Short-term clinical worsening is a clear predictor for worsening at 2 years in established knee and hip osteoarthritis

E.A.M. Mahler[1], A.A. den Broeder[1,2], N. den Broeder[1], J.W.J. Bijlsma[3], G.F. Snijders[1,2], F.H.J. van den Hoogen[1,2], C.H.M. van den Ende[1,2]

[1]Department of Rheumatology, Sint Maartenskliniek, Nijmegen, The Netherlands; [2]Department of Rheumatology, Radboud University Medical Center, Nijmegen, The Netherlands; [3]Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands.

Abstract

Objective

Our aim was to estimate the proportion of knee and hip OA patients showing worsening at 2 years, and to examine the additional predictive value of failure of optimised non-surgical treatment during 3 months for worsening at 2 years.

Methods

Data of patients participating in the longitudinal CONTROL-PRO study (patients fulfilling clinical ACR criteria for knee or hip OA) were used. Measurements of pain, functioning and patient global assessments were performed at baseline, 3 months and 2 years. Worsening at 2 years was defined as fulfilling the recently validated clinical worsening criteria for knee and hip OA, or total joint replacement (TJR). Logistic regression was performed with worsening at 2 years as the dependent variable.

Results

The 297 included patients were predominantly women (66%) with a mean age of 55 years. At 2 years, 61% showed worsening (knee 59%; hip 71%) and 24% had undergone a TJR (knee 19%; hip 51%). Clinical worsening at 3 months appeared to be a clear independent predictor for worsening at 2 years (OR 2.8 95% CI 1.5–5.2) with a moderate discriminative ability (AUC 0.68 95% CI 0.57–0.70). Similar results were obtained when only TJR at 2 years was used as the outcome measure (OR 4.1 95% CI 2.0–8.4) with good AUC (0.82 95% CI 0.76–0.87).

Conclusion

Our findings suggest that re-assessment of symptoms after optimised non-surgical treatment could be meaningful in clinical decision making for TJR. Furthermore, this information could be used to identify subgroups of patients potentially eligible for novel and advanced treatment options.

Key words

predictors/prediction, symptomatic/clinical progression, worsening, knee, hip, osteoarthritis, non-surgical treatment, clinical decision making, outcome assessment/outcome measures, TJR, longitudinal study, AUC

Elien A.M. Mahler, MD, MSc, PhD
Alfons A. den Broeder, MD, PhD
Nathan den Broeder, BSc
Johannes W.J. Bijlsma, MD, PhD,
Gijs F. Snijders, MD, PhD
Frank H.J. van den Hoogen, MD, PhD
Cornelia H.M. van den Ende, PT, PhD

Please address correspondence to:
Dr Elien A.M. Mahler,
Department of Rheumatology,
Sint Maartenskliniek Nijmegen,
PO Box 9011,
6500 GM Nijmegen,
The Netherlands
E-mail: e.mahler@maartenskliniek.nl

Received on March 18, 2018; accepted in revised form on July 2, 2018.

Competing interests: none declared.

Introduction

Osteoarthritis (OA) is considered to be the most prevalent chronic joint disease and is one of the leading causes of pain and disability worldwide, with knee and hip being frequently involved joints (1-4). Meanwhile, the incidence and prevalence of knee and hip OA is rising substantially due to the ageing population and the epidemic of obesity which portends the associated future economic burden (3, 5, 6). The natural course of pain and physical functioning in knee and hip OA is highly variable: most patients have been found to remain stable, while a subset will gradually worsen (7-10). Both the involvement of a high socioeconomic burden as well as the variability on the natural course, mandate that identification of risk factors for clinical decline are important (11). This could be used to inform both patients and healthcare professionals, to identify patients at risk for deterioration in order to adapt treatment or to select individuals potentially eligible for novel therapies.

Longitudinal studies on validated clinical outcomes in knee and hip OA are lacking and therefore, little is known about the course and determinants associated with clinical deterioration of knee and hip OA (12). In contrast, many studies have been performed to determine prognostic factors for radiographic progression of knee and hip OA (8, 13). However, a clear discordance between radiographic and symptomatic knee OA has been well established. This highlights the need to also focus on symptomatic rather than radiological outcomes (7, 12, 14, 15). Symptomatic progression of knee OA is most relevant for both patient and healthcare professionals. Therefore, an understanding of the risk factors that predict clinical worsening in knee and hip OA would be useful to give insight in daily clinical practice.

Validated clinical worsening criteria have not been available up to recently. This is corroborated by two recently published systematic reviews of prognostic factors for symptomatic progression of knee OA, concluding that it was impossible to properly summarise the evidence due to different ways of measuring clinical progression (12, 14). Recently, we validated clinical worsening criteria that have been proposed to identify patients who have been deteriorated, enabling longitudinal outcome studies on determinants for clinical worsening over time of knee and hip OA (16), which corresponds to the current opinion to use symptom progression as outcome measure (12, 13).

