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: 7 December 2013

Reviewer: Keith Bredemeier

Reviewer’s report:

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).

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).

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.

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).

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).

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.

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.

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).

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).

**Level of interest:** An article whose findings are important to those with closely related research interests

**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