Reviewer's report

Title: Validation of Public Health Competencies and Impact Variables for Low- and Middle-Income Countries

Version: 4 Date: 23 August 2013

Reviewer: Rohit Ramaswamy

Reviewer's report:

1. Is the question posed by the authors well defined?

The question posed by the authors is somewhat confusing. My understanding from the paper is that the authors were seeking to validate impact variables and that they needed a consensus on the competencies in order to do so. The logical basis for this is not clearly articulated in the paper, because it is not clear how the impact variables were defined. Were the impact variables based only on the short list of competencies on which consensus was achieved? How were the impact variables created based on the competencies? And what is the relationship between the impact variables and the competencies? Is the fact that there was a high degree of consensus on the impact variables mean that the competencies are appropriate? It is not clear to me that this causality can be established. So it is not clear to me what exactly the authors are setting out to do and why.

2. Are the methods appropriate and well described?

The authors claim that the reason for the use of the modified Delphi method is that “public health competencies had already been developed but not yet validated across LMICs, so a reasonable degree of concurrence could be expected”. This is not a clear statement. If the competencies were developed by each country separately, then what is the basis for assuming that concurrence could be expected? In any case, the authors seem to have decided a priori that they would only have one round of consensus with the experts before they even started the process. While it is important to get alumni viewpoints, it seems like the decision to stop after one round should have been done after and not before the first round was completed.

Also I was not clear about the scale. The authors state that they use the Likert Scale of “poor” to “excellent”. The actual question asked is not reported, but I assume from the scale that the question asked was whether the competency was relevant or not. But the impact variable uses the same scale and what is being measured is not clear. For example, if an impact variable is “Developed a small scale study or research proposal: it is not clear what a value of “poor” means in this case. It also appears that the Likert scale used for the alumni is different from that used for the experts, and therefore the consensus reached by the experts and alumni may not reflect the same interpretation of the question.
Finally, the process by which the competencies were consolidated or modified is not clearly described. For example, the authors state that “we decided to group ‘analytical/assessment skills’ with ‘public health sciences skills’, as it was felt there was too much overlap between the two competency groups”, but these are two distinct sets of competencies, and without further explanation the justification for why these were deemed to overlap is unclear. The authors need to describe the process through which 76 competencies were reduced to 23.

3. Are the data sound?

The paper describes the criteria for how the experts and the alumni were selected, but since there is no detail about how these people were recruited, there is no information on whether these people are representative of the general stakeholder population, or whether there are characteristics that might cause their opinions to be biased. At the end of the paper, the authors state that since the MPH program convenors selected both the experts and the alumni, there is the possibility of bias. There is no discussion about what might have been done to reduce this bias.

The authors use a median of 4 or 5 for consensus, but a review of the data indicates that for almost all the competencies and the impact variables, the primary ratings were a 3, 4 or 5. While this indicates consensus, the lack of variation makes it difficult to interpret the results. Since all the raters gave most of the variables a high rating, it is not clear whether the consensus came about because every one agreed that the variables were important or whether there was pressure not to give assign low ratings. There appears to be some country to country variability – it would be useful to compute the median for each country for each rating and identify any variables that have a low median score in one country – this will help to identify any cross country variability that is not presented or discussed. Also in the rating of the impact variables, there are several qualitative comments about how difficult it is to rate the impact variables broadly without understanding the context and these were not reported in the discussion.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?

Since the primary objective of this paper is consensus, there was little reporting of data other than the consensus ratings.

5. Are the discussion and conclusions well balanced and adequately supported by the data?

Part of the discussion sessions focus on the addition/deletion of competencies. This is not a conclusion for the data, but more an elaboration of the methods

6. Are limitations of the work clearly stated?

Yes, though these limitations could be a major source of bias.
7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
   Yes

8. Do the title and abstract accurately convey what has been found?
   Yes

9. Is the writing acceptable?
   yes