Reviewer's report

Title: Diagnostic utility of a one-item question to screen for depressive disorders: results from the MONICA/KORA study

Version: 2 Date: 12 September 2013

Reviewer: Tara Donker

Reviewer's report:

Review Manuscript BMC Psychiatry

This manuscript describes the diagnostic utility of a one-item question to screen for depressive disorders. The authors conclude that the single item screener is able to moderately decrease post-test probability of major depressive disorders, but that it may have limited utility in combination with additional screening tests or for selection of at-risk populations. Short screening questionnaires may reduce participants burden and increase willingness to undergo screening, while it may maintain adequate psychometric properties. The study has relevance to the field and is well written. The strength of the study is the large sample size and the appropriate analyses of psychometric properties. However, I do have several comments that the authors might wish to address.

Major compulsory Revisions

Background

1. The background is well-written and the research question is well-defined. The authors use a single-item screener developed by Williams et al. to assess diagnostic utility compared to the PHQ-9. However, a description of the development of this instrument is lacking. How was this measure developed?

Methods

Study design and subjects

2. Data in this study stems from the KORA F3 survey, but how participants were recruited is not described. Please give a short description of data recruitment and selection (inclusion/exclusion). Where did the standard interview took place? In a clinic, at participants 'home?

3. What was the delivery method of the single item screener? Face to face, self-rated?

4. Was the single item screener assessed within the standard interview? If not, what was the time frame between the two?

5. Was the assessor blind for the results of the single item screener and the interview? (If so, this is a major limitation and needs to be addressed in the discussion and abstract as well).

6. Were the screener and the interview always assessed in the same order? Could there be an order effect? (If so, please mention this in the discussion
section as well).

Instruments

7. Depression was assessed in an interview version of the PHQ-9. Who was the assessor? Was it a trained clinician, a student? How was the assessor trained?

8. PHQ-9 showed good sensitivity and specificity, please add the cut-off score for these coefficients.

Statistical analyses

9. Considering the impact of prevalence on sensitivity and specificity, the authors have carefully conducted analyses for different age groups and gender, and provided NPV and PPV measures as well, which are not influenced by prevalence rates. I would recommend to add short description of these validity measures (Sens, Spec, ROC) either in this section or in the results so audience who are unfamiliar with these terms can understand the manuscript (NPV, PPV, LR are adequately described in the result section already).

Results

10. Please add demographic and socioeconomic characteristics of the study sample in the result section. Only gender and age are presented in Table 1. More information on demographics/socioeconomic characteristics would be interesting (e.g. for interpretation of generalizability of results).

Discussion:

The discussion and conclusions are well balanced and adequately supported by the data. Limitations are adequately addressed. However, I do have some more comments.

11. My main concern is about blindness of assessors and participants. If the assessor/participants were not blind for the results of the interview (and the assessor for the result of the single item screener), this can have a major influence on the results and interpretation of the results.

12. The PHQ assesses depressive symptoms within the last 2 weeks, in line with a diagnosis of depression according to the DSM-IV or DSM-V. However, the Williams single item screener inquiries about the past year. Please elaborate on this subject.

13. Reliability of the single-item measure has not been tested in this study. Although Chronbach`s alpha cannot be assessed given the dichotomous outcome measure, test-retest reliability would have been a possibility. If a measure is not reliable, it cannot be valid. This should be addressed in the limitation section as well.

Minor essential Revisions

Abstract

14. Please add other relevant statistical analyses to the abstract as well (Likelihood Ration, AUC)

Please add to the abstract whether the PHQ is clinician-rated or self-rated
Sensitivity, specificity, PPV and NPV coefficients are usually reported in percentages instead of the current presentation of data (eg. Sensitivity 80% or sensitivity: 0.80) (also to the results).

Methods

15. Please add psychometric properties of the single item screener.

Results

16. Please add 95% CI to the coefficients reported, and the number of participants within each sub-analyses reported in the text as well, given the possibility of low prevalence of MDD.

Discussion

17. Please add information about the generalisability of results and the clinical impact of the study.

Minor issues not for publication

18. There are some typographical errors and grammatical throughout the manuscript

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

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**

I declare that I have no competing interests