LUND UNIVERSITY
Faculty of Medicine

LUP
Lund University Publications
Institutional Repository of Lund University

<u>**TITLE PAGE**</u>

The role of pain and functional impairment in the decision to recommend total joint replacement in hip and knee osteoarthritis: an international cross-sectional study of 1909 patients. Report of the OARSI-OMERACT Task Force on total joint replacement

Laure Gossec, Simon Paternotte, Jean Francis Maillefert, Christophe Combescure, Philip G. Conaghan, Aileen M. Davis, Klaus-Peter Gunther, Gillian Hawker, Marc Hochberg, Jeffrey N. Katz, Margreet Kloppenburg, Keith Lim, L. Stefan Lohmander, Nizar N. Mahomed, Lyn March, Karel Pavelka, Leonardo Punzi, Ewa M. Roos, Lydia Sanchez-Riera, Jasvinder A. Singh, Maria E. Suarez-Almazor, Maxime Dougados*, For the OARSI-OMERACT task force "total articular replacement as outcome measure in OA"*

AUTHORS AND INSTITUTIONS

Laure Gossec, MD, PhD, Simon Paternotte, MSc, Maxime Dougados, MD: Paris Descartes University, Medicine Faculty ; APHP, Rheumatology B Department, Cochin Hospital, Paris France.

Jean Francis Maillefert, MD, PhD, Department of Rheumatology, Dijon University Hospital, Dijon, F-21078, France, University of Burgundy, Dijon, F-21079, France, INSERM U887, Dijon, F-21079, France

Christophe Combescure, MSc PhD, University of Geneva, Switzerland

Philip G Conaghan, MD, PhD, Section of Musculoskeletal Disease, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, United Kingdom

Aileen M. Davis, PhD, Division of Health Care and Outcomes Research and Arthritis and Community Research and Evaluation Unit, Toronto Western Research Institute, Toronto, Ontario and Departments of Rehabilitation Science and Health Policy, Management and Evaluation, University of Toronto

Klaus-Peter Gunther, MD, Department of Orthopaedic Surgery, University of Dresden, Germany

Gillian Hawker, MD, PhD, Division of Rheumatology, Department of Medicine, Women's College Hospital; Clinical Epidemiology and Health Care Research Program, Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Canada

Marc Hochberg, MD, MPH, Division of Rheumatology & Clinical Immunology, Department of Medicine and Division of Gerontology, Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA

Jeffrey N. Katz, MD, MSc, Department of Orthopedic Surgery and Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School; and Department of Epidemiology, Harvard School of Public Health, Boston, MA USA.

Margreet Kloppenburg, MD, PhD, Department of Rheumatology, Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, Netherlands

Keith Lim, MD, FRACP, Department of Rheumatology, St Vincent's Hospital & Western Hospital, Melbourne, Australia

L. Stefan Lohmander, MD, PhD, Department of Orthopaedics, Clinical Sciences Lund, Lund University, Lund, Sweden

Nizar N. Mahomed, MD, ScD FRCSC, Division of Orthopaedic Surgery, University of

Toronto, Toronto Western Hospital, Toronto, Canada

Lyn March, MD, MSc, PhD, FRACP, FAFPHM Institute of Bone and Joint Research, University of Sydney, Royal North Shore Hospital, St Leonards, Australia

Karel Pavelka, MD, Ph.D, Institute of Rheumatology, Charles University Prague, Czech Republic

Leonardo Punzi, MD, PhD, Rheumatology Unit, Department of Clinical and Experimental Medicine, University of Padova, Italy

Ewa M. Roos, PT PhD, Institute of Sports and Clinical Biomechanics, University of Southern Denmark, Denmark

Lydia Sanchez-Riera, MD, Institute of Bone and Joint Research, University of Sydney, Australia.

Jasvinder A. Singh, MD, MPH, Birmingham VA Medical Center and University of Alabama, Birmingham, AL, USA

Maria E. Suarez-Almazor, MD, PhD, Section of Rheumatology, University of Texas, M.D. Anderson Cancer Center, Houston, Texas USA

CORRESPONDENCE

Dr. Laure Gossec, Service de Rhumatologie B, Hôpital Cochin, 27, rue du faubourg Saint-Jacques, 75014 PARIS.

Telephone: 33-158412606

Fax: 33-143549256

Email: laure.gossec@cch.aphp.fr

RUNNING HEAD   Pain and functional limitation in total joint replacement indication.

ABSTRACT (254 words)

Objective:

To assess the pain and functional disability levels corresponding to an indication for total joint replacement (TJR) in hip and knee osteoarthritis (OA).

Methods:

Design: international cross-sectional study in 10 countries. *Patients:* consecutive outpatients with definite hip or knee OA attending an orthopaedic outpatient clinic. *Gold standard measure for recommendation for TJR*: surgeon's decision that TJR is justified. *Outcome measures:* pain (ICOAP: intermittent and constant osteoarthritis pain, 0-100) and functional impairment (HOOS-PS/KOOS-PS: Hip/ Knee injury and Osteoarthritis Outcome Score Physical function Short-form, 0-100). *Analyses*: Comparison of patients with versus without surgeons' indication for TJR. ROC curve analyses and logistic regression were applied to determine cut-points of pain and disability defining recommendation for TJR.

Results:

In all, 1909 patients were included (1130 knee / 779 hip OA). Mean age was 66.4 (SD 10.9) years, 58.1% were women; 628/1130 (55.6%) knee OA and 574/779 (73.7%) hip OA patients were recommended for TJR. Although patients recommended for TJR (yes versus no) had worse symptom levels (pain, 55.5 [95% confidence interval 54.2, 56.8] vs. 44.9 [43.2, 46.6], and functional impairment, 59.8 [58.7, 60.9] vs. 50.9 [49.3, 52.4], respectively, both p<0.0001), there was substantial overlap in symptom levels between groups, even when adjusting for radiographic joint status. Thus, it was not possible to determine cut points for pain and function defining 'requirement for TJR'.

Conclusion:

Although symptom levels were higher in patients recommended for TJR, pain and functional disability alone did not discriminate between those who were and were not considered to need TJR by the orthopaedic surgeon.

