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

L. Gossec, S. Paternotte, J.F. Maillefert, C. Combescure, P.G. Conaghan, A.M. Davis, K.-P. Gunther, G. Hawker, M. Hochberg, J.N. Katz, et al.

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

L. Gossec †, , ,

S. Paternotte †,

J.F. Maillefert ‡, §, ‖,

C. Combescure ¶,

P.G. Conaghan#,

A.M. Davis † †, ‡ ‡, § §,

K.-P. Gunther‖‖,

G. Hawker ¶ ¶, ##,

M. Hochberg † † †, ‡ ‡ ‡,

J.N. Katz § § §, ‖‖‖, ¶ ¶ ¶,

M. Kloppenburg###, † † † †,

K. Lim ‡ ‡ ‡ ‡,

L.S. Lohmander § § § §,

N.N. Mahomed‖‖‖‖,

L. March ¶ ¶ ¶ ¶,

K. Pavelka####,

L. Punzi † † † † †,

E.M. Roos ‡ ‡ ‡ ‡ ‡,

L. Sanchez-Riera § § § § §,

J.A. Singh‖‖‖‖‖,

M.E. Suarez-Almazor ¶ ¶ ¶ ¶ ¶,

M. Dougados †,

for the OARSI-OMERACT Task Force "total articular replacement as outcome measure in OA"

† Paris Descartes University, Medicine Faculty, APHP, Rheumatology B Department, Cochin Hospital, Paris, France

‡ Department of Rheumatology, Dijon University Hospital, Dijon F-21078, France

§ University of Burgundy, Dijon F-21079, France

‖ INSERM U887, Dijon F-21079, France

¶ University of Geneva, Switzerland

Section of Musculoskeletal Disease, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, United Kingdom

† † Division of Health Care and Outcomes Research and Arthritis and Community Research and Evaluation Unit, Toronto Western Research Institute, Toronto, Ontario, Canada

‡ ‡ Department of Rehabilitation Science, University of Toronto, Canada

§ § Department of Health Policy, Management and Evaluation, University of Toronto, Canada

‖‖ Department of Orthopaedic Surgery, University of Dresden, Germany

¶ ¶ Division of Rheumatology, Department of Medicine, Women’s College Hospital, University of Toronto, Canada

Clinical Epidemiology and Health Care Research Program, Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Canada

† † † Division of Rheumatology & Clinical Immunology, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

‡ ‡ ‡ Division of Gerontology, Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

§ § § Department of Orthopedic Surgery, Harvard Medical School, Boston, MA, USA

‖‖‖ Division of Rheumatology, Immunology and Allergy, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA

¶ ¶ ¶ Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

† † † † Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

‡ ‡ ‡ ‡ Department of Rheumatology, St Vincent’s Hospital & Western Hospital, Melbourne, Australia

§ § § § Department of Orthopaedics, Clinical Sciences Lund, Lund University, Lund, Sweden

‖‖‖‖ Division of Orthopaedic Surgery, University of Toronto, Toronto Western Hospital, Toronto, Canada

¶ ¶ ¶ ¶ Institute of Bone and Joint Research, University of Sydney, Royal North Shore Hospital, St Leonards, Australia

Institute of Rheumatology, Charles University Prague, Czech Republic

† † † † † Rheumatology Unit, Department of Clinical and Experimental Medicine, University of Padova, Italy

‡ ‡ ‡ ‡ ‡ Institute of Sports and Clinical Biomechanics, University of Southern Denmark, Denmark

§ § § § § Institute of Bone and Joint Research, University of Sydney, Australia

‖‖‖‖‖‖ Birmingham VA Medical Center and University of Alabama, Birmingham, AL, USA

¶ ¶ ¶ ¶ ¶ Section of Rheumatology, University of Texas, M.D. Anderson Cancer Center, Houston, TX, USA

Received 20 May 2010. Accepted 26 October 2010. Available online 31 October 2010.

http://dx.doi.org/10.1016/j.joca.2010.10.025, How to Cite or Link Using DOI

Summary

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 vs without surgeons’ indication for TJR. Receiver Operating Characteristic (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 [standard deviation (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 vs 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.

Keywords

Knee;

Hip;

Osteoarthritis;

Joint replacement;

Surgery;

Symptom

Introduction

Osteoarthritis (OA) is a major cause of disability worldwide1. 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] and [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], [3], [4], [5], [6], [7], [8], [9] and [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] and [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], [15] and [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 two time-points16. 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; two per country in France and The Netherlands; three in Germany), Canada (two centers), the United States of America (two centers), and Australia (two centers).

Ethical approval was obtained from all participating centers.