Several international consensus-based clinical guidelines for the management of knee and hip OA are available, emphasising the importance and efficacy on non-surgical treatment modalities, which include education, exercise, step up analgesics, life style advice concerning physical activity and advice on weight loss in patients that are overweight (17). An important issue in clinical practice would be to evaluate whether failure of optimal standardised non-surgical treatment, is an additional risk factor for worsening over time, beyond history taking and physical examination. Therefore, the aims of this study are to estimate 1) the proportion of knee and hip OA patients showing worsening at 2 years, and 2) to examine the additional predictive value of failure of optimised standardised non-surgical treatment during 3 months for worsening at 2 years.

Materials and methods

Design, setting and participants

This study is part of the longitudinal study CONTROL-PRO (Cohort Of Non-invasively Treated Osteoarthritis of Lower Extremities – Pain, function and Radiological Outcome) (18). Consecutive patients of a specialised knee and hip OA outpatient clinic were invited to participate. All patients fulfilled the clinical American College of Rheumatology (ACR) criteria for knee or hip OA and were at inclusion deemed ineligible for total joint replacement (TJR) by their orthopaedic surgeon OA. The most symptomatic knee or hip at baseline was considered the index joint.

All patients received standardised non-surgical treatment during the first 3 months which included education, referral for physical therapy (aerobic and strengthening exercises), step-up analgesics using acetaminophen based

on the numeric rating scale (NRS) pain (patients were contacted every 4 weeks; next step only if NRS pain >4 unless contraindicated), followed by a first NSAID, substitution of NSAID and tramadol thereafter), and advice on weight reduction if indicated (goal 5% weight loss when BMI ≥28), as described elsewhere (18). Exclusion criteria were: other rheumatic or severe orthopaedic diseases leading to inflammatory arthritis or secondary OA, comorbidity exceeding the complaints or limitations of the knee or hip OA (paralysis due to a cerebrovascular event, severe fibromyalgia), orthopaedic procedures planned within the next three months, or cognitive or sensorimotor problems interfering with questionnaire completion. For the current study, patients were invited for a follow-up visit after 2 years if they completed both baseline and 3 months follow-up visits, and were included when they indeed completed the 2-year follow-up visit. The local Medical Research Ethics Committee, region Arnhem-Nijmegen (The Netherlands) approved the study design (study number 2009/095). The procedures followed were in accordance with the Declaration of Helsinki. All patients signed informed consent.

Data acquisition

Visits were scheduled at baseline, at 3 months, and at 2 years. Two-year visits were scheduled only for those patients who completed both baseline and 3-month questionnaires, and had not undergone a TJR. At inclusion, demographic and OA-related characteristics were collected, using a standardised interview and physical examination as described elsewhere (18). The number of comorbidities was assessed using the long version of the Dutch Arthritis Impact measurement Scales (19). At baseline, erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) were measured and radiographs were assessed. To examine structural abnormalities, knee (weight-bearing posterior-anterior fixed flexion) or hip radiographs (both anterior-posterior supine position) were obtained. Scoring of the index joint was completed by an experienced rheumatologist blinded to clinical data, using the atlas based Kellgren and Lawrence (K&L) grading system (20, 21). Previous intraobserver reliability (kappa) for K&L score ranged from 0.68 to 0.89 (re-scored in 20 participants (18). At all visits, patients completed a standardised set of patient reported outcomes measures.

Patient-reported Outcome Measures (PROMs)

Pain intensity and the patient global assessment (PGA) of OA impact during the last week were measured on a 0–10 point numeric rating scale (NRS) where 0 equals no symptoms. Patients also completed the Dutch Knee/Hip injury and Osteoarthritis Outcome Score (KOOS/HOOS) questionnaire (Likert scale version) (22). From the KOOS/HOOS, the Western Ontario and McMaster Universities (WOMAC) scores can be derived, with WOMAC pain, function, and stiffness subscales presented as standardised scores ranging from 0 to 100, where 100 equals no symptoms (23). Fatigue was measured at baseline and 3 months with the 8-itemed "Subjective Fatigue" subscale of the Checklist Individual Strength (CIS) (24). The total score can range from 8–56 points where scores of ≥35 represent severe fatigue. Fear of movement was measured with the Tampa Scale for Kinesiophobia (TSK) (25), where scores >37 represent excessive fear of movement. Mental health was measured with the mental component score and calculated with corresponding subscales of the 36-item Short Form Health Survey (SF-36), a widely used generic health status questionnaire comprising eight areas of health status, with higher scores indicating better health (range 0–100) (26, 27).

Primary outcome

Worsening at 2 years was operationalised as TJR in the index joint, or fulfilling recently validated clinical worsening criteria for knee and hip OA: worsening in: pain ≥20% and absolute change ≥20 *or* function ≥10% and absolute change ≥10 *or* PGA ≥10% and absolute change ≥1 (scale 0-10) compared to baseline values (16). We used two different pain outcome measures – NRS and WOMAC – and consequently two distinct sets of worsening criteria *i.e.* worsening using NRS pain and worsening using WOMAC pain respectively. As described elsewhere, these literature- and expert-group-based worsening criteria were first tested in a derivation cohort (n=219) and confirmed in a validation cohort (n=296). Both datasets incorporated observational data of patients with knee and hip OA who received standardised conservative treatment. This set performed best regarding sensitivity (59%) and specificity (74%) (16). Clinical worsening at 3 months was dichotomised similarly, but with different time points *i.e.* change between 3 months and baseline values.

