Outcome after total hip arthroplasty

Part II. Disease-specific follow-up and the Swedish National Total Hip Arthroplasty Register

Peter Söderman, Henrik Malchau, Peter Herberts, Roland Zügner, Hans Regnér and Göran Garellick

Department of Orthopedics, Sahlgrenska University Hospital/Sahlgrenska, SE-413 45 Göteborg, Sweden.
Tel +46 31–342 18 39. Fax –82 55 99
Submitted 00-04-30. Accepted 00-09-21

ABSTRACT – The Swedish National Total Hip Arthroplasty Register records primary hip replacements, revisions and surgical technique/environmental factors. The end-point for failure is revision. A prosthesis still in place, however, does not mean success. Clinical and radiographic outcomes should describe in more detail the efficacy of hip replacement surgery instead of the relatively blunt outcome measure that the register can provide. We performed a clinical outcome analysis on patients with primary total hip replacement thus testing the adequacy of the end-point for failure in the Swedish register. 1,113 randomly selected patients who had had total hip replacement surgery between 1986 and 1995 answered a disease-specific self-administered questionnaire (WOMAC). A cohort of 344 patients was studied, using the Harris Hip Score and a conventional radiographic examination as outcome measures. We found clinical failure rates of 13% and 20% for all implants after 10 years, using 60 points or revision as the definition of failure in the Harris Hip Score and WOMAC, respectively. The result, according to the register during the same period, was a 7% revision rate. The clinical failure rate depended on the type of evaluation tool, definition of failure and demographics, which made it difficult to decide whether there was a need for revision. With the exception of pain measured by the Harris Hip Score, the results showed no significant correlation between clinical failure and radiographic failure. Hence, with the knowledge that there is a difference between the revision rate according to the register and clinical outcome, the strict definition of failure in the register is useful as an end-point for primary hip replacement surgery.

The efficacy of total hip arthroplasty surgery (THA) can be measured in several ways, such as the clinical outcome using disease-specific or generic tools, radiographic and economic evaluations and survival analyses. The rate of revision surgery is another important parameter of quality. Two registers in Sweden give information about nationwide revision surgery. First, all hospitals report the personal identity number and the diagnosis and type of operation to the Discharge Register (the Swedish National Board of Health & Welfare). Secondly, a more detailed register, the Swedish National Total Hip Arthroplasty Register (the Swedish THA register), documents primary hip replacement, revision and environmental factors (Herberts et al. 1989, Malchau et al. 1993, Herberts and Malchau 1997). The end-point for failure in the Swedish THA register is revision—that is, exchange or removal of one or both components. This failure end-point gives no information about the clinical or radiographic outcome or the patient's subjective satisfaction.

In this study we did an outcome analysis using disease-specific analyses and radiographic evaluation, and thus examined the clinical relevance of the strict end-point for failure in the Swedish THA register.

Methods

■ The study was approved by the Ethics Committee at Sahlgrenska University Hospital.

1,113 patients subjected to primary THA in Sweden between 1986 and 1995 were randomly selected, stratified by age and gender, from the Discharge Register. These patients were asked to fill in a short questionnaire about any reoperations, the presence of hip pain and overall satisfaction with the hip replacement. The patients also answered a disease-specific self-administered questionnaire (WOMAC). Two reminder letters were sent out to those who did not answer within 3 weeks. Patients who still did not reply were contacted by phone and asked to send in the questionnaire.

The mean ages at index surgery and follow-up were 69 (27–95) years and 75 (33–97) years, respectively. 54% were women. 86% were operated on for osteoarthritis, 3% for rheumatoid arthritis and 2% for complications after hip fracture. The commonest prostheses used in the clinical investigations were the Charnley (DePuy Olmed), the Lubinus SP II (Link) and the Scan Hip (MITAB) implants.

