Complementary Therapies in Medicine (2005) **13**, 54—60

ELSEVIER

Complementary Therapies in Medicine

www.elsevierhealth.com/journals/ctim

Systematic literature reviews

Adrian White[a,*], **Katja Schmidt**[b,1]

[a] *Department of General Practice and Primary Care, Peninsula Medical School, Universities of Exeter and Plymouth, Institute of Health and Social Care, ITTC Building, Tamar Science Park, Plymouth PL6 8BX, UK*
[b] *Complementary Medicine, Peninsula Medical School, Universities of Exeter and Plymouth, Institute of Health and Social Care, 25 Victoria Park Road, Exeter EX2 4NT, UK*

Summary Systematic reviews retrieve, appraise and summarise all the available evidence on a specific health question. They are designed to reduce the effect of the reviewers' own bias, and a full protocol should be written to define and guide the process. The appropriate resources should be in place before undertaking a review. The steps of the review are: frame the question and choose appropriate methods; identify relevant work; extract relevant data on outcomes and quality; summarise the evidence; and, interpret the evidence. Reviews that combine valid, homogeneous studies of treatments that are relevant to health care, in patients who are typical, can provide good evidence to guide health care decisions.

Introduction

The existing literature on any subject can be reviewed for different reasons, ranging from providing the historical background on a particular style of reflexology to finding out how safe or effective a particular treatment is. A *systematic* review retrieves, appraises and summarises all the available evidence on a specific (health) question and then attempts to reconcile and interpret it. A systematic review may include a *meta-analysis*, which is a mathematical technique for pooling and combining the data. This article focuses on how to conduct a systematic review of the evidence of a therapy's effectiveness; it does not cover the subject of meta-analysis.

In the days before reviews became systematic, the writer was free to pick and choose the papers that supported his or her viewpoint. This is clearly a biased approach, and contrary to the principle of evidence based medicine. Using a review that has any kind of bias can lead to poor decisions in health care.[1] The systematic review was specifically developed to try to reduce the influence of the reviewer's own bias. It does this by deciding in advance what evidence to use and how to use it, so these decisions are not influenced by the evidence itself.[2] The Methods section of a systematic review should be as rigorous as that of any other piece of research.

Because a systematic review is a piece of research, make sure that you have sufficient time and other resources before you start. You will need

* Corresponding author. +44 1752 764448.
E-mail addresses: adrian.white@pms.ac.uk (A. White), Katja.Schmidt@pms.ac.uk (K. Schmidt).
[1] Tel: +44 1392 424839; fax: +44 1392 427562.

0965-2299/$ — see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ctim.2004.12.003

to be able to access databases and retrieve original articles, you will need various detailed reference sources on the methodology[3,4] (see also Other Resources at the end of this paper) and you will need co-reviewers, including ideally an expert on reviews, an expert in the field of health care being reviewed, a statistician if you are contemplating anything more than the most simple kind of meta-analysis - and ideally a consumer (patient). The QUOROM checklist for what needs to be reported is a vital tool.[5] Finally, you should make sure that another review group has not recently covered precisely the topic you have chosen.

Resources check list before conducting systematic literature review:

- Access to relevant databases
- Access to original articles
- Availability of translators for foreign language articles
- Collaboration with experienced reviewer
- Collaboration with content expert
- Collaboration with user

When you are conducting the review, you need to proceed in careful steps as shown here. However, you must plan all the steps from the outset: in other words, you have to understand the whole process in advance, in detail, and write a full protocol. Most examples we use are drawn from three reviews chosen to illustrate a range of techniques and problems, and we advise the reader to consult the originals for more detail.[6–8]

Step 1: Frame the question and choose appropriate methods

The introduction

The Introduction to your review needs to give the epidemiological and clinical background to the condition you are studying as well as the rationale for using the treatment. It has to convince the reader that your review is essential and your questions are exactly the ones they want to know the answers to.

From the work involved in writing the Introduction, by looking at previous reviews in this area, and by discussing the issues with your co-authors, you should evolve a clear and unambiguous research question. Express the essence of this question as an overall 'Aim', but as you develop the precise research questions so you will also be able to refine your ideas into a set of precise objectives which will determine exactly how you are going to conduct the review.

