Predictors of patient relevant outcome after total hip replacement for osteoarthritis
a prospective study

Nilsdotter, A-K; Petersson, I F; Roos, E M; Lohmander, L Stefan

Published in:
Annals of the Rheumatic Diseases

DOI:
10.1136/ard.62.10.923

2003

Citation for published version (APA):
Predictors of patient relevant outcome after total hip replacement for osteoarthritis: a prospective study

A-K Nilsdotter, I F Petersson, E M Roos, L S Lohmander

**PATIENTS AND METHODS**

**Patients**

Two hundred and nineteen patients (120 women) with a mean age at index surgery of 71 years (50–92) were consecutively included in the study. All patients had a primary unilateral THR performed because of primary OA between September 1995 and October 1998 at the department of orthopaedics in Halmstad, Sweden. Primary OA was defined as idiopathic OA in contrast to secondary OA caused by metabolic, anatomical, traumatic, and inflammatory conditions. Most (n=155) of the replacements were performed with both components cemented, while in 64 (29 women) the acetabular component was uncemented. The mean age of this subgroup at surgery was 62 years (50–72).

The cemented replacements were made using a cemented Lubinus acetabular component and a cemented Lubinus SP II femoral component (Link) and the hybrid replacements with an uncemented Trilogy or HGC acetabular component (Zimmer) and the cemented Lubinus SP II (Link) or Anatomic (Zimmer) femoral component.

Patients with the hybrid prosthesis were advised to partially bear weight for the first eight weeks after surgery. Surgical technique, cementing technique, rehabilitation, and follow up evaluation were otherwise identical for both groups.

Patients in need of bilateral surgery (n=14) were analysed separately at the final follow up. Patients who had recurrent dislocations of the prosthesis (n=1) during the first follow up year were excluded before analysis.

The preoperative hip radiographs were classified by two radiologists according to OARSI criteria with a radiographic atlas as a guide. OA severity was graded from 0 to 3 in accordance with the degree of joint space narrowing, where 3 indicates severe OA, 71% of the patients had severe radiological OA, 28% moderate OA, and 1% mild OA.

**Abbreviations:** OA, osteoarthritis; SF-36, Short Form-36; THR, total hip replacement; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index

---

**EXTENDED REPORT**

**Objectives:** To investigate prospectively long term patient relevant outcomes after unilateral total hip replacement (THR) for osteoarthritis (OA). To identify non-responders to this intervention and patient related predictors of unsatisfactory outcome.

**Methods:** A case-control study comparing health related quality of life of 219 patients (mean age 71) after THR with that of a matched reference group of 117 subjects without hip complaints recruited from the community. Patients and reference group answered SF-36 and WOMAC questionnaires preoperatively, at 3, 6, 12 months, and at 3.6 years (range 26–65 months) postoperatively. Supplementary questions were asked at the final follow up.

**Results:** 198/211 (94%) of the patients and 83/109 (76%) of the reference group participated at the final follow up. At follow up, the only difference between the two groups in the SF-36 was physical function, where patients scored worse. Patients also reported worse WOMAC function. 31% of the patients had improved by <10/100 WOMAC score points for pain and/or function at final follow up, compared with preoperatively. More pain preoperatively and higher age and postoperative low back pain predicted a worse outcome in WOMAC function.

**Conclusion:** 3.6 years after THR for OA, health related quality of life was similar for patients and reference group except for function, where patients had worse function. Higher age and more pain preoperatively predicted a poor outcome. Patients with hip OA with musculoskeletal comorbidities, such as low back pain and OA of the non-operated hip, have less long term functional improvement after THR.
Reference group
Two hundred and nineteen consecutive patients were included in the study. Eighty six of these were included from September 1997 to October 1998. During this period 258 subjects (three for each patient) were identified in the National Population Records. The subjects were matched to the patients by age, sex, and municipality. Short Form-36 (SF-36) and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaires were sent to these 258 subjects with an explanatory cover letter. In the cover letter they were told not to respond if they had hip complaints. Hip complaints were defined as pain or diminished range of motion in their hips. One hundred and seventeen (45%) answered the first inquiry. Their mean age was 72 years (range 52–92, 57% women, 43% men). One hundred and forty one subjects did not answer the first inquiry. Their mean age was 72 years (range 50–90, 52% women, 48% men). No reminder letters were sent as the number was regarded as sufficient for comparison of groups. Additional letters with the same questionnaires were then sent to these subjects at the same intervals as for the patients. At the follow up after 26–65 months eight subjects had died. 83 (76%) of the remaining 109 subjects responded to the mailing. Their mean age was 71 years (range 52–86, 55% women, 45% men). These 83 subjects were used for comparison with the patients at baseline as well as at follow up.

