Reviewer’s report

Title: Reliability of mobility measures in older medical patients with cognitive impairment

Version: 0 Date: 28 Aug 2018

Reviewer: Morten Tange Kristensen

Reviewer's report:

Review of Manuscript Number: BGTC-D-18-00444
Full Title: Reliability of mobility measures in older medical patients with cognitive impairment.

Thank you for giving me the opportunity to review your paper. The topic of your study is considered relevant in view of the limited evidence regarding reliability of outcome measures that also are being used in patients with cognitive impairment. The study is extensive and with many supplementary files. However, I do have a major concern specifically regarding the design of your study asking older participants to conduct 10 performance measures in the same session. This might have influenced your results and its very different from use in research and certainly when used in clinical practice. I will address my comments in the following:

Major comments:

Methods:
"In a single session, a comprehensive set of commonly used performance-based measurement instruments of mobility was administered in a standardised order, starting with the least physically challenging tests. In a sub-sample of eligible participants, all measures were repeated by the same assessor on the same day and in the same environment. The single independent rater was well informed of each participant's medical condition, such as diagnoses and level of cognitive impairment".

It's not specifically mentioned (sorry if I missed it) but I take for granted that the order of testing was as written, which means that testing started with DEMMI, followed by HABAM etc.:

"In this study, the following 10 performance-based measures of the mobility capacity of older people were applied: DEMMI [27, 34], HABAM [35, 36], POMA [37], SPPB [38], 4-meter gait speed, 5-times chair rise test (5xCRT), 2-minute walk test [39], TUG [40], Barthel Index mobility subscale [41], and Functional Ambulation Categories (FAC) [42]".

I fully agree with you that reliability estimates should be based on scores of patients who are considered stable (no one is completely stable), and whose medical condition has not changed. In that view, it seems obvious to me that introducing older patients with cognitive impairment to 10 performance based measures in the same session, and twice for one day, certainly will influence the stability of their performances, within the two sessions, and in favor of the measures applied as the first
and maybe second in your extensive test battery. Thus, your design potentially exhibits reliability estimates of tests applied as number 2/3-10 in a bad light, and mainly based on your design. It's therefore up for discussion if it's fair to report any reliability estimates for these outcome measures included in your study (my opinion is that only the measures applied as the first and maybe second in your sessions should be reported). It might still contribute valuable information to the literature.

Also, it's stated that the single rater was not blinded to the results of the first session:
"In the retest session, the rater was not blinded towards the results of the first session". Both, the intra-rater (compared to inter-rater) and the non-blinded design is a major limitation of you study findings, if one wishes to use your findings in one's own practice. I acknowledge that its mentioned as such under "strengths and limitations", but I suggest that it be highlighted as a major limitation, and I do not agree with you that: "the test-retest reliability of other assessors should be comparable if the same strict learning procedure is followed".

An inter-rater reliability design with raters blinded to each other's ratings until end of study is needed to answer that question!

Results:

Minimal detectable change was calculated as: \( MDC95 = 1.96 \times \sqrt{2} \times SEM \)

Based on your data in table 3, I re-calculated MDC95 as follows for the DEMMI:
\[ 1.96 \times \sqrt{2} \times 2.29 = 6.3 \] which is the same as reported in table 3.

I also calculated MDC95% as MDC divided by the mean of all scores for the sample, which is commonly used and much easier to interpret for clinicians, compared to estimates from B & A plots.

DEMMI, MDC% exact= \( \frac{6.3}{38.6} \) *100 = 16.3%. Rounded up estimate = 18%. This is a larger measurement error% than the 11% reported for a single person in your study.

I calculated the corresponding MDC% for the HABAM in the same way which gave a measurement error% > 20%.

B&A plots are excellent to illustrate the distribution of scores and if systematic between rater scores and heteroscedasticity are present, but interpreting measurement error from these requires specific knowledge, that is not common among clinicians.

I therefore, suggest that you report SEM% and MDC% in the same way for all measures and report these instead of those based on B&A plots, and adjust your discussion and conclusions, accordingly.

Especially these values are considered important for the interpretation of values in case of heteroscedasticity, which was present for some of the measures you evaluated.

Minor comments:
Discussion - Strengths and limitations:

You state that a strength of the study is the consecutive sample of 65 participants. I agree on the 65 for the first test session, but I do suggest that you re-consider if this can be called a "consecutive sample", considering that you assessed 153 participants with the measures (out of 259 with cognitive impairment) at baseline, but only 65 were included in your study.

Additional File 2: please check, it seems as if some of your values are placed in the wrong rows

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

Yes

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.

Yes

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