Aim

An example of a simple aim would be to determine the effectiveness of hypericum as a treatment for depression. In our example, this became refined into '... to investigate whether extracts of hypericum are more effective than placebo in the treatment of depression, are as effective as standard antidepressive treatment ...'[6]

Don't forget that you may want to take the opportunity to include secondary aims, such as 'to identify key features of the acupuncture treatment which are associated with positive results'.[9]

Thus, you will be able to define a series of precise research questions. Each of them will consist of four components:

- a *health condition*, for example moderate depression
- in a *population*. For example: adults (>18 years) with acute (<4 weeks), sub-acute (4–12 weeks) or chronic (>12 weeks) non-specific LBP; LBP is defined as pain localized from the costal margin or 12th rib to the inferior gluteal fold; non-specific indicates that no specific cause is detectable, such as infection, neoplasm, metastasis, osteoporosis, rheumatoid arthritis, fracture, inflammatory process or radicular syndrome.[8]
- the precise *intervention* you are studying for that health condition (e.g. massage in this review is defined as soft tissue manipulation using the hands or a mechanical device...),[8]
- one or several *outcomes*; these might still be broad at this stage, for example: pain, return to work or work status, Subjective change of symptoms, or Functional status expressed by validated instruments.[8] The 'validated instruments' are not pre-defined in this case.

This may seem a lot of fuss, but the discipline is well worthwhile, in order to define what you will do—and to show you where bias may enter your review later on, if you modify or further refine your questions.

Protocol

The next thing to do is to expand the research question into a full protocol, which will form the Methods section of the eventual paper. It will also be a source of reference for the reviewers to guide you through the review, so it is well worth spending time and effort to get it right. The protocol will describe how you will decide which studies to include, how you will assess their quality, and how you will summarise and interpret the results (i.e. Steps 2–4

below). In principle, try to establish the methods as fully as possible in advance, to reduce bias. In practice, you may need to refine your methods section when you see the papers themselves, but each decision can affect the outcome of the review, so be careful.[10]

Step 2: Identify relevant work

Search the databases

Decide beforehand which electronic databases you will search. AMED, CancerLit, CINAHL, CISCOM, Cochrane CENTRAL, Embase and Medline are standard but there may be some specific to your subject (e.g. CAMPAIN[7]). Some databases have free access, for others you will need a password and username or may have to pay a fee. Searching the databases successfully is a subject on its own, and it is worth piloting your searches to improve the rate of retrieval. For a systematic review, you need a search that is very sensitive rather than too specific: this will reduce the chance of missing studies but at the cost of long lists of abstracts to read through.

Try different search terms (text words) (e.g. "massage" AND "back pain") and keyword combinations to find the most sensitive. See whether you should set any limits for searching each database (e.g. in Medline you can limit the search to 'clinical trials'). However, this relies on the database coding studies accurately which is not always the case. It is reasonable to use 'clinical trial' as a search term but do not exclude all others purely because they have not been coded by Medline as such. Also, remember to retrieve the abstract not just the reference, and record which database you found each article, as this is useful to include in the report.

State explicitly where else you will search, for instance, how you will try retrieving non-indexed RCTs, e.g. from the reference lists of retrieved papers, hand-searches of journals, and contacting other resources relevant to the intervention such as manufacturers, research-active acupuncturists, or professional associations of massage therapists.

Retrieve original reports

The next step is for two authors carefully to review the abstracts of all search results and to decide whether to retrieve the full copy. Even if one reviewer is not absolutely sure that the study should be rejected, then you should obtain the full copy, in order not to overlook important evidence. Inevitably this means that many copies that you suspected are irrelevant indeed turn out to be excluded. This is the most chaotic part of the review so it is well worth keeping a master list, on a computer spreadsheet, of all the references that might be considered, then fill in the columns as you make decisions which to exclude and for what reason. It is advisable to always keep the papers or abstracts that you don't want to include in the review but that might be useful for the discussion.