344 patients from the randomly selected original group (1,113 patients) were examined in 9 cities, which included regional, county and rural hospitals. One independent physician (PS) or an independent physiotherapist (RZ) examined this cohort clinically using the Harris Hip Score system. The patients were also examined by an independent observer (HR) using conventional radiographic techniques. All investigations were done within 3–4 months of one another to reduce the risk of changes in the patient's status, and thus allow comparison between the WOMAC, the Harris Hip Score, and the radiographic examination. With the exception of survival statistics, only unrevised patients were included in the clinical evaluation.

All patients were classified in different clinical categories according to Charnley—that is, patients with one affected hip (A), both hips affected (B), multiple joint disease or other disabilities impairing gait (C), and gender.

Clinical failure was arbitrarily estimated at 50, 60 or 70 points, using the Harris Hip Score and the WOMAC, so that patients who were revised or who scored below these values were considered failures.

Clinical outcome measurement

The Western Ontario and McMaster University Osteoarthritis Index (WOMAC) is a disease-specific and self-administered health measure developed to study patients with osteoarthrosis in the hip or knee. The scoring system utilizes pain (5 questions), stiffness (2 questions) and physical function (17 questions) domains (Bellamy et al. 1988, Bellamy et al. 1992, Sun et al. 1997). The 3 domains in the WOMAC can be analyzed separately or combined in a single score. Each question has 5 alternative answers with a score of 0–4 points. Maximum scores are 20 points for pain, 8 for stiffness and 68 points for physical function. The WOMAC questionnaire has been extensively tested for its validity and reliability, and it is sensitive enough to detect clinically important changes in health status after surgical interventions (responsiveness) (Bellamy et al. 1988, Bellamy et al. 1991, McGrory and Harris 1996, Martin et al. 1997). It has been translated into several languages. Roos et al. (1999) translated the WOMAC into Swedish and that version was found to be valid and reliable (Söderman and Malchau 2000a).

The Harris Hip Score is a disease-specific staff-administered scoring system which was introduced to provide an evaluation system for various hip disabilities and methods of treatment (Harris 1969). This Score gives a maximum of 100 points, with domains for pain, function, deformity and motion. Pain and function were the two basic considerations and received the heaviest weighting (44 and 47 points, respectively). Range of motion and deformity are seldom of primary importance, and therefore received 5 and 4 points, respectively. Function was subdivided into activity of daily living (ADL, 14 points) and gait (33 points). With respect to validity, reliability and responsiveness, we found only a few evaluations in the literature (Laupacis et al. 1993, Söderman and Malchau 2000b).

Radiographic method

Standard anteroposterior frontal, pelvis (centered over the symphysis) and true lateral radiographs were taken. The Hodgkinson criteria for loosening of the cup were used (Hodgkinson et al. 1988). Postoperative radiographs were not saved by all the hospitals and therefore migration of the cup

Table 1. Mean total score for the Harris Hip Score and the WOMAC 2–10 years postoperatively

Year	2	3	4	5	6	7	8	9	10	all
n	31	29	30	26	34	26	25	31	25	257
HHS										
pain	42	38	38	42	42	42	41	40	38	40
function	39	39	37	37	39	38	36	37	35	37
deformity	3.9	3.8	3.7	3.9	3.7	3.6	3.6	3.7	3.4	3.7
motion	4.9	4.9	4.9	4.9	4.9	4.8	4.8	4.9	4.8	4.9
total	90	86	84	87	90	88	86	86	81	87
WOMAC										
n	113	107	94	142	124	92	90	99	102	963
pain	16	16	16	16	16	16	15	15	15	16
stiffness	6.2	6.1	5.8	6.0	5.9	5.7	5.5	5.5	5.4	5.8
function	51	50	47	49	49	47	45	44	42	47
total	74	74	68	71	71	69	66	65	63	69

could not be measured since the patients were examined only once, 2–10 years postoperatively. Failure of the cup was classified as Hodgkinson type 3 with a 100% circumferential radiolucent line. The criteria for stem failure were debounding, stem fracture, cement fracture or a 100% circumferential radiolucent line (Garellick et al. 1999a, Mulroy and Harris 1997). The patients were divided into two groups: one with and one without radiographic failure. The radiographic results were compared to pain, function and total score from the clinical investigations (Harris Hip Score and WOMAC). The frequency of revision was noted for each type of hospital.

