Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group

van Tulder, M.W.; Furlan, A.; Bombardier, C.; Bouter, L.M.

published in
Spine
2003

DOI (link to publisher)
10.1097/00007632-200306150-00014

document version
Publisher's PDF, also known as Version of record

SPINE Volume 28, Number 12, pp 1290–1299

Updated Method Guidelines for Systematic Reviews in the Cochrane Collaboration Back Review Group

Maurits van Tulder, PhD,*†‡ Andrea Furlan, MD,‡ Claire Bombardier, MD, FRCP,‡§‖¶#
Lex Bouter, PhD,* and the Editorial Board of the Cochrane Collaboration Back Review Group

Study Design. Descriptive method guidelines.

Objectives. To help reviewers design, conduct, and report reviews of trials in the field of back and neck pain.

Summary of Background Data. In 1997, the Cochrane Collaboration Back Review Group published method guidelines for systematic reviews. Since its publication, new methodologic evidence emerged and more experience was acquired in conducting reviews.

Methods. All reviews and protocols of the Back Review Group were assessed for compliance with the 1997 method guidelines. Also, the most recent version of the Cochrane Handbook (4.1) was checked for new recommendations. In addition, some important topics that were not addressed in the 1997 method guidelines were included (*e.g.*, methods for qualitative analysis, reporting of conclusions, and discussion of clinical relevance of the results). In May 2002, preliminary results were presented and discussed in a workshop. In two rounds, a list of all possible recommendations and the final draft were circulated for comments among the editors of the Back Review Group.

Results. The recommendations are divided in five categories: literature search, inclusion criteria, methodologic quality assessment, data extraction, and data analysis. Each recommendation is classified in minimum criteria and further guidance. Additional recommendations are included regarding assessment of clinical relevance, and reporting of results and conclusions.

Conclusions. Systematic reviews need to be conducted as carefully as the trials they report and, to achieve full impact, systematic reviews need to meet high methodologic standards. [Key words: systematic reviews, meta-analysis, Cochrane Collaboration, method guidelines, back pain, neck pain] **Spine 2003;28:1290–1299**

The current interest in evidence-based medicine has led to an extensive increase in the publication of systematic reviews. Recently, the QUOROM statement was developed to improve the standards for the report of systematic reviews.[1] Several leading medical journals (*e.g.*, *BMJ*, *JAMA*, *Lancet*) have adopted the QUORUM recommendations for the reporting of abstract, introduction, methods, results, and discussion sections of systematic reviews. However, it has been shown that many reviews in the field of back pain are of low methodologic quality and that their reports often lack essential components.[2,3]

In 1997, the editorial board of the Cochrane Collaboration Back Review Group (BRG) published method guidelines for systematic reviews in the field of spinal disorders.[4] These method guidelines addressed the main steps in conducting a systematic review: literature search, inclusion criteria, methodologic quality, data extraction, and data analysis. The purpose of the method guidelines was to offer guidance to researchers preparing, conducting, or reporting a systematic review and to readers evaluating these reviews. The guidelines were operationalized specifically for the field of back pain. They included certain minimum criteria for which empirical evidence existed that they are associated with bias in systematic reviews or for which consensus existed among the editorial board of the BRG that they are likely to be associated with bias. In addition, further guidance was presented to enhance the quality of systematic reviews.

The BRG was established in 1996 and did not have a systematic review published in the Cochrane Library at the time of publication of the method guidelines.[5] At present, 21 completed reviews and 7 protocols of various treatments for spinal disorders are published in the Cochrane Library (issue #4, 2002). Many of these reviews have a copublication in *Spine* (for more information, visit www.cochrane.iwh.on.ca). Because new evidence on review methodology has emerged since 1997 and the BRG has acquired ample experience in preparing, conducting, and reviewing systematic Cochrane reviews, it seems necessary to update the 1997 method guidelines. For example, the 1997 method guidelines included recommendations for quantitative (meta-)analysis. However, in many of the reviews within the BRG, the authors refrained from statistical pooling or concluded that insufficient data were provided to enable statistical pooling and consequently conducted a qualitative analysis (or best-evidence synthesis) for which no clear guidelines were provided. Also, only a few reviews currently published within the BRG include a discussion of the clinical relevance of trial results. Lastly, the reporting of conclu-

From the *Institute for Research in Extramural Medicine and the †Department of Clinical Epidemiology & Biostatistics, VU University Medical Center, Amsterdam, the Netherlands, the ‡Institute for Work & Health, §University of Toronto, ‖University Health Network, Toronto General Hospital, ¶Mount Sinai Hospital, and #Canada Research Chair in Knowledge Transfer for Musculoskeletal Care, Toronto, Ontario, Canada.
Funding provided by the Institute for Work & Health (Toronto, Ontario, Canada); CIHR (Canadian Institute of Health Research); and the Institute for Research in Extramural Medicine (EMGO-Institute), Amsterdam, The Netherlands.
The manuscript submitted does not contain information about medical device(s)/drug(s).
No funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.
Address correspondence and reprint requests to Maurits van Tulder, PhD, Institute for Research in Extramural Medicine, VU University Medical Centre, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands; E-mail: mw.van_tulder.emgo@med.vu.nl

Table 1. Comparison of the Minimum Criteria of the 1997 Cochrane Back Review Group Method Guidelines With the Cochrane Handbook (Version 4.1) and Compliance of Cochrane Back Review Group Protocols and Reviews With These Recommendations

Minimum Criteria	Cochrane Handbook	Protocols	Review
Literature search			
MEDLINE	Yes	27/28	96%
EMBASE	Yes	27/28	96%
Cochrane Library (CENTRAL)	Yes	24/28	86%
Sensitive Search strategy by UK Cochrane Centre	Yes	23/28	82%
Screening of references	Yes	25/28	89%
Handsearch Cochrane Centre Baltimore	No	0	0%
Personal communication with content experts	Yes	17/28	61%
Inclusion criteria			
Inclusion of randomized controlled trials	Yes	28/28	100%
Description of participants by age	Yes	23/28	82%
Description of participants by acute, subacute, chronic	Yes	16/28	57%
Description of participants by localization of pain	Yes	28/28	100%
Description of participants by type of symptoms	Yes	23/28	82%
Explicit description of type of intervention	Yes	26/28	93%
Explicit description of outcome measure	Yes	28/28	100%
Symptoms as outcome	Yes	27/28	96%
Overall improvement as outcome	Yes	19/28	68%
Function as outcome	Yes	26/28	93%
Well being as outcome	Yes	20/28	71%
Disability as outcome	Yes	22/28	79%
Methodological quality assessment			
Assessment by at least 2 independent reviewers	Yes	25/28	89%
Item: concealment of treatment allocation/randomization	Yes	23/28	82%
Item: drop-out rate	Yes	23/28	82%
Item: blinding of patients	Yes	17/28	61%
Item: blinding of observers	Yes	22/28	79%
Item: intention-to-treat analysis	No	18/28	64%
Data extraction			
Data extracted by at least 2 reviewers	Yes	25/28	89%
Data analysis			
Reason behind choice for quantitative or qualitative Review	Yes	18/28	64%
Explicit description of approach	Yes	26/28	93%

sions in reviews within the BRG varies widely. Therefore, we decided to modify existing and add some new recommendations in an update of the 1997 method guidelines.

