

# Reading systematic reviews

**BACKGROUND** While systematic reviews should be of assistance in decision making, currently many reviews are unreliable, poorly reported or not clinically useful. It is therefore important that they are read critically.

**OBJECTIVES** The objective is to provide a logical way to appraise systematic reviews. Particular objectives are to demonstrate how to use a validated general guide for appraising systematic reviews; amend the guide so that it can be used to appraise systematic reviews of controlled trials, prognostic studies and studies evaluating diagnostic tests; choose between discordant reviews and assess the usefulness of good systematic reviews.

**DISCUSSION** Guidelines for undertaking and reporting systematic reviews should eventually improve their quality and usefulness. In the meantime, reviews must be appraised critically, particularly those outside the Cochrane Library.

**Hedley Peach**, BSc, MB, BCh, PhD, FFPHM, is a Professorial Fellow of the University of Melbourne and an honorary visiting consultant in public health to Ballarat Health Services, Victoria.

When faced with inconsistent findings from multiple research studies, we all know how difficult it is to answer a particular question we might have, for example about the treatment of asthma or chronic pain, in a definitive way. Traditional (narrative) review articles are rarely of much value, often being too broad, too selective in the studies cited and not rigorous enough to provide a reliable answer to our specific clinical question.<sup>1</sup> It is here that systematic reviews of the literature on asthma and chronic pain should come into their own. By appraising and then summarising all the research evidence pertaining to a particular question in a systematic way, such reviews have the potential to be a reliable and useful source of information.<sup>2</sup>

However, systematic reviews are only as good as the primary research studies located, how well they are appraised and combined, and as valuable as their usefulness in clinical practice. As with primary research studies, a substantial proportion

of the systematic reviews undertaken to date have been found to be unreliable, poorly reported or not clinically useful. For example, 90% of systematic reviews on the treatment of asthma and chronic pain have been found to have methodological flaws that could limit their validity and 82% of reviews evaluating diagnostic tests omitted certain information which compromised their clinical usefulness.<sup>2-4</sup> It is therefore important that systematic reviews are read critically to distinguish the good reviews from the poor ones.

## Objectives

The objective of this article is to demonstrate how to appraise systematic reviews in a logical way. The article's particular objectives are:

- to show how to use a validated general guide for appraising systematic reviews
- amend the guide so that it can be used to appraise systematic reviews of controlled trials, prognostic studies and studies evaluating diagnostic tests

- choose between discordant reviews and
- assess the usefulness of good systematic reviews to clinical practice.

## General guides

A number of general guides to appraising systematic reviews have been published.<sup>5-9</sup> Although there is some overlap, there are marked differences between them. Only one, Oxman and Guyatt's, has been proven to distinguish between systematic reviews of good and poor quality, including reviews from a range of specialties and on a range of topics of clinical interest, such as therapy, diagnosis, prognosis and prevention.<sup>5</sup> The guide has also been proven to perform equally well in the hands of different individuals, including people without a detailed knowledge of the review topic or a background in research methodology.<sup>10</sup>

Table 1 shows the aspects of a systematic review covered by Oxman and Guyatt's guide.<sup>3</sup> The questions in the

**Table 1. Recommended guide to appraising systematic reviews<sup>3</sup> (modified)**

1.	Were the methods used to find primary research studies reported?	<b>No/Partially/Yes</b>
2.	Was the search comprehensive?	<b>No/Can't tell/Yes</b>
3.	Were the criteria used for deciding which studies to include in the review reported?	<b>No/Partially/Yes</b>
4.	Was bias in the selection of studies avoided?	<b>No/Can't tell/Yes</b>
5.	Were the criteria used for assessing validity of the studies reported?	<b>No/Partially/Yes</b>
6.	Was the validity of all studies assessed using appropriate criteria?	<b>No/Can't tell/Yes</b>
7.	Were the methods used to combine the findings of the studies to reach a conclusion reported?	<b>No/Partially/Yes</b>
8.	Were the methods appropriate?	<b>No/Can't tell/Yes</b>
9.	Was the conclusion supported by the data and/or analysis?	<b>No/Partially/Yes</b>
10.	What was the overall scientific quality of the review?	<b>Extensive flaws/Major flaws/ Minor flaws/Minimal flaws</b>

guide have been modified slightly to improve their clarity. One should use the answers to the first nine questions to judge whether a systematic review has