Introduction

Osteoarthritis (OA) is a major cause of disability worldwide [1]. Over the past years, interest has grown among the scientific community, pharmaceutical companies, and regulatory agencies in the development of drugs that might influence the natural history of structural changes in OA by preventing, retarding, or reversing cartilage breakdown. Interest exists, therefore, in identifying a valid, dichotomous outcome variable that reflects the natural history of structural changes in OA. In particular, interest has grown in using the requirement of total joint replacement (TJR) as a "hard" endpoint [2,3]. Limitations exist, however, in the use of such an outcome. Performance of TJR is a measure of utilization and not of a health state. Numerous non-health related factors have been shown to influence utilization including patient race, ethnicity, income, activity level and preferences among others, and other non-musculoskeletal health factors influence the decision to undergo TJR including comorbidity [2-10]. Thus, a better alternative might be to change "time to TJR" to "time to fulfill the criteria for TJR" [11]. In this context and as described elsewhere [12,13], an international working group was created under the auspices of Osteoarthritis Research Society International (OARSI) and Outcome Measures in Rheumatology Clinical Trials (OMERACT). The group's charge was to elaborate a set of criteria defining a state corresponding to recommendation for TJR in patients with symptomatic knee and hip OA, for use in clinical trials evaluating potential disease-modifying drugs and other interventions in OA. It was decided that the domains of pain, physical function and joint structure on radiographs [14-16] would be combined as a surrogate measure of outcome. The consensus was to consider the level of symptoms (i.e., pain and function) at one point, and a definition of radiological progression between 2 time-points [16].The final binary outcome could then be used as

a definition for "responders/non-responders" in OA clinical trials. For each of these domains, a categorical outcome needs to be used to render combination of the domains feasible. To this end, it is necessary to categorize or dichotomize the continuous variables pain and functional disability.

Thus, the objective of the present study was to define cut points for both pain and functional disability, leading to a joint replacement indication. To this end, a data-driven approach, based on real patient data, was chosen.

This article presents the results of a large cross-sectional study performed to define cut-point levels for pain and functional disability among patients with hip or knee OA being evaluated by orthopaedic surgeons for possible need of TJR. The goal was to use these cut offs to develop a theoretical indication for TJR, in hip and knee OA.

PATIENTS AND METHODS

Study design

This international prospective observational cross-sectional study was conducted in the orthopaedics departments of tertiary-care and secondary-care centers in Europe (12 centers, one per country in the Czech Republic, Italy, Spain, Sweden, and the United Kingdom; 2 per country in France and The Netherlands; 3 in Germany), Canada (2 centers), the United States of America (2 centers), and Australia (2 centers).

Ethical approval was obtained from all participating centers.

Study population

Consecutive outpatients consulting with an orthopedic surgeon in one of the participating centers and with a diagnosis of hip or knee OA (according to the orthopedic surgeon and based on symptoms and radiographs) were included. Only patients for whom the surgeon answered 'There are definite radiographic signs of OA of the target joint' were included. Exclusion criteria were: no definite diagnosis of OA, prior TJR or prior osteotomy of the target joint, concomitant inflammatory arthritis (e.g., rheumatoid arthritis, spondyloarthropathy), patient inability to fill in a questionnaire or patient refusal.

Gold standard: indication for TJR

The gold standard was defined by the orthopaedic surgeon's opinion regarding the recommendation for TJR, operationalized as the surgeon stating that (a) TJR was recommended for the patient or (b) the patient's pain and functional disability were

severe enough to indicate TJR but surgery was not indicated because of co-morbidity or patient declining surgery. These answers defined an 'indication for TJR', irrespective of whether the joint replacement surgery was performed or not.

Pain and functional disability

Two self-reported measures, pain and functional disability, were collected using the intermittent and constant osteoarthritis pain (ICOAP) score[17,18] for pain, and the Hip disability and Osteoarthritis Outcome Score Physical function Short-form (HOOS-PS) for hip, and Knee injury and Osteoarthritis Outcome Score Physical function Short-form (KOOS-PS) for knee for function [19-21]. All scores had Likert answer modalities. The scores were linearly transformed to 0-100 scores, where higher scores indicate worse status. These questionnaires previously underwent translation and cross-cultural adaptation into each of the participating countries' languages[18].

Clinical severity was also estimated through the pain, stiffness, and function subscales of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[22] with Likert answer modalities Results were also linearly transformed to a 0-100 score where higher scores indicate worse status.

Symptom duration

The duration of symptoms, at their current level, was collected by self-report.

Radiographic severity

The local investigator evaluated the radiographs of the target joint, recording joint space narrowing as categories (none, <25%, 25-50%, 50-75%, > 75%). Not all canters participated in the radiographic evaluation of severity.

Other clinical data collection.

Demographic data included age, and sex. Other information included weight and height (body mass index was then calculated), and date of onset of development of OA symptoms in the target joint.

Statistical analysis

1. Sample size: It was anticipated that 1000 knee OA and 1000 hip OA patients would be included, allowing the assessment of Receiver Operating Characteristic (ROC) curves and areas under the curve (AUCs) with a precision of 0.03 for an expected AUC of 0.80 [23]. Other sample size calculations based on expected sensitivities or specificities, led to smaller sample sizes (data not shown).

2. Descriptive analysis of pain and functional disability: the distributions of the 2 variables were analysed for both hip and knee OA, according to the gold standard outcome (recommendation for TJR yes/no) and compared using Student's t-test or the Wilcoxon rank test. Pain and function were also categorized in deciles, and the frequency of the positive gold standard was assessed (with exact confidence intervals, by the Clopper-Pearson method [24]) to describe the relation between pain, function and indication for TJR.

3. Univariate ROC curves: this was the main planned analysis to assess cut points for pain and functional disability. The ability of pain and functional disability to predict the gold standard was assessed in a univariate manner by a non parametric ROC curve [25] and its AUC was calculated. The null hypothesis was that pain and functional disability levels could not distinguish the groups 'recommended for TJR yes/no'. The criteria for accepting the null hypothesis were AUCs < 0.65. If the null hypothesis was rejected, it was planned to assess cut points to maximise specificity (for a specificity

of 90%, 95%, 98%) but also sensitivity, and for each cut point, the sensitivity, the specificity, and the likelihood ratios were assessed.

To take into account radiographic severity, the analyses were stratified on radiographic severity by analysing the relationship between symptoms and recommendation for TJR, for a radiographic joint space narrowing <50%, 50-75%, and more than 75% separately.

4. Correlation between pain and function was examined graphically and tested by Spearman's correlation coefficient.

5. Logistic regression. Pain and function were combined based on logistic regression. In the logistic regression, pain and function were entered into a bivariate model to predict theoretical indication for TJR, with a stepwise selection mode. The goodness-of-fit was checked with Hosmer-Lemeshow's test[26]. The regression parameters of the variables pain and function allowed the assessment of the relative importance (weight) of these variables versus the gold standard, thus allowing us to combine the 2 domains ($\beta_1 pain + \beta_2 function$ where β_1 and β_2 are the regression parameters of the variables *pain* and *function* respectively). The combination was then tested using non parametric ROC curves as described above, and stratified on radiographic severity as explained above.

6. Additional sensitivity analyses: Analyses were run separately for the hip and knee. Potential heterogeneity across centers (regrouped by country) was assessed. A modified version of the gold standard question was modelled ('surgeon saying the patient is referred for TJR', not taking into account patients not referred to surgery due to comorbidities or patient refusal). Another statistical technique involving the 75th percentile of the distribution of patients recommended for TJR was applied. The 75th percentile gave the value of the sum (pain+function) defining 75% of the

population which had an indication for TJR. Furthermore, the same analyses were performed using WOMAC pain and function subscales.