It should be emphasized that these guidelines are not a "gold standard" but merely an indication of the current "state-of-the-art" of review methods. The method guidelines are useful to plan, conduct, or evaluate other systematic reviews in the field of back pain within and outside the framework of the BRG. The usefulness of the 1997 method guidelines is reflected in the number of citations in published scientific articles (n = 63, November 10th, 2002).

■ Methods

In order to assess the need for an update, all reviews and protocols of the BRG were assessed for compliance with the 1997 method guidelines. Also, the most recent version of the Cochrane Handbook (4.1) was checked for new recommendations. In addition, the editorial board of the BRG recognized that some important topics were not addressed in the 1997 method guidelines (*e.g.*, methods for qualitative analysis, reporting of conclusions, and discussion of clinical relevance of the results).

The compliance of the reviews and protocols with the minimum criteria and further guidance recommendations of the 1997 method guidelines is shown in Tables 1 and 2, respectively. The Cochrane Library issue #4, 2002, contained 21 completed reviews and 7 protocols of the BRG.

In March 2002, a literature search was conducted in the Cochrane Database of Methodology (from 1997–2002) and in PubMed for articles dealing with methods of conducting a systematic review. A total of 584 titles and abstracts were screened, and hard copies of selected papers were retrieved. The recommendations were updated if new evidence was found, and the latest relevant publications were included in the references (even if the evidence did not change, we replaced older references by more recent ones).

In May 2002, preliminary results were presented and discussed in a workshop at the Fifth International Forum for Primary Care Research in Low Back Pain involving members of the BRG editorial board, (co)authors of reviews, and other experts on low back pain research interested in systematic review methodology. After the workshop, a list of all possible recommendations, including all items of the 1997 method guidelines plus additional items from the literature search, the Cochrane Handbook, and experience with BRG reviews, was circulated among the editors of the BRG. Each editor was asked to decide if a specific recommendation should be included as a minimum criterion or as further guidance, or if it should not be included at all. Six of the eight editors participated in this process. A final version of the updated method guidelines was prepared and circulated among all BRG editors for comments.

Table 2. Comparison of the Further Guidance of the 1997 Cochrane Back Review Group Method Guidelines With the Cochrane Handbook (Version 4.1), and Compliance of Cochrane Back Review Group Protocols and Reviews With These Recommendations

Further Guidance	Cochrane Handbook	Protocol	Review
Literature search			
Specific database for intervention at issue	Yes	23/28	82%
Handsearch key journals	No	5/28	18%
Citation tracking (Science Citation Index)	Yes	14/28	50%
Identification of trials by at least 2 reviewers	Yes	24/28	86%
Pilot test inclusion criteria	Yes	3/28	11%
Consensus method to solve disagreement	Yes	23/28	82%
Inclusion criteria			
Reason for inclusion of non-randomized trials	Yes	5/19	26%
Inclusion of physical examination as outcome	Yes	16/28	57%
Inclusion of other symptoms as outcome	Yes	12/28	43%
No language restrictions	No	17/28	61%
Reason for exclusion of non-English articles	No	3/11*	27%
Methodological quality assessment			
Assessment blinded for authors, institutions, journal	Yes	6/28	21%
Assessment by content expert and non-expert	Yes	2/28	7%
Pilot test of quality assessment	Yes	7/28	25%
Consensus method to solve disagreement	Yes	21/28	75%
Assessment of interobserver reliability	Yes	13/28	46%
Use of recommended criteria list	Yes	12/28	43%
Explicit operationalization of criteria	Yes	12/28	43%
Data extraction			
Extraction blinded for authors, institutions, journal	Yes	5/28	18%
Use of standardized form	Yes	17/28	61%
Pilot test the data extraction form	Yes	4/28	14%
Consensus method to solve disagreements	Yes	10/28	36%
Contact author for additional information	Yes	13/28	46%
Data analysis			
Explicit description of comparisons	Yes	18/28	64%
Analysis of clinical heterogeneity	Yes	16/28	57%
Statistical test for heterogeneity	Yes	15/28	54%
Use of random effects model	Yes	4/9†	44%
Subgroup analysis related to review objectives	Yes	17/28	61%

* Eleven reviews/protocols excluded non-English language trials.
† Nine reviews included a meta-analysis.

■ Method Guidelines

Literature Search

Minimum Criteria. Because one of the main principles of a systematic review is to include all available evidence, the literature search is the first and a very important step in conducting a systematic review. The starting point for the literature search is to decide which articles should be retrieved, ensuring that as many relevant trials as possible are identified. The search strategy should relate directly to the research question(s) of the review at issue and should be based on the inclusion criteria regarding study design, participants, interventions, outcomes, and language (see *Inclusion Criteria* section). A simple MEDLINE search is clearly insufficient, because it has been shown that in general only approximately half of the available randomized clinical trials (RCTs) will be identified if the identification of RCTs solely depends on MEDLINE searches.[6] It has been suggested that at least MEDLINE and EMBASE must be used to ensure a comprehensive literature search, because overlap between these databases is small.[7,8] Especially in the field of low back pain, EMBASE seems to retrieve more clinical trials than MEDLINE.[9]

Therefore, we recommend the following as a minimum search strategy:

1. A computer aided search of the MEDLINE and EMBASE databases since their beginning. The highly sensitive search strategies for retrieval of reports of controlled trials[10] should be run in conjunction with a specific search for spinal disorders and the intervention at issue (see *Appendix 1 and 2*).
2. A search of the Cochrane Central Register of Controlled Trials (Central) that is included in the latest issue of the Cochrane Library.[11]
3. Screening references given in relevant systematic reviews and identified RCTs.
4. Personal communication with content experts in the field.[12] It is to the discretion of the reviewers to identify who the experts are on a specific topic.