Statistics

For statistical analyses, the computerized statistical program SPSS (Chicago) was used. Total score, domain scores, mean, median, standard deviation (95% CI), minimum and maximum values (range) were calculated for all patients 2–10 years postoperatively. The Mann-Whitney U-test was used for statistical analysis. For the 10-year survival analysis, logistic regression analysis was used and the results were compared to the survival statistics in the Swedish THA register. To study whether the HHS and WOMAC fulfilled their purpose, the Spearman correlation was calculated between the same domains (convergent validity) and different (divergent validity) (Katz et al. 1996). Our hypothesis was that the same domains should show a greater correlation with one another than with different domains. The Cronbach alpha index was used to study the internal consistency reliability of the Harris Hip Score and WOMAC. To make the results comparable with other scoring systems, all domains and total scores were transformed to 0–100 point scales where 100 points meant the best health.

Results

The response rates were 93% and 84% for the postal survey (WOMAC) and for the clinical test (Harris Hip Score), respectively. The median total Score was 96 (39–100) for patients given the Charnley prosthesis (n 32), 92 (34–100) for the Lubinus SPII prosthesis (n 73), and 87 (35–100) for the Scan Hip (n 57). The mean total score for the Harris Hip Score and WOMAC declined during the follow-up (Table 1). The domain and total score of the Harris Hip Score were higher than the WOMAC scores (Table 1). The clinical study (Harris Hip Score) and the postal survey (WOMAC) showed significant differences between patients with one affected hip (Charnley A, total median Harris Hip Score 96 (37–100) points, and those with general disabling disease (Charnley C, total median Harris Hip Score 79 (34–98) points. The mean values are given in Table 2).

During the same period as this study, the Swedish THA register has shown a 93% survival rate on the national level after 10 years (Malchau et al. 1993). Patients who were revised or who scored lower than 60 points on the total score in the Har-

Table 2. Mean total score, standard deviation and mean transformed total score (T) for the Harris Hip Score and the WOMAC 2–11 years postoperatively. Charnley category A, B, C

	Charnley A			Charnley B			Charnley C		
HHS									
n	121			95			72		
pain	42	5.2	93	41	6.9	90	38	9.4	83
function	42	6.4	88	38	9.3	80	30	9.6	63
deformity	3.8	0.4	95	3.6	0.6	92	3.7	0.6	94
motion	4.9	0.1	99	4.9	0.1	98	4.8	0.2	97
total	92	10	92	87	13	87	76	16	76
WOMAC									
n	317			179			539		
pain	18	2.8	91	17	4.1	82	14	4.3	69
stiffness	6.8	1.4	85	5.9	1.8	74	5.2	1.9	65
function	58	11	86	51	14	74	40	16	58
total	84	14	88	73	20	77	59	21	64

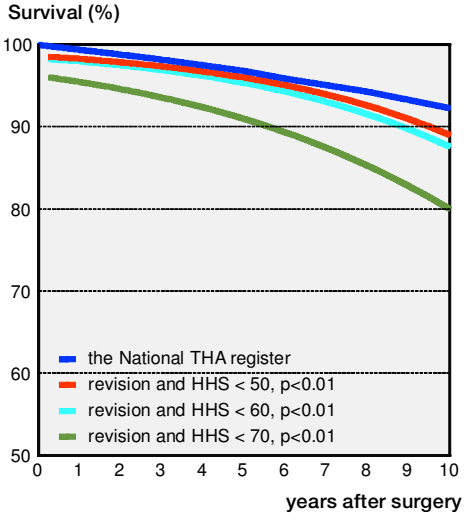

Figure 1. 10-year survival of all THA in the Swedish National Total Hip Arthroplasty Register (n 93,852) compared to the results of the clinical investigation evaluated with the Harris Hip Score (n 287).

