Scope for improvement in the quality of reporting of systematic reviews. From the Cochrane Musculoskeletal Group.

Beverley Shea, Lex M Bouter, Jeremy M Grimshaw, Daniel Francis, Zulma Ortiz, George A Wells, Peter S Tugwell and Maarten Boers

J Rheumatol 2006;33;9-15
http://www.jrheum.org/content/33/1/9

1. Sign up for our monthly e-table of contents
   http://www.jrheum.org/cgi/alerts/etoc

2. Information on Subscriptions
   http://jrheum.com/subscribe.html

3. Have us contact your library about access options
   Refer_your_library@jrheum.com

4. Information on permissions/orders of reprints
   http://jrheum.com/reprints.html

The Journal of Rheumatology is a monthly international serial edited by Duncan A. Gordon featuring research articles on clinical subjects from scientists working in rheumatology and related fields.
Scope for Improvement in the Quality of Reporting of Systematic Reviews.
From the Cochrane Musculoskeletal Group

BEVERLEY SHEA, LEX M. BOUTER, JEREMY M. GRIMSHAW, DANIEL FRANCIS, ZULMA ORTIZ, GEORGE A. WELLS, PETER S. TUGWELL, and MAARTEN BOERS

ABSTRACT. Objective. To assess the quality of reporting in Cochrane musculoskeletal systematic reviews (excluding back and injury reviews).

Methods. This study assessed all the Cochrane Musculoskeletal Group’s systematic reviews from Issue 4, 2002, of the Cochrane Library Database of Systematic Reviews. Two reviewers independently extracted data and assessed quality. Two assessment tools were used, including an 18 item checklist and flow chart developed by the Quality of Reporting of Meta-analysis (QUOROM) consensus group, and a 10 item scale, the Oxman-Guyatt Overview Quality Assessment Questionnaire (OQAQ). One question on the latter scale (item 10) scores overall quality on a 7 point scale, with high scores indicating superior quality. Data were analyzed using univariate approaches.

Results. The 57 systematic reviews assessed were found to have good overall quality, with scores on individual items revealing only minor flaws. Documenting the flow of included and excluded studies and summarizing the results are 2 areas needing improvement in reporting. According to the Oxman-Guyatt scale the overall scientific quality of the Cochrane musculoskeletal reviews was good [mean 5.02 (95% CI 3.71–6.32)].

Conclusion. Our study found that the reporting quality of Cochrane musculoskeletal systematic reviews was generally good, although there was room for improvement. For example, it might be feasible to develop specific guidelines for reporting protocols. Certainly more work is needed in reporting search results, documentation of the flow of studies, identification of the type of studies, and summarization of the key findings. (J Rheumatol 2006;33:9-15; First Release: Nov 1, 2005)

Key Indexing Terms:
MUSCULOSKELETAL
QUALITY ASSESSMENT
COCHRANE
SYSTEMATIC REVIEWS
REPORTING
musculoskeletal injuries. We assessed the quality of reviews conducted by the CMSG.

Growing recognition of the key role of reviews in synthesizing and disseminating research results has prompted careful scrutiny of the validity of reviews. In the 1970s and early 1980s, psychologists and social scientists drew attention to the systematic steps needed to minimize bias and random errors in literature reviews. In the late 1980s attention began to focus on the poor scientific quality of healthcare review articles. An appreciation of the quality of a systematic review is essential to assessment of whether its recommendation of the use or avoidance of an intervention should be followed. Two major areas are assessed in determining the quality of a systematic review. The first is its methodological quality, which is an assessment of how well the systematic review was conducted (literature search, pooling of data, etc.). The second is its quality of reporting, which is an assessment of how well its systematic reviewers have reported their methodology and findings. Separate tools were used to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Our objective was to review both the methodological and reporting quality of all published Cochrane musculoskeletal reviews. This review should serve as a baseline, enabling the CMSG to measure improvement in both the methodological and the reporting quality of its reviews over time.

MATERIALS AND METHODS
Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

An overview of the quality of a systematic review is essential to assessment of whether its recommendation of the use or avoidance of an intervention should be followed. Two major areas are assessed in determining the quality of a systematic review. The first is its methodological quality, which is an assessment of how well the systematic review was conducted (literature search, pooling of data, etc.). The second is its quality of reporting, which is an assessment of how well its systematic reviewers have reported their methodology and findings. Separate tools were used to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Our objective was to review both the methodological and reporting quality of all published Cochrane musculoskeletal reviews. This review should serve as a baseline, enabling the CMSG to measure improvement in both the methodological and the reporting quality of its reviews over time.