We recommend that two reviewers independently apply the inclusion criteria to select the potentially relevant trials from the titles, abstracts, and key words of the references retrieved by the literature search. We suggest pilot testing the inclusion criteria on a sample of articles, including some considered to be definitely eligible, some definitely not eligible, and some questionable.[13] Articles for which disagreement exists and articles for which title, abstract, and key words provide insufficient information for a decision on selection should be obtained to assess whether they meet the inclusion criteria. A consensus method should be used to solve disagreements between the two reviewers regarding the inclusion of studies. A third reviewer should be consulted if disagreements are not resolved in the consensus meeting.

Further Guidance. Depending on the intervention at issue, if available, a specific database should be searched; for example:

- Mantis (Manual Alternative and Natural Therapy Index System; formerly Chirolars) for chiropractic intervention (http://www.chiroaccess.com/MANTISAbout.html)
- Ciscom (Research Council for Complementary Medicine, London, UK) for complementary medicine interventions (http://www.rccm.org.uk/ciscom/CISCOM_intro.asp)
- PsycINFO for psychological interventions (http://www.apa.org/psycinfo)
- PEDro (Physiotherapy Evidence Database) for physiotherapy (http://ptwww.cchs.usyd.edu.au/pedro)

Citation tracking of the identified RCTs (use of Science Citation Index to search forward in time for subsequent articles that have cited the identified RCTs) is recommended. The value of using citation tracking has not yet been established, but it may be especially useful to identify additional studies on topics that are poorly indexed in MEDLINE and EMBASE.

Inclusion Criteria

Minimum Criteria

Study Design. Randomized clinical trials should be included. Nonrandomized controlled clinical trials (CCTs) may be included if there are less than five RCTs. If authors wish to extend a systematic review beyond RCTs by also including observational studies, the reason for this should be outlined.

Participants. Participants of trials that will be included in the systematic review should be defined explicitly in terms of age, gender, duration of symptoms, localization of symptoms, type of symptoms, and setting. It is especially recommended to clearly report if patients with acute (less than 6 weeks), subacute (6–12 weeks) and/or chronic (12 weeks or more) low back pain are included.

Interventions. The type, intensity, frequency, and duration of index interventions that will be included in the systematic review should be explicitly described as well as the specific comparison groups. (For example, spinal manipulation *versus* placebo manipulation, spinal manipulation *versus* other active interventions, or spinal manipulation *versus* no treatment.)

Outcomes. The outcome measures and instruments that will be included in a systematic review should be explicitly described. Outcomes of symptoms (*e.g.*, pain), overall improvement or satisfaction with treatment, function (*e.g.*, back specific functional status), well being (*e.g.*, quality of life), disability (*e.g.*, activities of daily living, work absenteeism), and side effects should always be included in a systematic review of back pain if they are reported in the original trials. Timing of outcome measures should be explicitly described, that is, measurement of outcomes less than 3 months after randomization (short-term follow-up), between 3 months and 1 year (intermediate follow-up), or 1 year or more (long-term follow-up).

For examples of descriptions of inclusion criteria, see the 1997 BRG method guidelines[4] or published Cochrane reviews.

Further Guidance

Study Design. Description of funding sources and potential conflict of interest of authors of trials is recommended.

Outcomes. Outcomes of physical examination (range of motion, spinal flexibility, degrees of straight leg raising, or muscle strength), other symptoms (medication use, health care utilization), and economic outcomes may be included where appropriate, considering the aim of the intervention at issue. Depending on the intervention specific outcomes may be relevant, for example depression for a review of antidepressants.

Language. The empirical evidence on exclusion of trials published in other languages than English is conflicting. Some authors suggested that it might be associated with bias,[14–16] whereas others showed that it has generally little effect on summary treatment effect estimates.[17,18] Inclusion of studies published in other languages than English is recommended, although we acknowledge that it may not always be feasible and may depend on the time and resources available. If trials published in other languages are excluded from the review, the reason for this decision should be given. The BRG strongly recommends having an international group of (co)authors with different language skills involved in a systematic review to enable including more languages than English. This is particularly recommended on topics where there are likely to be a significant number of non-English language publications (for example, the Asian literature on acupuncture); we recommend involving a collaborator with relevant language skills.

Methodologic Quality

Minimum Criteria. The methodologic quality of the studies should be independently assessed by at least two reviewers.

Currently, there is empirical evidence that inadequate concealment of treatment allocation[19–22] and inadequate double-blinding (of patients and outcome assessors)[20] are associated with bias. There is no evidence for any of the other methodologic criteria.[22] Furthermore, this evidence is collected in other fields than back pain, and it is still unclear whether this evidence is also valid for back pain studies.

The Editorial Board of the BRG strongly recommends including the following quality items: concealment of treatment allocation, dropout rate, blinding of patients, blinding of outcome assessor, and intention-to-treat analysis. The operationalization of the criteria should be described explicitly (for example, Table 3), and the criteria should be scored as positive, negative or unclear ("yes," "no," and "don't know").

Further Guidance. Some empirical evidence has been provided that blinded assessment of the methodologic quality of trials, that is, removing authors, institution, and journal from copies when assessing the methodologic quality, resulted in lower and more consistent scores than open assessment.[23] However, two other studies did not find an association between blinded assessment of studies and bias.[24,25] It is difficult to achieve true blinding, because experts are usually involved in the assessment of the methodologic quality of the studies. Therefore, the BRG leaves it to the discretion of the reviewers to perform a blinded quality assessment or not. Because the quality assessment of content experts may be biased by prior opinions, it may be desirable to have both a clinical content expert and a nonexpert (but with a methodologic background) assess the quality of the studies. In systematic reviews where there is likely to be a conflict of interest (*e.g.*, chiropractors reviewing spinal manipulation, or physiotherapists reviewing exercise therapy), it may be desirable to additionally mask the