'minimal', 'minor', 'major' or 'extensive' flaws. A review with only 'minimal' flaws would receive a score of 7 while reviews with 'minor', 'major' or 'extensive' flaws would receive scores of 5, 3 and 1 respectively. Reviews which fall between the four anchors (ie. between having 'minimal' and 'minor', 'minor' and 'major', or 'major' and 'extensive' flaws) should receive scores of 6, 4 or 2. Some readers may prefer to score a systematic review by following the steps in Table 2. These steps have been developed to improve the clarity of the published instructions.<sup>3</sup> Systematic reviews scoring **4** are unreliable.

Oxman and Guyatt's guide requires judgment as to whether the research for primary studies was comprehensive and whether the criteria used for assessing the validity of the studies and the methods used for combining them were appropriate. A comprehensive search should include at least bibliographic databases

and reference lists of retrieved articles.<sup>11</sup> So as not to label more primary studies as unreliable than is necessary, the focus should be on aspects of the studies that are known or strongly believed to affect their validity (see below). The methods used in combining primary studies to produce an overall summary statistic are complex. Clinical acumen should be used to decide whether the decisions to combine data and what to combine were correct.<sup>12</sup>

### Specialised questions

There are particular issues with systematic reviews of controlled trials, prognostic studies and studies evaluating diagnostic tests.<sup>4,13-15</sup> Tables 3-5 contain questions designed to help identify the aspects which are known or, in the case of prognostic studies, are strongly believed to affect their validity.<sup>13,14,16</sup> (The response in parenthesis indicates bias).

Table 4 also contains two questions (8A and 8B) designed to identify particular problems with combining data from primary prognostic studies.<sup>14</sup> There are special methods for combining primary studies evaluating diagnostic tests and one can construct a simple algorithm from the material in reference 4 to help judge whether a review has used the appropriate method. In addition, it is especially difficult to identify primary prognostic studies and studies evaluating diagnostic tests.<sup>4,14,15</sup> Therefore particular attention should be given to the first two questions of Oxman and Guyatt's guide when reading systematic reviews of these types of studies.

### Discordant reviews

As the number of systematic reviews has increased, so has the number of reviews addressing the same question while coming to different conclusions.<sup>17</sup> When discordance leads to contradictory clinical decisions, it is important the reasons for the discordance are explored and the most appropriate review chosen. This needs to be done in a logical way, for example, by following the steps in Table 6. These steps

**Table 2. Scoring of systematic reviews<sup>3</sup> (modified)**

<b>Step 1</b>	Is the 'No' option used for one or more of questions 2, 4, 6 and 8? <b>No</b> — Go to step 3 <b>Yes</b> — Go to step 2
<b>Step 2</b>	How often is the 'No' option used for questions 2, 4, 6 and 8? <b>Four times</b> — review scores 1 <b>Two/three times</b> — review scores 2 <b>Once only</b> — review scores 3
<b>Step 3</b>	Is the 'Can't tell' option used for one or more of questions 2, 4, 6 and 8? <b>Yes</b> — review scores 4 <b>No</b> — review scores 7

**Table 3. Specialised questions for systematic reviews of controlled trials<sup>13</sup> (modified)**

6A. Could patients and investigators enrolling patients have foreseen the treatment to be allocated?	(Yes)
6B. Could the allocation of treatment have been related to prognosis?	(Yes)
6C. Were both the patients and care providers blind to the treatment allocated?	(No)
6D. Were those assessing outcomes blinded?	(No)
6E. Were deviations from protocol and loss to follow up reported on?	(No)

