526) of Cochrane reviews and 48.1% (n=75) of paper based reviews, authors carried out the quality assessment using single components rather than a formal scale.

No significant differences emerged when Cochrane reviews and paper based reviews were analysed separately by type of intervention assessed—for example, drug compared with non-drug interventions.

We found that 496 systematic reviews (51.4%) linked quality to the interpretation of results, with no difference in the proportions of Cochrane reviews and paper based reviews (table 1). This also held true for the subgroup analysis of drug compared with non-drug interventions.

The authors of Cochrane reviews were more likely than those of paper based reviews to state that they would assess quality yet did not always do so (table 1). About 5% of systematic reviews in each group carried

out quality assessment despite not being explicitly stated in the methods. Finally, only 328 (33.9%) of the systematic reviews formally specified how they planned to use the quality assessment in the methods: 36.0% (n=291) of Cochrane reviews and 23.7% (n=37) of paper based reviews (P=0.79).

Discussion

More than 50% of systematic reviews (both Cochrane reviews and reviews based in paper articles) did not specify in the methods whether and how they would use quality assessment in the analysis and interpretation of results. Cochrane reviews fared better than paper based reviews in carrying out quality assessment but were equally unsuccessful in taking it into account.

During the past 15 years research has concentrated on two main issues: which components of the quality assessment (for example, allocation concealment) are predictive of valid results and what tool (scales or checklists) best assesses quality. In 2003 Egger et al found that allocation concealment and double blinding were strongly related to treatment effects.^{4 12 13} Despite the dozens of quality scales and checklists that have been proposed,^{5 7 14} the answer is still unclear and many doubt that a generic quality assessment tool valid in all cases can ever be found. In our study the most frequently used tool was the Jadad scale, a tool that has been criticised for its low sensitivity and which was developed before the importance of concealment was established.¹⁵ Moreover, less attention has been paid to exploring how quality can be used in the interpretation of the results of systematic reviews.^{8 9 16}

As Cochrane reviews are preceded by a published protocol which must state that quality assessment of primary studies is to be done, it is not surprising that authors state they will carry out quality assessment and do so. Yet when it comes to incorporating quality into the results, Cochrane reviews fared no better than their paper based counterparts.

Table 2 Summary of approaches to quality assessment and formal quantitative analyses related to quality assessment used in Cochrane systematic reviews and systematic reviews published in paper journals. Values are numbers (percentages) unless stated otherwise

Approaches used to incorporate quality assessment*	Cochrane systematic reviews (n=809)	Paper based systematic reviews (n=156)	Cochrane and paper based reviews	Difference between proportions (95% CI)	P value (χ^2 test)
Informal:					
General comment	732 (90.5)	81 (51.9)	813 (95)	38.6 (30.5 to 46.7)	<0.0001
Formal:					
Exclusion criteria	102 (12.6)	9 (5.8)	111 (12)	6.8 (2.5 to 11.2)	0.0142
Exploration of heterogeneity	63 (7.8)	11 (7.1)	74 (8)	0.7 (-3.7 to 5.2)	0.7517
Subgroup analysis	29 (3.6)	8 (5.1)	37 (4)	-1.5 (-5.2 to 2.1)	0.3580
Sensitivity analysis	89 (11.0)	9 (5.8)	98 (10)	5.2 (1.0 to 9.5)	0.0476
Weighting of estimates	1 (0.1)	0	—	0.1 (-0.1 to 0.4)	0.6604
Cumulative meta-analysis	1 (0.1)	0	—	0.1 (-0.1 to 0.4)	0.6604

*More than one answer possible.

These findings may have several explanations. That Cochrane reviews provide more details may be due to the absence of limitations on space in electronic publications; however, most of the medical journals now publish a web version of their papers. We found, however, that none of the paper based reviews was supplemented with an electronic appendix of quality assessment. It is also possible that authors are unaware of or that editors are not interested in publishing extensive electronic versions. Moreover, most authors of paper based articles may be aware of space limitations imposed by journals and omit details of quality assessment.

Limitations of the study

The Cochrane reviews were published between 1995 and 2002, whereas the paper based articles were first published in 2001. However, between 1995 and 2000 no major methodological advances or new consensus emerged in the literature on systematic reviews. Another limitation is whether the DARE database was an appropriate source from which to sample paper based articles as it may have led to the selection of a control group with better than average quality. Any selection bias would move our estimate towards the null effect.

