Reviewer's report

Title: Systematic reviews of adverse effects: framework for a structured approach

Version: 3 Date: 20 February 2007

Reviewer: Barnaby Reeves

Reviewer's report:

General

I would like to reiterate at the outset that I am, in principle, very supportive of the authors aims in this paper. I also think it is a very important topic. However, despite the revisions, I am not persuaded that the paper is anything like as helpful as it could be to prospective reviewers.

The authors appear to imply that changing the emphasis of the paper from a guideline, or guidance, to a framework of a structured approach deals at a stroke with the comments that I made on the previous version. Unfortunately, I do not think that this is the case. Rather, the revisions seem fairly superficial and many of the ambiguities in the precise wording of the text remain.

I am afraid that I don’t see it as my responsibility to identify all of the problems in the paper. Therefore, the following list of comments are only meant to be illustrative of the kinds of problem that I think remain with the manuscript.

-------------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Methods, context, page 5
In this paragraph the authors point out that a review might aim to carry out either an exhaustive analysis of all (any?) harmful outcomes or an analysis of only a limited set of adverse outcomes, citing reference 3. The key question is how to make this judgement. The cited paper does not deal with this question. Moreover, although this distinction is raised elsewhere in the paper, I don’t feel the authors help the reviewer to make this key choice. The next paragraph gives examples of different kinds of reviews but still gives no insight about how the reviewers in each case made the decision and whether the decision was appropriate.

Methods, page 6, para 3
Table 2 which is so non-specific that it really doesn’t help the reader to weigh up the pros and cons of broad and narrow questions. I am also concerned about careless use of key words such as applicability / generalisability, which have very specific meanings to systematic reviewers and epidemiologists. In my view, a narrowly focussed review that found high quality evidence and reviewed the evidence appropriately should reach conclusions that ARE widely applicable with respect to the research question that was posed at the outset. Conversely a broad review might find such diverse evidence that its conclusions would not be applicable or generalisable at all.

Methods, page 7, para 1.
I am not sure what the authors mean by “hypothesis-testing”. Should this be taken to imply that reviewers should only include comparative studies i.e. studies that compare at least two different interventions? This paragraph seems to state the obvious but without guiding reviewers about whether they ought to be reviewing specific safety issues or broader concerns about safety.

Methods, page 8, para 1.
The first line on this page provides another example of the use of an important word which is ambiguous in the text. By ‘comprehensiveness’ do the authors mean narrow versus broad, the rigour of searching, or of some other aspect of the review?

Methods, page 8, para 2.
The authors state that “many RCTs exclude high risk groups who are most likely to experience harms”. In fact, many RCTs recruit high risk groups, in terms of their disease severity, because they are likely to have ε
higher outcome frequency; these same RCTs may set eligibility criteria that exclude patients at high risk of specific, known adverse effects. The following sentence (“in addition, some subject areas ...”) is irrelevant since it applies to both harms and benefits. More generally I was surprised that the authors had not included some description of the way in which the choice of eligible studies might be expected to vary by scope.

Locating and selecting studies, pages 8 & 9
The authors spent a lot of time describing mesh and text word searches, which should be a standard part of an experienced reviewer's skill set. If the authors simply advised reviewers to seek help from an expert in information retrieval, this issue should be dealt with. Paradoxically, the authors give less help with respect to issues that affect adverse effects specifically. For example, table 3 simply shows database indexing terms when it could show the structured way in which adverse effects are indexed, as described in the text on page 9.

Methods, page 11, para 2.
The authors blithely suggest that reviewers might want to carry out searches using study design terms such as “trial” or “case-control” without any consideration of the usefulness of these study design labels in identifying studies that used the respective designs.

Methods, page 13, general principles. I feel that one general principle should be a requirement for primary studies to describe the adverse effects that they aimed to identify at the outset. This is different, I believe, simply to requiring definitions of reported adverse effects which are often only given in the results sections of primary studies.

This paragraph seems to me to be very important but doesn’t really take the reader very far forward. From the point of a systematic review, variability in detection methods contributes heterogeneity.