All analyses were performed using SAS, the Statistical Analysis System, version 9.1. Statistical significance was set at 0.05.

RESULTS

Patient characteristics (Table 1)

In all, 1974 patients were included; 1909 had an answer for the gold standard question and were analysed: 1130 knee OA and 779 hip OA patients (**Table 1**). The patients were included in Europe (N=1050), Australia (N=394), the United States of America (N=261), and Canada (N=204). **Supplementary file 1** shows the characteristics of the patients from the different centers.

Mean age of the patients was 66.4 (standard deviation, SD: 10.9) years, 58.1% were women, mean OA duration was reported as 5.4 (SD 6.9) years. Of the 1909 patients, 628 (55.6%) knee patients and 574 (73.7%) hip patients were recommended for TJR. The recommendation was mainly related to the surgeon stating TJR was indicated (91.7% of indications) and much less often to the answers 'although the symptoms are severe enough, the patient declined surgery' (4.0%) or 'there were comorbidities' (4.3%). The frequency of indication for TJR varied across countries, from 33.8% with an indication for surgery among the patients from the Italian center, to 87.9% among the patients from the Czech Republic center.

Pain assessed by ICOAP and functional disability by HOOS-PS / KOOS-PS (Table 2)

Scores for pain and functional disability were not normally distributed (**Figure 1 for online version only**), but showed a wide spread in severity of symptoms. Pain had the following distribution in knee OA: mean±SD 50.3±22.0, median 50.0 (first quartile=31.8, third quartile =68.2, range 0-100) and in hip OA mean±SD 53.3±22.6, median 54.5 (first quartile=4.1, third quartile=70.5, range 0-100). Functional impairment had the following distribution in knee OA: mean±SD 55.5±18.8, median 51.2 (first quartile=42.0, third quartile=66.6, range 0-100) and in hip OA mean±SD 57.8±21.5, median 55.9 (first quartile=41.7, third quartile=74.8, range 0-100).

Pain and functional disability levels and their duration, for those who did versus did not receive a TJR recommendation, are shown in **Table 2**. Patients meeting the gold standard had higher symptom levels. For knee/hip patients pooled, mean pain was 55.5 [95% confidence interval 54.2, 56.8] for those with TJR recommendation versus 44.9 [43.2, 46.6] for those without TJR recommendation ($p < 0.0001$). Mean functional impairment was 59.8 [58.7, 60.9] for those with TJR recommendation versus 50.9 [49.3, 52.4], for those without TJR recommendation ($p < 0.0001$). However, there was a wide overlap in symptom levels between groups: almost 50% of patients in the lowest decile of symptom scores were considered candidates for TJR, whereas only 75% of patients in the highest decile were considered candidates (**Figure 2 for online version only**).

Symptom duration

The duration of symptoms at their current level was longer for patients who did versus those who did not receive a TJR recommendation (**Table 2**).

Univariate ROC curves

Taking pain and function separately, in the pooled hip/knee population, it was not possible to determine relevant cut points defining recommendation for TJR (**Figure 1**). The AUCs for the ROC curves for pain and function versus the gold standard were 0.64 [95% confidence interval, 0.61, 0.67] and 0.63 [0.60, 0.66}, respectively. Thus, we had to accept the null hypothesis (i.e., that pain and functional disability levels do not distinguish patients with versus without a recommendation for TJR). The cut points had low diagnostic properties: e.g., for a specificity of 0.90, the sensitivity was only 0.23 for pain and 0.24 for function; i.e., the positive and negative

likelihood ratios were only (1.17; 0.43) for pain and (1.18; 0.42) for physical disability. *After stratifying on radiographic severity,* the AUCs were not much improved (AUCs ranging from 0.65 to 0.68 for pain, and 0.60 to 0.63 for function, respectively) and cut points assessed had low diagnostic properties (data not shown).

Correlation and relative importance of pain, functional disability and radiographic status

Pain and functional disability were only moderately correlated (R= 0.59, p<0.0001), in the pooled hip/knee population, indicating that these domains were not redundant.

In logistic regression, the coefficients of regression of pain and function were very similar (and significant), indicating pain and function are independent predictors of recommendation for TJR, with similar weights. The coefficients of regression were 0.015 for pain, and 0.013 for function, respectively (both p<0.0001). This result justified our combining pain and functional status additively with equal weights. Furthermore, radiographic severity was a significant independent predictor of recommendation for TJR (p<0.0001) in the pooled hip/knee population.

ROC curves for the sum (pain+function)

With the sum (pain+function), it was also not possible to determine cut-points leading to relevant sensitivity/specificity in the pooled hip/knee population: the AUC of the ROC curve was 0.64 [95% confidence interval, 0.61, 0.67, and for a specificity of 0.90 the sensitivity was 0.27 (i.e., positive and negative likelihood ratios were 2.70; 0.81).

When these analyses were stratified on radiographic severity, the AUCs were not improved (AUCs ranging from 0.62 to 0.65 in the different radiographic groups).

Sensitivity analyses

Several sensitivity analyses including use of alternate measures, specifically the WOMAC pain and function subscales (**supplementary file 2**); and changing the gold standard to true indication for TJR (i.e., not considering patients with severe status but comorbidites or patient refusal as recommendations for TJR), did not modify the results (data not shown).

The 75th percentile technique gave 89 as the value of the sum (pain+function) defining 75% of the population which had an indication for TJR (respectively, 87 and 92, for knee and hip). When applying the cut point of 89 to the whole population, 59% of the patients were above that level; specificity was 0.51, sensitivity was 0.66, the positive and negative likelihood ratios were (1.34; 0.86).

However, analysing the participants with hip or knee OA separately, the association between symptoms and surgery was stronger in the hip than in the knee (**Figure 2**).

The AUCs of the ROC curves of the sum pain+function were higher in hip OA (AUC, 0.70, 95% CI, 0.66-0.75) than in knee OA (AUC, 0.60, 95% CI, 0.56-0.64). However, even so, for hip patients the cut-points assessed had low diagnostic properties; e.g. the cutpoint leading to a sensitivity of 90% was 66 (sum pain+function); for that cutpoint, for a specificity of 0.92, the sensitivity was only 0.31; i.e., the positive and negative likelihood ratios were only (1.35; 0.23).

We also showed when analysing the centers separately (regrouped by country), that in certain centers pain and function were more strongly related to receipt of a TJR recommendation than in other centers (e.g., AUC of the ROC curves for the sum

pain+function in hip OA, 0.75 to 0.86 in centers in Canada, France, Germany and Australia as compared with 0.54 to 0.67 in centers in The United Kingdom, the United States, the Netherlands and the Czech Republic, **supplementary file 3**).

