Author's response to reviews

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

Authors:
Eva Blozik (e.blozik@uke.de)
Martin Scherer (m.scherer@uke.de)
Maria E Lacruz (elena_kt@yahoo.co.uk)
Karl-Heinz Ladwig (ladwig@helmholtz-muenchen.de)

Version: 3 Date: 28 November 2013

Author's response to reviews: see over
We thank the Editor and reviewers for their time and thoughtful critics of our manuscript. We have reviewed each reviewer comment carefully; our responses are directly beneath the original reviewer comment. Our changes in the manuscript are marked used the MS Word ‘Track changes’ feature.

Reviewers’ comments:

Reviewer #1:

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?

We added details on the development process of the single item screener (Please see Methods section, `Instruments`).

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).

A description of the recruitment process of the participants in the KORA F3 survey is described in the Methods section, paragraph `Study design and subjects`.

Where did the standard interview took place? In a clinic, at participants ‘home’?

We modified text to clarify that the interviews were done at the KORA Study centre in the city of Augsburg (Please see Methods sections, `Instruments`).

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

The one item screener and the subsequent PHQ-9 instrument were administered in a standard face-to-face interview. We clarified that (Please see Methods sections, `Instruments`).
4. Was the single item screener assessed within the standard interview? If not, what was the time frame between the two?

The PHQ-9 and single item screener were assessed in the same interview session (Please see Methods sections, `Instruments`).

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).

Both questionnaires were assessed in the same interview session, so the administrator could not have known the results of the depression screening (Please see Methods sections, `Instruments`).

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).

The single item screener was administered directly in advance to the PHQ-9 (Please see Methods sections, `Instruments`).

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?

We clarified that the interviews were performed by experienced study nurses. Before start of the study, they received an extended training program and were certified thereafter. All interviews were taped and subjected to a routine quality assessment in the KORA data center to avoid biases. At study halftime, all interviewers were recertified (Please see Methods sections, `Instruments`).

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

We are not aware of any established and widely accepted cut-offs for sensitivity or specificity. We therefore changed text by reporting the exact values (Please see Methods section, `Instruments`).

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).

We added an explanation of sensitivity, specificity and ROC to the Methods section (Please see `Statistical analyses`). For consistency, we also added a short description of PPV, NPV and LR.

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).

We added sociodemographic and clinical information of the study participants (Please see Table 1).

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.

Both questionnaires were administered during the same interview session. We clarified that in the Methods section (Please see `Instruments`).

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 inquires about the past year. Please elaborate on this subject.

Done (Please see last paragraph of Discussion section).

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.

Thank you for this important remark. We added this limitation to the last paragraph of the Discussion section.

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

Done.

Please add to the abstract whether the PHQ is clinician-rated or self-rated

Done (Please see Abstract, Methods section).

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).

Done throughout the manuscript.

Methods
15. Please add psychometric properties of the single item screener.
Sensitivity and specificity of the single item screener, as reported by Williams et al and Corson et al, are presented in the Introduction section. We added LR+ and area under the ROC as reported in the Corson publication (Williams et al only report sensitivity and specificity) (Please see Methods section, `Instruments`).

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.

Done (Please see Results section).

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

We added text to the last paragraph of the Discussion section to explicitly discuss the limitation of our study that participants stem from one geographic region only.

Minor issues not for publication
18. There are some typographical errors and grammatical throughout the manuscript

We carefully checked the manuscript for typographical and grammatical errors and removed them.

Reviewer #2:

The major strength of this manuscript is the large scale of the sample (over 3000 persons), although the capacity to generalize to the population as a whole is somewhat limited given that the data come from Bavaria only.

We now explicitly discuss the limitation of our study that participants stem from one geographic region only (Please see last paragraph of Discussion section).

The main research question addressed is clear and sound (can a single-item screen for depression be clinically useful in primary practice?). An important methodological limitation, however, is that the validity of the screen is assessed against estimates of the presence of mere depressive symptoms or major depression derived from a short (9-item) self-report questionnaire, which in itself has moderate sensitivity and specificity (~.88 in both cases). The authors rightly acknowledge in the Discussion that the specificity of the one-item screener may drop to as low as 66% when the criterion is clinically-assessed depression.

We agree with the reviewer that using the PHQ-9 as the reference standard is an important limitation of our study. We therefore discuss this issue in the last paragraph of the Discussion section. However, the PHQ-9 has been shown to have good concordance with clinical diagnosis of depression and is frequently used in clinical settings. Information on how the single items screener compares to a frequently used and well-known instrument may be helpful for deciding on the use of short questionnaires in clinical routine processes.

Paired with its low sensitivity (46% in the present study), it would mean that the response to a single question regarding presence of sad/depressed mood performs at near-chance levels.
We fully agree with the reviewer that the single item screener was able only to identify half of the persons with depressive mood which is far away from what a screening tool for depressive disorders should do. However, the picture is not consistent when different measures of test performance and different diagnostic standards (“depressive mood” vs. “major depressive disorder”) are evaluated. For example, sensitivity and specificity for detecting major depressive disorder of the single item screener were quite well. We are convinced that this detailed evaluation is helpful for persons searching for a very short instrument in very special situations such as specific trials with preselected patient populations.

Analyses are appropriate for the research question addressed. In addition to statistical indices of concurrent validity, however, I would like to see two 2x2 tables crosstabulating the numbers of participants who were screened positive and negative for (a) depressive mood and (b) major depression against those who replied “yes” and “no” on the single-item screener.

Done (please see new Table 2).

In view of these serious limitations, which are only in part addressed in the Discussion, the results do not support the validity of the one-item instrument in non-psychiatric practice.

We completely agree with the reviewer that the single item screener should not be used as a standard screening tool for depressive disorders in routine clinical practice. However, it may have limited utility for identification of populations that should undergo additional, more detailed evaluation for depressive disorders. We added text to the Conclusion section of the manuscript and the Abstract so that it becomes clear that routine use of the instrument cannot be recommended.