Reviewer's report

Title: Validation of the Western Ontario Rotator Cuff Index in patients with Arthroscopic Rotator Cuff Repair: A study protocol

Version: 1 Date: 11 March 2010

Reviewer: Susanne Wulff Svendsen

Reviewer's report:

The protocol describes a proposed study of a Dutch translation of the Western Ontario Rotator Cuff Index (WORC). The study aims to assess validity, reliability, and responsiveness of WORC in patients undergoing arthroscopic surgery for a rotator cuff tear. To achieve the aim, properties of WORC will also be investigated in patients with rotator cuff disorders without rupture and in shoulder instability patients – the two last-mentioned patient groups will be treated either conservatively or surgically. The performance of WORC will be compared to SF-36 and Constant Score.

Several instruments exist for assessing core outcome domains in patients with shoulder disorders, e.g. (1;2). To enable clinicians and researchers to choose the most appropriate instrument(s) depending on the purpose, evaluation of the instruments’ properties is important (3). Accordingly, I appreciate the research idea, but I have a number of questions and comments that I think should be addressed in a revised version of the study protocol.

Major Compulsory Revisions

1) I think that the rationale of the study needs to be clarified. By way of introduction, the authors refer to nine studies of arthroscopic cuff repair that have reported good to excellent results. As judged by the titles, these studies reported on case series of surgically treated patients without comparison to patients treated in other ways. Although such studies can be used to compare results e.g. between countries, they cannot determine the efficacy of arthroscopic rotator cuff repair because observed changes cannot necessarily be ascribed to surgery. I think that the proposed study would be better motivated by the need for validated outcome measures to be used in randomized controlled trials of interventions for rotator cuff tears. In such trials, shoulder specific quality of life could be a core outcome domain (4) in addition to pain, shoulder specific function, and work ability. The authors state that they want to be able to compare their results of arthroscopic rotator cuff repair with international literature (page 3, lines 4-5 and page 6, line 4), but for this purpose, WORC seems to be a less obvious choice than UCLA and Constant Score that have been much more frequently used.

2) The rationale for selecting WORC is not obvious at present. The use of disease-specific instruments has been recommended to improve sensitivity to change (4), but other rotator cuff specific questionnaires might be equally or more relevant – e.g. the rotator cuff quality-of-life-measure (RC-QOL) (5;6) and
Oxford Shoulder Score (OSS). OSS has been recommended for evaluation of patients having a shoulder operation other than stabilisation (1) and was recently selected as the primary outcome measure in a promising UK randomized controlled trial of interventions for rotator cuff tears (7). I would like the authors to elaborate on their reasons for selecting WORC as compared to these other possibilities. Specifically, it is not clear to me why the authors look for an outcome measure addressing rotator cuff repair as opposed to treatment for other rotator cuff disorders. What aspects of shoulder functioning in day-to-day activities and related quality of life are expected to differ between rotator cuff patients with or without a tear, in a way that will not be detected by e.g. Constant Score?

3) The hypotheses need to be specified. On page 3 it is hypothesized that WORC has a high validity compared to SF-36 and Constant Score in the two patient groups with rotator cuff disorders (with or without rupture) and, implicitly, not in the patient group with instability. Validity is a broad concept that can be divided into content validity, criterion validity, and construct validity (3;8). According to page 5, the authors focus on criterion validity (or more specifically concurrent validity), i.e. the extent to which WORC scores are related to gold standard instruments. However, since the authors are looking for a new instrument, it seems that they believe that SF-36 and Constant Score are suboptimal, meaning that these instruments cannot be considered gold standards. Maybe the authors are in fact aiming to assess construct validity, i.e. the extent to which WORC Scores relate to SF-36 and Constant Scores in a manner that is consistent with a priori hypotheses concerning the concepts that are being measured. Construct validity can be assessed by comparing the convergent/discriminant validity across instruments and patient groups (‘known-groups validation’, hence the three patient groups). It seems to me that the fourth hypothesis on specificity (page 3) belongs among the hypothesis on construct validity. It should be noted that the study will not truly be able to assess disease specificity of WORC (page 3, lines 17-19) because the three patient groups differ not only with respect to diagnosis, but also with respect to treatment modality. Hypotheses regarding construct validity need to be specified in detail, cf. (3).

4) It is hypothesized that WORC has a high responsiveness in the two patient groups with rotator cuff disorders. This means that it is expected that 1) the patients will improve (or get worse) in response to treatment and that 2) the according change in shoulder related quality of life can be detected by WORC. Implicitly, low responsiveness seems to be expected in the instability group. To find out if poor responsiveness among patients with instability is due to unchanged status or low responsiveness, it could be a idea to compare WORC to a disease specific instrument for instability, such as the Western Ontario Shoulder Instability Index (WOSI) (9;10) in addition to the intended global assessment of change at T1 (page 4, ‘patient routine’ section, lines 12-13). This is not meant as a required change of the study, only I would like the authors to consider the idea.
5) Title: I suggest that the title is changed to reflect the fact that not only validity, but also reliability and responsiveness will be assessed.

