Method Guidelines for Cochrane Musculoskeletal Group Systematic Reviews

LARA MAXWELL, NANCY SANTESSO, PETER S. TUGWELL, GEORGE A. WELLS, MARIA JUDD, RACHELLE BUCHBINDER, and the Editorial Board of the Cochrane Musculoskeletal Group

ABSTRACT. The Cochrane Musculoskeletal Group (CMSG), one of 50 groups of the not-for-profit international Cochrane Collaboration, prepares, maintains, and disseminates systematic reviews of treatments for musculoskeletal diseases. To enhance the quality and usability of systematic reviews, the CMSG has developed tailored methodological guidelines for authors of CMSG systematic reviews. Recommendations specific to musculoskeletal disorders are provided for various aspects of undertaking a systematic review, including literature searching, inclusion criteria, quality assessment, grading of evidence, data collection, and data analysis. These guidelines will help researchers design, conduct, and report results of systematic reviews of trials in the following fields of musculoskeletal health: gout, osteoarthritis, osteoporosis, pediatric rheumatology, rheumatoid arthritis, soft tissue rheumatism, spondyloarthropathy, systemic lupus erythematosus, systemic sclerosis, and vasculitis. Systematic reviews need to be conducted according to high methodological standards. These recommendations on developing and performing a systematic review will help improve consistency among CMSG reviews. (J Rheumatol 2006;33:2304–11)

Key Indexing Terms:
EVIDENCE-BASED MEDICINE REVIEW LITERATURE METAANALYSIS ARTHRITIS

The systematic review has emerged as an essential technology for managing the vast amounts of information generated on health treatments. Compared to a narrative literature review, the systematic review employs "scientific strategies that limit bias in the assembly, critical appraisal and synthesis of all relevant studies on a specific topic"[1].

The aim of the Cochrane Collaboration is to help physicians, patients, and policymakers make well informed decisions on healthcare treatments by preparing, maintaining, and disseminating high quality systematic reviews[2]. The Cochrane Musculoskeletal Group (CMSG), one of 50 international groups in the Cochrane Collaboration, synthesizes the results of high quality studies to determine the effectiveness and safety of interventions for the prevention, treatment, and management of musculoskeletal diseases including various forms of arthritis.

Registered in 1993 and currently based in Ottawa, Canada, with a satellite editorial office in Melbourne, Australia, the CMSG has become one of the largest Cochrane review groups. It has over 200 active researchers, healthcare professionals, and consumer representatives from 26 countries (including 8 developing countries) that conduct and disseminate research on musculoskeletal conditions. The satellite office supports Australian-based authors, reviews on soft tissue disorders, and works to disseminate CMSG reviews to Australian clinicians, consumers and policymakers. The scope of the CMSG includes: gout, lupus erythematosus, osteoarthritis (OA), osteoporosis, pediatric rheumatology, rheumatoid arthritis (RA), soft tissue rheumatism, spondyloarthropathy, systemic sclerosis, and vasculitis[3]. There are 93 completed CMSG reviews and 65 protocols in the Issue 3, 2006, edition of the Cochrane Database of Systematic Reviews.

It is known that literature reviews are susceptible to many types of bias and that one of the main ways of reducing this is to use a systematic approach that lays out the information on the major potential biases[4]. Within the scientific literature, there have been cases of literature reviews on the same topic arriving at different conclusions. One example is the 2 reviews in 2002 on the efficacy and safety of COX-2-selective nonsteroidal antiinflammatory drugs. The review by Deeks, *et al*[5] concluded that a COX-2 was effective and improved gastrointestinal safety, while the review by Wright[6] concluded the opposite. These inconsistent findings can result in confusion

From the Cochrane Musculoskeletal Group, Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada.

Supported by the Canadian Institutes of Health Research (CIHR) Knowledge Translation Branch, Canadian Agency for Drugs and Technologies in Health (CADTH), CIHR Institutes of Health Services and Policy Research, Musculoskeletal Health and Arthritis, Gender and Health, Human Development, Child and Youth Health, Institute of Nutrition, Metabolism and Diabetes, Institute of Infection and Immunity; NHMRC Practitioner Fellowship.

L. Maxwell, BSc, CMSG Review Group Coordinator; N. Santesso, RD, MLIS, Knowledge Translation Specialist, University of Ottawa; P.S. Tugwell, MD, MSc, Professor, Department of Medicine, University of Ottawa, Canada Research Chair in Health Equity; G.A. Wells, MSc, PhD, Professor, Department of Epidemiology and Community Medicine, University of Ottawa; M. Judd, MSc, Senior Program Officer, Canadian Health Services Research Foundation; R. Buchbinder, MSc, Associate Professor, Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Victoria, Australia.