Decide which studies to include

Inclusion/exclusion criteria ideally should be established in the protocol development stage and should inform the development of a checklist for deciding which studies to include and which to exclude. At least two authors should then independently judge which studies are going to be included or excluded based on these pre-established criteria. This can be done by each reviewer independently establishing which studies fit the criteria for inclusion followed by a meeting to discuss any discrepancies between reviewers. These should be resolved by discussion and reference to the original report. Inclusion criteria should be related to the appropriateness of the study not its availability. Thus, one should not seek to include, for example, non-randomised studies merely because very few randomised studies exist. Even in cases where no studies fit the inclusion criteria (commonly known as an empty shell review), inclusion criteria should not be altered.

Inclusion/exclusion criteria are closely linked to the research question you specify in your protocol. For example, they may be based on:

- Types of study design (e.g. randomised[7] or controlled)
- Types of participant (e.g. patients with pain longer than three months)
- Types of intervention precisely defined (e.g. used needles)
- Types of control (e.g. had a comparison group for which a between groups analysis was presented)
- Types of outcome (e.g. had a measurement for pain)

Specify that you will only include reports that contain data, as you will probably want to exclude brief conference abstracts; decide whether to include unpublished studies (some reviewers prefer

not to because they have not been peer-reviewed) and whether to restrict your papers to particular languages. It is best, naturally, to include all languages, but retrieving reports of foreign language articles and getting them translated need considerable resources. Any language restriction is a limitation: for example, if the review of St. Johns' wort in 1994 had only included reports in English, it would not have contained a single study.[6] You may also need to contact the authors of a paper to retrieve any vital data not included. This may not be possible but it cannot be assumed, for example, that adequate randomisation was used, unless the methods are explained.

Step 3: Extract relevant data on outcomes and quality

Decide beforehand what data are to be extracted, and how you will carry out the data extraction. Then you can set up the process, using either paper data extraction sheets, or an electronic data extraction spreadsheet. It is worth extracting all the data you may conceivably want, even if not all is included in the report: it is very time-consuming to have to go through the papers twice. The data extraction should be done independently by at least two authors and the results compared, as they will often interpret the results in a different way. Also decide in advance whether to use data that are presented only in graphic form, and have to be measured with a ruler and converted into units.

The data that can be extracted from each study are likely to include:

- names of the authors
- study design and setting
- an assessment of the quality, see below
- numbers and characteristics of the subjects and numbers in each arm (and the numbers analysed, if you are intending to carry out a meta-analysis)
- the condition treated, with diagnostic criteria, severity and duration
- details of the intervention, and/or control, with dosage and duration of treatment, and any other care given including co-interventions that patients were given or allowed to use
- the primary/secondary outcome measures used at various time-points
- dropouts and withdrawals (was intention to treat analysis done?)
- success of any blinding
- the changes on the various outcome measures, including follow-up
- author's interpretation of the results

The results can be extracted as changes or differences in means, or numbers of treatment responders,[10] for combining in a meta-analysis. They can also be extracted as significant differences between treatment and controls (e.g. *p*-values or odds ratios); or in categories such as positive, neutral or negative.[7] If you keep a space for 'comments', be aware that this is open to your bias, so that any information you put there should not be used in synthesising the data.

You may need to decide whether to contact authors for missing/additional data. One frequent problem arises when authors do not report standard deviations. There are methods of imputing SDs from the *p*-value, but you will need to involve a statistician. Another decision you might have to make is what to do with studies with a crossover design. If you believe there is significant likelihood of carry-over effects, you may decide to take the results of the first arm only.

Assess the quality

There are several quantitative methods for assessing the methodological quality of studies,[8,11–13] and you must decide which is most appropriate for your purposes. There is some doubt about how useful these 'quality scores' really are, since different methods do not produce consistent results,[14] and it is important to evaluate the studies critically using common sense. The simplest method introduced by Jadad assesses internal validity by the presence of a report of randomisation, blinding and dropouts.[11] This method has been validated, but it does not include some criteria that others consider are crucial for good quality, and operates more as an indicator of poor quality than a test of good quality. Full quality assessment usually involves the following headings as a minimum:

- patient selection (eligibility criteria, baseline comparability)
- patient allocation (randomisation, allocation concealment)
- blinding (patient, care-giver, observer, statistician, clinician)
- interventions (co-interventions, compliance)
- outcome measurement (adverse events)
- statistics (intention-to-treat analysis)

Note that the concept of quality, although only discussed here as criteria for Step 3, should run throughout the review. You may want to include

quality criteria in your research question, or to set a minimum quality for inclusion in Step 2, and you will use the quality in summarising (Step 4) and interpreting (Step 5) the evidence.

