J Orthop Sci (2009) 14:35–39
DOI 10.1007/s00776-008-1292-9

JOURNAL OF
ORTHOPAEDIC SCIENCE

Original article

Validity and responsiveness of the Oxford hip score in a prospective study with Japanese total hip arthroplasty patients

Yuko Uesugi[1], Kiyoko Makimoto[2], Kimie Fujita[3], Takashi Nishii[4], Takashi Sakai[5], and Nobuhiko Sugano[4]

[1] Department of Nursing, Kobe University Graduate School of Health Sciences, 7-10-2 Tomogaoka, Suma-ku, Kobe, Hyogo 654-0142, Japan
[2] Division of Health Sciences, Osaka University Graduate School of Medicine, Osaka, Japan
[3] Department of Nursing, Faculty of Medicine, Saga University, Saga, Japan
[4] Department of Orthopaedic Medical Engineering, Osaka University Graduate School of Medicine, Osaka, Japan
[5] Department of Orthopaedic Surgery, Osaka University Graduate School of Medicine, Osaka, Japan

Abstract

Background. With the increasing need for disease-specific health outcome measurements, the Oxford hip score was developed to measure health-related quality of life of total hip arthroplasty (THA) patients in the United Kingdom. The Oxford hip score comprises 12 items pertaining to pain and physical function, which are increasingly used to measure health outcomes of patients who have undergone THA. The purpose of this study was to establish the validity and responsiveness of the Oxford hip score in a prospective study of Japanese patients.

Methods. The study was conducted at two hospitals. The eligibility criterion for the study was consenting adult patients who underwent primary unilateral THA between April 2005 and October 2007. Three scales were self-administered at the preoperative stage and 6 months after THA. These scales were the Oxford hip score, the Short Form-36 (SF-36) version 2, and three activities requiring deep flexion of the hip (i.e., clipping one's toenails; use of a Japanese squat toilet; "seiza" — sitting on one's legs on the floor, a common posture while eating in Japan.

Results. A total of 224 consenting adult patients were recruited. Among them, 125 (61.9%) participated in pre- and postoperative surveys. Altogether, 108 (22 men, 86 women; mean age, 58.4 ± 12.5 years) of the 125 patients answered all the items. A significant improvement in the mean scores was observed in all scales. Correlation coefficients between the Oxford hip score and the SF-36 version 2 (physical functioning, role physical, bodily pain) ranged from 0.60 to 0.76 preoperatively and postoperatively. Effect size was 1.7 for pain and 1.3 for physical function. The effect size for seiza was small (0.3).

Conclusions. This study demonstrated the validity and responsiveness of the Oxford hip score in a prospective study. However, it does not measure activities requiring deep flexion of the hip joint, and the use of additional items is suggested.

Offprint requests to: Y. Uesugi
Received: June 27, 2008 / Accepted: September 29, 2008

Introduction

Total hip arthroplasty (THA) is a well-established, reproducible surgical technique for treating various hip disorders and improving quality of life (QOL).[1] In 2006, the number of THAs performed in Japan exceeded 35 000. THA is recognized as a good treatment option for hip disorders. To evaluate the outcomes of THA, self-descriptive QOL scores such as the Short Form-36 (SF-36),[2] the Western Ontario and McMaster universities (WOMAC) index,[3] and the Oxford hip score (OHS)[4,5] have been used.

The SF-36, a generic QOL scale,[6] is the most popular scale for evaluating the health outcomes of THA patients.[7,8] This scale comprises eight subscales, including those for evaluating physical functioning, bodily pain, and general mental health.

With the increasing need for disease-specific health outcome measurements, the WOMAC index,[9] a health-related QOL scale for osteoarthritis, became a popular tool for evaluating THA patients.[10,11] Furthermore, a THA-specific health outcome scale (i.e., the OHS) was developed by Dawson and colleagues in 1996.[4,5] The OHS comprises 12 items pertaining to pain and daily life activities. The reliability and validity of the OHS have been confirmed by large prospective studies.[12,13] The OHS is reported to have a greater effect size than the WOMAC index.[14] In addition, the OHS is in the public domain and can be used free of charge.

