Author's response to reviews

Title: To what extent does the anxiety scale of the Four-Dimensional Symptom Questionnaire (4DSQ) detect specific types of anxiety disorder in primary care? A psychometric study

Authors:

Berend Terluin (b.terluin@vumc.nl)
Desiree B Oosterbaan (d.oosterbaan@psy.umcn.nl)
Evelien PM Brouwers (e.p.m.brouwers@uvt.nl)
Annemieke van Straten (a.van.straten@psy.vu.nl)
Peter van de Ven (p.vandeven@vumc.nl)
Wendy Langerak (wendy.langerak@uwv.nl)
Harm WJ van Marwijk (hwj.vanmarwijk@vumc.nl)

Version: 2
Date: 25 March 2014

Author's response to reviews: see over
Dear Dr. Murray,

Herewith we would like to submit a revised manuscript (nr. 1309410465113617). In response to the reviewers’ comments, we have added more detailed information on the ROC analyses and a discussion of the background and merits of using two cut-offs instead of one. We hope we now have met with the reviewers’ concerns but we are open for further comments. We will address the comments point-by-point.

**Reviewer #1: Keith Bredemeier**

Overall, I feel that the authors have done a commendable job addressing the critiques and recommendations from the initial round of reviews. As a result, I feel that this version is both clearer and stronger than the original submission. More importantly, I am inclined to think this manuscript would make a worthwhile contribution to the literature. Nevertheless, I do have a few additional comments/suggestions, all of which I consider minor, and only the first two I consider essential for the authors to address.

Minor Essential Revision:

1) The last three sentences of the fourth paragraph of the Introduction (which begins “The 4DSQ anxiety scale employs two cut-off points...”) seems a bit confusing/misleading. Specifically, it sounds from the current wording like the authors are saying that there are existing, established cut-offs on this scale. My understanding is that this is not true, and if that is the case, I think the authors should clarify the wording in this section (e.g., by saying that the scale could employ two cut-off points).

*Reply: In 2002 two cut-off points have been assigned to the 4DSQ anxiety scale, based on clinical experience. We have added this information on page 6. The present study provides the opportunity to revise these cut-offs and base them on clinical evidence.*

2) I do not understand how the authors arrived at the 61% estimated “predictive value” provided in the last sentence of the second to last paragraph of the Discussion – I think they need to explain how they arrived at this number, or get rid of this point (my sense is that this estimation may misleading anyway, since it is likely based upon a number of assumptions, though it is hard to say without more information about how it was determined or computed).
Reply: We have added the estimation of likelihood ratios (LRs, Table 7) as well as an explanation of how knowledge of LRs helps to predict the probability of having a disorder in populations with different prevalences. The estimate of 61% is now in Table 7 (lower-right cell).

Discretionary Revisions:

3) I understand that obtaining secondary ratings on the interviews in order to compute interrater reliability was not feasible. While the reliability and validity of the instruments themselves does provide some assurance, it doesn’t guarantee that the interviewers were reliable. Thus, I would suggest that the authors acknowledge the lack of evidence for interrater reliability as a limitation of their data.

Reply: We have acknowledged the lack of information about interrater reliability of the diagnostic interviews in the Discussion section (p. 24) and included a short discussion on its relevance for the interpretation of the present results. In addition, we have now provided information on the reliability of the anxiety scale (p. 14-15).

4) I think that the authors’ response to the second reviewer’s suggestion to cross-validate the results of the ROC analyses is reasonable. That said, given the size of the total sample, this recommendation does still warrant consideration. I would argue that the primary concern regarding sensitivity and specificity is not inflation per se, but rather generalizability across samples/populations. On that note, an alternative to consider is to present the results from the ROC analyses both combined/pooled AND separately for each of the samples (perhaps in a table).

Reply: We have complied with this suggestion and supplied the results of the ROC-analyses both for the separate studies and the pooled studies in a new table (Table 5). There was quite some heterogeneity across the studies but this is, in our view, not a generalizability issue. Rather, the heterogeneity seems to be due to differences between the study samples and the relative instability of ROC-results in relatively small samples. We have discussed this in the Discussion section.

5) The organization of the “Anxiety Score Distribution” section (in the Results) is a bit confusing, and I wonder if it could be streamlined a bit (which alone would probably make it easier to follow).