Selective reporting, page 15.
The common usage of this term is to do with reporting of outcomes that are significant or interesting while not reporting other data collected which are not interesting or not significant. This highlights the importance of defining the adverse effects that researchers aim to identify at the outset (i.e. in the Methods of a primary research report) so that the reader can see which ones are subsequently reported. Paradoxically, this may mean that generic statements about "no adverse effects being observed" could be very useful given the usual constraints on space in peer reviewed publications.

Collecting data, pages 16-17.
I was really surprised that the authors don’t discuss what actual data reviewers might expect to have to collect. The data collected clearly depend upon the eligible studies but I have no feel from the paper about whether the authors recommend collecting just numerators and denominators, or traditional effect sizes comparing to normal groups (also probably dependent on study design). There are important considerations to do with risks and rates in studies that are not randomised, for example.

Analysing and presenting results, pages 19-20.
Like the section on data collection, I feel this section fails to consider some fundamental issues. How should reviewers go about presenting the data from several primary studies of different kinds. Should the studies be separated, for presentation, according to study design? In what circumstances should reviewers attempt to combine data across studies using meta-analytic methods? Is it only valid to combine data from comparative studies (as in conventional meta-analyses of benefit), or do the authors believe that it is appropriate to combine data across case studies or across single arms from a number of trials or other studies?

Interpreting results, pages 20-21.
This section presents a general discussion which is reasonable as far as it goes but I don’t feel that it provides anything like a framework for reviewers faced with interpreting concrete results from a review that they may have done. I am frustrated that the authors often pick up on small details, typically as particular examples, but completely fail to tackle the underlying issue.

------------------------------------------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Abstract results, page 3
The last sentence of the paragraph still talks about “evidence-based guidance”. I do not feel that the authors describe detailed evidence in relation to study bias, data collection, analysis, etc.

Abstract, conclusions, page 3
It is not at all clear to me how the authors intend the verb “should be” to be interpreted. Does this mean that readers “need to be able to”, or do the authors mean that readers “should have the competency because they have read the paper”. Given the difficulties in this area and the lack of high quality empirical evidence I think that the best the authors can aim to do is to help reviewers weigh up the pros and cons of different choices at each step in the review process.

Background, page 3
The authors describe omission of information on harmful effects as a form of bias. I do not like the use of the word bias in this everyday sense because it may confuse readers who come across the term in an epidemiological sense. The phrase really is not necessary here; the authors could simply say that “omission of information on harmful effects could misinform anyone trying to make a balanced treatment decision”.

Background, page 2, para 2
The authors claim that the manuscript was drafted with consultation of content experts in reviews methodology. However, I did not recognise those listed in the acknowledgements as systematic review methodologists. Rather, they may be experts in the methods used in systematic reviews, a different thing altogether.

Methods, page 6, para 1
At the end of this paragraph the authors raise the issue of dealing with two data sets and attribute this to the fact that reviewers aim to evaluate benefit and harm together. The authors do not define what they mean by a data set (is this to do with more than one outcome or more than one group of eligible studies). Lots of reviews of benefits only look at multiple outcomes, so I would argue that the key difference is to do with reviewing two different sets of studies for different outcomes. This is hinted at in the last sentence in the paragraph which only comes after the previous, rather confusing, sentence.

Methods, page 7 para 2.
I think this paragraph is a good example of the general ambiguity of the text. I think I understand why the lack of consistency in reporting adverse effects might particularly hinder a broad review. This is presumably because in a narrow review, reviewers would have the advantage of structuring the review around very specific named effects. However, I do not think this is necessarily obvious to a reader who is not familiar with the area.

Methods, page 12.
The authors switch between the Cochrane Collaboration’s preferred term of “assessment of susceptibility to bias”, and other, less appropriate terms such as methodological quality and quality assessment. It is not clear what the authors mean by “standard quality assessment tools”. Such tools hardly exist for RCTs, certainly do not exist for non randomised studies and have never been considered specifically for studies of adverse effects.

Methods, page 13, para 1.
The description of retrospective collection of adverse effects really has little to do with the applicability of tools for assessing susceptibility to bias. Retrospective data collection may mean that the frequency of adverse effects is underestimated but, providing masking between groups is maintained, should provide a valid estimate of relative effect.

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published
**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

The authors may consider that I have a non-financial competing interest, because I co-convene the Cochrane Collaboration's Non-Randomised Studies Methods Group.