Reviewer’s report

Title: The iPhone Measure App level function as a measuring device for the weight bearing lunge test in adults. A reliability study

Version: 0 Date: 03 Apr 2019

Reviewer: Vivienne Chuter

Reviewer's report:

Thank you for the opportunity to review this paper investigating the reliability and validity of an iPhone App for measuring ankle joint dorsiflexion in a weight bearing lunge test.

Overall the paper supports the use of this widely available App in clinical practice as an alternative to a digital inclinometer.

I have the following comments related to the manuscript:

Abstract
The methods section does not contain information about the number of trials done, methods for calculating reliability etc please update. Please review your abstract for consistency of terminology and letter capitalization.

Background
Your second statement needs references.
Second paragraph: You start the paragraph by identifying digital inclinometers are used in clinical practice to measure ankle joint range of dorsiflexion and are precise and repeatable. Then make the argument that these are expensive and clients/carers may wish to take the measurement at home. There is no mention of relevance of this to clinicians? Please update.
I note the senior author has published a similar study using the Tiltmeter App and comparing this to a digital inclinometer. It would be helpful if there was a clear explanation of why one App may be different from another and therefore why an additional reliability study is required.

Methodology
There are some aspects of the methodology that require further clarification.
In relation to the measurement of the ankle joint lunge test, previous studies have identified the likelihood of subtalar joint pronation affecting the measurement and used a vertical line placed perpendicular to the wall to limit the confounding effect of this on the measurement. This is done in addition to ensuring the knee drops over the second toe (as this could occur without the foot being perpendicular to the wall). Was this used in the present study or was a different method used to control foot alignment with the wall?
I am unclear from the methodology as to whether an average of multiple trials by the assessors was used or, if only one measurement of each lunge test by each as assessor was used to determine the
ankle joint motion and subsequent reliability. This needs to be specified as I am unsure from what has been provided whether the correct ICC model for interrater reliability has been used.

In relation to the positioning of the measuring device "one cm superior to the posterior calcaneal tuberosity"- the referenced paper by Hoch et al does not measure from this point but uses distance from the wall to calculated the range of ankle joint motion. The other reference is in relation to a similar reliability paper and does not support the method as being one to measure ankle joint ROM. Please update the references and provide an alternative supporting using this technique.

Who deemed the sample size big enough to prevent recall bias- is this the judgment of the authors? In which case this should be clearly stated.

The number of participants seems quite small. The sample size has been calculated for $\alpha = 0.05$, $\beta = 0.20$ for detecting a correlation of 0.6. I assume this means the target ICC was 0.6? or was this the lowest acceptable ICC? Was the power calculation based on two observations per participant? Was this an adequate sample size to determine measurement validity? More information needs to be provided as to how this number was calculated. As a target ICC for a clinical measure 0.6 would be quite low and not what would be considered a particularly clinically useful measurement. Please identify the target ICC and the lowest acceptable ICC and justify accordingly or revise the power calculation. References supporting the interpretation of the ICCs also need to be provided.

The SEM for the each rater for the intra-rater reliability has been calculated. This has not been calculated for inter-rater reliability from what I can see in your results table nor is it referenced (ie method used to calculate) in your statistical analysis section-please update this and calculate for all reliability measures.

It would be worthwhile calculating the minimal important change (MIC) to identify the magnitude of change required for this to be a true change in ROM rather than error. Alternatively, 95% limits of agreement would identify the difference required for a true change. This would provide more clinically useful data. Conducting t tests between tests and retest data or rater 1 and rater 2 for mean ankle joint dorsiflexion measurements would also identify statistically significant differences between measurements or raters and identify is there was any effect from multiple measurements.

I am a bit confused about the statistical analysis for the concurrent validity. It seems an ICC was performed and a Bland Altman plot. Limits of agreement are reported in the results but not included in the statistical analysis-please update. How were the Bland Altman plot calculated? Did the calculation need to be adjusted for the small sample size? I am not sure why all the measures where included in one plot given you found worse reliability in the straight leg lunge versus the bent knee? It would seem to be a clearer method to do this as two separate plots for the straight leg and bent knee data.

In addition there is no mention of whether these data were normally distributed? Were data assessed for these? Please update.

Results
You currently have two tables labelled Table 1, please update your reliability outcomes table to Table 2 and change text accordingly. Your reliability results table is a bit confusing. I assume the 95% confidence interval in the last column is for the ICC- this really needs to go in the column next to the ICC value with the SEM in the final column. In relation to the inter-rater reliability data I am not sure what is being reported here-it states it is for rater one versus rater two in straight or bent knee position but gives only one mean value for each knee position when there should be two? (one for each rater). Please revise this.

The methodology for the concurrent validity (line 46) should be in the methodology. This aspect of the study is not very well described, please review.

Why were 21 participants recruited if your power calc required 18?
Your mean ankle joint dorsiflexion seems very low given your population. Is this finding consistent with other studies using this method? In a healthy cohort of adults straight leg lunge means are typically closer to 40 degrees. Please ensure this is included in your discussion.

In relation to the inter-rater reliability data I am not sure what is being reported here— it states it is for rater one versus rater two in straight or bent knee position but gives only one mean value for each knee position when there should be two? (one for each rater). Please revise this.

The results for your 95%LOA for your concurrent validity are different in your results here to the results in your abstract, please clarify.

Discussion
The first paragraph of the discussion would work better in the introduction— it is justifying why the study was done. This paragraph in the discussion should provide a summary about what was novel about your research and overview of what you found.

The second paragraph of the discussion gives a broad overview of reliability results stating all but two scores for reliability were above 0.9. Is this relating to ICCs? There needs to be more precise differentiation between the results of intra-rater, inter-rater reliability results. Additional statistical analysis relating to MIC/95%LOAs as mentioned above would more adequately assess reliability of the App and allow for a more detailed discussion of what these results mean for clinical practice. You make a statement in this paragraph suggesting the iPhone App can be confidently included in clinical practice to measure ankle joint dorsiflexion. Given you only had healthy people you need to be cautious generalizing your results to clinical populations where more extreme measurement may occur. Please revise this.

There needs to be a discussion of the concurrent validity results the limits of agreement and what these mean.

There also needs to be discussion about the mean ankle joint lunge measures recorded in this study and how these compare to other studies in healthy populations and why your results may differ.

Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

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 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.