Address reprints requests to L. Maxwell, Institute of Population Health, University of Ottawa, 1 Stewart Street, Ottawa, Ontario K1N 6N5, Canada. E-mail: lmaxwell@uottawa.ca

Accepted for publication May 25, 2006.

and reduced confidence in literature reviews. The rigorous, systematic approach used by Cochrane reviews aims to provide a definitive statement on the effects of healthcare treatments.

The Cochrane Handbook for Systematic Reviews of Interventions[7] describes in detail the process of creating Cochrane systematic reviews. Although the Handbook is comprehensive, it is quite large (265 pages, including appendices) and it does not include the tailored recommendations for musculoskeletal reviews. Other groups[8,9] have found it useful to provide tailored guidance to those carrying out reviews. Tailored guidance should improve consistency among authors and thus facilitate comparison across reviews. Our report will provide direction for reviewers of musculoskeletal interventions within the CMSG scope on literature searching, inclusion criteria, quality assessment, grading of evidence, data collection, and data analysis. Conditions specific to the back and musculoskeletal injuries are addressed by the Cochrane Back Group (www.cochrane.iwh.on.ca) and the Bone, Joint, and Muscle Trauma Group (http://cmsig.tees.ac.uk).

Further to summarizing the best available evidence for therapies, the CMSG actively works to promote the dissemination and integration of quality evidence into health and healthcare decisions; we refer to this as "knowledge translation." In the accompanying report[10] there are recommendations to translate results into "usable" and "useful" formats.

CMSG GUIDELINES

Literature search

A crucial step in the preparation of a systematic review is to develop the method by which all relevant trials on the specific topic will be identified. First, however, the research question needs to be clearly formulated using the "PICO" framework, i.e., a clinical or research question that takes into account the Patient/Population, Intervention, Comparison, Outcomes (further defined under Inclusion Criteria, below).

The search strategy is defined *a priori* and needs to be documented and contain sufficient detail so that the search can be duplicated and retrieve comparable results[11]. Ideally, the complete Medline strategy should be available as well as any other modified search strategies developed for other databases. It is also important to identify who conducted the search.

The search strategy usually consists of 2 parts: the content search using the terms identified by PICO, and the study design search using filters. The search for content is developed by using appropriate controlled vocabularies (e.g., MeSH) as well as text. To search for study design, such as randomized controlled trials, filters should be used. At a minimum, one of the 2 search strategies developed to retrieve randomized controlled trials by members of the Cochrane Collaboration, Dickersin, *et al*[12] (1994) and Robinson, *et al*[13] (2002), are recommended for use. High quality literature searching and information retrieval requires expertise in this area. To help reviewers develop their search methods, the CMSG has the expertise of 2 trained librarians specializing in library and information science.

It is recommended that, at a minimum, the following 3 databases are searched: Medline, Embase, and the Cochrane Central Register of Controlled Trials (Central). To ensure that as many relevant studies as possible are identified, review authors are also encouraged to conduct additional searches by one or more of the following strategies: hand searching those high-yield journals and conference proceedings that have not already been hand searched on behalf of the Cochrane Collaboration; reviewing reference lists of all papers and relevant reviews; contacting authors of relevant papers and authors of other reviews or experts in the subject area; and searching citation databases (e.g., Web of Science or Science Citation Index) and other relevant bibliographic databases, such as Cinahl for nonpharmacological interventions. It is recommended that trials in languages other than English are not excluded from a review[14]. Some topics, such as certain alternative therapies for arthritis, may have a significant number of trials published in another language and the CMSG can assist with translation when necessary.

Two people should independently be involved in screening the titles and abstracts from the results of the searches for the selection of trials meeting the predefined inclusion criteria. The inclusion criteria should be pilot tested on a sample of articles. If consensus is not reached, a third reviewer is assigned to decide. Ideally, the level of interrater reliability using the Kappa statistic or intraclass correlation for discrete and continuous data, respectively, should be documented. The full text of those articles that meet the inclusion criteria should then be obtained.

Inclusion criteria

The minimum criteria for trial inclusion in the systematic review should be defined *a priori* and address the following items using the PICO framework and study design:

Population. Participants of trials should be defined by acceptable diagnostic criteria where possible, such as the American College of Rheumatology (ACR) criteria for fibromyalgia, gout, lupus, OA, and RA. These are listed together in the Arthritis Foundation *Primer on the Rheumatic Diseases*[15]. Specific exclusions, such as age, sex, and condition, must be detailed.

