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:** 24 March 2012

**Reviewer:** Felix Angst

**Reviewer’s report:**

Positive criticism
This article reports the data of a well conducted study of cross-cultural adaptation and validation of the LEFS, a standardized outcome measure of the lower extremity.

Specific positive issues are:
1. The use of the standardized methodology for cross-cultural adaptation according to Beaton 2000 (Ref 17) – however, one step is lacking (see below, item 7).
2. The use of a couple of validated instruments covering a broad spectrum of domains and constructs.
3. Good reporting of reliability data.
4. Short, concise discussion covering limitations.
5. Reporting of the final Dutch LEFS in the Appendix for free use.

Major compulsory revisions
1. The major concern is about the lack of basic descriptive data – a basic need for every study report. By that, it is impossible to get an overview and a feeling for the validation results and to rate the quality of the results of the whole study. Please add means, standard deviations (or min, max, medians, … if the score distributions are not approximately symmetric) for all instruments and scores used (incl. frequencies of the global health change ratings=transition item) and for all time points of assessments. Consistently to that, the following items 2 and 3 are major criticism:

2. Cross-sectional validity: Report the/all correlation data (e.g. in a matrix) for all scores used in one Table (as partly done in Tables 3 and 4). Then, convergent and divergent constructs can be overviewed. Further, there are some ambiguities, for example: HADS: depression or anxiety?, SF-36: other scales?, HADS: negative correlation: more depression -> better function ? Please explain!

3. Longitudinal validity: Report the “classical” measures of responsiveness, the effect size and the standardized response mean for all scores used. This is the methodology used over decades and by many hundreds of responsiveness
studies. Then, sort them by constructs (pain/function/mental health) and level. By that, longitudinal construct validity and (joint-/domain-) specificity can be quantified and overviewed. This is necessary in addition to the ROC method.

Then: Compare the responsiveness within the same constructs by the sensitive, modified Jacknife test; description and one example can be found in (1).

4. You cite the fundamental textbook for establishing and testing instruments, the “Health measurement scales” of Streiner and Norman by the old, obsolete 3rd version/edition (Ref 24). Cite and use the current 4th version of 2008 (2). For example, the concepts of items 2 and 3 above are well described in that book.

5. Validity testing: The use of the “hypotheses concept” method (Table 2).

You used the method as described by the COSMIN group. I understand that because the authors are your Dutch colleagues. You can do that in addition to the method of items 2 and 3 above, but do also the “classical” methodology.

5.1. Discuss shortly the advantages/disadvantages of the COSMIN method and cite the reference (3). Some problems (only some of many examples) are inherent in the COSMIN methodology:

5.2. You did not establish 16 hypotheses but each 8 hypotheses for knee and hip. Two versions of the same hypothesis for hip and knee have another dependence to each other than hypothesis 1 to hypothesis 2 etc.

5.3. The number of 16 hypotheses is arbitrary, why not less, why not more?

5.4. Hypothesis: correlation is low <0.50 (true/false): There is a difference if it is 0.49 or 0.01, Discuss!

5.5. Dependency of educational level / socio-demographic status. There is a huge number of epidemiologic literature that demonstrate that health and almost every outcome has a social gradient. Explain and discuss!

5.6. Pat. with complaints 5 years or longer… Why 5 years: arbitrary?

5.7. If 75% of the hypotheses were confirmed: Why 75%: arbitrary?

6. Construct validity of the LEFS for function. Test the uni-dimensionality for function by (for example) factor analysis.

7. Translation: Stage/step VI of the process is lacking: “sending all versions and the protocols of steps I to V to the developer of the original questionnaire” (Ref 17: Beaton 2000). Explain!

8. A Cronbach’s alpha of 0.96 means that some items of the scale are redundant. Discuss! See also ref. 27 Bot et al.: a value in the range 0.70-0.90 would be good. See also in Streiner (2).

9. Results, Score distribution of the LEFS (p. 11). Analyze and report the characteristics of the score distribution: normal (Gauss), symmetric? This belongs also to validity. In this context: If the scores will not be symmetrically distributed one should rather use Spearman correlations that Pearson correlations (Tables 3 and 4).
Minor essential revisions

10. Abstract: Match the aims with the title. You did much more than only discriminant validity.

11. Introduction: As you described, the WOMAC is the most often used tool for the leg. State why you did not use the WOMAC (License problems, prohibited for validation studies….).

12. Patients, inclusion/exclusion, p.6: State explicitly that you only included patients with osteoarthritis.

13. Measures. Report the “missing rules”, i.e., how many completed items are necessary to determine the scale for all instruments, especially for the LEFS.

14. Methods, discriminant validity, p. 9: Cite a reference for the Meng’s test to compare two correlations or list the formula. Moreover, is this comparison of correlations a established method? Give a reasoning and cite a reference for that.

15. Methods, Validity (p. 9). Citing of the reference 27 on p. 9. This paper of Bot et al. is a review of shoulder instrument studies. This seems not be appropriate to explain the 75% proportion of positive hypotheses. Explain of look for that rather in (3)!

16. Results, Patient selection (p. 11). 401 patients at baseline, only 120 at follow-up: How were they selected? Randomly? Is there possible bias? Explain!

17. Reliability. You reported the MDC90% but the 95%-CI for the Bland-Altman plot. To compare, report also the MDC95%.

18. Reference 2. For the WOMAC, cite the WOMAC manual (4) not the preliminary study Ref 2.

19. Discussion, p. 16. “Convergent validity should be assessed in future studies.” Yes, but you also did that by the correlation data of Tables 3 and 4 and by the responsiveness data that will given (item 3): the LEFS shows convergent validity to H/KOOS function and SF-36 PF. Adapt the discussion!

Discretory revisions

none

References


3. COSMIN-manual:


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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.