

# Validation of the Danish version of the Oxford Elbow Score

Hans Christian Plaschke, Andreas Jørgensen, Theis Muncholm Thillemann, Stig Brorson & Bo Sanderhoff Olsen

## ABSTRACT

**INTRODUCTION:** The Oxford Elbow Score (OES) is a patient-related outcome measure quantifying quality of life in relation to elbow disorders. This 12-item patient-administered English questionnaire comprises three domains: function, social-psychological status and pain. The purpose of this study was to examine the structure and reliability of the Danish version of the OES (D-OES).

**MATERIAL AND METHODS:** A total of 130 patients who had total elbow arthroplasty (TEA) surgery performed in the period from 1981 to 2008 completed the D-OES. The Disability of the Arm, Shoulder and Hand score (DASH) and the Mayo Elbow Performance Score (MEPS) were also completed with a view to examining the convergent validity of the D-OES. In 45 cases, the questionnaire was completed twice to allow for calculation of test-retest reliability. The structure of the questionnaire was tested using Rasch analysis.

**RESULTS:** Overall internal consistency tested by Cronbach's alpha was 0.99. The test-retest reliability correlation coefficient was 0.99. Expressed in terms of Pearson's correlation coefficient, the convergent validity of the D-OES's functional, social-psychological and pain domain was 0.78, 0.80 and 0.81, respectively, for the MEPS and  $-0.66$ ,  $-0.58$  and  $-0.49$ , respectively, for the DASH. The 12 items of the D-OES fitted the Rasch model, and the multidimensionality of the OES was confirmed.

**CONCLUSION:** The Danish 12-item OES is a valid and reliable-patient administered questionnaire that can be used to quantify quality of life in patients with TEA.

**FUNDING:** not relevant.

**TRIAL REGISTRATION:** not relevant.

Patient-reported outcome measures (PROMs) quantify the patients' own experience in relation to a health condition and its therapy [1]. PROMs are useful for the evaluation of surgical interventions because they express how the patient feels about a certain treatment. Since PROMs are independent of the surgical team, they are less biased than standard clinical assessments [2].

Many scoring systems have been used in elbow disorders. However, only few of these have been validated and many assess only few aspects of elbow function [3]. The Oxford Elbow Score (OES) is a 12-item patient-administered questionnaire that measures the quality of

life of patients with disorders of the elbow joint (Table 1). The OES comprises three domains: function, pain and social-psychological status. Each domain is covered by four questions whose answers are recorded on a five-point Likert scale. Each item within a domain contributes equally to the total score of the given domain. Each domain is transformed into a 100-point metric scale with 0 being the worst and 100 being the best score [4].

The Disability of the Arm, Shoulder and Hand questionnaire (DASH) is a 30-item questionnaire for measuring the symptoms and function in patients with musculoskeletal disorders of the upper limb. The DASH is measured on a 100-point scale with 100 being the worst and 0 being the best score [5, 6]. The Danish version of the DASH has previously been validated [7]. The DASH is not specific for the elbow.

The Mayo Elbow Performance Score (MEPS) is a physician-administered tool that covers four domains: pain (45 points), ulno-humeral movement (twenty 20 points), stability (ten points) and the patient's ability to accomplish five everyday functional tasks (25 points) [8].

The OES has been found to be reliable and valid in previous studies from England and the Netherlands [4, 9, 10]. In a study by Dawson et al [11], the OES was also found to have good responsiveness and ability to detect changes six months after surgery. The purpose of this study was to translate the OES into a Danish version (D-OES) for publication, to validate the Danish translation and to retest the psychometric properties of the D-OES within a Danish setting for patients with TEA.

## MATERIAL AND METHODS

A total of 150 patients who had undergone TEA surgery were identified. In all, 130 patients (87%) were included at Herlev Hospital from October 2010 to October 2011. The study group comprised 156 TEAs inserted in Eastern Denmark from 1981 to 2008. Both primary and revision TEAs were included. The average age was 68 years (range 41-91 years).

