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: 2 Date: 2 November 2012

Author's response to reviews: see over
Dear Editor,

We thank both reviewers for their valuable comments on our manuscript *The performance of the K10, K6 and GHQ-12 to screen for present state DSM-IV disorders among disability claimants*. We have carefully discussed and considered our responses on all comments and hereby submit our revised manuscript. We feel that the recommendations of both reviewers have resulted in a substantial improvement of our article.

We welcome any further comments and feedback regarding our revised manuscript. 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

-----------------------------------------------------------------------------------------------

In the following we respond to all comments of each reviewer point-by-point. Page numbers refer to the *originally* submitted manuscript.

**Reviewer 1 (Genevieve Gariepy)**

**Minor essential revisions**

**Comment 1**
The introduction mainly provides general information on mental disorders and disability. However, the objective of the study is to investigate psychometric properties of the K6, K10 and GHQ-12; it is not to study the association between mental health and disability. Therefore, the introduction needs to make a stronger case for why it is important to investigate psychometric properties specifically in the population of disability claimants, particularly since the validity of the K10, K6 and GHQ-12 has already been studied extensively (see references 20-29 in the paper). For instance, do the authors hypothesize that individuals with disability have different cut-offs on screening scales of mental disorders than those from the general population? Is it thought that the validity of the instruments is different in disability claimants than in the general population?

**Authors response**
We thank the reviewer for this valuable comment and agree that the importance of studying psychometric properties of K10, K6 and the GHQ-12 needs stronger emphasis in *Introduction* (pg. 3, 4). Accordingly, we have added new paragraphs on psychometrics in a totally restructured introductory section, see the underlined paragraph *Introduction* in the revised manuscript.

**Comment 2**
Although under-recognition and under-diagnosis of mental disorders are highlighted in the introduction, it is not clear if this is because IPs fail to assess mental disorders or because screening instruments (K6, K10 or GHQ-12) are not
adequately used. Perhaps providing readers with some background information on how disability claimants are routinely assessed by IPs would be relevant to provide context.

**Authors response**
We agree with this comment and have added (pg. 5) the following to Setting and procedures:

In the Dutch social security system, one can apply for disability benefit after two years of continuous sick leave. Medical aspects of disability are then assessed by IPs employed by the Dutch Social Security Institute (SSI) in face-to-face interviews and examinations. For their assessment of diagnosis and treatment of the disorder(s) related to the disability claimed, IPs rely additionally in part on historic and actual medical data provided by occupational physicians who have assessed the sickness absence in the period preceding the disability claim.

**Comment 3**
Please provide more information on participant selection. It is not clear how the 1544 eligible disability claimants were selected. Please indicate whether they are the complete sample of eligible disability claimants in the province, a random sample of claimants or a non-random sample/other selection scheme. It is also not clear how participants were contacted to be invited into the study.

**Authors response**
We agree and have added (pg. 5) the following to Setting and procedures:

All persons claiming disability benefit at the SSI office in the city of Groningen in the period Oct 1st 2008 until Jan 1st 2010, were eligible to participate in the present study. As a result, all diagnoses were included, both mental and physical. The recruitment procedure was organised in two steps. As a first step, a SSI research assistant contacted eligible claimants by telephone asking permission to sent information about the study and a consent form. When permission was granted, name and address were given by the SSI assistant to the researcher, who then sent an information letter and a consent form as a second step. If eligible persons could not be contacted by telephone, the information letter and the consent form were sent by the SSI. Persons willing to participate returned signed consent forms to the researcher.

**Comment 4**
In the original GHQ-12, the four-point responses are: less than usual OR not at all, same as usual, rather more than usual, much more than usual. However, the manuscript suggests that the first response category is limited to “not at all” in this study. If this is the case, than the instrument is a different test than the GHQ-12. The authors should justify why they changed the response categories and express caution when comparing their results to those of other studies that have used the original GHQ-12.

