THE JOURNAL OF BONE & JOINT SURGERY (Br)

Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement

J. Dawson, R. Fitzpatrick, S. Frost, R. Gundle, P. McLardy-Smith, D. Murray

From the Institute of Health Sciences, University of Oxford and the Nuffield Orthopaedic Centre, Oxford, England

The Oxford hip score (OHS) is a patient-based instrument for assessment of outcome which is often used after total hip replacement, and the EuroQol 5D (EQ5D) is a patient-based generic questionnaire for health assessment. In an analysis of the outcome at one year of 609 revision hip replacements (RHRs), we compared the OHS and EQ5D scores, postoperative patient satisfaction and change in pain. About 25% of the operations were repeat RHRs. At one year, 57% of patients were very pleased with their operation. The correlation between preoperative and postoperative scores and change scores for the OHS and EQ5D was high. For both instruments the effect sizes were large, but the greater effect size of the OHS suggests that it is particularly sensitive to improvements after RHR. The effect scores of the OHS declined with the number of previous RHRs, while those for the EQ5D seemed less sensitive. Our results confirm the value of the OHS in assessing outcome after RHR.

J Bone Joint Surg [Br] 2001;83-B:1125-9.
Received 18 August 2000; Accepted 17 January 2001

Approximately 14% of total hip replacements (THRs) carried out in the National Health Service (NHS) in the UK are revision procedures. This represents a high cost, and it is therefore important to have effective ways of evaluating and comparing outcomes after revision hip replacement (RHR).

J. Dawson, DPhil, Senior Research Fellow
Oxford Centre for Health Care Research and Development, School of Health Care, Oxford Brookes University, 44 London Road, Headington, Oxford OX3 7PD, UK.

R. Fitzpatrick, PhD, Professor of Public Health and Primary Care
Department of Public Health, Institute of Health Sciences, University of Oxford, Old Road, Headington, Oxford OX3 7LD, UK.

S. Frost, Clinical Data Manager
R. Gundle, FRCS Orth, Consultant Orthopaedic Surgeon
P. McLardy-Smith, FRCS, Consultant Orthopaedic Surgeon
D. Murray, FRCS Orth, Professor of Orthopaedic Surgery
Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford OX3 7LD, UK.

Correspondence should be sent to Dr J. Dawson.

0301-620X/01/811643 $2.00

Many long-term (five- to ten-year) follow-up studies of clinical and technological features of RHR define outcome in terms of the annual rate of subsequent revision.[1] This oversimplifies the matter by implying that patients who do not need a further revision have had a good outcome.[2]

Information may be obtained directly from patients, whose judgements are often valid, reliable and free from clinical bias.[3-5] Patient-based measures of health-related quality of life allow large-scale long-term studies of outcome to be carried out.[6-8] Either condition-specific or generic assessments can be used.

The Oxford hip score[4-9] (OHS) is a 12-item patient-based questionnaire developed and validated specifically to assess function and pain after THR. The EuroQol 5D (EQ5D), a generic instrument for assessing quality of life, identifies 243 possible health states.[10,11] It is based on five questions about mobility, self-care, usual activity, pain/discomfort and anxiety/depression. There are three possible levels of response for each item.[10,11] Each state carries a utility value, which is calculated using time trade-offs. Perfect health and death have utility values of one and zero, respectively, and states worse than death (< 0) are possible.

We have evaluated the OHS as an outcome assessment instrument after RHR. First, we compared its validity with that of the EQ5D. Secondly, we examined the sensitivity to change of the two instruments. Sensitivity to change is an important property of an outcome assessment instrument,[12] particularly when, as in the case of RHR, improvement may be marginal.[13-16] Thirdly, we examined the sensitivity to change of both instruments among subgroups characterised by the number of previous RHRs which had been undertaken.

Patients and Methods

Between September 1996 and April 1999, 609 RHRs were performed at a specialist orthopaedic centre. RHR was defined as removal and replacement of the cup, the stem or both components. To obtain data we used a variety of prepiloted questionnaires and also drew information from the hospital patient administration system.

Patients' preoperative questionnaires. About two weeks before surgery, patients completed the OHS and EQ5D questionnaires.

