Author’s response to reviews

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

Authors:
John Delaney (jdel2642@uni.sydney.edu.au)
Rebecca Cui (cuireb@gmail.com)
Alexander Engel (alexander.engel@sydney.edu.au)

Version: 2 Date: 11 Mar 2019

Author’s response to reviews:

Reviewer Comment: Authors contributions and Conflict of interest should not be in the Title page
Author Comment: Amended

Reviewer Comment: Running head is not needed
Author Comment: Deleted

Reviewer Comment: Complete all affiliations with city and country
Author Comment: Amended

Reviewer Comment: The excluded 22 studies under PRISMA may be specified with reasons.
Author Comment: Included

Reviewer Comment: Figure 4 may be avoided, it may be explained only under text. Sometimes such figures might provide deceptive understanding.
Author Comment: Deleted – the quality score has been removed.

Reviewer Comment: 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.
Author Comment: Title amended
Reviewer Comment: 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.).

Author Comment: Text updated to ‘informs clinical practice’ or ‘does not inform clinical practice’

Reviewer Comment: 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.'

Author Comment: Updated numbers. Please revert is still unclear. 'Papers' updated 'Study'.

Reviewer Comment: 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.

Author Comment: Author Comment: Analysis revised with “low risk” as the comparator. Additionally, “high risk” has been assessed independent of “unclear risk”.

Reviewer Comment: 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
Author Comment: Thank you for this. A thorough review of all statistical analysis has been undertaken.

Reviewer Comment: 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).

   Author Comment: Amended

Reviewer Comment: Page 7: Quality score assessment

   Author Comment: Based on commentary from reviewer #3 regarding the uneven impact of each ROB category, the quality score described in the initial manuscript has been removed from this submission. 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

   Author Comment: Author Comment: A thorough review of all statistical analysis has been undertaken.

Reviewer Comment: 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'.

   Author Comment: The bias created by these missing judgements is again highlighted. Unfortunately time did not permit us to review all studies again.

Reviewer Comment: 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?

   Author Comment: Cochrane’s commentary on bias within the surgical domain often noted the improbability of blinding, though there was no mention of a default risk. We have not discussed this with the CCCG but perhaps could. The systemic challenges of surgical research are highlighted again.

Reviewer Comment: Page 3: "the Cochrane Colorectal Cancer Group (5)." Reference 5 cites the Cochrane Eyes and Vision Group.

   Author Comment: Amended, apologies for this error

Reviewer Comment: 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.

   Author Comment: Amended
Reviewer Comment: Page 4: "and the provided database checked for accuracy." I don't understand what this means.

Author Comment: We have clarified the language, please revert if still unclear

Reviewer Comment: 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.

Author Comment: An example has been provided. If it remains unclear we can explain further in the text.

Reviewer Comment: 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.

Author Comment: More detail has been provided.

Reviewer Comment: 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!)

Author Comment: We have provided more background and explanation of the subgroup assessment of laparoscopic surgery, along with more detail on the sample.

Reviewer Comment: 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

Author Comment: Quality score has been removed from this review.

Reviewer Comment: Methods: Presumably two or more reviewers independently extracted information? This is implied throughout the methods but not actually stated.

Author Comment: Updated

Reviewer Comment: 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

Author Comment: We have added further explanation here

Reviewer Comment: Page 7: Quality score assessment, normalized mean score. It isn't necessary to provide both an SD and a 95% C

Author Comment: Removed

Reviewer Comment: 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.

Author Comment: Citations updated and standardised.
Reviewer Comment: 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.

Author Comment: Amended

Reviewer Comment: 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?

Author Comment: Discussion has been revised.

Reviewer Comment: 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. 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

Author Comment: Revised

Reviewer Comment: Table 3: Please define the Rec. abbreviations

Author Comment: Updated

Reviewer Comment: Please clarify what the line bars on the graphs correspond to, presumably standard error bars?

Author Comment: Updated

Reviewer Comment: Throughout the results section, results should be presented alongside information about (a) the direction of the effect and (b) the size of the effect. Presenting p values alone is insufficient to judge the difference between categories

Author Comment: Updated

Reviewer Comment: * The discussion section is very descriptive, and could better showcase the findings of the review by highlighting key evidence and discussing the implications. Some examples:

Author Comment: We have substantially revised the discussion and conclusion. We would welcome any further comments.

Reviewer Comment: Unclear why the authors do not also report the relationship between risk of bias and recommendation type. If this data is available to the authors, this would increase the interest of this paper.

Author Comment: Added

Reviewer Comment: Authors should emphasise in abstract and throughout study aims that RCTs were the focus of this study and no 'real world data' were included.

Author Comment: Added

Reviewer Comment: Description of the risk of bias categories in introduction is very basic and well known to many readers of Systematic Reviews. It would be stronger to add in some references demonstrating the relative importance of these factors for the validity of research; e.g. association
between allocation concealment and treatment effects (Schulz et al 1995).

Author Comment: Introduction has been revised

Reviewer Comment: Add start date of search to methods (lit search and inclusion criteria)

Author Comment: Added

Reviewer Comment: Clarify description of table 1, on page 5 line 6: this table is about recommendation strength, not a description of how reviews were able to inform clin practice.

Author Comment: In response to commentary from reviewer #2 the wording around conclusion type has been changed

Reviewer Comment: How does the word cloud contribute to the aims of the paper?

Author Comment: We have offered more background and support to the laparoscopic subgroup and the word cloud

Reviewer Comment: Computation of the quality score is pragmatic and explained well, however it should be noted that this approach is limited as it does not take into consideration the relative importance of each category. Not all categories contribute equally to the validity of the research.

Author Comment: On review, we have removed the quality score. In order to provide a useful metric it needs to be able to incorporate a weighting. We will continue to work on this.

Reviewer Comment: Change referencing style in discussion to match rest of paper

Author Comment: Updated

Reviewer Comment: Weighted kappa indicates some disagreement between authors on the conclusiveness of the reviews. In how many instances? Was there a pattern in the disagreements?

Author Comment: We have provided some more information on this area

Reviewer Comment: Clarify inclusion of papers that did not "specifically address questions surrounding colorectal cancer per se" (page 8, beginning line 49). Surely these did not meet the inclusion criteria?

Author Comment: The inclusion criteria was for reviews produced by the CCCG, which included non-cancer reviews.

Reviewer Comment: Insert paragraph after "higher risk of bias." (page 11, line 6)

Author Comment: Added

Reviewer Comment: The authors may wish to mention in their summary of the review strengths that this is a large review across cochrane reviews, which are viewed as being a gold standard of systematic reviewing. That in all cases, the same validated checklist was used to evaluate risk of bias, which is also gold standard and provides a method for comparison across trials. In limitations, should mention limitation of the summary quality score.

Author Comment: Added