Example of definition of type of participants[16]:
Patients at least 16 years of age meeting the ACR 1987 revised criteria for RA[17]. These patients must have evidence of active disease as demonstrated by at least 2 of the following:

1. Tender joint count
2. Swollen joint count
3. Duration of early morning stiffness > 30 minutes
4. Acute phase reactants such as Westergren erythrocyte sedimentation rate (ESR) or C-reactive protein (CRP)

Intervention. The intervention must be explicitly described. If

applicable, the route of administration, dose, timing, duration of treatment, and concomitant treatments should be outlined.

Example of definition of type of intervention[18]:
Infliximab and methotrexate to methotrexate alone or comparing infliximab alone to placebo were eligible for inclusion. Patients could also be taking other disease modifying antirheumatic drugs (DMARD) or corticosteroids provided they were taking stable doses and were randomly allocated to treatment with infliximab or to treatment without infliximab. Doses of infliximab eligible for inclusion include 1, 3, 5, and 10 mg/kg with a minimum trial duration of 6 months.

Comparison. The comparison intervention should be explicitly defined (e.g., placebo, another treatment).

Example of definition of type of comparison[16]:
Treatment with adalimumab 20, 40, or 80 mg subcutaneously every to every other week, alone or in combination with DMARD versus placebo or DMARD were eligible for inclusion.

Outcomes. Where available, standardized, validated, established outcome measures appropriate for the disease condition and the medical or surgical intervention should be used. For example, in RA, the ACR core set of disease activity measures for RA clinical trials endorsed by the World Health Organization and the Outcome Measures in Rheumatology (OMERACT) group are recommended. OMERACT and their associated groups have established preliminary outcomes for OA, osteoporosis, lupus, and ankylosing spondylitis, and are actively developing consensus on outcomes for fibromyalgia, systemic sclerosis, and vasculitis[19-33]. With regional disorders such as shoulder and elbow disorders, a set of standardized measures does not exist. For these cases, we favor a description of the most commonly used outcome measures in the trials as well as the most relevant ones from the patient's perspective. Examples for RA and OA are given in Table 1.

The CMSG accepts surrogate outcomes if other outcomes are not available, providing that they have been shown to be causally linked to patient outcomes, that the change in the surrogate largely captures the intervention's effect on the outcome, and that their limitations are explicitly acknowledged. For example, it is tempting to make inferences about the antifracture efficacy of pharmacotherapies on the basis of their effects on bone mineral density (BMD). However, there are many limitations associated with using BMD for this purpose. Studies using efficacy estimates from metaanalyses of randomized controlled trials (RCT) of antiresorptive therapies to explore the relationship between BMD and fractures using logistic regression analysis have demonstrated that the increases in BMD do not fully explain the reduction in fracture risk. It is important that the limitations are clearly outlined in the review.

Information on adverse effects as well as benefits should routinely be collected from the controlled trials. The Cochrane Adverse Effects Subgroup have developed recommendations for evaluating adverse effects[34]. The minimum recommendation is to collect the adverse effects reported in the trials included in the systematic review. However, where toxicity is a major concern, it is appropriate to do a more comprehensive review of adverse effects. One example is the review in progress by Rostom, *et al*, "Adverse gastrointestinal effects of COX-2 inhibitors for inflammatory diseases"[35].

The CMSG recommends that the categories of "major" and "secondary" are used to define various outcomes. The major category should include the core set of outcomes for the disease in question along with any adverse events. Other outcomes of interest should be placed under the secondary category.

Example of definition of outcomes:
Major
ACR 20 response — An ACR 20 response represents a 20% improvement in tender and swollen joint counts plus a 20% improvement in 3 of the 5 following remaining core measures: patient and physical global assessments, pain, functional status, and an acute phase reactant.
Low disease activity state[36] —
Adverse events —

Table 1. Core set outcome criteria for rheumatoid arthritis and osteoarthritis[20,26].

Rheumatoid Arthritis	Osteoarthritis
Improvement is denoted as ACR 20, ACR 50, or ACR 70 reflecting an improvement to the 20%, 50%, or 70% level in the following indicators: Tender joint count Swollen joint count And improvement in 3 of the following: Patient global assessment Physician global assessment Pain Disability Acute phase reactant (ESR rate)	OMERACT-OARSI set of responder criteria for definition of improvement: Improvement in at least 2 of the 3 following: Pain ≥ 20% and absolute change ≥ 10 Function ≥ 20% and absolute change ≥ 10 Patient global assessment ≥ 20% and absolute change ≥ 10

OMERACT-OARSI: Outcome Measures in Rheumatology Clinical Trials-Osteoarthritis Research Society International; ESR: erythrocyte sedimentation rate.