**Authors response**
Indeed, our description of the 4-point response scale we used for the GHQ-12 is not entirely correct. We used the response categories of the original GHQ-12. Only questions 2, 5, 6, 9, 10 and 11 have a response category -not at all-. The other questions 1,3,4,7,8 and 12 have -better than usual- at the ‘better’ end of the scale.

In the section Methods, we added (pg. 6) the following to better describe our use of the GHQ-12:
(For the GHQ-12 we used the 0-1-2-3 scoring method with a four-point response scale: 'not at all' *(for questions 1,3,4,7,8 and 12: 'better than usual') (0), 'same as usual' (1), 'rather more than usual' (2), 'much more than usual' (3)).

Comment 5
In table 1 and in the text, it is stated that 25.9% of the sample has any mental disorder. However, 39.5% of the sample has anxiety disorders, therefore implying that at least 39.5% of the sample has any mental disorder. Please correct.

Authors response
We are very grateful for this comment and fully agree. These percentages were mistakenly copied from an earlier draft of the manuscript and are obviously not correct. We have corrected them in the text (pg. 8) and in Table 1: any (one or more) 30-day DSM-IV: 25.9%; any (one or more) DSM-IV mood disorder: 10.2%; any (one or more) DSM-IV anxiety disorder: 20.1%; any (one or more) substance use disorder: 2.7%. We have checked whether this error had any consequence for the results of our validity analysis and are happy to report that it had not.

Comment 6
If the number of people with "any mood disorder" is calculated by adding the number of people with each type of mood disorders, than the proportion should be 17.7% instead of 25.9%. Please check and correct.

Authors response
We have checked the prevalence of all specific 30-day DSM-IV classifications mentioned in Table 1 and found them correct. They add up to more than the (corrected, see our response to comment 5) prevalence of any DSM-IV classification due to comorbidity.

Comment 7
The category “adult separation anxiety disorder” should be dropped from the study since it is not a DSM-IV disorder.

Authors response
Although not listed in the DSM-IV, Adult Separation Anxiety Disorder is commonly looked upon as a relevant syndrome and the WMH-CIDI includes the assessment of it. Therefore, we are a bit reluctant to drop it. Instead, we added a short footnote to Table 1 on this matter.

Comment 8
The discussion could be significantly shortened. Discussion on the selection of cut-offs (bottom of p.10- top of p.11) from different economic perspectives is beyond the scope of the paper. The authors should instead report and discuss on how their results, such as cut-offs values, compare to others found in the literature.

Authors response
As we have pointed out in our paper, an extensive discussion on how to choose optimal cut-off values of the screeners is beyond the scope of our study. We have shortened text dealing with this (pg. 10). Still, we would appreciate drawing at least some attention to this issue, i.e. that there is no consensus on what cut points are optimal, as they depend on the test’s aim, costs and benefits.

Building on our totally revised introductory section (see also our response to comment 1 of this reviewer), we have added (pg. 11) to Discussion the following:
Since the psychometric properties of the GHQ-12 are clearly inferior to those of the K10 and the K6, we limit our discussion on how our validity findings compare to the literature to the K10 and the K6. We found the optimal cut-off score of the K10 to be 24 with sensitivity (SE): 0.724 and specificity (SP): 0.779, and of the K6 to be 14 (SE: 0.684 and SP: 0.770). As we point out in the introductory section, it is difficult to compare the validity estimates we found for the K10 and K6 with those found in other studies, conducted in other populations, using other interviewing methods as golden standards, assessing different sets of DSM-IV classifications with different time-frames and using different scoring methods. The optimal cut-off value (24) we found for the K10 is higher than found by Donker et al. (2009) in a Dutch primary care sample (optimal cut-off point 20; SE: 0.80; SP: 0.81) and by Fassaert et al. (2008) in a general population sample of ethnic Dutch (optimal cut-off point 16.5; SE: 0.792; SP: 0.768). It seems that in a population of disability claimants, the threshold for caseness is higher compared to the general population and primary care. This may primarily be based on population differences. First, it is well known that among long-term disabled persons psychosocial factors interplay with mental health related factors in sustaining long-term sickness absence and disability [26-28]. Second, the prevalence of mental disorder in our sample of disability claimants is much higher than found in other populations [39,40]. The optimal cut-off value (14) we found for the K6 almost equals the cut-off point found by Kessler et al. (2003) in a community sample, i.e. 13 (SE: 0.36 and SP: 0.96), while a higher cut-off point was to be expected. This may in our view primarily be explained by methodological differences: Kessler et al. used another structured psychiatric interview, assessing 12-month, not present state DSM-IV disorders and excluded substance-use disorders.

