

# Validation of the Danish version of Oxford Shoulder Score

Lars Henrik Frich<sup>1</sup>, Peter Moensted Noergaard<sup>2</sup> & Stig Brorson<sup>3</sup>

## ABSTRACT

**INTRODUCTION:** The Oxford Shoulder Score (OSS) is a patient-administered condition-specific questionnaire for patients with degenerative or inflammatory shoulder disease. The purpose of this study was to validate a Danish translation of the OSS and to compare it with the Constant Score (CS).

**MATERIAL AND METHODS:** A total of 102 consecutive patients referred to our shoulder unit were recruited for the study. The OSS was translated into Danish according to recommendations presented by Guillemin and co-workers. We established the psychometric properties of the scoring system. Test-retest reliability was assessed by inviting 32 patients to complete another OSS 72 hours after the first test. The Bland-Altman plot was used to show absolute differences between test and retest.

**RESULTS:** The validity expressed as a Spearman's rank correlation coefficient comparing the OSS with the CS was 0.74. The test-retest reliability correlation coefficient was 0.98. Bland-Altman plots revealed limits of agreement between the scores of -4.5 - -5.4. Internal consistency tested by Cronbach's alpha was 0.93.

**CONCLUSION:** The psychometric properties of the Danish version of OSS showed good validity and reliability with a substantial correlation between the OSS and the CS. Internal consistency was high. The OSS is recommended for the evaluation of patients with degenerative or posttraumatic shoulder diseases.

**FUNDING:** not relevant.

**TRIAL REGISTRATION:** not relevant.

The prevalence of shoulder pain is increasing, and shoulder pain is the third most frequent disorder of the locomotor system after back pain and neck problems [1]. Several questionnaires have been developed to assess shoulder function [2] but the performance of these scoring systems has not been validated within a Danish setting. Results have traditionally been assessed using observer-administered measures. Scoring systems such as Constant Score (CS) tend to be inaccurate and affected by surgeons bias [3]. The existence of discrepancies between patients' and surgeons' perception of outcome is now generally accepted [2]. This difference in perception of outcomes has stimulated research into and development of patient-administered scoring systems. The

Oxford Shoulder Score (OSS) was introduced in 1996 [4] and is designed to measure outcomes following treatment of shoulder conditions. The questionnaire contains 12 items to be completed solely by the patient. Each item has five response categories. The questionnaire includes a mix of pain-related questions and questions about functional ability. Scores range from 12 (best; no pain or functional limitation) to 60 (worst). When using this instrument it is necessary to take into account that lifestyle and cross-cultural differences may affect scores in different countries. We have translated the OSS into Danish language according to the criteria's presented by Guillemin et al [5]. The objective of the study was to validate the Danish translation and retest the psychometric properties of the OSS in a Danish setting.

## MATERIAL AND METHODS

A total of 102 consecutive patients who had suffered from shoulder pain for at minimum of eight weeks were included at Odense University Hospital in December 2008. The study group comprised a variety of diagnoses of which rotator cuff related disorder made up the largest group (Table 1). All patients in the rotator cuff group were diagnosed with impingement syndromes and seven patients had a full-thickness rotator cuff tear. The fracture group comprised conservatively treated fractures of the proximal humerus and clavicle. Exclusion criteria were recent surgery, shoulder pain originating from neoplasms, and systemic arthritic

## ORIGINAL ARTICLE

- 1) Department of Orthopaedic Surgery, Odense University Hospital,
- 2) Department of Orthopaedic Surgery, Regional Hospital Horsens, and
- 3) Department of Orthopaedic Surgery, Herlev Hospital

Dan Med Bul  
2011;58(11):A4335

Force measurement  
is part of the Constant  
Score.



 TABLE 1

Demographics of the patients.