Data will be collected on: (1) total withdrawals; (2) withdrawals due to adverse effects; (3) withdrawals due to inefficacy; (d) all reported adverse effects such as infections and allergic reactions.

Secondary

Health-related quality of life such as the SF-36

Radiographic bone changes as measured by Sharp scores or Larsen scale for studies with a minimum duration of 12 months

STUDY DESIGN

Review authors should consider what study designs are likely to provide reliable data to answer their questions. Randomized controlled trials (RCT), where 2 or more groups are formed by randomly allocating participants so that any differences between groups can be attributed to the intervention, are preferred. In some treatment settings there is a danger that the intervention offered to the control group will be contaminated by individuals delivering the treatment option, for example, educational interventions. Cluster RCT (C-RCT) overcome this contamination by randomizing the different individual practices to different groups. Controlled clinical trials (CCT) are trials where allocation to treatment and control groups is quasi-random, for example, alternation, date of birth, or case record number.

Where no RCT or CCT have been identified, the following study designs may be considered for inclusion in a CMSG systematic review[37]:

• Controlled before and after studies (CBA) — CBA incorporate a nonrandomized control group. Data are collected on the control and intervention groups before the intervention is introduced, and then further data are collected after the intervention has been introduced. The reliability of the estimate of effect is questionable because there may be unidentified differences between the intervention and control groups that may have contributed to the effect.

• Interrupted time series (ITS) — ITS designs provide a method to measure the effect of an intervention when randomization or identification of a control group is impractical. Multiple data points are collected before and after the intervention. For inclusion in a CMSG review, measurements are required for at least 3 time points before the intervention and 3 time points after the intervention. The intervention effect is measured against the preintervention trend.

While RCT may be the gold standard for determining the benefits of an intervention, they may not adequately capture information about adverse effects of an intervention. The Cochrane Non-Randomised Trial Methods Group have developed guidelines for nonrandomized studies to standardize searching for rare and delayed adverse effects that will not be detected in short-term trials[34]. For reporting data on adverse effects, the CMSG recommends including other study designs such as pharmacovigilance reports; observational studies — cohort studies (prospective or retrospective), case-control studies; uncontrolled studies—simple case series, before-after comparison (no control group) and case reports.

For example, the US Food and Drug Administration website contains important data on tuberculosis and fungal infections from the use of biologics.

Example of study design. All RCT or CCT will be included. Observational studies will be included to address the issue of rare and delayed adverse effects.

Methodological quality assessment

All CMSG reviews require a standard measure of quality assessment. For controlled trials, many review authors opt to describe each component from the validated tool by Jadad, *et al*[38] that assesses randomization, blinding, and withdrawals and dropouts, supplemented by the allocation concealment assessment of Schulz, *et al*[39]. In addition, review authors may also use other quality assessment checklists.

Independent quality assessment using separate pre-piloted forms should be undertaken by at least 2 reviewers. Where differences in assessment cannot be resolved, arbitration by a third person, who could be a CMSG editor, is warranted. At present, masking of trial identifiers such as authors' and journals' names is not required; the only requirement is a statement in the Methods section of the review of whether masking was done.

In general, empiric research has shown that quality scores (numeric scores based on arbitrary weights given to each item in a scale) are arbitrary, unreliable, and hard to interpret[40,41]. Our suggestion, therefore, is to avoid using quality scores and use (at a minimum) the following individual components of a checklist: concealment of treatment allocation; blinding of intervention provider, recipient, and outcome assessment; handling of withdrawals and dropouts.

The Cochrane Handbook suggests summarizing individual quality criteria using three categories—low, moderate, and high risk of bias—corresponding to all individual criteria met, one or more criteria partially met, and one or more criteria not met[42]. Since allocation concealment is often not understood, detailed definitions are provided below. The "Characteristics of Included Studies" table in RevMan, the review manager software for Cochrane (http://www.cc-ims.net/RevMan), includes a column for description of allocation concealment with a classification of: adequate (A), unclear (B), inadequate (C), or that allocation concealment was not used (D) as a criterion to assess validity. The definitions provided are as follows:

(A) Adequate:

• centralized (e.g., allocation by a central office unaware of subject characteristics) or pharmacy-controlled randomization

• pre-numbered or coded identical containers which are administered serially to participants

• on-site computer system combined with allocations kept in a locked unreadable computer file that can be accessed only after the characteristics of an enrolled participant have been entered

• sequentially numbered, sealed, opaque envelopes.

