Reviewer's report

Title: Is the Beck Anxiety Inventory a good tool to assess the severity of anxiety? A primary care study in The Netherlands Study of Depression and Anxiety (NESDA)

Version: 1 Date: 17 June 2011

Reviewer: Paul M Kohn

Reviewer's report:

Major Compulsory Revisions

I have no such revisions to suggest.

Minor Essential Revisions

The following are changes which I believe are necessary or highly desirable:

1) The sentence in the Results section of the Abstract which refers "co-morbid anxiety or depression" is ambiguous and potentially subject to misinterpretation. I would rephrase into two sentences, one one on co-morbid anxiety vs. anxiety which is not co-morbid and one on co-morbid depression vs. depression which is not co-morbid.

2) The final sentence in the Abstract should specify clinical populations because the Beck Anxiety Inventory (BAI) does appear to distinguish effectively between anxiety and depression in nonclinical populations.

3) When you state the test-retest reliability of the BAI, you should specify the test-retest interval to which this finding applies.

4) Your use of "p = 0.00" is misleading. When SPSS shows a p value of .000, this does not mean that the sample result is impossible under the null hypothesis. It means rather that the p value is < 0.0005. (If it were 5 or greater at the fourth decimal place, p would round up to 0.001.)

5) Reference to anxious and depressed patients’ having equally high BAI scores is inaccurate at both the sample and population levels. First, the sample values of 13.34 and 13.94 are very close but not identical. Second, the absence of a significant effect does not mean that the null hypothesis is literally true at the population level. There is such a thing as a Type 2 error. One can fail to reject the null hypothesis, but one is never in a position to accept it.

6) On p.16, para 2, the word, "population", appears incorrectly twice where the word, "sample", properly applies. A population is the exhaustive set of elements from some category, e.g. all Bipolar 2s in the Netherlands, whereas a sample is a selected subset from the corresponding population.

Discretionary Revisions

In my view, the analysis of data is suboptimal. The authors report using one
multiple regression for each comparison, statistically controlling for age and gender. First, the mechanics are unclear: Was there prior hierarchical entry of age and sex or was there simultaneous forced entry of these variables with the immediate comparison under consideration? Second, was "statistical control" even justified? I see two possible justifications, age and sex were differentially distributed over values of the primary comparison variable (hence justification as a statistical control against confounding) or age and sex had significant effects (hence justification by reduction of error variability). (Interestingly, the authors did not report on either tests for the significance of the associations between age and sex on one hand and the main variables of interest on the other.) Third, analysis of variance (ANOVA) or analysis of covariance (ANCOVA) would provide what is presently lacking: an omnibus test of the relation between diagnostic category and BAI score. ANCOVA could use age as a covariate and gender as a second independent variable. A significant F for diagnostic category (which I'm fairly sure there would be) would occasion either planned orthogonal comparisons or more likely post hoc comparisons which approximate a constant experimentwise alpha, e.g. via Tukey's Honestly Significant Difference test. Also, the output for ANOVA or ANCOVA would facilitate size-of-effect estimates via eta-square.

Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.