

BJR



■ HIP

Translation, cross-cultural adaptation and validation of the Danish version of the Oxford hip score

ASSESSED AGAINST GENERIC AND DISEASE-SPECIFIC QUESTIONNAIRES

**A. Paulsen,
A. Odgaard,
S. Overgaard**

From Odense
University Hospital,
Odense, Denmark

Objectives

The Oxford hip score (OHS) is a 12-item questionnaire designed and developed to assess function and pain from the perspective of patients who are undergoing total hip replacement (THR). The OHS has been shown to be consistent, reliable, valid and sensitive to clinical change following THR. It has been translated into different languages, but no adequately translated, adapted and validated Danish language version exists.

Methods

The OHS was translated and cross-culturally adapted into Danish from the original English version, using methods based on best-practice guidelines. The translation was tested for psychometric quality in patients drawn from a cohort from the Danish Hip Arthroplasty Register (DHR).

Results

The Danish OHS had a response rate of 87.4%, no floor effect and a 19.9% ceiling effect (as expected in post-operative patients). Only 1.2% of patients had too many items missing to calculate a sum score. Construct validity was adequate and 80% of our predefined hypotheses regarding the correlation between scores on the Danish OHS and the other questionnaires were confirmed. The intraclass correlation (ICC) of the different items ranged from 0.80 to 0.95 and the average limits of agreement (LOA) ranged from -0.05 to 0.06. The Danish OHS had a high internal consistency with a Cronbach's alpha of 0.99 and an average inter-item correlation of 0.88.

Conclusions

This Danish version of the OHS is a valid and reliable patient-reported outcome measurement instrument (PROM) with similar qualities to the original English language version.

Keywords: Patient-reported outcome measurement instrument, PRO, PROMs, Oxford hip score, OHS, Validation, Total hip replacement, THR, Danish Hip Arthroplasty Registry (DHR)

■ A. Paulsen, MD, PhD Student
■ S. Overgaard, MS, DMSc, Head of Research, Professor, Orthopaedic Surgeon
Odense University Hospital, Department of Orthopaedics and Traumatology, Institute of Clinical Research, University of Southern Denmark, Sdr. Boulevard 29, 5000 Odense C, Denmark.

■ A. Odgaard, MD, DMSc, Consultant Orthopaedic Surgeon
Aarhus University Hospital, Department of Orthopaedics, Nordre Ringgade 1, 8000 Aarhus, Denmark.

Correspondence should be sent to Mr A. Paulsen; e-mail: akselpaulsen@gmail.com

10.1302/2046-3758.19.2000076 \$2.00

Bone Joint Res 2012;1:225–33.
Received 16 March 2012; Accepted after revision 30 July 2012

Article focus

- The use of patient-reported outcome measurement instruments (PROMs) in orthopaedics is increasing
- Their development is laborious and costly and therefore translation, cross-cultural adaptation and validation of established outcome measures is sensible and also facilitates international comparisons
- To date, no validated version of the Oxford hip score (OHS) has existed in the Danish language

Key messages

- The Danish language version of the OHS proved to be a valid PROM with similar qualities to the original English language version

Strengths and limitations

- This large validation study, which included 1992 post-operative patients, followed the principles of best practice for the translation and cultural adaptation process for PROMs and validated the Danish language version of the OHS against several frequently used generic (EuroQol 5D-3L (EQ-5D) and Short-Form 12 (SF-12)) and disease-specific (Hip dysfunction and Osteoarthritis Outcome Score (HOOS)) PROMs in the context of a hip arthroplasty registry
- The inclusion of patients from 30 to 80 years of age increases the external validity of the psychometric findings, as did measuring PROMs at one to two, five to six and ten to 11 years following total hip replacement

■ We included solely post-operative patients and further studies on the responsiveness/sensitivity of the Danish language version of the OHS are warranted. Patients received two disease-specific questionnaires: they answered the HOOS at a median post-operative time period of 4.9 years (0.9 to 10.5) and the OHS at a median of 7.1 years (3.1 to 12.8). Presumably both PROMs measured the patient's health status during a period in which their hip function was in the same steady state. We did not exclude patients who had undergone revision surgery or THR of the contralateral hip

Introduction

Total hip replacement (THR) – ‘the operation of the century’¹ – is a successful orthopaedic procedure with respect to survival of the implant.¹⁻⁵ Implant survival and surgeon-reported outcomes have traditionally been used to evaluate success. However, not everyone who has a failing arthroplasty is willing or able to go through with revision surgery. The recent shift towards a more patient-centric perspective has led to a change of focus from traditional clinical outcomes to patient-reported outcomes, which has revealed a much higher proportion of operations with inferior outcomes.^{6,7} By using patient-reported outcomes the results of THR can be monitored to an entirely different degree, potentially leading to improvements in the treatment of these patients.

The Oxford hip score (OHS)⁸ is a frequently used patient-reported outcome measurement instrument (PROM) developed for patients undergoing THR. However, no adequately translated and culturally adapted Danish version of the score exists. We aimed to develop such a version for use in the Danish Hip Arthroplasty Register (DHR).

Materials and Methods

The study was performed in two phases. In 2009, the original OHS was translated into Danish and cross-culturally adapted. Secondly, in 2011, following implementation of the Danish version, data from that version were tested for psychometric quality.

