Author's response to reviews

Title: Intra-rater reliability of hallux flexor strength measures using the Nintendo Wii Balance Board

Authors:

June Quek (j.quek@uq.edu.au)
Julia Treleaven (j.treleaven@uq.edu.au)
Sandra G Brauer (s.brauer@uq.edu.au)
Shaun O'Leary (s.oleary@uq.edu.au)
Ross A Clark (ross.clark@acu.edu.au)

Version: 2
Date: 2 June 2015

Author's response to reviews: see over
We thank Reviewer 1 for his or her comments which have contributed to a better paper. Please note that all amendments in the revised manuscript are in red.

Some comments from the authors relating to validity are certainly warranted. To what extent are the authors confident that the new proposed method, which is evaluating force (or pressure) under the interphalangeal joint of the hallux, actually measuring flexor hallucis brevis strength? Can the anatomy of the plantar aspect of the hallux (e.g. ipj prominence) affect this? They need to make it clear to the reader that increased force is directly related to increased FHB strength, and not to other factors.

Authors rightly commended that they attempted to minimize compensatory strategies from the trunk and ankle; however, no mention at all as to whether foot posture was controlled.

Since foot posture is directly related to arch height and consequently arch length, this could have a significant effect on results.

We fully agree with the reviewer that many factors can influence the force output. We have considered the potential compensatory strategies participants may engage and have put in place measures to optimize selective activity of the FHB. We want to assure the reviewer that all effort was taken to ensure that the experimental set-up was consistent and well controlled between days and across all participants by 1) standardizing all instructions, 2) securing the foot on the platform using a Velcro strap across the forefoot (please see figure 2) and fixing the heel on the platform using a Velcro strap over the knee. The trial is discarded if any compensatory strategies are observed. However, we acknowledge the limitations of this study that it is not possible to know for certain if this method was successful to isolate FHB without EMG measures and have previously included in the limitations section that this needs to be further investigated.

It reads as follows;

“Secondly, although all effort was taken to ensure that participants did not use compensatory strategies by using appropriate physical restraints and standardised instructions, it was not possible to know if the experimental set-up was able to separate the extrinsic and intrinsic foot muscle activity, which will need to be investigated in future studies.”

To address the influence of the IPJ prominence and foot dimensions on FHB strength, the purpose of this study was to test for reliability within-subjects and hence did not need to take into account between-subject variability. However, we have included under future considerations;

“Given that the purpose of this study was to test the reliability of within-subject measures across 2 time-points, we did not need to include variables such as foot arch height and length, the size of interphalangeal joint prominence, the length of the first MTP shaft and the range of motion of MTP joint. However these are important when comparing between subjects, and should be considered in future studies.”

With regards to stabilizing the foot in the experimental set-up, we have included under procedures,
“To minimize movement from the foot, a Velcro strap was firmly secured across the foot and fixed onto the platform between the first metatarsophalangeal joint to the middle of the arch (see figure 2).”

Although ‘healthy individuals’ were recruited, there are no comment whether other related factors such as pinch callus and hyperextension of the interphalangeal joint were controlled for. Also an assessment for normal first metatarsophalangeal joint range of motion is considered an essential requisite before actual testing. Did any of the participants have hallux limitus or rigidus?

We thank the reviewer for raising these valid considerations. We want to assure the reviewer that we did not include anyone with foot deformities (observed or reported).

As such we modified the manuscript which reads;

“Participants were excluded if they reported any foot deformities or ankle problems.”

We did not measure the metatarsophalangeal joint range of motion and acknowledge that this is important for future studies although not crucial for our study because this is an intra-subject reliability study and the strength test was isometric in nature. Hence, we do not think that the range of the MTP joint would have a significant impact on the reliability of the measures. Nevertheless, we acknowledge that this is important for future studies using this method and have included it under a new section “future considerations”. It reads as follows;

“Given that the purpose of this study was to test the reliability of within-subject measures across 2 time-points, we did not need to include variables such as foot arch height and length, the size of interphalangeal joint prominence, the length of the first MTP shaft and the range of motion of MTP joint which are important when comparing between subjects and ought to be considered in future studies.”

With regard to hyperextension of the interphalangeal joint, because of the design of the experimental set-up, where the toe was well fixed in neutral position, we did not observe this to be an issue.

How was data extracted? Was the balance board connected to the Wii, or was it interfaced to a pc or other logging equipment? What was actually measured, force (page 5, para 3) or pressure (page 8 para 2)? How was force represented on the device, in Newtons? How was pressure represented on the device, in KiloPascals, or in any other format?

To clarify for the reader, we have included details about how the data was extracted and calibrated for the purpose of this study. This has been included in the methodology as follows:

“The NWBB was interfaced via Bluetooth with a customized software program on a Windows computer, and the force data (in kg) was derived using the internally stored calibration information
unique to each NWBB. To improve the accuracy of these data, the value derived from the load cell used in the measurement of force was re-calibrated in the testing position. This consisted of applying a range of known loads (0.1-20 kg) to the load cell.”

