Reviewer's report

Title: Methodological quality of systematic reviews of animal studies: a survey of reviews of basic research.

Version: Date: 7 October 2005

Reviewer: Maria Ospina

Reviewer's report:

General

This paper addresses an important issue, that of the methodological quality of systematic reviews (SR) of animal studies. The research question is novel and very relevant to the field. The manuscript is clearly written in a concise manner. However, this manuscript would benefit from substantial revisions before being considered for publication. There are some serious methodological deficiencies that need to be corrected that would improve the quality of the research work. Some of the deficiencies are related to the methods used for an overview of the methodological quality of SRs, and others are related to deficiencies in the reporting of the review process and statistical methods.

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

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

GENERAL COMMENTS:

The research question is new, relevant and well-defined. Nevertheless, it would be important to provide a more acceptable rationale for the selection of SRs of bench studies as the group to compare the methodological quality of SRs of animal studies. Little is known about the methodological quality of SRs of bench studies, so no firm conclusions can be made in terms of how “poor” is poor quality of SRs of animal studies.

The methodological quality of SRs in human/clinical studies has been extensively evaluated in a variety of biomedical areas and they constitute a more appropriate group to compare the quality of SRs of animal studies (if direct comparisons want to be made). On another hand, bench studies and animal studies may overlap as they are not often mutually exclusive. How did the authors deal with reviews that included studies reporting both bench and animal results? The comparison between human/clinical vs. animal/lab studies would have been more relevant. I would suggest reviewing this approach if comparison data will be reported. Otherwise, it is important to recognize the limitations of the approach chosen for the study.

Another main limitation is related to the quality assessment instrument (e.g., a quality components approach) that was chosen for this study. Please, provide a rationale for not using a validated scale to assess the methodological quality of the SRs that is available in the scientific literature and widely used in similar methodological studies (e.g., the Overview Quality Assessment Questionnaire (OQAQ), (Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44:1271-1278)) (see below specific comments).

SPECIFIC COMMENTS

1) Page 4 under "Methods". Please, clarify if grey literature sources were searched (e.g. conference
proceedings, non-indexed sources, hand searches, etc); otherwise, report this under the “Limitations” section. It gives the impression that the search strategies were not comprehensive enough to identify all the SRs conducted in the field of animal research. This is an important issue as the authors highlight in the Discussion session the problem of publication bias in animal research. If publication bias has an impact on the reporting of primary studies, it is likely that it will also affect the reporting/publication of SRs in the field.

2) Page 5; paragraph 2. The methods section does not provide sufficient details to replicate the work: a) How many reviewers participated in the selection process?, b) If two (LM, and KSK), did they independently applied the criteria for inclusion of SRs in the overview?, c) How disagreements were resolved, if any? Were measurements of agreement among reviewers (e.g. kappa) performed?. If so, please report them. If not, please address this in the Limitations section. d) Please, clarify what type of data were extracted (e.g. variables).

3) Page 5; line 4. “We selected all reviews of animal studies (n = 30)…”. This is part of the reporting of literature search results. Please, move to the corresponding section of “Results”.

4) Page 5; line 4. “From this list we randomly selected reviews of bench studies”. Please, clarify why the authors used a random sample of SRs of bench studies instead of using the total group of SRs of bench studies. If the objective is to map the evidence regarding the quality of SRs in this field, it is more appropriate to work with the actual number of reviews of animal and bench studies that exist out there. There is no justification for using a sample size calculation in this case. It is not clear why sample size calculations were performed. Why a 50% difference was chosen? (based on what available evidence that this is a minimally significant difference in quality). Does the 50% difference relates to a minimally significant difference to be detected regarding ALL the methodological features?.

5) Page 5; under Methods. Please, provide a rational for selecting a quality component approach (e.g. individual quality features). How and why they were selected?. It is surprising that the authors did not use any of the validated tools available in the scientific literature to assess the methodological quality of SRs (Overview Quality Assessment Questionnaire (OQAQ, (Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol 1991; 44:1271-1278), the Barnes & Bero score, the Russell criteria, the QUOROM guidelines, etc.). Particularly, the psychometric properties (e.g. validity, reliability) of the OQAQ have been published (Oxman AD, Guyatt GH, Singer J, Goldsmith CH, Hutchison BG, Milner RA et al. Agreement among reviewers of review articles. J Clin Epidemiol 1991; 44:91-98) and by far, it is the most used quality assessment instrument for SRs in the biomedical literature. The OQAQ is not content-dependent, so it can be easily applied to SRs of animal and bench studies. I strongly recommend its use in this study. It will allow direct comparisons with SRs in other areas as it provides both overall score and individual scores for each of the quality components. It is also important to remind the authors to report how many reviewers performed the quality assessment, the level of agreement, and how discrepancies were solved.

6) Page 5. Line 9. “Group comparisons were made using Chi-square or Fisher’s exact tests”. The description of the statistical analysis should be expanded. Please, clarify your “outcome measure” (e.g. differences in proportions/percentages), and if the statistical tests of significance chosen were applied to differences in proportions. What level of significance was considered? (I assume 0.05, but please, report it).

   The results are reported largely as P values. Recommend that instead they be reported as point and interval estimates of effect magnitude (e.g., the difference in proportions, and the CI on the difference).

7) Page 5 under “Results”. The reporting of the selection process is not detailed enough (both in text and figure 2. Try to follow a QUOROM flow diagram). The authors should report on the reasons for
exclusion of SRs from the overview. If two reviewers participated in the selection process, please report the level of agreement for the inclusion/exclusion of SRs.

8) Page 6 under Discussion. The results on how the quality of SRs of animal studies compares to that of reviews of clinical trials (e.g. Cochrane reviews) should not be reported in the “Discussion” section. The authors made indirect comparisons between their results and those from Jadad et al 1998 (e.g. comparison of proportions and calculated p values for the differences). These indirect comparisons are inappropriate as different methods (study selection, quality assessment of SRs) were used in the study reported here and the Jadad overview.

9) A section of “Limitations” of the study would be helpful for the readers. Please, include it.

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

10) Page 5. Lines 3 and 7. Please, clarify if “study validity” and “study quality” are equivalent terms. If so, please, choose “study quality” as described in Figure 2. Same for “heterogeneity” (line 4) and “differences between studies” (line 8).

11) Page 6. Line 4 and below. Please, report the results using 95% CI for the differences in proportions. P values are not enough.

12) Figure 2 and Appendix 1: Figure 2 is very well organized but it is important to keep consistency between Figure 2 and Appendix 1 in the terms selected to describe the results of the quality assessment. For example; under “assessment of risk of missing studies”, is “Not stated” (Appendix 1) the same as “Unclear” (Figure 2)?

13) Appendix 1: Please, use headings for each page of Appendix 1.

****************************************************************************************
Discretionary Revisions (which the author can choose to ignore)

None

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

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

I declare that I have no competing interests.