Study population

Consecutive outpatients consulting with an orthopaedic surgeon in one of the participating centers and with a diagnosis of hip or knee OA (according to the orthopaedic 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 (1) TJR was recommended for the patient or (2) the patient’s pain and functional disability were severe enough to indicate TJR but surgery was not indicated because of comorbidity 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] and [18] for pain, and the Hip disability and

Osteoarthritis Outcome Score (HOOS) for hip, and Knee injury and Osteoarthritis Outcome Score (KOOS) for knee for function [19], [20] and [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' languages18.

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 0.8023. 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 two 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 method24) 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 curve25 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 test26. The regression parameters of the variables pain and function allowed the assessment of the relative importance (weight) of these variables vs the gold standard, thus allowing us to combine the two domains (β1pain + β2function where β1 and β2 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. Finally, a sensitivity analysis was performed, excluding patients from the United Kingdom since for these patients, the questionnaires had been administered differently.

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

Results

Patient characteristics (Table I)

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 I). 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.

Table I. 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

	All patients N = 1909	Knee patients N = 1130	Hip patients N = 779
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/m2	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 ± SD unless otherwise mentioned. Pain and functional disability were linearly transformed to 0–100 scores where 100 = worst state.

*

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

Full-size table

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 II)

Scores for pain and functional disability were not normally distributed (Fig. 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).

Table II. Symptom levels and radiographic severity according to recommendation for TJR

	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 I.

*

% of available data.

Full-size table

Pain and functional disability levels and their duration, for those who did vs did not receive a TJR recommendation, are shown in Table II. 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 vs 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 vs 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 (Fig. 2 for online version only).

Symptom duration

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

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 (Fig. 1). The AUCs for the ROC curves for pain and function vs 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 vs 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.

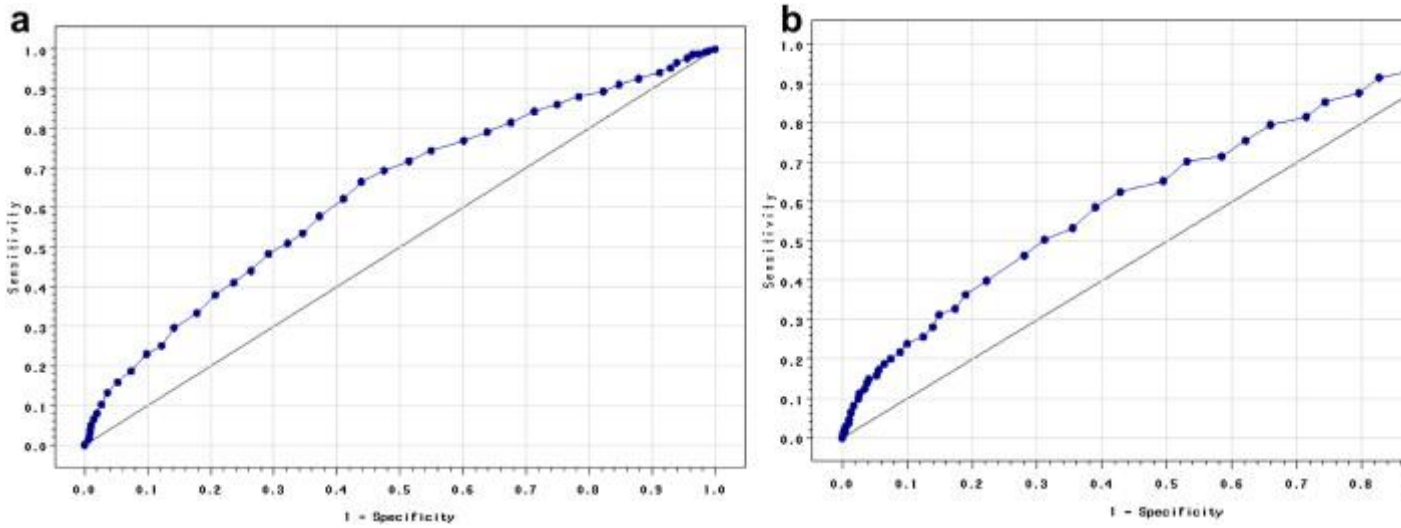

Fig. 1. Ability of pain and functional impairment severity to predict indication for TJR in 1909 hip or knee OA patients. (a) pain, AUC for curve: 0.64 (95% confidence interval, 0.61–0.67), (b) functional disability, AUC for curve: 0.63 (95% confidence interval, 0.60–0.66).

View thumbnail images

View high quality image (281K)

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 comorbidities 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).

Excluding patients from the United Kingdom did not modify the conclusions (data not shown).

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 (Fig. 2).