MATERIALS AND METHODS
Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Our objective was to review both the methodological and reporting quality of all published Cochrane musculoskeletal reviews. This review should serve as a baseline, enabling the CMSG to measure improvement in both the methodological and the reporting quality of its reviews over time.

MATERIALS AND METHODS
Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Our objective was to review both the methodological and reporting quality of all published Cochrane musculoskeletal reviews. This review should serve as a baseline, enabling the CMSG to measure improvement in both the methodological and the reporting quality of its reviews over time.

Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.

Choosing the assessment instruments. A review of published scales and checklists was performed to inventory the instruments available for assessing the quality of systematic reviews. Each item of a scale is scored numerically, and individual numerical scores are combined to generate an overall quality score. To be considered a scale, an instrument should be able to assess each quality area to obtain a comprehensive quality assessment. Assessors were sensitive to the fact that although methodological quality and reporting quality are intrinsically linked, a review may be strong in either area and weak in the other. It was also recognized that poor reporting makes it difficult to assess the methodological quality of a review.
All CMSG systematic reviews reported the criteria used for
reported whether study selection bias was avoided (item 4). 95% of the reviews (item 7), while 97% of CMSG reviews
reported whether the search strategy for the evidence was reasonably comprehensive (item 2) and 53%
reported whether study selection bias was avoided (item 4). All CMSG systematic reviews reported the criteria used for
deciding which studies to include in the overview (item 3) and combined the findings of the relevant studies appropriately relative to the primary question addressed (item 8). The methods used for combining studies were reported in 95% of the reviews (item 7), while 97% of CMSG reviews reported the criteria used for assessing the validity of included studies (item 5) and drew conclusions that were support-

similar quality (Table 1). Of the 10 items making up the scale, the Cochrane musculoskeletal systematic reviews scored poorly on items 2 and 4: 63% of CMSG systematic reviews reported whether the search strategy for the evidence was reasonably comprehensive (item 2) and 53% reported whether study selection bias was avoided (item 4). All CMSG systematic reviews reported the criteria used for

Table 1. Scores for Cochrane musculoskeletal group (CMSG) systematic reviews from the Overview Quality Assessment Questionnaire (OQAQ).

<table>
<thead>
<tr>
<th>Questions</th>
<th>Yes, n (%)</th>
<th>Partially or can’t tell, n (%)</th>
<th>No, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1 1. Were the search methods used to find evidence reported?</td>
<td>50 (88)</td>
<td>3 (5)</td>
<td>4 (7)</td>
</tr>
<tr>
<td>Item 2 2. Was the search strategy for evidence reasonably comprehensive?</td>
<td>36 (63)</td>
<td>3 (5)</td>
<td>18 (32)</td>
</tr>
<tr>
<td>Item 3 3. Were the criteria used for deciding which studies to include in the overview reported?</td>
<td>57 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Item 4 4. Was bias in the selection of studies avoided?</td>
<td>30 (53)</td>
<td>23 (40)</td>
<td>4 (7)</td>
</tr>
<tr>
<td>Item 5 5. Were the criteria used for assessing the validity of the included studies reported?</td>
<td>55 (97)</td>
<td>2 (3)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Item 6 6. Was the validity of all the studies referred to in the text assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited)?</td>
<td>41 (72)</td>
<td>4 (7.0)</td>
<td>12 (21)</td>
</tr>
<tr>
<td>Item 7 7. Were the methods used to combine the findings of the relevant studies (to reach a conclusion) reported?</td>
<td>54 (95)</td>
<td>1 (2)</td>
<td>2 (3.5)</td>
</tr>
<tr>
<td>Item 8 8. Were the findings of the relevant studies combined appropriately relative to the primary question the overview addressed?</td>
<td>57 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Item 9 9. Were the conclusions made by the author(s) supported by the data and/or analysis reported in the overview?</td>
<td>55 (97)</td>
<td>1 (2)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Item 10 10. How would you rate the scientific quality of this overview?</td>
<td></td>
<td></td>
<td>5.02, 95% CI 3.71, 6.32</td>
</tr>
</tbody>
</table>

Table 2. Scores for Cochrane musculoskeletal group (CMSG) systematic reviews from Quality of Reporting of Meta-analysis (QUOROM).

