Author's response to reviews

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

Authors:
Yoon K Loke (Y.Loke@uea.ac.uk)
Deirdre Price (deirdre.price@clinpharm.ox.ac.uk)
Andrew Herxheimer (a.herxheimer@ntlworld.com)

Version: 4 Date: 23 May 2007

Author's response to reviews:

Dear Editor,

Despite our efforts to carry out all the proposed revisions, we were disheartened to find additional new criticisms directed at existing sections of our manuscript (which had not previously been commented on unfavourably by the reviewer). We find it particularly disappointing that target or goalposts seem to have shifted.

Nevertheless, we have made considerable new revisions in response to these new comments, as listed below.

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.

We believe there is a fundamental misunderstanding here. Our framework is not aimed at dictating right or wrong (or whether reviewers' decisions were appropriate). We provide advice on the available options, and the pros and cons of each approach, so that reviewers can make their own judgement on which adverse effects are important to them and their patients. As each intervention has its own safety issues (which may be very different from other interventions), it would be very wrong for us to lay down a rigid set of rules. Instead we have provided practical advice to reviewers under the Scope section, pg 6.

"The research question about safety and tolerability in a review may be broad or narrow in scope. For example, a review with a broad scope might ask "what adverse effects are associated with antidepressant therapy in humans?" Or, a more narrowly focused review might examine the risk of suicide and suicidal behaviour in adolescents taking a serotonin reuptake inhibitor. Table 2 describes the advantages and disadvantages of addressing broad and narrow questions.

[TABLE 2 near here]

In general, reviewers who have already identified important safety concerns (for instance, from the knowledge of the pharmacology, or anatomical site of the intervention) should carry out a narrow-focused evaluation covering particular aspects of the relevant adverse effects. On the other hand, reviewers who are not aware of any specific safety problems, could start with a general overview of the range of adverse effects associated with an intervention."

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.

We agree with the reviewer's comments and have re-drawn Table 2 with specific examples, and also
removed the careless use of words such as applicability and generalisability.

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.

We have re-written this paragraph to remove the wording regarding 'hypothesis-generating', and also
provided clearer guidance on which approach is preferable in certain situations.

"In general, reviewers who have already identified important safety concerns (for instance, from the
knowledge of the pharmacology, or anatomical site of the intervention) should carry out a narrow-focused
evaluation covering particular aspects of the relevant adverse effects. On the other hand, reviewers who are
not aware of any specific safety problems, could start with a general overview of the range of adverse
effects associated with an intervention."

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?

We have re-written the sentence to clarify that rigour of searching is one of the important elements.

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 a 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.

We have re-written the sentence to highlight that RCTs may exclude patients who are considered likely to
experience harm, and removed the sentence concerning surgery and complementary medicines. We are
uncertain what the reviewer wishes us to do in describing the choice of eligible studies and scope - the
choice of the eligible studies may vary considerably for each individual review, depending on how long the
intervention has been available, the ease or interest for researchers in studying its safety, and whether any
safety concerns have previously been raised. It is not possible for us to tell what the choice would be.

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.

While Mesh and text word searches may well be part of the experienced reviewer's skill set for assessing
the beneficial effects, we believe that any particular reviewer may not have conducted a specific adverse
effects analysis in the past. Table 3 provides information on search terms that are specific to adverse
effects.

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.

Contrary to the reviewer's comments we did not, and have not made any unqualified recommendation on
searching using study design terms. We have now changed the wording to further emphasize the difficulties
involved.

"While a more specific and less onerous search can be performed using study design terms such as 'trial' or
'case-control', the disparate designs of safety evaluations (for instance, research using 'prescription event
monitoring') and the different terms for describing non-randomized studies means that reviewers may miss potentially relevant data."

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.

Given that potential adverse effects may be new or unrecognized to start with, we cannot agree with the reviewer's suggestion that studies should describe the adverse effects at the outset.

Detection methods, page 14. 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.

We have included a sentence on the potential for 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.

We have changed the term to 'Incomplete Reporting'. As discussed above, pre-defining adverse effects is possible only in the limited instances where we know what the problems are going to be. However, there are numerous instances when a spontaneous monitoring system (even within a clinical trial) can pick up important new adverse effects.

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.

The data collection varies immensely depending on study design, and research question. It is not possible to make any recommendation here as individual reviews would have very different requirements in collecting different sets of data.

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?

We have now added a paragraph on the potential value of presenting data according to study design. Nevertheless, we would like to stress that there are many other ways of presenting data, and that there is no single 'correct' method. With regards to meta-analytical techniques, the issues mentioned by the reviewer are generally applicable to reviews of benefit or harm, and there are no special aspects to this with respect to adverse effects.

"Analysis and presentation of results categorized by study design can potentially provide useful insights into a particular adverse effect. In a systematic review evaluating pancreatitis with statins, data from case-control studies were pooled in a meta-analysis to yield a numerical estimate of the relative risk, and to allow assessment of statistical significance. [25] Data from case series and case reports were then used to identify important characteristics such as patients' age and gender, dose and duration of drug therapy, and other susceptibility factors involved in the adverse reaction."

Interpreting results, pages 20-21. This section presents a general discussion which is reasonable as far as
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.

We believe that a general discussion is the most appropriate choice here. Again, a multitude of results are possible in any given review, and our aim is to highlight the options, without being prescriptive.

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.

This sentence has been re-written - Reviewers of adverse effects are given general guidance on the assessment of study bias, data collection, analysis, presentation and the interpretation of harms in a systematic review.

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.

This sentence has been re-written - "Readers need to be able to recognize..."

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".

This sentence has been re-written as suggested.

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.

Amended as suggested.

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.

Amended as suggested. "Reviews that aim to evaluate benefit and harm together will usually require a more complex design that can efficiently handle different sets of studies for various outcomes. Using different search strategies and/or eligibility criteria for studies of benefit and harm will generate two or more diverse groups of eligible studies."

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.

Paragraph re-written to clarify the problem. "Whilst reviewers carrying out a narrow focused review may have to concentrate only on specific named adverse effects, those performing a broad review may be confronted with an unstructured mix of lists, tables and text covering many diverse adverse outcomes. This
difficulty is compounded by the lack of consistency in reporting adverse effects and the absence of a common format for doing so."

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.

Paragraph re-written. Any available quality assessment tools should be used cautiously. However, we disagree with the reviewer's assertion that such tools hardly exist - what about the Oxford (Jadad) score, and the Newcastle Ottawa scale?

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.

Amended to clarify differences in how adverse effects data are collected. Our point remains though that the quality assessment tool may measure only the methods used to assess beneficial outcomes, and the tool may not measure how adverse effects were assessed.