Statistical analysis

• *Patient characteristics, follow-up at 3 months and 2 years*

Descriptive statistics were provided as mean and standard deviation (SD) or median and interquartile range (IQR) or numbers with percentages when appropriate. T-tests or chi-squared tests were used to compare baseline and 3 month values between patients who were included in the current analysis and patients who were lost to follow-up and to compute differences in knee and hip OA patients for worsening at 2 years.

• *Additional predictive value of failure of optimised non-surgical treatment during 3 months*

Multivariate logistic regression analyses were performed with clinical worsening at 2 years as the dependent variable. According to the TRIPOD statement, backward logistic regression analyses – guided by the Akaike information criterion (p=0.157) – was used to build the full baseline model for worsening at 2 years (dependent variable) (28). Based on the literature and clinical relevance, the following independent variables were selected: age, gender, BMI, affected joint, comorbidities, pain, PGA, function, stiffness, CRP, fatigue, mental component scale of SF-36 and K&L score and used in developing the full multivariable model. Separate models were run for NRS pain and for WOMAC pain, where both the independent variable for pain and

the outcome measure for pain differed. The results are presented as odds ratio (OR) with 95% confidence interval (95% CI). Secondly, clinical worsening at 3 months despite optimised non-surgical treatment was added as the independent variable to the model. For including variables, we used a rule of thumb as recommended by various authors (29), that a minimum of 10 events per variable is required to obtain a reliable and concise prediction model. To reduce the impact of missing data, data at baseline and 3 months was imputed using multiple imputations to create 20 datasets and results were combined using Rubin's rules (30, 31). The discriminatory ability of the final model was estimated using the area under (AUC) the receiver operating characteristic (ROC) curve, which is similar to the concordance-statistic (c-statistic). An AUC of 1 indicates perfect discrimination, while an AUC of 0.5 indicates discrimination no better than chance. Moreover, the positive and negative predictive values (PPV and NPV) as well as sensitivity and specificity of fulfilling clinical worsening criteria at 3 months and worsening at 2 years were estimated. Finally, the pre- and post-test probability was calculated and considered clinically relevant when the increase was above 15% (32, 33). Furthermore, we performed two sensitivity analyses; one with TJR at 2 years as the dependent variable and one on the subgroup of patients with knee OA. All analyses were performed using STATA 13.1.

Results

Patient characteristics

No relevant and significant differences were found between the patients included in the analyses (n=297) and the patients who did not reply to the invitation for the 2 years assessment (Supplementary Fig. 1, n=142, 32.4%) with regard to all baseline values presented in Table I, except for the proportion of patients with baseline K&L ≥2 (included 74% *vs*. not-replying 60%, p=0.06). Three months data were available for 54 out of 142 not-replying patients and we found no significant difference in proportion with clinical worsening between patients not-replying and patients included in the current analysis (24%; 95% CI 22–27% *vs*. 28%; 95% CI 25–31% respectively). The cohort consisted predominantly of women (66%), with a mean age of 55 years and median BMI of 28 kg/m^2 who are moderate to severely disabled by their disease considering the relatively high scores for pain and PGA, and the high proportion of patients showing fear of movement and severe fatigue.

Table I. Baseline characteristics of 297 patients with knee or hip osteoarthritis

Sociodemographic characteristics	
Age, years	55.0 (9.6)
Women, n (%)	195 (65.6)
Body mass index, kg/m^2, median (IQR)	27.9 (25.3 – 32.9)
Duration of symptoms , years, median (IQR)	3.8 (1.6 – 10.4)
Index joint knee, n (%)	252 (84.8)
Education, low/middle, n (%)	214 (72.1)
Comorbidities, >1, n (%)	119 (40.1)
Clinical parameters	
NRS pain (0-10)	5.9 (1.8)
NRS PGA (0-10)	6.1 (2.1)
WOMAC pain (0-100)	48.8 (19.2)
Function (0-100)	47.9 (19.7)
Stiffness (0-100)	45.9 (23.4)
ESR (mm/h), above upper limit, n (%)	30 (12.1)
C-reactive protein above upper limit, n (%)	16 (6.5)
Severe fatigue (CIS ≥35), n (%)	114 (43.9)
Fear of movement (TSK >37), n (%)	147 (49.5)
SF-36 mental component score (range 2-71)	51.2 (11.6)
Kellgren and Lawrence ≥2, n (%)	209 (74.1)[#]

Values are mean (SD) unless stated otherwise; IQR: interquartile range; NRS: numeric rating scale; PGA: patient global assessment; WOMAC pain, function and stiffness, Western Ontario and McMaster University Osteoarthritis Index scale; ESR: Erythrocyte sedimentation rate; CIS: Checklist Individual Strength; TSK: Tampa Scale for Kinesiophobia; SF-36: Short Form 36 Health Survey; Higher scores indicate more NRS pain, worse PGA, better scores for WOMAC pain, function and stiffness, better mental and physical health SF-36.
[#]15 missing values.