Mean age (range), years	49 (18-91)
Male:female, n	56:46
Dominant hand, right/left, n	90/15
Affected shoulder, right/left, n	67/35
Rotator cuff disease, n	72
Glenohumeral arthrosis, n	2
Shoulder instability, n	16
Fractures, n	12
Total, n	102

condition or neuropathic pain. Three patients treated with subacromial injection after completion of the scores were excluded from the re-test cohort. Patients' functional status was measured using the CS and the

OSS. Data were collected by three bachelor students in rehabilitation. The OSS was translated into Danish according to international recommendations [4]. This process involved translation of the questionnaire by two bilingual persons who were Danish native speakers and had clinical experience. After this process, a reverse translation was made by an English native speaker. Similarities were evaluated during a conference and consensus was achieved on the final Danish version (**Figure 1**).

CS contains subjective assessment of pain and activities of daily living which are allocated a maximum of 15 and 20 points, respectively. An additional 65 points for active motion and strength yields a final score maximum of 100 points. The CS score is gold standard in all European shoulder clinics and guidelines exist for its use [6].

### Statistics

The psychometric properties of the Danish version were tested in terms of reliability and validity:

 FIGURE 1

Danish version of the Oxford Shoulder Score. The questionnaire is available for download free of charge from skulderalbue.dk.

**Oxford shoulder score** Patient Label

Problemer med din skulder Dato: \_\_\_\_\_

præoperativ  6 mdr  12 mdr  Venstre  Højre

Sæt et **X** ved hvert spørgsmål

1. I løbet af de sidste 4 uger ...  
 Hvordan vil du beskrive den værste smerte du har haft i din skulder?

Ingen  Mild  Moderat  Svær  Uudholdelig

2. I løbet af de sidste 4 uger ...  
 Har du haft svært ved at tage tøj på, på grund af din skulder?

Intet besvær  Lidt besvær  Besvær  Meget besvær  Umuligt

3. I løbet af de sidste 4 uger ...  
 Har du haft svært ved at komme ind og ud af en bil eller ved at bruge offentlig transport på grund af din skulder?

Intet besvær  Lidt besvær  Besvær  Meget besvær  Umuligt

4. I løbet af de sidste 4 uger ...  
 Har du været i stand til at bruge kniv og gaffel - på samme tid?

Ja, let  Lidt besvær  Besvær  Meget besvær  Umuligt

5. I løbet af de sidste 4 uger ...  
 Kunne du selv klare de daglige indkøb?

Ja, let  Lidt besvær  Besvær  Meget besvær  Umuligt

6. I løbet af de sidste 4 uger ...  
 Kunne du bære en bakke med en tallerken med mad gennem et lokale?

Ja, let  Lidt besvær  Nogen besvær  Meget besvær  Umuligt

7. I løbet af de sidste 4 uger ...  
 Kunne du børste/rede dit hår med den dårlige arm?

Ja, let  Lidt besvær  Nogen besvær  Meget besvær  Umuligt

8. I løbet af de sidste 4 uger ...  
 Hvordan vil du beskrive den smerte, du normalt har haft i din skulder?

Ingen  Mild  Moderat  Svær  Uudholdelig

9. I løbet af de sidste 4 uger ...  
 Kunne du hænge dit tøj op i en garderobe, med din dårlige arm?

Ja, let  Lidt besvær  Nogen besvær  Meget besvær  Umuligt

10. I løbet af de sidste 4 uger ...  
 Har du været i stand til at vaske og tørre dig selv under begge arme?

Ja, let  Lidt besvær  Nogen besvær  Meget besvær  Umuligt

11. I løbet af de sidste 4 uger ...  
 Hvor meget har smerten fra din skulder forstyrret dit normale arbejde (inkl. husligt arbejde)?

Slet ikke  En lille smule  Moderat  Meget  Totalt

12. I løbet af de sidste 4 uger ...  
 Har du været besværet af smerter i din skulder i din seng om natten?

Ingen nætter  1 til 2 nætter  Nogle nætter  De fleste nætter  Hver nat

Test-retest reliability is the ability to obtain the same score repeatedly and independently of time. This was analyzed in a subgroup of 32 patients selected at random from the rotator cuff group. Patients in this group were tested twice at a 72-hour interval. Such interval allowed us to anticipate that the patient's health was in a stable period and no specific shoulder treatment was given between the two evaluations. The Bland-Altman plot was used to show absolute differences between the test- and the retest dataset.