The OHS was originally written in English, and one of the authors translated the OHS into Japanese after obtaining permission from Dawson. The translation was validated by back translation. The reliability and validity of the OHS were established by a cross-sectional survey in Japanese THA patients;[15] however, the responsiveness of the OHS has not yet been established. Responsiveness refers to the efficiency with which a scale measures the changes in the condition of a patient after an intervention.

Western studies have shown significant improvements in physical functions in THA patients.[5,10,16,17] However, several surveys conducted in Japan revealed that post-THA patients reported difficulty performing activities that required deep flexion of the hip joint, such as "seiza" — sitting on one's legs on the floor, a common posture while eating in Japan — use of a Japanese toilet, and clipping one's toenails.[18–20] This suggests a need for evaluating these activities before and after THA that are not covered by the OHS.

The aims of this study were as follows: (1) to establish the validity and responsiveness of the OHS and (2) to evaluate the improvements in the physical functions of Japanese THA patients in a prospective study.

Patients and methods

The study was conducted at O university hospital and K community hospital, which is affiliated to O university. The eligibility criterion for study participation was all adult patients who underwent primary unilateral THA at the two study sites between April 2005 and October 2007. The questionnaires were handed to the subjects on the day of admission for THA, which was 2 days prior to surgery at both hospitals. Subsequently, 6 months after THA, the questionnaires were mailed to the subjects at K community hospital. To increase the response rate, the researcher at O university hospital directly gave the questionnaire to the subjects at the outpatient department and requested that they mail the completed questionnaire to the hospital. The following information was obtained from medical records: major diagnosis, type of procedure, duration of hospital stay, and rehabilitation.

Questionnaires

In addition to the OHS and SF-36v2,[21] three additional items were administered to measure the response to activities requiring deep flexion of the hip. These three items were based on a survey of THA patients,[18–20] and the results were confirmed by personal interviews of 20 THA patients, part of which was previously published.[22] These were clipping one's toenails, use of a Japanese squat toilet, and seiza. Seiza, which refers to sitting on one's legs on the floor, is a common posture while eating in Japan.[23] THA patients tend to remodel their houses to Western style to accommodate their restricted ROM. However, access to Western toilets is limited when they go out. Similarly, seiza is required at ceremonial occasions. As reported by the THA patients, the posture employed while clipping one's toenails is one of the most difficult as podiatry services are not readily available. The OHS items were based on a five-point Likert scale. Each item is ranked from 1 to 5, with a higher score representing greater difficulty.

Ethical considerations

This study was approved by the ethics committee of the Osaka University Graduate School of Medicine (421-1). The following points were explained to all the eligible patients: (1) participation was voluntary; (2) nonparticipation did not affect subsequent treatment; and (3) participants could withdraw from the study at any time. Only those who gave written permission to the study were enrolled in the study.

Data analysis

The internal reliability of the OHS, its concurrent validity, and internal responsiveness were examined. Cronbach's α was calculated for checking internal reliability. Next, the concurrent validity of the OHS was tested by examining the correlation coefficients between the OHS subcategories and SF-36v2 subscales before and after THA. Because the OHS and SF-36 comprises ordinal variables, Spearman's correlation coefficients were used to determine the correlation coefficients between the two scales. Internal responsiveness was examined based on the changes in the mean OHS scores between the pre- and post-THA periods using Wilcoxon's test. After THA, most of the items were not expected to demonstrate normal distributions. Hence, the effect size and floor and ceiling effects were calculated to examine the changes in the distribution of the scores.