Reply: We apologize for not seeing how this section could be further streamlined.
6) The authors might consider presenting the “ROC Analyses” section before the “Anxiety Score Distribution” section (again in the Results), as it seems a bit confusing to reference “low” and “high” scores based upon cut-offs that have been introduced or justified yet.

*Reply:* We do not agree with Dr. Bredemeier here. We feel it is important to maintain the “Anxiety Score Distribution” section before the “ROC Analyses” section because the former section shows that the anxiety score is less able to detect GAD and specific phobia. On that basis, we decided to limit the ROC-analysis to those anxiety disorders that are best detected: panic disorder, agoraphobia, social phobia, OCD and PTSD. Our consideration was that there is little point in determining best cut-off points for disorders (GAD and specific phobia) that are relatively poorly detected by the 4DSQ anxiety scale. Rather, we chose to let the scale do what it can do best.

7) I would recommend getting rid of the sentence in the first paragraph of the Discussion that starts “The specific anxiety disorders, as conceptualized in DSM-IV…” – this point seems to go beyond the data presented in this manuscript.

*Reply:* We agree and we have discarded the sentence.

8) I still feel that the clinical recommendations provided by the authors in the third paragraph of the Discussion regarding patients who score in the moderate range (previously described as “watchful waiting”) seem a bit narrow or oversimplified. Further, I agree with the other reviewer that the data from this paper don’t directly speak to the stability of the scores from this measure, and thus the suggestion that false positive scores might “abate” after a few weeks seems speculative. While I recognize the utility in providing a very specific set of concrete instructions about how to use such cutoff scores, my recommendation is to provide a slightly more nuanced and/or tentative discussions of how physicians could respond to patients who score in this range (e.g., administering other screening measures, conducting a more circumscribed follow-up diagnostic assessment) and how this might be influence by practical considerations (e.g., the availability of diagnostic specialists to conduct follow-up assessments).

*Reply:* We have acknowledged that there is currently no firm evidence supporting our recommendations, but that they are based on clinical experience in general practice (p. 22-23). We have added a discussion of the meaning of moderate anxiety scores.
9) I still find Table 2 a bit confusing when viewed in isolation, and I suspect most readers would as well. The description the authors added to the text certainly helps, but I wonder if the authors might be able to provide some clarity in the title of the table, or even a note underneath. Alternatively, I think the authors could consider getting rid of this table altogether, as I would argue that these data are marginally relevant to the paper (at least, at the level of detail provided in the current version of the table).

*Reply: We feel that information on the way and extent to which the anxiety disorders are interwoven with each other and with depressive disorder is vital for understanding the performance of the 4DSQ anxiety scale. We have rephrased the caption to improve the clarity of Table 2.*

Reviewer #2: Thomas Ehring

The authors have generally been responsive to the comments and suggestions from the first round of reviews. However, some concerns remain, as detailed below.

*Major compulsory revisions*

(1) In their response to my first comment on the original manuscript (inconsistency between stated study aims and actual analyses), the authors clarify the aims of their study. However, in the manuscript there are still a number of sentences that in my opinion are easily misunderstood. Please make clear throughout the manuscript that the study is NOT about distinguishing between different anxiety disorders (which the 4DSQ is not able to do) but about distinguishing anxiety disorders from non-anxiety disorders, whereby the performance of the 4DSQ is tested for each disorder separately. The sentences that I think should be modified include:
- “To what extent does the anxiety scale detect each of the specific DSM-IV anxiety disorders” (p.3 = Abstract)
- “Research has shown that general practitioners (GPs) do recognize mental health problems in most of their anxiety disorder patients but they experience difficulty in establishing specific anxiety disorder diagnoses. A solution to this problem might be the use of a case finding instrument.” (p. 5)
- “As noted above, GPs usually recognize most patients with anxiety disorder as having emotional problems but they fail to recognize that some of these patients have specific anxiety disorders in need of specific treatment.” (p. 7)
Reply: We were not aware that the confusion still existed. We have rephrased the sentences mentioned. We have changed the first sentence in agreement with the manuscript title. In addition, we have explicitly added to the Abstract conclusion that the 4DSQ anxiety scale is “not able to distinguish between the different anxiety disorder types”. We have changed the second sentence into an almost exact citation of reference [3] and added to the next sentence that the case finding instrument should distinguish between patients with high risk of having an anxiety disorder and patients with low risk. We have changed the third sentence into: “As noted above, GPs usually recognize non-specific emotional problems in patients with an anxiety disorder without recognizing that these patients actually have an anxiety disorder that needs specific treatment. The 4DSQ, as a case finding instrument, could assist GPs in separating patients with high risk of having an anxiety disorder from patients with low risk.” In addition, we have added to the Conclusions that “It should be noted that the 4DSQ anxiety score is not able to distinguish between the separate anxiety disorder types.”