Design of the study
Self report with the patient administered questionnaires SF-36 and WOMAC was obtained preoperatively, at 3, 6, 12 months and at a final follow up at 3.6 years (26–65 months, mean 43 months, median 40 months) after the index THR surgery.

Questionnaires
The SF-36 measures three major health attributes (functional status, wellbeing, overall health) in eight subscales. These include PF (physical function), RP (role limitations due to physical health), BP (bodily pain), GH (general health), VT (vitality), SF (social function), RE (role limitations due to emotional health), and MH (mental health). The SF-36 scores are calculated on a 0–100 worst to best scale. SF-36 is translated and validated for Swedish conditions. It has previously been used in follow up studies of THR.

WOMAC (LK 3.0) was used as the disease-specific instrument. WOMAC is a self administered instrument validated for OA in the lower extremities and for evaluating outcome after THR. It consists of 24 items grouped into three categories: pain (five questions), stiffness (two questions), and physical function (seventeen questions). It is reliable and valid for Swedish conditions. To enhance the interpretation, WOMAC is transformed to a 0–100 worst to best scale. Because this instrument was not available and validated for Swedish conditions when the study was started, it was used at baseline for the last 92 patients only. There were no differences in age and sex between these 92 patients and the 106 who were included earlier. For cross sectional analysis of outcome at 3.6 years results from 198 patients were used and for longitudinal analysis 92 patients were used.

Supplementary questionnaire
Questions about postoperative complications, preoperative and postoperative comorbidity, social circumstances, and patient satisfaction were asked at the final follow up. The patients and the reference group received the same questions except those concerning postoperative complications and patient satisfaction.

Postoperative complications
Three questions were asked about serious postoperative complications: dislocation of the prosthesis, deep infection in the hip joint, and reoperation. The self reported data were compared with data from the patients’ case records. The definition of postoperative complication in this study referred to a positive answer to one of the three questions about postoperative complications.

General comorbidity
Fourteen questions were asked about intercurrent diseases preoperatively and in the present situation. Questions were asked about the presence of 12 comorbid conditions or body areas with problems (heart, hypertension, peripheral arteries, lung, diabetes, neurological problems, cancer, ulcer, kidney disease, vision, back pain, and psychiatric disease). The questions were multiple choice (yes, no, don’t know). The total number of conditions or problems reported was used as a summary variable (0, 1, 2, or more), a method shown to be valid in this kind of follow up.

Musculoskeletal comorbidity
Two questions were asked about the need of walking assistance and walking distance, preoperatively and in the present situation, two questions were asked about the need for analgesics due to pain in the operated hip joint or due to pain elsewhere. One question was asked about the experience of regional or widespread pain lasting more than three months during the past 12 months. One question was asked about joint replacement in the contralateral hip or in the knees since the THR. The final questions concerned fractures in the spine, wrist, hip, or elsewhere.

Social circumstances
One question was asked about the living circumstances preoperatively and in the present situation. One question was asked about the civil status and one about the main profession and the present profession or occupation.

Patient satisfaction
One question was asked about how satisfied the patients were with the result of the operation. The question was: “Overall, how satisfied are you with the result of your hip replacement surgery?” The alternative answers were: very satisfied, satisfied, dissatisfaction, very dissatisfaction.

Test-retest reliability
One hundred and ninety eight patients received questions about comorbidity at the final follow up. Of those, 60 patients selected at random received the same questionnaire three months later; 54 (90%) answered. In the analysis of the answers the answer “don’t know” was interpreted as “no”. The measurement of agreement, k, value, was 0.609–1.000.

