

Specific or general health outcome measures in the evaluation of total hip replacement

A COMPARISON BETWEEN THE HARRIS HIP SCORE AND THE NOTTINGHAM HEALTH PROFILE

Göran Garellick, Henrik Malchau, Peter Herberts

From Göteborg University, Sweden

We assessed 100 patients with a primary total hip replacement using the Harris hip score and the Nottingham Health Profile at one, three and five years after operation. They were derived from two prospective randomised series of cemented and uncemented replacements.

Both scoring systems correlated highly and were each heavily influenced by the system of functional classification defined by Charnley. After five years both reflected the function of the implant and the general state of the patient. A higher degree of sensitivity is needed to show differences in the performance of an implant in the short and medium term. We do not yet have an ideal system of clinical assessment and the overall function must always be properly assessed.

J Bone Joint Surg [Br] 1998;80-B:600-6.

Received 5 September 1997; Accepted after revision 9 January 1998

Specific scoring systems have been widely used to assess the clinical results¹⁻⁴ after total hip replacement. Most attempt to measure the outcomes of technique and procedure and do not assess the effect on general function and the satisfaction of the patient.⁵

We have evaluated the quality of life before and after operation over a minimum of five years and compared the results with the information obtained from a specific scoring system in order to see whether a general health profile was more sensitive than a conventional hip score in describing the outcome of hip replacement.

Patients and Methods

We made a prospective study of 100 patients, 54 of whom had had a cemented total hip replacement and 46 an unce-

G. Garellick, MD, PhD

H. Malchau, MD, PhD

P. Herberts, MD, PhD

Department of Orthopaedics, Institute of Surgical Sciences, Sahlgrenska University Hospital, Göteborg University, S-413 45 Göteborg, Sweden.

Correspondence should be sent to Dr G. Garellick.

©1998 British Editorial Society of Bone and Joint Surgery
0301-620X/98/48345 \$2.00

mented prosthesis. Both groups were cohorts of larger prospective and randomised studies comparing two cemented (Charnley and Spectron) and two uncemented (Harris-Galante I and PCA) systems. The groups were classified into clinical categories according to Charnley¹ in which category A describes unilateral hip disease, category B bilateral hip disease and category C multiple joint disease or other disabilities impairing walking ability. The details of the patients are shown in Table I. Patients were examined before operation and at one, three, and five years after. An independent physiotherapist performed the clinical follow-up.

Harris hip score (HHS).² In the HHS, activities of daily living and gait account for 47 points, pain for 44, joint movement for 5 and absence of deformity for 4 in a total of 100. Harris defined 90 to 100 points as excellent, 80 to 90 as good, 70 to 80 as fair and below 70 as poor.

Nottingham Health Profile (NHP). This is a health-related quality-of-life score⁶ developed and tested for its validity and reliability in the UK^{7,8} and in Sweden in a translated form.⁹ It consists of two parts. The first contains 38 questions requiring the answer 'yes' or 'no', dealing with six aspects of health, namely pain, energy, sleep, mobility, emotional reaction and social isolation. The items are weighted and each yields a value of between 0 and 100, with the worst state being 100. The second part has seven sections, answered as yes or no, which reflect the frequency of problems with occupation, housework, social life, family life, sexual function, hobbies and holidays. The answers can be compared with the average scores in a population matched for age and gender. The values of all six items in part 1 are added and the total divided by six to give the global score.

Patient satisfaction. At each follow-up the patients were asked if they were satisfied, uncertain or dissatisfied with the result of the operation.

Radiological evaluation. At each review an antero-posterior (AP) radiograph of the pelvis was taken together with AP and true lateral views of the hip. Radiological evaluation of the cup and stem was performed according to the criteria of Hodgkinson, Shelley and Wroblewski¹⁰ and Harris, McCarthy and O'Neill¹¹ (modified system), respectively.

Statistical analysis. All statistical calculations were made on a personal computer using SPSS for Windows 95 (Ver.