Other approaches may include approaches similar to ones listed above, along with reassurance that the person who generated the allocation scheme did not administer it. Some schemes may be innovative and not fit any of the approaches above, but still provide adequate concealment.

(B) Unclear: When studies do not report any concealment approach, adequacy should be considered unclear. Examples include merely stating that a list or table was used, only specifying that sealed envelopes were used, and reporting an apparently adequate concealment scheme in combination with other information that leads the reviewer to be suspicious.

(C) Inadequate: includes alternation; the use of case record numbers, dates of birth or day of the week, and any procedure that is entirely transparent before allocation, such as an open list of random numbers.

(D) Not used: clearly stated that allocation concealment was not used.

Nonrandomized studies. Assessment of quality of nonrandomized (observational) studies is more difficult than assessment of quality of RCT. Quality assessment methods for nonrandomized studies are still under development. Recently, an evaluation of quality assessment tools in systematic reviews of nonrandomized studies was performed for the NHS Health Technology Assessment Programme[43]. Six tools were identified as potentially useful for systematic reviews. The CMSG recommends 2 of these tools for assessing the quality of nonrandomized studies in metaanalyses. The first, the Newcastle-Ottawa Scale, is a shorter tool taking 5-10 minutes to complete. The Powerpoint presentation, manual, and scale are available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm. The second, by Downs and Black[44], is a longer tool taking about 10–20 minutes to complete.

Once quality is assessed (using individual components of one of the available checklists), the influence of quality on effect estimates (summary relative risk, etc.) can be evaluated by sensitivity analysis (stratifying by criteria met or not met). Separate summary effect estimates can be generated for studies that meet and do not meet the individual quality criterion. Only when a large number of studies are identified for inclusion, approaches such as metaregression might be useful. The metaregression analysis models the outcome (odds ratio, for example) of each study as the dependent variable and will include quality variables as covariates (independent variables). Incorporating quality scores in the analysis as weights is not recommended.

Grading of the evidence

In an effort to make it easier for the end user to understand the strength of the quality of the evidence included in the review, we now recommend that an overall grade of the evidence for each major outcome is provided in each review. A simplified grading system that is easily understood and used by the consumer as well as the clinician was derived by the the editors of *Evidence-based Rheumatology*[45]. This simplified grading system focuses on a few validated criteria to decide which studies warrant the highest levels of Gold and Platinum; namely adequate sample size, completeness of followup, blinding of outcome assessors and patients, and concealment of allocation. There are 4 categories to rank the evidence from research studies: Platinum, Gold, Silver, and Bronze (Table 2).

Data Collection

It is the CMSG's policy to recommend that at least 2 reviewers independently select the studies, assess the methodological quality, and extract the data.

Data collection forms should be used on all CMSG reviews and it is recommended that they are piloted on a sample of studies. At the editorial office we have developed a collection of data collection forms which can be modified for each of the reviews. These forms include the following general items:

Identification: review author; study identifier; article title; journal reference; source of funding; contact author address

Study design: experimental design (parallel group, crossover); study duration

Methodology: allocation concealment; blinding (specifying blinded individuals such as assessor, patient, investigator, clinician); withdrawals and drop outs; quality criteria

Data analysis: whether or not intention-to-treat (ITT) analysis

Participants: trial inclusion criteria, exclusion criteria, diagnostic criteria, pretreatment group differences; country; setting

Intervention: intervention type (combination/single intervention, placebo-controlled); description of intervention and comparator treatment; dosage; any cointerventions; method of delivery; duration of treatment

Results: outcome measures (continuous/dichotomous); time intervals of endpoints; compliance to treatment; outcome measures for patient subgroups; note if data were extracted from graphs; separate transformations on raw data

Review authors are encouraged to contact trial authors to retrieve unpublished data or to clarify uncertainties in the reported trial whenever possible.

Data analysis

The CMSG suggests review authors reference Section 8: Analysing and Presenting Results of the Cochrane Handbook for Systematic Reviews of Interventions[46].

Metaanalysis should only be undertaken when data are clinically and statistically homogeneous. This may not be an easy decision to make and reviewers must explicitly state their reasons for combining results of studies in a metaanalysis. If data are available, sufficiently similar, and of sufficient quality, statistical analyses should be performed using RevMan. Review authors are encouraged to contact the CMSG coordinator should they have any questions.