Assess the validity

For assessing the validity of a clinical trial one needs to assess (1) whether the study sample was relevant, (2) whether the intervention was appropriate and (3) whether the outcome measure was suitable.[15] Recently, Smith et al.[12] created the Oxford Pain Validity Scale (OPVS), specifically for pain conditions, which can also be adapted and modified for other conditions.

According to the authors full validity assessment usually involves:

- Blinding

(a) Was the trial convincingly double-blind?
(b) Was the trial convincingly single-blind or unconvincingly double-blind?
(c) Was the trial either not blind or the blinding is unclear?

- Size of trial group

(a) Was the group size $\geq$40?
(b) Was the group size 30—39?
(c) Was the group size 20—29?
(d) Was the group size 10—19?

- Outcomes

Look at pre hoc list of most desirable outcomes relevant to the review question:

(a) Did the trial include results for at least one pre-hoc desirable outcome, and use the outcome appropriately?
(b) There were no results for any of the pre-hoc desirable outcomes, or a pre-hoc desirable outcome was used inappropriately.

- Baseline and internal sensitivity

Look at the baseline levels for the outcomes relevant to the review question:

(a) For all treatment group baseline levels were sufficient for the trialist to be able to measure a change following an intervention. Alternatively, did the trial demonstrate internal sensitivity?
(b) For all treatment groups baseline levels were insufficient to be able to measure a change following the intervention or baseline levels could not be assessed or internal sensitivity was not demonstrated.

- Data analysis

(i) Definition of outcomes

(a) Did the paper define the relevant outcomes clearly, including where relevant exactly what 'improved', 'successful treatment', etc. represented?
(b) The paper failed to define the outcomes clearly.

(ii) Data presentation: location and dispersion

(a) Did the paper present either mean data with standard deviations or dichotomous outcomes or median with range or sufficient data to enable extraction of any of these?
(b) The paper presented none of the above.

(iii) Statistical testing

(a) Did the trialist choose an appropriate statistical test with correction for multiple tests where relevant?
(b) Inappropriate statistical tests were chosen and/or multiple testing was carried out but with no correction or no statistics were carried out.

(iv) Handling of dropouts

(a) The dropout rate was either $\leq$10% or was >10% and includes an intention-to-treat analysis in which dropouts were included appr -opriately.
(b) The dropout rate was >10% and dropouts were not included in the analysis or it is not possible to calculate a dropout rate from data presented in paper.

Step 4: Summarise the evidence

Make a table of the study characteristics as the first part of the Results.[6] The accompanying text may briefly describe the individual studies, drawing attention to similarities and differences between the studies but not repeating the data that are in the tables. Then you need to assess, though not yet judge, the differences between study results, and how this may be due to clinical differences, or 'heterogeneity', between the studies, e.g. single hypericum preparations compared with combined preparations.[6]

Decide what is the most appropriate way to combine the results of the studies. The options are usually either to combine them in narrative form only, or using the best evidence synthesis,[8] or the so-called vote-count, or in a meta-analysis.

You may also want to look at subgroups—of studies, or of patients. For example, you might want to combine just the studies of one particular type of acupuncture, or just the results in female patients. Decide beforehand — or you will be accused of bias — what subgroup analyses you will do to answer your research questions: for example, if when you look at the results you realise for the first time that the treatment might be more effective for men with acute pain, this is a post hoc analysis and must be interpreted very cautiously. But you could specify in advance that you will do a subgroup analysis to examine the effect in different patients, e.g. acute, subacute and chronic low back pain.[8] You may wish to perform a sensitivity analysis, for example, to test whether the overall result is affected by study quality e.g. study outcome versus quality criterion.[7]

If you decided to carry out a meta-analysis, you must consider how to identify and deal with statistical heterogeneity (i.e. results of trials differ by more than can be explained by chance). This is too complex to be covered in detail here, but usually you need to try and explain the heterogeneity of the studies and then analyse using a random effects model.

