Reviewer's report

Title: Measurements of Knee Rotation - Reliability of an External Device in Vivo

Version: 1 Date: 12 September 2011

Reviewer: Thomas Linding Jakobsen

Reviewer's report:

Review of “Measurements of Knee Rotation - Reliability of an External Device in Vivo”

This study examined the (1) intra-day, intratester reliability, (2) 1-week apart, intratester reliability, and (3) intra-day, intertester reliability of knee rotation of an external device (Rottometer) on healthy subjects. They concluded that the Rottometer was a reliable device to assess knee rotation at 3 different flexion angles (90, 60 and 30 degrees), when 6 Nm, 9 Nm torques and the examiner’s apprehension of end-feel were applied. Generally, the manuscript is well written, but it needs to be revisited. Besides, there are a number of major concerns related to the construction of the study, the statistics used, and, subsequently, the conclusions of the study.

Major Compulsory Revisions

Methods - Paragraph 1.

You had 3 repeated measurements of internal and external rotation. “Three repeated measurements of internal and external rotation with 3, 6 and 9 Nm torque as well as the examiner’s apprehension of end-feel (16) were recorded in both knees”. It is unclear, whether you performed 3 repeated measurement at each specific torque or performed 3 test trials of each torque or you refer to 1 measurement at 90, 60, 30 degrees, which is in total 3 measurements?

If 3 repeated measurement at each torque at each position (90, 60 and 30 degrees) was performed, it should be reflected in the ICC formula used - ICC (2,1) or ICC (2,3). Please explain why you chose the total internal-external knee rotation, and not divided the measurement of the knee rotation into internal and external rotation. In clinical practice, clinicians may be interested in only one direction e.g. if a patients suffer of an injury of the posterior lateral corner of the knee.

Please explain why zero degree of knee rotation was defined as the subject’s resting position? What if resting position of the subject was 50 degrees of external rotation and may present an impaired knee? To my knowledge, in most of the literature assessing range of motion (ROM) of e.g. shoulder joint, hip joint, a zero degree is defined from “neutral zero method” and ‘a body position in which the extremity joints are halfway between internal and external is 0 degrees for the ROM in rotation’ (Norkin CC, White DJ. Measurement of Joint Motion. A Guide to...
Methods - Paragraph 2

Was the order of the measurements randomised? It is important as it may influence the reliability of the different torques and apprehension.

Was the recordings blinded? Again, the tester may remember the degrees recorded which favour a high degree of reliability, even though you performed many measurements.

You wrote in the Discussion section, it was trained testers. Please specify the tester characteristics. It is of importance as the ICC model you use in the data analysis. The ICC model is determined whether the tester is ‘handpicked’ - model 1, or the tester could be randomly selected from group of testers – model 2. Thereby, your reliability estimate, ICC, may change and so would your results.

Methods - Paragraph 3.

I know that researchers are using the term test-retest reliability to describe intratester reliability. You should strongly consider dividing the study into intra-day or within-day, intratester reliability; 1-week apart, intratester reliability; and intertester reliability.

Methods - Paragraph 4

Was the order of the testers randomised? Why did the subject only wait 10 minutes – as the ROM rotation of the knee may increase due to the numerous rotations of the knee? Did subject leave the setup or stayed in the chair?

Methods - Paragraph 5

A healthy population of 10 subjects and both knees were measured, which equals n=20 knees. Firstly, the number of subjects is small, which limits the generalisability of the results. Secondly, by using both legs on healthy subjects, it may influence your recordings, as large ROM in one leg, the tester would suppose that the same subject would have a large ROM in the other leg. Thirdly, a power calculation of the sample size should be provided with such a small number of subjects. Fourthly, you refer to an earlier study (Almquist PO, Arnbjornsson A, Zatterstrom R, Ryd L, Ekdahl C, Friden T. Evaluation of an external device measuring knee joint rotation: an in vivo study with simultaneous Roentgen stereometric analysis. J Orthop Res. 2002;20:427-432.), where you examined the validity of the Rottometer in 5 subjects with a reconstructed anterior cruciate ligaments knee. Please explain why you chose healthy and not reconstructed anterior cruciate ligament subjects? The reliability is population-specific and often the reliability is lower in subjects with pathology compared to healthy subjects.

Methods - Paragraph 6

Please explain why the different statistical reliability estimates are used? The Intraclass Correlation Coefficient, ICC (2,1) should be justified as well as the Limits of Agreement Ratio (LOA ratio) and 95% Confidence Interval.
When the ICC is used to assess relative reliability, a 95% confidence interval of the ICC is recommended. This is especially warrant, when sample size is small. The LOA ratio is seldom used in reliability studies, and is not recommended in the referenced citation by Bland and Altman (Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8:135-160.). The LOA is recommended, when measuring agreement in method comparison studies, not in reliability studies. Please revisit the article page 149, paragraph 4.1 Estimating repeatability. If you assess the absolute reliability, the measurement error of the Rottometer should be calculated at a group level (research) and at an individual (clinical practice) level. This is crucial as you conclude, “we believe that the Rottometer has the potential to be used in research and clinical practice, due to its validity (13) and reliability”. Please explain why the 95% confidence interval is calculated, and is it the 95% confidence interval of difference of the measurement 1 and 2.

Finally, I strongly recommend that the systematic bias between test trials, and testers should be calculated. Systematic bias is a great concern in clinical reliability studies and affects the internal validity. How do we know when we have measured the most valid knee rotation? Is 1, 2 or 3 test trials necessary to measurement the largest knee rotation?

Results- Paragraph 1
According to table 1, the highest ICC was 0.84 and not 0.82. I suppose that all CI’s should be in degrees.

Results –Paragraph 3. Please rewrite this paragraph as misinterpretation may occur. Example: ICC(2,1) ranged from 0.77 to 0.87 in all applied torques, except 3 Nm at 90# and 60# knee flexion angles. All paragraphs in the results section are difficult to read. The same order of presentation of data in all paragraphs may help the reader.

Discussion
Discussion - paragraph 4 and 6.
Please state when you determine the Rottometer is reliable, before any judgement is done. An ICC of 0.5 is not reliable according to Fleiss or Landis & Koch (Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-174). The reliability cannot be fully assessed, if the absolute reliability (e.g. Standard Error of Measurement, smallest real difference etc.) is not calculated in a proper manner, and analysed according to the measurement under investigation. E.g., is a LOAratio of 0.65-1.39 and 95% CI of -6.95.6 acceptable in research or clinical practice?

Results – paragraph 5
As pain may be a risk in the intertester reliability study, the authors should present pain measurement (e.g. visual analogue scale) if they exist.

Table
The table is difficult to read. I suggest that the table is divided into 3 tables (test-retest), within-day reliability and intertester reliability. A clear separation between the three different degrees (90, 60 and 30) in each table. 95% CI of the ICC or just the lower limit of it should be calculated.

Minor Essential Revisions
Please be coherent with the term used.
Please choose knee joint rotation or knee rotation
Please choose tester or examiner
Remember that ICC is an abbreviation of Intraclass Correlation Coefficient.
I would recommend editing the paper and adjusting spelling mistakes and unclear sentences.

**Level of interest:** An article of limited interest

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**
I declare that I have no competing interests