Table I. Details of the patients in both groups

	Cemented (n = 54)	Uncemented (n = 46)
Mean age in years (median; range)	71 (71; 58 to 85)	49 (52; 21 to 68)
Gender (%)		
Female	57	50
Male	43	50
Diagnosis (%)		
Osteoarthritis	68	39
Rheumatoid arthritis	13	-
Fracture of the hip	17	2
Avascular necrosis	-	13
Childhood disease	2	26
Other	-	20
Clinical category (%)		
A	50	54
B	26	41
C	24	4
Other major medical condition (%)		
Yes	57	33
No	43	67
Mean duration of symptoms in the hip in years (median; range)	6.8 (4.0; 0 to 45)	11.4 (5.0; 0 to 50)
Occupational status (%)		
Retired	100	37
Unfit for work	-	39
Working	-	24
Marital status (%)		
Married or cohabiting	41	37
Single, divorced or widowed	59	63
Implant (%)		
Charnley	54	-
Spectron	46	-
Harris-Galante I	-	52
PCA	-	48

Table II. Details of the number of patients who had died or had revision at each follow-up for both groups

Group	Number of patients	Follow-up (year)								
		1			3			5		
		Number of patients	Died	Revised	Number of patients	Died	Revised	Number of patients	Died	Revised
Cemented	54	52	2	-	50	3	1*	45	8	1*
Uncemented	46	46	-	-	45	-	1†	45	-	1†

* late infection

† aseptic loosening

7.5; SPSS, Chicago, Illinois). The statistical methods used are given in the text. Two-tailed tests were performed.

Results

No patient was lost to follow-up. Eight died and two hips were revised within the five years (Table II).

HHS and NHP before operation. The mean HHS was 38 (8 to 76) and the mean pain score 12 (0 to 30). The lowest HHS was recorded for category-C patients (Table III). Part 1 of the NHP before operation is compared with a population matched for age and gender in Figure 1. The most pronounced problems were with pain, energy and mobility. Patients had difficulty in daily activities especially housework, holidays and hobbies (Fig. 2). Since 61% of the

Table III. Mean values for the HHS, pain score and the NHP global score before operation

Group	Number of patients	HHS	Pain score	NHP global score
All patients	100	38	12	49
Category				
A	52	40	13	45
B	33	38	12	52
C	15	31	9	55

patients were retired we excluded them from assessment regarding problems with paid employment. The mean global score was 49 (13 to 95) (Table III).

Results at one and three years. One year after operation the mean HHS was 88 (61 to 100) and the mean pain score

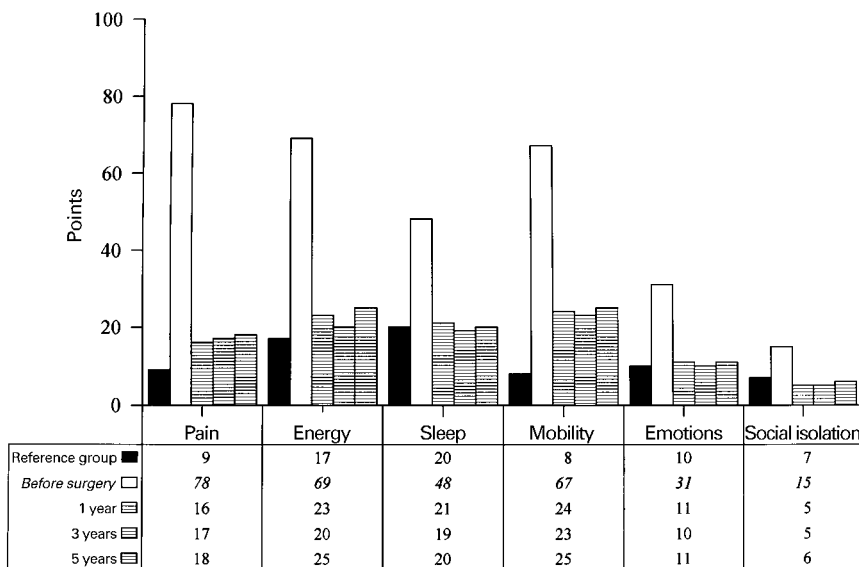


Fig. 1

NHP part 1 before operation and at one, three and five years compared with an age- and gender-matched population.