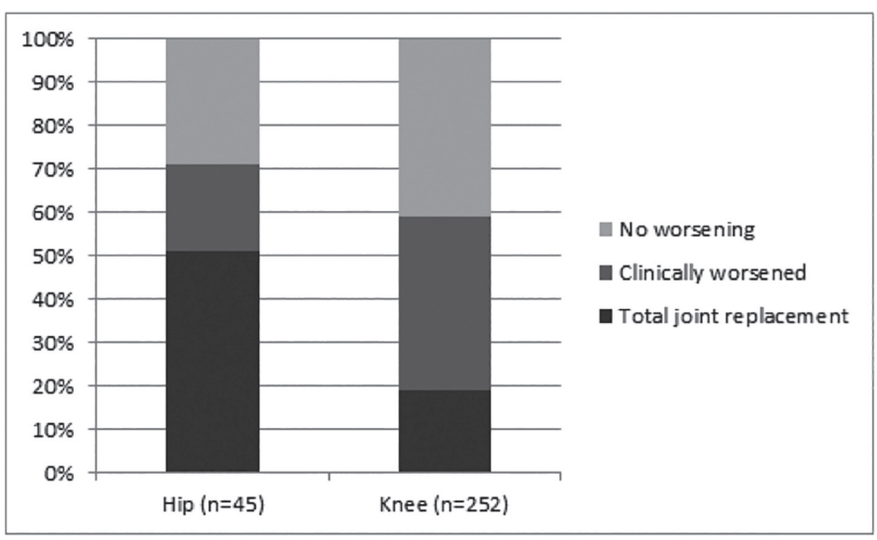

Fig. 1. Proportions of worsened patients at 2 years per index joint.
Testing proportions between knee and hip OA patients: no worsening; p=0.13, clinically worsened; p<0.01, and total joint replacement; p<0.0001.

Follow-up at 3 months and 2 years

Of the 297 patients in the cohort, a total of 85 (29%) and 181 patients (61%) clinically worsened at 3 months and 2 years respectively (see also Table III). A total of 71 out of 181 patients (39%) who worsened at 2 years – *i.e*. 24% of

Table II. Multivariate logistic regression model with determinants presented as OR (95% CI) for worsening at 2 years in knee and hip OA patients.

	Worsening using NRS pain (n_{yes}=181; n_{no}=116)			Worsening using WOMACpain (n_{yes}=173; n_{no}=123)*		
	OR	95% CI	AUC	OR	95% CI	AUC
Baseline model			0.63 (0.57-0.70)			0.67 (0.61-0.73)
Patient global assessment (range 0-10)	0.81	0.71-0.93		0.85	0.75-0.96	
Kellgren & Lawrence score ≥2 (range 0-4)	1.62	0.94-2.81		2.13	1.20-3.78	
WOMAC stiffness (0-100)	0.99	0.98-1.00				
BMI (kg/m^2)				1.06	1.01-1.10	
Hip index joint				1.98	0.97-4.06	
Baseline model with clinical worsening at 3 months			0.68 (0.62-0.74)			0.70 (0.64-0.76)
Clinical worsening at 3 months	2.84	1.54-5.22		2.47	1.34-4.55	
Patient global assessment baseline (range 0-10)	0.83	0.72-0.96)		0.87	0.77-0.99	
Kellgren & Lawrence scale ≥2 (range 0-4)	1.69	0.96-2.99		2.26	1.26-4.06	
WOMAC stiffness baseline (0-100)	0.99	0.98-1.00				
BMI baseline (kg/m^2)				1.05	1.01-1.10	
Hip index joint				1.96	0.95-4.07	

OA: osteoarthritis; n: number; OR: odds ratio; CI: confidence interval; AUC: area under ROC curve/c-statistic; NRS: numeric rating scale; WOMAC: Western Ontario McMaster University Index of osteoarthritis; PGA: patient global assessment; BMI: body mass index.
*Total number of patients with complete WOMACpain = 296, due to missing values.
Higher scores indicate worse PGA, more NRS pain, and better scores for WOMAC stiffness

Table III. 2 x 2 table of clinical worsening at 3 months and worsening at 2 years.

		Worsening at 2 years		
		Yes	No	Total
Clinical worsening at 3 months	Yes	66	19	85
	No	115	97	212
	Total	181	116	297

Values are n (%).
Positive predictive value 77.6% (95% CI 67.1-85.7); negative predictive value 47.8% (95% CI 39.0-52.7); sensitivity: 36.5% (95% CI 29.5-44.0); specificity: 83.6% (95% CI 75.3-89.6); positive likelihood ratio 2.23 (95% CI 1.41-3.51); negative likelihood ratio 0.76 (95% CI 0.68-0.85).

the whole group of patients – had undergone a TJR in the index joint on average 1.1 years (SD 0.5) after inclusion. As shown in Figure 1, a higher proportion of knee OA patients showed clinical worsening at 2 years compared to hip OA patients (40% *vs*. 20%, $p<0.01$). However, the proportion of patients who underwent a TJR was lower for knee than hip OA (19% *vs*. 51% respectively, $p<0.0001$). There was no difference in the proportion of patients who did not worsen at 2 years for those with knee compared with hip OA (41% and 29%, respectively, $p=0.13$). Of 85 patients who demonstrated clinical worsening at 3 months; 29 (34%) had progressed to TJR, and 36 (42%) maintained their "clinically worsened" classification at 2 years. Median BMI did not change from baseline to 3 months and 2-year follow-up.

