

Translation and validation of the Western Ontario Osteoarthritis of the Shoulder (WOOS) index – the Danish version

Jeppe V Rasmussen¹

John Jakobsen²

Bo S Olsen¹

Stig Brorson¹

¹Department of Orthopedic Surgery, Herlev University Hospital, Herlev, Denmark; ²Department of Orthopedic Surgery, Aalborg University Hospital, Thisted, Denmark

Background and purpose: The Western Ontario Osteoarthritis of the Shoulder (WOOS) index is a patient-reported, disease-specific questionnaire for the measurement of the quality-of-life in patients with osteoarthritis. The purpose of this study was to describe the process used to translate the WOOS into Danish and to test the translation in a Danish population, in terms of validity, reliability, and responsiveness.

Material and methods: The translation of the WOOS was done according to international standardized guidelines. The psychometric properties were tested in 20 consecutive patients. The eligibility criteria were: a diagnosis of osteoarthritis without symptomatic rotator cuff pathology and treated with primary shoulder replacement. Patients were excluded only in the case of other pathology of the upper extremity or in the case of cognitive or linguistic impairment compromising the ability to complete the questionnaires.

Results: The Pearson's correlation coefficient between the WOOS and the Constant–Murley score (CMS), preoperatively was 0.62 ($P = 0.004$) and the correlation between the changes of score for the WOOS and CMS was 0.73 ($P < 0.001$). The correlation coefficient between the WOOS and the CMS, SF-36, and the Oxford Shoulder Score postoperatively was 0.82 ($P < 0.001$), 0.48 ($P = 0.03$), and 0.82 ($P < 0.001$), respectively. There were no floor and ceiling effects. The Cronbach's alpha was 0.98. The intraclass correlation coefficient between test and retest was 0.96. The standardized response mean was 1.41, and effect size was 2.32.

Conclusion: We have shown that the Danish version of the WOOS, translated according to international standardized guidelines, has substantial statistical and clinical psychometric properties at the same level as was described for the original version.

Keywords: outcome assessment, cross-cultural adaption, questionnaire

Introduction

With the advance of modern shoulder surgery, patient-reported outcome has become popular and is increasingly used. Since the early 1980s, generic questionnaires, such as the Short Form (SF)-36[®] Health Survey (QualityMetric Inc, Lincoln, RI, USA), the Sickness Impact Profile, and the Nottingham Health Profile have been available.^{1–3} More recently a wide variety of shoulder-specific and disease-specific questionnaires have been developed.^{4–8} All of these patient-reported outcomes were developed in Anglo-Saxon countries and tested in native English-speaking populations.

Patient-reported outcomes are often used in countries with languages and with cultural traditions other than those in which they were originally developed and tested. Despite this, there are relatively few translations and validations of shoulder-specific questionnaires. It is essential to use questionnaires that have been translated

Correspondence: Jeppe V Rasmussen
Herlev University Hospital, Department
of Orthopedic Surgery, Herlev Ringvej
75, 2730 Herlev, Denmark
Tel +45 40 56 18 59
Fax +45 44 88 30 33
Email jevera01@heh.regionh.dk

according to international standardized guidelines and with psychometric properties that have been retested and culturally adapted. A standardized translation and evaluation of the Western Ontario Osteoarthritis of the Shoulder (WOOS) index is important, not only because it is used as a patient-reported outcome in scientific literature, but in particular, because it is used as a patient-reported outcome in the Danish Shoulder Arthroplasty Registry.⁹

The purpose of this study was to describe the process used to translate the WOOS into Danish and to test the translation in a Danish population, in terms of validity, reliability, and responsiveness.

Material and methods

Translation

The translation of the WOOS was done according to the recommendation of Guillemin et al.¹⁰ First, two bilingual orthopedic surgeons with Danish as their first language, working independently, translated the original English version into Danish. In the translation process, equality of sense rather than equality of word was given priority. Then, during a conference, consensus was achieved on a first preliminary Danish version based on the two translations. Subsequently, two professional translators with English as their first language translated this version back into English. Neither of these two professionals had any medical knowledge or knew anything about the WOOS. Finally, a committee compared the source and the final translated Danish version. The committee consisted of orthopedic surgeons with special interest in shoulder surgery. For a preliminary test, the final Danish version was tested for comprehensibility in a group of 20 consecutive patients, and no further changes were required.

