Reviewer's report

**Title:** comparing the accuracies of brief versus longer depression screening instruments that have been validated in low and middle income countries; a systematic review.

**Version:** 1  **Date:** 29 July 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

This is an important topic, particularly in light of the new WHO’s mental health Gap Action Programme which includes an evidence-based ‘intervention guide’ for the management of depression in primary care settings. Detection of depression in the PHC setting is a prerequisite for being able to deliver care and there is a question around whether or not training of PHC workers should be augmented by use of short screening scales. This paper has the potential to contribute but needs a lot of work.

Conceptually I don’t think it makes sense to include the validation studies for pregnant and postnatal women. Although these are women attending health facilities, they are not help-seeking for a complaint in the same way that people with undetected depression in PHC clinics might be. Also we know that the EPDS has different psychometric properties in pregnant and postnatal women. All this will lead to heterogeneity in the ‘accuracy’ of the scales.

**MAJOR ESSENTIAL REVISIONS**

Many well-conducted validation studies from PHC settings in LAMICs have been excluded because the criterion was ‘psychiatric morbidity’ or ‘caseness’ or ‘depression and anxiety’. Scales such as the GHQ and SRQ were developed to be able to detect such ‘common mental disorders’ rather than ‘depression’ per se, based on the evidence that anxiety, depressive and somatic symptoms frequently co-occur in the PHC setting. (See Goldberg and Huxley). It could be argued that the paper would have more real-world applicability if the authors included this broader concept of CMD rather than MDD. (If you dropped perinatal validations then this would be manageable).

Some studies that were included don’t fit the inclusion criteria:

1) Hong Kong is a high income country (the World Bank has classified it as high income for at least the last 25 years) and so the studies from this country should be excluded.
2) The validation study by Adewuya et al. is in a sample of students and not based in a health facility so this should also be excluded.

3) The validation study by Baggeley et al. is nested in a cohort study and I don’t believe it is a treatment-seeking group of women.

Some of the other studies aren’t clearly from health facility settings going on the title alone; please double-check. If you are broadening to the community setting there are many more studies that you should have included.

There are several facility-based validation studies that I am aware of from my knowledge of perinatal mental health in Africa and studies relating to the SRQ-20. I believe they should have been included. This raises concerns about the comprehensiveness of the literature search. The studies are:


(a ROC curve and validity coefficients are presented for a criterion of MDD)


The results section is a little thin, mostly just describing the studies without any analytic approach. Although a meta-analysis was not possible (due to the unsurprising heterogeneity), a narrative review of the systematically-ascertained studies would be useful.

Discussion

The references cited to make the point that the prevalence of depression seemed high in the included validation studies don’t seem to be appropriate – they are community samples where the prevalence of depression would be expected to be lower than facility-based samples.

The authors talk about the implications of their findings for screening in PHC but some of the included studies were from other settings.

Before concluding that EPDS is highly accurate, it would be worth skimming the literature. For example, in rural Ethiopia (a community sample) the EPDS had poor criterion validity whereas it worked fine in the capital city. This suggests contextual factors (e.g. mental health literacy, education) are important.
Some of the brief measures have the drawback of not including a question on suicidal ideation. For example, Kessler-6 and 10. This makes the scales less useful for the PHC setting.

The references need attention. The author names are not presented in the correct format.

**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 declare that I have no competing interests, except that I am co-investigator with JJ and DS on an NIH grant to develop hubs for mental health research in LAMICs. We are not directly involved in the same research component of the hub grant.