DISCUSSION

This large-scale international study was launched to determine whether self-reported measures of pain and function could be used to accurately identify patients with OA whose surgeons recommended them for total hip or knee arthroplasty. The first conclusion of this work is that, indeed, among patients with hip and knee OA referred to an orthopedic surgeon, the level of symptoms was higher among patients for whom TJR was indicated by the orthopedic surgeon. Both the level of pain and self-reported functional impairment were independently, though weakly, predictive of the surgeon's recommendation for TJR. The second conclusion is that we could not find a cut point for pain and or physical disability that accurately discriminated across different countries, patients who did versus did not receive a TJR recommendation, as the AUCs for ROC curves were low (< 0.65). Radiographic severity, when available, was a strong predictor of recommendation for TJR but stratifying by radiographic joint status did not modify our conclusions.

Factors consistently predicting TJR are symptom levels and radiographic severity. Less consistent predictors have included gender, age and current treatment [2, 4-9]. In the present study, the mean values of pain and function in the group of patients considered candidates for surgery by the surgeons were consistent with previously reported data in this area [27-32]. We confirmed here that both pain and functional disability are independent predictors of recommendation by a surgeon for TJR; however, previous studies did not include a control group to attempt to determine cut points for patient-reported outcomes. In this study, using the ICOAP and HOOS-PS/KOOS-PS, though pain and function were correlated (as could be expected), the correlation was only moderate, which indicates that pain and function using these

scores are not redundant when analysing OA patients. Furthermore, we also found that the duration of the symptoms at their current level was an important factor explaining indication for TJR. Other predictors included radiographic severity, stiffness (assessed by WOMAC) and OA disease duration (data not shown).

Despite the fact there was a difference in the level of symptoms between the two groups (candidate for surgery yes/no) the overlap between the two groups prevented us from proposing a specific cut-off. Indeed, among these OA patients referred to an orthopaedic surgeon, most patients were symptomatic. However, the surgeons often decided that surgery was warranted even among the less symptomatic patients (around 50% of the patients in the lower decile of symptoms were recommended for TJR, **Figure 2 online**), or that surgery was not warranted even if the symptoms were severe (only around 75% of these patients were considered surgery candidates). This indicates that the level of symptoms in this population was not the only driver for such a TJR indication [30-32]. Possibly, the surgeons paid greater attention to the radiographic severity than to the symptom levels [10, 27, 33] and several studies have indicated a discordance between radiographs and symptoms in lower-limb OA [27, 34-37]. In the present study however, stratifying the analyses on radiographic severity did not modify our conclusions. Finally, the present results indicated a stronger relationship between symptoms and surgical indication in hip OA than in knee OA.

It is possible that the questionnaires used, the ICOAP and KOOS-PS/HOOS-PS [17-21] (which were not seen by the orthopaedic surgeon) may assess different aspects of symptoms, than what the orthopaedic surgeon usually assesses in the clinic; however, it is reassuring to note that these new tools gave results very similar to the WOMAC sub-scales. Indeed, the sensitivity analyses performed using WOMAC data

confirmed our main results. Perhaps also, other data related to patient-reported outcomes could be relevant in the indication for TJR, such as worsening of symptoms (e.g., minimal clinically important deterioration); however, we did not collect change in status in this study, but only status at one time point, and persistence of that status, since we felt that a decision for TJR would be more strongly based on status than on change. Clearly, in addition to symptomatic severity, many other factors are as strong or stronger determinants of surgery [9,38-41]. Furthermore, perhaps other aspects of symptomatic severity are taken into account in the surgeons' decision, e.g. the duration of symptoms (whereas questionnaires have a short time-frame), or the ongoing symptomatic treatment of the patient that may influence his/her current level of symptoms[4]. Finally, the surgeons may have based their surgical decision on joint mobility or peri-articular amyotrophy [38], which were not assessed here. There were clear differences across centers and countries; these might be explained by several elements, including differences in the health care systems, or characteristics / training of the surgeons.

This study has strengths and weaknesses. It is a large, international study which enhances the external validity of our results. On the other hand and as could be expected, there were differences across centers and countries in terms of symptomatic severity and in terms of the frequency of indication for TJR as assessed by the orthopaedic surgeon [9,32]. We do not believe this is an important limitation to the present results. Indeed, the objective here was to develop international criteria reflecting a level of OA symptoms and disability at which point TJR should be considered, for use as outcome measure in clinical trials. In this context, it was necessary to include patients from different backgrounds. In this study, one possible

bias is that only symptomatic OA patients were included since the patients had to have definite OA to be included, and were in fact seeing an orthopaedic surgeon, generally to discuss a surgical indication for their target joint (we do not have information regarding if the patients were coming for the first time, or for return visits). Therefore the present study did not include many asymptomatic patients which may explain the low predictive power of symptomatic severity here. Indeed, symptom thresholds associated with TJR in a more heterogeneous sample (including asymptomatic patients) might be relevant for defining endpoints for observational studies. Nevertheless, the patients in the study presented with a wide range of symptomatic severity, and only about half of them were considered candidates for TJR. Several statistical techniques and sensitivity analyses were performed, to further confirm the internal validity of our results; and the study was not underpowered.

In this study, the gold standard was the surgeon's opinion regarding need for TJR [13, 28]. We considered that if surgery was recommended or if the surgeon considered symptoms were severe enough for surgery (although because of comorbidity [10] or patient refusal [9], the patient was not referred for surgery), a state of indication for TJR was attained. However, we did not collect data regarding actual carrying-out of surgery in these patients, which may differ widely[2-10].

In conclusion, this large study indicates that among patients referred to an orthopaedic surgeon to discuss TJR, the level of symptoms was higher among patients for whom TJR was indicated by the surgeon, but there was no cut point for pain and functional disability allowing to discriminate between patients with or without

an indication for TJR.

REFERENCES

1. Elders MJ. The increasing impact of arthritis on public health. J Rheum 2000; 60 suppl: 6-8.

2. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, et al. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis? J Rheumatol. 1999;26:855-61.

3. Altman RD, Abadie E, Avouac B, Bouvenot G, Branco J, Bruyere O, et al; Group for Respect of Excellence and Ethics in Science (GREES). Total joint replacement of hip or knee as an outcome measure for structure modifying trials in osteoarthritis. Osteoarthritis Cartilage. 2005;13:13-9.

4. Maillefert JF, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, et al. A composite index for total hip arthroplasty in patients with hip osteoarthritis. J Rheumatol. 2002;29:347-52.

5. Maillefert JF, Gueguen A, Monreal M, Nguyen M, Berdah L, Lequesne M, et al. Sex differences in hip osteoarthritis: results of a longitudinal study in 508 patients. Ann Rheum Dis. 2003;62:931-4.

6. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, et al. Differences between men and women in the rate of use of hip and knee arthroplasty. N Engl J Med. 2000;342:1016-22.

7. Gossec L, Tubach F, Baron G, Ravaud P, Logeart I, Dougados M. Predictive factors of total hip replacement due to primary osteoarthritis: a prospective 2 year study of 505 patients. Ann Rheum Dis. 2005;64:1028-32. Epub 2005 Jan 7.

