Reviewer's report

Title: Detecting Functional Change in Response to Exercise in Knee Osteoarthritis: A Comparison of Two Computerized Adaptive Tests

Version: 0 Date: 30 Jun 2017

Reviewer: Gregor Liegl

Reviewer’s report:

This study aims to compare the performance of two computer adaptive tests (PROMIS PF-CAT; OA-CAT) for the assessment of self-reported physical function (PF) in detecting function change in knee osteoarthritis patients. Generally, the manuscript is well written and clearly structured and the authors are dealing with a relevant and contemporary topic. However, there are some issues that should be resolved prior to publication:

Major comments

1. The sensitivity to change of an instrument is directly related to measurement precision (reliability; standard error of measurement, SEM). Highly precise measures are more sensitive to detect change; therefore, it would be interesting to compare the measurement precision between the two instruments (PROMIS PF-CAT; OA-CAT) and discuss respective results in the paper. The same applies to floor- and ceiling effects which may negatively affect the sensitivity to change of an instrument in a patient population of interest.

2. Moreover, sensitivity to change is highly related to the latent construct of an instrument. For example, differences in effect sizes can simply be caused by the assessment of different (sub-)constructs (e.g., Mobility versus Upper Function; or disease-specific versus generic PF). Thus, a direct comparison of two measures only makes sense if it can be presumed that both instruments show sufficient construct validity and are assessing the same latent construct. Are there previous studies that indicate that the instrument-specific PF construct definitions are similar for PROMIS PF and the OA-CAT? This seems to be of particular relevance because the differential impact of disease-specific (as in OA-CAT) versus generic conceptualization (as in PROMIS PF) of self-reported PF has been discussed by a number of authors before; related literature should be discussed in the manuscript. Moreover, the correlation between PROMIS PF and OA-CAT scores should be investigated for the study sample to ensure that a similar concept of PF was assessed.

3. Sensitivity to change can be defined as the ability of an instrument to detect true change in the latent construct of interest. As a consequence: shouldn't there be any kind of indicator for the true PF change (e.g., a 'gold standard' PF measure)? Alternatively, I think the expected ("known") effect size should be prespecified, based on findings of previous studies (if existing). Otherwise, it is not clear if the calculated effect sizes reflect any true change in physical function or if it is a result of response bias or response-shift.
instead. Moreover, according to the paper of Stratford and Riddle (2005) on choosing the appropriate change coefficient, which was cited in the manuscript, using effect sizes such as Cohen's d is appropriate only if an equal amount of change can be assumed for all patients. Is this assumption justified/realistic for the sample?

4. In this context: How do the authors distinguish between the treatment effect (= true PF change) and the instruments' sensitivity to change (ability to detect this change), if both things were assessed by the same indicator (Cohen's d)? (In my opinion, the 'usual' interpretation of the study's main results would be that the intervention is somewhat more effective for improving OA-specific PF than for improving general PF). Please explain.

5. In the specific context of computerized adaptive assessments using predefined automatized item selection algorithms which are usually based on large heterogeneous samples, other aspects are interesting to compare between the instruments, especially when used in a clinically homogeneous sample, e.g., construct coverage/content of selected items, (average) number of presented items, time to complete the assessment, etc.

6. I think the authors should provide more detailed information on the development and psychometric/technical background of the CAT instruments in the methods section (original calibration samples, underlying IRT models used for item parameter estimation, scoring methods, psychometric characteristics/measurement performance found in previous studies, etc.) and take regarding similarities and differences between PROMIS PF-CAT and OA-CAT into account when discussing the results.

7. p.7, lines 141 & 144: To allow for interpreting the standard error of a scale it would be useful to inform the reader about the measure's metric. For PROMIS PF, a T-metric is used, linked to a U.S. general population mean of T=50 with an SD of 10. What metric is used for OA-CAT scores?

8. Appropriate references are missing for several statements, e.g., p.4/lines 63-65. Moreover, most of the provided references are rather old and some of them seem to be outdated given that the field of IRT-based and computerized adaptive assessment of health-related constructs is quite new and constantly developing.

9. p.13, line 222: A more intensive treatment may lead to a greater "true" change, which is not equal to "better responsiveness" of the specific outcome measure assessing this change.

10. p. 6, line 116: It is not clear how there was a control group included. The authors state that the "CAT measures were administered to all participants prior to participating in a 6-week structured exercise program and again after the exercise program was complete [...]", indicating that all participants were participating in the exercise intervention.

Minor comments:
10. p.3, line 59: physical functioning is usually referred to as a subdomain of health-related quality of life rather than as a symptom.

11. p.6, line 109: Can the authors provide a reference to the larger study the present study is part of?

12. The PROMIS v1.0 PF item bank consists of 124 items (see Rose et al. 2014). PROMIS v1.2 consists of 121 items. Which version was used in the present study?

13. "Sensibility to change" and "responsiveness" should not be used interchangeably. For more detailed information, see Streiner et al. (2015)

References:
Rose, M., Bjorner, J. B., Gandek, B., Bruce, B., Fries, J. F., & Ware, J. E. (2014). The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. Journal of clinical epidemiology, 67(5), 516-526.


Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Unable to assess

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Unable to assess

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:
Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

Yes