Results

A total of 224 patients were found to be eligible and 90% of them consented to enroll in the study. Of the questionnaires that were returned, the completion rate for the OHS was higher than that for the SF-36v2 (96.5% and 90.1%, respectively). The response rate of the second survey was 61.9% of that of the first survey, and the completion rate for both questionnaires were comparable (approximately 96%). Table 1 lists the characteristics of the subjects. The characteristics of the subjects who did not respond to the post-THA survey were similar to those of the subjects who did respond. In terms of internal consistency, Cronbach's α was 0.91 for the pre-THA period and 0.92 at 6 months after THA.

A significant improvement in the mean scores was observed in the SF-36v2 subscales, the OHS items, and the three activities requiring deep flexion of the hip (Table 2). Effect sizes for the SF-36v2 varied from 0.3 to 1.8. Bodily pain had the highest effect size, and general

Table 1. Patient characteristics

	University hospital	Community hospital	Total	%
No. of eligible patients	98	126	224	
First survey	94	108	202	90.0
Second survey	61	64	125	61.9
Effective	54	54	108	
Sex				
Males	10	12	22	20.4
Females	44	42	86	79.6
Mean age (years)	57.0	59.8	58.4 ± 12.5	
Diagnosis				
Osteoarthritis	44	47	91	84.3
Avascular necrosis	9	0	9	8.3
Others	1	7	8	7.4

Table 2. Improvement in OHS scores, 8 SF-36v2 subscale scores, and items requiring deep flexion of the hip, sorted in descending order of effect size

Item	Description	Preop mean score	SD	Difference		Postop mean score	SD	Difference		Effect size
				Floor	Ceiling			Floor	Ceiling	
OHS1	Usual pain	3.8	1.0	2.8	22.2	1.7	0.9	48.1	–21.3	2.0
OHS8	Pain while standing	2.7	1.0	10.2	1.9	1.3	0.6	68.5	–1.9	1.4
OHS6	Pain while walking	2.6	1.0	17.6	1.9	1.3	0.7	63.0	–1.0	1.3
OHS9	Limp	3.1	1.4	12.0	22.2	1.4	0.7	61.1	–22.2	1.3
OHS11	Pain interfering with work	2.8	1.1	13.0	4.6	1.3	0.6	61.1	–4.6	1.3
OHS7	Ascending stairs	2.9	1.2	10.2	9.3	1.6	0.9	50.9	–8.4	1.2
OHS10	Sudden pain	2.7	1.3	27.8	9.3	1.2	0.7	60.2	–9.3	1.1
OHS4	Putting on socks	2.7	1.1	15.7	5.6	1.5	0.8	46.3	–5.6	1.0
OHS3	Getting in/out of a car	2.5	1.1	21.3	1.9	1.5	0.8	46.3	–1.0	0.9
OHS12	Pain at night	2.5	1.3	34.3	7.4	1.3	0.6	48.1	–7.4	0.9
OHS2	Washing oneself	2.1	0.9	30.6	0.9	1.3	0.6	44.4	–0.9	0.8
OHS5	While shopping	2.3	1.3	34.3	10.2	1.4	0.9	40.7	–8.3	0.7
OHS total		32.6	9.8	0.0	0.0	16.8	6.6	22.2	0.0	1.6
OHS pain		17.5	5.5	0.0	0.0	8.3	3.4	39.8	0.0	1.7
OHS function		15.1	5.1	0.9	0.0	8.6	3.7	0.0	0.0	1.3
Three items requiring deep flexion of the hip	Clipping toenails	3.4	1.3	9.5	23.8	2.0	1.1	29.4	–18.2	1.1
	Japanese-style toilet (squat on the floor)	4.0	1.2	1.9	47.1	3.0	1.4	13.2	–24.5	0.8
	Seiza (kneel on the floor)	2.6	1.6	36.2	20.0	2.1	1.4	14.7	–7.0	0.3
SF-36v2 (0–100 score)										
BP	Bodily pain	39.1	19.4	6.5	0.0	73.1	22.0	–6.5	23.1	–1.8
PF	Physical functioning	39.7	21.7	2.8	0.0	71.9	19.8	–2.8	1.9	–1.5
RP	Role physical	49.4	30.1	10.2	8.3	78.5	25.3	–9.3	30.6	–1.0
SF	Social functioning	57.9	28.7	3.7	16.7	82.4	25.2	–2.8	36.1	–0.9
RE	Role emotional	61.6	31.7	9.3	21.3	84.4	24.5	–9.3	41.7	–0.7
VT	Vitality	53.2	21.7	0.0	0.9	67.8	19.4	0.0	4.7	–0.7
MH	Mental health	64.0	22.2	0.0	3.7	75.0	18.6	0.0	6.5	–0.5
GH	General health	58.3	19.7	0.0	0.0	63.8	19.3	0.9	2.8	–0.3

