Reviewer's report

Title: A cross-validation of the Provisional Diagnostic Instrument (PDI-4)

Version: 2 Date: 17 August 2012

Reviewer: Kurt Kroenke

Reviewer's report:

This is a useful study evaluating the operating characteristics of the PDI-4, a recently developed brief screen for 4 common mental disorders. Strengths include a good sample size, appropriate psychometric analyses, and an articulate presentation of study methods, results and discussion. The major limitation is the estimation of PDI-4 operating characteristics (sensitivity, specificity and predictive value) against other rating scales rather than a criterion standard diagnostic interview. Even with this limitation, however, the data in this study provides valuable data about how PDI-4 scores correspond to well-established comparator scales, the substantial overlap of symptom comorbidity among the 4 disorders studied, and the strengths and caveats regarding future use of the PDI-4. Other comments are relatively minor.

Minor Essential Revisions

1. On p. 8, it is stated: “In addition, the sensitivity, specificity, positive predicted value, and negative predicted value of each PDI-4 subscale, using the corresponding patient scale as the gold standard (as formal diagnoses were not performed in this study), were also computed.” Since the corresponding scales used in the study are not criterion standard structured diagnostic interviews, the sentence above needs clarification. Were specific cutpoints used on the corresponding scales and, if so, what were these cutpoints. Were they the cutpoints provided in the following paragraph? If so, information provided in these 2 paragraphs needs to be better integrated so what is meant by “using the patient scale as the gold standard” is clear.

2. Related to point #1, the authors should report, in the Methods section, the sensitivity and specificity for the diagnostic cutpoints for the comparator scales. For example, what is the sensitivity and specificity of a PHQ-9 score # 12 for major depression, of a MDQ # 7 for mania/bipolar, etc.

3. There is good data on the operating characteristics of PHQ-9 cutpoints for major depression compared to structured diagnostic interviews. However, have HADS-A cutpoints been used specifically for generalized anxiety disorder, or instead for an anxiety disorder in general. If the latter (rather than specifically GAD), this should be noted. Likewise, have CAARS-S cutpoints been tested against structured interviews for ADHD and bipolar disorder, respectively. Noting these points in the Methods is important when trying to interpret the data on sensitivity, specificity and predictive value.
4. Related to point #1 above, cutpoints on the non-PDI-4 validating scales have imperfect sensitivity and specificity, so there are very important limitations in calculating sensitivity, specificity and predictive value against these measures which themselves are not criterion standard diagnostic measures. Thus, I have several recommendations so this point is clear to the reader.

a. For Table 4, I would title column 1 “PDI-4 Scale”. I would then delete the corresponding scale currently listed in parentheses. Instead, I would add a second column that “Comparator Scale Diagnostic Cutoff”. Then, I would put the diagnostic scale and cutpoint in this column (e.g., PHQ-9 # 12, HADS-A # 14, etc.). Then I would consider adding an overarching column heading over the next 4 columns and title it something like “Operating Characteristics of PDI-4 Scale Using the Comparator Scale Diagnostic Cutoff”.

b. Be very careful using terms like “met diagnostic criteria for”. Thus, the last sentence of the first paragraph of the Results on p. 9 should be revised to read “For instance, among patients with a provisional GAD diagnosis based on the HADS-A, 89.5% also equaled or exceed the diagnostic cutoffs for 1 or more additional conditions (MDE by the PHQ-9, ADHD by the CAARS-S, and/or mania by the MDQ). Indeed, 82% met attained or exceed the PHQ-9 cutpoint for MDE.”

5. On p. 9, it is stated: “Thus, 2 follow-up exploratory factor analyses were conducted: a standard analysis using only PDI-4 items and an analysis including items from the PDI-4 and the corresponding scales.” I am confused about the second analysis. Does this mean items from on the PDI-4 subscales and each of the other 4 corresponding rating scales was pooled? If so, please clarify this.

6. On p. 9, it is stated: “The comparative fit indices were inconclusive (Bentler’s Comparative Fit Index = 0.92; the Goodness of Fit Index = 0.90; Bollen Normed Index = 0.89), suggesting a borderline fit of the model.” For the reader unfamiliar with these fit indices, what scores on these indices would be considered “conclusive”. I would provide this information to the reader to show how “close” each index was to being conclusive.

7. Table 5 needs clarifications on several points:

a. From which model were the F-values for gender, age, and ethnicity derived (the model with Cohort A, Cohort B, or Cohort C)?

b. The last sentence in the footnotes is “Cohort term for interactions is the main Cohort definition of with vs. without the provisional diagnoses.” What interactions are referred to here? This should be specified.

c. Also, what does the Omnibus F-test refer to? This should be clarified in a footnote.

d. Also, the results of Table 5 need more explanation in the Results rather than mentioned in a single sentence. This is the most complex table in the paper, and needs explanation to walk the reader through its key findings.

8. Figure 1 legend needs more detail to adequately explain the Figure to the reader. For example, do the labels on the x-axis (and corresponding bars) refer
to the particular disorder as defined by the comparator scale cutpoint (instead of self-identified patient groups). In other words, does “Anxiety” mean those with a HADS-A # 14, and “Depression” those with a PHQ-9 # 12, etc.? Likewise, do the “additional diagnoses” (and corresponding percentages) in the red part of each bar represent diagnoses according to the corresponding scale cutpoints. I presume this is the case, but make this clearer in the Figure legend.

9. Predictive values will be inflated in this particular sample because the majority of patients had a self-reported physician diagnosis of one of the 4 disorders. Thus, positive predictive values would be considerably lower in a primary care or population-based sample (where prevalences of disorders would be much lower), and negative predictive values would be somewhat higher. This point needs to be emphasized in the Discussion.

10. I would interpret the factor analysis findings in a slightly more favorable light. For example, the fact that exploratory factor analyses suggested a potential 5-factor structure is not inconsistent with the fact that depression and anxiety are highly comorbid (which is why DSM-V often considers included a mixed depression-anxiety disorder), and that ADHD (by its very name) is a mixed disorder of attention-deficit and hyperactivity symptoms and patients not infrequently manifest one part of this dyad predominantly. Thus, the 5-factor solution may indicate less a deficit of the scales than the definition and nature of these disorders, both epidemiologically and conceptually. The authors might note these points.

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