Reviewer's report

Title: When Should Observational Studies Be Afforded More Weight in Evidence Reviews? Case Examples

Version: 1 Date: 10 December 2013

Reviewer: Holger Schünemann

Reviewer's report:

Major comments and suggested revisions

1. We believe that the submitted article will provide some excellent examples and guidance for those interested in systematic review methodology, however we do think that the readers of this manuscript will benefit from a more structured approach to illustrating the examples and making these examples easier to follow. This would include the description of which criteria were considered for determining that the strength of evidence is increased for these “bodies of evidence”. It would also include grouping the criteria appropriately, such as all criteria being evaluated under risk of bias that all belong into the risk of bias category without mentioning them twice and separately. The authors have already attempted to do this in some examples. We would just suggest that this is being done consistently.

2. One of the fundamental discoveries for us reviewing this paper was that it appears that the confidence or the strength of evidence for these observational studies was high because criteria that were previously considered by Bradford Hill or the GRADE approach or the modified GRADE approached used by ARC were already present. It is our view that there haven’t been any new criteria for increasing the strength of evidence identified. This, in fact is a very important finding for this article because it suggests that the work that has so far been done in looking at criteria that increase the strength of evidence in is complete. The authors may wish to comment on this particular aspect in their discussion.

3. On Page 8 in the Results and Discussion – one of the major suggestions for us is that for each of the examples should provide a clear PICO on top of the table so that it will be easier to follow for the actual authors. We have another suggestion for the structure and that relates to the sections about potential sources of bias, best evidence and strength of evidence domains and criteria for causation are placed. To us, there is possibly some confusion. To us it seems it would be most helpful to always discuss the risk of bias before presenting the case examples and then discuss how the cases deal with the risk of bias or at a separate discussion section after presenting the case example and illustrate how the examples of the different scenarios respond or control potential sources of bias.

4. The authors conclude with recommendations or suggestions for how to approach the problem. So, for instance, the authors provide criteria for how to
improve the strength of evidence but we think that this information is hard to follow for readers at this point and there is insufficient instructions for how to use these criteria in future work. Please see our first comment in that we feel there is possibly not sufficient new information about criteria and how new criteria would be applied for increasing the strength of evidence. Once again, one way of rephrasing the paper is to focus on the examples and how, when there is a general feeling that the strength of evidence is higher, usually certain criteria for upgrading the strength of evidence that have already been described by other authors are applied. It would also be helpful to see if and how this approach could be integrated with the flowchart on how to address the quality of evidence or strength of evidence in non-randomized studies provided by Schünemann and all in Research Synthesis and Methods 2013, issue 4, pages 49 – 62.

5. Our next major comment relates to how to assess the risk of bias. The authors provide detailed descriptions on the types of bias within the risk of bias domain and a case-by-case analysis of controlling risk of bias. For example on page 15, the table clearly identifies the risk of bias that may be present in what appears to be high quality primary studies. They describe that there are four different sub-types of risk of bias present. We believe that it would be helpful to apply a structured general risk of bias tool to describe the situations for when risk of bias is present for certain examples and when risk of bias is not present for certain examples. Perhaps the use of the draft version of the non-randomized study risk of bias tool will be helpful for the reader and it would be a good first validation exercise for the newer risk of bias tool that has been put forward under leadership of the Cochrane Collaboration but with involvement of some of these authors here as well.

6. Criteria to elevate the strength of evidence for observational studies

7. The authors propose criteria based on examples to elevate the strength of evidence. Based on our review of this article in all but one of the examples are the already established criteria for upgrading the quality of evidence present. In the one example where it is possible that there was no upgrading criterion (the example of the effects of bariatric surgery on mortality) there may be specific study design issues that reduce the risk of bias. What would be extremely helpful is to find out if there are criteria that can be operationalized that lead to a higher trust in the actual design of and the avoidance of risk of bias in the observational study that lead to increasing the quality of evidence. In other words, it seems to, as I have said, in all of the other examples a structured evaluation according to GRADE would have led to exactly the same results in terms of increase quality of evidence or strength of evidence. And this may be the only example where there may be design issues that would lead to increased quality. It is also helpful to describe how the authors are dealing specifically with the criteria that in the current GRADE/ARC approach lead to downgrading such as lack of precision, lack of directness, rather than increasing the quality of evidence. It seems to us, as mentioned above, that this paper does not provide the detailed guidance in terms of how these criteria are applied and they are applied inconsistently across the different examples. This is not to say that the examples are extremely useful
and we need to look at these examples but the whole framing of the article might be improved by taking this more structured approach.

We will provide some minor comments that could be considered essential.

On page 4 in the Methods, we believe that there may be some insufficient description of the methodology. The reader would benefit from describing the sources of the case examples and how they were selected. The authors state that the intent of this report is to illustrate the role of observational studies according to the current evidence based practice center methods guidance through case examples. It appears to us that these methods have actually not been described clearly, previously and more importantly it is not applied consistently across the examples that are described here. It might be helpful, although not mandatory to just describe how the judgment of about the eligibility of the systematic review and individual studies were done. In general we would suggest obviously to focus on systematic review rather than describing increasing strength of evidence on the basis of individual studies as that is really the purpose of the strength of evidence assessment.

On pages 8, limitation of randomized controlled trial data, the included systematic review of observational studies do not necessarily improve the strength of evidence. As the author stated in both reviews the inclusion of observational data did not significantly improve the strength of evidence for treatment effectiveness, however the authors chose to include them to highlight the need for stronger studies in the future in order to increase the strength of evidence. Does that mean that systematic review of observational studies do not sufficiently address the concern of the randomized controlled trial data that are also limited? Perhaps the authors could clarify this further.

Also on page 8 there is insufficient patient information on the included comparative observational studies. The authors provided information on randomized controlled trials where the evidence is not sufficient but they have not provided details about the included observational studies. The authors described it as a result, the review authors made a decision to include comparative observational studies that reported benefit outcomes and so on. It is not clear to us whether the included comparative observational studies actually do address the concern and a slight description there would be helpful.

Page 11 – Coherence

This usually refers, as the authors describe to whether results are congruent with the known biological course of history of a condition. As the authors stated, coherence is a controversial criterion. We think it is very important but even without knowing the biological course or history of a condition can we have high confidence in the estimate. Furthermore we would suggest that coherence is already addressed by asking the right PICO questions. Without knowing the biology in many instances would we not conduct research, explore the question or try to answer the question. It would be helpful to describe if the example has criterion of coherence better.
Dose response relationship on page 16, there are primary study examples which demonstrated dose response relationships and that they may be helpful to elevate the strength of evidence. However the primary studies are possibly biased. It would be helpful to see if this criterion of dose response relationship should also be used in the context of possibly biased observational studies. In other words, would the authors suggest upgrading or increasing the strength of evidence if there was studies that suffered from risk of bias if they showed a dose response relationship and if so if they could provide a rationale for their assumption?

**Level of interest:** An article of importance in its field

**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:**

I am co-chair of the GRADE working group.