$P < 0.01$ indicates significance. All comparisons were significant at $P < 0.01$ (Wilcoxon's test)

health had the smallest effect size. Similarly, the effect sizes of the OHS items varied; items related to pain tended to have a greater effect size than those related to an instrumental activity of daily living (IADL) (Table 2). Regarding the three activities that required deep flexion of the hip, the effect size of seiza was much smaller than that of the other two activities. The effect sizes did not correspond to the increased proportion of ceiling or floor effects in either scale.

The 12 OHS items were grouped into pain and function subscales.[24,25] Using these two subscales, concurrent validity was tested. During the pre-THA period, moderate to strong correlation was observed between the two OHS subscales and the three SF-36v2 pain and physical

Table 3. Spearman's correlation coefficient between the SF-36v2 and OHS subscales before and after THA

	SF-36v2							
OHS	PF	RP	BP	VT	SF	RE	MH	GH
Pre-THA								
Total	(0.72)**	(0.70)**	(0.76)**	(0.39)**	(0.52)**	(0.50)**	(0.35)**	(0.16)
Pain	(0.58)**	(0.60)**	(0.76)**	(0.39)**	(0.47)**	(0.46)**	(0.31)**	(0.14)
Function	(0.75)**	(0.70)**	(0.66)**	(0.33)**	(0.50)**	(0.46)**	(0.35)**	(0.16)
Post-THA								
Total	(0.66)**	(0.64)**	(0.60)**	(0.47)**	(0.51)**	(0.64)**	(0.54)**	(0.43)**
Pain	(0.45)**	(0.56)**	(0.59)**	(0.44)**	(0.53)**	(0.61)**	(0.42)**	(0.40)**
Function	(0.69)**	(0.62)**	(0.48)**	(0.40)**	(0.45)**	(0.59)**	(0.49)**	(0.35)**

THA, total hip arthroplasty; PF, physical functioning; RP, role physical; BP, bodily pain; VT, vitality; SF, social functioning; RE, role emotional; MH, mental health; GH, general health.
** $P < 0.01$

scales, and a moderate correlation was observed between the OHS subscales and the four SF-36v2 subscales (Table 3). The SF-36v2 general health subscale was the only subscale that did not correlate with the OHS in pre-THA period. However, the association between general health and the two OHS subscales became significant. Conversely, the correlation coefficients between the OHS and three SF-36v2 subscales (physical functioning, role physical (actual activity required by position in society), bodily pain) decreased but remained highly significant after THA.

Discussion

This prospective study demonstrated that the OHS had good internal reliability for Japanese patients who had undergone THA. This indicates that the OHS can serve as a valuable tool for evaluating the health outcomes in Japanese THA patients. The large, significant decreases in the mean scores of the 12 OHS items suggest substantial improvement in the physical functioning and bodily pain after THA. Furthermore, the substantial increases in the floor effect indicated that most of the respondents no longer experienced difficulty with regard to most of the OHS items.

Significant improvements were also observed in the activities that require deep flexion of the hip joint. However, 6 months after THA, more than 20% of the respondents still reported extreme difficulty when using a Japanese squat toilet. Clipping one's toenails and seiza were also rated as extremely difficult tasks. This suggests that the post-THA patient may experience difficulty performing activities that are not covered by the OHS. Health care professionals need to inform preoperative patients regarding the potential difficulty when performing daily life activities. Furthermore, when evaluating the health outcomes in Japanese THA patients, items that measure the deep flexion of the hip joints need to be included.