Table I. Questionnaire response rates in 609 patients who had RHR. For the postal follow-up n = 496

Questionnaire	Number	Percentage
Patients' preoperative	541	*89*
Surgeons at operation (check)	585	*96*
Anaesthetists at operation	416	*68*
Postoperative complications	402	*66*
One-year postal follow-up of those eligible* and still alive at one year	435	*88*
Operation details obtained from medical notes	32	*5*
Complications details obtained from medical notes	139	*23*
Patients known to have died within first year after RHR	20	*3*
Patients who refused to be followed up	5	*<1*

* having reached 12 months after surgery and not known to have had further revision surgery at the time

Table II. Characteristics of patients immediately before RHR

	Number	Percentage
Male	260	*43.0*
Working status (n = 522)		
Working full- or part-time	101	*19.3*
Unable due to disability	59	*11.3*
Retired, unemployed, not looking	362	*69.3*
Main reported diagnosis (n = 588):		
1° or 2° OA	463	*78.7*
Inflammatory arthritis	33	*5.6*
Other	92	*15.6*
Reason for revision (n = 589)		
Aseptic loosening	426	*72.3*
Sepsis	58	*9.8*
Recurrent dislocation	40	*6.8*
Other (e.g. fracture, pain)	65	*11.0*
Patients' assessment of their general health (n = 525):		
Excellent	71	*13.5*
Very good	192	*36.6*
Good	179	*34.1*
Fair	67	*12.8*
Poor	16	*3.0*
Patient reported another lower limb problem (n = 514)	289	*56.2*
Mean age in years (n = 60) (95% CI; range)	68.1 (67.1 to 69.2; 25.1 to 95.1)	
Mean number of days on waiting list* (n = 306 (95% CI; range)	316.3 (296.9 to 335.7; 14 to 987)	
Mean number of GP visits in last 12 months (n = 513) (95% CI; range)	5.5 (4.9 to 5.7; 0 to 25)	
Mean Oxford Hip Score (OHS) (n = 498) (95% CI; range)	43.0 (42.3 to 43.8; 14 to 60)	
Mean EuoQol 5D (EQ5D) score (95% CI; range)	0.32 (0.29 to 0.36; -0.43 to 1.00)	

* excludes private cases and emergency admissions

Surgeons' questionnaire. Surgeons provided details of the implant and surgical technique which they used, and of their own grade.

Anaesthetists' questionnaire. Anaesthetists recorded the patient's weight and physical status, using the rating of the American Society of Anaesthesiologists (ASA).

Questionnaire on postoperative complications. About nine weeks after surgery a clinician in the outpatient department noted whether or not patients had developed any of a number of listed major postoperative complications.

Patient-based questionnaires. One year after surgery we asked patients whether they had undergone any operations since their RHR. They were asked also whether they were very pleased, fairly pleased, not very pleased or very disappointed with their operation, and whether their hip was much improved, slightly improved, unchanged, slightly worse or much worse than before surgery. Transition questions, asking patients to compare a current health state with a specified past state, have been shown to be a valid way of assessing outcome.[3,4,17] Patients again completed the OHS and EQ5D questionnaires.

For each instrument, we calculated change scores (postoperative score minus preoperative score) and effect sizes. The latter measures the extent of change in a standardised way which allows comparison between instruments.[18] To calculate the effect size, we divided the difference between

mean preoperative and postoperative scores by the SD of preoperative scores. An effect size of 1.0 indicates a change of one SD in the sample. Effect sizes of 0.2, 0.5 and 0.8 generally indicate small, medium and large changes, respectively. We predicted that the effect sizes 12 months after RHR would be large for both the OHS and the EQ5D.

Statistical analysis. We used SPSS for Windows (SPSS Inc, Chicago, Illinois). Statistical tests included the chi-squared test, *t*-tests, paired *t*-tests and analysis of variance (ANOVA).

Results

Table I shows the questionnaire response rates, which were generally high.

Table II gives preoperative details of the patients. For 20 hips, the data were missing. In the remaining 589 hips, the reason for RHR was aseptic loosening in 426 (72%), sepsis in 58 (9.8%) and recurrent dislocation in 40 (6.8%). A further 65 prostheses (11%) were replaced for other reasons, including fracture and pain. Repeat RHR was performed on 152 hips (26%), with 52 patients (9%) undergoing their second, third or fourth RHR.