OHS. The OHS is a short, 12-item questionnaire for patients undergoing THR.⁸ It was designed as an intervention- and site-specific outcome measure to assess functional ability, daily activities and pain from the patient's perspective. Items are answered by ticking a box on a five-point Likert scale. Originally, the raw scores were added to obtain a sum score between 12 and 60, with higher scores being better. Due to modifications, the sum score is now described as ranging from 0 (worst) to 48 (best).^{9,10} The OHS has been translated into different languages and used in several clinical studies and registry settings. It has been shown to be consistent, reliable, valid and sensitive to clinical change following THR.¹¹⁻¹⁸ Thresholds in the OHS associated with patient satisfaction with post-surgical outcomes have been estimated.¹⁹ A

license for the study and translation was obtained from Isis Innovation (Oxford, United Kingdom).

Procedure for translation and cross-cultural adaptation. The translation and cross-cultural adaptation process for the OHS was carried out in accordance with a recommended best-practice methodology,²⁰ and involved the following steps:

1. An uninformed forward-translation from English to Danish (by translator T1, Associate Professor in English Language in Denmark (mother tongue Danish, fluent in English)).

2. An uninformed back-translation from Danish to English (by translator T2, Associate Professor in English Language in Denmark (mother tongue English, fluent in Danish)).

3. An expert panel consensus meeting, during which the original and back-translated English versions were compared, and clinical/linguistic issues in the Danish forward-translated version were resolved.

4. Five new individually uninformed back-translations from Danish to English (by five members of a multidisciplinary group that included professional translators and experienced health professionals; two with English as their mother tongue, three with Danish as their mother tongue, and all bilingual).

5. A new expert panel consensus meeting with translators and coordinators, where the versions were reviewed, reconciled and harmonised, and the back-translations compared with the original English version and prior translations. This resulted in consensus on the Danish version of the OHS.

6. The final Danish language version was tested for understanding on 24 patients (ten men and 14 women) with a mean age of 65 years (24 to 86), with hip dysfunction, hip osteoarthritis or THR, by experienced health professionals. After completing the OHS, the respondents were systematically interviewed and debriefed on their thoughts concerning the relevance of the questions, the specific wording of each item, any difficulties in understanding the questions, the readability, and their experience in answering the questionnaire. The interviewing health professionals also assessed the patient's ability to complete the PROM, using the same criteria.

7. We used the PROMs in their standard lay-out, and based on the testing, we made minimal adjustments to optimise readability for elderly patients, and to facilitate automated forms processing. Written instructions for the PROM were added, layout, font, text size and points in correspondences were adjusted after consulting typographers and educationalists, and these final modifications were incorporated after examination of the outcomes from the debriefing. The Danish language version of the OHS was proofread by a key in-country consultant and project manager, and a report prepared of the translation process.

Data collection. We used a cross-sectional design. Our study was a secondary analysis of data from a previous study of a cohort of 5777 patients from the Danish Hip

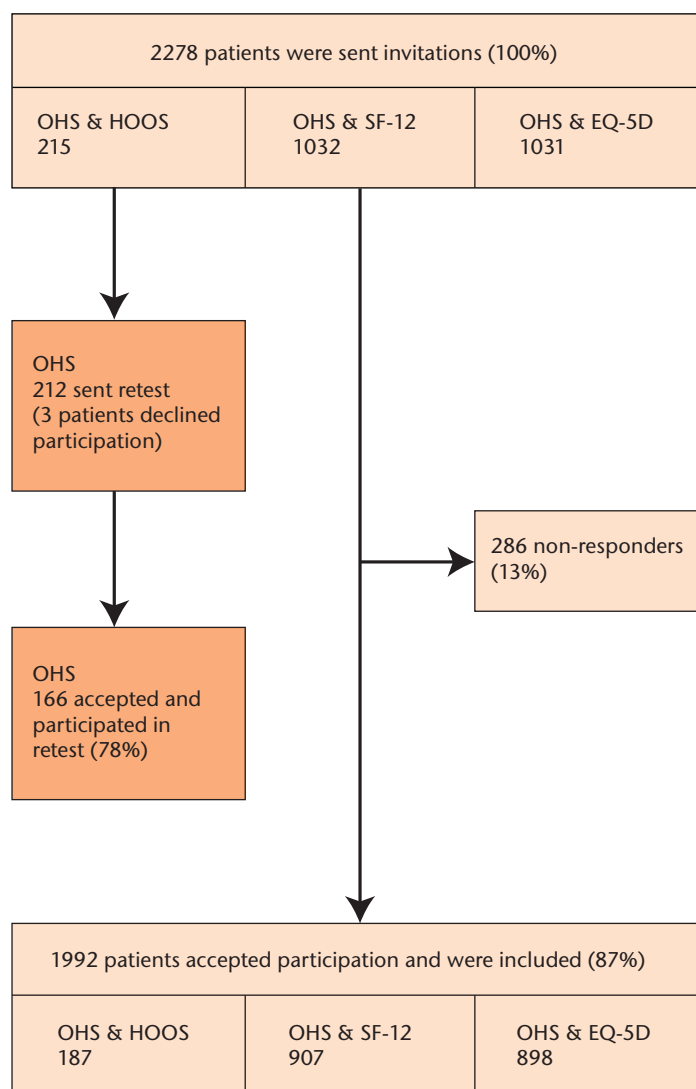


Fig. 1

Flow chart showing participation (OHS, Oxford hip score; HOOS, Hip dysfunction and Osteoarthritis Outcome Score; SF-12, Short-Form 12; EQ-5D, EuroQol 5D-3L).

Registry.²¹ The patients had received a primary THR either one to two, five to six, or ten to 11 years before dispatch of the PROMs. Patients who had revision surgery, or received contralateral THR following the index operation, were not excluded from the study. For the current study, we included the subgroup of all patients between the ages of 30 and 80 years who had answered the OHS. These patients, as part of the original cohort, had received one generic PROM, either the EuroQol 5D-3L (EQ-5D)²² or Short-Form 12 (SF-12),²³ and one disease-specific PROM, either the Hip dysfunction and Osteoarthritis Outcome Score (HOOS)²⁴ or the OHS.⁸ We also included 215 patients who had previously answered the HOOS. These patients were asked to also complete the OHS to enable comparison of disease-specific PROMs. They were randomly selected from the original cohort,

and received the OHS at two years after completing the HOOS. This gave us a total of 2278 patients for the current study. For test-retest validation, 212 patients received the OHS twice within two weeks, at a median of 7.1 years (3.1 to 12.8) after their index operation (Fig. 1). We assumed the patients to be in the same state regarding their hip when answering the questionnaires.

We included between 187 and 907 patients for each combination of PROMs to calculate construct validity and internal consistency, and 166 patients completed the test-retest. These numbers are all higher than the recommended minimum proposed by Terwee et al.²⁵ All PROMs were posted to the patients with a return-addressed and stamped envelope. Paper questionnaire formats were used, and up to two reminder letters were sent.²⁶ All returned PROMs were scanned electronically,

using a validated automated forms processing technique.²⁷ The study was conducted in accordance with the STROBE statement.²⁸

Other PROMs. As a part of the validation, the Danish OHS was compared with two generic outcome measures (EQ-5D and SF-12) and a disease-specific outcome measure (HOOS).

The EQ-5D is a standardised generic measure of health outcome.^{22,29} The EQ-5D gives a summary index (EQ-5D Index) and a VAS score (EQ-VAS). Its psychometric properties have been validated for THR³⁰ and for patients with rheumatoid arthritis.³¹ A license for the study was obtained from The EuroQol Group.³²

The SF-12 is a short form of the SF-36 with 12 items,^{23,33} a generic measure of health status. The SF-12 gives a physical component summary score (PCS) and a mental component summary score (MCS). Its psychometric properties have been validated for osteoarthritis patients.³⁴ A license for the study was obtained from The Medical Outcomes Trust Health Assessment Lab and Quality Metric Incorporated.³⁵

The HOOS includes five subscales: pain, other symptoms, function in daily living, function in sport and recreation, and hip-related quality of life.²⁴ The HOOS-Physical Function Short form (HOOS PS) is a five-item short form of the two HOOS subscales: function in daily living and sport and recreation function. The HOOS PS has been validated for THR.³⁶ We used the HOOS subscales Pain (HOOS Pain), HOOS PS and Hip-related Quality of Life (HOOS QoL).³⁷

Psychometric properties. The Danish OHS was examined for response rate, floor and ceiling effects, skew of the distribution, missing items, construct validity, reliability (intraclass correlation, limits of agreement and inter item correlation reliability), and internal consistency. We defined response rate as the percentage of patients who agreed to participate and answer the questionnaire, missing items as the percentage of all questionnaires with too many items missing to calculate a sum score, as recommended,¹⁰ floor and ceiling effects as the percentage of patients at the extreme ends of the PRO (no possibility to measure a meaningful deterioration of, or improvement in, their condition), calculated as the percentages of patients with the lowest (0) or highest (48) possible sum score out of the total number of patients. Construct validity was tested by comparing the Spearman's correlation coefficients of the OHS scores with the domains of the HOOS, EQ-5D, and SF-12. We hypothesised that the OHS should have moderate to high (0.50 to 0.80) correlations with HOOS Pain, HOOS PS and HOOS QoL; the pain/discomfort domain, mobility, current state of health and the usual activities domain from the EQ-5D; and the general health, physical component score and body pain domains from the SF-12, since these domains are similar to those of the OHS. We also hypothesised that the OHS should show lower (< 0.50) correlations with the anxiety/depression and self-care domains of the EQ-5D, and the

mental component score, vitality and social functioning domains from the SF-12, since these domains are not directly a part of the OHS. Reliability was measured as the intraclass correlation coefficients (ICC), and the limits of agreement (LOA). The time period between the repeated administrations was 2 weeks. Internal consistency was determined by calculating Cronbach's Alpha³⁸ for the OHS. A value for Cronbach's Alpha > 0.8 was considered "good" while a value ≥ 0.9 was considered "excellent".³⁹ We used COSMIN definitions and taxonomy to describe psychometric properties.⁴⁰

Statistical analysis. Descriptive statistics were used to describe patient characteristics. Response rate, floor and ceiling effects, and missing items were calculated as proportions with 95% confidence intervals (CI). We used a chi-squared test to compare proportions. A p-value < 0.05 was considered significant. For test-retest, we used the STATA 'sample' command to draw random samples of the original cohort from the Danish Hip Registry. Construct validity was tested by comparing the Spearman's correlation coefficients. Internal consistency was determined by calculating Cronbach's Alpha. Intraclass correlation (ICC) was calculated as ICC[2,1] with STATA 'icc23' command (two-way random effects model). Bland and Altman's limits of agreement were calculated by STATA 'concord' command. The STATA software Version 11.0 (StataCorp LP, College Station, Texas) was used for all statistical analyses.

Ethics. The study was approved by the Danish Data Protection Agency (number 2008-41-2593), the Danish National Board of Health, and DHR. The study was presented for the Science Ethics Committee of the Region of Southern Denmark. They declared that the study did not require acceptance from the committee due to no intervention or human material were included. All patients gave informed written consent and the study was carried out in accordance with the Declaration of Helsinki.

Results

Patients. Patient characteristics are listed in Table I, and their mean scores for PROMs are listed in Table II. Non-responders were predominantly younger patients and had the diagnoses 'low impact fractures' and 'other arthritis' more often than responders. The mean OHS score was 40. Post-operative follow-up was a median of 4.9 years (0.9 to 10.5).

Translation and cross-cultural adaptation. The translation process revealed minor discrepancies in wording and comprehension for items 1 (Usual level of hip pain), 8 (Pain on standing up from sitting), 9 (Limping when walking), 11 (Work interference due to pain), 12 (Pain in bed at night) and option 4 in item 6 (Walking time before severe pain), so these were rephrased in the translation process. Some patients had problems with item 3 (Trouble with transport), which was resolved by adding a written instruction for the questionnaire.

Table I. Patient characteristics of responders and non-responders

	Responders	Non-responders	p-value (chi-squared test)
Number (%)	1992 (87.4)	286 (12.6)	
Female (n, %)	1088 (54.6)	165 (57.7)	0.329
Median age (yrs) (range)*	68.8 (31 to 80)	66.9 (32 to 80)	0.004 (Student's t-test)
Age group (n, %)			
30 to 49 years	138 (6.9)	35 (12.2)	0.002
50 to 70 years	955 (47.9)	133 (46.5)	0.649
71 to 80 years	899 (45.1)	118 (41.3)	0.218
Diagnosis (n, %) ^{†‡}			
Idiopathic osteoarthritis	1598 (80.6)	186 (65.3)	< 0.001
Low-impact fractures	116 (5.9)	38 (13.3)	< 0.001
Childhood diseases	113 (5.7)	16 (5.6)	0.953
Other arthritis	53 (2.7)	18 (6.3)	0.001
High-impact injuries	20 (1.0)	3 (1.1)	0.945
Atraumatic necrosis of femoral head	62 (3.1)	17 (6.0)	0.015
Other	20 (1.0)	7 (2.5)	0.035
Prostheses design (n, %) [‡]			
Uncemented	1091 (55.0)	154 (54.0)	0.749
Cemented	433 (21.8)	63 (22.1)	0.921
Hybrid	458 (23.1)	68 (23.9)	0.779

* age of patients on date of dispatch of the patient-reported outcome measures

† other arthritis (including rheumatoid arthritis, Bechterew's disease), childhood diseases (congenital hip dislocation, Calvé-Legg-Perthes, epiphysiolysis, acetabular dysplasia), high-impact injuries (fracture of acetabulum, traumatic hip dislocation) and low-impact fractures (fresh fracture of proximal femur, late sequel from fracture of proximal femur)

‡ data on diagnosis and prosthesis design was only available for 1982 responders and 285 non-responders

Table II. Scores of the patient-reported outcome measures (PROMs) for the total population (CI, confidence interval)

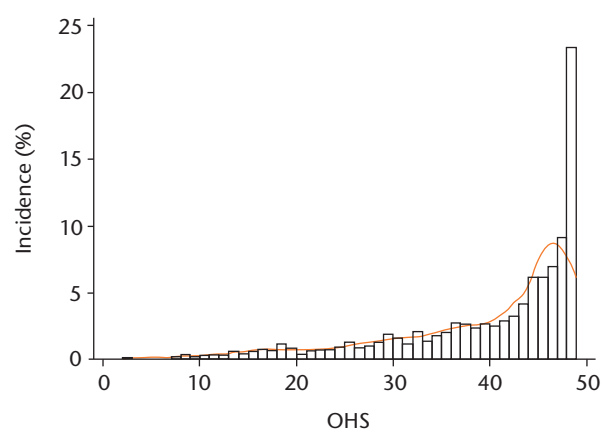
PROM	Mean (95% CI)
OHS (n = 1992)*	39.8 (39.3 to 40.2)
HOOS (n = 187) [†]	
Pain	91.4 (89.3 to 93.5)
Physical function	86.7 (84.2 to 89.3)
Hip-related quality of life	82.1 (79.2 to 85.0)
SF-12 (n = 907) [‡]	
Physical component	38.5 (38.2 to 38.8)
Mental component	46.8 (46.5 to 47.2)
EuroQol 5D-3L (n = 898) [§]	
EQ-5D Index	0.85 (0.84 to 0.86)
EQ-VAS	79.7 (78.3 to 81.1)

* OHS, Oxford hip score (from 0 (worst) to 48 (best))

† HOOS, Hip dysfunction and Osteoarthritis Outcome Score (from 0 (worst) to 100 (best))

‡ SF-12, Short-Form 12 (from 0 (worst) to 100 (best), by computation with a standardised scoring algorithm developed to get a mean of 50 (sd 10) in the United States 1998 general population value set

§ EuroQol 5D-3L. The EQ-5D Index ranges from -0.624 (worst) to 1.000 (best), using a Danish value set.⁴⁵ EQ-VAS, visual analogue scale for current state of health (from 0 (worst) to 100 (best))

**Fig. 2**

Bar chart showing the skew of distribution of the Oxford hip score (OHS).

Psychometric properties. The OHS had a response rate of 87.4%, no floor effect and 19.9% ceiling effect in our post-operative patients. The frequency distribution of the scores was negatively skewed (Fig. 2), with a skew value

of -1.39. Only 1% of patients had too many items missing to calculate a sum score (Table III). Regarding construct validity, OHS showed the highest correlations with the HOOS Pain, HOOS PS and HOOS QoL; the pain/discomfort domain, mobility, current state of health and the usual activities domain from the EQ-5D; and the body pain domain from the SF-12 ($\rho = \pm 0.51$ to 0.62) (Table IV, V and VI). The OHS showed the lowest correlations with the anxiety/depression and self-care domains of the EQ-5D; and the mental component score, vitality and social functioning domains from SF-12 ($\rho = \pm 0.32$

Table III. Response rate, floor effect, ceiling effect and missing items of the Oxford hip score (OHS) compared with the Hip dysfunction and Osteoarthritis Outcome Score (HOOS), Short-Form 12 (SF-12) and EuroQol 5D-3L (EQ-5D)

	OHS (n = 1992)	HOOS (n = 187)			SF-12 (n = 907)		EQ-5D (n = 898)	
		Pain	Physical function	Quality of life	Physical component	Mental component	Index	VAS
Response rate*	87.4 (86.1 to 88.8)	87.0 (82.4 to 91.58)			87.9 (85.9 to 89.9)		87.1 (85.1 to 89.1)	
p-value	Reference	0.843			0.782		0.721	
Floor effect†	0.0	0.0	0.0	0.0	0.1 (-0.1 to 0.3)	0.1 (-0.1 to 0.3)	0.0	0.1 (-0.1 to 0.3)
p-value	Reference	-	-	-	0.138	0.138	-	0.136
Ceiling effect‡	19.9 (18.1 to 21.6)	46.0 (38.8 to 53.2)	39.6 (32.5 to 46.6)	30.0 (23.3 to 36.6)	8.3 (6.5 to 10.1)	8.3 (6.5 to 10.1)	49.9 (46.6 to 53.2)	12.0 (9.9 to 14.2)
p-value	Reference	< 0.001	< 0.001	0.001	< 0.001	< 0.001	< 0.001	< 0.001
Missing items§	1.2 (0.7 to 1.6)	2.1 (0.5 to 4.2)	2.1 (0.5 to 4.2)	0.5 (-0.5 to 1.6)	0.8 (0.2 to 1.3)	0.8 (0.2 to 1.3)	1.0 (0.3 to 1.7)	4.2 (2.9 to 5.6)
p-value	Reference	0.245	0.245	0.437	0.345	0.345	0.717	< 0.001

* defined as percentage who accepted participation and answered the patient-reported outcome measures out of total number. No subscale calculations

† floor effect defined as percentage of worst possible outcome out of total number

‡ ceiling effect defined as percentage of best possible outcome out of total number

§ too many items missing to calculate a sum score (if more than 2 items were omitted, the response was considered invalid and no overall score was calculated, in accordance with the User Manual)

Table IV. Construct validity: Spearman's correlation coefficient between the Oxford hip score (OHS) and the Hip dysfunction and Osteoarthritis Outcome Score (HOOS)

n = 177	OHS	HOOS		
		Pain	Physical function	Quality of life
OHS	1.00			
HOOS				
Pain	0.53*	1.00		
Physical function	0.51*	0.75*	1.00	
Quality of life	0.51*	0.72*	0.22*	1.00

* statistically significant (p ≤ 0.05)

Table V. Construct validity: Spearman's correlation coefficient between the Oxford hip score (OHS) and the EuroQol 5D-3L (EQ-5D)

n = 941	OHS	EQ-5D					
		Mobility	Self-care	Usual activities	Pain/discomfort	Anxiety/depression	Current state of health
OHS	1.00						
EQ-5D							
Mobility	-0.56*	1.00					
Self-care	-0.44*	0.47*	1.00				
Usual activities	-0.61	0.62*	0.44*	1.00			
Pain/discomfort	-0.62*	0.62*	0.40*	0.61*	1.00		
Anxiety/depression	-0.32*	0.34*	0.31*	0.36*	0.37*	1.00	
Current state of health	0.61*	-0.54*	-0.39*	-0.59*	-0.59*	-0.37*	1.00

* statistically significant (p ≤ 0.05)

to 0.46). SF-12 general health, body pain domain and physical component score had a correlation of 0.38 to 0.49. Thus 12 of the 15 predefined hypotheses about the strength of correlation were confirmed. The test-retest reliability of the OHS sum score was established with an ICC of 0.96 (95% CI 0.94 to 0.97), and limits of agreement was -0.05 (95% CI -4.67 to 4.58) (Table VII). For

internal consistency, the overall Cronbach's alpha was 0.99, and the average inter-item correlation was 0.88 (Table VIII).

Discussion

PROMs are an important addition to measuring implant survival, and essential for patient perspectives of outcome.^{41,42}

Table VI. Construct validity: Spearman's correlation coefficient between the Oxford hip score (OHS) and the Short-Form 12 (SF-12)

n = 380	OHS	SF-12					
		Physical component	Mental component	Vitality	Body pain	Social functioning	General health
OHS	1.00						
SF-12							
PCS	0.39 [†]	1.00					
MCS	0.46 [†]	0.30 [†]	1.00				
Vitality	0.37	0.57 [†]	0.66 [†]	1.00			
Body pain	0.49 [†]	0.78 [†]	0.53 [†]	0.56 [†]	1.00		
Social functioning	0.38 [†]	0.36 [†]	0.76 [†]	0.46 [†]	0.46 [†]	1.00	
General health	0.38 [†]	0.73 [†]	0.50 [†]	0.60 [†]	0.66 [†]	0.41 [†]	1.00

* statistically significant ($p \leq 0.05$)

Table VII. Intraclass correlation (ICC) and limits of agreement (LOA) of the Oxford hip score (OHS)

Question	Content [†]	ICC (95% CI)	Average LOA (95% CI)
1	Usual level of hip pain	0.87 (0.83 to 0.91)	-0.02 (-1.05 to 1.01)
2	Trouble with washing and drying	0.85 (0.80 to 0.89)	0.01 (-0.71 to 0.72)
3	Trouble with transport	0.81 (0.74 to 0.85)	-0.02 (-0.94 to 0.91)
4	Putting on socks/stockings/tights	0.85 (0.80 to 0.88)	0.06 (-0.86 to 0.99)
5	Doing household shopping alone	0.95 (0.93 to 0.96)	-0.01 (-0.51 to 0.48)
6	Walking time before severe pain	0.80 (0.73 to 0.85)	-0.01 (-1.10 to 1.09)
7	Difficulty going up stairs	0.84 (0.79 to 0.88)	-0.05 (-0.96 to 0.86)
8	Pain on standing up from sitting	0.83 (0.77 to 0.87)	0.00 (-0.89 to 0.89)
9	Limping when walking	0.81 (0.75 to 0.85)	0.06 (-0.94 to 1.07)
10	Sudden, severe pain from hip	0.86 (0.82 to 0.90)	0.04 (-0.79 to 0.88)
11	Work interference due to pain	0.85 (0.80 to 0.88)	-0.02 (-0.95 to 0.91)
12	Pain in bed at night	0.86 (0.81 to 0.89)	-0.01 (-0.91 to 0.89)
OHS sum score		0.96 (0.94 to 0.97)	-0.05 (-4.67 to 4.58)

* the wording of each item reported in this table is in abridged form

Table VIII. Internal consistency of the Oxford hip score. The mean inter-item correlation is 0.88 and the overall Cronbach's alpha is 0.99

Question	Content [†]	Obs.	Mean (SE) score	Item total correlation	Alpha if item removed
1	Usual level of hip pain	1980	3.55 (0.04)	0.94	0.93
2	Trouble with washing and drying	1988	3.90 (0.03)	0.95	0.94
3	Trouble with transport	1989	3.77 (0.04)	0.95	0.94
4	Putting on socks/stockings/tights	1988	3.63 (0.04)	0.93	0.91
5	Doing household shopping alone	1978	3.93 (0.04)	0.94	0.93
6	Walking time before severe pain	1986	3.84 (0.04)	0.94	0.93
7	Difficulty going up stairs	1963	3.71 (0.04)	0.95	0.94
8	Pain on standing up from sitting	1973	3.78 (0.04)	0.96	0.95
9	Limping when walking	1969	3.61 (0.04)	0.91	0.89
10	Sudden, severe pain from hip	1969	3.76 (0.04)	0.94	0.92
11	Work interference due to pain	1968	3.64 (0.04)	0.97	0.96
12	Pain in bed at night	1975	3.74 (0.04)	0.93	0.91

* the wording of each item reported in this table is in abridged form

We translated and cross-culturally adapted the OHS into Danish, and the subsequent validation showed similar psychometric properties to the original OHS.

This translation of the OHS into Danish used a robust methodology that maximised linguistic accuracy and cross-cultural adaptation. There were only minor discrepancies concerning wording and understanding in the

translation process, probably due to the relatively small cultural differences between the United Kingdom and Denmark. In item 6 instead of the original option 4, "around the house only", we chose to focus on walking distance ("only very short distances") because of differences in the size and the number of floors in homes in Denmark compared with the United Kingdom.

Item 3 is complex and comprises three different questions: “Have you had any trouble getting in a car because of your hip?”, “Have you had any trouble getting out of a car because of your hip?” and “Have you had any trouble using public transport because of your hip?”. The testing showed that some patients were unsure of how to answer, if they answered yes to only one or two of these questions. To resolve this problem, we added Danish written instructions to the OHS, as an addendum.

The OHS had an excellent response rate of 87%. We consider a response rate of 80% as being sufficiently representative of the sample studied. We found no floor effect but a ceiling effect that was beyond the recommended 15% ceiling. Others have found a similar ceiling effect for the OHS.^{25,43,44} Since the current results could be explained by the median post-operative follow-up period of five years in our study and the good overall clinical outcome from THR,⁴³ it could be argued that the finding is merely a degree of skew, which is to be expected given the timing of measures relative to the intervention, and this can explain the skew in sum score distribution. Consistent with this, Naal et al¹⁸ found a lower ceiling effect with the pre-operative OHS. In contrast, SF-12 PCS and SF-12 MCS had lower ceiling effects, as reported by others and explained by computation of a norm-based value set.³¹ Considering the good outcome of THR, low floor effects and high ceiling effects can be expected; therefore, the criterion of having the best possible score in less than 15% of patients following THR might be too restrictive.²¹ Concerning missing items, the OHS performed similarly to the other PROMs in our study. We have followed the instructions given in the 2010 User Manual for the OHS¹⁰ for dealing with missing data. However, imputed data can be problematic to use for assessing the measurement properties of an instrument, as imputed data will artificially reduce variation in overall scores.

Convergent and divergent construct validity were adequate with over 75% of the predefined hypotheses confirmed.²⁵ de Groot et al⁴⁴ also found a moderate to high correlation between the OHS and the HOOS Pain (-0.85) and HOOS QoL(-0.62). The correlation of the OHS with the SF-36 has also been found to be moderate to high (± 0.53 to 0.71) for the physical function and bodily pain domains in post-operative patients.^{10,17}

The Danish OHS translation was found to have acceptable test-retest reliability, with an ICC > 0.70.²⁵ The ICC of the different items ranged from 0.80 to 0.95, and the OHS sum score had a LOA of -0.05 (-4.67 to 4.58) and an ICC of 0.96 (0.94 to 0.97), which is better than the original OHS and other language versions.^{10,16,18} This might be explained by the post-operative administration of the OHS in our study.

Internal consistency of the OHS was found to be very high as expected, with a Cronbach's alpha of 0.99. A Cronbach's alpha over 0.95 could be explained by a

possible redundancy in one or more items.²⁵ Cronbach's alphas of 0.87 to 0.89 have been reported in preoperative patients,^{17,18} 0.89 at six months post-operatively,^{8,13} and 0.93 to 0.92 at one to two years post-operatively,¹³ and seems to rise directly in line with the length of follow-up. We therefore believe the very high alpha found is almost certainly due to the long follow-up period, where patients are likely to have few or no symptoms giving a suboptimal timeframe to assess the Cronbach's alpha, and not due to item redundancy – in the usual sense of the term.

We found an excellent response rate. We included a wide age range of patients from 30 to 80 years; most patients undergo THR in this age range. Our study population is slightly younger than the Danish THR population, but we believe that our results have high external validity since the gender ratio and diagnoses are similar between our study population and the Danish THR population. The Danish OHS was validated in the context of a registry of hip replacements, compared with both generic and disease-specific PROMs, and examined at one to two, five to six, and ten to 11 years following THR.

Several methodological limitations have to be taken into consideration when interpreting our results. This is a secondary data analysis and we have solely included post-operative patients. The psychometric properties of PROMs used in elective surgical contexts are usually largely evaluated on pre-operative data, making the interpretation of our ceiling effect, skew and internal consistency more demanding. Since the patients are all post-operative, we expected the OHS to be highly skewed, and it could therefore be argued that referring to ceiling effects could be misleading. We argue that it is important to assess post-operative development, and have chosen to report the percentage of ceiling at PROM level, even though this characteristic would more often be assessed at the individual item level in the development of a PROM. Further studies on the responsiveness/sensitivity to the Danish version of the OHS are warranted. Patients who received two disease-specific PROMs answered the HOOS at a median of 4.9 years (0.9 to 10.5) post-operatively and the OHS at a median of 7.1 years (3.1 to 12.8) post-operatively, when both PROMs presumably measured the patient's health status during a period in which their hip function was in the same steady state. We did not exclude patients who had undergone revision surgery, or received contralateral THR following the index operation

The Danish version of the OHS had good feasibility, an excellent response rate, no floor effect, but a high ceiling effect as was expected with our post-operative patients, and few patients missed too many items to calculate a sum score. The Danish version of the OHS is a valid and reliable tool for outcome studies on THR patients, in comparison with the HOOS, EQ-5D and SF-12, and can be used in a hip registry setting.

Supplementary material



The Danish version of the Oxford hip score with instructions is available alongside this article at www.bjr.boneandjoint.org.uk

References

1. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. *Lancet* 2007;370:1508–1519.
2. No authors listed. National Joint Registry for England and Wales: 6th Annual Report, 2009. <http://www.njrcentre.org.uk/njrcentre/> (date last accessed 13 July 2012).
3. Garellick G, Kärrholm J, Rogmark C, Herberts P. Swedish Hip Arthroplasty Register: Annual Report, 2008. <http://www.shpr.se/en/Publications/DocumentsReports.aspx> (date last accessed 13 July 2012).
4. No authors listed. Danish Hip Arthroplasty Register: Annual Report, 2009 (in Danish). <http://www.dhr.dk/> (date last accessed 13 July 2012).
5. No authors listed. Norwegian Arthroplasty Register: Annual Report, 2010. http://nrlweb.ihelse.net/eng/Report_2010.pdf (date last accessed 13 July 2012).
6. Nikolajsen L, Brandsborg B, Lucht U, Jensen TS, Kehlet H. Chronic pain following total hip arthroplasty: a nationwide questionnaire study. *Acta Anaesthesiol Scand* 2006;50:495–500.
7. Judge A, Cooper C, Williams S, Dreinhofer K, Dieppe P. Patient-reported outcomes one year after primary hip replacement in a European Collaborative Cohort. *Arthritis Care Res (Hoboken)* 2010;62:480–488.
8. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg [Br]* 1996;78-B:185–190.
9. Murray D, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *J Bone Joint Surg [Br]* 2007;89-B:1010–1014.
10. Dawson J, Fitzpatrick R, Churchman D, Verjee-Lorenz A, Clayson D. *User manual for the Oxford hip score (OHS): Version 1.0*. Oxford, Isis Innovation Ltd, 2010.
11. Uesugi Y, Makimoto K, Fujita K, et al. Validity and responsiveness of the Oxford hip score in a prospective study with Japanese total hip arthroplasty patients. *J Orthop Sci* 2009;14:35–39.
12. Garbuz DS, Xu M, Sayre EC. Patients' outcome after total hip arthroplasty: a comparison between the Western Ontario and McMaster Universities index and the Oxford 12-item hip score. *J Arthroplasty* 2006;21:998–1004.
13. Gosens T, Hoefnagels NH, de Vet RC, et al. The "Oxford Heup Score": the translation and validation of a questionnaire into Dutch to evaluate the results of total hip arthroplasty. *Acta Orthop* 2005;76:204–211.
14. Kalairajah Y, Azurza K, Hulme C, Molloy S, Drabu KJ. Health outcome measures in the evaluation of total hip arthroplasties: a comparison between the Harris hip score and the Oxford hip score. *J Arthroplasty* 2005;20:1037–1041.
15. Wylde V, Learmonth ID, Cavendish VJ. The Oxford hip score: the patient's perspective. *Health Qual Life Outcomes* 2005;3:66.
16. Delaunay C, Epinette JA, Dawson J, Murray D, Jolles BM. Cross-cultural adaptations of the Oxford-12 Hip score to the French speaking population. *Orthop Traumatol Surg Res* 2009;95:89–99.
17. Martinelli N, Longo UG, Marinuzzi A, et al. Cross-cultural adaptation and validation with reliability, validity, and responsiveness of the Italian version of the Oxford Hip Score in patients with hip osteoarthritis. *Qual Life Res* 2011;20:923–929.
18. Naal FD, Sieverding M, Impellizzeri FM, et al. Reliability and validity of the cross-culturally adapted German Oxford hip score. *Clin Orthop Relat Res* 2009;467:952–957.
19. Arden NK, Kiran A, Judge A, et al. What is a good patient reported outcome after total hip replacement? *Osteoarthritis Cartilage* 2011;19:155–162.
20. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005;8:94–104.
21. Paulsen A, Pedersen AB, Overgaard S, Roos EM. Feasibility of four patient-reported outcome measures in a registry setting: a cross-sectional study of 6000 patients from the Danish Hip Arthroplasty Registry. *Acta Orthop* 2012;83:321–327.
22. No authors listed. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
23. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–233.
24. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS): validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord* 2003;4:10.
25. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
26. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev* 2009;3:MR000008.
27. Paulsen A, Overgaard S, Lauritsen JM. Quality of data entry using single entry, double entry and automated forms processing: an example based on a study of patient-reported outcomes. *PLoS One* 2012;7:35087.
28. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;4:296.
29. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
30. Dawson J, Fitzpatrick R, Frost S, et al. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Joint Surg [Br]* 2001;83-B:1125–1129.
31. Linde L. *Health-related quality of life in patients with rheumatoid arthritis: a comparative validation of selected measurement instruments* [PhD Thesis]. Copenhagen: University of Copenhagen, 2009.
32. No authors listed. The EuroQol Group: EQ-5D. <http://www.euroqol.org/> (date last accessed 17 August 2012).
33. Ware JE, Kosinski M, Bayliss MS, et al. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care* 1995;33:AS264–AS279.
34. Gandhi SK, Salmon JW, Zhao SZ, et al. Psychometric evaluation of the 12-item short-form health survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. *Clin Ther* 2001;23:1080–1098.
35. No authors listed. Short-Form-36: SF-36. <http://www.sf-36.org/> (date last accessed 17 August 2012).
36. Davis AM, Perruccio AV, Canizares M, et al. Comparative, validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. *Osteoarthritis Cartilage* 2009;17:843–847.
37. No authors listed. Koos. <http://www.koos.nu/index.html> (date last accessed 17 August 2012).
38. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull* 1955;52:281–302.
39. Feinstein AR. *Clinimetrics*. New Haven: Yale University Press, 1987.
40. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–745.
41. Söderman P, Malchau H, Herberts P. Outcome after total hip arthroplasty: Part I. General health evaluation in relation to definition of failure in the Swedish National Total Hip Arthroplasty register. *Acta Orthop Scand* 2000;71:354–359.
42. Söderman P, Malchau H, Herberts P, et al. Outcome after total hip arthroplasty: Part II. Disease-specific follow-up and the Swedish National Total Hip Arthroplasty Register. *Acta Orthop Scand* 2001;72:113–119.
43. Ostendorf M, van Stel HF, Buskens E, et al. Patient-reported outcome in total hip replacement: a comparison of five instruments of health status. *J Bone Joint Surg [Br]* 2004;86-B:801–808.
44. de Groot IB, Reijman M, Terwee CB, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthritis Cartilage* 2007;15:104–109.
45. Witttrup-Jensen KU, Lauridsen J, Gudex C, Pedersen KM. Generation of a Danish TTO value set for EQ-5D health states. *Scand J Public Health* 2009;37:459–466.

Funding statement:

- This study was supported by the Region of Southern Denmark [095382], The Danish Rheumatism Association [R63-A830-B373, R71-A1043, R69-A1427, R80-A1282-B373], and Odense University Hospital [2-41-4-00065-2010]. The funding bodies did not play a role in the investigation.

Author contributions:

- A. Paulsen: Study design, Data collection, Data analysis, Statistical analysis, Writing the paper, Manuscript preparation, Project manager for the translation and cross-cultural adaptation process
- S. Overgaard: Study design, Manuscript preparation
- A. Odgaard: Study design, Manuscript preparation, Key in-country consultant for the translation and cross-cultural adaptation process

ICMJE Conflict of Interest:

- None declared

©2012 British Editorial Society of Bone and Joint Surgery. This is an open-access article distributed under the terms of the Creative Commons Attribution licence, which permits unrestricted use, distribution, and reproduction in any medium, but not for commercial gain, provided the original author and source are credited.