Validity
The questions used in the questionnaire about comorbidity have in previous studies been tested for validity.

Non-responders
Non-responders were identified according to three different sets of criteria. Firstly, as the patients who scored worst (the lowest quartile) in WOMAC function at follow up. Secondly, as the patients with the least absolute improvement, defined as improvement of <20/100 score units in WOMAC function between the preoperative value and that at follow up at 3.6 years. Thirdly, criteria have been suggested for the identification of non-responders and responders in a clinical trial of drug treatment for hip OA, the OARSI set of criteria. These criteria are based on pain, function, and the patient’s global assessment. High improvement in pain is defined as an absolute change of 30/100 score units and a relative change of 50%. High improvement in function is defined as an absolute change of 20/100 score units and a relative change of 50%. A
paring the frequency of comorbidities in subgroups the
ationnaire data Wilcoxon’s signed rank test was used. For com-
comparison between groups, the Mann-Whitney test was
Statistical analysis was done with the SPSS 10.0 package. For
Statistics
year (for age) or 0–100 scale unit increase.
this regression analysis. Odds ratios were expressed for a one
circumstances were included in a stepwise multivariate logis-
ence of type of pain, preoperative scores of SF-36 BP (bodily
was used. Age, sex, body mass index, number of comorbid
conditions, occurrence of postoperative complications, pres-
ence of type of pain, preoperative scores of SF-36 BP (bodily
pain), PF (physical function), MH (mental health), and social
circumstances were included in a stepwise multivariate logistic
regression analysis. Odds ratios were expressed for a one
year (for age) or 0–100 scale unit increase.

RESULTS
Of the 219 patients, eight died during the follow up period and
13 did not participate. Thus the results for 198 patients (94%,
106 women), with a mean age at the time of surgery of 71
years, are presented for an average follow up time of 3.6
years, range 26–65 months, median 40, mean 43 (tables 1 and 2).
The patients, evaluated as a group, almost doubled their abso-
lute scores in all WOMAC subscales (p<0.0001) at the final
follow up compared with the preoperative scores (table 1).
They also reported improved scores at the final follow up in all
SF-36 subscales (p<0.0001) except GH (general health)
(p=0.2) (table 2).
Twenty three per cent of the patients reported a WOMAC
function score <30 and 57% a score <40 at the final
follow up.

Individual change in pain and function
The individual change in pain and function measured by
WOMAC was calculated as the difference between preopera-
tive and 3.6 years postoperative score (fig 1). The cut off level
for a significant improvement was set at 10/100 units,
previously defined as the smallest detectable clinical
improvement. At 3.6 years after surgery 31% of the patients
had improved by ≤10 score units for pain and/or function
compared with their value before surgery. Thus, 20/92 patients
had improved by ≤10 score units for pain only, eight patients
by <10 score units for function only, and one patient by <10
units for both pain and function. Twenty of 92 patients
improved by <20/100 score units for pain only and 15 patients
by <20 units for function only, while 5 improved by <20 units

Statistics
Statistical analysis was done with the SPSS 10.0 package. For
comparison between groups, the Mann-Whitney test was used. For comparison preoperative and postoperative ques-
tionnaire data Wilcoxon’s signed rank test was used. For com-
paring the frequency of comorbidities in subgroups the χ² test
was used. Age, sex, body mass index, number of comorbid
conditions, occurrence of postoperative complications, pres-
ence of type of pain, preoperative scores of SF-36 BP (bodily
pain), PF (physical function), MH (mental health), and social
circumstances were included in a stepwise multivariate logistic
regression analysis. Odds ratios were expressed for a one
year (for age) or 0–100 scale unit increase.

Table 1
Preoperative and 3.6 year postoperative mean scores and (standard
deviations) for the three WOMAC subscales, for the patients, mean age 71 (50–88),
and for the reference group, mean age 71 (52–86). The scale is 0–100, worst to
best.

