Author's response to reviews

Title: The 12-item World Health Organization Disability Assessment Schedule II (WHO-DAS II): A nonparametric item response analysis

Authors:

Juan V Luciano (jvluciano@sjd-ssm.com)
Jose L Ayuso-Mateos (joseluis.ayuso@uam.es)
Jaume Aguado (jaguado@sjd-ssm.com)
Ana Fernandez (afernandez@sjd-ssm.com)
Antoni Serrano-Blanco (aserrano@pssjd.org)
Miquel Roca (mroca@uib.es)
Josep M Haro (jmharo@fsjd.org)

Version: 3 Date: 28 April 2010

Author's response to reviews: see over
We greatly appreciate the reviewers’ kind and encouraging comments about the quality of our manuscript. We have followed their suggestions, trying to incorporate them into the revised version of our manuscript, as follows:

RESPONSE TO REVIEWER 2 (Russell Steele)

Reviewer 2:
Discretionary Revisions:

As for my limited criticisms of the paper, they would be merely of style rather than of substance. In general, I prefer parametric methods combined with assessment of parametric assumptions rather than non-parametric methods mostly due to power arguments. In a perfect world, one would use both methods (parametric and non-parametric) and they would agree on the general interpretation of the available evidence, giving lots of strength to the results. Here the authors only present the non-parametric results (and somewhat weak reasoning for doing so). It would have been nice to see the parametric IRT results (or at least for the authors to have tried them and mention a brief comparison in the text). In particular, the null DIF results could be due to lack of power due to the use of the non-parametric smoothing technique (in place of a parametric IRT curve). Given the modest evidence for DIF here, it is difficult to conclude one way or the other. A parametric IRT analysis would have potentially had more power and given a different (albeit not necessarily correct) answer. I had wished that the authors would have presented that result in their work here.

The only other complaint that I had about the paper is that they seem to under-interpret the results with respect to the items. They merely report parameter estimates and summaries of the various discrimination and difficulty ratings for the items. I would have liked to see more about the specific estimates for the difficulties and discrimination in relating the items to each other. For example, the ICC and OCC curves by item seem somewhat similar in shape. What does that say about the items themselves? Are they all of the same difficulty or are they measuring the same thing? Is there some way with the non-parametric approach to summarize differences amongst items in the way that one could with a 2 or 3-parameter IRT model? I want to stress that I'm not saying that non-parametric method they've used here is not the correct one to use; rather, that it would have made for a stronger and more interesting paper with more interpretation comparing curves between items. Similarly, the options comparisons
seem important in that many of the plots are similar in Figure 4, but can we thusly conclude that the items all have similar discriminatory patterns with respect to the options?

In conclusion, I think this is a nice paper that really only needs to delve a bit more deeply into interpretations of their advanced (and justifiably so) statistical analyses in order to become a very good paper.

Authors:

In the present work, we performed nonparametric item response analyses of the 12-item WHO-DAS II instead of parametric analyses taking into account the following considerations:

Some experts on IRT (e.g. Sijtsma, K. and Molenaar, I. W. 2002. *Introduction to Nonparametric Item Response Theory*, Measurement Methods for the Social Sciences Series Volume 5. London: Sage Publications; Lei P, Dunbar SB, Kolen MJ. *A comparison of parametric and nonparametric approaches to item analysis for multiple-choice tests*. *Educ Psychol Meas* 2004, 64: 565-587) have pointed out that assumptions of parametric models impose very stringent restrictions on the items, which usually causes the rejection of frequently scarce and costly items. In contrast, nonparametric Item Response Theory (NIRT) models do not require complex estimation procedures, can be applied to relatively small data sets, are less imposing concerning distributional form of Item Response Functions and help to avoid misleading results obtained from parametric IRT models (these arguments have been included in the first paragraph of page 8).

In order to justify a bit better the selection of a NIRT technique (the kernel-smoothing technique), we have also added the following information in the paper:

*With the kernel-smoothing technique implemented by TESTGRAF, the researcher determines the item response functions directly from the data without forcing the data to conform to a logistic IRT model* (see first paragraph of page 8).

Finally, we have modified the second paragraph of the Discussion (see page 11) in order to delve a bit more into the interpretation of the results:

*Overall, the results obtained in the present research indicate that all WHO-DAS II items performed well at discriminating varying levels of disability. The inspection of the ICCs and the OCCs indicated that all items assessed well the entire continuum of disability. We could see that certain items, for instance items 8 and 9 (bathing and dressing, respectively), ask about activities in which patients clearly experience fewer problems, whereas in others, for instance items 5 and 12 (emotionally affected by the...*
health problem and work, respectively), patients report more problems or difficulties. Finally, the weights assigned to the individual item options are appropriate for measuring the underlying trait due to the absence of overlap between adjacent OCCs.

Other minor changes
- The following email addressees have been updated (see page 2):
  - JVL: jvluciano@pssjd.org
  - JA: jaguado@pssjd.org
  - AF: afernandez@pssjd.org
  - ASB: aserrano@pssjd.org

- The following reference has been updated (see page 14):