<table>
<thead>
<tr>
<th>Questions</th>
<th>Yes, n (%)</th>
<th>No, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1 1. Does the title identify the report as a metaanalysis (or systematic review) of randomized trials?</td>
<td>54 (95)</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Item 2 2. Is the abstract in a structured format?</td>
<td>56 (98)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Item 3 3. Do the objectives describe the clinical question explicitly?</td>
<td>56 (98)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Item 4 4. Are the data bases (i.e., list) and information sources described?</td>
<td>55 (96.5)</td>
<td>2 (3.5)</td>
</tr>
<tr>
<td>Item 5 5. Are the selection criteria (population, intervention, outcome, and study design), methods for validity assessment, data abstraction, study characteristics, and quantitative data synthesis described in sufficient detail to permit replication?</td>
<td>47 (82)</td>
<td>10 (18)</td>
</tr>
<tr>
<td>Item 6 6. Is there a description of the main results?</td>
<td>49 (86)</td>
<td>8 (14)</td>
</tr>
<tr>
<td>Item 7 7. Are the conclusions presented?</td>
<td>57 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Item 8 8. Is the clinical, biologic rationale for the intervention and rationale for the review provided?</td>
<td>54 (95)</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Item 9 9. Were the information sources in detail (e.g., databases, registers, personal files, expert informants, agencies, hand-searching), and any restrictions (years considered, publication status, language of publication) provided?</td>
<td>52 (91)</td>
<td>5 (9)</td>
</tr>
<tr>
<td>Item 10 10. Were the inclusion and exclusion criteria (defining population, intervention principal outcomes, and study design) presented?</td>
<td>57 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Item 11 11. Were the criteria and process used (e.g., masked conditions, quality assessment and their findings) in the validity assessment reported?</td>
<td>57 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Item 12 12. Was the process provided (e.g., completed independently, in duplicate)?</td>
<td>50 (88)</td>
<td>7 (12)</td>
</tr>
<tr>
<td>Item 13 13. Were the type of study design, participants’ characteristics, details of intervention, outcome definitions, and how clinical heterogeneity was assessed reported?</td>
<td>56 (98)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Item 14 14. Were the principal measures of effect (e.g., relative risk), method of combining results (statistical testing and confidence intervals), handling of missing data, how statistical heterogeneity was assessed, a rationale for any a priori sensitivity and subgroup analyses, and any assessment of publication bias reported?</td>
<td>53 (93)</td>
<td>4 (7)</td>
</tr>
<tr>
<td>Item 15 15. Was a metaanalysis profile summarizing trial flow provided?</td>
<td>0 (0)</td>
<td>57 (100)</td>
</tr>
<tr>
<td>Item 16 16. Were the descriptive data for each trial presented?</td>
<td>57 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Item 17 17. Was the agreement on the selection and validity assessment reported?</td>
<td>56 (98)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Item 18 18. Were a summarization of the key findings, discussion of clinical inferences based on internal and external validity, interpretation of the results in light of the totality of available evidence, description of potential biases in the review process (e.g., publication bias), and suggestion of a future research agenda presented?</td>
<td>47 (82)</td>
<td>10 (18)</td>
</tr>
</tbody>
</table>
ed by the data and/or analysis reported in the overview (item 9). Eighty-eight percent of the reviews reported the search methods used to find evidence (item 1). However, only 72% reported whether the validity of all the studies referred to in the text was assessed using appropriate criteria (either in selecting studies for inclusion or in analyzing the studies that are cited) (item 6).

Scores on individual QUOROM items ranged from 5.0% (item 1) to 100% (items 7, 10, 11, 16) (Table 2). Only 5% of CMSG reviews identified the review as a metaanalysis or systematic review of randomized trials in their title (item 1). Almost all CMSG reviews had an abstract with a structured format (item 2) and included objectives (item 3) and data sources (item 4) in the abstract. Items on which they were less likely to report adequately were results (item 6) and selection criteria (item 5) (i.e., population, intervention, outcome, study design, methods for validity assessment, data abstraction, study characteristics, and quantitative data synthesis).

Almost 90% of the reviews described their method of data abstraction (item 12), while no review provided a flow chart for the included and excluded studies (item 15).

More than half the CMSG reviews received a rating of “adequate” on 50% or more of the 10 OQAQ quality items. On the overall quality item (range 0–7) (item 10), the CMSG reviews scored relatively well, with only minor flaws identified [mean 5.02 (95% CI 3.71–6.32)] (Table 1). Of the 18 QUOROM items, the CMSG systematic reviews scored more than 50% on all but 2 of the items (items 1 and 15).

One item (18) was noted as being more difficult to assess and certainly needs further exploration [i.e., the summarization of the key findings, discussion of clinical inferences based on internal and external validity, interpretation of the results in light of the totality of available evidence, description of potential biases in the review process (e.g., publication bias), and suggestion of a future research agenda presented].

Documenting the flow of included and excluded studies and summarizing the results are 2 areas needing improvement in reporting.

