Author's response to reviews

Title: The performance of the K10, K6 and GHQ-12 to screen for present state DSM-IV disorders among disability claimants

Authors:

Bert Cornelius (l.r.cornelius@planet.nl)
Johan G Groothoff (j.w.groothoff@med.umcg.nl)
Jac JL van der Klink (j.j.l.van.der.klink@med.umcg.nl)
Sandra Brouwer (Sandra.Brouwer@med.umcg.nl)

Version: 3 Date: 23 December 2012

Author's response to reviews:

Dear Editor,

We thank both reviewers for their scrutiny in reviewing our revised manuscript. The performance of the K10, K6 and GHQ-12 to screen for present state DSM-IV disorders among disability claimants. We have carefully considered our responses to the comments of the first reviewer and hereby submit our revised manuscript. We feel that the additional recommendations substantially further improved our article.

If you require more information, please do not hesitate to contact me.

Thank you for your kind consideration and we look forward to hearing from you.

Sincerely,

Bert Cornelius Johan Groothoff Jac van der Klink Sandra Brouwer

# 1 Reviewer's report

Version: 2 Date: 13 November 2012; Reviewer: Genevieve Gariepy

Reviewer's report:

The manuscript reads much better, with notable improvements in the introduction and the discussion. Some additional comments and suggestions are provided below.

We thank this reviewer for her positive comment. In the following all comments are addressed point-by-point. Page numbers refer to the revised manuscript. In the revised manuscript, added and omitted text is indicated by Track Changes.
Comment 1.
The evidence for external validity of the instruments is not clear. External validity is “the ability to produce accurate predictions among patients not included in the development of the system but from the same population” (Justice et al, Assessing the generalizability of prognostic information, Annals of Internal Medicine, 1999). Although the prevalence of mental disorders in the sample is similar to those of the target population, this does not show external validity of the instruments per say. For instance, the study sample could have less severe symptoms than the target population. The distribution of other risk factors could also vary between the 2 groups, such as age and sex distribution. Methods to check for external validity include internal validation tests, such as bootstrapping, and external validation tests, such as testing cut-offs on different samples of the same target population. One suggestion is to drop external validity from the objective of the study. Information on the comparability of prevalence of mental disorders in the sample vs target population could be provided in the Methods section, under Setting and Procedure. This would reassure readers on the representativeness of the sample.

Response 1
As in the study sample, the target population consists of persons claiming disability benefit after 2 years of sickness absence. So, target population and sample are similar as to a very important risk factor for mental health problems, i.e. the duration of long-term sickness absence. Combined with the similar prevalence of mental disorder as prime cause of disability, we are fairly confident that our results are externally valid. Nonetheless, we agree with this reviewer that external validity is not proven. Therefore, we have replaced the term external validity by representativeness (page 5) and moved information on this to Methods, subheading Setting and Procedures. We have omitted a line on this in section Strengths and Limitations (page 12).

Comment 2.
In the discussion, the authors state that “both the K10 and the K6 outperform the GHQ-12 as to validity”. There is also a use strong language to highlight the inferior performance of the GHQ-12 such as “The GHQ-12 may not be suited for screening a population of long-term disabled persons [...]” and “Since the psychometric properties of the GHQ-12 are clearly inferior to those of the K10 and the K6, [...]”. However, it is not clear from results if the AUC of the GHQ-12 is
statistically significantly lower than the K10 and the K6, as confidence intervals overlap. And table 3 shows that the GHQ-12 cut-off may have better specificity than the 2 other instruments. The authors might consider conducting significance testing comparing the AUC of their instruments. Alternatively, the authors could use a more conservative language when comparing the performance of the instruments.

Response 2

We agree that the difference we found between the K10/6 and the GHQ-12 should have been formulated more conservatively, since CI’s indeed overlap indicating non-significant difference between the K10 and GHQ-12. We have changed our wording accordingly in Discussion (page 10, 11) and Conclusion (page 13).

Comment 3.

It is interesting that the threshold cut-off for the K10 is higher in the study sample than in samples from the general population and primary care. However, the explanations for this are not obvious. First, it is not clear why the presence of psychological factors in long-term disability explains the higher cut-offs. Please add 1 or 2 sentences to explain.

Response 3

To better explain the higher cut-offs, we have added the following to Discussion (page 11):

The importance of these psychosocial factors increase with the duration of sickness absence [26]. Therefore, distress found in the study sample may also be associated with psychosocial factors related to the sickness absence duration of two years, adding to the distress caused by the mental disorder itself.

Comment 4

Second, it is not clear how prevalence differences in outcome influences cut-offs. Cut-offs are determined from sensitivity and specificity which are insensitive to outcome prevalence (compared to NPV and PPV which are sensitive to outcome prevalence). An additional alternative explanation might be that disability claimants score higher on the K10 because of concurrent physical symptoms (e.g., fatigue, lack of energy) that are items on the K10.

Response 4

To better explain the relation between diagnostic test accuracy and prevalence we added in Discussion the following (page 12, with reference on page 20):

Although a higher prevalence does not systematically result in either higher or lower sensitivity and specificity, diagnostic test accuracy may vary with
prevalence [41]. The study sample with a higher prevalence of mental disorder may include more severe disorders, resulting in higher cut-off scores for the K10.

Minor Essential Revisions

Comment 5.
p.11, “However, selection bias is not likely, since we found no significant difference as to the prevalence of most frequent mental disorders [...]”. As discussed in point 1, similar prevalence of mental disorders does not guarantee generalizability, although it adds evidence for it. Authors could use “less likely” instead of “not likely”.

Response 5
We agree, see our response on comment 1.

Comment 6.
Table 1, the column heading should be n(%) and the table should include the number of subjects (n) in each category

Response 6
We added this to the table.

Comment 7.
Table 2, the title should read “... and in the total population of disability claimants”.

Response 7
We have added this to the table.