Reviewer’s report

Title: Selective citation in the literature on swimming in chlorinated water and childhood asthma: a network analysis

Version: 0 Date: 27 May 2017

Reviewer: Adam Dunn

Reviewer's report:

Thank you for the opportunity to review this manuscript. The authors have used citation networks to examine the potential for citation bias in studies on swimming in chlorinated water and childhood asthma, finding that authors were more likely to cite themselves and that articles confirming the association. The strengths of the research include the number of factors that were considered, the neat and clear structure, the reporting of a protocol, and the systematic identification of literature. The weaknesses include the small number of articles that were included, which may have affected the ability to draw conclusions about the differential effects of self-citation versus citation bias (selecting for result) for other reasons. In what follows, I have made comments and suggestions generally in chronological order but because they relate to sections of the manuscript I haven't provided line/page numbers.

Major comments:

1. Abstract: when stating odds ratios, always include 95% confidence intervals.

2. Extracting articles only from Web of Science Core Collection for clinical articles may have been a problem. It would be important to expand the search to include at least PubMed and preferably Embase as well. I agree that reference checking may affect the validity of the study but there is a strong chance many articles will have been missed. Details of the systematic identification of the literature should be described in detail, and the search strategy, number of identified articles, and the search date should all be included (or replicated from the protocol in supplementary information).

3. Methods: The details of how the citations were identified was not clear and should be the first part of the methods given its focus in the aim of the study. If this was done using CitNetExplorer and not manually validated (with proof that it correctly retrieves all articles correctly), then there is a major risk that citations were missed. Since there are only a small number of articles and potential citations, this could easily be done manually (we have previously done this for more than 10,000 citations in 152 articles: 10.1016/j.jclinepi.2014.09.014).
4. Methods: Explain in more detail how differences were resolved when the publication date of a study was shortly before the publication date of an article and may have been published after the final submission date of the other article. Were submission dates examined? Because authors know their own articles are accepted before they are published, this may have skewed the potential for self-citation due to differences in timing.

5. Methods: Similarly, the explanation of the method for statistically analysing the selection of citations was brief and not clear. This could be expanded given that it is different from the rest of the analyses.

6. Figure 1: A co-authorship network is fine given the results, but given the objective of the manuscript, why not include a figure for the citation paths? There are neat ways of representing potential citation paths that also include information about the publication date. These go back to Garfield's work, but an example of one that includes temporal information and would be suitable for this size network is here: 10.1016/j.jclinepi.2013.11.015 (disclosure - I authored this).

7. Consider a measure in which the conclusion of the cited article matches the conclusion of the citing article. Citation bias makes more sense when the two are aligned and you could define a measure which indicates the over-representation of "positive" articles cited in articles with positive conclusions compared to the distribution of citations of those "positive" articles in articles that were "negative", each given as a proportion of the number of times an article *could* have been cited.

8. Using only 36 articles and a substantial number of characteristics is problematic because of multiple comparisons and this could have been determined at the protocol stage. First, state the number of different predictors next to the number of articles (which is the denominator of the analysis). Then decide if the 36 articles are able to say anything useful about these measures. This was a major limitation of the study. While there is not much that can be done about it now, it seems unfortunate that all the statistics are reported and then the limitations get a few sentences at the end. I have no problem with the way the conclusions are stated in relation to the study.

Minor comments:

1. In the abstract, it would be worthwhile stating the number and proportion of the 570 citation paths that were "realised" (i.e. the 191; 34%). The problem, of course, is that it is not always clear when the date of revised submissions and publication dates were incomplete or ambiguous.
2. In the background, try re-writing the definition for citation bias more precisely. Readers trying to understand the difference between the selective citation (affected by lots of factors) and citation bias (selecting studies based on their results or conclusions) may struggle.

3. In the background I felt the description of a QRP did not add much to the manuscript and was a little out of place. To motivate the study, explain what the potential impact of citation bias can have and use examples where citation bias is known to have affected the research consensus.

4. There are some other examples where citation networks have been used to evaluate bias in research more recently than Greenberg, though that is a seminal example.

5. Methods: I have never seen "in duplo" before and I'm not sure if it is a word.

6. Methods: No need to repeat the definition of citation bias here. In addition, it should be defined rather than "spoken of".

7. Methods: How was "gender" assessed? How was "country" assessed where authors had multiple affiliations in different countries?

8. State p-values exactly, not as "p<0.05" and especially when the data are not shown.

9. When referencing a study by an author's name, include the reference after the name so it is clear to the reader which reference is being discussed.

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.