To test the convergent validity, we compared the D-OES to the MEPS, which has been a standard tool for measuring outcome after total elbow arthroplasty. We further compared the D-OES to the Danish validated DASH score. Patients were evaluated with the MEPS within a clinical setting, whereas the patients had filled

## ORIGINAL ARTICLE

Shoulder and Elbow Clinic, Orthopaedic Department, Herlev Hospital

Dan Med J  
2013;60(10):A4714

 TABLE 1

The 12 items of the Oxford Elbow Score.

“During the past 4 weeks ...”

1. Have you had difficulty lifting things in your home, such as putting out the rubbish, because of your elbow problem?
2. Have you had difficulty carrying bags of shopping, because of your elbow problem?
3. Have you had any difficulty washing yourself all over, because of your elbow problem?
4. Have you had any difficulty dressing yourself, because of your elbow problem?
5. Have you felt that your elbow problem is “controlling your life”?
6. How much has your elbow problem been “on your mind”?
7. Have you been troubled by pain from your elbow in bed at night?
8. How often has your elbow pain interfered with your sleeping?
9. How much has your elbow problem interfered with your usual work or everyday activities?
10. Has your elbow problem limited your ability to take part in leisure activities that you enjoy doing?
11. How would you describe the worst pain you had from your elbow?
12. How would you describe the pain you usually had from your elbow?

out the DASH and the D-OES at home. The OES was translated into Danish according to guidelines provided by Beaton et al [12]. Two bilingual persons who were Danish native speakers translated the OES into Danish. One of the translators had clinical experience and one did not. Two independent bilingual native English speakers then back-translated the Danish version into English. This back translation (**Figure 1**) was then compared to the original OES, evaluated for mistakes and finally accepted by the authors of the original OES.

The median follow-up time from surgery to evaluation was 6.7 years (range 2-20 years). The patients were all considered to be in a stable state concerning the affected elbow.



The Oxford Elbow Score can easily be filled in at home in a few minutes.

A total of 97% of the D-OES questionnaires and 92% of the DASH questionnaires were completed correctly. All the patients were evaluated with the MEPS. In all, 50 patients were asked to fill in the D-OES two times at different occasions at a 14-day interval. In all, 45 patients (90%) completed both of the D-OES questionnaires.

### Statistics

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

Test-retest reliability is a measure of the consistency of a psychological test. The test-retest reliability is expressed as the intraclass correlation coefficient. This type of reliability assumes that the condition is stable. No treatment was given between the two evaluations. Cronbach's alpha was calculated as a measure of internal consistency and the Pearson correlation coefficient was calculated between the D-OES, the MEPS and the DASH. SPSS Statistics version 20.0 was used to calculate reliability, internal consistency and correlation.

Rasch analysis was performed using a rating scale model with Winsteps Rasch Measurement Version 3.75.1. The following analyses were performed: Construction of the person and item map (Wright map), testing of the fit between the data and the model, estimation of the person and item reliability and separation coefficient, testing of the ordering of the categories, and analysis of the dimensionality.

*Trial registration:* not relevant.

### RESULTS

The internal consistency calculated as Cronbach's alpha was 0.998 (95% confidence interval (CI): 0.997-0.999). Expressed by the Pearson's correlation coefficient, the convergent validity of the D-OES functional, social-psychological and pain domains were 0.78, 0.80 and 0.81, respectively, for the MEPS and -0.66, -0.58 and -0.49, respectively, for the DASH. Between the MEPS and the DASH, the correlation coefficient was -0.572.

The intraclass correlation coefficient was 0.998, 0.996 and 0.996 for the functional, social-psychological and pain domains, respectively.

The Wright map values for items and patients were in order based on the logit scale and the default mean difficulty was set to zero. The majority of the patients were located opposite and above the items which indicates an overall good D-OES score. The mean person estimate was 2.2 (95% CI: 1.8-2.6).

In **Table 2**, the items are placed according to item difficulty with the most difficult item at the top. Chi-square fit statistics were calculated to determine how well the data fit the Rasch model. The in-fit mean square (MNSQ) represents the information-weighted mean

square residuals difference between observed and expected responses. These statistics are sensitive to unexpected responses near the person's ability level. The outfit MNSQ represents the usual unweighted mean square residual and is more sensitive to outliers than the infit MNSQ. Values should range between 0.6 and 1.4 for rating scales or 0.5 to 1.7 for clinical observations [13]. All the items in the OES had infit and outfit MNSQ values between 0.74 and 1.36.