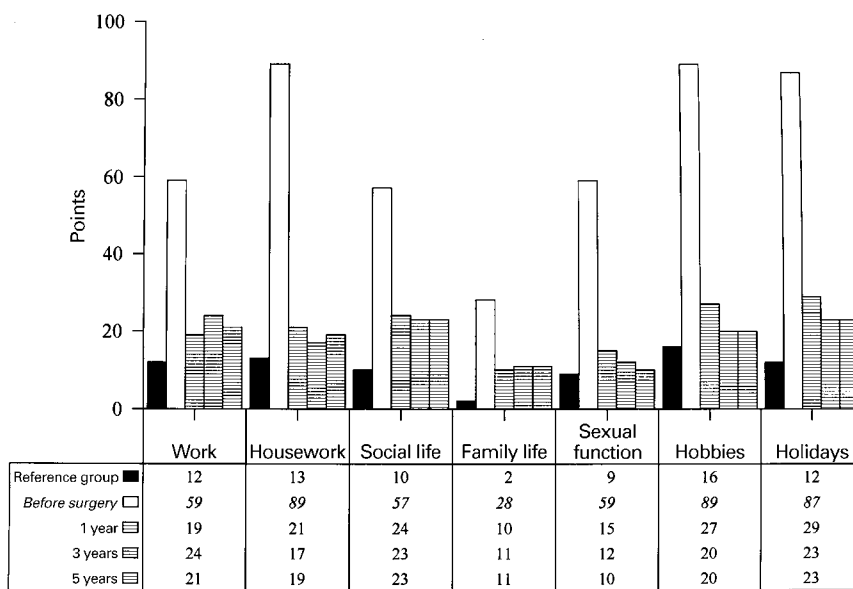


Fig. 2

NHP part 2 before operation and at one, three and five years compared with an age- and gender-matched population.

42 (20 to 44). The patients had progressed within all aspects of the NHP part 1 (Wilcoxon rank test, $p < 0.001$ for all items except for social isolation in which $p = 0.001$; Fig. 1). There was significant improvement in all items of the activities of daily living (McNemar test, $p < 0.001$ except for family life in which $p = 0.05$; Fig. 2).

At three years the results for both methods of assessment were similar to those at one year (Figs 1 and 2).

Results at five years

Harris hip score. The mean HHS after five years was 89 (44 to 100), varying from 96 (65 to 100) in category A to 78 (49 to 100) in category C (Mann-Whitney U test,

Table IV. Mean values for the HHS, pain score, the NHP global score and satisfaction at five years

Group	Number of patients	HHS	Δ -value*	Pain score	Good to excellent (%)	NHP global score	Δ -value*	Satisfaction (%)
All patients	90	89	52	42	78	18	30	92
Category								
A	34	96	55	43	91	5	38	100
B	27	91	53	42	89	14	36	89
C	29	78	50	41	55	37	16	86