Kneeling and squatting are still commonly practiced in Asian countries as well. However, neither WOMAC nor the OHS has subconstructs to measure activities requiring deep flexion of the hip, and WOMAC validation studies in Korea and Singapore suggest the need to evaluate these activities.[26,27] For future validation of the WOMAC and OHS in non-Western countries, each culture needs to evaluate non-Western lifestyle activities that require a greater range of motion than Western lifestyle activities.

During the pre-THA period, the correlation coefficients between the OHS and SF-36v2 PF subscales were high; this fulfills the concurrent validity criteria.[28] Our findings were similar to those of previous studies.[5] After THA, correlation coefficients between the OHS and SF-36v2 pain and PF-related subscales are reduced, which may be a reflection of a large floor effect after surgery. In contrast, an association between the OHS and general health subscale became significant; this suggested that THA subjects responded in a manner similar to the general population to a certain extent after experiencing significant reductions in pain.

Because of its brevity, the OHS is reported to have a higher completion rate and higher responsiveness than the SF-36.[5] The completion rate in our study was comparable to that in previous studies. However, the effect size of the OHS in our study was lower than that of the OHS in Western countries, where the effect size was around 2.5 during a post-THA period of 3 months to 1 year.[5,12,29] This discrepancy can be explained by the differences in the preoperative scores: European studies reported substantially higher mean preoperative scores than those of our study.[5] This could be interpreted that Western patients have to wait longer than Japanese patients to have THA. Alternatively, Japanese patients may underrate their pain when compared with their Western counterparts. A series of personal interviews with Japanese THA patients revealed that Japanese patients tended to postpone surgery until the pain became unbearable.[22] This implies that THA patients

underrate their pain when the scale is self-administered. In future studies, cross-cultural comparisons of clinical presentations and self-rating of pain in THA patients are required to explore the differences in THA outcomes.

Conclusion

The current study demonstrated the responsiveness of the OHS, although additional items that evaluate the difficulty of executing activities that require deep flexion of the hip need to be included. The OHS has the potential to serve as a responsiveness survey tool for monitoring changes in the health-related QOL of THA patients; additionally, it can serve as an indicator for revision in Japan. However, it is necessary to exercise caution when making cross-country comparisons of patient health outcomes with the OHS.

Acknowledgments. This project was funded by a grant from the Hip Joint Foundation of Japan in 2005.