Table 2
Preoperative and 3.6 year postoperative mean scores and (standard
deviations) for the SF-36 subscales, for the patients, mean age 71 (50–88), and
the reference group, mean age 71 (52–86). The scale is 0–100, worst to
best.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Patients</th>
<th>Patients</th>
<th>Reference group at</th>
<th>Reference group at</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>preoperative</td>
<td>3.6 years</td>
<td>baseline</td>
<td>3.6 years</td>
</tr>
<tr>
<td></td>
<td>(n=198)</td>
<td>postoperative</td>
<td>(n=83)</td>
<td>postoperative</td>
</tr>
<tr>
<td>WOMAC pain</td>
<td>45 (17.7)</td>
<td>88 (17.7)</td>
<td>87 (19.8)</td>
<td></td>
</tr>
<tr>
<td>WOMAC stiffness</td>
<td>39 (18.6)</td>
<td>88 (18.6)</td>
<td>84 (20.6)</td>
<td></td>
</tr>
<tr>
<td>WOMAC function</td>
<td>38 (21.7)</td>
<td>87 (18.9)</td>
<td>84 (21.8)</td>
<td></td>
</tr>
<tr>
<td>SF-36 PF</td>
<td>30 (19.6)</td>
<td>60 (25.1)</td>
<td>75 (20.8)</td>
<td>71 (26.7)</td>
</tr>
<tr>
<td>SF-36 BP</td>
<td>9 (21.4)</td>
<td>48 (44.0)</td>
<td>65 (42.3)</td>
<td>58 (44.8)</td>
</tr>
<tr>
<td>SF-36 GH</td>
<td>30 (16.8)</td>
<td>66 (26.2)</td>
<td>69 (27.9)</td>
<td>68 (27.6)</td>
</tr>
<tr>
<td>SF-36 VTI</td>
<td>68 (19.9)</td>
<td>66 (22.0)</td>
<td>66 (24.4)</td>
<td>62 (24.8)</td>
</tr>
<tr>
<td>SF-36 SF</td>
<td>49 (20.9)</td>
<td>64 (24.3)</td>
<td>65 (25.8)</td>
<td>65 (26.4)</td>
</tr>
<tr>
<td>SF-36 RE</td>
<td>64 (26.2)</td>
<td>84 (22.6)</td>
<td>85 (22.2)</td>
<td>80 (27.2)</td>
</tr>
<tr>
<td>SF-36 MH</td>
<td>70 (21.0)</td>
<td>78 (20.1)</td>
<td>82 (19.0)</td>
<td>79 (19.4)</td>
</tr>
</tbody>
</table>
Figure 1 Preoperative (baseline) and 3.6 year postoperative THR scores of WOMAC function for OA. The scale is 0–100, worst to best. Each line represents one patient (n=92). The blue line with triangle symbols represents the average scores for the patient group. The red line with circle symbols represents average scores for WOMAC function for the reference group (n=83).

for both pain and function. Among those subjects who had improved by <20 units for pain or function at 3.6 years, 6 had also improved by <20 units at the one year follow up.

Patients who received bilateral THR
The patients who received bilateral THR during the first follow up year were examined at the 3.6 year follow up. A comparison showed no difference in the SF-36, between the patients who had received bilateral THR and those with unilateral THR. However, there was a difference in the WOMAC subscale pain, with the patients who received bilateral THR scoring better (p=0.008) than the patients who received unilateral THR (table 3).

Comparisons between patients and reference group
Pain and physical function
At 3.6 years after THR the patients on average reported worse WOMAC pain (82 ± 87, p=0.006) and worse WOMAC function (74 ± 84, p<0.0001) than the reference group. There were no differences in the SF-36 subscales between the two groups except PF (physical function) where the patients reported worse function (60 ± 71, p<0.0001) than the reference group (tables 1 and 2).

Frequency of comorbidities
Before THR the patients, in comparison with the reference group, reported a higher prevalence of low back pain (patients 32% vs reference 15%, p=0.001), but a lower prevalence of pulmonary disease (patients 1% vs reference 7%, p=0.005). No other differences between the patient and reference groups before surgery were found.