The effect sizes for dichotomous outcomes may be expressed in terms of relative risk (RR), odds ratio (OR), or

Table 2. CMSG recommended grading system.

Platinum level	The Platinum ranking is given to evidence that meets the following criteria, as reported: that has at least 2 individual randomized controlled trials, each satisfying the following: • Sample sizes of at least 50 per group. If they do not find a statistically significant difference, they are adequately powered for a 20% relative difference in the relevant outcome • Blinding of patients and assessors for outcomes • Handling of withdrawals > 80% followup [imputations based on methods such as last observation carried forward (LOCF) acceptable] • Concealment of treatment allocation
Gold level	The Gold ranking is given to evidence if at least one randomized controlled trial meets all of the following criteria: • Sample sizes of at least 50 per group. If they do not find a statistically significant difference, they are adequately powered for a 20% relative difference in the relevant outcome • Blinding of patients and assessors for outcomes • Handling of withdrawals > 80% followup (imputations based on methods such as LOCF acceptable) • Concealment of treatment allocation
Silver level	The Silver ranking is given to evidence from a randomized trial that does not meet the above criteria. Silver ranking would also include evidence from at least one study of nonrandomized cohorts who did and did not receive the therapy or evidence from at least one case-control study. A randomized trial with a "head-to-head" comparison of agents is considered Silver level ranking unless a reference is provided to a comparison of one of the agents to placebo showing at least a 20% relative difference
Bronze level	The Bronze ranking is given to evidence if at least one case series without controls (including simple before/after studies in which the patient acts as their own control) or is derived from expert opinion based on clinical experience without reference to any of the foregoing (for example, argument from physiology, bench research, or first principles)

risk difference (RD). The CMSG recommends that RR is used to express dichotomous outcomes. RD is very vulnerable to baseline rates, while RR is less so. In cases when events are very rare, the Peto odds ratio is recommended.

For continuous outcomes, weighted mean differences (WMD) between the postintervention values, or the difference between baseline values and postintervention values, of the intervention and control groups should be used to analyze the size of the effects of the interventions. Standardized mean differences (SMD) should be used when results for continuous outcomes are presented on different scales; for example, the visual analog scale (VAS) and Likert pain scales.

Cluster randomized trials are often incorrectly analyzed. These trials frequently include a "unit of analysis error" meaning the data analysis is conducted at the individual level rather than the unit of allocation level. This error can be addressed by using statistical methods that correct for the clustering effect. When including cluster randomized trials in a review, we recommend that the CMSG statistical editor is consulted.

It is important to test for heterogeneity to determine whether the observed variation in study results is compatible with the variation expected by chance alone. When appropriate, publication bias should be investigated. Section 8 of the Cochrane Handbook for Systematic Reviews of Interventions provides further details on testing for heterogeneity and publication bias.

Subgroup and sensitivity analyses can be performed to examine the robustness of the results and the influence of other variables and should be specified *a priori*. For example, a sensitivity analysis could be carried out to see if effect sizes vary on pain, function, and radiologic measures when only those studies with adequate allocation concealment (score of A) are analyzed. A subgroup analysis could be planned to determine the effects of dosage or disease severity on the response to treatment. However, review authors must be aware of the limitations of undertaking subgroup and sensitivity analyses and the CMSG recommends that authors refer to Section 8 of the Cochrane Handbook for Systematic Reviews of Interventions.

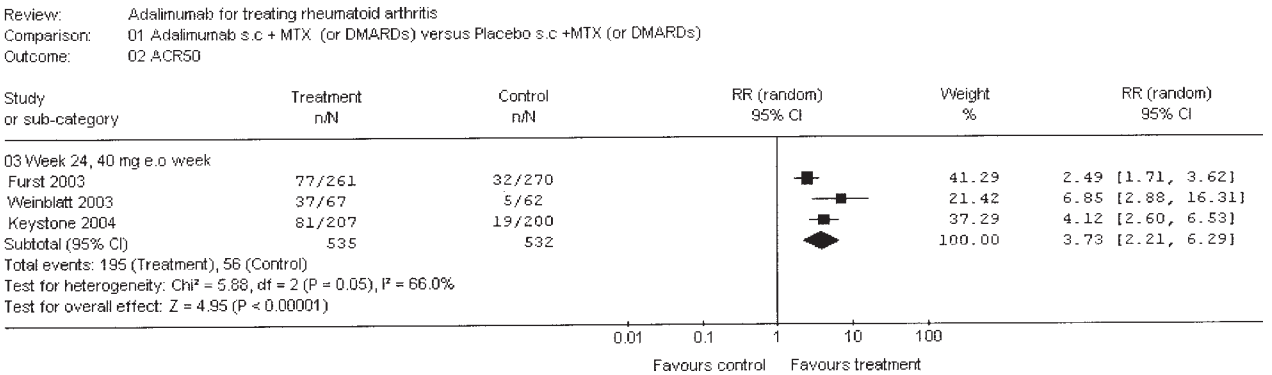

Figure 1. Forest plot from RevMan for adalimumab for treating rheumatoid arthritis.