(2) In my third comment on the original manuscript, I highlighted the fact that either sensitivity or specificity appear unacceptably low (.47-.52) for the different cutoffs suggested for the 4DSQ. In their reply, the authors argue that this is not problematic as two different cutoffs are presented that show either high sensitivity or high specificity. I disagree. If the higher cutoff is used, the very low sensitivity leads to a large number of cases in need of immediate treatment being missed as they now fall within the middle category described by the authors. The low specificity of the lower cutoff, on the other hand, means that too many individuals are falsely screened positive, which leads to unnecessary further assessments and/or referrals. Thus, the low sensitivity and specificity values respectively come at a high cost in clinical practice. I therefore still think that this limits the usefulness of the screener in clinical practice, which should be discussed. Specifically, I suggest that the authors

(a) report positive predictive power and negative predictive power in addition to sensitivities and specificities as these will provide important information with the regard to the issue described above;
(b) critically discuss the costs that are related to using cutoffs with either low sensitivity or low specificity;
(c) discuss possible reasons for the occurrence of either low sensitivity or low specificity. Although the authors are of course correct in stating that the selection of cutoffs always means a tradeoff between sensitivity and specificity, it is on the other hand of course not impossible to develop screening instruments that show high sensitivity and at least acceptable specificity (or vice versa). In my opinion, one reason for the current findings is the fact that the samples comprised treatment-
seeking individuals only. In such a situation, it is much more difficult to find high sensitivity and specificity than when distinguishing between clinical vs. non-clinical groups.

Reply: We are afraid that we continue to disagree about the unacceptability of the low sensitivity/specificity values connected to the higher and lower cut-offs that we propose. It has stimulated us, though, to discuss more extensively the background of our recommendations to use two cut-off points instead of one. We are certain that the selection of more extreme groups would yield better sensitivities and specificities, but the relevant population for the 4DSQ anxiety scale is a help seeking population. In response to professor Ehring’s specific suggestions, we have (a) reported positive predictive values, negative predictive values, sensitivities and specificities for all cut-off points in the individual studies and the pooled sample (Table 5), (b) discussed the costs related to using two cut-offs instead of one, and (3) discussed possible reasons for low sensitivity and low specificity extensively in the Discussion section (pages 20-23).

(3) In my fourth comment to the original manuscript, I argued that it is problematic when cutoffs are empirically derived and tested within the same sample. I therefore suggested means to cross-validate the findings. In their response to my comment, the authors argued that they cross-validated the operational characteristics of their screener in the four different samples. I would like to highlight that this is only true for the test of the dimensionality of the measure (Table 3). However, when it comes to information on diagnostic properties (i.e., sensitivity and specificity) of the 4DSQ for identifying anxiety disorders only results for the total sample are presented, but not for the separate subsamples. I would therefore like to reiterate my original comment that cross-validation within the current study appears desirable as we currently do not know whether the cutoffs established in the current study work equally well in the different subsamples (and for that matter even more so in independent samples).

Reply: We have added all information on the separate studies, revealing much heterogeneity across the studies due to differences in prevalence and severity spectrum and the fact that ROC analysis is unduly sensitive to chance irregularities in the distribution of test scores, especially in relatively small samples. In the end, we had to sensibly choose a set of cut-off points, assuming that the truth was to be found somewhere “in the middle”. Only by pooling the samples, it was possible to obtain fairly stable results.

*Minor essential reviews
The authors may want to briefly acknowledge somewhere in the manuscript that PTSD and OCD are not classified as anxiety disorders any more in the DSM-5.

Reply: We have added a few words on this issue at the end of the Discussion section.