* gain after operation

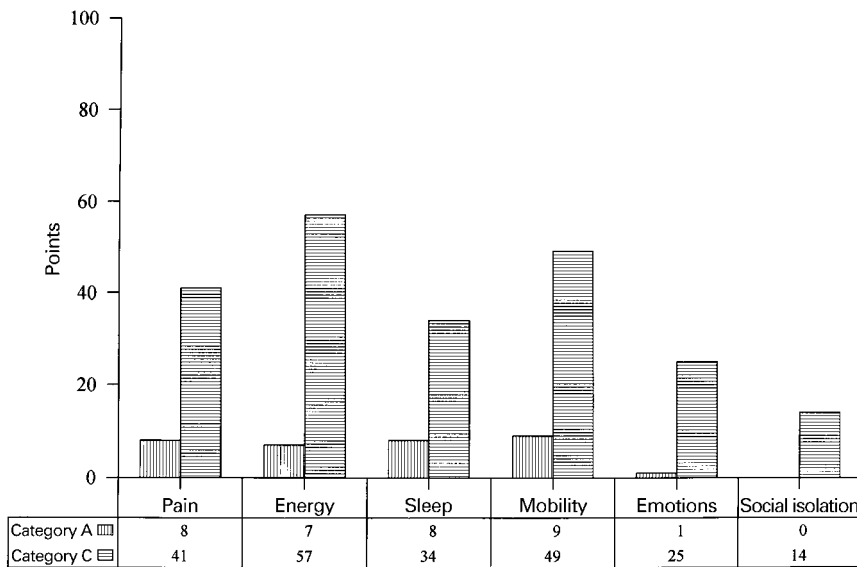


Fig. 3

NHP part 1 at five years. Charnley category A is compared with category C.

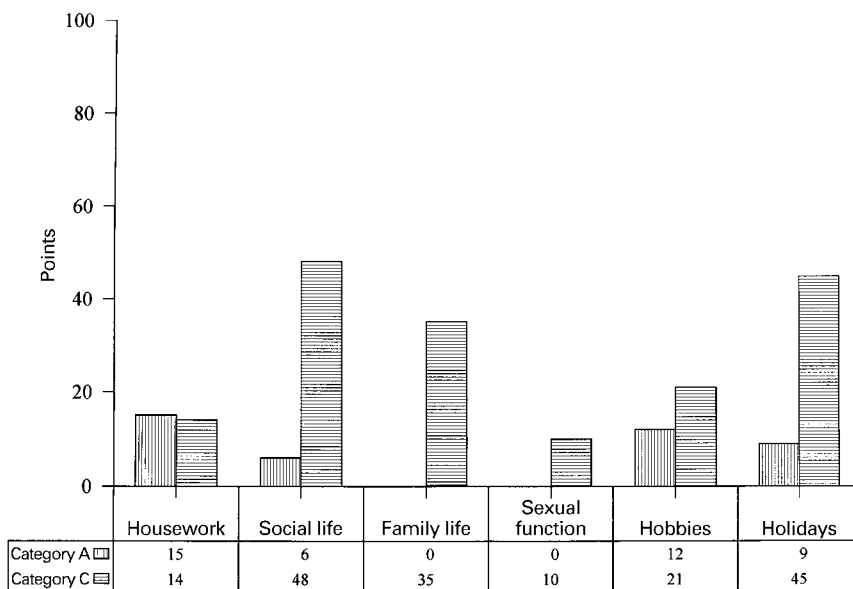


Fig. 4

NHP part 2 at five years. Charnley category A is compared with category C. The ability to work is not considered since most of the patients in category C were retired.

$p < 0.001$). The difference between the HHS and the pain score before and after operation (Δ -value) did not differ significantly between clinical categories. According to the Harris definition of success 78% of the patients were classified as good or excellent. In categories A and C 91% and 55% were good or excellent, respectively (Fisher's exact test, $p < 0.001$). Despite the low scores in category C, most of these patients were satisfied and free from pain (Table IV).

Nottingham Health Profile. The gain in the quality of life continued over the five years and was not affected by the length of follow-up (Figs 1 and 2). The global score varied significantly between categories A and C (Mann-Whitney U test, $p < 0.001$; Table IV) and in contrast to the HHS had a significant gain between category A and category C (Mann-Whitney U test, $p < 0.001$).

Charnley classification. The distribution of patients in the Charnley categories varied over time. Before operation 15% were classified in category C increasing to 32% by five years. This trend was most pronounced in the older, cemented group with a change from 24% to 56%.

