Reviewer's report

Title: Automating Data Extraction in Systematic Reviews: A Systematic Review

Version: 1 Date: 15 April 2015

Reviewer: Zachary Munn

Reviewer's report:

This is a very interesting and timely paper. I congratulate the authors on their work and believe they have developed a useful overview of current approaches to the automation of data extraction in systematic reviews. Comments are below:

- Major Compulsory Revisions

1. As a systematic review, this report should be in line with the PRISMA standards. As such, please clarify the following:
   • Was there a protocol for this review?
   • Discuss risk of bias and justify why this was not done.

2. I am not sure what this statement means in the methods section, page 14: We conducted our study procedures based on formally defined processes and instruments that were drafted and piloted by one of us (S.R.J.) What are the formally defined processes and instruments?

3. Study selection: Justify why you had a second author independently review two random samples of 100 citations instead of the full list of titles and abstracts. The agreement between these two is mentioned later but more detail is needed.

4. I'd like to see more information in the discussion about the importance of the findings of this review. What are the implications? The final sentence in the conclusion hints at this, but I would like to see this discussed further.

- Minor Essential Revisions

1. Page 5, line 3: what and who are the expert review committee? How was this group formed?

2. General gramma and writing needs to be improved, and the paper would benefit from a professional copyedit, for example (not exhaustive):

   a. Paragraph 3, introduction, line 7: There is limited knowledge and methods on how to automate the data extraction 7 phase of the systematic reviews

   Page 15, line 25: to classify medical abstracts whether they contain patient-oriented evidence

3. Page 13, line 9: Why is the Dawes study included if there was no automation?
To me it doesn’t seem to meet the eligibility criteria

4. Appendix 2 is referred to in the text before appendix 1.

- Discretionary Revisions

1. Define natural language processing in the background (page 3).

2. Paragraph 2, introduction: you discuss automation and then leap into details about the length of time it takes for published studies to be included in systematic reviews and that within 2 years systematic reviews are out of date. I think you should state here why this is important, and how automated data extraction might help.

3. The statement in paragraph 2 that it may take 2.5 to 6.5 years for a primary study to be included and published in a systematic review does not seem to be mentioned in the Bragge reference, but rather in the paper by Julian Elliott on living systematic reviews.

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

**Quality of written English:** Not suitable for publication unless extensively edited

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I am a director and senior research fellow at the Joanna Briggs Institute.