Table 3. Operationalization of the Criteria List

A	A random (unpredictable) assignment sequence. Examples of adequate methods are computer generated random number table and use of sealed opaque envelopes. Methods of allocation using date of birth, date of admission, hospital numbers, or alternation should not be regarded as appropriate.
B	Assignment generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about eligibility of the patient.
C	In order to receive a "yes," groups have to be similar at baseline regarding demographic factors, duration and severity of complaints, percentage of patients with neurologic symptoms, and value of main outcome measure(s).
D	The reviewer determines if enough information about the blinding is given in order to score a "yes."
E	The reviewer determines if enough information about the blinding is given in order to score a "yes."
F	The reviewer determines if enough information about the blinding is given in order to score a "yes."
G	Cointerventions should either be avoided in the trial design or similar between the index and control groups.
H	The reviewer determines if the compliance to the interventions is acceptable, based on the reported intensity, duration, number and frequency of sessions for both the index intervention and control intervention(s).
I	The number of participants who were included in the study but did not complete the observation period or were not included in the analysis must be described and reasons given. If the percentage of withdrawals and drop-outs does not exceed 20% for short-term follow-up and 30% for long-term follow-up and does not lead to substantial bias a "yes" is scored. (N.B. these percentages are arbitrary, not supported by literature).
J	Timing of outcome assessment should be identical for all intervention groups and for all important outcome assessments.
K	All randomized patients are reported/analyzed in the group they were allocated to by randomization for the most important moments of effect measurement (minus missing values) irrespective of noncompliance and cointerventions.

studies also for results and conclusions or to include as a reviewer someone who has no conflict of interest in the assessment of the quality of the studies. Also, if one of the reviewers is a/an (co)author of one of the included trials, this person should not be involved in the quality assessment of the trial at issue.

We recommend that reviewers pilot-test the methodologic quality assessment on some similar articles (regarding another intervention or disorder) that will not be included in the review. It is important for reviewers to agree on a common interpretation of the items and their operationalization.

We recommend using a consensus method to discuss and solve the disagreements between the reviewers. If disagreement persists, an additional independent person should be consulted who is an expert in review methodology. The initial interobserver reliability (*e.g.*, Kappa) of the quality assessment may be evaluated and reported.

A recent study in the field of rheumatology showed that some trials that inadequately reported the method of randomization and allocation concealment had actually been performed adequately.[26] Therefore, if the article does not contain information on (one or more of) the methodologic criteria, the authors may be contacted for additional information. If the authors cannot be contacted or if the information is no longer available, the criteria should be scored as "unclear." If many studies are scored as "unclear" for the same criterion, the reviewers could consider dropping that criterion *post hoc* (Chapter 6 of the Cochrane Handbook).[13]

We recommend using the list presented in Table 4 and the operationalization presented in Table 3. These criteria have already been used in a number of systematic reviews within the BRG. Compared to the 1997 method guidelines, this list includes only the internal validity criteria (n = 11) that refer to characteristics of the study that might be related to selection bias (criteria a and b), performance bias (criteria d, e, g, and h), attrition bias (criteria i and k) and detection bias (criteria f and j). The internal validity criteria should be used to define methodologic quality in the meta-analysis. Some recent evidence suggests that using a different summary quality score may lead to a different interpretation in systematic reviews.[27] Therefore, the impact of individual internal validity criteria on the overall effect estimate should be evaluated if possible. Unfortunately, this will often not be the case as statistical pooling (in subgroups) is usually not feasible in back pain reviews.

Differences in methodologic quality may explain variation in the results of the studies included in a systematic review and can result in over- or underestimation of the effectiveness of the intervention at issue. However, there are no strict guidelines for the use of methodologic quality assessment in systematic reviews. In general, the BRG recommends choosing from the options listed below and clearly describing the reasoning behind the choice.[28–30] First, the methodologic quality can be used as an additional criterion for inclusion of studies in the review based on one or more items (*e.g.*, only inclusion of RCTs or double-blind RCTs) or based on a summary quality score (*e.g.*, only studies that adequately fulfill 50% or more or 6 out of 11 of the validity criteria). Choosing cut-off points for inclusion or exclusion of studies remains arbitrary. Second, a stratified analysis can be performed in which the results are separately presented for different strata of studies (*e.g.*, studies

Table 4. Criteria List for the Methodological Quality Assessment

A	Was the method of randomization adequate?	Yes/No/Don't know
B	Was the treatment allocation concealed?	Yes/No/Don't know
C	Were the groups similar at baseline regarding the most important prognostic indicators?	Yes/No/Don't know
D	Was the patient blinded to the intervention?	Yes/No/Don't know
E	Was the care provider blinded to the intervention?	Yes/No/Don't know
F	Was the outcome assessor blinded to the intervention?	Yes/No/Don't know
G	Were cointerventions avoided or similar?	Yes/No/Don't know
H	Was the compliance acceptable in all groups?	Yes/No/Don't know
I	Was the drop-out rate described and acceptable?	Yes/No/Don't know
J	Was the timing of the outcome assessment in all groups similar?	Yes/No/Don't know
K	Did the analysis include an intention-to-treat analysis?	Yes/No/Don't know

that do or do not meet a specific quality item or low and high methodologic quality studies defined by a preset cut-off point on a summary quality score). Third, a sensitivity analysis can be performed to determine whether the overall results are the same when studies above different methodologic cut-off points are analyzed. Fourth, weights can be applied in the analysis to studies according to the methodologic quality, so that studies of higher quality have more impact on the overall results. Obviously, choosing weights involves additional arbitrary decisions. Fifth, a cumulative meta-analysis can be performed by examining the impact on the overall results as studies of decreasing quality are included one at a time. And lastly, a meta-regression can be performed to explore the relation between methodologic quality items and the magnitude of effect across studies. The first four options are also available when statistical pooling is not feasible, the last two specifically apply to statistical pooling.

Data Extraction

Minimum Criteria. At least two reviewers should independently extract the data. The data that should be extracted and presented in a Table with study characteristics including characteristics of participants, interventions, outcomes, and results (see *Inclusion Criteria*).

Further Guidance. The BRG recommends using a standardized form for data extraction that will facilitate the comparison process. It is advisable to pilot test the data extraction form to minimize misinterpretations or later disagreements. If there are any disagreements, consensus should be achieved by discussion among the reviewers. If disagreements persist, an additional independent person should be consulted. If the article does not contain sufficient information, the authors may be contacted.

