Reviewer's report

Title: Risk of bias judgements and strength of conclusions in meta-evidence from the Cochrane Colorectal Cancer Group.

Version: 0 Date: 02 Jan 2019

Reviewer: Sarah Nevitt

Reviewer's report:

I have performed a statistical review of the manuscript "Risk of bias judgements and level of equivocation in meta-evidence from the Cochrane Colorectal Cancer Group. A review of reviews."

This is a very interesting topic which is well worthwhile researching and I read with interest how the authors approached this topic.

I have the following comments on the manuscript:

Major comments

1) I do not think that the title accurately reflects the research for two reasons. Firstly, the authors are examining the type and strength of the conclusions made rather than equivocation (i.e. ambiguity or evasiveness). Also, I wouldn't say it is a 'review of reviews' that the authors have done. Some of the considerations are made on a review level, others within a study within a review level and others based on subgroups of reviews by intervention type. So, unless the second part of the title is required under formatting guidelines of the journal, I suggest deleting 'A review of reviews'. For me, the first part of the title (with 'level of equivocation' reworded, see above) sufficiently describes the work.

2) Throughout the manuscript, the authors refer to 'recommendations' made by the Cochrane Reviews considered. I understand what the authors mean by this language and I appreciate that some of the reviews published several years ago may use the term 'recommend.'

However, Cochrane Reviews should not be making 'recommendations' (see R102 of the latest version of MECIR standards), instead conclusions should concentrate on whether evidence informs or could impact on clinical practice. Therefore, I suggest that the term 'recommend' should not be used in this manuscript either. Please reword - e.g. categories could become 'Informs clinical practice - Firm, Does not inform clinical practice - firm' etc.).

3) It is a little difficult to follow some of the numbers throughout the manuscript in terms of the number of Cochrane Reviews, number of studies within those reviews and number of risk of bias judgements made.

Please make sure that it is clear exactly what a number refers to and what the denomination of all percentages are. I also suggest adding further details to the PRISMA diagram as the number of
Cochrane Reviews does not tell the full story.

I suggest that 'study' is a better word that 'Papers' included in a review as Cochrane methods group together all citations (e.g. journal articles) of a 'Study' under a single ID and the risk of bias judgement is made based on the 'Study' taking into account all 'papers' published of that 'Study.'

4) Throughout the results and conclusions, the authors refer to the group of reviews with 'high and unclear risk of bias' as a combined group which I interpret the authors to mean 'bias present', with the comparator group being the reviews with low risk of bias (i.e. no bias present).

I do not consider this combined high/unclear to be intuitive for interpretation, as the authors highlight in the discussion, a large proportion of this combined group is from 'unclear' judgements such as 'study was described as randomised but no detailed of randomisation method given.' This really reflects poor methodological reporting rather than 'risk of bias' necessarily, and such poor reporting is likely more common within older studies published before the widespread use of CONSORT statements. Furthermore, as the authors are giving high and unclear different 'scores' in their quality scoring total, I question whether it is appropriate to consider these two judgements as a combined group.

Therefore I suggest that it would be more intuitive to interpret the results with low risk of bias as the 'reference group', i.e. Figure 2 should display the % of low risk of bias, the text should describe that M has the highest percentage of low risk of bias judgements etc. Also the testing should be based on the three categories (see comment 5 below).

Certainly the discussion is helpful for further describing what constitutes high and unclear risk of bias for the different domains, but overall I suggest it would be easier to follow and interpret the findings of this work if the low risk of bias reviews were the main focus.

5) Page 6: "A chi-square test was performed, and a significant relationship found between the intervention type and the likelihood of a paper to show a high or unclear risk of bias (Chisquare, df 2, 36.533, p < 0.001)."

Related to comment 4, if df=2 here then you are not testing high and unclear risk of bias combined vs low as this implies. A chi squared test with df=2 is testing three categories, presumably low vs high vs unclear, so this p value corresponds to the difference between the three categories. Please check your statistical testing and interpretations throughout the results text again

6) Page 6: Good that the authors are aware of multiple testing and have employed a Bonferroni correction which adjusts for the three intervention types. However, please note that it is actually seven different tests of the risk of bias domains in the three intervention types so it should be alpha / 21 rather than alpha / 3 (so approx. 0.002).