Details of the component replaced were missing for four of the 609 hips. In 346 (57%) it was necessary to replace both stems and cups. Only the stem was replaced in 112 (19%) and only the cup in 147 (24%). In 54% of first-time RHRs, both components were replaced.

Table III shows the high correlation between the preoperative scores for the OHS and EQ5D ($r = -0.67$, $p < 0.001$), postoperative scores ($r = 0.77$, $p < 0.001$) and change scores ($r = 0.59$, $p < 0.001$).

Overall outcomes are given in Table IV. By about nine weeks after surgery, 172 of 516 patients (33%) reported at least one postoperative complication. At one year, 246 of 434 patients (57%) stated that they were very pleased with the results of RHR and 321 (75%) that their hip pain was much improved since the operation. Neither the OHS nor the EQ5D reflected any important age- or gender-related differences.

Table V compares patients' answers about satisfaction and retrospective change with preoperative and postoperative OHS and EQ5D scores, change scores and effect

Table III. Correlation between mean preoperative and one-year postoperative OHS and EQ5D scores and change scores after RHR

	OHS		
	Preop	**Postop**	**Change score**
EuroQol (EQ5D)			
Preop	-0.67*	-0.37*	-0.20†
Postop	-0.32*	-0.77*	0.51*
Change score	0.22*	-0.37*	0.59*

* $p < 0.001$
† $p < 0.05$

Table IV. Postoperative complications and patient-based outcomes at one year

	Number	**Percentage**
Reported at median 9 weeks unless stated otherwise:		
Known to have died (n = 609)	26	*4.3*
Any postoperative complications (n = 516)	172	*33.3*
Any cardiorespiratory postoperative complications (n = 513)	26	*5.1*
Wound infection requiring antibiotics (n = 517)	18	*3.5*
DVT/PE (n = 497)	15	*3.0*
Dislocation of new hip (n = 523)	50	*9.6*
Any other surgery on same hip within first year (incl revision) (n = 437)	59	*13.5*
Revision of stem or cup within first year (n = 609)	17	*2.8*
Dislocation within first year (n = 609)	71	*11.7*
Patient rating of hip operation at one year (n = 434)		
Very pleased	246	*56.7*
Fairly pleased	121	*27.9*
Not very pleased	26	*4.1*
Very disappointed	41	*9.4*
Patient rating of pain at one year compared with before operation at one year (n = 431)		
Much improved	321	*74.5*
Slightly improved	70	*16.2*
Unchanged	14	*3.2*
Slightly worse	14	*3.2*
Much worse	12	*2.8*
Continuous measures		
Mean length of hospital stay in days (n = 599) (95% CI; range)	17.1	*(16.0* to *18.3; 0* to *128)*
Mean OHS at one year (n = 397) (95% CI; range)	26.4	*(25.3* to *27.4; 12* to *60)*
Mean OHS change score at one year (n = 355) (95% CI; range)	15.5	*(14.3* to *16.6; -27* to *44)*
Mean EuroQol 5D score (EQ5D) at one year (95% CI; range)	0.62	*(0.59* to *0.65; -0.59* to *1.00)*
Mean EQ5D change score (95% CI; range)	0.27	*(0.21* to *0.32; -1.18* to 1.43)

* this includes patients who died as well as those who were transferred without any complication

Table V. Comparison (ANOVA) between responses to items regarding satisfaction and retrospective change in hip pain after RHR and pre- and postoperative OHS and EQ5D scores, change scores and effect sizes. Higher OHS scores indicate poorer results, while higher EQ5D scores show better results

