Reviewer's report

Title: A rapid screening tool for psychological distress in children 3-6 years old: results of a validation study

Version: 1 Date: 11 October 2011

Reviewer: Mark Jordans

Reviewer's report:

Overall, the authors present a much needed study into the validation of a screening tool for young children in humanitarian settings. The conducted study is of importance in the effort to increase availability of contextually valid tools to assess psychosocial distress of children. While of importance, there are several key issues that need to be addressed.

Major compulsory revisions:

- The authors do not present a clear rationale for why one should conduct this type of screening, nor do they convincingly describe what happens after scoring above a validated cut-off score. This is of crucial importance, because it determines the need to go through the effort of development and validation of instruments in the first place. Not in the least place because of ethical concerns. One of the major criticism in the field has been that there is no use in screening if there are no clear interventions available that follow-up on the screening outcomes. The authors do not provide a clear framework for how to deal with this important public health criteria for screening. There are examples of programs for children in humanitarian settings that include systematic screening, yet this is always linked to subsequent service provision. The authors are invited to elaborate on this point. Of importance here is further the lack of clarity what intervention one might do for children with generic psychosocial distress of that young age (3-6 years).

- A second crucial point of needed revisions, in my opinion, is that currently the authors seem to combine three different objectives into one overarching objective of ‘validation’, i.e. adaptation of the instrument, concurrent validation and item reduction. However, these represent two separate and distinct phases and objectives: instrument development and adaptation (including the item reduction question and cultural adaptations question) and validation (assessment of the psychometric properties of the final instrument). In the present form, the paper combines these processes. Furthermore, if item reduction is aimed for, it is strange that the authors do not use any factor analyses in the process. Similarly, for the item adaptation question, the authors do not present clearly how they have gone about this (beyond the translational issues and FGDs – and even with the conducted FGDs it is not entirely clear how the data was used in the item adaptation process).

- While I think that the emphasize on non-diagnostic or non-disorder-focused
instruments is very valid in low-income settings where resources are unavailable to provide disorder specific care, it is still necessary to provide a framework for what is being assessed. The authors refer to generic psychological distress, but fail to define this. Furthermore, from the interpretation of the results it appears that it is a collection of domains including depression, PTSD and anxiety, however it is not clear whether these are subscales or whether it is a list of items that corresponds with symptoms from these diagnostic clusters. Nor is there information on how this was perceived for the original instrument. The authors refer to ‘several areas and domains of psychopathology’ but do not clarify this statement, and at other times refer to it as ‘traumatic exposition’ (e.g. Page 11). Finally, the authors refer to using the QGE scores to make inferences of psychological difficulties, followed by a list of specific disorders ranging from depression to PTSD (page 5). In addition, it is striking that the authors have not included a construct validity component to this study, nor make mention of the lack thereof in the limitations section (i.e. what is being measured?). This is important in light of the theoretical issue above, but also for methodological reasons (i.e. can one use composite total scores using this scale?).

- Finally, the use of the golden standard is not clear. The authors describe a combination of clinical interview and administration of the CGI-S, but it is not clear how the indication for need for psychological/psychiatric care was made. Furthermore, given that this is the golden standard it would be good to know how this method has been applied in other settings. Finally, it is unclear to me what type of psychological/psychiatric care the psychologist is screening for (again, especially given the age group 3-6 years old)? And while the presented AUC with sensitivity and specificity are presented, it would be good to also include other metrics of concurrent validity, such as positive and negative predictive values.

Minor compulsory revisions:
- In the results section of the abstract the authors refer to construct validity; however, this study does not assess construct validity, only concurrent validity. This should be corrected.

- A few times the authors present the conducted study as the first validation of a screening scale with cross cultural validation for use in a humanitarian context. That is not correct (given also the references that the authors use, i.e. Jordans et al.). The authors should correct this for the fact that this might be the case for the 3-6 years old age group.

- The authors report a method for assessing test-retest reliability that seriously allows recall bias. This should at least be included as a limitation.

- It would be good to include a table with overview of main validation results.

Discretionary revisions:
- Why was the tool initially developed in French?
- Why was language itself a reason for population selection, the fact that it is a language spoken in different parts of sub Saharan Africa does not entail that there are no cultural variations between these settings. I would opt for omitting this reason for sample selection. In addition, the second reason for choosing this
population is not clearly formulated and could deserve some reformulation.
- The sampling strategy makes mention of exclusion criteria being MR, development disorder and psychosis; but how were these identified?
- The first of the presented limitation seems more like a strength of the study.

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

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

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**
I declare that I have no competing interests