Additional predictive value of failure of optimised non-surgical treatment during 3 months

The prediction models for worsening at 2 years (defined as clinical worsening or TJR) are shown in Table II. Significant independent baseline predictors were: PGA and K&L score ≥2. Furthermore, BMI turned out to be an independent predictor when using WOMAC, but not NRS pain. The higher the baseline BMI, the greater the risk of worsening at 2 years. Adding clinical worsening at 3 months (yes/no) as the independent variable to the baseline model, resulted in an additional predictor for worsening at 2 years with an adjusted OR of 2.8 (95% CI 1.5–5.2 in NRS pain model). Overall, the discriminative ability of the model was fair with an AUC of 0.68 (95% CI 0.62–0.74), indicating a moderate ability to discriminate between patients with and without (clinical) worsening at 2 years (Fig. 2). Table III shows a high positive predictive value, low negative predictive value, low sensitivity, and high specificity for clinical worsening at 3 months and outcome at 2 years. The positive likelihood ratio of 2.2 (95% CI 1.4–3.5) suggests that taking clinical worsening at 3 months despite optimised non-surgical treatment into account increases the pre-test probability of 61% for worsening at 2 years to a post-test probability of 78% (Table III).

Sensitivity analyses using TJR as dependant outcome, yielded similar independent clear predictors for TJR: clinical worsening at 3 months, K&L score ≥2, and affected joint (Table IV). The adjusted OR for clinical worsening at 3 months for having a TJR at 2 years was 4.1 (95% CI 2.0–8.4 in NRS pain model). Overall, the discriminative ability of this model, showed a good AUC of 0.82 (95% CI 0.77–0.87), indicating good ability to discriminate between patients with and without TJR. Similar conclusions could be drawn for the sensitivity analyses using only knee OA patients (OR clinically worsened at 3 months 5.2 95% CI 2.2–11.9 and OR K&L score increased to 8.1 95% CI 2.2–29.3).

Table IV. Multivariate logistic regression model with determinants presented as OR (95% CI) for total joint replacement in knee and hip OA patients.

	Total joint replacement (using NRSpain) (n_{yes}=71; n_{no}=226)			Total joint replacement (using WOMACpain) (n_{yes}=71; n_{no}=226)		
	OR	95% CI	AUC	OR	95% CI	AUC
Baseline model			0.77 0.71-0.83			0.77 0.72-0.83
Patient global assessment (range 0-10)	1.20	1.01-1.42		1.15	0.98-1.36	
Kellgren & Lawrence score ≥2 (range 0-4)	4.32	1.66-11.26		4.38	1.20-3.78	
WOMAC stiffness (0-100)	0.98	0.97-1.00		0.99	0.97-1.00	
Hip index joint	6.61	3.03-14.42		5.60	2.62-11.95	
Comorbidities >1				1.90	0.88-4.09	
Duration of symptoms (years)	1.03	1.00-1.06		1.03	0.99-1.06	
BMI (kg/m^2)	1.04	0.99-1.09				
Age (years)	1.02	0.99-1.06				
Baseline model with clinical worsening at 3 months			0.82 0.76-0.87			0.82 0.77-0.87
Clinical worsening at 3 months	4.11	2.00-8.48		4.06	1.98-8.34	
Patient global assessment baseline (range 0-10)	1.30	1.08-1.55		1.28	1.07-1.54	
Kellgren & Lawrence score 2 (range 0-4)	4.72	1.79-12.43		3.24	1.84-11.95	
WOMAC stiffness baseline (0-100)	0.98	0.97-1.00		0.99	0.97-1.00	
Hip index joint	7.38	3.22-16.94		6.21	2.78-13.89	
Comorbidities >1				1.78	0.81-3.92	
Duration of symptoms (years)	1.03	1.00-1.07		1.03	0.99-1.07	
BMI (kg/m^2)	1.03	0.98-1.08				
Age (years)	1.03	1.00-1.06				

OA: osteoarthritis; n: number; OR: odds ratio; CI: confidence interval; AUC: area under ROC curve/c-statistic; NRS: numeric rating scale; WOMAC: Western Ontario McMaster University Index of osteoarthritis; PGA: patient global assessment; BMI: body mass index.
Higher scores indicate worse PGA, more NRS pain, and better scores for WOMAC stiffnes.

Discussion

To our knowledge, this is the first longitudinal study on the additional predictive value of failure of optimised non-surgical treatment during 3 months for worsening at 2 years in knee and hip OA. Our results show that more than half of patients with established knee and hip OA in secondary care have worsened at 2 years, despite optimised non-surgical treatment. We also found that clinical worsening at 3 months is a clear independent predictor for worsening at 2 years.

How could our results be used in clinical practice? Our results could be used for patient information and to guide both patients and (orthopaedic) surgeons in decision making about the appropriate timing of TJR. This study adds that patients who are clinically worsened at 3 months despite optimised non-surgical treatment, have an almost threefold increased odds ratio for worsening at 2 years. This corresponds to an increase of pre and post-test probability from 61 to 78%, which is above the considered clinically relevant cut-off of 15% improvement in probability of response after a positive test (32, 33).