<table>
<thead>
<tr>
<th>Comorbid condition</th>
<th>Patients</th>
<th>Controls</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart</td>
<td>9.2</td>
<td>10.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Hypertension</td>
<td>27.8</td>
<td>26.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Peripheral arteries</td>
<td>12.7</td>
<td>13.3</td>
<td>0.4</td>
</tr>
<tr>
<td>Lung</td>
<td>0.5</td>
<td>7.2</td>
<td>0.004*</td>
</tr>
<tr>
<td>Diabetes</td>
<td>7.1</td>
<td>6.0</td>
<td>0.8</td>
</tr>
<tr>
<td>Neurological disease</td>
<td>2.0</td>
<td>3.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Cancer</td>
<td>5.6</td>
<td>3.6</td>
<td>0.5</td>
</tr>
<tr>
<td>Peptic ulcer</td>
<td>1.5</td>
<td>2.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Kidney disease</td>
<td>0</td>
<td>1.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Vision problems</td>
<td>1.5</td>
<td>2.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Low back pain</td>
<td>21.5</td>
<td>16.9</td>
<td>0.2</td>
</tr>
<tr>
<td>Psychiatric disease</td>
<td>2.6</td>
<td>2.4</td>
<td>0.3</td>
</tr>
<tr>
<td>Unilateral hip pain</td>
<td>20.2</td>
<td>4.8</td>
<td>0.001*</td>
</tr>
<tr>
<td>Regional pain</td>
<td>59.1</td>
<td>43.9</td>
<td>0.002*</td>
</tr>
<tr>
<td>Widespread pain</td>
<td>6.6</td>
<td>7.3</td>
<td>0.8</td>
</tr>
</tbody>
</table>

At the final follow up after THR there was no difference between the two groups in the prevalence of low back pain or in widespread pain. The prevalence of patients’ low back pain had decreased from 32% to 21.5%. However, at this time patients had more regional pain (patients 59%, reference 44%, p=0.002) and unilateral hip pain (patients 20%, reference 5%, p=0.001). The prevalence of pulmonary disease was still lower among the patients than the reference group (0.5% and 7%, respectively, p=0.004) (table 4). At the 3.6 year follow up 33% of the patients used walking assistance, while the corresponding proportion for the reference group was 20% (p=0.03). Consistent with this, 42% of the patients and 57% of the reference group reported an unlimited walking distance, respectively (p=0.03).

Non-responders
As noted in the “Patients and methods” section, non-responders were identified in three different ways. Firstly, the lowest quartile in WOMAC function at the final follow up (25% are defined as non-responders). Secondly, an absolute improvement between baseline and 3.6 years follow up in WOMAC function of <20 score units (22% are defined as non-responders). Thirdly, by the OARSI criteria (9% are defined as non-responders) (table 5). Figure 3 illustrates the overlap between these three groups. There was no difference in age and sex between these three subgroups. We have chosen to report the results based on the first definition (lowest quartile) (see “Discussion”).

Preoperative predictors for non-responders
The preoperative data for the patients who scored worst (the lowest quartile, mean score 44.8 (SD 11.4, range 15.6–61.8)) in