We further calculated Cronbach's alpha for inter-correlation among the items in the OSS (internal consistency). Cronbach's alpha indirectly measures the extent to which each of the 12 items of the OSS measure the same construct.

Validity is an index showing how well the test measures what it is supposed to measure. The Spearman's rank correlation coefficient was calculated for the OSS and the CS to assess correlation between the two outcome measures.

*Trial registration:* not relevant.

## RESULTS

The Spearman rank correlation coefficient between the OSS and the CS was 0.74. The test for internal consistency resulted in a total Cronbach's alpha of  $R = 0.93$ . Elimination of each of the 12 items did not result in values below 0.916. Except for item 1 (pain), all items showed a correlation exceeding 0.62 (Table 2).

The time used to complete the forms was recorded for 80% of the test persons. The mean completion time was 2.25 min. (range 1.5-12.0 min.) for the OSS questionnaire and 7.1 min (range 5-17 min.) for the CS.

87% of the forms were completed correctly. A total of 32 patients were asked to fill-in the OSS questionnaire twice at different occasions. In all, 23 (72%) returned the questionnaire. The test-retest reliability expressed as a Spearman rank correlation coefficient was 0.98. The difference plot (Figure 2) between the two datasets revealed that the test-retest results were not strictly centered.

## DISCUSSION

The psychometric properties of the original OSS score shows that it is a valid, reliable and sensitive tool for the assessment of patients with various shoulder conditions [4]. The OSS has been tested for responsiveness to clinical change, not only in postoperative cases but also in non-operative cases [7] and the patient's independent completion of the questionnaire does not require detailed supplementary instructions. This tool was, however, developed for the Anglo-Saxon area and based on English culture. The challenge of this study was to trans-

late into Danish the OSS in order to address Danish patients' perception of shoulder conditions. The validity of patient-administered outcome questionnaires is hampered if questions are misinterpreted. An accurate and approved method of translation is therefore required to ensure that respondents understand the questions as intended [5]. The mean time for completion of the OSS was 2:25. This implies that the questionnaire was well comprehended and easy to fill out.

The translation and use of standardized questionnaires also enables comparison of national studies with results from international studies and the implementation of international multi-centre studies [5]. This implies some cross cultural adaption and modification of some of the items of the OSS. For example, knowledge of national traditions and characteristics such as differences in the behavior and eating habits of families may have an unintended impact on scores. Item six in the OSS, for instance, deals with the way a dinner plate is carried through the dining hall. Italian life style rarely involves this task and the OSS therefore needs adaptation before it may be used in Italy [8].

Patients with rotator cuff related disease made up the majority of the tested patient group. We believe that this cohort is representative because those suitable for surgery were included. Our results relate to a population of patients whose age ranged from 18 to 91 years (Table 1). Results indicate that the OSS may be used to follow a cohort over a longer period of time without the need for patients to show up in the clinic facilitating cost-effective follow up. However, the results of this study are not necessarily valid in other patient groups.

For comparison we chose the Constant Score.



