Dear Editor

This is to re-submit a revised version of our paper (MS: 4481253281026597): “Combining evidence and values in priority setting: testing the balance sheet method in a low-income country”

We appreciate your interest in considering our paper for publication. We are grateful to the two reviewers for their highly constructive comments.

Here follows a detailed response to the reviewers’ comments:

We respond to each point in the order they appeared.

A. Response to Baltussen

Major comments

1. a. We agree that the criteria used in this study could be seen to belong under a broader framework than the twin objectives of efficiency and fairness (for example "ease of implementation"). However, we believe that the criteria actually used here (prevalence, burden, coverage, clinical outcomes, cost-effectiveness, and social group characteristics) fit nicely within this framework.

   b. We did not mean to say that the criteria used are "constraints to policymaking". We meant to say that there exist political constraints - such as dominant interest groups, donor agencies' priorities, multiple levels of government and dysfunctional financing systems - which influence policy-making. These contextual factors are so important that they should be mentioned, although we did not wish to discuss them in this paper. We have changed the paragraph accordingly.

   2. We agree that all non-comprehensive methods (including the balance sheet method) that rely on evidence and comparisons of alternatives are incomplete if they do not include all relevant alternatives. Although we discussed the problem, we thought that this point was almost self-evident. We have added the following sentences in the discussion: "Results from using the Balance Sheet Method will obviously change when other alternatives are introduced. All methods for priority setting that rely on evidence and comparisons of alternatives are incomplete if they do not include all relevant alternatives. What our study shows is that discussion of evidence and values may change priorities, and this may be an advantage also in real-world situations where the alternatives are many and complex."

2. We agree that all non-comprehensive methods (including the balance sheet method) that rely on evidence and comparisons of alternatives are incomplete if they do not include all relevant alternatives. Although we discussed the problem, we thought that this point was almost self-evident. We have added the following sentences in the discussion: "Results from using the Balance Sheet Method will obviously change when other alternatives are introduced. All methods for priority setting that rely on evidence and comparisons of alternatives are incomplete if they do not include all relevant alternatives. What our study shows is that discussion of evidence and values may change priorities, and this may be an advantage also in real-world situations where the alternatives are many and complex."

3. The notions of internal and external accountability are not central to the theoretical framework of this paper. In stead of discussing them in the introduction (as the reviewer suggests), we therefore decided just to revise the sentence by adding an explanation: "Requirements other than participation are probably needed for securing external accountability, that is, accountability to the public."

4. The criteria used in the balance sheet method overlaps considerably with those used by David Eddy (who developed the method), but the researchers made the final decision.

5. We chose homogenous groups because they are, in qualitative research, typically considered better to avoid dominance from single individuals with expert status. We agree, however, that mixed groups would
6. The reviewer comments that we did not ask respondents about re-allocations, and that this could be a major drawback of the tool. We disagree. There is nothing in the tool itself that prohibit rephrasing the question posed to the respondents - so that re-allocation is also considered.

7. We presented evidence on clinical outcomes based on results from meta-analyses and studies used in key CEAs. Performing a systematic search and evaluation of all possibly relevant clinical studies for all nine interventions is extremely resource and time demanding, and we therefore assumed that key studies had already been identified in such a manner in the meta-analyses and CEAs that we used.

8. The study was not designed for making conclusions about which criterion was considered most important - although this would have been very interesting. A key idea of our design was to test how a set of several criteria, when made explicit, could be used when ranking health care interventions. We did find, however, that overall ranking differed only slightly from ranking according to cost-effectiveness.

Detailed comments

1. Title is changed accordingly
2. Misleading part of sentence is removed.
3. Done
4. Sentence deleted
5. "clinical evidence" is changed to "evidence on clinical outcomes"
6. The phrase "combining evidence and values" should be clearer now after change in 5
7. Intro to paragraph is changed to "In particular, key research issues to consider when testing a method for combining evidence and values may include......"
8. We agree. The stated aim now reads: "We aimed at testing out the balance sheet method as a model for incorporating scientific evidence and public values in priority ranking of nine health care interventions in a setting of extreme resource scarcity."
9. Information added
10. Considered interventions are now listed, and software for safe water (educational tools, not plumbing etc) is explained.
11. Sentence rephrased
12. Sentence rephrased
13. Sentence deleted
14. Ok
15. Done
16. Sentence rephrased to avoid confusion. We did not mean WHO-CHOICE data here.
17. Done
18. Tanzania has a "non-discriminatory" TB policy. Sentence changed by adding parenthesis.
19. Yes, WHO-CHOICE inserted
20. Last sentence changed

B. Response to Thomson

1. We agree that we could have been clearer about the motivation for this study and possible applications of the approach. We have therefore added the following paragraph in the abstract: "Our use of the balance sheet method was meant as a demonstration project, but could if properly developed be feasible for health planners, experts and health workers, although more work is needed before it can be used for laypersons."

2. Analysis of the audio-recorded group sessions will be published elsewhere. We therefore followed Baltussen's suggestion not to mention this part of the method since it is not further utilised here. We included a short discussion: "Although the respondents gave reasons for their ranking and these would add understanding of the rankings, we do not report analysis of them here. This paper is the first exploration of this tool and its application. Details of the reasons and discussions during the deliberation will be presented elsewhere."

3. We also agree that the difference in rankings between professionals and lay people is interesting, but for clarity we chose to focus mainly on the effect of introducing evidence and discussion in the ranking exercise.
4. We have added a brief discussion of our search strategy: "A less narrow search strategy could also have changed our results, i.e. we could have included studies from outside sub-Saharan Africa, and cost-effectiveness studies with other outcome measures than life years or DALYs."

5. The presentation of evidence was made by quite detailed explanation of the information provided in the balance sheets. The information was written on blackboards or on a flip-over. We agree that the way this information was presented could significantly affect or bias the responses. We chose to use a less technical presentation of evidence when presenting to lay persons. We have edited this sentence: "Presentation of the evidence (the wording and use of technical terms such as risk reduction or odds ratios) varied according to the informant's level of expertise. Where appropriate, discussions were carried out in Kiswahili."

6. In particular we are aware that presentation of relative risk can be perceived as different from absolute risk, but not all studies we used had reported absolute risk reductions. Whenever possible, however, we also presented ARR's.

7. The nominal group technique has been used both with open and secret voting. We designed the study with open voting because we thought that would follow naturally after open debate. Both choices have advantages and disadvantages.

8. Excluded responses. We have added the following sentence under "analysis": "Eight groups were convened with a total of 85 respondents, but three responses (two from the village group), had to be excluded because of obvious misunderstandings and inconsistencies. We present descriptive statistics for the remaining 82 respondents."

9. We mainly presented results for all groups taken together. A separate presentation for each group was considered, but this would require many more figures or tables. And as mentioned above we made a decision not to focus on these differences here. We therefore added the following sentences under "analysis": "While aggregation of the ranks masks the individual- and group-differences, the purpose of this paper was to test the application of the tool. Detailed discussion of the differences between individual groups will be presented elsewhere."

10. We added this sentence under "discussion": "Many expressed the need for capacity building with respect to summarising the evidence base."

FORMATTING CHANGES

The following formatting changes have been incorporated into the manuscript:
- Unnecessary capitalisation in the title is taken out
- Authors' qualifications are deleted from the title page
- Sections on competing interests and author's contribution are added
- The table is now included in the manuscript file after the References
- The legends of the figures are now listed in the manuscript file after the References
- The legends are no taken out from the figure files

Yours
Ole Frithjof Norheim
(Corresponding author)