Reviewer’s report

Title: Comparing the accuracy of brief versus long depression screening instruments which have been validated in low and middle income countries; a systematic review.

Version: 2 Date: 30 August 2012

Reviewer: Charlotte Hanlon

Reviewer’s report:

COMPARING THE ACCURACIES OF BRIEF VERSUS LONGER DEPRESSION SCREENING INSTRUMENTS THAT HAVE BEEN VALIDATED IN LOW AND MIDDLE INCOME COUNTRIES; A SYSTEMATIC REVIEW

Reviewer response to revision

Thank you to the authors for their responses to my comments.

Response 1
I accept the inclusion of pregnant / postnatal women. I recommend including the sample type (an extra column) in Table 1 to aid interpretation of the generalisability i.e. whether male/female, pregnant, postnatal clinic, delivery ward, primary care, HIV facility, general medical in-patients, general medical out-patients). I note that the Stewart (Malawi) validation was conducted in postnatal women – infants 10 months old. It would be good to highlight this.

Response 2
OK

Response 3
OK

Response 4
As I noted previously, the Tesfaye study did use MDD as criterion (and presents ROC curves for minor and major depression). I believe it should, therefore, be included.

Response 5
The new point about weighting is not clear to me. Could additional explanation be added, perhaps in the methods?

Response 6
I still recommend including the point about ultra-brief screens not including questions about suicidal ideation as a potential disadvantage.

I am concerned about the interpretation that there is no significant difference
between brief and longer scales. Due to the heterogeneity you were unable to conduct sub-group analyses so I don’t know how this statement can be justified. See below.

Response 7
There are still problems with the references. I leave it to the editorial team to liaise with you about this.

New comments that are essential to address (major compulsory revisions)

It is inconsistent to say that you identified substantial heterogeneity and then go on to present a pooled estimate (and also to conduct sub-group analyses). With substantial heterogeneity the implication is that a meta-analysis cannot be carried out. Therefore I think the pooled estimate should be removed from the figure + table 2 should be omitted.

In the methods you need to be more specific about the diagnostic criteria that you accepted as ‘depression’. I presume you accepted DSM-IV and ICD-10 criteria but you also need to state which diagnoses were included e.g. was dysthymia included, was it ‘any depressive disorder’, etc. It would also be helpful to include in the presentation of the results.

You don’t appear to have given references for the gold standard measures i.e. for CISR, MINI, SCID, etc.

Page 3 – the EPDS is a 10-item scale (not 15-item) and so should not be considered as ‘long’. The Hanlon reference is not needed (please delete) – we validated using ‘common mental disorders’ as gold standard and so the paper is not relevant to your paper.

Editorial issues (minor essential revisions)
Page 4, para 2 – ‘plugged’ should be ‘plagued’ (I assume?)
Page 7 – add a % to the prevalence figures

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: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:
I am a collaborator with JJ, SM and DS on an NIH ‘hubs’ grant (a network of academic institutions across sub-Saharan Africa for improving mental health
intervention research), although am not directly working with any of these co-authors on a specific research project.