Reviewer’s report

Title: Comparative efficiency research: Meta-analysis of cost-effectiveness studies

Version: 2 Date: 6 October 2014

Reviewer: Inigo Gorostiza

Reviewer’s report:

Minor Essential Revisions

1. Abstract / Methods / 1st paragraph/ line 29.
A re-wording of the first line is suggested as the term “meta-analysis” constitutes a broader concept than the one that authors used.

2. Abstract / Conclusions / 1st paragraph/ line 46.
I don’t think that this conclusion could be drawn from the article in so a taxative way. Maybe a less assertive conclusion would be advisable.

3. Abstract / Results / 1st paragraph/ line 44.
Maybe this statement “The TINB test was confirmed when increasing the sample size” seems to be obvious, as if TINB is supposed to follow a normal distribution (assumed by authors; line 259), p-values should decrease as sample size increases.

4. Background / 2nd paragraph/ line 92.
It is not clear from the text if the term COMER has been used before or it is a new term proposed by the authors for their cost-effectiveness meta-analysis new method

5. Background / 4th paragraph/ line 112.
Although authors declare that the COMER method is suitable for individual and aggregated data, it seems that they only work on individual data meta-analysis in the remainder of the article. Maybe in the Discussion section a commentary could be inserted on that issue.

6. Methods / 1st paragraph/ line 117.
For those not familiar with cost-effectiveness analyses and the ICER, perhaps it should be advisable to insert here an introductory paragraph on the problem that arises when having a ratio with 2 parameters with very different distributions; costs and effects. The ICER is introduced in lines 235-241, but maybe it would be better to explain it earlier.

7. Methods / 2nd paragraph/ line 126.
Should it the words “joint distribution” be added to this line, for a better
comprehension?

8. Methods / line 139.
A “C” is lacking: C(u,1)=u and C(1,v)=v

9. Methods / line 141.
Shouldn’t it be needed a capital “C” in the third one?
C(u2,v2)-C(u2,v1)-C(u1,v2)+C(u1,v1)>-0.

10. Methods / lines 164-166.
The data from the allergic rhinitis study are apparently used for creating the hypothetical population and applying the whole COMER process to them, not only for estimating which is the most appropriate copula. Maybe I am wrong but I have understood it this way. So I suggest that perhaps, a better approach to the paragraph could be to move the first line (“Because it is crucial to know the most appropriate copula for this study”) to line 169, after the “…COMER.” sentence. So the logic could be: “as we need to create a hypothetical population where to prove/develop our COMER method proposal, we use data from this study. we will also…”

I suggest that authors gave more details about how the process of the creation of cohorts was made: how many cohorts were created (500-15= 485 cohorts?); how the cohorts were assigned to each study (Randomized?). It’s not clear how if sample sizes ranged from 15 to 500 individuals, there could be studies with sample sizes with as few as 3 individuals each. That’s why I think that lines 214-217 would deserve a more detailed explanation.

12. Methods / Creation of cohorts/ line 128.
Perhaps it would be useful at this point to insert a paragraph where authors explain why they are going to consider those different scenarios. Readers could not be so familiar with the cost-effectiveness plane so as to readily understand what those three scenarios mean

Wouldn’t it be the term “effectiveness” better than “effectivities”?

Authors assume that “TINB follows a normal distribution” but maybe authors should justify this assumption.

15. Results / line 279.
Previous to delivering results based on significance (p-values), perhaps information on the generated cohorts/studies should be advisable; something similar to the “Table 1” of clinical articles. The first 7 columns of Table 2 could serve as a guide. Although there would not be a good idea to submit information on all the studies, at least it would fine to watch at 10-15 of them; just to see how
the generation process worked.

16. Results / line 298.

It is not clear from this statement (and more on the rest of the document) if this sample size refers to all the individuals of the study or is a sample size for each cohort of the alternatives being compared.

17. Discussion / paragraph lines 376-383.

The method that authors present here for generating the patient cohorts for feeding the studies seems to “produce” very homogenous cohorts in parameters other than cost/effects. But this is not the actual scenario when performing meta-analyses in real life: patients differ in basal age, basal severity, come from different health-systems, there are variability among centers, and so on, and those parameters might have an impact on the ICER (inter-studies variability).

Authors talk on homogenization as a potential solution to variability in cost structures or health-systems, but perhaps, the influence of other variables as for example, disease severity, use of resources or the age should be taken into account with a different approach (for example, something similar to meta-regression techniques).

Probably the authors had this fact in mind and it could be a new development line of research for the COMER method; but it would be fine if the authors could also insert this idea as a limitation in this article.

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

**Quality of written English:** Needs some language corrections before being published

**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