Author's response to reviews

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

Authors:

  Douglas E Faries (faries_douglas_e@lilly.com)
  John P Houston (john.houston@incresearch.com)
  Ellen M Sulcs (esulcs@harrisinteractive.com)
  Ralph W Swindle (swindle_ralph@lilly.com)

Version: 3 Date: 14 September 2012

Author's response to reviews: see over
Response to Reviewers

Reviewer: Editorial Office

Please declare all authors’ affiliations or other relationship to Harris Interactive in the Competing Interests section.

AUTHOR RESPONSE: The change to the Competing Interests section has been made to adhere to editorial requirements.

Please remove additional file 2 (Harris consent info) from the revised submission as this was required for editorial assessment only.

Please note that Additional files should not include letters to the editor, response to reviewers, copies of the manuscript in foreign languages, copies of other articles, identifiable patient information etc.

AUTHOR RESPONSE: Thank you for the helpful submission information. We have removed additional file 2 from the revised submission.

Please expand the consent statement in the manuscript to indicate that the consent was written.

AUTHOR RESPONSE: We have revised the consent statement; however, consent was not technically written. Respondents were screened to determine if they qualified for the study. Those who qualified were asked to give consent within the survey. Those who selected “Yes, I agree to participate in this study” after viewing the consent question were allowed to proceed with the survey while all other respondents did not proceed with the survey and were not included in the study. Note that those asked to participate in the survey were all participants in the Harris panel members who have previously provided contact information and agreed to participate in Harris panel surveys.

Reviewer: Michael Von Korff

This paper assesses the concurrent validity of a brief screening scale (PDI-4) for identifying four psychological disorders: major depression; generalized anxiety disorder; attention deficit hyperactivity disorder; and bipolar I disorder. This paper assesses the PDI-4 in an independent sample, replicating prior work. The main finding was that the PDI-4 shows reasonable agreement with validating scales, but that there is considerable symptom overlap across the diagnostic categories. While there is an
abundance of psychological screening scales, a brief scale covering these four disorders has value.

**AUTHOR RESPONSE:** Thank you for taking the time to review and improve this manuscript.

Major compulsory revisions - The rating scale is a screening scale that identifies persons with potential disorders. The paper should avoid the use of the term "provisional diagnoses". The amount of data presented is excessive. The confirmatory factor analysis and Table 2 should be dropped. Table 5 is also not essential. Figure 2 could also be dropped.

**AUTHOR RESPONSE:** We have made several major changes to the manuscript in response to these comments. First, we cannot completely remove the word ‘provisional’ as it is part of the name of the scale itself that is already published in two other manuscripts. However, we understand the reviewer’s concern and have revised all of the text (that is not referring to the name of the scale) to avoid the use of the term ‘provisional’. Second, we have removed the factor analysis from this manuscript. This includes text from the methods, results, discussion, as well as removal of Table 2. We considered the factor analysis as ‘nice to have’ but not as essential as the other analyses in evaluating the PDI-4 operating characteristics. Third, both Table 5 and Figure 2 were used to describe the results of the discriminant validity analyses. Upon reviewing this again, we feel that Figure 2 alone would suffice to describe the results of the analyses. Much of the information in Table 5, such as the coefficients of terms in the model other than cohort, is not essential. Thus, we have deleted Table 5 and retained Figure 2 in the revised manuscript.

Minor essential revisions - The length of the manuscript should be reduced.

**AUTHOR RESPONSE:** We have made several major changes to the manuscript, described above, which have reduced the length of the manuscript (removal of all text and tables surrounding the factor analyses as well as removal of Table 5).

**Reviewer:** Kurt Kroenke

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.

AUTHOR RESPONSE: Thank you for taking the time to review and improve this manuscript. We agree that the lack of a gold standard diagnostic interview for these subjects is indeed the major limitation of this work. While this was mentioned in the original version, it was only mentioned in a brief sentence at the very end of the discussion. We concur with the reviewer that this is not a sufficient discussion of this limitation. In the revised version we have first added clarification of this fact in the statistical methods section. Then, we have added comment on this limitation both early in the discussion section and expanded comments in the limitations section. In addition, by following other suggestions from the reviewer below, we have modified text and tables to better clarify how the cohorts were formed (e.g., meeting a cutoff score on a patient-rated scale).

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.

AUTHOR RESPONSE: We have revised the statistical methods section to address the use of the patient-rated scales and the cutoff scores in the first rather than the second paragraph on page 8. The additional text now makes it clearer to understand the formation of the cohorts for assessing the operating characteristics of the PDI-4.

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.

AUTHOR RESPONSE: We have added text showing the sensitivity and specificity of the comparator scales in a prior office-based study (Houston et al 2011). While we could add results from other studies in the literature for each scale, the Houston et al. study has the advantage of evaluating all 4 of these scales in the same study, using the
same cutoffs as we incorporated here, as well as having formal diagnostic interviews (SCID and ACDS) as the comparison. In addition, each scale assessed the same specific diagnoses as we assessed in this work (e.g., GAD rather than a related anxiety diagnosis).

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.

**AUTHOR RESPONSE:** See the response to #2 above.

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”.

**AUTHOR RESPONSE:** We agree that these changes will provide clarity to the information in Table 3 (previously Table 4), and these revisions have been made.

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.

AUTHOR RESPONSE: We have revised the text mentioned on page 9 and in a couple of other locations to avoid vague/suggestive language such as “met diagnostic criteria.” Thank you for the suggested edits and we have utilized this with only slight modification. Regarding the questions on the factor analyses, in response to another reviewer, we have removed the factor analysis and all such discussion from this text. Thus, no clarification is needed for this version of the manuscript. However, your interpretation was correct – the second exploratory factor analysis we conducted included items from the PDI-4 and items from each of the corresponding patient-rated scales. We wanted to see whether in the full pool of items the correlations and groupings would be as expected.

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.

AUTHOR RESPONSE: In response to another reviewer, we have removed Table 5 from this version of the manuscript. Figure 2 also described the results of the discriminant analysis – and did so in a more focused (directly presented cohort differences) and less confusing manner. Thus, Figure 2 remains and Table 5 was removed. We also added text in the Results to further explain these results as well as expanded the legend of Figure 2 to include the details of the statistical model.

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.

AUTHOR RESPONSE: We have expanded the legend to Figure 1 to better explain the graph. Specifically the following text was added: “The labels on the x-axis denote
patients meeting the cutoff score for the corresponding patient-rated scale: PHQ-9 score \( \geq 12 \) for MDE, HADS-A score \( \geq 14 \) for GAD, CAARS-S score \( \geq 28 \) for ADHD, MDQ score \( \geq 7 \) with functioning item score of \( \geq 3 \) for mania. The same definitions are used in computing the percentages of patients with other diagnoses denoted inside each histogram bar.”

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.

AUTHOR RESPONSE: It is true that positive and negative predictive values are dependent on the prevalence of the diagnoses in the population under study. Thus, one will obtain different values for in different populations (even if sensitivity and specificity do not change). Thus, we have added a sentence and a reference to this in the Discussion: “Note that the positive predictive values in this type of sample – where patients are self-selected to have greater rates of mental health disorders – would be expected to be lower in a general population sample where the prevalence of mental disorders is lower [23].”

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

AUTHOR RESPONSE: We have no disagreement with the above comments regarding the factor analysis results. However, given responses to another reviewer, we removed the factor analysis and all such discussion from this version of the manuscript.