Background variables. The marital and socioeconomic status and the duration of symptoms in the hip were evaluated and did not correlate with either the Harris or the NHP ratings. When comparing category C with category A there was a significant deterioration in all items in both the NHP part 1 (Mann-Whitney U test, $p < 0.001$) and part 2 (chi-squared test, $p < 0.001$, sexual function $p = 0.05$), except for housework and hobbies (Figs 3 and 4).

Comparison of the HHS with the NHP. The HHS had a strong correlation with the NHP global score at the follow-up at five years (Spearman correlation, $r = -0.821$,

Table V. Follow-up of the different implants at five years

Implant	Number of patients	HHS	Pain score	NHP global score	Charnley category C (%)	Satisfaction (%)
Charnley	22	86	42	25	50	96
Spectron	23	85	43	24	61	87
Harris-Galante I	23	92	42	12	9	96
PCA	22	92	41	8	9	91
Cemented	45	86	42	24	56	91
Uncemented	45	92	41	10	9	93

Table VI. Clinical outcome of patients with radiologically loose components at five years

Case	Prosthesis	Component	Loosening	HHS	Pain score	NHP global score	Satisfaction	Charnley category
1	Spectron	Cup	Grade 3	94	44	5	Satisfied	B
2	Spectron	Cup	Grade 3	89	44	52	Satisfied	C
3	Spectron	Cup	Grade 3	100	44	0	Satisfied	A
4	Spectron	Cup	Grade 3	73	44	52	Satisfied	C
5	PCA	Stem	Definitely	95	44	24	Satisfied	B
Mean	-	-	-	90	44	27		

$p < 0.001$). Stepwise linear regression analysis with the HHS as the dependent variable showed that only the Charnley category significantly influenced the total HHS ($p < 0.001$). The age, diagnosis, marital or socioeconomic status and duration of hip disease were not significantly associated with the HHS. Regression analysis with the global score as a dependent variable showed significance for the Charnley category ($p < 0.001$) and age ($p = 0.035$), but not for the other background variables.

Implant groups. Within the two groups there were no differences in either the HHS or the NHP global score between patients receiving a Charnley and a Spectron prosthesis or a Harris-Galante I and a PCA implant. When the cemented and uncemented groups were compared, the HHS (Mann-Whitney U test, $p = 0.034$) and the NHP global score (Mann-Whitney U test, $p = 0.004$) differed significantly, but there were no differences in the pain score or in overall satisfaction. In the cemented group 56% of the patients were in category C after five years compared with only 9% in the uncemented group (Fisher's exact test, $p < 0.001$; Table V).

Radiological loosening and clinical outcome. Five implants were loose at five years (Table VI). None of these patients had clinical symptoms and there were no significant differences in the HHS, the pain score or the NHP global score compared with the 85 patients without radiological signs of loosening. The mean value for pain in the NHP part 1 was 28 (0 to 65) in patients with radiological loosening and 20 (0 to 100) in the remainder (Mann-Whitney U test, $p = 0.22$). We found no significant differences for any other item in the NHP when comparing loose and sound implants.

Discussion

We had felt that the NHP would give a more sensitive and consistent evaluation than the HHS and that it would be

more independent of the different Charnley categories. We found, however, that the clinical category significantly influenced both methods of assessment. This was most apparent in the NHP and significantly affected the gain (Δ -value) of the global score.

There have been several critical studies which have analysed and compared different methods of assessing hip replacements.¹²⁻¹⁶ Many have concluded that the different hip scores give conflicting appraisals of success in the same patient and that they lack validation. All measurements of outcome need testing for validity, reliability and responsiveness, but few of those used for assessing the hip have been examined in this way.

The HHS is commonly used for clinical evaluation after total hip replacement, but the method has not been critically assessed or compared with age- and gender-matched control series. The suggested subgrouping of results into poor, fair, good and excellent is questionable since most of the patients in our series who scored less than 80 points were free from pain and satisfied. It may be more appropriate to measure improvement in the score rather than the total, but this requires a prospective study.