	OHS				EQ5D			
Postoperative item	**Mean preop score (95% CI)**	**Mean postop score (95% CI)**	**Mean change score (95% CI)**	**Effect size**	**Mean preop score (95% CI)**	**Mean postop score (95% CI)**	**Mean change score (95% CI)**	**Effect size**
Patient satisfaction								
Very pleased	41.1 (39.9 to 42.3)	20.6 (19.8 to 21.5)	20.2 (18.9 to 21.4)	2.36	0.40 (0.33 to 0.46)	0.75 (0.72 to 0.78)	0.35 (0.28 to 0.42)	1.17
Fairly pleased	43.5 (42.0 to 45.0)	30.1 (28.6 to 31.6)	12.8 (10.9 to 14.6)	1.70	0.33 (0.24 to 0.42)	0.54 (0.49 to 0.59)	0.27 (0.17 to 0.36)	0.61
Not very pleased	44.4 (41.1 to 47.6)	38.4 (34.9 to 41.9)	6.5 (3.3 to 9.6)	0.76	0.26 (0.07 to 0.45)	0.42 (0.29 to 0.54)	0.09 (-0.21 to 0.40)	0.50
Very disappointed	44.1 (41.2 to 47.1)	42.8 (38.9 to 46.6)	0.8 (-3.0 to 4.6)	0.16	0.13 (-0.02 to 0.28)	0.15 (0.04 to 0.27)	-0.06 (-0.23 to 0.12)	0.08
	f = 3.2 df3	f = 122.5 df3	f = 55.10 df3		f = 3.8 df3	f = 67.0 df3	f = 7.27 df3	
	p < 0.05	p < 0.001	p < 0.001		p < 0.05	p < 0.001	p < 0.001	
Retrospective change in hip pain								
Much improved	41.8 (40.8 to 42.9)	22.4 (21.5 to 23.3)	19.0 (17.8 to 20.1)	2.26	0.38 (0.33 to 0.44)	0.70 (0.67 to 0.73)	0.31 (0.25 to 0.37)	0.99
Slightly improved	43.6 (41.8 to 45.4)	34.1 (32.1 to 36.2)	8.8 (6.9 to 10.7)	1.36	0.23 (0.12 to 0.35)	0.50 (0.44 to 0.57)	0.25 (0.14 to 0.37)	0.87
Unchanged	38.5 (30.6 to 46.4)	38.7 (30.7 to 46.7)	-2.9 (-11.5 to 5.8)	-0.02	0.27 (-0.16 to 0.69)	0.35 (0.15 to 0.54)	0.05 (-0.31 to 0.40)	0.19
Slightly worse	45.4 (41.5 to 49.3)	45.7 (41.3 to 50.1)	0.0 (-5.3 to 5.3)	-0.05	0.10 (-0.25 to 0.46)	0.24 (0.05 to 0.43)	0.04 (-0.30 to 0.31)	0.46
Much worse	45.7 (39.5 to 51.9)	51.1 (46.6 to 55.6)	-3.3 (-7.3 to 0.6)	-1.04	0.17 (-0.27 to 0.61)	-0.16 (-0.34 to 0.03)	-0.39 (-1.28 to 0.49)	-1.20
	f = 2.09 df4	f = 86.8 df4	f = 46.84 df4		f = 2.57 df4	f = 38.21 df4	f = 5.12 df4	
	p = 0.087	p < 0.001	p < 0.001		p < 0.05	p < 0.001	p < 0.01	

Table VI. Short-term and one-year outcomes according to number of previous revisions on same hip reported at a median of nine weeks unless stated otherwise