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Comparison between outcomes at 3.6 years after unilateral or bilateral THR. Results are shown as mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WOMAC subscale</td>
<td>Unilateral THR (n=198)</td>
</tr>
<tr>
<td>Pain</td>
<td>82 (20)</td>
</tr>
<tr>
<td>Stiffness</td>
<td>78 (22)</td>
</tr>
<tr>
<td>Physical function</td>
<td>74 (21)</td>
</tr>
<tr>
<td>Unilateral THR, mean age 71, bilateral THR mean age 74. The bilateral surgery was made during the first follow up year. The scale is 0–100, worst to best. *p= 0.008.</td>
<td></td>
</tr>
<tr>
<td>Table 4</td>
<td>Comparison of the prevalence of comorbid conditions between patients (n=198) and controls (n=83) at the 3.6 year follow up analysed with a χ² test. Results are shown as percentages</td>
</tr>
<tr>
<td>Comorbid condition</td>
<td>Patients</td>
</tr>
<tr>
<td>Heart</td>
<td>9.2</td>
</tr>
<tr>
<td>Hypertension</td>
<td>27.8</td>
</tr>
<tr>
<td>Peripheral arteries</td>
<td>12.7</td>
</tr>
<tr>
<td>Lung</td>
<td>0.5</td>
</tr>
<tr>
<td>Diabetes</td>
<td>7.1</td>
</tr>
<tr>
<td>Neurological disease</td>
<td>2.0</td>
</tr>
<tr>
<td>Cancer</td>
<td>5.6</td>
</tr>
<tr>
<td>Peptic ulcer</td>
<td>1.5</td>
</tr>
<tr>
<td>Kidney disease</td>
<td>0</td>
</tr>
<tr>
<td>Vision problems</td>
<td>1.5</td>
</tr>
<tr>
<td>Low back pain</td>
<td>21.5</td>
</tr>
<tr>
<td>Psychiatric disease</td>
<td>2.6</td>
</tr>
<tr>
<td>Unilateral hip pain</td>
<td>20.2</td>
</tr>
<tr>
<td>Regional pain</td>
<td>59.1</td>
</tr>
<tr>
<td>Widespread pain</td>
<td>6.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Stratified baseline data of WOMAC function and SF-36 PF and the percentage of patients with an improved or worse score at the final follow up</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;20 Score change</td>
<td>&lt;20 Score change</td>
</tr>
<tr>
<td>Baseline WOMAC function (n=92)</td>
<td>83</td>
</tr>
<tr>
<td>0–29.9 (25%)</td>
<td>78</td>
</tr>
<tr>
<td>30–59.9 (66%)</td>
<td>62</td>
</tr>
<tr>
<td>60–100 (9%)</td>
<td>27</td>
</tr>
<tr>
<td>Baseline SF-36 PF (n=198)</td>
<td>73</td>
</tr>
<tr>
<td>0–29.9 (51%)</td>
<td>70</td>
</tr>
<tr>
<td>30–59.9 (38%)</td>
<td>29</td>
</tr>
</tbody>
</table>
the WOMAC subscale physical function at 3.6 years after THR were compared in a multivariate analysis with the preoperative data for the three quartiles of patients who scored better. A higher degree of pain, as measured by the SF-36 preoperatively, and higher age predicted worse WOMAC function outcome at 3.6 years after surgery (table 6). It should be noted that the odds ratios are expressed for one year or scale unit difference. The number of preoperative comorbid conditions did not predict worse WOMAC function outcome.

In a univariate analysis, a high body mass index and worse physical function preoperatively as measured by SF-36 were predictors for worse outcome in WOMAC function at 3.6 years (table 6).

**Postoperative characteristics of patients with poor function**

Univariate analyses were made to compare the postoperative data at the final follow up for the patients who scored worst (lowest quartile) in the WOMAC subscale physical function with the patients who scored better. Low back pain and postoperative complications were associated with worse outcome. In a multivariate analysis, low back pain postoperatively was the only significant characteristic of the patients with a non-successful result (table 6).

**DISCUSSION**

This is, to our knowledge, the first prospective long term follow up study of pain, function, and health related quality of life in patients with THR for OA using validated patient relevant outcome measures, including a comparison with a matched reference group. In this report we focus on the patient relevant outcome 3.6 years after unilateral THR for OA, and attempt to identify any predictors of poor outcome.

Treatment outcome can be assessed in different ways. Improvement, not just the final outcome score, is an important measure for patients receiving a THR. In our study the WOMAC pain and/or function score of almost one third of the patients did not improve by more than 10/100 units during the 3.6 year follow up period (figs 1 and 2). Surgeons should thus carefully consider other options before operating on patients with less preoperative pain or function scores because these patients will improve only slightly.

**Methodological issues**

We used both a generic and a disease-specific instrument because SF-36 shows a better gradient with comorbidities than the WOMAC on all three dimensions, while the WOMAC scores provide a better gradient by the current physical condition. This is the reason for using SF-36 in the comparisons between the patients and the reference group and WOMAC when trying to identify the predictors of poor outcome in hip OA. Figures 1 and 2 illustrate this difference between the WOMAC and SF-36. In fig 1, related to WOMAC, it is clear that the intervention for almost all patients has an effect on physical function (83/92), while this is more doubtful in fig 2, related to SF-36 (145/184). For the SF-36 subscale physical function (fig 2), we would expect factors other than the operated hip to influence outcome to a higher degree than for WOMAC function (fig 1).