Outcome assessment tools

Western Ontario Osteoarthritis of the Shoulder Index

The WOOS is a patient-administrated, disease-specific questionnaire for measurement of the quality-of-life of patients with osteoarthritis.⁷ It provides scores on four domains: (1) physical symptoms; (2) sport, recreation, and work; (3) lifestyle; and (4) emotions. Each question is answered using a visual analog scale with a possible score ranging from 0 to 100. There are 19 questions, and the total score ranges from 0 to 1900. A score of 1900 signifies that the patient has an extreme decrease in the shoulder-related quality of life, whereas a score of 0 signifies that the patient has no decrease in shoulder-related quality of life. For simplicity of presentation, the raw scores are often converted to

a percentage of the maximum score, as was done in this present study.

Constant–Murley Score (CMS)

The Constant–Murley Score (CMS) includes an assessment of: (1) pain; (2) activities of daily living (ADL); (3) range of motion; and (4) strength. There are a possible 35 points given for the subjective assessment of pain and the ability to perform ADL. There are also a possible 65 points given for an objective assessment, of which 40 points are allocated to range of motion and 25 points are allocated to strength. The maximum of 100 point indicates a shoulder with no disability. We used the modified version described by Constant and colleagues in 2008.¹¹ The CMS was not adjusted for age or sex.

Oxford Shoulder Score

The Oxford Shoulder Score (OSS) was conceived as a measurement tool for the assessment of outcomes of shoulder surgery.⁵ It has been tested and validated in patients with primary or secondary arthritis, inflammatory arthritis, and in patients with impingement with or without rotator cuff pathology. The OSS is a 12-item questionnaire, with each item scored from 1 to 5; thus the total score ranges from 12 (best score) to 60 (worst score). For simplicity of presentation, the raw scores are converted to a percentage of the maximum score. We used a recently validated Danish version.¹²

The SF-36

The SF-36 is a 36-item questionnaire, widely used to assess general health. It provides scores on different domains: (1) physical symptoms; (2) limitations caused by emotional problems; (3) general health; (4) vitality; (5) pain; and (6) perception of general health. The questions require different types of answers: some are to be answered in two parts, whereas others require answers from a scale with up to six parts. The total score is converted to a percentage of a maximum score.^{3,13} We used a validated Danish version.¹⁴

Patients

The eligibility criteria were the diagnosis of osteoarthritis without symptomatic rotator cuff pathology and treatment with primary shoulder replacement. Patients were excluded only in the case of another pathology of the upper extremity or in the case of cognitive or linguistic impairment compromising the ability to complete the questionnaires.

We included 20 consecutive patients diagnosed with osteoarthritis and treated with shoulder replacement

between May 2010 and April 2012 at the Department of Orthopedic Surgery, Herlev Hospital, Denmark. Information about other pathology of the upper extremity or cognitive or linguistic impairment as a consequence of medical disease was obtained through medical records, and patients were asked whether they could read Danish adequately before they were included in the study. There were six men and 14 women participants, with median age 69 years (range 46–89 years). The patients were evaluated preoperatively using the WOOS and CMS. At the 1-year follow-up examination (range 10–14 months postoperatively), the patients were evaluated using the WOOS, CMS, OSS, and SF-36. After another 7 days (3–10 days), the patients were once again evaluated using the WOOS and CMS. The CMS was evaluated by a single surgeon, and the patients themselves completed the questionnaires (WOOS, SF-36, and OSS). The sequence in which the three questionnaires were administered was random. There were no missing values for any measures.

Psychometric testing and statistics

Construct validity compares the outcome measurement tool to a gold standard when no “true value” is available. Pearson’s correlation coefficient was used to correlate the preoperative measurement and the changes of score for the WOOS and CMS. Pearson’s correlation coefficient was also used to correlate the WOOS, CMS, SF-36, and OSS postoperatively.

Content validity assesses whether the items measure the full range of the actual question. We used the “floor and ceiling effect” to assess this. The distribution of the results of each item, both pre- and postoperatively, were presented and evaluated, with the “floor” (worst) considered to be a score between 0 and 1 and the “ceiling” (best) score to be between 99 and 100.