How was it possible to ensure that participants actually applied 50% of their perceived maximal effort? (Page 6, last para)

This served the purpose of a warm-up and hence we did not think that it was important to monitor accurately the participant’s performance at 50% of their perceived maximal effort and an estimate was sufficient.

It would be of interest to the reader to know why ICC (3,3) was chosen instead of other ICC classes.

Based on the definition by Portney and Watkins (2000) [1], ICC (3,3) was selected because we are interested in the relative reliability of specific rater (hence variance is excluded), and the average of three measures was obtained.
We thank the second reviewer for his valuable insights. Please note that all amendments in the revised manuscript are in red.

**Major concerns**
- The publication does not follow the GRRAS checklist for reporting of studies of reliability and agreement, which can be found on [www.equator-network.org/](http://www.equator-network.org/). Thus, please revise the paper according to the GRRAS checklist.

We have amended the manuscript accordingly. Please note that all amendments in the revised manuscript are in red.

- Please provide data on SEM and MDC - the statistical section states that they will be calculated. However, I could not find these data in the results section (although %error might be a similar outcome to SEM).

We apologise for this oversight and have included SEM and MDC in Table 1.

- Finally, I’m missing a section in the discussion on the clinical and research implications with respect to the agreement outcomes found in the current study.

We want to bring to the reviewer’s attention the implications as stated in the discussion and limitations, which reads;

“From a clinical perspective this new method permits an accurate and selective measure of flexor hallucis brevis muscle performance, and could have positive implications for clinical assessment and rehabilitation. This is particularly relevant in light of evidence suggesting that improved intrinsic foot muscle strength is associated with improved dynamic support of the medial longitudinal arch and balance [2].”

“…Secondly, although all effort was taken to ensure that participants did not use compensatory strategies by using appropriate physical restraints and standardised instructions, it was not possible to know if the experimental set-up was able to separate the extrinsic and intrinsic foot muscle activity, which will need to be investigated in future studies.”

**Please describe the method for enrolment – was it random, consecutive or convenient?**

In response, we have modified the manuscript, which reads;

“Thirty healthy individuals (age: 34.9 ± 12.9 years, height: 170.4 ± 10.5 cm, weight: 69.3 ± 15.3 kg, female= 15) were recruited by convenient sampling.”

**How was the sample size determined?**

We have included in the manuscript an added section under methods, which reads as follows;
“Sample Size

We determined the sample size of 30 participants using the method proposed by Walter et al [3]. This was calculated based on a minimum acceptable reliability of 0.70 and an expected ICC value of 0.90 in a test-retest (k=2) design, with a level (1-tailed) of 0.05 and power of 95%.”

Test for heteroscedasticity in the data and choose appropriate statistical analysis accordingly

We have included in the manuscript, which reads as follows;

Under statistical analysis
“Visual inspection for bias and heteroscedasticity was performed by examining a generated Bland-Altman plot for the difference between the scores obtained on both days against their means.”

Under results,
“This new method of hallux flexor strength measurement demonstrated excellent intra-rater reliability (ICC=0.982, CI= 0.96-0.99) with percentage error of 12% and no presence of systematic bias and no evidence of heteroscedasticity.”

Please describe the experience of the rater

We have included in the manuscript which reads as follows;

“All procedures on both days were identical and supervised by the same physiotherapist (JQ) who has postgraduate qualifications and 13 years of clinical experience in the musculoskeletal field.”

Please state what the measuring unit was for the procedure – in addition this information should be added to the y-axis in figure 3

The measuring unit was in kilograms and have included this in Figure A.

Figure A

Bland Altman Plot
Editor’s comment: Assessment of strength: It is not clear what the units of measurement were. Can you please state this in the Procedures subsection and in Table 1.

Under procedure, the manuscript is amended as follows; “Strength was measured in kilograms.”

Table 1 now includes the unit of measurement in kilograms.
Table 1. Mean strength recordings (± SD) and reliability coefficients for hallux flexor muscle strength measurement

<table>
<thead>
<tr>
<th></th>
<th>Wii Board Day 1 (Mean ± SD)</th>
<th>Wii Board Day 2 (Mean ± SD)</th>
<th>ICC (3,3)</th>
<th>Spearman’s ρ*</th>
<th>95% CI</th>
<th>Systematic Bias (CI)</th>
<th>Width of 95% LoA</th>
<th>% Error†</th>
<th>SEM</th>
<th>MDC</th>
<th>Prop Bias*</th>
<th>Fixed Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toe Strength (kg)</td>
<td>8.12 ± 3.86</td>
<td>8.28 ± 3.80</td>
<td>0.982</td>
<td>0.933</td>
<td>0.96 to 0.99</td>
<td>None</td>
<td>2.01</td>
<td>12</td>
<td>0.5</td>
<td>1.4</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

SD = standard deviation; ICC = intra-class coefficients; LoA = limits of agreement; SEM = standard error of measurement; MDC = minimal detectable change; *Prop Bias= Proportional Bias
†% Error= 0.5*Width of 95% LoA/[(MeanDay1+MeanDay2)/2]*100
*All correlations were p<0.001
*Prop and fixed bias were determined from ordinary least product analysis.
References