Reviewer’s report

Title: Data extraction from machine-translated versus original language randomized trials: a comparative study

Version: 1 Date: 17 July 2013

Reviewer: Alexis Llewellyn

Reviewer’s report:

The authors addressed an interesting and relevant question. The methods were well described and it appears that sufficient details were provided to replicate the work (taking into account the original AHRQ report to which the paper refers). Attempts were made to distinguish between issues associated with poor translating, poor reporting and errors unrelated to translation in extraction. Relevant limitations of the study were acknowledged, and cautious attempts were made to explain variations in findings relating to accuracy across languages.

I have the following suggestions for revision:

A) Major compulsory revisions

1) Methods, “Data extraction form and comparison”, first sentence: I would at least nuance the statement about which domains are the most important for systematic review. I think judgement about importance and detail given to each domain depends amongst other things on the review question, and one could argue that extracting at least some basic data on participant characteristics and comparator is often equally important.

2) Conclusions: The authors clearly attempted to provide nuanced and balanced conclusions. However, I do not think that the conclusions sufficiently reflect the evidence presented: concerns about the inaccuracy of extraction across all foreign languages regarding results and/or outcomes data do not suggest, in my opinion, that basic translation software is sufficiently accurate for data extraction in systematic reviews of health interventions. I am not convinced that the better results in other domains, or the relatively limited resources and effort needed for using the tool may sufficiently outweigh the inaccuracies found in the analyses.

3) Conclusions: 3rd sentence: the threshold according to which extraction is considered “fairly accurate” is unclear. Some clarification is required to strengthen this statement.

4) Discussion: Despite the fact that a power calculation was performed, I still have some concerns about power/precision of estimates, notably given the wide confidence intervals of ORs presented. I think a brief comment on this (similar to the one main in p22 of you original AHRQ report) would be appropriate.
B) Minor essential revisions

5) Results, “Comparison of extractions from translated to original articles”: For consistency, and as for Chinese articles, the % of items that are each #98% correct for Spanish papers should be reported.

C) Discretionary revisions

6) Abstract: The results section of the abstract currently reports the percentage of items that were correctly extracted more than half the time. I think this could be seen as a low threshold for medical research. I would prefer to see the % of items that are each #91% (or #98%).

7) Background: For readability, I suggest breaking the last sentence of the background into two.

8) Methods: Figure 1 in the original AHRQ report may make the translation/extraction process more explicit. I would consider adding it to the appendices.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

I declare that I have no competing interests.