DISCUSSION

Assessments made using both quality instruments indicated that the quality of Cochrane musculoskeletal systematic reviews was good, although minor flaws were observed. This is important to users of CMSG reviews, as it provides assurance that their results are relatively reliable. Although their methodological quality and quality of reporting were found to be fair to good, there is room for improvement. For example, it might be feasible to develop specific guidelines for reporting protocols and improve on reporting search results, and documentation of the flow of studies.

The quality of systematic reviews requires examination in order to substantiate the claim that they are the best evidence available to clinicians, health policymakers, and consumers. The use of assessment tools to structure peer review systems can encourage quality improvement in systematic reviews. The Cochrane Collaboration has begun to achieve this objective through continual peer review of protocols, reviews, and updated reviews from the analytical process through to the report. The use of evidence-based criteria such as the QUOROM statement can contribute to the improvement of reporting quality over time by establishing consistent guidelines for the conduct of systematic reviews. At least 2 studies have addressed improvement or lack of improvement over time. A review of 86 English-language metaanalyses assessed every report on 14 items from 6 content areas believed to be critical in the conduct and reporting of metaanalyses. These items included study design, comparability, control of bias, statistical analysis, sensitivity analysis, and problems of applicability. They found that only 24 of the 86 (28%) metaanalyses addressed all 6 content areas75. This survey was updated in 1992 with little change in the results76. A similar study by the authors of this paper showed that the quality of systematic reviews does improve over time, but that the differences on specific items remain variable77. A comparison of Cochrane versus paper-based reviews revealed similar results78.

Inadequate reporting79,80–82 is a significant impediment to the assessment of the quality of systematic reviews! Essential criteria may be met in a given study without being adequately reported in a review. In such a case, a study of high quality may appear to be poor in a review. It may be inaccurate to assume that items not included in a systematic review were missing in the study it reviews, but this is what users of systematic reviews are likely to do. Assessors of a systematic review may invest the time and effort required to obtain additional data directly from the investigators who conducted the systematic review, but ordinary readers cannot reasonably be expected to do so.

The ongoing use of the quality of reporting checklists6,80–82 is to be encouraged, because it will facilitate assessment of study validity by ensuring a high level of congruence between the quality of individual studies and their depiction in systematic reviews. Such an improvement in the quality of reporting of reviewed studies will serve to make systematic reviews more accurate, reliable, persuasive, and useful to those who depend on them.

We tried to address the issue of potential conflict of interest by inviting someone from outside the Musculoskeletal group who would be willing to work with the team on a voluntary basis to carry out this study, a common practice for Cochrane work, especially methods work. One of the main reviewers of the studies is from Brisbane, Australia, and had worked with the Respiratory Airways group for about 3 years. He is very well informed of the format of a Cochrane review. The authors felt that because he was not involved in any of the CMSG reviews he would be considered a nonbiased reviewer. We believe this will negate any potential biases.
Although the 2 instruments used in this study proved useful, the challenges that had to be overcome in applying them clearly demonstrated the need for better measurement instruments. Both instruments proved to be difficult to apply; the main problem encountered was a lack of published guidance on their application. It was only after 3 rounds of pilot testing and resolution of several questions that arose concerning the application of the OQAQ assessment tool that the assessors were ready to apply it. Moreover, while using this scale, reviewers continued to encounter difficulties with its application. Although the QUOROM checklist was rather long and time-consuming to apply, it was accompanied by more detailed directions regarding its use.

One additional item that was not addressed adequately because the measurement tool does not address this well is the inferences based on the results of the systematic reviews. More work is needed in this area.

The statistical analysis revealed that the reliability was poor to fair. There may be several explanations for this. First, the relative magnitude of the kappa value (i.e., the proportion of agreement beyond that expected by chance alone) is difficult to interpret. In the pilot test, the kappa values were categorized and labeled as suggested by Fleiss, but this classification is purely arbitrary. Kappa coefficient is a popular measure for chance-corrected nominal scale agreement between 2 raters. Future methodological work must include alternative options for calculating agreement among raters, such as exploring Bayesian inferences.

This study found that the overall quality of reports of Cochrane musculoskeletal systematic reviews was generally good, although there was room for improvement. Areas in particular that need special attention include the title and protocol, documentation of the flow of the studies, and inferences made by the conclusions. A recent study reported that the methods for assessment of methodological quality of systematic reviews are still in their infancy and there is substantial room for improvement.

ACKNOWLEDGMENT

The authors express their appreciation to David Moher for his helpful suggestions and to Ashley Porter and Thelma Hasson for their comments on the manuscript.

REFERENCES

11. National Health Service (NHS) Centre for Reviews and Dissemination (CRD). York, UK: University of York; Y010 5DD.