A third possible limitation is that incomplete reporting might have influenced our assessment. Evidence, however, suggests that what is reported about important aspects of the conduct of a study typically do reflect what is done.^{17 18}

Finally, we assessed quality assessment using a checklist that had been developed ad hoc. Although the lack of validation may be criticised, we believe that the items have good face validity, and the inter-rater reliability was acceptably high. We did have trouble with the classification of quality items, tools, and approaches, as there are innumerable ways to define study quality.^{7 12} It is possible that we recorded quality data with slightly different meanings from those intended by the authors of the studies.

Within the Cochrane Collaboration there is room for improvement in the approaches to quality assessment, and the Cochrane handbook should provide clearer guidelines on how to do it. Less clear is how to improve quality assessment in paper based systematic reviews. We believe that more methodological research is needed to understand how best to assess and to incorporate the methodological quality of primary studies into the results of systematic reviews.

We thank Sabrina Bidoli and Sheila Iacono for their help with data entry and organisational activities; Simone Mangano and Luca Clivio for their support in developing the database; the other members of the Cochrane Collaboration who commented on the study and our interpretation of it when it was presented at the International X and XI Cochrane Colloquia (Stavanger, Norway, 2002 and Barcelona, Spain, 2003); and David Moher, Peter Tugwell, Keith O'Rourke, and Dave Sackett for commenting on an earlier version of this paper. See [bmj.com](http://www.bmj.com) for members of the Metaquality Study Group.

Contributors: See [bmj.com](http://www.bmj.com)

Funding: Ministero della Istruzione, dell'Università e della Ricerca, Rome, Italy (protocol grant No 2002061749 COFIN 2002).

Competing interests: All members of the research team are members of the Cochrane Collaboration. ET, IM, RD'A, LC, and AL are authors of Cochrane reviews published in the Cochrane Library. The views and opinions expressed are those of the authors and do not necessarily reflect those of the Cochrane Collaboration.

What is already known on the topic

Appraisal of the methodological quality of primary studies is essential in systematic reviews

No consensus exists on the ideal checklist and scale for assessing methodological quality

The Cochrane Collaboration encourages a simple approach to quality assessment based on individual components such as allocation concealment

What this study adds

Approaches to quality assessment of primary studies by systematic reviews are heterogeneous and reflect a lack of consensus on best practice

Cochrane reviews assess methodological quality more often than paper based reviews

Both types of review failed to link the quality assessment to the interpretation of results in almost half of cases

Ethical approval: Not required.

- 1 Juni P, Altman DG, Matthias E. Assessing the quality of randomised controlled trials. In: Egger M, Davey Smith G, Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*, 2nd ed. London: BMJ Books, 2001.
- 2 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. Quality of reporting of meta-analyses. *Lancet* 1999;354:1896-900.
- 3 Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *CMAJ* 1988;138:697-703.
- 4 Egger M, Juni P, Bartlett C, Hohenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;7:1-76.
- 5 Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054-60.
- 6 Alderson P, Green S, Higgins JPT, eds. Assessment of study quality. Cochrane Reviewers' Handbook 4.2.2 [updated March 2004]; Section 6. In: The Cochrane Library, Issue 1. Chichester: Wiley, 2004. www.cochrane.org/resources/handbook/hbook.htm (accessed 1 Apr 2005).
- 7 Verhagen AP, de Vet HC, de Bie RA, Boers M, van den Brandt PA. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol* 2001;54:651-4.
- 8 Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2001;2:463-71.
- 9 Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992;45:255-65.
- 10 Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials. Current issues and future directions. *Int J Technol Assess Health Care* 1996;12:195-208.
- 11 Petticrew M, Song F, Wilson P, Wright K. Quality-assessed reviews of health care interventions and the database of abstracts of reviews of effectiveness (DARE). NHS CRD review, dissemination, and information teams. *Int J Technol Assess Health Care* 1999;15:671-8.
- 12 Balk EM, Bonis PA, Moskowitz H, Schmid CH, Ioannidis JP, Wang C, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973-82.
- 13 Juni P, Egger M. Allocation concealment in clinical trials. *JAMA* 2002;287:2408-9.
- 14 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.
- 15 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
- 16 Trichter D. Modelling study quality in meta-analysis. *Stat Med* 1999;18:2135-45.
- 17 Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. *Lancet* 1990;335:149-53.
- 18 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.

(Accepted 2 March 2005)

doi 10.1136/bmj.38414.515938.8F