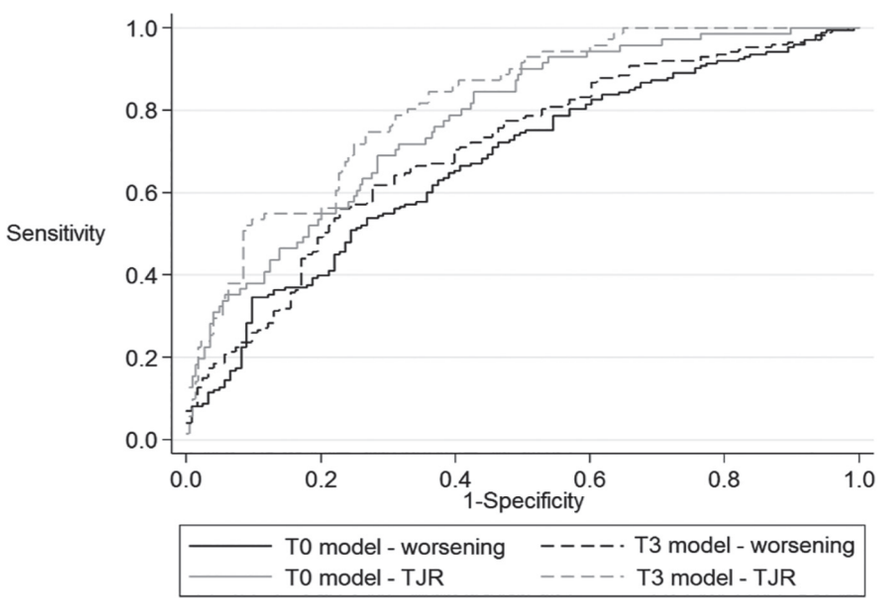

Fig. 2. ROC curves showing the sensitivity and 1-specificity for both the baseline model (T0) and the baseline model with clinical worsening at 3 months added (T3), for 2 different dependent outcome measures at 2 years: 1) worsening (TJR or fulfilling clinical worsening criteria, Table II), and 2) TJR (Table IV) in patients with knee or hip OA.
Pain was calculated using WOMAC pain.
WOMAC: Western Ontario and McMaster University Osteoarthritis Index scale; TJR: total joint replacement.

In addition, our worsening criteria are easy to assess in clinical practice. These advantages favour the use of clinical worsening criteria to monitor the symptoms of patients with established OA and suggest that patients with persisting symptoms after optimised non-surgical treatment, should be referred back to

the (orthopaedic) surgeon to reconsider the TJR indication. This predictor could be used to identify a more severely affected subgroup of patients that would be eligible for TJR (34).

Lastly, using clinical worsening criteria could support the identification of subgroups of patients potentially eligible for novel and advanced treatment options (35-38).

A remarkable finding is the relatively high proportion of worsened patients on the short term, compared with previous OA cohorts. This is not surprising, since most of the well-known OA cohorts focus on early OA patients (7, 8, 10, 14, 39). This difference is most likely explained by the selection of patients who were not yet deemed eligible for TJR by their surgeon. This homogenous population might have led to a selection of patients with a relatively high clinical burden and may hamper the generalisability of our results. Therefore, our study population is not representative for the general OA population, but generalisable to this more established OA population. Therefore, future research is warranted and should aim to investigate other OA populations and settings, for example in OA patients from primary care who are referred to secondary care.

An interesting finding is the lower proportion of TJR in knee OA patients compared to hip OA patients, whilst the total proportion of worsened patients between knee and hip OA patients was similar. This could be explained by the better longterm outcomes of a TJR of the hip than the knee (for example limited lifespan, and higher risk for serious adverse events for TJR of the knee) (38, 40-46), whereby clinicians might be more reluctant to decide for TJR in knee than hip OA. Furthermore, as expected, K&L score turned out to be a strong predictor, especially for predicting TJR at 2 years, which might be explained by the influence the K&L score has on the decision of an orthopaedic surgeon to propose a TJR. Moreover, a remarkable finding is the direction of the association between baseline PGA and the probability of worsening at 2 years. Where one might expect a worse PGA to correspond to a higher chance of worsening at 2 years, the opposite was found. This finding might be explained by regression to the mean effect.

Several strengths of this study should be considered. Overall, our study was well-powered and we chose a validated dichotomous measure for worsening combining arthroplasty with clinical worsening at 2 years incorporating the domains pain, function and PGA, the outcome measures advised according to the current opinion to use symptom progression as outcome measure (12, 13). Considering our homogenous population, the results of our study seem to be generalisable to patients with established knee and hip OA for whom decision making about TJR is forthcoming.

Some limitations that we faced should be reflected on. Firstly, estimated risks from prediction models have the tendency to be overestimated and thus further validation is required. In addition, most of the patients in our study had knee OA, so generalisability for hip OA is not assured. Furthermore, the substantial proportion of patients not replying to our invitation for the 2-year assessment could have influenced the results, although this seems unlikely, considering the lack of relevant differences in most relevant baseline variables and the similarity in proportion of replying and non-replying patients showing clinical worsening at 3 months. In addition, while the non-replying rate was quite high, it was comparable with other OA studies in which patients are not remaining under medical treatment (39). Lastly, adherence to treatment in clinical practice is quite challenging (47, 48) and subsequently, non-adherence to treatment could have influenced our results. However, we can only speculate about the potential direction. Nevertheless, given the challenge of adherence in both our study and clinical practice, our results are more likely to be representative for daily clinical practice, which strengthens the external generalisability of our results.