The case mix is probably one reason for inconsistency in results in scoring systems. Brinker et al¹⁷ studied 200 adults with no history of hip disease and a mean age of 65 years (50 to 100) using three different hip scores, including the HHS. Advanced age and the presence of two or more major medical conditions had a significant negative correlation with the hip score. They observed that the case mix must be considered when comparing different studies and that hip scores may decline over a long period due to change in the patient's age and/or medical condition rather than to any factor relating to the hip replacement. The proportion of category-C patients increases with time; in our study this was from 24% to 56% over five years in the cemented group. Elderly patients may become less fit due to ageing, giving a

lower score despite unchanged good function of the hip implant.

Because of restriction of costs and resources in health care there has been increasing interest in methods of measuring the health-related quality of life. The advantage of a generic compared with a disease-specific approach is the ability to compare outcomes across different populations and methods of treatment, which may assist studies in cost-effectiveness, policy analysis and decision-making. A number of ways of measuring the subjective quality of life are now available.^{6,18-20} The length and complexity of some of these questionnaires, however, may present difficulty. The NHP and the additional questions (background variables) request answers to 50 questions at every follow-up. This is laborious to carry out and elderly patients may have problems in understanding and answering some of the questions.

Franzen, Johnsson and Nilsson²¹ posted the NHP questionnaire to 225 patients with a mean age of 81 years (61 to 98); 187 patients (83%) answered. This is a low response rate in a long-term assessment of total hip replacement which has failure rates of 10% to 15% after 15 years.²²⁻²⁵ Extensive questionnaires can be used in controlled trials especially for effectiveness studies, but they are probably less suitable for long-term follow-up of randomised trials or routine quality control in larger populations. It may be sufficient to assess pain relief and overall satisfaction in larger long-term trials. Table VI shows significant differences between the two cohorts for both the HHS and NHP after five years, but this may only reflect the functional category and the difference in age.

Nilsson et al²⁶ used the NHP to evaluate the clinical outcome of 49 total hip replacements at five years. They found a significant reduction in the score in eight patients with radiological evidence of loosening but not enough clinical symptoms to consider revision. They concluded that this was a more sensitive method of assessment than a disease-specific score. We have not been able to confirm these findings. The five patients with radiological signs of loosening at the five-year follow-up in our series had no symptoms and the scores were not affected. In both groups, however, the number of patients with loosening was small. Several of the studies have shown a low correlation between early radiological findings and clinical symptoms.^{27,28}

Conclusion. The HHS and the NHP correlated well with each other and were both heavily influenced by the Charnley system of classification. Both scores reflected general health and function but were not able to detect subtle differences in the performance of two different types of hip replacement at five years.

The classification proposed by Charnley should be recorded in outcome studies of total hip replacements and the case mix must be considered to avoid demographic bias.^{17,29} A short, simple and internationally standardised hip score is needed; it must be sound scientifically and

include generic and specific questions,³⁰ together with those on overall satisfaction and pain relief.