Data extraction forms will vary across different systematic reviews, but there will also be similarities among the forms needed for reviews on back pain. Because designing a data extraction form is time consuming, and given the important function of data extraction forms, it may be helpful to profit and learn from experiences of others. Examples of data extraction forms used in reviews can be obtained at the BRG Web site: www.cochrane.iwh.on.ca

Data Analysis

Minimum Criteria. The BRG recommends performing a quantitative or qualitative analysis only on groups of studies that have been judged sufficiently clinically similar to yield meaningful results. This means reviewers should avoid grouping studies that are clinically heterogeneous. The comparisons with regard to the outcomes and subgroups that will be analyzed should relate directly to the objective or research question(s) of the review. The analysis should include an explicit description of the comparisons (Table 5).

As in every empirical study, statistical methods can be used to analyze and summarize the data in a systematic review. The objective of most systematic reviews will be to provide a reliable estimate of the effects of an intervention in comparison with alternatives for clinically relevant subgroups of studies. There is consensus among the Editorial Board of the BRG that if relevant, valid data are lacking (data are too sparse or of inadequate quality) or if data are statistically too heterogeneous (and the heterogeneity cannot be explained), statistical pooling (meta-analysis) should be avoided and reviewers should perform a qualitative analysis. In these instances, a qualitative analysis may be performed by attributing various levels of evidence to the effectiveness of a treatment, taking into account the participants, interventions, outcomes, and methodologic quality of the original studies. Also, if only a subset of available trials provide sufficient data for inclusion in a meta-analysis (*e.g.*, some trials do not report standard deviations [SDs]), both a quantitative and qualitative analysis could be used in addition to each other. Because various approaches are possible in analyzing the data in a systematic review, we suggest that authors clearly outline the reasoning behind the approach they use.

Table 5. Example of an Analysis for a Systematic Review on Traction for Low Back Pain

- Subgroup 1: Acute low back pain without neurological symptoms
  - Comparison 1.1: traction vs. placebo/sham/no treatment
    - Outcome 1.1.1: pain intensity
      - Follow-up:
        - Short-term
        - Intermediate
        - Long-term
    - Outcome 1.1.2: functional status
      - Follow-up:
        - Short-term
        - Intermediate
        - Long-term
    - Outcome 1.1.3: .....
  - Comparison 1.2: traction vs. active treatments
    - Outcome 1.1.1: pain intensity
      - Follow-up:
        - Short-term
        - Intermediate
        - Long-term
    - Outcome 1.1.2: functional status
      - Follow-up:
        - Short-term
        - Intermediate
        - Long-term
    - Outcome 1.1.3: .....
- Subgroup 2: Acute low back pain with neurological symptoms
  - Comparison 2.1
    - Outcome 2.1.1
      - Follow-up: .....
- Subgroup 3: Chronic low back pain without neurological symptoms

Further Guidance. The Editorial Board of the BRG suggests referring to the recommendations of Chapter 8 of the Cochrane Handbook for further guidance on data-analysis.[13]

Quantitative Analysis. If it is clinically relevant and statistically justified to combine the results, statistical pooling should be performed that provides an overall effect estimate, with a 95% confidence interval for each comparison.[31,32] The BRG recommends contacting a statistician before performing a quantitative analysis. A meta-analysis should start with examining potential publication and other biases by a funnel plot exploring asymmetry among trial results.[33] If asymmetry is present, potential reasons should be explored. However, funnel plots may be misleading and should be interpreted cautiously.[34] Formal statistical tests also exist, but there is no consensus regarding the strengths and weaknesses of these tests.[33,35]

For the meta-analysis of dichotomous outcomes, the relative risk, risk difference, or odds ratio can be used to summarize the effect. Empirical evidence from 125 meta-analyses showed that

summary odds ratios and risk differences usually lead to similar conclusions about treatment effect, but that risk differences are substantially more heterogeneous.[36] For continuous outcomes, mean differences from each trial can be combined. If the continuous outcomes are not directly combinable—that is, if different instruments are used for the outcome measurements—standardized mean differences (effect sizes) might be used.[31,32] If necessary, the authors of the original studies may be contacted to provide relevant information.

There are two statistical models for combining data in a meta-analysis: the fixed effects model and the random effects model.[31] Although there are arguments favoring each model, in general, the clinical heterogeneity of the back pain literature suggests that the assumptions underlying the random effects model are better suited to statistical combinations of different trials in this field. However, the random effects model does not account for the heterogeneity, does not explain it, and does not take it away. Careful analysis of heterogeneity, that is, of study characteristics that might explain differences among the results, is always important.[37] An explicit list should be given of the characteristics of participants in, the types of interventions for, and the exact outcome values from, each study of a group of studies that are combined. Sensitivity analyses should be performed to examine the impact of variation in summary validity scores or individual validity items (see *Methodologic Quality* section).

It may be difficult sometimes for reviewers to decide whether it is clinically relevant to combine the results of a group of studies in a meta-analysis—for example, studies of patients with different types of treatments, different types of comparison groups, or different clinical characteristics of patients studied. There are no simple answers here, and reviewers will need to be explicit about their decisions so that others may judge for themselves whether these choices were clinically sensible. A related but separate issue concerns statistical homogeneity. A test for the statistical homogeneity of studies may be performed to evaluate whether the differences among the results of the studies are greater than by chance alone (Cochrane Handbook, Chapter 8.3).[13] However, the test is not very powerful, and failure to reject the hypothesis of homogeneity is not proof that the studies are homogeneous. If the hypothesis of homogeneity is rejected, or if the reviewers decide, on clinical grounds, that the studies are too heterogeneous to support statistical combinations, then the potential sources of heterogeneity should be examined because the observed differences might be caused by factors other than chance, such as differences in methodologic quality, characteristics of participants, interventions, control groups, or outcomes.[38] If the heterogeneity can be explained, reviewers should separately present the results of the relevant subgroups at issue. The subgroup analyses should be kept to a minimum and should be defined *a priori*, because subgroup analyses can be informative but also misleading.[39,40] If the heterogeneity cannot be explained, reviewers should perform a qualitative analysis (see *Minimum Criteria*).

Qualitative Analysis. Although no empirical evidence exists that supports the methods of qualitative analysis, most of the 21 completed reviews that have been published in the Cochrane Library (issue #3, 2002) included some type of qualitative analysis. The methods used varied, and the consensus among the BRG is that more consistency is needed.