Using RevMan, the results of individual studies can be presented graphically in forest plots (Figure 1). The effect size is shown as a square and the horizontal line determines the 95% confidence interval. The treatment effect is determined by the location of the square in relation to the vertical middle line. An effect size is considered to have no statistical significance when the confidence interval crosses the vertical middle line. When appropriate, data from more than one trial may be pooled in a metaanalysis and the diamond at the bottom of the graph provides an estimate of effect of this pooled data.

In addition to relative measures, it is important that absolute measures are also included in the review. The absolute risk reduction should be provided and for those outcomes that are statistically significant, the number needed to treat should be calculated. Further details for presenting the results are provided in the accompanying article in the section "Summarizing the evidence for the end user"[10].

CONCLUSION

Systematic reviews need to be conducted according to high methodological standards. Designed to accompany the detailed Cochrane Handbook for Systematic Reviews of Interventions[7], this report provides guidelines tailored to authors undertaking a review within the Musculoskeletal Group scope. These recommendations to develop and perform a review will help improve consistency among CMSG reviews.

ACKNOWLEDGMENT

The Editorial Board of the Cochrane Musculoskeletal Review Group: Peter Tugwell (Co-ordinating Editor), Rachelle Buchbinder (Co-ordinating Editor of Australian Satellite), Ann Cranney, Rob de Bie, Louise Falzon, Renea Johnston, Jessie McGowan, Beverley Shea, Maria Suarez-Almazor, George Wells, Gustavo Zanoli, Jim Davies (Consumer Editor), and Ann Qualman (Consumer Co-ordinator). Thanks also to the CMSG subgroup facilitators. For a list of our team of editors and subgroup facilitators visit http://www.cochranemsk.org/cmsg/who

REFERENCES

1. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomised control trials in health care from the Potsdam Consultation on Meta-Analysis. J Clin Epidemiol 1995;48:167-71.
2. Clark M. Cochrane Collaboration–Systematic reviews and the Cochrane Collaboration. Available at http://www.cochrane.org/docs/whycc.htm [accessed July 19, 2006]
3. Tugwell P, Maxwell L, Buchbinder R, et al. Musculoskeletal Group. About the Cochrane Collaboration [Collaborative Review Groups (CRGs) 2006, Issue 2.]
4. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994;272:1367-71.
5. Deeks JJ, Smith LA, Bradley MD. Efficacy, tolerability, and upper gastrointestinal safety of celecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials. BMJ 2002;325:619.
6. Wright J. The double-edged sword of COX-2 selective NSAIDs. CMAJ 2002;167:1131-7.
7. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005]. In: The Cochrane Library, Issue 3. Chichester, UK: John Wiley & Sons, Ltd.; 2005.
8. van Tulder MW, Assendelft WJ, Koes BW, et al. Method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group for Spinal Disorders. Spine 1997;22:2323–30.
9. van Tulder M, Furlan A, Bombardier C, Bouter L; Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine 2003;28:1290–9.
10. Santesso N, Maxwell L, Tugwell PS, Wells GA, O'Connor AM, Judd M, Buchbinder R, and the Editorial Board of the Cochrane Musculoskeletal Group. Knowledge transfer to clinicians and consumers by the Cochrane Musculoskeletal Group. J Rheumatol 2006;33:2312-8.
11. Tugwell P, Shea B, Boers M, et al (editors). Evidence Based Rheumatology. London: BMJ Books; 2003.
12. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. BMJ 1994;309:1286-91.
13. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of controlled trials using PubMed. International Journal of Epidemiology 2002;31:150-3.
14. 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;347:363-6.
15. Klippel JH, Goronzy JJ, Weyand CM, Anderson RJ, editors. Matteson primer on the rheumatic diseases. 12th ed. Atlanta, Ga: Arthritis Foundation; 2001.
16. Blumenauer B, Judd M, Cranney A, et al. Infliximab for the treatment of rheumatoid arthritis. The Cochrane Database of Systematic Reviews 2002, Issue 3. Art. No.: CD003785. DOI: 10.1002/14651858.CD003785.
17. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315-24.
18. Navarro-Sarabia F, Ariza-Ariza R, Hernandez-Cruz B, Villanueva I. Adalimumab for treating rheumatoid arthritis. The Cochrane Database of Systematic Reviews 2005, Issue 3. Art. No.: CD005113. DOI: 10.1002/14651858.CD005113.pub2.
19. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT filter for outcomes measures in rheumatology. J Rheumatol 1998;25:198–9.
20. Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum 1993;36:729-40.
21. Boers M, Tugwell P, Felson DT, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol 1994;21 Suppl 41:86-9.
22. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
23. van der Heijde D, Bellamy N, Calin A, Dougados M, Khan MA, van der Linden S. Preliminary core sets for endpoints in ankylosing spondylitis. Assessments in Ankylosing Spondylitis Working Group. J Rheumatol 1997;24:2225-9.
24. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitits. Progress report of the ASAS Working Group. J Rheumatol 1999;26:951-4.
25. Anderson JJ, Baron G, van der Heijde D, Felson D, Dougados M. Ankylosing spondylitis assessment group: preliminary definition of