Step 5: Interpret the evidence

Finally, in the Discussion, you should restate the findings, summarise the strengths and weaknesses both of the original data and your own review, interpret the evidence with a discussion of the mechanisms if appropriate, and set out the implications for practice and research.

A review will be strong if it includes several trials that are clinically homogeneous, are of high validity, and have consistent findings. You should assess the risk of publication bias, i.e. the tendency of authors to write (and editors to accept) only those trials that have positive results. Limitations in the weight of the evidence presented (i.e. study design, quality, and volume) reduce the impact of the findings.

The results are likely to be generalisable to clinical practice if they use interventions that are widely available, in patients who are typical of normal practice. Keep in mind the variation in settings of the trials, and the compliance with the treatment. Discuss the issues of cost and safety of the therapy compared with other available therapies. You should discuss each of these issues taking into account previous research that has been carried out. Additionally, try to address what is the trade off between benefits on the one hand, and costs and safety on the other.

Not every review can give the answers to all its questions, and you will raise new questions during the process of the review. Address those and make suggestions for future directions of research.

Conclusion

Reviews are complex and time-consuming. They require considerable resources, and importantly, knowledgeable co-authors. Reviews are sometimes accepted as the definitive statement of the evidence. Often the abstract is quoted by people who have not read the full text and who do not really understand the issues. Therefore, the reviewer has great responsibility to ensure that any limitations on the conclusion are very clearly indicated, particularly if the quality of the primary studies is not good. A well-conducted review of good primary data is both satisfying to the authors and beneficial to patient care.

Other resources

Websites:

www.systematicreviews.com Egger, Smith O'Rourke

The Cochrane Handbook, http://www.cochrane.org/

CRD Report 4: Undertaking Systematic Reviews of Research on Effectiveness. At The York Centre for Reviews and Dissemination, http://www.york.ac.uk/inst/crd/

References

1. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA* 1992;**268**(2): 240—8.
2. Mulrow CD. Rationale for systematic reviews. *BMJ* 1994;**309**(6954):597—9.
3. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized controlled trials in health care from the Potsdam consultation on meta-analysis. *J Clin Epidemiol* 1995;**48**:167—71.
4. NHS Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness* (CRD report no. 4). York: CRD; 2001.
5. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM

statement. Quality of reporting of meta-analyses. *Lancet* 1999;**354**(9193):1896—900.
6. Linde K, Ramirez G, Mulrow CD, Pauls A, Weidenhammer W, Melchart D. St John's wort for depression—an overview and meta-analysis of randomised clinical trials. *BMJ* 1996;**313**:253—8.
7. Ezzo J, Berman B, Hadhazy V, Jadad AR, Lao L, Singh BB. Is acupuncture effective for the treatment of chronic pain? A systematic review. *Pain* 2000;**86**:217—25.
8. Furlan AD, Brosseau L, Imamura M, Irvin E. Massage for low back pain (Cochrane review). *The Cochrane library*, Issue 2. Oxford: Update Software; 2003.
9. Berman B, Ezzo J, Hadhazy V, Swyers JP. Is acupuncture effective in the treatment of fibromyalgia? *J Fam Pract* 1999;**48**:213—8.
10. Linde K, Willich SN. How objective are systematic reviews? Differences between reviews on complementary medicine. *J R Soc Med* 2003;**96**(1):17—22.
11. Jadad AR, Moore RA, Carrol D, Jenkinson C, Reynolds DJ, Gavaghan D. Assessing the quality of reports of randomised clinical trials: is blinding necessary? *Controlled Clin Trials* 1996;**17**:1—12.
12. Smith LA, Oldman AD, McQuay HJ, Moore RA. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. *Pain* 2000;**86**(1/2):119—32.
13. Koes BW, Assendelft JJ, van der Heijden GJ, Bouter LM, Knipschild P. Spinal manipulation and mobilisation for back and neck pain: a blinded review. *BMJ* 1991;**303**:1298—303.
14. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;**282**(11):1054—60.
15. Ernst E. "Flower remedies": a systematic review of the clinical evidence. *Wien Klin Wochenschr* 2002;**114**(23/24):963—6.