The item reliability coefficient was 0.97, and the item separation coefficient was 5.37. The person reliability index was 0.90 and the person separation index was 3.04. The person separation index was used to calculate the number of distinct levels of quality of life (strata) that could be distinguished (strata =  $(4 \times \text{person separation index} + 1)/3$ ) = 4.4.

The five categories of the D-OES were represented with increasing count (Table 3). The observed average is the average of the (person measures – item difficulties) modelled to produce the responses observed in the category. Outfit MNSQ is the average of the outfit MNSQs associated with the responses in each category. The expected values for all categories are 1.0. This statistic is sensitive to grossly unexpected responses. Only values greater than 1.5 are problematic. The Rasch Andrich thresholds are the calibrated measures of the transitions between adjacent categories. They are the points on a Likert scale where probability of being observed in either of two adjacent categories is equal. The Rasch Andrich threshold is expected to increase monotonically on a rating scale when categories are ordered.

A Rasch principal component analysis (PCA) of the residuals of the OES was performed. The raw variance of the OES explained by the Rasch measure was 73.3% (73.1% was expected by the model). The unexplained variance in the first contrast was 9.3% (4.2 eigenvalue units), and the second contrast was 5.5% (2.5 eigenvalue units). The first contrast consisted of the four pain items. The second contrast consisted of the four social psychological items. When data fit the Rasch model, the Rasch dimension is the only dimension. Any other dimension in the data must explain at least two items (2.0 eigenvalue units) worth of variance.

## DISCUSSION

We found a high internal consistency and a high test-retest reliability for all three domains of the D-OES. In the Dutch validation study [9], the investigators also found a high test-retest reliability for the function (0.87), the pain (0.89) and the social social-psychological domain (0.87). Dawson et al [11] found equally high reliabilities of 0.89, 0.98 and 0.87, respectively.

The D-OES corresponds well with the MEPS in all three domains, whereas the DASH had less correspond-

 FIGURE 1

The Danish version of the Oxford Elbow Score. Sæt kun et x per spørgsmål.