In conclusion, in light of our findings, we suggest that re-assessment of OA symptoms after optimised non-surgical treatment could be meaningful for both patients and surgeons in clinical decision making for TJR. Furthermore, this information could be used to identify subgroups of patients potentially eligible for novel and advanced treatment options.

Acknowledgements

We thank all the patients who participated.

References

1. KURTZ S, ONG K, LAU E, MOWAT F, HALPERN M: Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am* 2007; 89: 780-5.
2. DESHPANDE BR, KATZ JN, SOLOMON DH *et al*.: Number of persons with symptomatic knee osteoarthritis in the us: impact of race and ethnicity, age, sex, and obesity. *Arthritis Care Res* (Hoboken) 2016; 68: 1743-50.
3. BIJLSMA JW, BERENBAUM F, LAFEBER FP: Osteoarthritis: an update with relevance for clinical practice. *Lancet* 2011; 377: 2115-26.
4. SINGH JA, VESSELY MB, HARMSEN WS *et al*.: A population-based study of trends in the use of total hip and total knee arthroplasty, 1969-2008. *Mayo Clin Proc* 2010; 85: 898-904.
5. MURPHY L, SCHWARTZ TA, HELMICK CG *et al*.: Lifetime risk of symptomatic knee osteoarthritis. *Arthritis Rheum* 2008; 59: 1207-13.
6. RAVI B, CROXFORD R, REICHMANN WM, LOSINA E, KATZ JN, HAWKER GA: The changing demographics of total joint arthroplasty recipients in the United States and Ontario from 2001 to 2007. *Best Pract Res Clin Rheumatol* 2012; 26: 637-47.
7. VAN DIJK GM, DEKKER J, VEENHOF C, VAN DEN ENDE CH: Course of functional status and pain in osteoarthritis of the hip or knee: a systematic review of the literature. *Arthritis Rheum* 2006; 55: 779-85.
8. WESSELING J, BIERMA-ZEINSTRA SM, KLOPPENBURG M, MEIJER R, BIJLSMA JW: Worsening of pain and function over 5 years in individuals with 'early' OA is related to structural damage: data from the Osteoarthritis Initiative and CHECK (Cohort Hip & Cohort Knee) study. *Ann Rheum Dis* 2015; 74: 347-53.
9. HOLLA JF, VAN DER LEEDEN M, HEYMANS MW *et al*.: Three trajectories of activity limitations in early symptomatic knee osteoarthritis: a 5-year follow-up study. *Ann Rheum Dis* 2014; 73: 1369-75.
10. WHITE DK, KEYSOR JJ, LAVALLEY MP *et al*.: Clinically important improvement in function is common in people with or at high risk of knee OA: the MOST study. *J Rheumatol* 2010; 37: 1244-51.
11. CONAGHAN PG, KLOPPENBURG M, SCHETT G, BIJLSMA JWJ: EULAR osteoarthritis ad hoc committee. *Ann Rheum Dis* 2014; 73: 1442-5.
12. BASTICK AN, RUNHAAR J, BELO JN, BIERMA-ZEINSTRA SM: Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies. *Arthritis Res Ther* 2015; 17: 152.
13. BELO JN, BERGER MY, REIJMAN M, KOES