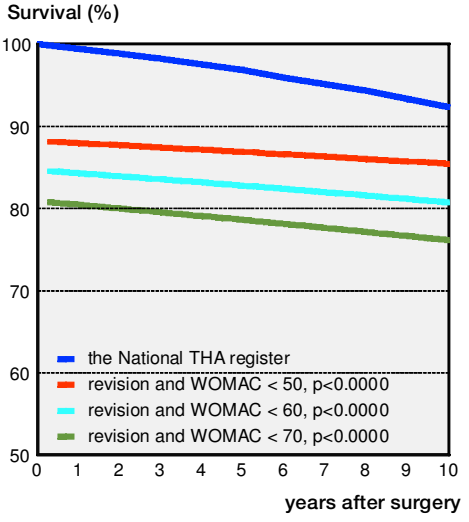

Figure 2. 10-year survival of all THA in the Swedish National Hip Arthroplasty Register (n 93,852) compared to the results of the clinical investigation with the WOMAC (n 1,000)

ris Hip Score had an 87% 10-year survival rate in this study (Figure 1). In the WOMAC, the corresponding result was 80% (Figure 2). When using the same definition of clinical failure, the Harris Hip Score generally showed higher scores than the WOMAC.

The correlation between equal domains in the Harris Hip Score and the WOMAC was significant ($p<0.01$). The pain domain in the Harris Hip Score correlated almost equally with pain in the WOMAC (Spearman's rho= 0.40) and function in the WOMAC (rho= 0.38). The same results were obtained when comparing the domains of function in the two scores with pain domains (Harris Hip Score function vs. WOMAC function, rho= 0.57, compared to Harris Hip Score function vs. WOMAC, rho= 0.38). The scoring system used showed high convergent and divergent construct validity. Good correlation was also found between items in the different domains in the Harris Hip Score and WOMAC (Cronbach alpha index 0.76–0.94).

Table 3. Correlation between clinical results (estimated with the HHS and the WOMAC for pain, function and total score) and radiographic failure (stem and cup). Mann-Whitney U-test for analysis of significance. The domains in HHS are transformed to a 0–100 point scale

Score	Failure	n	Median	Mean	SD	P-value
HHS						
pain	No	166	98	92	14	0.05
	Yes	27	98	86		
function	No	166	87	82	19	0.2
	Yes	26	82	78		
total	No	166	92	88	13	0.1
	Yes	26	90	84		
WOMAC						
pain	No	160	17	16	4	0.3
	Yes	27	16	15		
function	No	158	54	49	16	0.09
	Yes	26	44	44		
total	No	157	78	71	21	0.08
	Yes	26	67	65		

Radiographic follow-up

One hospital declined to help with the radiographic examinations, and patients who replied incompletely to the Harris Hip Score examinations and the WOMAC questionnaire were not included in the statistics for the radiographic results. 76% of the patients were examined using radiographic analysis. The total number of operations in each type of hospital was low; regional and rural hospitals each had a radiographic failure rate of 9% (5 failures of 53 primary prostheses in regional hospitals, and 7 failures of 79 primary operations in rural hospitals). The county hospitals had a 23% failure rate (15 failures of 65 primary operations). We also studied radiographic failure of the cup and/or the stem. With the exception of pain measured by the HHS, no significant difference was found (on the 5% level) between the failure and the non-failure groups in pain, function and total score in the two clinical studies (Table 3).

Discussion

We found that the results of the Swedish THA register are reliable. The clinical failures depend on demographics (comorbidity) and the scoring system and definition of clinical failure used.

Others have shown out a discrepancy between the failure rate, as described with disease-specific scoring systems, and the number of revised patients (Garellick et al. 1999b, Söderman et al. 2000). The clinical failure rates 10 years postoperatively in those studies and in ours were at least twice as high as in the register and this discrepancy is discussed below.