Oxford Elbow Score	Patient label	
Problemer med Deres albue		7. Inden for de sidste 4 uger ...
<i>inden for de sidste 4 uger</i>		Har De da været besværet af smerter fra Deres albue i sengen i løbet af natten?
Højre <input type="checkbox"/> Venstre <input type="checkbox"/> Dato:		Slet ikke <input type="checkbox"/> Nogle nætter <input type="checkbox"/> 1-2 nætter <input type="checkbox"/> De fleste nætter <input type="checkbox"/> Alle nætter <input type="checkbox"/>
1. Inden for de sidste 4 uger ...		8. Inden for de sidste 4 uger ...
Har De da haft problemer med at løfte ting i hjemmet, såsom at bære skraldet ud pga. problemer med Deres albue?		Hvor ofte har Deres albuesmerter forstyrret Deres søvn?
Ingen problemer <input type="checkbox"/> Få problemer <input type="checkbox"/> Moderate problemer <input type="checkbox"/> Svære problemer <input type="checkbox"/> Det er umuligt <input type="checkbox"/>		Slet ikke <input type="checkbox"/> Lidt af tiden <input type="checkbox"/> Noget af tiden <input type="checkbox"/> Det meste af tiden <input type="checkbox"/> Hele tiden <input type="checkbox"/>
2. Inden for de sidste 4 uger ...		9. Inden for de sidste 4 uger ...
Har De da haft problemer med at bære indkøbsposer pga. problemer med Deres albue?		Hvor meget har Deres albue påvirket deres arbejde eller daglige aktiviteter?
Ingen problemer <input type="checkbox"/> Få problemer <input type="checkbox"/> Moderate problemer <input type="checkbox"/> Svære problemer <input type="checkbox"/> Det er umuligt <input type="checkbox"/>		Slet ikke <input type="checkbox"/> Lidt <input type="checkbox"/> Moderat <input type="checkbox"/> Meget <input type="checkbox"/> Fuldstændig <input type="checkbox"/>
3. Inden for de sidste 4 uger ...		10. Inden for de sidste 4 uger ...
Har De da haft problemer at vaske Dem selv over det hele pga. problemer med Deres albue?		Har Deres albue da begrænset Dem i deltagelse i fritidsaktiviteter, som De holder af?
Ingen problemer <input type="checkbox"/> Få problemer <input type="checkbox"/> Moderate problemer <input type="checkbox"/> Svære problemer <input type="checkbox"/> Det er umuligt <input type="checkbox"/>		Slet ikke <input type="checkbox"/> Lidt af tiden <input type="checkbox"/> Noget af tiden <input type="checkbox"/> Det meste af tiden <input type="checkbox"/> Hele tiden <input type="checkbox"/>
4. Inden for de sidste 4 uger ...		11. Inden for de sidste 4 uger ...
Har De da haft problemer med at klæde dem selv på pga. problemer med Deres albue?		Hvorledes vil De beskrive den værste smerte fra Deres albue?
Ingen problemer <input type="checkbox"/> Få problemer <input type="checkbox"/> Moderate problemer <input type="checkbox"/> Svære problemer <input type="checkbox"/> Det er umuligt <input type="checkbox"/>		Ingen smerter <input type="checkbox"/> Mild smerte <input type="checkbox"/> Moderat smerte <input type="checkbox"/> Voldsom smerte <input type="checkbox"/> Uudholdelig smerte <input type="checkbox"/>
5. Inden for de sidste 4 uger ...		12. Inden for de sidste 4 uger ...
Har De da følt, at Deres albue "styrer Deres liv"?		Hvorledes vil De beskrive den sædvanlige smerte fra Deres albue?
Slet ikke <input type="checkbox"/> Indimellem <input type="checkbox"/> Nogle dage <input type="checkbox"/> De fleste dage <input type="checkbox"/> Hver dag <input type="checkbox"/>		Ingen smerter <input type="checkbox"/> Mild smerte <input type="checkbox"/> Moderat smerte <input type="checkbox"/> Voldsom smerte <input type="checkbox"/> Uudholdelig smerte <input type="checkbox"/>
6. Inden for de sidste 4 uger ...		Oxford Elbow Score © Isis Innovation Limited, 2008. All rights reserved.
Hvor meget har De tænkt over Deres albue?		
Slet ikke <input type="checkbox"/> Lidt af tiden <input type="checkbox"/> Noget af tiden <input type="checkbox"/> Det meste af tiden <input type="checkbox"/> Hele tiden <input type="checkbox"/>		

ence with both the MEPS and the three domains of the D-OES. The majority of our patients were operated with TEA due to rheumatoid arthritis. These patients often also had affected wrist or shoulder problems to which the DASH score is sensitive as it is not a specific elbow score. This could explain the differences in correspondence between the DASH and both the D-OES and the MEPS as the two latter measures are specific to the elbow.

The original OES is valid, reliable and sensitive as a tool for assessing patients with elbow disorders [4, 10]. To ensure the validity for patients operated with TEA, we performed the Rasch analysis.

The data fitted the stringent Rasch model with good infit MNSQ and outfit MNSQ values. This indicates that

 TABLE 2

Item statistics of the Danish version of the Oxford Elbow Score. Measures are listed as the logits (natural logarithm of the odds).

Item	Count	Measure	Infit MNSQ	Outfit MNSQ
2	156	1.13	0.88	0.90
6	156	1.13	0.96	0.93
3	156	0.70	0.88	0.86
9	156	0.65	0.78	0.75
11	156	0.63	1.27	1.36
10	156	0.56	0.77	0.74
7	156	-0.27	1.30	1.27
4	156	-0.37	1.17	1.17
1	156	-0.53	0.94	0.90
5	156	-0.61	0.87	0.78
8	156	-1.10	1.16	0.91
12	156	-1.91	1.02	0.82

MNSQ = mean square.

 TABLE 3

Summary of the category structure of the Danish version of the Oxford Elbow Score.