References

1. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet 2007;370:1508–19.
2. Stucki G, Liang MH, Phillips C, Katz JN. The short form-36 is preferable to the SIP as a generic health status measure in patients undergoing elective total hip arthroplasty. Arthritis Care Res 1995;8:174–81.
3. Laupacis A, Bourne R, Rorabeck C, Feeny D, Wong C, Tugwell P, et al. The effect of elective total hip replacement on health-related quality of life. J Bone Joint Surg Am 1993;75:1619–26.
4. 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.
5. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. Qual Health Care 1996;5:81–8.
6. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). 1. Conceptual framework and item selection. Med Care 1992;30:473–81.
7. Towheed TE, Hochberg MC. Health-related quality of life after total hip replacement. Semin Arthritis Rheum 1996;26:483–91.
8. Quintana JM, Escobar A, Arostegui I, Bilbao A, Azkarate J, Goenaga I, et al. Health-related quality of life and appropriateness or knee or hip joint replacement. Arch Intern Med 2006; 166:220–6.
9. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
10. Rorabeck CH, Bourne RB, Laupacis A, Feeny D, Wong C, Tugwell P, et al. A double-blind study of 250 cases comparing cemented with cementless total hip arthroplasty: cost-effectiveness and its impact on health-related quality of life. Clin Orthop 1994;298: 156–64.
11. McGrory BJ, Freiberg AA, Shinar AA, Harris WH. Correlation of measured range of hip motion following total hip arthroplasty and responses to a questionnaire. J Arthroplasty 1996;11:565–71.
12. Fitzpatrick R, Morris R, Hajat S, Reeves B, Murray DW, Hannen D, et al. The value of short and simple measures to assess outcomes for patients of total hip replacement surgery. Qual Health Care 2000;9:146–50.
13. Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. J Bone Joint Surg Br 2001;83:1125–9.
14. Ostendorf M, van Stel HF, Buskens E, Schrijvers AJ, Marting LN, Verbout AJ, et al. Patient-reported outcome in total hip replacement. J Bone Joint Surg Br 2004;86:801–8.
15. Uesugi Y, Fujita K, Okumiya A. Quality of life (QOL) of total hip arthroplasty patients, and reliability and validity of the Oxford hip score (OHS) Japanese version. J Jpn Soc Nurs Res 2006;29:81–7 (in Japanese).
16. Patil S, Garbuz DS, Greidanus NV, Masri BA, Duncan CP. Quality of life outcomes in revision vs primary total hip arthroplasty. J Arthroplasty 2008;23:550–3.
17. Wiklund I, Romanus B. A comparison of quality of life before and after arthroplasty in patients who had arthrosis of the hip joint. J Bone Joint Surg Am 1991;73:765–9.
18. Izumi K, Hiramatsu T, Tsuchiya T, Kanai K, Kanagawa K. The study on recovery process of daily living in patients and quality of life with total hip arthroplasty. J Jpn Soc Nurs Res 1994;17(2): 9–19 (in Japanese).
19. Kanzaki H, Irie S, Ibuki S, Kato N, Nankaku M, Nakamura T, et al. Physical therapy for the patient with cemented total hip arthroplasty. Hip Joint 2003;29:29–34 (in Japanese).
20. Satoh M, Kawaguchi T, Shimada T, Tani K, Nakayama M. Study on environmental transition of patients who underwent total hip arthroplasty. J Jpn Soc Nurs Res 2005;28:41–50 (in Japanese).
21. Fukuhara S, Suzukamo Y. Manual of the SF-36v2 Japanese version. Kyoto: Institute for Health Outcomes & Process Evaluation Research; 2004 (in Japanese).
22. Fujita K, Makimoto K, Hotokebuchi T. Qualitative study of osteoarthritis patients' experience before and after total hip arthroplasty in Japan. Nurs Health Sci 2006;8:81–7.
23. Mulholland SJ, Wyss UP. Activities of daily living in non-Western cultures: range of motion requirements for hip and knee joint implants. Int J Rehabil Res 2001;24:191–8.
24. Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C. Comparing alternative Rasch-based methods vs raw scores in measuring changes in health. Med Care 2004;42:25–36.
25. Fitzpatrick R, Norquist JM, Jenkinson C, Reeves BC, Morris RW, Murray DW, et al. A comparison of Rasch with Likert scoring to discriminate between patients' evaluations of total hip replacement surgery. Qual Life Res 2004;13:331–8.
26. Bae SC, Lee HS, Yun HR, Kim TH, Yoo DH, Kim SY. Cross-cultural adaptation and validation of Korean Western Ontario and McMaster universities (WOMAC) and Lequesne osteoarthritis indices for clinical research. Osteoarthritis Cartilage 2001;9: 746–50.
27. Thumboo J, Chew LH, Soh CH. Validation of the Western Ontario and McMaster University osteoarthritis index in Asians with osteoarthritis in Singapore. Osteoarthritis Cartilage 2001;9:440–6.
28. Fukuhara S. Why do we need QOL now: QOL as a patient-oriented health outcome. In: Ikegami N, Fukuhara S, Simozuma K, Ikeda S, editors. QOL evaluation handbook for clinicians. Tokyo: Igakusyoin; 2001. pp. 2–7 (in Japanese).
29. Garbuz DS, Xu M, Sayre EC. Patient' 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.