The first consideration using revision as an endpoint is the surgeon's willingness to revise the prosthesis. The decision to revise depends on several factors: age and comorbidity of the patient and the competence and resources of the surbeon. Furthermore, the set-up of the health care system may entail a long waiting-list, providing better results of the surgery because the operation is delayed. The second consideration is that the outcome findings in this study depended on the questionnaire used and the choice of clinical failure end-point. Both the Harris Hip Score and the WOMAC are valid and reproducible, but two problems using these scoring systems are the choice of level for clinical failure and demographic bias. The Harris Hip Score was developed and initially tested on young people with severe pelvic injury and below 70 points was considered a poor result (Harris 1969). Today, the Harris Hip Score is often used for outcome evaluation after THA in an older population, with less activity and higher comorbidity. According to Harris, the level for "poor results" should be less than 70 points. The level for clinical failure should also be nuanced with respect to various Charnley categories, such as patients with one hip affected and those with multiple joint disease (Garellick et al. 1999b). Separation into different Charnley categories for small and case-mixed studies could lead to difficulties in statistical analyses. However, it is important to state the number of patients in each Charnley category because patients with one affected hip have better clinical results than those with several joints affected (Callaghan et al. 1990, Brinker et al. 1996, Garellick et al. 1999b).

The response rates in this study were high, but in theory, non-responders could be those with the best clinical results. Most of these patients were phoned, however, and the non-responders were found to be a mixture of patients with good or bad results. Therefore, the results whould not be much

affected, if such patients were included.

Some authors have found little, if any, correlation between clinical and early radiographic failure (Gustilo and Pasternak 1988, Maloney et al. 1990, Wixson et al. 1991, Kwong et al. 1992, Boeree and Bannister 1993). Longitudinal radiographic examinations, however, are important for periprosthetic bone loss, which could lead to substantial bone loss with a high risk of fracture and difficulties in revision. The follow-up in the present study was short (2–10 years) and the patients were examined on only one occasion postoperatively. This led to difficulties in deciding whether an implant was loose, and the number of radiographic examinations that could be used for statistical analysis was too low to detect any level for clinical failure that could be correlated to radiographic failure. Patients with a very loose stem had more pain and less function than those with 100% radiolucent lines around the cup or those without severe radiographic complications, indicating that clinical scoring systems can detect failure in some hip replacements.

In summary, the failure end-point in the Swedish THA register is valid and very exact. The failure end-point for clinical outcome measurements should be more extensively evaluated to give reference material for studies of THA, such as validations of the register with survival analyses.

The authors thank the Volvo Research Foundation, the Swedish National Board of Health and Welfare, the Gothenburg Medical Association, the Dr Felix Neuberg Foundation, the Greta & Einar Asker Foundation and the Hjalmar Svensson Foundation for financial support.

Bellamy N, Buchanan W, Goldsmith C H, Campbell J, Stitt L W. 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 and the knee. J Rheumatol 1988; 15: 1833-40.

Bellamy N, Wells G, Campbell J. Relationship between severity and clinical importance of symptoms in osteoarthritis. Clin Rheumatol 1991; 10 (2): 138-43.

Bellamy N, Kean W F, Buchanan W W, Gerecz-Simon E, Campbell J. Double-blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): Postvalidation reapplication of the WOMAC osteoarthritis index. J Rheumatol 1992; 19 (1): 153-9.

Boeree N R, Bannister G C. Cemented total hip arthroplasty in patients younger than 50 years of age. Clin Orthop 1993; 287: 153-9.

Brinker M R, Lund P J, Cox D D, Barrack R L. Demographic biases found in scoring instruments of total hip arthroplasty. J Arthroplasty 1996; 11 (7): 820-30.

Callaghan J J, Dysart S H, Savory C F. Assessing the results of hip replacement. A comparison of five different rating systems. J Bone Joint Surg (Br) 1990; 72 (6): 1008-9.

Garellick G, Malchau H, Regnér H, Herberts P. The Charnley versus the Spectron Hip Prosthesis. Radiographic evaluation of a randomized, prospective study of 2 different hip implants. J Arthroplasty 1999a; 14 (4): 414-25.