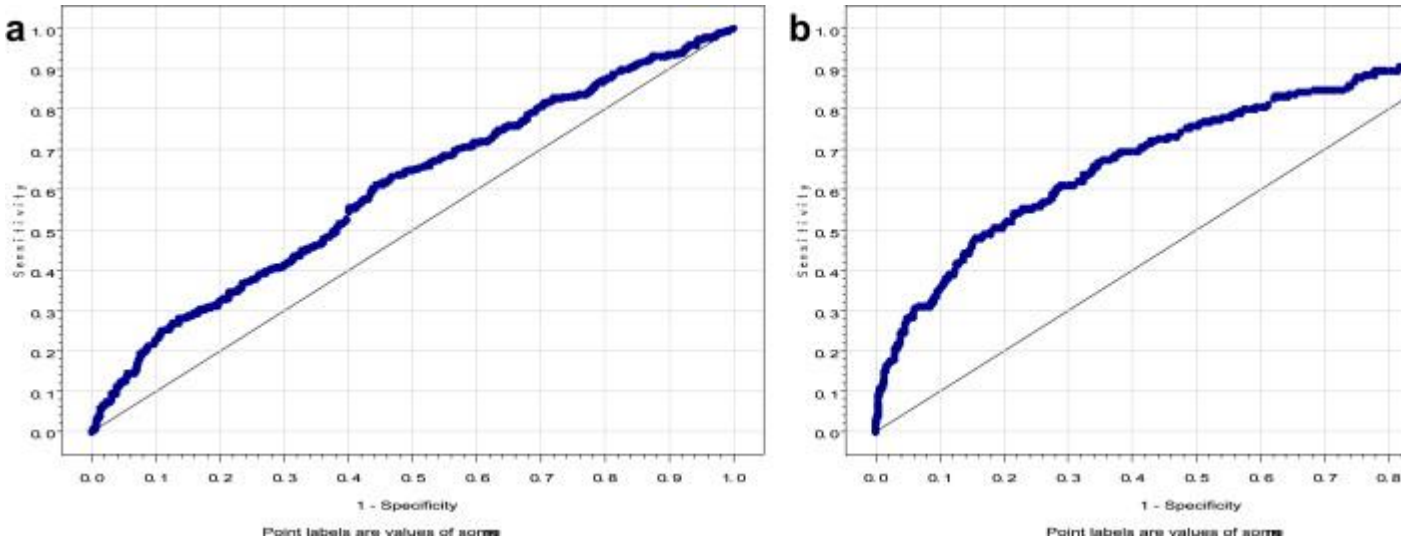

Fig. 2. Ability of the sum (pain + function) to predict indication for TJR, in knee and hip OA separately (a) Knee, AUC for the ROC curve: 0.60 (95% confidence interval, 0.56–0.64) (b) Hip, AUC for the ROC curve: 0.70 (95% CI, 0.66–0.75).

View thumbnail images

View high quality image (305K)

The AUCs of the ROC curves of the sum pain + function were higher in hip OA [AUC, 0.70, 95% confidence interval (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 cut point leading to a specificity of 90% was 66 (sum pain + function); for that cut point, 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, 0.75–0.86 in centers in Canada, France, Germany and Australia as compared with 0.54–0.67 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 orthopaedic surgeon, the level of symptoms was higher among patients for whom TJR was indicated by the orthopaedic 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 vs 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], [5], [6], [7], [8] and [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], [28], [29], [30], [31] and [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, Fig. 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], [31] and [32]. Possibly, the surgeons paid greater attention to the radiographic severity than to the symptom levels [10], [27] and [33] and several studies have indicated a discordance between radiographs and symptoms in lower-limb OA [27], [34], [35], [36] and [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], [18], [19], [20] and [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 subscales. 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], [39], [40] and [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 symptoms4. Finally, the surgeons may have based their surgical decision on joint mobility or peri-articular amyotrophy38, 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. In all, the current study confirms the wide variability in the indication for TJR suggested by studies of who actually receives a TJR; these results provide evidence that variability in surgeons' recommendations and practices is an important contributor to the clinical variability among TJR recipients.

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] and [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] and [28]. We considered that if surgery was recommended or if the surgeon considered symptoms were severe enough for surgery (although because of comorbidity10 or patient refusal9, 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], [3], [4], [5], [6], [7], [8], [9] and [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.

Sources of support

This work was supported by unrestricted grants from both scientific societies (OARSI and OMERACT) and pharmaceutical companies: Pfizer, Expansciences, Novartis, Negma Lerads, Rottapharm, Fidia and Pierre Fabre Santé laboratories). LSL was supported by the Swedish Research Council. JAS was supported by the National Institute of Health (NIH) Clinical Translational Science Award 1 KL2 RR024151-01 (Mayo Clinic Center for Clinical and