As a consequence there is a more obvious ceiling effect when using the WOMAC as outcome measurement after THR than with the SF-36 (figs 1 and 2).

Several reasons motivated the identification and use in this study of a reference group matched for age, sex, and municipality. Firstly, patients compare themselves, not with other disabled people, but with how they were before the onset of symptoms. Because such information is not available in a prospective study of the present kind, we chose instead to examine a matched reference group and to use these data as a surrogate. Secondly, a lack of normative data for WOMAC...
necessitated generating a matched reference group. This reference group also enabled us to gather data for SF-36, which complemented the available normative Swedish SF-36 data.\textsuperscript{10} The reference group was examined at the same times, at the same frequency, and with the same questionnaires as the THR study group. Hip problems are common in the community.\textsuperscript{16–20,22} However, subjects with hip problems were asked not to return the questionnaire. Thus, the reference data used here should represent a “best” case.

**Complications**

Postoperative complications such as recurrent dislocations, deep infection, and revision surgery were as frequent as expected from Swedish national average data for THR (http://www.jru.orthop.gu.se), which suggests that the study cohort was representative in this respect. Consequently, the number of complications in relation to the number of patients was few, which makes it difficult to draw any conclusions about postoperative complications. However, patients with treated postoperative complications in this study seemed to attain a similar level of function at 3.6 years after THR as the patients without complications (data not shown). This is noteworthy, because in most follow up studies implant failure and revision are the end points for follow up.\textsuperscript{24}

**Non-responders**

Non-responders were identified in three different ways. The first definition identified non-responders as the patients who scored worst (lowest quartile) in WOMAC function at the 3.6 year follow up. The second definition identified the patients with an absolute change of <20 score units on a 0–100 scale. The reason for the choice of an absolute change of 20 score units is the knowledge that in clinical trials of rehabilitation intervention and medical treatment the smallest detectable clinical improvement in WOMAC function and pain is 9–12 score units.\textsuperscript{30–33} Knowing that the responsiveness of medical treatment in OA is about half as high (standardised response mean 0.5–0.8) as for causal surgical treatment (standardised response mean 0.8–3.1),\textsuperscript{32–34} we decided to double the smallest detectable clinical improvement for medical treatment. This cut off point also concurs with the OARSI definition of high improvement in function.\textsuperscript{39}

The third definition was an attempt to use an established definition of non-responders that takes into account both pain, function, and patient global assessment. The problem with using this definition in this study is that it was developed for clinical trials in pharmacological treatment of hip OA, and not for surgical interventions. When these criteria were used only 8/92 (9%) patients were defined as non-responders at the 3.6 year follow up. Of those eight subjects, two reported that they were very satisfied, one satisfied, one dissatisfied with the outcome of the THR (four were missing values). Thus, even if the patient reports a bad outcome in pain and function she/he may be satisfied with the result.\textsuperscript{39} Only 4% of all the patients reported that they were dissatisfied with the result of the THR. This is a higher degree of satisfaction after THR than that reported for total knee replacement in OA.\textsuperscript{40} It has been stated that satisfaction is a wide concept, not necessarily relevant in outcome after THR.\textsuperscript{37,38}

When comparing the three methods of identifying non-responders we found that the first criterion identified the greatest number of patients with a poor outcome in WOMAC function at 3.6 years. However, the overlap with the two other definitions was considerable. Because the two other methods of identifying non-responders to a great extent identified the same patients, the same predictors were identified when using these criteria (fig 3).

**Determinants of poor outcome**

In the present study old age predicted a poor postoperative outcome, compared with the younger patients. This finding is consistent with a report that older people with self reported conditions restricting mobility in addition to arthritic pain in the hip or knee are at higher risk of psychological distress and physical dysfunction.\textsuperscript{47}

We found that the number of comorbid conditions preoperatively did not predict a worse outcome postoperatively when measured by the WOMAC function. The same result was found when using the SF-36 PF (physical function) as outcome measure, even though it is expected to show a better gradient with comorbidities.