	0		**1**		**2+**		**Chi-squared**	
	Number	**Percentage**	**Number**	**Percentage**	**Number**	**Percentage**	**df**	**p value**
Reported at a median of nine weeks unless otherwise stated								
Any postoperative complications (n = 515)	137	*35.0*	18	*23.1*	17	*37.0*	4.47 (2)	NS
Any cardiorespiratory post operative complications (n = 512)	24	*6.2*	2	*2.6*	1	*2.2*	2.63 (2)	NS
Wound infection (n = 516	15	*3.8*	2	*2.5*	1	*2.2*	0.59 (2)	NS
DVT/PE (n = 510)	12	*3.1*	2	*2.6*	1	*2.2*	0.16 (2)	NS
Dislocation of new hip (n = 518)	36	*9.2*	6	*7.5*	8	*17.4*	3.70 (2)	NS
Any other surgery on same hip within first year (incl revision) (n = 435)	34	*10.3*	12	*17.6*	13	*36.1*	18.43 (1) chi-squared trend	<0.001
Dislocation within first year (n = 600)	45	*10.0*	12	*12.0*	14	*26.9*	10.09 (1) chi-squared trend	0.001
Revision of stem or cup within first year (n = 600)	11	*2.5*	4	*4.0*	2	*3.8*	0.92 (2)	NS
Death within first year (n = 600)	16	*3.6*	4	*4.0*	0	*0.0*	2.01 (2)	NS
'Very pleased' with hip operation at one year (n = 431)	195	*59.3*	34	*50.2*	17	*50.0*	2.73 (2)	NS
Pain much better than before hip operation at one year (n = 428)	252	*77.1*	46	*67.6*	22	*66.7*	3.89 (2)	NS
Continuous measures (ANOVA)	Mean (95% CI; range)		Mean (95% CI; range)		Mean (95% CI; range)		F	
Mean LOS (n = 438)	16.8 (15.2 to 18.5; 0 to 128)		19.4 (15.6 to 23.2; 0 to 111)		21.8 (16.8 to 26.7; 6 to 87)		2.69 (2)	0.069
OHS at one year (n = 396)	25.1 (23.9 to 26.3; 12 to 60)		29.1 (26.4 to 31.8; 12 to 57)		31.6 (27.6 to 35.5; 14 to 55)		8.26 (2)	0.001
OHS change score at one year (n = 347)	16.4 (15.1 to 17.7; -15.0 to 17.7)		12.8 (9.6 to 16.0; -27 to 33)		11.6 (8.1 to 15.2; -5 to 30)		4.48 (2)	0.012
EQ5D at one year (n = 138)	0.64 (0.61 to 0.67; -0.59 to 1.00)		0.58 (0.50 to 0.65; -0.24 to 1.00)		0.51 (0.40 to 0.63; -0.24 to 1.00)		3.22 (2)	0.041
EQ5D change score (n = 128)	0.29 (0.61 to 0.67; -0.59 to 1.00)		0.19 (0.50 to 0.65; -0.24 to 1.00)		0.17 (0.40 to 0.63; -0.24 to 1.00)		1.43 (2)	NS

sizes. In line with diminishing patient satisfaction, OHS postoperative scores deteriorated. Both OHS and EQ5D change scores clearly reflected the different levels of patient satisfaction, but there was some overlap in the scores of patients who were not very pleased or very disappointed. The overall effect size was 1.93 and -0.89 for the OHS and the EQ5D, respectively. For those patients who were most pleased, positive changes were reflected in both OHS and EQ5D effect sizes.

There were differences between the OHS postoperative scores and change scores of patients whose hip pain had improved since the operation and those who said it was

much worse. The OHS effect size was approximately zero when pain was unchanged and was negative when pain was much worse.

We divided patients into subgroups according to whether their hips had previously been replaced once, twice or more and summarise the results of our analysis in Table VI. With each RHR, the risks both of suffering hip dislocation and of having to undergo another RHR increased. In the first year after surgery, dislocation of the new implant occurred in over a quarter of patients who had undergone two or more previous RHRs and over a third needed a further operation.

The OHS postoperative scores at one year deteriorated progressively in line with the number of previous RHRs ($p < 0.001$) and the OHS change scores also differed ($p = 0.012$). The difference in change scores (16.4-11.6 = 4.8) was greatest between patients undergoing RHR for the first time and those who had previously undergone RHR twice or more. The difference remained significant after adjustment for age and gender ($p = 0.017$). The corresponding effect sizes were 2.11, 1.68 and 1.58 for groups who had undergone RHR once, twice or more, respectively. The EQ5D was less sensitive than the OHS in assessing change if there had been previous RHR. While EQ5D scores progressively deteriorated ($p = 0.041$), differences in change scores were not significant. The corresponding EQ5D effect sizes were -0.89, -0.78 and -0.87 for patients who had undergone RHR once, twice or more, respectively.

Discussion

For any specific application, the value of an outcome assessment instrument should be tested rather than assumed. To be relevant and useful, it must be demonstrably sensitive to clinically important change.[19] We examined aspects of the OHS by comparing OHS and EQ5D scores, change scores and effect sizes.