TABLE 2

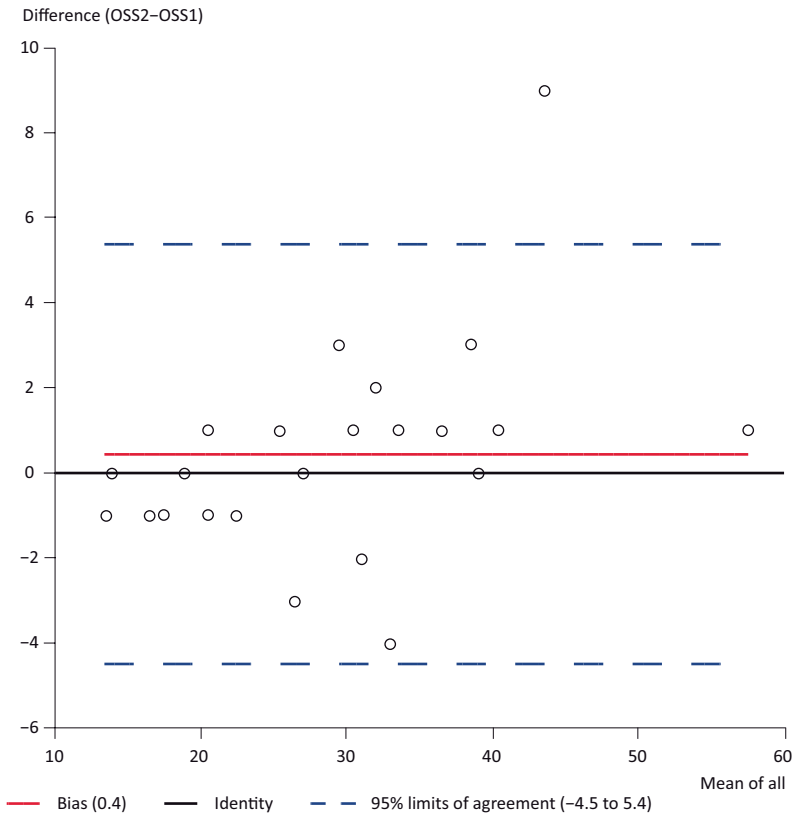
Internal consistency of the 12 items comprising the Oxford Shoulder Score.

OSS question	Mean score ( $\pm$ SD)	Corrected item-total correlation	Alpha if item removed
1	3.43 ( $\pm$ 1.04)	0.54	0.93
2	2.31 ( $\pm$ 0.91)	0.73	0.92
3	1.65 ( $\pm$ 0.97)	0.66	0.92
4	1.48 ( $\pm$ 0.91)	0.63	0.92
5	1.75 ( $\pm$ 1.13)	0.70	0.92
6	2.04 ( $\pm$ 1.27)	0.77	0.92
7	2.47 ( $\pm$ 1.26)	0.74	0.92
8	3.08 ( $\pm$ 0.98)	0.65	0.92
9	2.74 ( $\pm$ 1.35)	0.70	0.93
10	1.99 ( $\pm$ 1.13)	0.76	0.92
11	3.05 ( $\pm$ 1.21)	0.76	0.92
12	3.51 ( $\pm$ 1.29)	0.62	0.92
Total alpha	–	–	0.93

OSS = Oxford Shoulder Score; SD = standard deviation.


**FIGURE 2**

Bland-Altman plot showing the difference between the Oxford Shoulder Scores of the first (OSS1) and second visit (OSS2).



Although never validated, the CS has prevailed as the most used outcome score in Europe for more than two decades. Like other authors (0.64 Berendes et al [9]; 0.73 Murena et al [8]) we found that the OSS and the CS correlated well within the population studied. A correlation coefficient of 0.74 indicates, however, that these two outcome measures are not identical and probably also reflects the differences of perception of pain and function between interviewer and patient. Our results are in-line with those of the validated Dutch [9] and German [10] versions and also correspond well with the results of the original OSS study [4].

Reliability was high in our setup. The use of a 72-hour interval between test and retest may be questioned and the results might have been different if the patients had been tested at a 14-day interval as often chosen for patients with hip disorders [11]. In patients with shoulder disorders, symptoms are more volatile and patients' pain perception may vary from one day to the next. A shorter test interval is therefore desirable for this group of patients (24-48 h, Berendes et al [9]; 24-72 h, Huber et al [10]; 48 h, Murena et al [8]). 76% of patients returned the questionnaire for retest. All the re-

turned questionnaires were complete in the test and the retest. A Bland Altman difference plot of the test and retest scores indicated a low bias between the two tests and no discrepancies between scores at either end of the scales. There were no floor or ceiling effects and there is no reason to question the use of the Danish OSS, even in some of highly stigmatized patients of our study group.