8. Boutron I, Rannou F, Jardinaud-Lopez M, Meric G, Revel M, Poiraudeau S. Disability and quality of life of patients with knee or hip osteoarthritis in the primary

care setting and factors associated with general practitioners' indication for prosthetic replacement within 1 year. Osteoarthritis Cartilage. 2008;16:1024-31. Epub 2008 Feb 13.

9. Hawker GA, Guan J, Croxford R, Coyte PC, Glazier RH, Harvey BJ, et al. A prospective population-based study of the predictors of undergoing total joint arthroplasty. Arthritis Rheum. 2006;54:3212-20.

10. Maillefert JF, Roy C, Cadet C, Nizard R, Berdah L, Ravaud P. Factors influencing surgeons' decisions in the indication for total joint replacement in hip osteoarthritis in real life. Arthritis Rheum. 2008 Feb 15;59(2):255-62.

11. Maillefert JF, Hawker GA, Gossec L, Mahomed NN, Lohmander S, Dieppe PA, et al. Concomitant therapy: an outcome variable for musculoskeletal disorders? Part 2: total joint replacement in osteoarthritis trials. J Rheumatol 2005 ;32:2449-51.

12. Gossec L, Hawker G, Davis AM, Maillefert JF, Lohmander LS, Altman R, et al. OMERACT/OARSI initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. J Rheumatol. 2007;34:1432-5.

13. Dougados M, Hawker G, Lohmander S, Davis AM, Dieppe P, Maillefert JF, et al. OARSI/OMERACT criteria of being considered a candidate for total joint replacement in knee/hip osteoarthritis as an endpoint in clinical trials evaluating potential disease modifying osteoarthritic drugs. J Rheumatol. 2009;36:2097-9.

14. Gossec L, Jordan JM, Lam MA, Fang F, Renner JB, Davis A, et al ; For The Oarsi-Omeract Task Force Total Articular Replacement As Outcome Measure In Oa. Comparative evaluation of three semi-quantitative radiographic grading techniques for hip osteoarthritis in terms of validity and reproducibility in 1404 radiographs: report of the OARSI-OMERACT Task Force. Osteoarthritis Cartilage. 2009;17:182-7. Epub 2008 Aug 8.

15. Gossec L, Jordan JM, Mazzuca SA, Lam MA, Suarez-Almazor ME, Renner Jbet al; For the OARSI–OMERACT task force "total articular replacement as outcome measure in OA". Comparative evaluation of three semi-quantitative radiographic grading techniques for knee osteoarthritis in terms of validity and reproducibility in 1759 X-rays: report of the OARSI-OMERACT task force Osteoarthritis Cartilage. 2008; 16: 742-748. Epub 2008 Apr 14.

16. Ornetti P, Brandt K, Hellio-Le Graverand MP, Hochberg M, Hunter DJ, Kloppenburg M, et al. OARSI-OMERACT definition of relevant radiological progression in hip/knee osteoarthritis. Osteoarthritis Cartilage. 2009;17:856-63. Epub 2009 Feb 9.

17. Hawker GA, Davis AM, French MR, Cibere J, Jordan JM, March L, et al. Development and preliminary psychometric testing of a new OA pain measure--an OARSI/OMERACT initiative. Osteoarthritis Cartilage. 2008;16:409-14.

18. Maillefert JF, Kloppenburg M, Fernandes L, Punzi L, Günther KP, Martin Mola E, et al. Multi-language translation and cross-cultural adaptation of the OARSI/OMERACT measure of intermittent and constant osteoarthritis pain (ICOAP). Osteoarthritis Cartilage. 2009;17:1293-6. Epub 2009 Apr 17.

19. Perruccio AV, Lohmander SL, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthritis Cartilage. 2008;16:542-50. Epub 2008 Feb 21.

20. Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an OARSI/OMERACT initiative. Osteoarthritis Cartilage. 2008;16:551-9. Epub 2008 Mar 4.

21. Davis AM, Perruccio AV, Canizares M, Hawker GA, Roos EM, Maillefert JF, et al. Comparative validity and responsiveness of the HOOS-PS and KOOS-PS to the WOMAC physical function subscale in total joint replacement for osteoarthritis. Osteoarthritis Cartilage. 2009;17:843-7. Epub 2009 Jan 31.

22. Bellamy N. WOMAC osteoarthritis index. A user's guide. London Health Services Centre; McMaster University: Ontario; 1996.

23. Zhou XH, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine, John Wiley & Sons, Inc., New-York. 2002.

24. Clopper C, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; 26: 404-413.

25. Hanley JA, Mac Neil BJ. The meaning and use of the area under a receiver operating curve (ROC) curve. Radiology 1982; 143: 29-36.

26. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York (1989): John Wiley & Sons, Inc.

27. Dieppe P, Judge A, Williams S, Ikwueke I, Guenther KP, Floeren M, et al; EUROHIP Study Group. Variations in the pre-operative status of patients coming to primary hip replacement for osteoarthritis in European orthopaedic centres. BMC Musculoskelet Disord. 2009;10:19.

28. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, et al. Determining the need for hip and knee arthroplasty: the role of clinical severity and patients' preferences. Med Care. 2001;39:206-16.

29. Fortin PR, Clarke AE, Joseph L, Liang MH, Tanzer M, Ferland D, et al. Outcomes of total hip and knee replacement: preoperative functional status predicts outcomes at six months after surgery. Arthritis Rheum. 1999;42:1722-8.

30. Nilsdotter AK, Petersson IF, Roos EM, Lohmander LS. Predictors of patient

relevant outcome after total hip replacement for osteoarthritis: a prospective study. Ann Rheum Dis. 2003;62:923-30.

31. Judge A, Welton NJ, Sandhu J, Ben-Shlomo Y. Modeling the need for hip and knee replacement surgery. Part 1. A two-stage cross-cohort approach. Arthritis Rheum. 2009;61(12):1657-66.

32. Ackerman IN, Dieppe PA, March LM, Roos EM, Nilsdotter AK, Brown GC, et al. Variation in age and physical status prior to total knee and hip replacement surgery: a comparison of centers in Australia and Europe. Arthritis Rheum. 2009;61:166-73.

33. Fox KM, Hochberg MC, Resnik CS, Kenzora JE, Hebel JR, Zimmerman SI, et al. Severity of radiographic findings in hip osteoarthritis associated with total hip arthroplasty. J Rheumatol. 1996;23:693-7.

34. Cicuttini FM, Baker J, Hart DJ, Spector TD. Association of pain with radiological changes in different compartments and views of the knee joint. Osteoarthritis Cartilage. 1996;4:143-7.

35. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol. 2000;27:1513-7.