There was no difference between the patient group and the reference group in the number of comorbidities at the beginning of the study or at the follow up. This is consistent with previous observations which showed that OA is not predictive for the development of future comorbidities.\textsuperscript{31,40}

Our results indicate that low back pain and pain in the hip not operated on is characteristic for patients who do not reach the same level of function postoperatively as the matched reference group (table 4). These findings are in concordance with a recently published study,\textsuperscript{40} and indicate that WOMAC captures not only knee or hip pain and dysfunction but is influenced by the presence of low back pain.\textsuperscript{39} The decreased prevalence of reported low back pain after THR seen in the patient cohort could in part be explained by a changed postoperative pain threshold,\textsuperscript{44} and in part by an improved pain-free range of motion in the operated hip. All patients were offered physical therapy postoperatively and almost everyone accepted.

The general health status of people with pain in the hip or knee is comparable to that of a reference group without such pain,\textsuperscript{45} but the health status is worse when pain in the hip or knee occurs in combination with other mobility restricting conditions—for example, pain in other joints and other musculoskeletal problems such as back pain. Pain in the hip not operated on may be a symptom of bilateral OA in these patients,\textsuperscript{39} even in the absence of radiological change.

**Limitations**

A limitation of the study is the variable follow up time (26–65 months). Thus, the patients with the longest follow up time have reached an older age than the patients with a shorter follow up time. On the other hand, these patients have had a longer time for rehabilitation and recovery. However, there was no correlation between follow up time and function.

The 14 patients who had been operated on the contralateral side during the first follow up year were analysed separately at the final follow up. There was no difference in outcome for

Figure 3: Non-responders to THR for OA identified in three different ways: (a) Defined as patients who scored worst (lowest quartile) in WOMAC function at 3.6 year follow up, (circle A, 23/92 [25%]); (b) defined as patients who reported an absolute improvement from baseline to 3.6 year follow up of <20 score units (circle B, 20/92 [22%]); (c) defined according to OARSI criteria (circle C, 8/92 [9%]). The patients included in this figure are based the longitudinal data (n = 92).
these patients, compared with the main group with unilateral THR, except that they had less pain (as measured by the WOMAC pain subscale). Pain in the non-operated hip was also more common among patients with unilateral THR than controls, suggesting the presence of bilateral OA.

CONCLUSIONS

In this prospective cohort study we have shown that patients with hip OA over 50 years of age at 3.6 years after unilateral THR report similar pain but have a lower level of physical function than a matched reference group without hip complaints. The difference in function is explained, at least in part, by the presence of musculoskeletal comorbidities such as low back pain and pain in the non-operated hip. Most patients have a significant improvement in pain and function after THR. However, at 3.6 years after surgery almost one third of the patients report only a low degree of improvement or have worsened after the THR. Some of this lack of response may be explained by old age and a high degree of pain at the time of surgery. This is supported by previous reports that have suggested surgery earlier rather than later in the course of OA. A report of low back pain has a significant impact on postoperative function after THR, which is of importance when planning rehabilitation.

A major proportion of the variability in outcome after THR for OA remains unexplained. Further patient relevant outcome data collected by the use of self administered questionnaires, and a better understanding of the role of expectations and satisfaction is essential to improve our understanding of who does or does not benefit from these surgical procedures. Much effort is focused on the “technical” aspects of this intervention to improve its success rate further. Perhaps we stand to gain as much or more from a better understanding of these “other” factors.

ACKNOWLEDGEMENTS

The authors acknowledge Birgit Ljungquist, PhD for excellent assistance with the statistical work. Financial support was received from the Scientific Council, Province of Halland, Council for Medical Health Research in South Sweden, Swedish Medical Research Council, Swedish Rheumatism Association, Lund University Hospital and Medical Faculty, Thelma Zoega Foundation, and the King Gustaf 80-Year Fund.

Authors’ affiliations

A-K Nilsdotter, I F Petersson, EM Roos, LS Lohmander, Department of Orthopaedics, Lund University Hospital, Sjöbom Hospital of Rheumatic Diseases, Halmstad, Sweden

REFERENCES


