Reviewer's report

**Title:** The Dutch Lower Extremity Functional Scale was highly reliable, valid and responsive in individuals with hip/knee osteoarthritis: a validation study.

**Version:** 3  **Date:** 23 December 2011

**Reviewer:** Anton Lenssen

**Reviewer's report:**

First I want to state I regret that the authors did not include the original WOMAC in their project.

They state that the WOMAC is the most widely used measure to evaluate function, but it’s ability to discriminate between pain and physical functioning is questioned. Authors had a perfect opportunity to answer whether the LEFS is better than the WOMAC in this discrimination. Instead of doing this they chose to compare the LEFS to two other new self report measures.

**Major compulsory revisions**

1. On page 9 authors state they chose to use the MDC 90 as measure on minimale detectable change and they refer tot Donaghue and Stokes. Although the latter report that 90% MDC might be acceptable when making decisions on effectiveness of interventions, they do not add a solid foundation for this statement. I’m not convinced 90% is better and believe the authors aren’t either since in the result section the report on the MDC 90 and the 95% limits of agreement. Please elaborate why 90% CI would be appropriate when making decisions on effectiveness and formulate the result section accordingly or change the method section and report that both 90% MDC and 95% MDC will be reported in the results.

2. Authors chose to report on responsiveness as the capability to distinguish between improved and not improved patients on the basis of the ROC curves. In doing so it is difficult to interpret the difference between true change and measurement error as measured by the smallest detectable difference. I suggest authors add a distribution based method (responsiveness ratio).

3 Authors report that 401 patients returned the baseline questionnaire. They do not report on loss to follow up which seems to be 39% (table 1, follow up 246 participants). Please elucidate on the figures.

4. I would expect responsiveness to be calculated on 246 subjects responding at follow up. I miss the point of reporting 246 participants at follow up (table 1) since 108 and 106 participants responded in the test retest reliability study and responsiveness study respectively.

Please explain why responsiveness was calculated on 108 subjects where information on 246 patients seems to be available
Minor essential revisions

1. Page 7 concerning the Sf 36; it is widely used, reliable, validated into Dutch and is easy to complete. Higher scores indicate better health. Please add references

Please add information on psychometric properties of the HADS.

Age 7 Please add information on psychometric properties of the HADS, if available.

2. Authors suggest that the HOOS and KOOS include the full WOMAC physical functioning scale. In my opinion WOMAC asks for PF in the last four weeks whilst HOOS and KOOS use one week intervals. Furthermore, the questions on HOOS and KOOS are imbedded in a larger set of questions which might be of influence in patients responses. I’m not convinced authors can draw conclusion on the responsiveness of the WOMAC by extracting the WOMAC PF questions out of the KOOS HOOS questionnaire.

3. Page 9 to evaluate …and other instruments (see Table 1) Please change table 1 in table 2

Discretionary revisions

1. Page 4 line one of those new measures, the LEFS, showed promise as a competitive alternative to the WOMAC-PF. Please add references to this statement

2. Table 3 does not seem to add a lot to the author’s arguments, consider skipping table 3.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

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