**Discretionary Revision**

**Comment 9**
P.13, please define RTW

**Authors response**
We agree and have clarified the abbreviation -RTW- by -return to work- (pg. 13).

**Comment 10**
The first three lines of the introduction do little to sale the study and could be dropped.

**Authors response**
We agree, see our response to comment 1 of this reviewer.

**Comment 11**
Table 3 may be dropped since results are already written in the manuscript.

**Authors response**
We agree. We dropped Table 3.

**Comment 12**
It is not clear how specific phobia, such as spider phobia, is relevant to work disability. Perhaps distinguishing between types of phobia was not possible. This may be a limitation of the study.

**Authors response**
WMH-CIDI allows to distinguish between different types of phobias. Since some specific phobias are relevant to work disability (heights, water, seeing blood/injury), we have added to Table 1 prevalence of 30-day social phobia (6.1%); of agora phobia (3.0%) and of specific phobia (7.2%).
Comment 13
The strengths and potential limitations of the study are well described. Another potential limitation: since the K6 was embedded within the K10, it is possible that results could be different had the K6 been administered alone, as neighbouring questions of a questionnaire item may influence the way individuals respond to the item.

Authors response
We agree. We have already hinted at this limitation in the last three lines of the section Discussion. We have added (pg. 12) the following:

It is possible that results could have been different had the K6 been administered as stand-alone.

Reviewer 2 (Jennifer Green)

Major compulsory revisions

Comment 1
The authors describe the use of mental health screeners in disability claims cases. However, I think it would be important to include a discussion of how the results in this disability claimant study compare to validity estimates in other populations and how the cut-points established here differ from cut-points previously established. This seems particularly relevant in the choice of the K6 cut-point of 14, which differs only marginally in its dichotomous AUC value ((SN+SP)/2)) from the Kessler et al. (2003) K6 cut-point of 13 in a community sample.

Authors response
We are grateful for this valuable comment and added (pg. 11) a paragraph on this in Discussion, as already noted in our response to comment 8 from reviewer 1.

Comment 2
More information is needed about sample selection and study design. First, the authors describe this study as part of a larger prospective cohort study with one-year follow-up. Were these data collected as part of the initial wave, or in the follow-up? What was the purpose of the larger study and might that have influenced the respondent sample? The authors note that there were a total of 1544 "eligible" disability claimants. What were the eligibility criteria? How do the 293 participants in this study compare to the 1544?

Authors response
We agree that more information is needed on recruitment and study design. See also our response on comment 3 of reviewer 1. We added (pg. 5) to Methods:

(Data) were collected in the initial wave (of a larger prospective cohort study)

We realize that the low response rate of 24.3% makes the question very relevant whether the external validity of our results is compromised by selection bias. We wanted our results as to the performance of the screening questionnaires to be valid for a larger population. However, we were not so keen on comparing our sample with the population of 1544 eligible disability claimants in the province of Groningen, but were much more interested in whether our results could be generalized to the population of disability claimants in the whole of the Netherlands, i.e. the population we wanted the screeners (if proven valid) eventually to be used in. Data on diagnostic categories as cause of disability are readily available from the national Social Security Institute. So, we don’t know how the 293 participants in our study compare to the 1544 eligible claimants.
Instead, we thought it more relevant to report on how our sample compares to the national population and that no significant differences as to diagnostic registry codes were found.