7) Page 7: Quality score assessment

I would also be interested in whether the conclusions made by the reviews are supported by the quality scores e.g. what were the (normalized) mean quality scores and 95% CIs for each subgroup of reviews by conclusion type? A one way ANOVA of this could also be done

8) I assume that all of the risk of bias judgements considered within this work were taken directly from the Cochrane Reviews and assumed to be 'correct.'
While Cochrane reviews do generally employ rigorous methods of double quality assessment and often editorial team members also checking these judgements, as the authors highlight, these judgements are subjective and in this case, not perfect. As the authors note, over 5000 judgements that could have been made were not made. I appreciate that it would have been a lot of work for the present authors to make these 5000+ missing judgements as well as checking the 7000+ judgements made, but this is a 'limitation' of the work.

I also wonder if there were any systematic differences in the way the judgements were made. For example, in Cochrane Reviews of other clinical areas involving surgical interventions, I have seen high risk of bias due to lack of blinding applied automatically and also I have seen low risk of bias for blinding applied to objective outcomes where interventions such as surgery cannot be blinded. The authors do mention the difficulty of blinding of surgical interventions in the discussion, but did the authors notice that this lead to any differences in the way risk of bias was considered for this domain (or any other domains) within the reviews examined?

Minor comments

1) Page 3: "the Cochrane Colorectal Cancer Group (5)."

Reference 5 cites the Cochrane Eyes and Vision Group.

2) Page 3: "The combined data will provide a view of the quality of a selection of colorectal literature over time"

Sample may be a better word than selection here. Selection implies that the literature has been 'selectively' chosen here which it has not.

3) Page 4: "and the provided database checked for accuracy." I don't understand what this means.

4) Page 4: "An MS paper was one where a surgical intervention was assessed in the setting of medical intervention or vice versa."

Could an example be given of this? As a non-clinical expert, I'm not sure what this means - and I assumed that Medical & Surgical would be combination of medical and surgical intervention on first reading.

5) Page 5: "Reviews that were classified as O (that is, reviews that considered an intervention that was neither surgical nor medical) were excluded." Please explain why these reviews were excluded.

6) Page 5: "A subgroup of reviews concerning laparoscopic interventions was isolated and assessed. A graphical representation of the commentary made on evidence within the conclusions of those reviews (a "word cloud") was generated using Microsoft Word. " Please state how many reviews, why specifically laparoscopic interventions were used for this and why such a representation is useful (For information, I do consider such representations of evidence extremely useful!)

7) Quality score (page 5): Please explain why the score is normalized (rather than expressed on a 0 to 21 scale). Also, I assume that the quality score for a review (e.g. the numbers in Table 3) is the average of the scores across all studies included in the review. Please add this into the methods
8) Methods: Presumably two or more reviewers independently extracted information? This is implied throughout the methods but not actually stated.

9) Page 6: "The CCCG made a combined total of 7,564 judgements across the seven ROB categories. In 5,680 instances, a ROB judgement was not recorded." Please explain this further.

10) Page 7: Quality score assessment, normalized mean score.

It isn't necessary to provide both an SD and a 95% CI.

11) Discussion: For all of the Cochrane Reviews used as examples in this discussion section, please either add the citation to the reference list within the same style as the other references used in earlier sections of the paper (e.g. Vancouver style I think, rather than author year), or refer to Appendix 2 rather than author year.

12) Discussion: The word 'predictive' or 'predictor' is used several times. This term is usually used for relationships between independent and/or controlled variables and outcomes (e.g. does treatment predict survival?) Risk of bias is not an outcome so using 'predict' here looks strange to me. I suggest 'association' or 'associated' would be more appropriate.

13) Page 11: "This suggests that a reader of surgical colorectal meta-evidence may expect an increased likelihood of high or unclear risk of bias influencing the review's original input papers, but the likelihood of a clinical recommendation being made will be similar to that of a medical meta-analysis or systematic review comprised of papers with less risk of bias." I don't understand this sentence?

14) Table 2: it is not clear what n refers to here, please clarify.

Also, I assume that the % is based on the number of risk of bias judgements (from the text above) rather than number of papers or number of patients. Please clarify.

I suggest it would be helpful to have a table which summarises all of the results in a single table. This could be done by extending table 2 to have the columns (overall and by intervention type): reviews, studies, patients, low risk of bias, unclear risk of bias, high risk of bias, average quality score and the conclusion type categories.

15) Table 3: Please define the Rec. abbreviations.

16) Please clarify what the line bars on the graphs correspond to, presumably standard error bars?

Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests.
Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Were you mentored through this peer review?
No