A qualitative analysis is not similar to "vote counting," which involves comparing the number of studies with a positive outcome and a negative outcome. In general, vote counting should be avoided. A qualitative analysis consists of using various levels of evidence (Table 6) regarding the effectiveness of a treatment, taking into account the participants, interventions, controls, outcomes, and methodologic quality of the original studies. Clinical homogeneity should also be a prerequisite for a qualitative analysis, and the same structured approach should be applied to a qualitative analysis as is applied to a quantitative analysis.

Table 6. Levels of Evidence

Strong—consistent findings among multiple high quality RCTs*
Moderate—consistent findings among multiple low quality RCTs and/or CCTs and/or one high quality RCT
Limited—one low quality RCT and/or CCT
Conflicting—inconsistent findings among multiple trials (RCTs and/or CCTs)
No evidence from trials—no RCTs or CCTs
Consistency and quality should be clearly defined *a priori*

* There is consensus among the Editorial Board of the BRG that strong evidence can only be provided by multiple high quality trials that replicate findings of other researchers in other settings.

Clinical Relevance

Further Guidance. The BRG recommends including an assessment of clinical relevance of study results in systematic reviews. The conclusions regarding the effectiveness of the intervention should contain all the important information to enable users to make a decision if the results apply to their population. The clinical relevance of the studies should be independently assessed by at least two reviewers. The following five questions are recommended. They are based on previous guides to critical reading of the medical literature[41,42] and two systematic reviews on back pain[43,44]:

1. Are the patients described in detail so that you can decide whether they are comparable to those that you see in your practice?
2. Are the interventions and treatment settings described well enough so that you can provide the same for your patients?
3. Were all clinically relevant outcomes measured and reported?
4. Is the size of the effect clinically important?
5. Are the likely treatment benefits worth the potential harms?

Reporting of Results and Conclusions

Further Guidance. The BRG noticed that there is wide variation among reviews published in their module of the Cochrane Library (issue #4, 2002) with regard to the style they use to report results and conclusions. This further guidance is intended to increase consistency among future reviews and updates of reviews.

The reporting of results in the text of reviews should contain the following items (Table 7): the intervention, the type of patients, the comparison group, the number of trials (number of patients), results of quantitative analysis (effect size plus confidence interval [CI]), results of qualitative analysis (level of evidence, direction of the effect [more effective, no difference]), the outcome measures, and the timing (short-term or long-term) of the outcome measure.

Table 7. Recommendation for Authors' Conclusions in Systematic Reviews

No difference between index and comparison group(s)
- Quantitative analysis:
  "The pooled analysis of X trials (no. of people) failed to show a difference in short-term and/or long-term outcome Z between the index and comparison group with RR 1.1 (95% CI 0.8 to 1.4) for patients with acute/subacute/chronic back/neck pain with/without neurologic symptoms."
- Qualitative analysis:
  "There is strong/moderate/limited evidence (X trials; no. of people) that there is no difference in short-term and/or long-term outcome Z between the index and comparison group for patients with acute/subacute/chronic back/neck pain with/without neurologic symptoms."

Index is more/less effective than comparison group(s)
- Quantitative analysis:
  "The pooled analysis of X trials (no. of people) shows that the index is more/less effective than the comparison group for patients with acute/subacute/chronic back/neck pain with/without neurologic symptoms on outcome A (short-term follow-up) with RR 4.0 (95% CI 3.0 to 5.0) and outcome B (long-term follow-up) with RR 4.0 (95% CI 3.0 to 5.0)."
- Qualitative analysis:
  "There is strong/moderate/limited evidence (X trials; no. of people) that the index is more/less effective than the comparison group for patients with acute/subacute/chronic back/neck pain with/without neurologic symptoms on outcome A, B, and C (short or long-term follow-up)."

Contradictory findings across trials
- Qualitative analysis:
  There is conflicting evidence (X trials; no. of people) on the effectiveness of intervention A compared to control B for patients with acute/subacute/chronic back/neck pain with/without neurologic symptoms on outcome A, B, and C (short or long-term follow-up).

No evidence
There is no evidence from RCTs or CCTs on the effectiveness of treatment A for patients with acute/subacute/chronic back/neck pain with/without neurologic symptoms.

The comparison group should be explicitly described: treatment B, placebo, no treatment, waiting list controls, or active or inactive treatment.

Examples are:

1. The pooled analysis of 7 trials (1268 people) shows that behavioral treatment is more effective than no treatment for patients with chronic back pain without neurologic symptoms on short-term pain relief with standardized mean difference (SMD) 0.62 (95% CI 0.25 to 0.98) and short-term behavioral outcomes with SMD 0.40 (95% CI 0.10 to 0.70).

2. There is limited evidence (4 trials; 354 people) that there is no difference in short-term pain relief between acupuncture and placebo or sham acupuncture for patients with chronic back pain with or without neurologic symptoms.

■ Discussion

The Editorial Board of the BRG believes that systematic reviews represent one of the key advances in medical science in the past 10 years and offer the real opportunity to lead to changes in medical practice worldwide. Obviously, one of the major challenges for the future is to increase implementation of the results of systematic reviews. Some initiatives have been developed that try to make systematic reviews more easily available for clinicians in daily practice. Recently published clinical guidelines on the management of low back pain have used the evidence from systematic reviews as basis for their recommendations.[45] Also, the BMJ Publishing Group has issued "Clinical Evidence," which is a summary of the current state of knowledge based on Cochrane and other systematic reviews on prevention and treatment of a wide range of clinical conditions (www.clinicalevidence.com). Whether these and other implementation efforts indeed result in a change in clinicians' behavior and in improved patient outcomes remains unclear.

Systematic reviews need to be conducted as carefully as the trials they report and, to achieve full impact, systematic reviews need to meet high methodologic standards. The objective of these method guidelines is to help reviewers to design, conduct, and report reviews of trials in the field of back pain systematically and explicitly. These guidelines are not intended to set a gold standard or to discourage people from doing a systematic review. On the contrary, we encourage people to undertake a systematic review in collaboration with others. For more guidance on systematic reviews of back pain, we recommend referring to the Cochrane Handbook or looking at the Web site of the BRG: www.cochrane.iwh.on.ca. Address: Cochrane Collaboration Back Review Group, Institute for Work & Health, Toronto, Ontario, Canada, M5G 2E9. Telephone: (416) 927-2027; fax: (416) 927-4167.

