Reviewer's report

Title: The Conceptually Equivalent Dutch version of the Western Ontario Rotator Cuff Index (WORC)(c)

Version: 2 Date: 13 June 2013

Reviewer: Ole Marius Ekeberg

The paper has now been improved. Most of the criticisms from reviewers have been adequately addressed. The comments to the revised items are in general adequate. There are however still concerns regarding the quality of the article. The editor should be aware of that recently there is published a dutch version of WORC (Wiertsema SH et al, J Shoulder Elbow Surg. 2013 Feb;22(2):165-70.) . The authors should comment on this other version in their article and revise accordingly.

Major Compulsory Revisions

Construct validation:

1. In the revised version, the authors have rightly chosen abolish the term criterion validation, and use construct validation (except in the abstract were criterion validation is still used, needs correction). They adopt a priori hypotheses of the original article, and leave a reference to the original article, but without stating what the a-priori hypothesis were other than the correlation coefficients. The theory behind the expected correlation coefficients needs to be discussed more in detail. (For example; 1) the correlation with the hinderence-scale of -0.75 is high as expected, but the a-priori expectation is not described in detail in the methods section, as it should. 2) The rationale for correlation between the worc total score and domain scores in RAND 36 is not explained, why not correlate worc subscores with RAND 36 domain scores?).

2. Further it must be outlined that finding the same size of correlation coefficients as the original article would not be evidence for validation of the translation. On the other side, replicating the analysis from the original article does not increase our knowledge and interpretation of what constructs the WORC index is measuring. Therefore, it is difficult to see that the construct validation analysis is contributing anything to the article. In light of that there is already published a dutch version which is tested for reliability, but not validity, a thorough validation
would make this article more interesting.
3. The sentence “Since there are no gold in which to measure…..” needs to be rephrased.
4. Concluding that the dutch version is a valid measurement tool based on this analysis is to strong and needs revision.
5. My advice would be to rewrite the validation study with a clear description of the apriori expectations in the methods section, test interesting hypothesis of how the expected behaviour of the worc index and to discuss your findings or to remove the construct validation study from the article.
6. The constant score is widely used, but this is not proof of that the dutch version is properly translated. If there is no evidence of properly translation of the constant score, the reader of the article should be alerted.

Minor Essential Revisions
1. The floor or ceiling analysis is discussed under the header “construct validity”, and should be placed elsewhere. The sentence “This indicates that patients with lowest and highest scores could be distinguished from each other, and therefore supports content validity” is not correct (“and” should be replaced with “or” as lowest and highest scores will be differentiated from each other, but not extreme values).
2. The header on page 6 “validation” should probably be “patients”?

Discretionary Revisions

Agreement and reliability:
1. ICC estimates are sample dependant. High estimates of ICC may be due to high total variation in the sample (Range in WORC total score in table 3 supports this). It would be nice to discuss this fact in the discussion section. I agree that the measurement error (SEM) is acceptable, but in clinical use the measurement error in all PROMs are quite high (15-20%) and therefore these questionnaires are better suited to analysis on the group-level. I am not convinced that more specific patient groups (based on structural findings?) will result in smaller SEM, previous studies have supported that pain, function, emotional and sociodemographic variables affect total variation in scores.

Level of interest: An article of limited interest

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