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. Charnley J. *Low friction arthroplasty of the hip*. Berlin: Springer Verlag, 1979:246.
2. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fracture: treatment by mold arthroplasty. *J Bone Joint Surg [Am]* 1969;51-A:737-55.
3. Merle d'Aubigné R, Postel M. Functional results of hip arthroplasty with acrylic prosthesis. *J Bone Joint Surg [Am]* 1954;36-A:451-75.
4. Pellici PM, Wilson PD, Sledge CB, et al. Long-term results of revision total hip replacement: a follow-up report. *J Bone Joint Surg [Am]* 1985;67-A:513-6.
5. Gartland JJ. Orthopaedic clinical research: deficiencies in experimental design and determinations of outcome. *J Bone Joint Surg [Am]* 1988;70-A:1357-64.
6. Hunt SM, McKenna SP, McEwen J, Williams J, Papp E. The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med A* 1981;15:221-9.
7. Hunt SM, McKenna SP, McEwen J, et al. A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health* 1980;34:281-6.
8. Hunt SM, McKenna SP, Williams J. Reliability of a population survey tool for measuring perceived health problems: a study of patients with osteoarthritis. *J Epidemiol Community Health* 1981;35:297-300.
9. Wiklund I, Romanus B, Hunt SM. Self-assessed disability in patients with arthrosis of the hip joint: reliability of the Swedish version of the Nottingham Health Profile. *Int Disabil Stud* 1988;10:159-63.
10. Hodgkinson JP, Shelley P, Wroblewski BM. The correlation between the roentgenographic appearance and the operative findings at the bone-cement junction of the socket in Charnley low friction arthroplasties. *Clin Orthop* 1988;228:105-9.
11. Harris WH, McCarthy JC, O'Neill DA. Femoral component loosening using contemporary techniques of femoral cement fixation. *J Bone Joint Surg [Am]* 1982;64-A:1063-7.
12. Bryant MJ, Kernohan WG, Nixon JR, Mollan RAB. A statistical analysis of hip scores. *J Bone Joint Surg [Br]* 1993;75-B:705-9.
13. Andersson G. Hip assessment: a comparison of nine different methods. *J Bone Joint Surg [Br]* 1972;54-B:621-5.
14. Britton AR, Murray DW, Bulstrode CJ, McPherson K, Denham RA. Pain levels after total hip replacement: their use as endpoints for survival analysis. *J Bone Joint Surg [Br]* 1997;79-B:93-8.
15. Callaghan JJ, Dysart SH, Savory CG, Hopkinson WJ. Assessing the results of hip replacement: a comparison of five different rating systems. *J Bone Joint Surg [Br]* 1990;72-B:1008-9.
16. Galante J. Editorial. The need for a standardized system for evaluating results of total hip surgery. *J Bone Joint Surg [Am]* 1985;67-A:511-2.
17. Brinker MR, Lund PJ, Cox DD, Barrack RL. Demographic biases found in scoring instruments of total hip arthroplasty. *J Arthroplasty* 1996;11:820-30.
18. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787-805.
19. Euroquality of life Group: Euroqol. A new facility for the measurement of health related quality of life. *Health Policy* 1990;16:199-208.
20. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF36). I. conceptual framework and item selection. *Med Care* 1992;30:473-83.
21. Franzen H, Johnsson R, Nilsson LT. Impaired quality of life 10 to 20 years after primary hip arthroplasty. *J Arthroplasty* 1997;12:21-4.
22. Garellick G, Herberts P, Strömberg C, Malchau H. Long-term results of Charnley arthroplasty: a 12-16 year follow-up study. *J Arthroplasty* 1994;9:333-40.

23. **Kavanagh BF, Wallrichs S, Dewitz M, et al.** Charnley low-friction arthroplasty of the hip: twenty year results with cement. *J Arthroplasty* 1994;9:229-34.
24. **Madey SM, Callaghan JJ, Olejniczak JP, Goetz DD, Johnston RC.** Charnley total hip arthroplasty with use of improved techniques of cementing: the results after a minimum of fifteen years of follow up. *J Bone Joint Surg [Am]* 1997;79-A:53-64.
25. **Malchau H, Herberts P, Ahnfelt L.** Prognosis of total hip replacement in Sweden. *Acta Orthop Scand* 1993;64:497-506.
26. **Nilsson LT, Franzén H, Carlsson AS, Onnerfalt R.** Early radiographic loosening impairs the function of a total hip replacement: the Nottingham Health Profile of 49 patients at five years. *J Bone Joint Surg [Br]* 1994;76-B:235-9.
27. **Carlsson AS, Gentz CF.** Radiographic versus clinical loosening of the acetabular component in noninfected total hip arthroplasty. *Clin Orthop* 1984;185:145-50.
28. **Devitt A, O'Sullivan T, Quinlan W.** 16- to 25-year follow-up study of cemented arthroplasty of the hip in patients aged 50 years or younger. *J Arthroplasty* 1997;12:479-89.
29. **Greenfield S.** The state of outcome research: are we on target? *N Engl J Med* 1989;320:1142-3.
30. **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.