6) Study population (or patient assignment): Information needs to be provided on the setting where the data will be collected. Will consecutive patients be invited? Will there be any restriction on age? Language abilities? I recommend that a flow diagram is provided in the protocol with empty spaces to be filled in with numbers of patients assessed for eligibility and numbers of excluded patients according to reasons for exclusion.

7) Top of page 5: It says that patients with failed previous surgery will be excluded – why not exclude all patients who have had shoulder surgery?

8) Page 4: Under ‘patient routine’ (or in the appendix), please describe the procedures used to assess the different components of the Constant Score. I think that the most recent authoritative description of these procedures should be included among the references (11). Do you intend to adjust Constant Score for age and gender, cf. (12)?

9) Page 5, lines 3-4: Although it is not the purpose of the study to compare results of surgery and conservative treatment, I think that the study may convey an impression of surgery that is too favourable because patients who are treated surgically will not be assessed until three months after treatment whereas other patients will be assessed already after six weeks. I suggest that all T2 assessments are performed either six weeks or three months after treatment is started.

10) Page 5, second paragraph under the subheading ‘responsiveness’: It is not clear to me if you will focus on clinically important differences occurring over time for individual patients or on differences in scores between the three patient groups (or both)? As a generic instrument SF-36 cannot be expected to be very responsive at the individual level, and it may not be the best instrument to discriminate between the three patient groups, either. How do you intend to judge minimal clinical important differences by comparing with SF-36? I suggest that you use the main complaint score (see below), (13) may be inspirational.

11) Page 5, lines 13-15: It seems to me that the suggested anchors of the ‘severity of symptoms scale’ represent different outcome dimensions rather than different severities of the same dimension (complaints => hindrance). I suggest that you ask the patients to evaluate their main complaint on a scale ranging from worst possible to best possible, cf. (13).

12) How many patients do you intend to include from each group? The study protocol should include power considerations.

13) Will the study need approval from an ethical committee?

14) My general impression of the text is that it repeats itself several times, which calls for revision. E.g. part of the last paragraph on page 2 is repeated at top of page 6, part of the second paragraph on page 3 is repeated in lines 11-13 on
page 6, and most of the background section’s description of WORC is repeated in the discussion (top of page 6) and in the appendix (top of page 7). Moreover, I think that parts of the discussion and appendix would be better placed in the background section, e.g. page 6, lines 14-17 regarding the translation process.

Minor Essential Revisions

15) Abstract: Under the subheading ‘methods/design’ it is stated that an approved translation of WORC was used. How and by whom was the translation approved (I do not think that this appears from the main text)?

16) References should be numbered according to first appearance in the text. On page 2, the first paragraph of the background section refers to reference numbers 1-9 and 11-13, whereas reference number 10 does not appear until the following paragraph. I have not checked the numbering further on.

17) Page 2, background, paragraph 2, line 5. The sentence starting here should read: The Constant Score also consists of ... (not exists of).

18) In the background and methods/design sections, reference should be made to the appendix.

19) Page 4, line 3: The last word should be ‘signs’ not ‘symptoms’ (symptoms are assessed subjectively by the patient). Will all of the clinical tests be performed in all patients or will different surgeons select from the test battery according to their own clinical practice?

20) Page 4, line 6 in the ‘patient routine’ section: ‘A complete physical examination’ is an ill-defined concept.

21) Page 4, line 13 in the ‘patient routine’ section: How do you intend to ask patients whether their symptoms have changed? (Wording of the question; response options).

22) Page 5, line 5: Please, be as specific as possible when you state the patient characteristics that will be included in the analyses (or considered for inclusion).

23) Page 5, line 9: Please, provide references regarding missing rules for each instrument.

24) Page 5, lines 23-30: I believe that the changes in scores are calculated from T0 to T2, not from before T0 to T2? I suggest that you write T0 instead of baseline in order not to confuse the reader. Please, provide a reference regarding Cohen’s standard.

25) Appendix: The appendix could be shortened considerably – in my view, it will suffice if you provide key references for each clinical test and describe any deviations from published procedures.

26) References: Reference number 26 is incomplete.

Discretionary Revisions
27) Page 2, last paragraph, second sentence: I suggest that the wording is changed to something like: ‘Therefore, it is important to assess quality of life when such therapies are evaluated.’

28) Page 3, paragraph 2, line 1: WORC was also used as an outcome measure for rotator cuff repair in (14;15).

29) Table 1. I do not think that this table is really to the point of the paper, cf. the first paragraph of my ‘major compulsory revisions’ section.

Reference List


(6) Razmjou HF, Bean AF, van O, V, MacDermid JC FAU - Holtby R, Holtby R. Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures.(1471-2474 (Electronic)).


**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.