**Discretionary Revisions**

**Comment 1**
The authors note that the imperfect overlap of the 30-days assessed by the CIDI and 30-day screeners could potentially decrease validity estimates, but are unlikely to because of the nature of chronic disability conditions. To test whether this is the case, the authors could also compare the screening scales to 12-month CIDI DSM-IV diagnoses. This comparison may, actually, be a more important one than comparison with the 30-day CIDI diagnoses, because it would provide information about the extent to which screeners identified disorders affecting claimants in the past year, which is perhaps a more useful time-period for service planning.

**Authors response**
Following the valuable suggestion to test whether imperfect overlap influenced validity estimates, we compared both K10 and K6 sum scores with 12-month DSM-IV classifications. For the K10, we found an optimal cut point of 23, equalizing a SE of 0.649 and a SP of 0.842, with an AUC of 0.798. For the K6, we found an optimal cut-off value of 13, with SE 0.746, SP 0.711 and AUC 0.787. For comparison: corresponding estimates for 30-day classifications are: cut point 24; SE: 0.724; SP: 0.779; AUC: 0.806 (K10) and cut-point 14; SE: 0.684; SP: 0.770; AUC 0.796 (K6). This illustrates our assumption that imperfect overlap of recall periods have only very marginally influenced K10 and K6 validity estimates. We have added (pg. 12) this information to the paper:

*To test this assumption, we compared the K10 and K6 sum scores with 12-month DSM-IV classifications present in the year preceding the interview. For both the K10 and the K6, we found validity estimates for 12-month classifications only to differ marginally from those for 30-day classifications, showing our assumption is likely to be right (K10: optimal cut-off point 23; SE: 0.649; SP: 0.842; AUC:0.798; K6: optimal cut-off point 13; SE: 0.746; SP: 0.771; AOC:0.787).*

As to the second part of this comment 1: in most social security systems, social insurance physicians are not focused on service planning, but on whether a claimant should be awarded disability benefit. To be awarded benefit, the presence of at least one health disorder is required at the moment of claim assessment. While a time frame of 12 months for the prevalence of DSM-IV classifications may be more useful for service planning, for the assessment of functional impairment and disability within the legal context of disability benefit claim assessment after two years sickness absence, social insurance physicians are by law primarily focused on present state disorders as cause of present state disability, not on disorders, mental or physical, that may have caused disability in the past. That is why we focused on 30-day disorders.

**Comment 2**
Readers may benefit from a discussion of whether to use the K6 or K10 – in terms of the balance of screener length vs. validity results.

**Authors response**
We agree. We added (pg. 13) the following:

*The K10 and the K6 are both very short scales and both take only a few minutes to administer. While the validity of the K10 is slightly better than that of the K6, we advice to use the K10 instead of the K6 with cut-off values suitable for this particular population.*
Minor essential revisions

Comment 1
Abbreviations such as IP and RTW should be defined.
Authors response
The abbreviation IP (insurance physician) was already defined (pg 3). We have clarified the abbreviation -RTW- by -return to work- (pg. 13).

Comment 2
On page 7 and in Table 2, the authors note that ICD-10 classifications are “a primary cause for disability.” Are they the only cause of disability, or one of many? What are the other causes of disability?
Authors response
We agree that this needs clarification. We have added (pg. 7) to the section Administrative data:

In the Dutch social security legislature, to be granted benefit, the disability must be caused exclusively by a medical condition, whether somatic or mental.

And:

The registry of the Dutch Social Security Institute allows one diagnosis code for any (somatic or mental) disorder as primary cause of disability, and two additional codes for any comorbid disorders as secondary or tertiary cause of disability.

Comment 3
On page 9, it would be helpful to clarify that the analysis of internal consistency is again using the sample of 293. Related to this, the sample size should be included in Tables 3 and 4 for clarity.
Authors response
We agree and have added (pg. 9) this to the manuscript. Table 3 was dropped, see comment 11 of reviewer 1.

Comment 4
In Table 1, types of mood disorders add up to 17.7%, but any mood disorder is listed as 25.2%.
Authors response
We agree. See our response to comment 5 and 6 of reviewer 1.

Comment 5
“Educational level”, included in Table 1, should be defined.
Authors response
We agree and have added this information in a footnote to Table 1.