Category, label/score	Observed count	Observed count, %	Observed average	Outfit MNSQ	Threshold
0	37	2	-2.55	1.18	-
1	160	9	-1.55	0.87	-3.54
2	344	18	0.25	0.76	-1.36
3	471	25	2.83	0.82	1.14
4	860	46	4.81	1.32	3.76

MNSQ = mean square.

the data were neither underfit and thereby lacking predictability, nor overfit and thereby over-predictable of any of the items. The person separation index and the person reliability index were excellent. The item separation coefficient and the item reliability coefficient were also excellent supporting that the sample size was large enough to confirm the item difficulty hierarchy. The category rating scale of the D-OES worked well. The patients were able to discriminate the five levels of the items, and the Rasch Andrich threshold increased monotonically. PCA showed that the Danish version of the OES was multidimensional. The four items of the pain domain and the four items of the social psychological domain were recognized as separate domains. The original version of the OES was based on English culture. The purpose of this study was to translate the original OES into a Danish version including any necessary cross-cultural adaptation to avoid misinterpretations. To ensure that respondents understand the questions as intended, it is important to translate using an accurate and approved method. We did not encounter questions that challenged national traditions or characteristics as the functional tasks, the social-psychological concerns and

the perception of pain are not thought to be different between the UK and Denmark. The OES was therefore not subject to any cross-cultural adaptation.

### Limitations

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. Even though the OES PROM is targeted towards the affected elbow, we did investigate whether or not the patients performed the functional tasks with the unaffected side instead. The study group consisted of patients with TEA only, and the D-OES is therefore valid for this category of patients only. The value of the D-OES in evaluating patients with other elbow disorders needs to be studied in future studies.

### CONCLUSION

The Danish 12-item OES is now published as a valid and reliable multidimensional elbow-specific questionnaire that can be used as a quality of life measure in TEA patients.

**CORRESPONDENCE:** Hans Christian Plaschke, Skulder- og Albuesektionen, Ortopædkirurgisk Afdeling, Herlev Hospital, 2730 Herlev, Denmark. E-mail: hcplaschke@dadinet.dk

**ACCEPTED:** 14 August 2013

**CONFLICTS OF INTEREST:** none. Disclosure forms provided by the authors are available with the full text of this article at [www.danmedj.dk](http://www.danmedj.dk).

**ACKNOWLEDGEMENTS:** Many thanks to Tobias Wrenfeldt Klausen for statistical analysis.

### LITERATURE

- Patrick D, Guyatt GH, Acquadro C. Patient-reported outcomes. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Chichester, UK: John Wiley & Sons, 2008:531-45.
- Walters SJ. *Quality of life outcomes in clinical trials and health-care evaluation: a practical guide to analysis and interpretation*. Chichester, UK: John Wiley & Sons, Ltd, 2009.
- Longo UG, Franceschi F, Loppini M et al. Rating systems for evaluation of the elbow. *Br Med Bull* 2008;87:131-61.
- Dawson J, Doll H, Boller I et al. The development and validation of a patient-reported questionnaire to assess outcomes of elbow surgery. *J Bone Joint Surg Br* 2008;90:466-73.
- Hudak PL, Amadio PC, Bombardier C. The Upper Extremity Collaborative Group (UECG). Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). *Am J Ind Med* 1996;29:602-8. Erratum in: *Am J Ind Med* 1996;30:372.
- Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery, *BMC Musculoskelet Disord* 2003;4:11.
- Herup A, Merser S, Boeckstyns M. Validation of questionnaire for conditions of the upper extremity. *Ugeskr Læger* 2010;172:3333-6.
- Morrey BF, Adams RA. Semiconstrained arthroplasty for the treatment of rheumatoid arthritis of the elbow. *J Bone Joint Surg Am* 1992;74:479-90.
- De Haan J, Goei H, Schep N et al. The reliability, validity and responsiveness of the Dutch version of the Oxford elbow score. *J Orthop Surg Res* 2011;6:39.
- De Haan J, Schep N, Tuinebreijer W et al. Rasch analysis of the Dutch version of the Oxford elbow score. *Patient Relat Outcome Meas* 2011;2: 145-9.
- Dawson J, Doll H, Boller I et al. Comparative responsiveness and minimal change for the Oxford Elbow Score following surgery. *Qual Life Res* 2008;17:1257-67.
- Beaton DE, Bombardier C, Guillemin F et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186-91.
- Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57:1358-62.