36. Lethbridge-Cejku M, Scott WW Jr, Reichle R, Ettinger WH, Zonderman A, Costa P, et al. Association of radiographic features of osteoarthritis of the knee with knee pain: data from the Baltimore Longitudinal Study of Aging. Arthritis Care Res. 1995;8:182-8.

37. Szebenyi B, Hollander AP, Dieppe P, Quilty B, Duddy J, Clarke S, et al. Associations between pain, function, and radiographic features in osteoarthritis of the knee. Arthritis Rheum. 2006;54:230-5.

38. Hadorn DC, Holmes AC. The New Zealand priority criteria project. Part 1:

Overview. BMJ. 1997; 314(7074):131-4.

39. Conner-Spady BL, Arnett G, McGurran JJ, Noseworthy TW; Steering Committee of the Western Canada Waiting List Project. Prioritization of patients on scheduled waiting lists: validation of a scoring system for hip and knee arthroplasty. Can J Surg. 2004;47:39-46.

40. McHugh GA, Campbell M, Silman AJ, Kay PR, Luker KA. Patients waiting for a hip or knee joint replacement: is there any prioritization for surgery? J Eval Clin Pract. 2008;14:361-7. Epub 2008 Mar 24.

41. Hoogeboom TJ, van den Ende CH, van der Sluis G, Elings J, Dronkers JJ, Aiken AB, et al. The impact of waiting for total joint replacement on pain and functional status: a systematic review. Osteoarthritis Cartilage. 2009;17:1420-7. Epub 2009 May 20.

Table 1. Patients' characteristics.

	All patients N=1909	Knee patients N=1130	Hip patients N=779
Age, years	66.4±10.9	67.5±10.4	64.9±11.4
Sex, N (%) women	1086 (58.1)	657 (58.9)	429 (56.9)
OA symptom duration, years	5.4±6.9	6.3±7.7	4.1±5.5
Body mass index, kg/m^2	29.9±6.3	31.0±6.8	28.3±5.2
Pain, ICOAP score	51.6±22.3	50.3±22.0	53.3±22.6
Functional disability, HOOS-PS /KOOS-PS scores	56.5±20.0	55.5±18.8	57.8±21.5
Pain, WOMAC subscale	54.0±21.0	52.5±20.8	56.3±21.1
Function, WOMAC subscale	57.0±20.5	55.2±20.2	59.5±20.8
Radiographic joint space narrowing, N (%)* < 25%	95 (10.7)	67 (13.0)	28 (7.8)
25-50%	131 (14.7)	95 (18.5)	36 (9.6)
50-75%	274 (30.8)	159 (31.0)	115 (30.6)
> 75%	389 (43.8)	192 (37.4)	197 (52.4)

Results are presented as mean±standard deviation unless otherwise mentioned. Pain and functional disability were linearly transformed to 0-100 scores where 100=worst state.

HOOS-PS/KOOS-PS : Hip/knee knee disability and Osteoarthritis Outcome Score physical function short-form. WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

*X-ray scoring was only available for 889 patients and % are % of available data.

Table 2. Symptom levels and radiographic severity according to recommendation for TJR in 1909 OA patients

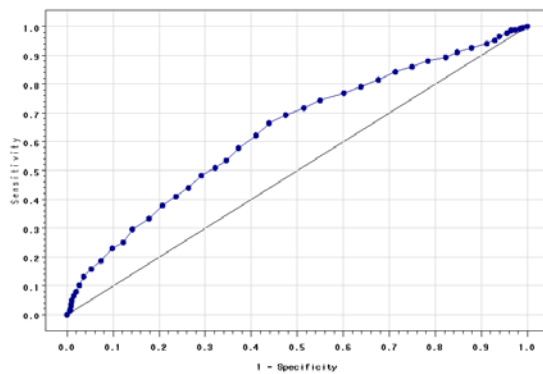|	Knee OA : TJR+ N=628	Knee OA: TJR – N=502	Hip OA: TJR+ N=574	Hip OA: TJR – N=205
Pain, ICOAP score	53.7 (52.0, 55.5)	45.9 (44.0, 47.9)	57.3 (55.6, 59.1)	42.4 (39.1, 45.7)
Functional disability, HOOS-PS/ KOOS-PS scores	58.1 (56.6, 59.7)	52.3 (50.5, 54.0)	61.4 (59.8, 63.1)	47.4 (44.1, 50.7)
Pain, WOMAC subscale	56.4 (54.8, 57.9)	47.3 (45.3, 49.4)	59.8 (58.3, 61.4)	45.9 (42.4, 49.3)
Function, WOMAC subscale	59.0 (57.4, 60.5)	50.3 (48.3, 52.4)	63.3 (61.7, 64.9)	48.7 (45.3, 52.2)
Duration of symptoms at the current level, months	11.0 (6.2,15.2)	5.9 (2.3, 10.2)	6.9 (3.5, 11.5)	5.9 (2.1, 10.6)
Radiographic joint space narrowing, N (%)*				
< 25%	3 (1.2)	64 (23.6)	3 (1.1)	25 (22.3)
25-50%	24 (10.0)	70 (25.8)	14 (5.3)	22 (19.6)
50-75%	85 (35.4)	74 (27.3)	75 (28.5)	39 (34.8)
> 75%	128 (53.3)	63 (32.3)	171 (65.0)	26 (23.2)

Results are presented as mean (95% confidence interval) except for radiographic results. Pain and functional disability were linearly transformed to 0-100 scores where 100=worst state. TJR+ : indication for TJR. TJR- : no indication for TJR. For other abbreviations please see Table 1.
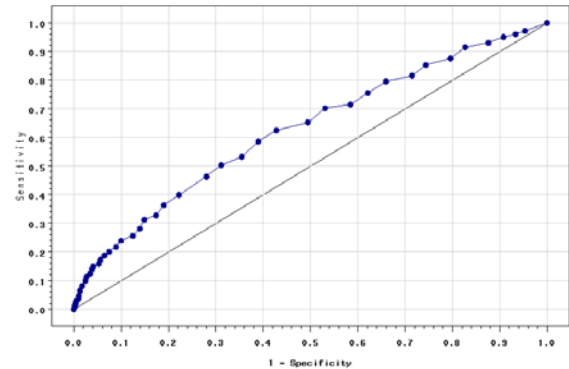
* % of available data