Internal consistency designates the correlation between items that make up the score and is assessed using Cronbach’s alpha. The range of scores in this test varies between 0 and 1, and higher scores are better; however, a score above 0.95 is not necessarily desirable since it may indicate that some questions deal with the same parameter. We evaluated the internal consistency of the postoperative measurement.

The test-retest reliability was measured as the agreement between two measurements taken 7 days apart and expressed as the intraclass correlation coefficient (ICC). The ICC was measured for the total score and for the four domains.

The responsiveness (sensitivity) of the instrument to changes occurring between baseline and posttreatment was

analyzed using: the standardized response mean (SRM), calculated as the difference between the preoperative mean score and the postoperative mean score divided by the standard deviation (SD) of the difference; and the effect size (ES), calculated as the difference between the postoperative mean score and the preoperative mean score divided by the preoperative SD. We compared the results of the WOOS with the results of the CMS. Furthermore, floor and ceiling effects also have an effect on the responsiveness – this is so for the floor effect, when an individual scores at the bottom of the scale and no further decline is possible and for the ceiling effect, when an individual scores at the top of the scale and no further improvement is possible.

The paired *t*-test was used to examine the improvement in the WOOS and CMS between the preoperative and the postoperative measurement and to compare the test and retest measurements.

The analysis was performed with use of SPSS (version 19.0; IBM Corporation, Armonk, NY, USA). The level of statistical significance was set at 0.05.

Results

Construct validity

The Pearson’s correlation coefficient between the WOOS and CMS preoperatively was 0.62 ($P = 0.004$), and the correlation between the changes of score for the WOOS and CMS was 0.73 ($P < 0.001$). The correlation coefficient between the WOOS and the CMS, SF-36, and OSS postoperatively was 0.82 ($P < 0.001$), 0.48 ($P = 0.03$), and 0.82 ($P < 0.001$), respectively (Table 1).

Content validity

There was no floor and ceiling effect preoperatively or postoperatively for the total WOOS and an adequate effect for some of the domains (Table 2).

Table 1 Correlation between measures

	WOOS	CMS	SF-36	OSS
WOOS	–	0.82 $P < 0.001$	0.48 $P = 0.03$	0.82 $P < 0.001$
CMS	0.82 $P < 0.001$	–	0.35 $P = 0.02$	0.90 $P < 0.001$
SF-36	0.48 $P = 0.03$	0.35 $P = 0.02$	–	0.57 $P = 0.01$
OSS	0.82 $P < 0.001$	0.90 $P < 0.001$	0.57 $P = 0.01$	–

Note: Pearson’s correlation coefficient between the WOOS, CMS, SF-36®, and the OSS.

Abbreviations: CMS, Constant–Murley Score; OSS, Oxford Shoulder Score; SF, Short Form Health Survey; WOOS, Western Ontario Osteoarthritis of the Shoulder (index).

Table 2 Floor and ceiling of WOOS

Domains	Floor/ceiling effect	
	Preoperative	Postoperative
Physical symptoms	None	None
Sport, recreation, and work	5% (1) F	5% (1) C
Lifestyle	5% (1) F	5% (1) C
Emotions	None	10% (2) C
Total WOOS	None	None
CMS	None	None

Note: Floor and ceiling of WOOS given as the percentage of answers with floor or ceiling effect with number in brackets.

Abbreviations: C, ceiling; CMS, Constant–Murley Score; F, floor; WOOS, Western Ontario Osteoarthritis of the Shoulder index.

Internal consistency

The Cronbach's alpha was 0.98. Elimination of one item in all 19 cases resulted in values between 0.97 and 0.98. Questions 5 and 6 had correlations, with a total score of 0.52 and 0.64 respectively. All other items had correlations with a total score of >0.75 .

Test-retest reliability

The mean WOOS for the first and second measurement were 72.9 and 73.7, respectively, with a mean difference of 0.82 and SD 7.1 (95% confidence interval [CI]: -4.12 ; 2.48) ($P = 0.61$). The ICC for the total WOOS was 0.96 (95% CI: 0.91; 0.99), for the domains physical symptoms was 0.94 (95% CI: 0.85; 0.98), for sport/recreation/work was 0.88 (95% CI: 0.73; 0.95), for lifestyle was 0.95 (95% CI: 0.87; 0.98), and for emotions was 0.97 (95% CI: 0.92; 0.99). All values were highly statistical significant ($P < 0.001$).