■ Key Points

- Many reviews of therapeutic interventions for spinal disorders have been published. It is important that these reviews use adequate, systematic methods to avoid bias.
- Previous method guidelines for systematic reviews in the field of spinal disorders were updated.
- These method guidelines include recommendations that are mandatory (minimum criteria) and optional (further guidance) for reviewers conducting reviews within the Cochrane Collaboration Back Review Group.
- Others may find them useful to plan, conduct, or evaluate systematic reviews in the field of spinal disorders.

Acknowledgments

The Editorial Board of the Cochrane Back Review Group: Coeditors: Claire Bombardier, MD, FRCP; Lex M. Bouter, PhD; Other members: Richard A. Deyo, MD, MPH; Paul G. Shekelle, MD, MPH; Rob A. de Bie, PhD; Gordon Waddell, DSc, MD, FRCS; Francis Guillemin, MD, PhD; and James N. Weinstein, DO, MS. Appendix

References

1. Moher D, Cook DJ, Eastwood S, et al. 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.
2. Assendelft WJ, Koes BW, Knipschild PG, et al. The relationship between methodological quality and conclusions in reviews of spinal manipulation. *JAMA* 1995;274:1942–8.

3. Furlan AD, Clarke J, Esmail R, et al. A critical review of reviews on the treatment of chronic low back pain. *Spine* 2001;26:E155–E162.
4. 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.
5. Bombardier C, Esmail R, Nachemson AL. The Cochrane Collaboration Back Review Group for spinal disorders. *Spine* 1997;22:837–40.
6. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;309:1286–91.
7. Minozzi S, Pistotti V, Forni M. Searching for rehabilitation articles on MEDLINE and EMBASE. An example with cross-over design. *Arch Phys Med Rehabil* 2000;81:720–2.
8. Woods D, Trewheellar K. MEDLINE and EMBASE complement each other in literature searches. *BMJ* 1998;316:1166.
9. Suarez-Almazor ME, Belseck E, Homik J, et al. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Control Clin Trials* 2000;21:476–87.
10. Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol* 2002;31:150–3.
11. Dickersin K, Manheimer E, Wieland S, et al. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. *Eval Health Prof* 2002;25:38–64.
12. Avenell A, Handoll HH, Grant AM. Lessons for search strategies from a systematic review, in The Cochrane Library, of nutritional supplementation trials in patients after hip fracture. *Am J Clin Nutr* 2001;73:505–10.
13. Clarke M, Oxman AD. Cochrane Reviewers' Handbook 4.1.5 [updated April 2002]. Oxford: Update Software. Updated quarterly, 2002.
14. Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? *Int J Epidemiol* 2002;31:1–5.
15. Egger M, Zellweger-Zahner T, Schneider M, et al. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;350:326–9.
16. Gregoire 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.
17. Juni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;31:115–23.
18. Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000;53:964–72.
19. Chalmers TC, Celano P, Sacks HS, et al. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358–61.
20. Colditz GA, Miller JN, Mosteller F. How study design affects outcomes in comparisons of therapy. *Medical Stat Med* 1989;8:441–54.
21. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185–90.
22. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408–12.
23. Jadad AR, Moore A, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1–12.
24. Berlin JA. Does blinding of readers affect the results of meta-analyses? *Lancet* 1997;350:185–6.
25. Verhagen AP, de Vet HCW, de Bie RA, et al. Balneotherapy and quality assessment: interobserver reliability of the Maastricht criteria list and the need for blinded quality assessment. *Clin Epidemiol* 1998;51:335–41.
26. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol* 2002;55:783–6.
27. Juni P, Witschi A, Bloch R, et al. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
28. Detsky AS, Naylor CD, O'Rourke K, et al. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992;45:255–65.
29. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987;107:224–33.
30. Verhagen AP, de Vet HCW, de Bie RA, et al. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol* 2001;54:651–4.
31. Normand ST. Tutorial in biostatistics. Meta-analysis: formulating, evaluating, combining, and reporting. *Stat Med* 1999;18:321–59.
32. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. *Stat Med* 1991;10:1665–77.
33. Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101–5.
34. Tang JL, Liu JLY. Misleading funnel plots for detection of bias in meta-analysis. *J Clin Epidemiol* 2000;53:477–84.
35. Sterne JAC, Gavaghan DJ, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;53:1119–29.
36. Engels EA, Schmid CH, Terrin N, et al. Heterogeneity and statistical significance in meta-analysis: an empirical study of 125 meta-analyses. *Stat Med* 2000;19:1707–28.
37. Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol* 1999;150:469–75.
38. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 1995;48:167–71.
39. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992;116:78–84.
40. Yusuf S, Wittes J, Probstfield J, et al. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA* 1991;266:93–8.
41. Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA* 1994;271:59–63.
42. Shekelle PG, Andersson G, Bombardier C, et al. A brief introduction to the critical reading of the clinical literature. *Spine* 1994;19:2028S-31S.
43. Guzman J, Guzman J, Karjalainen K, et al. Multidisciplinary rehabilitation for chronic low back pain: systematic review. *BMJ* 2001;322:1511–6.
44. Karjalainen K, Malmivaara A, van Tulder M, et al. Multidisciplinary biopsychosocial rehabilitation for subacute low back pain in working-age adults: a systematic review within the framework of the Cochrane Collaboration Back Review Group. *Spine* 2001;26:262–9.
45. Koes BW, van Tulder MW, Ostelo R, Burton KA, Waddell G. Clinical guidelines for the management of low back pain in primary care: an international comparison. *Spine* 2001;26:2504–13.