Figure 1. Ability of pain and functional impairment severity to predict indication for TJR in 1909 hip or knee OA patients.
Figure 1a: pain, figure 1b: functional disability
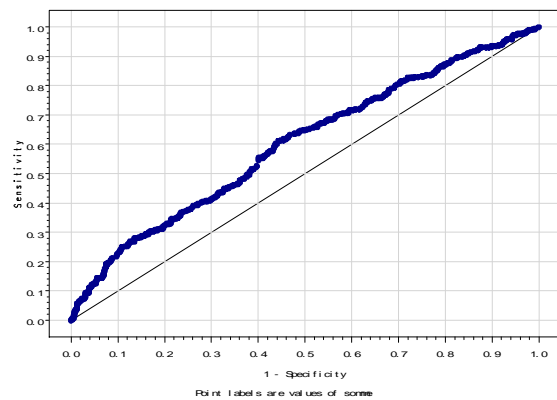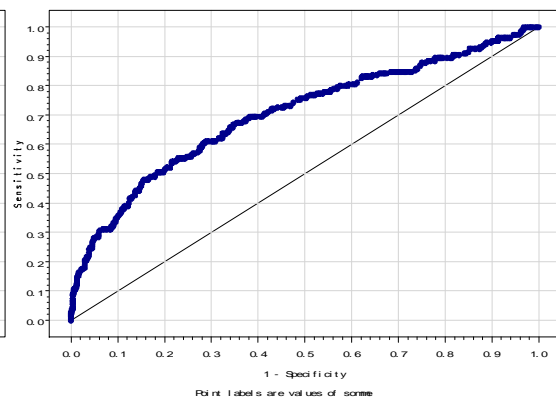1a   1b

AUC for curve 1a: 0.64 (95% confidence interval, 0.61-0.67), AUC for curve 1b: 0.63 (95% confidence interval, 0.60-0.66).

Figure 2. Ability of the sum (pain+function) to predict indication for TJR, in knee and hip OA separately
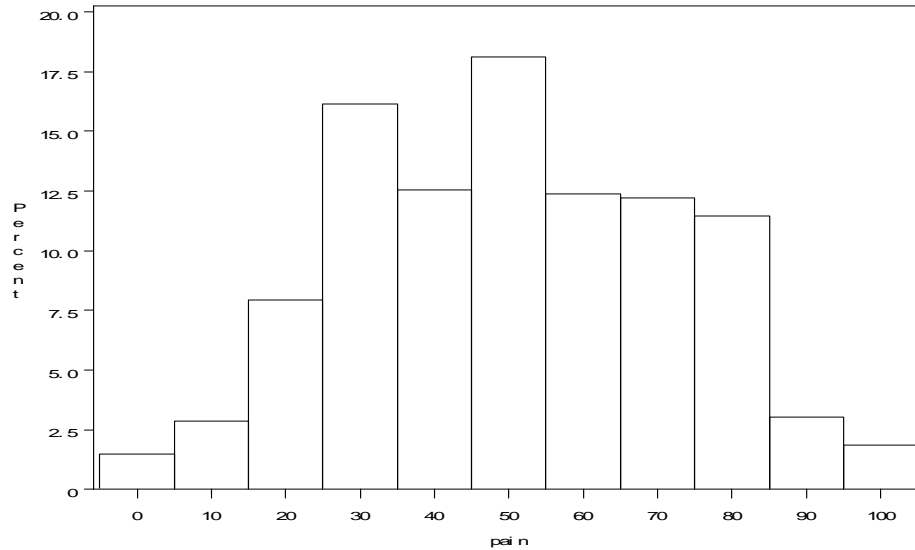
Figure 1a: knee, figure 1b; hip.

1a

1b

AUCs for the ROC curves: 1a 0.60 (95% confidence interval, 0.56-0.64), 1b 0.70 (95% confidence interval, 0.66-0.75).
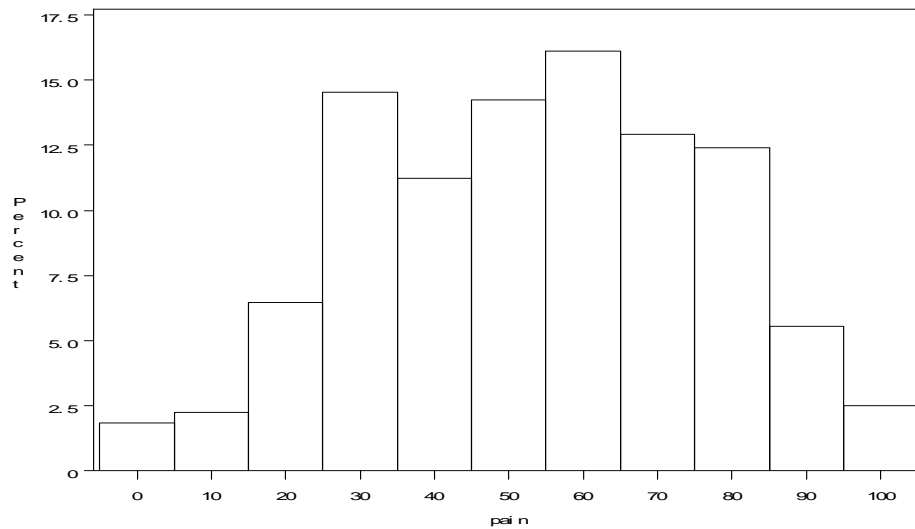
Figure 1 online only. Distribution of pain (assessed by ICOAP) and functional disability (assessed by HOOS-PS/KOOS-PS) in 1909 patients with hip/knee OA seeing an orthopaedic surgeon in outpatient clinic.
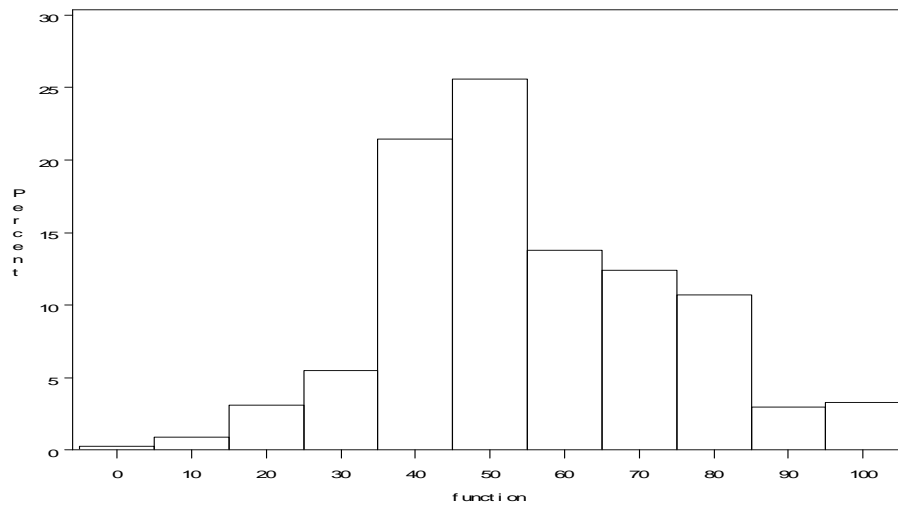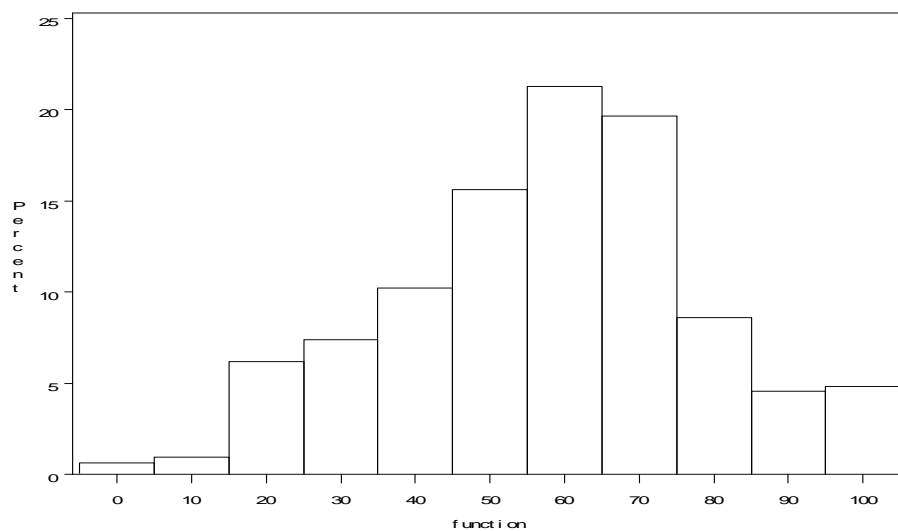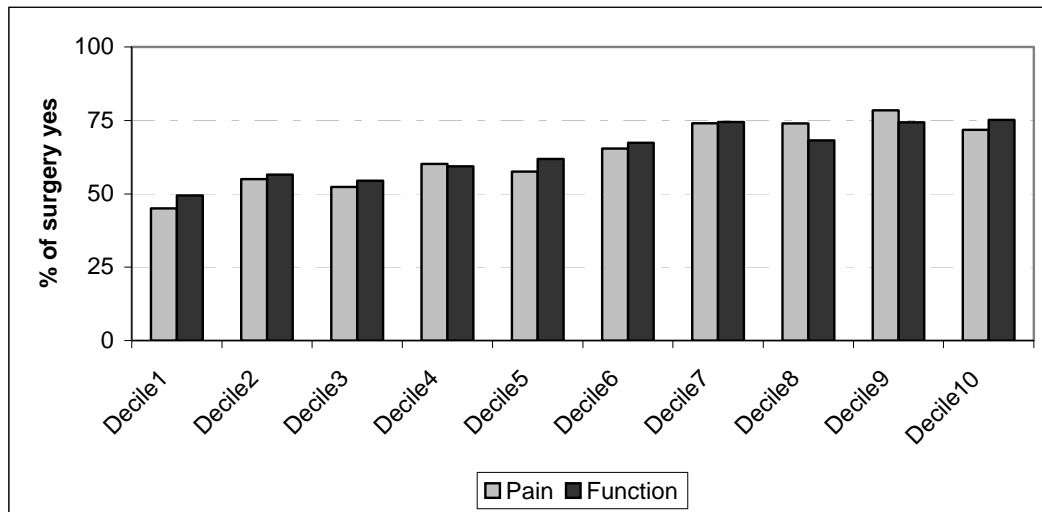
1a pain in knee OA