Responsiveness

A total of 19 out of 20 patients reported an improvement in the WOOS. The mean WOOS score preoperatively was 34.1 and postoperatively was 72.9, with a mean improvement of 38.7 (95% CI: 25.8; 51.6) ($P < 0.0001$). This can be compared with a mean CMS preoperatively of 25.8 and postoperatively of 57.1, with a mean improvement of 31.3 (95% CI: 22.6; 40.0) ($P < 0.0001$). The SRM was 1.41 and ES was 2.32 for the WOOS and was 1.69 and 1.98, respectively, for the CMS. The SRM and ES for the domains are presented in Table 3.

Discussion

The WOOS was translated into Danish, according to international standardized guidelines. There was no need for substantial changes compared with the original English version, and the Danish translation of the WOOS had psychometric properties at the same level as was described for the original English version.⁷

Table 3 The responsiveness of the WOOS and CMS

Domains	SRM	ES
Physical	1.46	2.05
Sport, recreation, and work	1.36	2.54
Lifestyle	1.13	1.96
Emotions	0.89	1.07
Total WOOS	1.41	2.32
CMS	1.69	1.98

Note: The SRM and ES of the four domains of the WOOS, the total WOOS, and the CMS.

Abbreviations: CMS, Constant–Murley Score; ES, effect size; SRM, standardized response mean; WOOS, Western Ontario Osteoarthritis of the Shoulder index.

The Pearson's correlation coefficient between the WOOS and CMS scores preoperatively was 0.62, and the correlation between the changes of score was 0.73. The correlations were strong and equivalent to the results presented for the original English version reporting 0.73 and 0.69, respectively. The postoperative evaluation showed a similar strong correlation between the WOOS, CMS, and OSS, of more than 0.80 and was highly statistical significant. The SF-36, which is a global measure of health, was found to have a rather poor correlation with the shoulder-specific measures. This is similar to the findings in the original English paper, where the correlation in the change of score between the WOOS and SF-36 was 0.29. The Cronbach's alpha was higher than 0.95, and thus, some questions may be redundant. This needs to be further examined and confirmed in studies with larger sample sizes. The test–retest reliability of the WOOS was high, with an excellent ICC for the domains and for the total score, similar to the results presented in the original English version, which reported an ICC of the domains between 0.87 and 0.95 and an overall ICC value of 0.96. The value for the SRM in the current study was similar to the SRM of 1.910 reported in the original English version. The original English version did not report ES. In this present study the SRM for WOOS was similar to that of CMS. Furthermore, there were no floor and ceiling effect for total WOOS and no or an adequate effect for the domains.

An outcome measurement tool, such as the WOOS, is validated by a comparison against a gold standard since no "true value" is available; however, there is no consensus of a gold standard to evaluate shoulder function either. In this present study, we chose to compare the Danish version of the WOOS with the CMS because the European Society of Shoulder and Elbow Surgery has promoted the use of the CMS in manuscript submission, making it one of the most widely used shoulder measures during the last decades. Furthermore, the CMS was also used to test the psychometric properties of the original English version of the WOOS.

Nevertheless, the Danish and even the international version of the CMS, based on the modified guidelines described by Constant et al,¹¹ have never been validated.

In the original English version of the WOOS, the responsiveness was analyzed using preoperative and 3-month postoperative measurements. We used a preoperative and a 1-year postoperative measurement instead. In cases where a patient-reported outcome measure is used to detect the benefit from an operation with shoulder replacement, a 3-month postoperative evaluation may more often be influenced by temporary pain related to rehabilitation exercise or by continuing use of pain medication that could make the analysis of responsiveness imprecise. Furthermore, some patients have a protracted rehabilitation, with only small changes the first 3 months, and the majority of patients may experience changes in shoulder function until 1 year postprocedure. Finally, the WOOS is used to measure outcomes 1 year postoperatively in the setting of the Danish Shoulder Arthroplasty Registry, and the responsiveness of the Danish version of WOOS throughout this time interval is desirable. Nevertheless, the use of different time intervals may influence the responsiveness, and a direct comparison between the responsiveness found in this study and in the original English version may not be justified.