Garellick G, Herberts P, Malchau H. The value of clinical data scoring systems. Are traditional hip scoring systems adequate to use in evaluation after total hip surgery? J Arthroplasty 1999b; 14 (8): 1024-9.

Gustilo R B, Pasternak H S. Revision total hip arthroplasty with titanium ingrowth prosthesis and bone grafting for failed cemented femoral component loosening. Clin Orthop 1988; 235: 111-9.

Harris W H. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. J Bone Joint Surg (Am) 1969; 51 (4): 737-55.

Herberts P, Malchau H. How outcome studies have changed total hip arthroplasty practices in Sweden. Clin Orthop 1997; 344: 44-60.

Herberts P, Ahnfelt L, Malchau H, Strömberg C, Andersson B J. Multicenter clinical trials and their value in assessing total joint arthroplasty. Clin Orthop 1989; 249: 48-55.

Hodgkinson J P, Shelly P, Wroblewski B M. The correlation between the roentgenographic appearance and operative findings at the bone-cement junction of the socket in Charnley low friction arthroplasties. Clin Orthop 1988; 228: 105-9.

Katz J N, Phillips C R, Poss R, Harrast J J, Fossel A H, Liang M H, Sledge C B. Correspondence. J Bone Joint Surg (Am) 1996; 78 (9): 1445-8.

Kwong L M, Jasty M, Mulroy R D, Maloney W J, Bragdon C, Harris W H. The histology of the radiolucent line. J Bone Joint Surg (Br) 1992; 74 (1): 67-73.

Laupacis A, Bourne R, Rorabeck C, Feeny D, Wong C, Tugwell P, Leslie K, Bullas R. The effect of elective total hip replacement on health-related quality of life. J Bone Joint Surg (Am) 1993; 75 (11): 1619-26.

Malchau H, Herberts P, Ahnfelt L. Prognosis of total hip replacement in Sweden. Follow-up of 92,675 operations performed 1978-1990. Acta Orthop Scand 1993; 64: 497-506.

Maloney W J, Jasty M, Rosenberg A, Harris W H. Bone lysis in well-fixed cemented femoral components. J Bone Joint Surg (Br) 1990; 72 (6): 966-70.

Martin D P, Engelberg R, Agel J, Swiontkowski F. Comparison of the musculosketal function assessment questionnaire with the Short form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile Health-Status Measures. J Bone Joint Surg (Am) 1997; 79 (9): 1323-35.

McGrory B J, Harris W H. Can the Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index be used to evaluate different hip joints in the same patients? J Arthroplasty 1996; 11 (7): 841-4.

Mulroy W F, Harris W H. Acetabular and femoral fixation 15 years after cemented total hip surgery. Clin Orthop 1997; 337: 118-28.

Roos E M, Klässbo M, Lohmander L S. WOMAC osteoarthritis index - reliability, validity, and responsiveness in patients with arthroscopically-assessed osteoarthritis. Scand J Reumatol 1999; 28: 210-215.

Sun Y, Sturmer T, Gunther K P, Brenner H. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee. A review of the literature. Clin Rheumatol 1997; 16 (2): 185-98.

Söderman P, Malchau H. Validity and reliability of Swedish WOMAC osteoarthritis index. A self-administered disease-specific questionnaire (WOMAC) versus generic instruments (SF-36 and NHP). Acta Orthop Scand 2000a; 71 (1): 39-46.

Söderman P, Malchau H. Is the Harris Hip Score system useful to study the outcome of total hip replacement? Accepted for publication in Clin Orthop, July. 2000b.

Söderman P, Malchau H, Herberts P, Johnell O. Are the findings in the Swedish National Total Hip Arthroplasty Register valid? A comparison between the Swedish National Total Hip Arthroplasty Register, the National Discharge register and the National Death register. J Arthroplasty 2000; 15 (7): 884-9.

Wixson R L, Stulberg S D, Mehlhoff M. Total hip replacement with cemented, uncemented and hybrid prostheses. A comparison of clinical and radiographic results at two to four years. J Bone Joint Surg (Am) 1991; 79 (2): 257-69.