1b pain in hip OA

1c function in knee OA

1d function in hip OA

Figure 2 online only. Frequency of indication for TJR, in symptomatic knee or hip OA patients, according to the deciles of symptomatic severity (light bar, pain distribution, dark bar, function distribution, both shown by deciles)**.**

Results obtained for pain and function assessed by WOMAC subscales.

Pain and functional disability assessed by WOMAC subscales were more strongly correlated than assessed by ICOAP and HOOS/KOOS (R= 0.84, p<0.0001), in the pooled hip/knee population.

In univariate logistic regression, the coefficients of regression of pain and function assessed by WOMAC, were 0.026 for pain, p<0.0001, and 0.028 for function, p<0.0001. However in multiple regression including as explanatory variables, pain and function, the coefficients of regression changed to 0.006 for pain, p=0.197, and 0.022 for function, p<0.0001. This indicated that although WOMAC pain and function analysed separately were predictors of recommendation for TJR, when analysed together, WOMAC pain was not a significant predictor.

Supplementary file 2 online only. Patients' characteristics, by center (presented as centers from one country).

	Australia N=396	Canada N=214	Czech R. N=199	France N=123	Germany N=193	Italy N=70	Netherlands N=133	Sweden N=128	UK N=194	USA N=261
Age, years	69.1±10.2	68.3±14.1	67.5±9.7	66.7±11.6	62.8±11.5	NA	68.8±10.7	67.0±9.9	66.5±11.3	60.4±9.2
Sex, N (%) women	230 (58.9)	108 (54.0)	129 (65.8)	79 (65.3)	117 (61.3)	43 (61.4)	77 (63.1)	72 (57.1)	103 (55.7)	120 (46.3)
OA symptom duration, years	3.9±4.3	5.9±7.2	2.3±4.0	3.4±4.6	4.1±7.6	1.2±0.4	2.2±2.2	4.2±5.3	7.63±7.9	7.9±9.0
Body mass index, kg/m^2	29.8±6.1	32.3±6.8	28.5±4.6	27.0±4.9	28.8±4.9	NA	28.2±4.4	27.7±4.5	31.1±6.5	34.6±7.5
Pain, ICOAP score	45.8±24.2	56.7±22.6	57.8±19.9	48.3±18.9	54.8±20.4	49.1±19.7	50.1±21.0	59.0±21.5	44.0±21.2	53.3±21.7
Functional disability, HOOS-PS /KOOS-PS scores	54.9±20.5	56.7±21.0	60.6±19.9	49.3±19.0	55.0±18.4	NA	56.0±21.2	57.0±18.6	62.1±18.1	54.8±19.5

Results are presented as mean±standard deviation unless otherwise mentioned. Pain and functional disability were linearly transformed to 0-100 scores where 100=worst state.

NA: not available.

Supplementary file 3 online only. Predictive capacities of the sum (pain+function) for the indication of TJR in hip OA, by center (presented as centers from one country).

	Australia N=100	Canada N=70	Czech R. N=99	France N=69	Germany N=122	Netherlands N=61	Sweden N=53	UK N=82	USA N=75
AUC	0.753	0.863	0.676	0.759	0.765	0.666	0.724	0.539	0.638
Cut-point	66	66	41	56	44	66	74	62	35
Sensitivity, %	97.6	96.7	98.9	99.0	99.0	95.6	95.0	93.8	100.0
Specificity, %	32.8	60.0	18.2	35.0	21.1	37.5	23.1	16.7	8.0
Positive LR	1.45	2.41	1.21	1.52	1.25	1.53	1.23	1.13	1.08
Negative LR	0.07	0.05	0.06	0.02	0.05	0.12	0.22	0.37	0

The AUCs of the ROC curves reflect the properties of (pain+function). For indicative purposes, cut-points were chosen to obtain 95% specificity; values are shown for that cut-point. LR: likelihood ratio. Values could not be calculated for Italy due to missing data regarding function scores.