The included patients were treated with shoulder replacement, and all the included patients except one had an improvement in the total WOOS score between the preoperative and the postoperative measurements. As a consequence, we cannot justify any conclusion about the ability of the Danish version of the WOOS to detect changes when the perceived shoulder function decreases. To our knowledge, the responsiveness of WOOS has only been tested using preoperative and postoperative measurements with an expected improvement in WOOS. The responsiveness of the WOOS in a population that also includes patients with an expected decrease in the perceived shoulder function is a subject for future research.

Patient-reported outcome has become popular and is increasingly used. The most important advantages are that questionnaires do not require the time of an orthopedic surgeon and that they can be completed by the patient and returned by mail without attending the hospital. Thus, a questionnaire is likely to have a high compliance compared with radiological and clinical examinations, such as the CMS. Furthermore, any influence of interobserver reliability is eliminated when questionnaires are used. Finally, it is cost effective and suitable in studies with large populations, such as in registry studies. There has been dispute about which patient-reported outcome is most appropriate to use. The WOOS has some advantages compared with other shoulder-specific questionnaires, such

as the OSS. Having a visual analog scale may be preferable to questionnaires with predefined options to select from since some patients might find that his or her situation does not fit into one of the predefined options. The WOOS also evaluates the shoulder function during the preceding week, whereas the OSS evaluates shoulder function during the previous 4 weeks. One could argue that patients might have difficulties remembering 4 weeks back and that the shoulder function may vary during a 4-week period. Nevertheless, there are also some potential limitations of the WOOS. Some patients may find the principle of a visual analog scale difficult, and the WOOS is restricted to patients diagnosed with osteoarthritis only.

One limitation of our study is that we did not perform a power analysis when initiating the study and that rather few patients were included. Furthermore, we did not register the time employed in filling out the WOOS. The strength of this study is that we were able to compare the results of the WOOS with an established outcome measure, the CMS. Furthermore, we included a population comparable with the population tested in the original English version.

We tested the Danish version of the WOOS with classical test theory, analyzing reliability, validity, and responsiveness. Modern test theory, analyzing the dimensional structure of the WOOS using Rasch analysis is a subject for future research.

In clinical research, it is important to define the minimal clinically important difference of the measures used. The minimal clinically important difference was not defined in the publication describing the original English version of WOOS, but it has recently been suggested to be 190 points, equivalent to 10% of a maximum score.¹⁵ In future, consensus of this limit needs to be established.

In summary, we have shown that the Danish version of the WOOS, translated according to international standardized guidelines, has substantial psychometric properties, at the same level as was described for the original version. We recommend the WOOS when evaluating patients with osteoarthritis of the shoulder.

Acknowledgments

The authors thank Tobias Winfeldt Klausen for statistical advice.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care*. 1981;19(8):787–805.

2. Hunt SM, McEwen J. The development of a subjective health indicator. *Sociol Health Illn.* 1980;2(3):231–246.
3. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30(6):473–483.
4. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res.* 1987;(214):160–164.
5. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br.* 1996;78(4):593–600.
6. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med.* 1996;29(6):602–608.
7. Lo IK, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: The Western Ontario Osteoarthritis of the Shoulder (WOOS) index. *Osteoarthr Cartil.* 2001;9(8):771–778.
8. Richards RR, An KN, Bigliani LU, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg.* 1994;3(6):347–352.
9. Rasmussen JV, Jakobsen J, Brorson S, Olsen BS. The Danish Shoulder Arthroplasty Registry: clinical outcome and short-term survival of 2,137 primary shoulder replacements. *Acta Orthop.* 2012;83(2):171–173.
10. 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(12):1417–1432.
11. Constant CR, Gerber C, Emery RJ, Søjbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. *J Shoulder Elbow Surg.* 2008;17(2):355–361.
12. Frich LH, Noergaard PM, Brorson S. Validation of the Danish version of Oxford Shoulder Score. *Dan Med Bull.* 2011;58(11):A4335.
13. Ware JE. SF-36 health survey update. *Spine.* 2000;25(24):3130–3139.
14. Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P. The Danish SF-36 Health Survey: translation and preliminary validity studies. *J Clin Epidemiol.* 1998;51(11):991–999.
15. Polk A, Rasmussen JV, Brorson S, Olsen BS. Reliability of patient-reported functional outcome in a joint replacement registry. A comparison of primary responders and non-responders in the Danish Shoulder Arthroplasty Registry. *Acta Orthop.* 2013;84(1):12–17.

Patient Related Outcome Measures

Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>

Dovepress

The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.