These instruments reflected considerable variation in outcome after RHR. There was a high level of agreement between OHS and EQ5D preoperative, postoperative and change scores. For both instruments the effect sizes were large, but the greater effect size of the OHS suggests that it is particularly sensitive to improvements after RHR. This is further evidence of its construct validity.[4,9]

One year after RHR we compared the OHS and EQ5D scores with patients' responses about postoperative satisfaction and changes in pain levels. There was a high level of agreement in all four measures. At one year, pain was reported to be much improved in 75% of hips. In the OHS developmental study, dealing chiefly with primary THA, pain was reported to be much improved in 86%.[9]

We further examined the sensitivity of the OHS to marginal change by analysing the results of subgroups of patients who had previously undergone RHR. Despite overall improvement, OHS postoperative and change scores progressively deteriorated with the number of previous RHRs. The EQ5D showed less sensitivity to change.

We believe that patients are generally likely to cooperate in providing information to the hospital at which they are treated, and that it is feasible to use postal questionnaires to collect informative data.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References

1. **Izquierdo RJ, Northmore-Ball MD.** Long-term results of revision hip arthroplasty: survival analysis with special reference to the femoral component. *J Bone Joint Surg [Br]* 1994;76-B:34-9.
2. **Herberts P, Malchau H.** How outcome studies have changed total hip arthroplasty practices in Sweden. *Clin Orthop* 1997;344:44-60.
3. **Fitzpatrick R, Fletcher A, Gore S, et al.** Quality of life measures in health care. I: Applications and issues in assessment. *BMJ* 1992;305:1074-7.
4. **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.
5. **Rothwell PM, McDowell Z, Wong CK, Dorman PJ.** Doctors and patients don't agree: cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. *BMJ* 1997;314:1580-3.
6. **Dawson J, Fitzpatrick R, Murray D, Carr A.** The problem of 'noise' in monitoring patient-assessed outcomes: generic, disease-specific and site-specific instruments for total hip replacement. *J Health Serv Res Policy* 1996;1:224-31.
7. **Bombardier C, Ware J, Russell IJ, et al.** Auranofin therapy and quality of life in patients with rheumatoid arthritis: results of a multicenter trial. *Am J Med* 1986;81:565-78.
8. **Guyatt GH, Townsend M, Pugsley SO, et al.** Bronchodilators in chronic air-flow limitation: effects on airway function, exercise capacity, and quality of life. *Am Rev Respir Dis* 1987;135:1069-74.
9. **Dawson J, Fitzpatrick R, Carr A, Murray DW.** Questionnaire on the perception of patients about total hip replacement. *J Bone Joint Surg [Br]* 1996;78-B:185-90.
10. **Brazier J, Jones N, Kind P.** Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. *Qual Life Res* 1993;2:169-80.
11. **Wolfe F, Hawley DJ.** Measurement of the quality of life in rheumatic disorders using the EuroQol. *Br J Rheumatol* 1997;36:786-93.
12. **Guyatt GH, Feeny DH, Patrick DL.** Measuring health-related quality of life. *Ann Intern Med* 1993;118:622-9.
13. **Robinson AH, Palmer CR, Villar RN.** Is revision as good as primary hip replacement?: a comparison of quality of life. *J Bone Joint Surg [Br]* 1999;81-B:42-5.
14. **Espehaug B, Havelin LI, Engesaeter LB, Langeland N, Vollset SE.** Patient satisfaction and function after primary and revision total hip replacement. *Clin Orthop* 1998;351:135-48.
15. **Hozack WJ, Rothman RH, Albert TJ, Balderston RA, Eng K.** Relationship of total hip arthroplasty outcomes to other orthopaedic procedures. *Clin Orthop* 1997;344:88-93.
16. **Gore DR, Murray MP, Gardner GM, Mollinger LA.** Comparison of function two years after revision of failed total hip arthroplasty and primary hip arthroplasty. *Clin Orthop* 1986;208:168-73.
17. **Patrick DL, Deyo RA.** Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27:(3 Suppl):217-32.
18. **Kazis LE, Anderson JJ, Meenan RF.** Effect sizes for interpreting changes in health status. *Med Care* 1989;27:178-89.
19. **Guyatt G, Walter GS, Norman G.** Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171-8.