Internal consistency of the Danish OSS expressed as Cronbach's alpha was high ( $R = 0.93$ ). Except for Item 1, the correlation between the individual items making up the OSS was also high ( $> 0.62$ ). Item 1 concerns pain experienced while doing daily activities and showed the lowest corrected item-total correlation (0.54) of the present study. This suggests that pain has less internal consistency than the rest of the OSS items. Of paramount importance is the fact that the OSS report pain within the past four weeks. Thus, the OSS is not a "here and now" test.

A potential drawback of patient-administered questionnaires such as the OSS is that these may be unsuited in contexts where patient and surgeon perceptions differ. Consequently, the value of self-reported outcome may be limited in insurance cases or when litigation is considered. The OSS may also be of limited usefulness in situations where workers' compensation claims are in play. Such cases need further study.

Another limitation of this study was our inability to compare the OSS and the CS with other established scores such as the Disability of Arm Shoulder and Hand (DASH) or the Short Form (SF)-36.

Finally we did not include sensitivity to change in our study. The ability to detect changes to intervention such as surgery would have added strength to the validation process.

## CONCLUSION

The ideal scoring system should be simple, effective and easy to use. It should also be strongly weighted towards functional outcome. The psychometric properties of the Danish version of OSS showed good validity with a substantial correlation between the OSS and the CS. It was easy to use and the time to perform the tests and to fill out the form was acceptably short. We also found a high reliability over time. Based on our findings we believe that the Danish translation of the OSS is reliable, and the simplicity of the questionnaire makes it a valuable tool in the assessment of patients with degenerative or post-traumatic shoulder diseases.

**CORRESPONDENCE:** Lars Henrik Frich, Department of Orthopaedic Surgery, Odense University Hospital, 5000 Odense C, Denmark.  
E-mail: lars.henrik.frich@ouh.regionsyddanmark.dk

**ACCEPTED:** 24 August 2011

**CONFLICTS OF INTEREST:** none

**ACKNOWLEDGEMENTS:** We thank *Tobias Wirenfeldt Klausen* for statistical analysis and *Niels Fryd, Peter Elkjær* and *Heidi Egmosse Busch* for data collection.

#### LITERATURE

1. Rekola KE, Keinänen-Kiukaanniemi S, Takala J. Use of primary health services in sparsely populated country districts by patients with musculoskeletal symptoms: consultations with a physician. *J Epidemiol Com Health* 1993;47:153-7.
2. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. *J Shoulder Elbow Surg* 1998;7:565-72.
3. Ragab AA. Validity of self-assessment outcome questionnaires: patient-physician discrepancy in outcome interpretation. *Biomed Sci Instrum* 2003;39:579-84.
4. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br* 1996;78:593-600.
5. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993;46:1417-32.
6. Constant CR, Gerber C, Emery RJ et al. A review of the Constant score: modifications and guidelines for its use. *J Shoulder Elbow Surg* 2008;17:355-61.
7. Dawson J, Rogers K, Fitzpatrick R et al. The Oxford shoulder score revisited. *Arch Orthop Trauma Surg* 2009;129:119-23.
8. Murena L, Vulcano E, D'Angelo F et al. Italian cross-cultural adaptation and validation of the Oxford shoulder score. *J Shoulder Elbow Surg* 2010;19:335-41.
9. Berendes T, Pilot P, Willems J et al. Validation of the Dutch version of the Oxford Shoulder Score. *J Shoulder Elbow Surg* 2010;19:829-36.
10. Huber W, Hofstaetter JG, Hanslik-Schnabel B et al. The German version of the Oxford shoulder score cross-cultural adaptation and validation. *Arch Orthop Trauma Surg* 2004;124:531-6.
11. Gosens T, Hoefnagels NH, de Vet RC et al. The "Oxford Hip Score": the translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. *Acta Orthop* 2005;76:204-11.