Reviewer's report

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

Version: 1 Date: 23 December 2013

Reviewer: Thomas Ehring

Reviewer's report:

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)

(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.

(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).

*Minor essential reviews

(4) 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.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests.