Acta Orthop. Belg., 2017, **83**, 98-109

ORIGINAL STUDY

Comparison between the Harris- and Oxford Hip Score to evaluate outcomes one-year after total hip arthroplasty

Hanneke Weel, Robert Lindeboom, Sander E. Kuipers, Ton M. J. S. Vervest

From the Tergooiziekenhuizen, Department of Orthopaedics, Hilversum

Harris Hip Score (HHS) is a surgeon administered measurement for assessing hip function before and after total hip arthroplasties (THA). Patient reported outcome measurements (PROMs) such as the Oxford Hip Score (OHS) are increasingly used. HHS was compaired to the OHS assessing whether the HHS can be replaced by the OHS for clinical evaluation of THAs.

All 155 patients (167 THAs) were asked to complete an OHS before and one-year after surgery. The surgeon independently scored the HHS at the same time points. We examined and compared the clinimetric properties of both instruments.

Internal consistency reliability of the OHS was notably higher than that of the HHS at all occasions. HHS had a higher effect size (4.1) than the OHS (2.1). Ceiling effect at follow up was 55.6% (HHS) and 36.4% (OHS). Spearman's rank correlation between HHS and OHS was 0.57 at baseline and 0.65 and after one year. The correlation between the change scores was 0.50.

The Oxford Hip Score is of good use in quality assessment after THA.

Keywords : total hip arthroplasty ; patient reported outcome ; harris ; oxford ; hip score.

INTRODUCTION

The increased aging of the population has led to a rise in chronic degenerative diseases including osteoarthritis ; up to 40% of the persons aged over 65 years have an effected joint *(6,1)*. The hip is often affected with about 500.000 total hip arthroplasties (THA) performed each year in the USA and the UK *(1,30)*. THA is a successful treatment to reduce pain and to improve mobility. An implant survivorship greater than 80% was observed at 25 years follow-up *(30)*.

To evaluate patients in clinical practice or clinical research, several outcome instruments are designed over the years. Among the existing hip scoring systems, the Harris Hip Score (HHS) *(13)* is the instrument of choice to measure the results after of THA. The HHS is a surgeon filled joint-specific rating scale designed by Harris in 1969 that is validated several times from then in THA populations *(34,33,15)*, although it was originally

■ Hanneke Weel[1,2] M.D.
■ Robert Lindeboom[3] Ph.D.
■ Sander E. Kuipers[1] M.D.
■ Ton M.J.S. Vervest[2] M.D., Ph.D.

[1]*Tergooiziekenhuizen, Department of Orthopaedics, Hilversum.*

[2]*Academic Medical Center, Department of Orthopaedics, Amsterdam.*

[3]*University of Amsterdam, Department of Clinical Epidemiology, Biostatistcs and Bioinformatics, Amsterdam.*

Correspondence : Hanneke Weel, ZiekenhuisTergooi, Department of Orthopaedics ziekenhuis Tergooi, tav Secretariaat orthopedie, 10016 1201 DA Hilversum, The Netherlands. E-mail : hannekeweel@hotmail.com.

No benefits or funds were received in support of this study. The authors report no conflict of interests.

designed for young patients with secondary osteoarthritis after severe acetabulum fractures. An important limitation of the HHS is a large 'ceiling effect' associated with the overall score. A recent review observed a pooled ceiling effect of 20% *(38)* indicating a poor discriminative ability of the test to detect clinically relevant changes. Another shortcoming of the traditional 'objective' outcome measures like the HHS is the absence of the effect of the intervention from the patient's perspective *(31,16)*.

Today, patient reported outcome measurements - PROMs *(12,17)* are valuable in both clinical and orthopaedic research settings. Studies are performed to investigate whether a PROM can be used as a profitable complement or replacement for the former rating scales that rely on the judgment of the surgeon. Several PROMs already have been validated in THA *(9,25)*. One of these is the Oxford Hip Score (OHS) *(24,10,14)*, to measure outcomes after total hip and knee surgery *(5)*. The OHS has some items in common with the HHS in terms of measuring pain and daily functioning, but does not evaluate range of motion and hip deformity.

To our knowledge, there are two studies comparing the HHS and OHS questionnaires after surgery showing a reasonable correlation. These studies did not include a baseline OHS measurement *(18,27)*. Previous research found that 12 months results also depend on baseline outcomes of the OHS *(17,2)*. Furthermore Parsons et al. *(27)* compared the outcome measurements after hip resurfacing arthroplasty, which represents a younger, different population than THA. In addition to this and to substantiate the current trend towards the use of PROMs instead of surgeon-filled questionnaires *(32)*, we compared the patient reported OHS with the clinician reported HHS in a THA population on some important clinimetric features. Therefore, we collected pre- and one year postoperative scores of both instruments. The primary objective was to examine the reliability and validity of the OHS and HHS in terms of their floor-, ceiling effects, internal consistency reliability and the sensitivity to change, as these properties are essential for their usefulness as an outcome instrument in clinical practice and research. In addition, we examined the strength of the relation between (change) scores of the OHS and HHS and assessed the concordance between the five common items ("Pain", "Distance", "Socks & Stockings", "Stairs", and "Limp") present in the HHS and the OHS.

PATIENTS AND METHODS

This prospective cohort study was performed in the orthopaedic department of (blinded), a teaching hospital in (blinded).

Patients

Between January 2006 and January 2011 in total 167 primary THAs were performed in 155 patients. Inclusion criteria were : osteoarthritis on the affected site. Exclusion criteria were : other hip-related problems on the same or contra-lateral leg. Having a prosthesis on the contra-lateral hip was not an exclusion criterion. All patients were implanted a Zweymueller prosthesis by the same surgeon. At baseline, before surgery, patients completed the Oxford Hip Score and the surgeon filled out the Harris Hip Score to measure their hip function and experienced hip problems, in addition to other clinical and socio-demographic data as a routine outcome measurement. One year postoperatively again patients and surgeon filled out the instruments. Patient characteristics included gender, age at surgery, BMI and surgery site (Table I).

Instruments

The OHS is a 12-item, joint-specific, self-administered questionnaire that was introduced in 1996. The OHS provides a single summed score, which reflects the severity of the patient's hip problems and is also sensitive to change *(5)*. The maximum score is 60 points, representing a poor hip function ; the minimum score is 12 points, representing a good hip function. (Appendix 1 OHS)

The HHS is a surgeon administered joint-specific rating. Raw item scores are weighted to achieve a 0-100 point scale. The maximum score possible is

100 points, representing an excellent hip function ; the minimum score is 0 points. The questions on pain (44 points) and function (47 points), receive the heaviest weight. Function items evaluate activities of daily living (14 points) and gait/mobility (33 points). Range of motion (ROM) and deformity of the hip are also evaluated and received only five and four points respectively. (Appendix 2 HHS)

Clinimetric evaluation

Clinimetric properties of the HHS en OHS as discriminative measures of functioning were compared by examining floor and ceiling effects, internal consistency reliability at baseline and follow-up one year after the operation and their sensitivity to capture within patient change after THA ("responsiveness").

Floor and ceiling effects were defined as the percentage of the sample with the lowest or highest score possible on an instrument. When more than 15% of the participants achieved the lowest or highest score possible, a floor respectively ceiling effect was considered present *(23)*.

Internal consistency reliability

The validity of change scores is dependent on the reliability of the score at both time points. We calculated the internal consistency reliability (Cronbach alpha coefficient) of the OHS and HHS and examined the relative contribution of their items to the reliability of the total score by inspection of the item-rest correlation and "alpha if item deleted". In addition, we also calculated the internal consistency reliability of the unweighted HHS. According to Terwee et al *(36)*, a good internal consistency exists when Cronbach's alpha is between 0.7 and 0.95. A Cronbach's alpha exceeding 0.95 indicates high correlation between the items in the scale suggesting redundancy of one or more items. Item rest correlations indicate the relation between an item score and the total scores of the remainder of the items. It is used to judge the appropriateness of the item to measure the construct of interest. Items with rest correlations ≤ 0.20 *(8)* in general do not contribute to the discriminative capacity of the scale and should be deleted. This is reflected in the "alpha if item deleted" coefficient that should increase when such an item is omitted from the scale.

Validity

Validity of the measurements was evaluated by comparing their sensitivity to change ("responsiveness") in hip functioning after THA. Responsiveness has been defined as the ability of a questionnaire to detect clinically important changes over time *(11)*. The sensitivity to change of both instruments was measured by calculating effect sizes. We used the 'standardized response mean' effect size defined as the mean change found in a variable divided by the standard deviation of that change *(19)*. An effect size > 0.8 is considered large, 0.4-0.8 moderate and below 0.2 is considered small *(7)*.

Relation between (change) scores on HHS and OHS and concordance between common items

The strength of the relationship between the OHS and the HHS at baseline and one year postoperatively was evaluated by calculating a Spearman's rank correlation coefficient. We also examined the strength of the relationship between the change scores of the instruments. A correlation coefficient of ≥ 0.7 is considered a strong relationship, between 0.5 and 0.7 moderate and below 0.5 a weak relationship.

A Somer's D correlation examines the concordance between patient and surgeons ratings of functioning as measured with items present in both the weighted- and unweighted HHS and the OHS. Somer's D compares for all possible pairs of subjects scored with the items, whether there is concordance between the rank order of an OHS item score (e.g. "Stairs") to that of the same item present in HHS. A value of 1 indicates that there are only concordant pairs, i.e., the higher ranked subject on an OHS item is also the higher ranked subject on the same HHS item. A value of 0 indicates no relation between the variables. A given Somer's D value of 0.50 indicates that 50% of the pairs are concordant pairs (or ties, equally ranked) and 50% are discordant.

All statistics were performed in SPSS 19.0. For all analyses, the OHS was rescored in the same direction as the HHS with higher scores representing a better hip function. For the HHS results are presented for the weighted scoring and the raw, unweighted scoring.

RESULTS

Of the 167 THAs (placed in 155 patients) 116 had all data complete at the final follow-up one-year after surgery. There were 135 completed HHSs one year after surgery and 132 OHSs. Six patients were lost to follow up for unknown reason, two were seen by another surgeon unaware of the study, two patients representing three THA died (one malignancy and one unknown cause), one was diagnosed Parkinson disease and one suffered an CVA one day postoperatively. The flowchart is presented in Figure 1. The response rate for the HHS was therefore 87.7% and 88.0% for the OHS. Characteristics of our study population at baseline and one year after surgery are presented in Table I.

Clinimetric properties of the HHS and OHS

Percentage floor and ceiling, internal consistency statistics and sensitivity to treatment effects are summarized in table 2. Floor effects at baseline (T0) were 0% for both HHS and OHS. At 1-year after surgery (T1) the HHS and OHS both showed ceiling effects. The ceiling effect was 55.6% for the HHS and 36.4% for the OHS. There were no patients with the lowest possible score on the OHS or HHS ; floor effect was 0%.

The baseline weighted HHS had a Cronbach's alpha of 0.53 and unweighted 0.69 ; the baseline OHS had a Cronbach's alpha of 0.88. The median item rest (min-max) correlation at baseline for the weighted HHS was 0.31 (0-0.58) and 0.40 (0-0.60) for the unweighted HHS. The HHS item "public transport" had the lowest item rest correlation (0.00) : removal of "public transport" increased the internal consistency reliability in both the weighted and unweighted HHS. The HHS item "pain" had an item rest correlation of 0.38 weighted and 0.40 when unweighted. In the weighted score removal of "pain" increased the internal consistency reliability

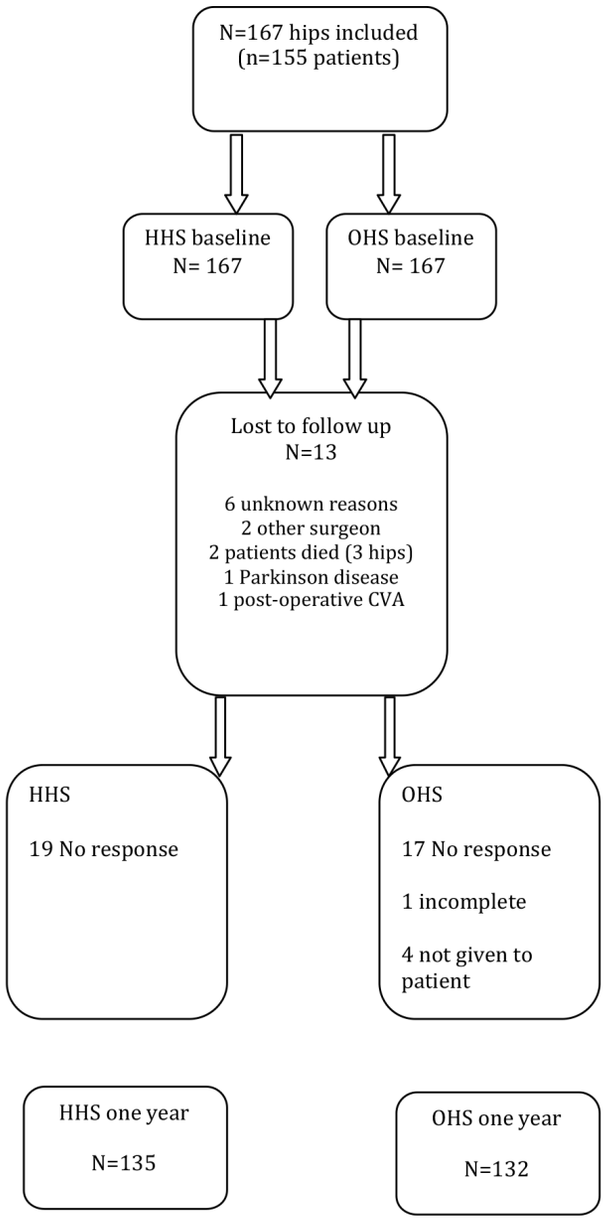

Fig. 1. — Flowchart, selection of patients

to 0.60 (when unweighted, an decrease of the reliability was observed). All other items contributed positively to the reliability of the HHS, removing an item resulted in a lower reliability. The OHS had a median item rest correlation of 0.65 at baseline. The item "limp" had the lowest item rest correlation (0.31), item "work" the highest (0.76) : removal of "limp" increased the internal consistency reliability

from 0.88 to 0.89. All other items contributed positively to the reliability of the OHS.

One year after surgery the weighted HHS had a Cronbach's alpha of 0.60 and unweighted 0.71 ; the OHS had a Cronbach's alpha of 0.96. The median item rest correlation was 0.47 (range 0-0.65) for the weighted and 0.38 (0-0.63) for the unweighted HHS. The item "public transport" again had the lowest item rest correlation and removing this item increased the internal consistency reliability of the unweighted HHS from 0.71 to 0.72 and the weighted HHS from 0.60 to 0.61. The item "pain" had an item rest correlation of 0.38 when weighted and 0.50 unweighted, removing this weighted item the Cronbach's alpha improved to 0.67. Removing this unweighted item decreased the reliability to 0.70. The item "limp" had the highest item rest correlation (0.63 unweighted and 0.65 when weighted) and all other items contributed positively to the reliability of the HHS. The OHS had a median item rest correlation of 0.82 (0.74-0.90). The item "walking" had the lowest item rest correlation, but removal of this item did not increase the reliability of the OHS. The effect sizes (Table II) indicate a good sensitivity to change for both summed scales.

Relation between (change) scores on HHS and OHS and concordance between common items

Table III and IV summarizes the relations between HHS and OHS at baseline and follow-up, the change scores on both instruments and the concordance between common HHS and OHS items. The Spearman's rank correlation coefficient when correlating the HHS to the OHS was 0.57 ($p < 0.001$) at baseline and 0.65 ($p < 0.001$) one year after surgery. The Spearman's rank correlation coefficient for the change scores on HHS and OHS was 0.50 ($p < 0.001$).

The concordance between ratings of items on HHS and OHS that evaluate the same aspect of hip functioning expressed as Somer's D correlations ranged from 0.02 to 0.56. Most items had con-

Table I. — Study characteristics

	Baseline (N**= 167)	Follow up HHS available (N=135)	Follow up OHS available (N=132)
Male, N (%)	68 (40.4)	53 (39.3)	53 (40.2)
Site n* (%) Left	74 (44.3)	64 (47.4)	59 (44.7)
Right	69 (41.3)	57 (42.2)	59 (44.7)
Both	12 (14.4)	7 (10.4)	7 (10.6)
Diagnosis type N (%) Coxarthritis	157 (94.0)	126 (93.3)	123 (93.2)
Dysplasia	1 (0.6)	1 (0.7)	1 (0.8)
Head necrosis	4 (2.4)	3 (2.2)	4 (3)
Post traumatic	3 (1.8)	3 (2.2)	2 (1.5)
Rheumatoid arthritis	1 (0.6)	1 (0.7)	1 (0.8)
M. Paget	1 (0.6)	1 (0.7)	1 (0.8)
Mean (SD) age years	68.8 (8.7)	69.0 (8.2)	68.3 (8.7)
Median length cm (range)	170 (42)	174 (31)	173 (31)
Median weight kg (range)	75.5 (72)	77.8 (73)	78.1 (73)
Median Body Mass Index (range)	25.7 (29.9)	25.9 (30)	26.1 (30)
Mean (SD) HHS	56.1 (11.2)	96.5 (6.7)	96.5 (6.3)
Mean (SD) OHS	36.5 (8.2)	16.4 (7.2)	16.9 (7.6)

* n represents the number of patients
** N represents the number of operatively treated hips

Table II. — Clinimetric properties of HHS and OHS

Clinimetric property	HHS	HHS unweighted	OHS
Floor& Ceiling effect :			
% Floor T0	0	-	0
% Floor T1	0	-	0
% Ceiling T0	0	-	0
% Ceiling T1	55.6	-	36.4[1]
Internal consistency :			
Cronbach a T0	0.53	0.69	0.88
Cronbach a T1	0.60	0.71	0.96
Median (range) item-rest correlation T0	0.31 (0-0.58)	0.40 (0-0.60)	0.65 (0.31-0.76)
Median (range) item-rest correlation T1	0.47 (0-0.65)	0.38 (0-0.63)	0.82 (0.74-0.90)
Sensitivity to treatment effects :			
SRM	4.1	-	2.1

SRM : standardized response mean
Dash (-) indicates not evaluated
1. OHS converted into an ascending scale, like the HHS : a higher score represents a better hip function

cordance values around 0.50. The item 'limp' showed the lowest concordance : 0.03 (p = 0.72) when unweighted and 0.02 (p = 0.93) for the weighted item at baseline. One year after surgery, Somer's D correlation for this item was 0.11 (p = 0.28) and 0.04 (p = 0.64) respectively. The item "distance" at follow-up had the highest Somer's D value (0.56).

DISCUSSION

Our results showed that the HHS had a better responsiveness, but the patient reported OHS showed higher reliability and had less ceiling effects. Both instruments had similar response rates. Response rate and postoperative correlation of the sum scores in this study are comparable to earlier comparisons *(18,27)*. In addition, this study also examined floor and ceiling effects, effect sizes after surgery, reliability, and concordance between individual items that are present in both the HHS and OHS.

Both scores showed a ceiling effect after THA, but the OHS showed this to a lesser extent. Generally the results of THA are very satisfactory *(30,39)* thus questionnaires with low ceiling effects are favorable to compare the results amongst patients with good results after THA. The responsiveness of the HHS was much higher than that of the OHS (4.1 versus 2.1). However, the HHS had a lower reliability and thus its higher responsiveness should be taken into doubt. At baseline and follow-up, the Cronbach's alpha of the HHS was even below the recommended level of 0.7 *(36)*. Reliability of the OHS was high, with a Cronbach's alpha at baseline of 0.88 and 0.96 one year after surgery. This study identified the item "public transport" of the HHS to threat the internal consistency in both the weighted and unweighted scoring with improved internal consistency reliability after removal of this item. The reliability of the HHS with the weighted scoring improved when deleting the item "pain", but not in the unweighted, raw scoring. This item is weighted in the HHS with a maximum of 44 points (no pain), where the minimum is 0 points (pain in rest, unable to walk). This weighting seems disproportional, especially because for the unweighted item "pain" Cronbach's alpha decreased when removing it, while when weighted the reversed was observed. Possibly, the heterogeneous item weighting of the HHS is the cause of the differential contribution of individual items to the scale's reliability in the weighted and unweighted HHS versions. In the original design of the HHS, it was not intended to measure effect of THA in elderly, which is the

predominant population of the patients receiving THA nowadays *(6,1)*. It was originally designed for young patients receiving THA because of osteoarthritis after severe acetabulum fractures *(13)* whereupon the current weighting of the score is based. In daily practice the weighted HHS is the standard, while we noticed that there is a difference in reliability comparing the weighted and unweighted score of the HHS ; the weighting did not contribute to the internal consistency of the measurement according to our results. The item "public transport" showed no variation in this study and therefore did not contribute to the reliability of the HHS. To be included in our cohort the patient had to be visiting the outpatient clinic at baseline, therefore the question "is it possible to travel by public transport?" was 100% answered with 'yes' at that time point. Possibly, at 1 years follow up a patient was not able to travel, but these patients were less likely to be seen in the outpatient clinic, so also at this time point the question was answered with 'yes' in 100% of the cases.

The concordance between patient and surgeons ratings of hip functioning on the common items of the HHS and OHS showed a moderate correlation. An exception was the item 'limp' with a Somer's D approaching zero at both time points. In the HHS the surgeon assesses the degree of limping with none, slightly, moderate to severe. This in contrast to the OHS where the patient is asked "have you been limping when walking, because of your hip'' and possible answers are frequencies : "never, sometimes, often, mostly, always". Therefore, the low concordance on this item between patient and surgeon may be because patients are not aware of their (minimal) limping. Furthermore, the concordance between surgeon and patients ratings on the common items of the HHS and OHS was better in the unweighted version. Again, this difference in concordance between HHS versions

Table III. — Relation between total scores and concordance between scores on common items of HHS and OHS at baseline

	HHS	HHS / HHS unweigted				
		dist	limp	pain	socks	stairs
OHS	0.57*					
dist.		0.44/ 0.46*				
limp			0.02/ 0.03			
pain				0.31/ 0.33*		
socks					0.37/ 0.41*	
stairs						0.36/ 0.39*

* Significant at p< 0.001

Table IV. —Relation between total scores, change scores and concordance between scores on common items of HHS and OHS 1 year postoperatively

	HHS	HHS / HHS unweigted				
	HHS diff	dist	limp	pain	socks	stairs
OHS / OHS diff	0.65* 0.50*					
dist.		0.49/ 0.56*				
limp			0.04/ 0.11			
pain				0.41/ 0.52*		
socks					0.49/ 0.49*	
stairs						0.54/ 0.54*

* Significant at p< 0.001

may be the extremely heterogeneous weights of the HHS items. For example, the item 'pain' has a maximum weight of 44 points, while the items 'shoes and stockings' or 'stairs' both have a maximum of 4 points.

With regard to the usefulness of a questionnaire to measure the effect of an operation, the response rate is also important. In the present time, the rising amount of mandatory quality registrations, questionnaires and other forms that have to be filled could provide extra workload for surgeons. The HHS is totally surgeon filled out and thus causes administrative burden. The OHS on the other hand is patient reported and yielded similar response rates as the HHS. This is in accordance with previous studies, showing good response rates for PROMs in general *(4,25)* and also for the OHS in particular *(29,22)*. Because of its the patient reported nature, the OHS can be administered by post or telephone as well as in the clinical setting, for example in the waiting room. No more assessment by a clinician is needed *(27)*.

A limitation of this study is that all patients were operated and evaluated by the same surgeon, so we could not make an assessment of consistency or reproducibility of the HHS made by different observers measuring the same quantity. This could hamper the generalizability of our results, but also may be a source of bias in the very high sensitivity to change (and higher ceiling effects) we observed for the HHS as compared to the OHS. A surgeon could have the tendency to evaluate his own results less critically and may be inclined to give more favorable scores to patients after surgery *(26)*. Therefore independent outcome assessors would be preferred to analyze the clinimetric properties of the HHS for the assessment of THA. However, our study surgeon is representative for general orthopaedic surgeons according to his experience with THA and with the HHS instrument. This study focused on only one follow-up moment ; one year after surgery. For example, a previous study found that patients with unfavorable scores on a PROM at three months follow up were more likely to not return their questionnaires at one-year follow up *(25)*. This may have inflated the effect size for the sensitivity to change of the OHS. However, we observed a major standardized effect (2.1) : effect sizes exceeding > 0.80 are generally considered as large *(7)*.

Quality assessment using PROMs are recommended *(20,35)* and getting more important and established also in orthopaedic surgery *(32)*. A lot of different PROM's to measure quality after total hip replacement surgery are available *(29,4,2,37)*. Further research, on which PROM is the best to be used in daily care, is needed in order to use an international uniform, reliable, valid and easy to administer outcome measurement in clinical practice.

CONCLUSION

This study compared the Oxford Hip Score to the Harris Hip Score, showing a moderately high correlation between summed scores at baseline and one year postoperatively as well as in the difference of the summed score between the time points. In addition the sensitivity to change of the HHS was better, but the patient filled OHS was more reliable with less ceiling effects, while having the same response rate. The OHS is patient filled and thus also capturing surgeons administrative burden. Therefore according to our results the Oxford Hip Score is of good use in quality assessment after total hip replacement surgery.

Acknowledgements

We would like to thank (blinded), for importing all OHS data into the database.

REFERENCES

1. **Alviar MJ, Olver J, Brand C, Hale T, Khan F.** Do patient-reported outcome measures used in assessing outcomes in rehabilitation on after hip and knee arthroplasty capture issues relevant to patients? Results of a systematic review and ICF linking process. *J Rehabil Med*. 2011 ; 43 : 374-81.
2. **Arden NK, Kiran A, Judge A et al..** What is good patient reported outcome after total hip replacement? *Osteoarthritis Cartilage*. 2011 ; 19 : 155-62.
3. **Bond M, Davis A, Lohmander s, Hawker G**. Responsiveness of the OARSI-OMERACT osteoarthritis pain and function measures. Osteoarthritis Cartilage. 2012 ; 20 : 541-7.

4. **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. *Ostoarthritis Cartilage*. 2009 ; 17 : 843-7.
5. **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 : 185-90.
6. **Dawson J, Linsell L, Zondervan K et al**. Epidemiology of hip and knee pain and its impact on overall health status in older adults. *Rheumatology*. 2004 ; 43 : 497–504.
7. **Fayers PM, Machin D**. Quality of life : The assessment, analysis and interpretation of patient-reported outcomes. *2nd ed. New York : John Wiley-Sons* ; 2007.
8. **Fitzpatrick R, Davey C, Buxton MJ, Jones DR**. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment*. 1998 ; 2 : i-iv,1-74.
9. **Fitzpatrick R, Morris R, Hajat S et al**. The value of short and simple measures to assess outcomes for patients of total hip replacement surgery. *Quality in Health Care*. 2000 ; 9 : 146-50.
10. **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 Orthopaedica*. 2005 ; 76 : 204-11.
11. **Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A.** Responsiveness and validity in health status measurement : a clarification. *J Clin Epidemiol*. 1989 ; 42 : 403-8.
12. **Hamilton DF, Lane JV, Gaston P et al.** What determines patient satisfaction with surgery? A prospective cohort study of 4709 patients following total joint replacement. *BMJ open*. 2013 :9 ; 3 : e002525.
13. **Harris WH**. Traumatic arthritis of the hip after dislocation and acetabular fractures : treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am*. 1969 ; 51 : 737-55.
14. **Haverkamp D, Breugem SJ, Sierevelt IN, Blankevoort L, van Dijk CN**. Translation and validation of the Dutch version of the Oxford 12-item knee questionnaire for knee arthroplasty. *Acta Orthopaedica*. 2005 ; 76 : 347-52.
15. **Hoeksma H, Van Den Ende CH, Ronday HK, Heering A, Breedveld FC.** Comparison of the responsiveness of the Harris Hip Score with generic measures for hip function in osteoarthritis of the hip. *Ann Rheum Dis*. 2003 ; 62 : 935-8.
16. **Jockowiski D, Guyatt G.** A guide to health measurement. *Clin Orthop Relat Res*. 2003 ; 413 : 80-9.
17. **Judge A, Arden NK, Kiran A et al.** Interpretation of patient-reported outcomes for hip and knee replacement surgery : identification of thresholds associated with satisfaction with surgery. *J Bone Joint Surg Br*. 2012 ; 94 : 412-8.
18. **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 : 1037-41.
19. **Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH**. Comparative measurement sensitivity of short and longer health status instruments. *Medical Care*. 1992 ; 30 : 917-25.
20. **Lamberts MP, Drenth JP, Van Laarhoven CJ, Westert GP**. Uitkomst van behandeling volgens patiënten. Instrument om variaties in klinisch handelen terug te dringen [Outcome of treatment reported by patients : instrument to reduce variations in clinical practice]. *Ned Tijdschr Geneeskd*. 2013 ; 157 : A5369. Dutch.
21. **Lohr KN, Aaronson NK, Alonso J et al**.. Evaluating quality-of-life and health status instruments : development of scientific review criteria. *Clinical Therapeutics*. 1996 ; 18 : 979-92.
22. **Martinelli N, Longo UG, Marinozzi A, Franceschetti E, Costa V, Denaro V**. 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 : 923-9
23. **McHorney CA, Tarlov AR**. Individual-patient monitoring in clinical practice : are available health status surveys adequate? *Qual Life Res*. 1995 ; 4 : 293-307.
24. **Murray DW, Fitzpatrick R, Rogers K et al**. The use of the Oxford hip and knee scores. *J Bone Joint Surg Br*. 2007 ; 89 : 1010-4.
25. **Ostendorf M, van Stel HF, Buskens E, Schrijvers et al**. Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Joint Surg Br*. 2004 ; 86 : 801-8.
26. **Paradis C.** Bias in surgical research. *Ann Surg*. 2008 ; 248 : 180-8
27. **Parsons NR, De Souza RM, Oni T et al**. A comparison of Harris and Oxford hip scores for assessing outcome after resurfacing arthroplasty of the hip : can patients tell us everything we need to know? *Hip Int*. 2010 ; 20 : 453-9.
28. **Paulsen A, Odgaard A, Overgaard S**. Translation, cross-cultural adaptation and validation of the Danish version of the Oxford hip score : Assessed against generic and disease-specific questionnaires. *Bone Joint Res*. 2012 ; 1 : 225-33.
29. **Paulsen A, Pedersen AB, Overgaard S, Roos EM**. Feasibility of 4 patient-reported outcome measueres in a registry setting. *Acta Oprthopaedica*. 2012 ; 83 : 321-7.
30. **Pivec R, Johnson AJ, Mers SC, Mont MA**. Hip arthroplasty. *Lancet*. 2012 ; 380 : 1768-77.
31. **Poolman RW, Swiontkowski MF, Fairbank JC, Schemitsch et al**. Outcome instruments : rationale for their use. *J Bone Joint Surg Am*. 2009 ; 91 : 41-9.
32. **Pynsent PB**. Choosing an outcome measure. *J Bone Joint Surg Br*. 2001 ; 83 : 792-4.
33. **Shi HY, Mau LW, Chang JK, Wang JW, Chiu HC.** Responsiveness of the Harris Hip Score and the SF-36 : five years after total hip arthroplasty. *Qual Life Res*. 2009 ; 18 : 1053-60.

34. Söderman P,Malchau H. Is the harris hip score system useful to study the outcome of total hip replacement? *Clin Orthop Rel Research*. 2001 ; 384 : 189-97.
35. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol*. 2006 ; 59 : 160-7.
36. 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.
37. Thorborg K, Tijssen M, Habets B et al. Patient-Reported Outcome (PRO) questionnaires for young to middle-aged adults with hip and groin disability : a systematic review of the clinimetric evidence. *Br J Sports Med*. 2015 ; 49 : 812.
38. Wamper KE, Sierevelt IN, Poolman RW, Bhandari M, Haverkamp D. The Harris hip score : Do ceiling effects limit its usefulness in orthopedics? *Acta Orthopaedica*. 2010 ; 81 : 703-7.
39. Whitehouse MR, Aquilina AL, Patel S, Eastaugh-Waring SJ, Blom AW. Survivorship, patient reported outcome and satisfaction following resurfacing and total hip arthroplasty. **J Arthroplasty.** 2013 ; 28 : 842-8.

APPENDIX

Appendix 1. — Oxford Hip Score (OHS)

1) How would you describe the pain you usually have in your hip?

1. None
2. Very mild
3. Mild
4. Moderate
5. Severe

2) Have you been troubled by pain from your hip in bed at night?

1. No nights
2. Only 1 or 2 nights
3. Some nights
4. Most nights
5. Every night

3) Have you had any sudden, severe pain-' shooting ', 'stabbing', or 'spasms' from your affected hip?

1. No days
2. Only 1 or 2 days
3. Some days
4. Most days
5. Every day

4) Have you been limping when walking because of your hip?

1. Rarely/never
2. Sometimes or just at first
3. Often, not just at first
4. Most of the time
5. All of the time

5) For how long have you been able to walk before the pain in your hip becomes severe (with or without a walking aid)?

1. No pain for 30 minutes or more.
2. 16 to 30 minutes
3. 5 to 15 minutes
4. Around the house only
5. Not at all

6. Have you been able to climb a flight of stairs?

1. Yes, easily
2. With little difficulty
3. With moderate difficulty
4. With extreme difficulty
5. No, impossible

7) Have you been able to put on a pair of socks, stockings or tights?

1. Yes, easily
2. With little difficulty
3. With moderate difficulty
4. With extreme difficulty
5. No, impossible

8) After a meal (sat at a table), how painful has it been for you to stand up from a chair because of your hip?

1. Not at all painful
2. Slightly painful
3. Moderately painful

4. Very painful
5. Unbearable

9) Have you had any trouble getting in and out of a car or using public transportation because of your hip?

1. No trouble at all
2. Very little trouble
3. Moderate trouble
4. Extreme difficulty
5. Impossible to do

10) Have you had any trouble with washing and drying yourself (all over) because of your hip?

1. No trouble at all
2. Very little trouble
3. Moderate trouble
4. Extreme difficulty
5. Impossible to do

11) Could you do the household shopping on your own?

1. Yes, easily
2. With little difficulty
3. With moderate difficulty
4. With extreme difficulty
5. No, impossible

12) How much has pain from your hip interfered with your usual work, including housework?

1. Not at all
2. A little bit
3. Moderately
4. Greatly
5. Totally

Appendix 2. — Harris Hip Score (weighted scores between brackets)

Pain

1. (44) None, or ignores it
2. (40) Slight, occasional, no compromise in activity
3. (30) Mild pain, no effect on average activities, rarely moderate pain with unusual activity, may take aspirin
4. (20) Moderate pain, tolerable but makes concessions to pain. Some limitations of ordinary activity or work. May require occasional pain medication stronger than aspirin
5. (10) Marked pain, serious limitation of activities
6. (0) Totally disabled, crippled, pain in bed, bedridden

Function : Distance walked

1. (11) Unlimited
2. (8) Six blocks (30 min)
3. (5) Two or three blocks (10-15 min)
4. (2) Indoors only
5. (0) Bed and chair only

Function : Support

1. (11) None
2. (7) Cane/Walking stick for long walks
3. (5) Cane/Walking stick most of the time
4. (3) One crutch
5. (2) Two canes/Walking sticks
6. (0) Two crutches or not able to walk

Function : Limp

1. (11) None
2. (8) Slight
3. (5) Moderate
4. (0) Severe or unable to walk

Function : shoe, socks

1. (4) With ease
2. (2) With difficulty
3. (0) Unable to fit or tie

Function : Sitting

1. (5) Comfotably, ordinary chair for one hour
2. (3) On a high chair for 30 minutes
3. (0) Unable to sit comfortably on any chair

Function : Public transportation

1. (1) Able to use transportation (bus)
2. (0) Unable to use public transportation (bus)

Function : Stairs

1. (4) Normally without using a railing
2. (2) Normally using a railing
3. (1) In any manner
4. (0) Unable to do stairs

Deformity

(1) < 30 degrees of flexion contracture
(1) < 10 degrees of adduction contracture
(1) < 10 degrees of endorotation contracture in extension
(1) < 3.2 cm discrepancy in leg length

Motion

Degrees of motion : Multiply by : Count all values, multiply by 0.05 = score of motion

0-45º flexion	(1.0)
45-90º flexion	(0.6)
90-110º flexion	(0.3)
110-130º flexion	(0.0)
0-15º abduction	(0.8)
15-20º abduction	(0.3)
20-45º abduction	(0.0)
0-15º external rotation	(0.4)
> 15º external rotation	(0.0)
any internal rotation	(0.0)
0-15º adduction	(0.2)
> 15º adduction	(0.0)