**Table 4. Specialised questions for systematic reviews of prognostic studies<sup>14</sup> (modified)**

6A. Were the subjects well defined, representative of the population of interest and assembled at a common point in the disease?	(No)
6B. Were subjects followed up for a sufficiently long period of time and completely?	(No)
6C. Was the outcome objective, free of bias and known for a high proportion of subjects? (No)	
6D. Was the prognostic variable well defined, precisely measured and available for a high proportion of subjects?	(No)
6E. Was any treatment standardised or determined by randomisation?	(No)
8A. Were any continuous predictor variables analysed appropriately?	(No)
8B. Were other important predictor variables accounted for?	(No)

**Table 5. Specialised questions for systematic reviews evaluating diagnostic tests<sup>16</sup> (modified)**

6A. Was the study a case control or cohort study?	(Case control)
6B. Was the study population adequately described?	(No)
6C. Was the test under evaluation adequately described?	(No)
6D. Were details of the cut-off for the reference test reported?	(No)
6E. Were different reference tests used for positive and negative results of the test under evaluation?	(Yes)
6F. Was the reference test interpreted with knowledge of the result of the test under evaluation?	(Yes)

**Table 6. Distinguishing between discordant reviews<sup>17</sup> (modified)**

<b>Step 1</b>	Do the reviews lead you to make different decisions? <b>No</b> – Stop <b>Yes</b> – Go to step 2
<b>Step 2</b>	Do all reviews score >4? <b>No</b> – Select reviews scoring >4 and go to step 3 <b>Yes</b> – Go to step 3
<b>Step 3</b>	Do the reviews address the same question? <b>No</b> – Select the review most relevant to the question you are asking <b>Yes</b> – Go to step 4
<b>Step 4</b>	Do all reviews include the same primary studies? <b>No</b> – Go to step 5 <b>Yes</b> – Go to step 6
<b>Step 5</b>	Have all reviews identified and selected the primary studies in the same way? <b>No</b> – Select the review with the least biased approach <b>Yes</b> – Select the review that has assessed and taken account of publication bias and the quality of the primary studies
<b>Step 6</b>	Are the reviews of the same quality? <b>No</b> – Select the review of the highest quality <b>Yes</b> – Select the review with data most relevant to your question and a sensible choice between a meta-analysis or qualitative approach

which are applicable to a broad range of systematic reviews were developed from a published algorithm devised to interpret discordant meta-analyses of intervention studies.<sup>17</sup>

### Questions about usefulness

The type of questions which were designed to help decide whether the results of primary controlled trials, prognostic studies and studies evaluating diagnostic tests would be useful in clinical practice (Table 7, where the response in parenthesis favours usefulness of a study), will also help in assessing the usefulness of good systematic reviews in these areas to your practice.<sup>18-20</sup>

### Comments

One should not be discouraged from reading systematic reviews by the currently high proportion of unreliable or badly reported reviews. The development of guidelines for undertaking and reporting systematic reviews, together with similar

**Table 7. Assessing the usefulness of systematic reviews to clinical practice<sup>18-20</sup> (modified)**

• Can the results be generalised to my setting and patients? Were the results consistent across many studies covering a diverse range of settings and patients?	(Yes)
• Were all clinically important outcomes considered?	(Yes)
• Does a diagnostic test add information beyond that otherwise available to me?	(Yes)
• Will the review results lead to a change in my management that is beneficial to my patients?	(Yes)
• Are the results useful for reassuring or counselling my patients or their relatives?	(Yes)
• Are the benefits of using an intervention or providing my patients with information about their prognosis or the result of a diagnostic test worth the harm and costs?	(Yes)

guidelines for primary studies, will eventually increase the proportion of good useful reviews. However, it is likely that systematic reviews will need to be assessed carefully for some years yet, particularly reviews will need to be scrutinised outside the Cochrane Library which are likely to be of poorer quality.<sup>3</sup> Systematic reviews held by the Cochrane Library are produced using a rigorous standardised methodology and are peer reviewed. This ensures their findings are valid and free from bias in the publication and selection of studies. They are also regularly updated and corrected. However, the Library's focus is on controlled trials so that it contains few reviews of prognostic and diagnostic studies.