short-term improvement in ankylosing spondylitis. Arthritis Rheum 2001;44:1876-86.
26. Brandt J, Listing J, Sieper J, Rudwaleit M, van der Heijde D, Braun J. Development and preselection of criteria for short term improvement after anti-TNF alpha treatment in ankylosing spondylitis. Ann Rheum Dis 2004;63:1438-44.
27. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
28. Pham T, van der Heijde D, Altman RD, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004; 12:389-99.
29. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40:1202–9.
30. Strand V, Gladman D, Isenberg D, Petri M, Smolen J, Tugwell P. Outcome measures to be used in clinical trials in systemic lupus erythematosus. J Rheumatol 1999;26:490-7.
31. Smolen JS, Strand V, Cardiel M, et al. Randomized clinical trials and longitudinal observational studies in systemic lupus erythematosus: consensus on a preliminary core set of outcome domains. J Rheumatol 1999;26:504-7.
32. Gladman DD, Urowitz MB, Goldsmith CH, et al. The reliability of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index in patients with systemic lupus erythematosus. Arthritis Rheum 1997;40:809-13.
33. Sambrook PN, Cummings SR, Eisman JA, et al. Guidelines for osteoporosis trials. J Rheumatol 1997;24:1234-6.
34. Loke Y, Price D, Herxheimer A, editors, on behalf of the Cochrane Adverse Effects Subgroup. Including adverse effects in your review. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005]; Appendix 6b. In: The Cochrane Library, Issue 3. Chichester, UK: John Wiley & Sons, Ltd.; 2005.
35. Rostom A, Dube C, Boucher M, et al. Adverse gastrointestinal effects of COX-2 inhibitors for inflammatory diseases. The Cochrane Database of Systematic Reviews 2003, Issue 4. Art. No.: CD004519. DOI: 10.1002/14651858.CD004519.
36. Wells G, Boers M, Shea B, et al. Low disease activity state for rheumatoid arthritis: a preliminary definition. J Rheumatol 2005;32:2016-24.
37. Methodology in the Cochrane Effective Practice and Organisation of Care Group. http://www.epoc.uottawa.ca/FAQincludedstudies2003.pdf [accessed July 19, 2006]
38. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomised clinical trials: Is blinding necessary? Controlled Clin Trials 1996;17:1-12.
39. 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.
40. 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.
41. Greenland S. Invited commentary: a critical look at some popular meta-analytic methods. Am J Epidemiol 1994;140:290-6.
42. Higgins JPT, Green S, editors. Assessment of study quality. Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005]; Section 4. In: The Cochrane Library, Issue 3. Chichester, UK: John Wiley & Sons, Ltd.; 2005.
43. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003; 7:1-173. http://www.ncchta.org/execsumm/summ727.htm
44. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomised studies for health care interventions. J Epidemiol Commun Health 1998;52:377-84.
45. Tugwell P, Shea B, Boers M, et al, editors. Evidence-Based Rheumatology. London: BMJ Books; 2004.
46. Deeks JJ, Higgins JPT, Altman DG, editors. Analysing and presenting results. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.5 [updated May 2005]; Section 8. In: The Cochrane Library, Issue 3. Chichester, UK: John Wiley & Sons, Ltd.; 2005.