Appendix 1. Search Strategy in MEDLINE

Phase 1. Sensitive Search for randomized controlled trials.[45]

For OVID-MEDLINE (Version 4.0)

01 randomized controlled trial.pt ***02*** controlled clinical trial.pt ***03*** Randomized Controlled Trials/ ***04*** Random Allocation/ ***05*** Double-Blind Method/ ***06*** Single-Blind Method/ ***07*** or/1-6 ***08*** Animal/ not Human/ ***09*** 7 not 8 ***10*** clinical trial.pt ***11*** explode Clinical Trials/ ***12*** (clinic$ adj25 trial$).tw ***13*** ((singl$ or doubl$ or trebl$ or tripl$) adj (mask$ or blind$)).tw ***14*** Placebos/ ***15*** placebo$.tw ***16*** random$.tw ***17*** Research Design/ ***18*** (latin adj square).tw ***19*** or/10-18 ***20*** 19 not 8 ***21*** 20 not 9 ***22*** Comparative Study/ ***23*** explode Evaluation Studies/ ***24*** Follow-Up Studies/ ***25*** Prospective Studies/ ***26*** (control$ or prospective$ or volunteer$).tw ***28*** Cross-Over Studies/ ***28*** or/22-28 ***29*** 28 not 8 ***30*** 29 not (9 or 21) ***31*** 9 or 21 or 30

For Pubmed (http://www.ncbi.nlm.nih.gov/entrez/)

(randomized controlled trial [pt] OR controlled clinical trial [pt] OR randomized controlled trials [mh] OR random allocation [mh] OR double-blind method [mh] OR single-blind method [mh] clinical trial [pt] OR clinical trials [mh] OR ("clinical trial" [tw]) OR ((singl* [tw] OR doubl* [tw] OR trebl*[tw] OR tripl*[tw] AND (mask*[tw] OR blind*[tw]) OR ("latin square"[tw]) OR placebos [mh] OR placebo*[tw] OR random*[tw] OR research design [mh:noexp] OR comparative study [mh] OR evaluation studies [mh] OR follow-up studies [mh] OR prospective studies [mh] OR cross-over studies [mh] OR control* [tw} OR prospective* [tw] OR volunteer* [tw] NOT (animal [mh] NOT human[mh])

Phase 2. Specific terms for spinal disorders

Below is a list of suggested terms that can be used when searching Medline. Reviewers should choose only those pertinent to their topic
When using OVID-Medline, it is important to search the same term as both MeSH (/) and text word (tw). When using Pubmed, the term will result on a search of both MeSH (exploded) and the text word.

- Spine/
  - Intervertebral disk/
  - Lumbar vertebrae/
  - Cervical vertebrae/
- Spinal Diseases/
  - Intervertebral disk displacement/
  - Spinal curvatures/
  - Spinal osteophytosis/
  - Spondylitis/
  - Spondylarthritis/
  - Spondylolisthesis/
  - Spinal Stenosis/
- Anatomy/
  - Body Regions/
    - Back/
    - Neck/
  - Musculoskeletal System/
    - Muscles/
      - Neck muscles/
    - Cartilage/
      - Intervertebral disk/
- Pain/
  - Back Pain/
    - Low-back pain/
  - Neck Pain/
  - Neuralgia/
    - Sciatica/
- Disorders of Environmental Origin
  - Neck Injuries/
    - Whiplash/
  - Back Injuries/
    - Spinal Injuries/

Example: OVID-Medline <1966 to October Week 5 2002>

01 low back pain/	4628
02 low back pain.tw	6127
03 backache.tw	1129
04 lumbago.tw	643
05 or/1-4	9585

Example: Pubmed <December 5th, 2002>

1 LOW BACK PAIN	8046 (equals to ("low back pain" [mh] OR "low back pain" [tw]))
2 BACKACHE [tw]	1119
3 LUMBAGO [tw]	653
4 #1 OR #2 OR #3	9468

Phase 3. Specific search for intervention at issue

All Phases. 1 and 2 and 3

Apendix 2. Search Strategy for EMBASE

Phase 1. Search strategy for randomized controlled trials (adapted from 1997 Method Guidelines) [4]

For OVID-EMBASE (Version 4.0)

01. clinical article/	*09.* single blind procedure/	*17.* compar$.ti,ab.	*25.* ((singl$ or doubl$ or trebl$ or tripl$) adj25 (blind$ or mask$)).ti,ab.	*32.* animal/
02. clinical study/	*10.* crossover procedure/	*18.* control$.ti,ab.	*26.* trial.ti,ab.	*33.* animal experiment/
03. clinical trial/	*11.* placebo/	*19.* cross?over.ti,ab.	*27.* (versus or vs).ti,ab.	*34.* 31 or 32 or 33
04. controlled study/	*12.* or/1-11	*20.* factorial$.ti,ab.	*28.* or/13-28	*35.* 30 and 34
05. randomized controlled trial/	*13.* allocat$.ti,ab.	*21.* follow?up.ti,ab.	*29.* 12 or 28	*36.* 29 not 34
06. major clinical study/	*14.* assign$.ti,ab.	*22.* placebo$.ti,ab.	*30.* human/	*37.* 29 and 35
07. double blind procedure/	*15.* blind$.ti,ab.	*23.* prospectiv$.ti,ab.	*31.* nonhuman/	*38.* 36 or 37
08. multicenter study/	*16.* (clinic$ adj25 (study or trial)).ti,ab.	*24.* random$.ti,ab.		

Phase 2. Specific MeSH terms for spinal disorders

Below is a list of suggested terms that can be used when searching Embase. Reviewers should choose only those pertinent to their topic

- Body Regions/
  - Back/
    - Back Muscles/
    - Spine/
  - Neck/
    - Neck Muscles/
- Spine/
  - Cervical Spine/
  - Thoracic Spine/
  - Thoracolumbar Spine/
  - Lumbar Spine/
  - Lumbosacral Spine/
  - Intervertebral Disk/
  - Vertebra/
  - Vertebra body/
  - Vertebra Canal/
  - Nucleus Pulposus/
  - Odontoid Process/
  - Atlas/
  - Axis/
- Spine Disease/
  - Cervical Spondylosis/
  - Spine Instability/
  - Spondylarthrosis/
  - Spondyloarthropathy/
  - Spondylolisthesis/
  - Spondylolysis/
  - Spondylosis/
  - Vertebral Canal Stenosis/
  - Intervertebral Disk Disease/
    - Diskitis/
    - Intervertebral Disk Degeneration/
  - Intervertebral Disk Hernia/
    - Lumbar Disk Hernia/
  - Kyphosis/
  - Scoliosis/
  - Lordosis/
  - Spine Injury/
  - Spine Malformation/
  - Spine Tumor/
  - Spondylitis/
- Pain/
  - Backache/
    - Low back pain/
  - Neuralgia/
    - Ischialgia/
  - Neck Pain/
- Head and Neck Disease/
  - Head and Neck Injury/
    - Neck Injury/
      - Whiplash Injury/

Example: OVID-Embase <1980 to 2002 Week 47>

01 low back pain/	9907
02 low back pain.tw	6097
03 backache	629
04 lumbago.tw	395
05 or/1-4	11691

Phase 3. Specific search for intervention at issue

All Phases. 1 and 2 and 3