BW, BIERMA-ZEINSTRA SM: Prognostic factors of progression of osteoarthritis of the knee: a systematic review of observational studies. *Arthritis Rheum* 2007; 57: 13-26.
14. DE ROOIJ M, VAN DER LEEDEN M, HEYMANS MW *et al*.: Prognosis of pain and physical functioning in patients with knee osteoarthritis: a systematic review and meta-analysis. *Arthritis Care Res* (Hoboken) 2016; 68: 481-92.
15. KIM C, NEVITT MC, NIU J *et al*.: Association of hip pain with radiographic evidence of hip osteoarthritis: diagnostic test study. *BMJ* 2015; 351: 5983.
16. MAHLER E, DEN BROEDER AA, WOODWORTH TG *et al*.: How should worsening in osteoarthritis be defined? Development and initial validation of preliminary criteria for clinical worsening in knee and hip osteoarthritis. *Scand J Rheumatol* 2017; 46: 396-406.
17. MCALINDON TE, BANNURU RR, SULLIVAN MC *et al*.: OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis Cartilage* 2014; 22: 363-88.
18. SNIJDERS GF, DEN BROEDER AA, VAN RIEL PL *et al*.: Evidence-based tailored conservative treatment of knee and hip osteoarthritis: between knowing and doing. *Scand J Rheumatol* 2011; 40: 225-31.
19. RIEMSA RP, TAAL E, RASKER JJ, HOUTMAN PM, VAN PAASSEN HC, WIEGMAN O: Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. *Br J Rheumatol* 1996; 35: 755-60.
20. ALTMAN RD, GOLD GE: Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage* 2007; 15 (Suppl A): A1-56.
21. KELLGREN JH, LAWRENCE JS: Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957; 16: 494-502.
22. DE GROOT IB, FAVEJEE MM, REIJMAN M, VERHAAR JA, TERWEE CB: The Dutch version of the Knee Injury and Osteoarthritis Outcome Score: a validation study. *Health Qual Life Outcomes* 2008; 6: 16.
23. MCCONNELL S, KOLOPACK P, DAVIS AM: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum* 2001; 45: 453-61.
24. VERCOULEN JH, SWANINK CM, FENNIS JF, GALAMA JM, VAN DER MEER JW, BLEIJENBERG G: Dimensional assessment of chronic fatigue syndrome. *J Psychosom Res* 1994; 38: 383-92.
25. KORI SH, MILLER RP, TODD DD: Kinesiophobia: A new view of chronic pain behavior. *Pain Management* 1990; 3: 35-43.
26. AARONSON NK, MULLER M, COHEN PD *et al*.: Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998; 51: 1055-68
27. WARE JE, JR., SHERBOURNE CD: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83
28. MOONS KG, ALTMAN DG, REITSMA JB *et al*.: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; 162: 1-73.
29. STEYERBERG EW: Clinical prediction models, New York: Springer-Verlag, 2009
30. STERNE JA, WHITE IR, CARLIN JB *et al*.: Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; 338: b2393.
31. VON HIPPEL PT: Regression with missing ys: an improved strategy for analyzing multiply imputed data. *Sociological Methodology* 2007; 37: 83-117.
32. CUPPEN BV, WELSING PM, SPRENGERS JJ *et al*.: Personalized biological treatment for rheumatoid arthritis: a systematic review with a focus on clinical applicability. *Rheumatology* (Oxford) 2016; 55: 826-39.
33. DE VET HCW, TERWEE CB, MOKKINK LB, KNOL DL: Measurement in medicine. Cambridge University Press 2011.
34. FERKET BS, FELDMAN Z, ZHOU J, OEI EH, BIERMA-ZEINSTRA SM, MAZUMDAR M: Impact of total knee replacement practice: cost effectiveness analysis of data from the Osteoarthritis Initiative. *BMJ* 2017; 356: 1131.
35. FRENCH SD, BENNELL KL, NICOLSON PJ, HODGES PW, DOBSON FL, HINMAN RS: What do people with knee or hip osteoarthritis need to know? An international consensus list of essential statements for osteoarthritis. *Arthritis Care Res* (Hoboken) 2015; 67: 809-16.
36. HAGEN KB, SMEDSLUND G, OSTERAS N, JAMTVEDT G: Quality of community-based osteoarthritis care: a systematic review and meta-analysis. *Arthritis Care Res* (Hoboken) 2016; 68: 1443-52.
37. FELSON DT: Clinical practice. Osteoarthritis of the knee. *N Engl J Med* 2006; 354: 841-8.
38. SKOU ST, ROOS EM, LAURSEN MB *et al*.: A randomized, controlled trial of total knee replacement. *N Engl J Med* 2015; 373: 1597-606.
39. EYLES JP, MILLS K, LUCAS BR *et al*.: Can we predict those with osteoarthritis who will worsen following a chronic disease management program? *Arthritis Care Res* (Hoboken) 2016; 68: 1268-77.
40. HAWKER GA, BADLEY EM, BORKHOFF CM *et al*.: Which patients are most likely to benefit from total joint arthroplasty? *Arthritis Rheum* 2013; 65: 1243-52.
41. CUSHNER F, AGNELLI G, FITZGERALD G, WARWICK D: Complications and functional outcomes after total hip arthroplasty and total knee arthroplasty: results from the Global Orthopaedic Registry (GLORY). *Am J Orthop* (Belle Mead NJ) 2010; 39: 22-8.
42. BOURNE RB, CHESWORTH BM, DAVIS AM, MAHOMED NN, CHARRON KD: Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? *Clin Orthop Relat Res* 2010; 468: 57-63.
43. RISSANEN P, ARO S, SLATIS P, SINTONEN H, PAAVOLAINEN P: Health and quality of life before and after hip or knee arthroplasty. *J Arthroplasty* 1995; 10: 169-75.
44. BULLENS PH, VAN LOON CJ, DE WAAL MALEFIJT MC, LAAN RF, VETH RP: Patient satisfaction after total knee arthroplasty: a comparison between subjective and objective outcome assessments. *J Arthroplasty* 2001; 16: 740-7.
45. KATZ JN: Parachutes and preferences--a trial of knee replacement. *N Engl J Med* 2015; 373: 1668-9.
46. BESWICK AD, WYLDE V, GOOBERMAN-HILL R, BLOM A, DIEPPE P: What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* 2012; 2: e000435.
47. PISTERS MF, VEENHOF C, DE BAKKER DH *et al*.: Behavioural graded activity results in better exercise adherence and more physical activity than usual care in people with osteoarthritis: a cluster randomised trial. *J Physiother* 2010; 56: 41-7.
48. BEINART NA, GOODCHILD CE, WEINMAN JA *et al*.: Individual and intervention-related factors associated with adherence to home exercise in chronic low back pain: a systematic review. *Spine J* 2013; 13: 1940-50.