**References**

1. Cook D J, Mulrow C D, Haynes R B. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997; 126:376-380.
2. Jadad A R, Moher M, Browman G P, et al. Systematic reviews and meta-analyses on treatment of asthma : critical evaluation. *Br Med J* 2000; 320:537-540.
3. Jadad A R, McQuay H J. Meta-analyses to evaluate analgesic interventions: a systematic qualitative review of their methodology. *J Clin Epidemiol* 1996; 49:235-243.
4. Irwig L, Tosteson A N A, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994; 120:667-676.
5. Oxman A D, Guyatt G H. Validation of an

index of the quality of review articles. *J Clin Epidemiol* 1991; 44:1271-1278.

6. Greenhalgh T. Papers that summarise other papers (systematic reviews and meta-analysis). *Br Med J* 1997; 315:672-675.
7. NH&MRC. How to use the evidence: assessment and application of scientific evidence. Canberra: NH&MRC, 2000; 16.
8. Centre For Reviews And Dissemination. What are the criteria for the inclusion of reviews in DARE? York: University of York, 2001.
9. Oxman A D, Cook D J, Guyatt G H. How to use an overview. *JAMA* 1994; 272:1367-1371.
10. Oxman A D, Guyatt G H, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol* 1991; 44:91-98.
11. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. *Ann Intern Med* 1997; 127:380-387.
12. Lau J, Ioannidis J P A, Schmid C H. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997; 127:820-826.
13. Jüni P, Altman D G, Egger M. Assessing the quality of controlled clinical trials. *Br Med J* 2001; 323:42-46.
14. Altman D G. Systematic reviews of evaluations of prognostic variables. *Br Med J* 2001; 323:224-228.
15. Deeks J J. Systematic reviews of evaluations of diagnostic and screening tests. *Br Med J* 2001; 323:157-162.
16. Lijmer J G, Mol B W, Heisterkamp S, et al. Empirical evidence of design: related bias in studies of diagnostic tests. *JAMA* 1999; 282:1061-1066.
17. Jadad A R, Cook D J, Browman G P. A guide to interpreting discordant systematic reviews. *Can Med Assoc J* 1997; 150:1411-1416.

18. Dans A L, Dans L F, Guyatt G H, Richardson S. How to decide on the applicability of clinical trial results to your patient. *JAMA* 1998; 279:545-549.
19. Laupacis A, Wells G, Richardson W S, Tugwell P. How to use an article about prognosis. *JAMA* 1994; 272:234-237.
20. Jaeschke R, Guyatt G H, Sackett D L. How to use an article about a diagnostic test. *JAMA* 1994; 271:703-707.

AFP

**SUMMARY OF IMPORTANT POINTS**

- Systematic reviews appraise and summarise all the research evidence pertaining to a particular question in a systematic way.
- A substantial proportion of reviews undertaken to date have been found to be unreliable, poorly reported or not clinically useful.
- Systematic reviews must be read critically to distinguish good from poor reviews.
- One of the published guides to appraising reviews has been proven to distinguish between good and poor reviews in the hands of people without a research background.
- This general guide can be supplemented with more specialised questions to identify flaws affecting reliability of systematic reviews of controlled trials, prognostic studies and studies evaluating diagnostic tests.
- The number of reviews addressing the same question but coming to different conclusions is increasing.
- An algorithm can be used to explore reasons for the discordance and choose the most appropriate review.
- Guidelines for undertaking and reporting systematic reviews will eventually improve their quality and usefulness.

**CORRESPONDENCE**

Professor Hedley Peach  
 The University of Melbourne  
 Ballarat Health Services Base Hospital  
 PO Box 577  
 Ballarat, Vic 3353