RHEUMATOLOGY

Rheumatology 2023;62:147–157
https://doi.org/10.1093/rheumatology/keac292
Advance access publication 16 May 2022

Original article

Predicted and actual 2-year structural and pain progression in the IMI-APPROACH knee osteoarthritis cohort

Eefje M. van Helvoort [1], Mylène P. Jansen [1], Anne C. A. Marijnissen[1], Margreet Kloppenburg[2,3], Francisco J. Blanco[4], Ida K. Haugen[5], Francis Berenbaum[6,7], Anne-Christine C. Bay-Jensen[8], Christoph Ladel[9], Agnes Lalande[10], Jonathan Larkin[11], John Loughlin[12], Ali Mobasheri[13,14,15,16,17], Harrie H. Weinans[17], Pawel Widera[18], Jaume Bacardit[18], Paco M. J. Welsing[1] and Floris P. J. G. Lafeber[1]

Abstract

Objectives. The IMI-APPROACH knee osteoarthritis study used machine learning (ML) to predict structural and/or pain progression, expressed by a structural (S) and pain (P) predicted-progression score, to select patients from existing cohorts. This study evaluates the actual 2-year progression within the IMI-APPROACH, in relation to the predicted-progression scores.

Methods. Actual structural progression was measured using minimum joint space width (minJSW). Actual pain (progression) was evaluated using the Knee injury and Osteoarthritis Outcomes Score (KOOS) pain questionnaire. Progression was presented as actual change (Δ) after 2 years, and as progression over 2 years based on a per patient fitted regression line using 0, 0.5, 1 and 2-year values. Differences in predicted-progression scores between actual progressors and non-progressors were evaluated. Receiver operating characteristic (ROC) curves were constructed and corresponding area under the curve (AUC) reported. Using Youden's index, optimal cut-offs were chosen to enable evaluation of both predicted-progression scores to identify actual progressors.

Results. Actual structural progressors were initially assigned higher S predicted-progression scores compared with structural non-progressors. Likewise, actual pain progressors were assigned higher P predicted-progression scores compared with pain non-progressors. The AUC-ROC for the S predicted-progression score to identify actual structural progressors was poor (0.612 and 0.599 for Δ and regression minJSW, respectively). The AUC-ROC for the P predicted-progression score to identify actual pain progressors were good (0.817 and 0.830 for Δ and regression KOOS pain, respectively).

Conclusion. The S and P predicted-progression scores as provided by the ML models developed and used for the selection of IMI-APPROACH patients were to some degree able to distinguish between actual progressors and non-progressors.

Trial registration. ClinicalTrials.gov, https://clinicaltrials.gov, NCT03883568.

Key words: Knee osteoarthritis, clinical trials and methods, study design, biomarkers

CLINICAL SCIENCE

[1]Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, [2]Department of Rheumatology, [3]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, [4]Grupo de Investigación de Reumatologia (GIR), INIBIC-Complejo Hospitalario Universitario de A Coruña, SERGAS, Centro de Investigación CICA, Departamento de Fisiotherapia y Medicina, Universidad de A Coruña, A Coruña, Spain, [5]Division of Rheumatology and Research, Diakonhjemmet Hospital, Oslo, Norway , [6]Department of Rheumatology, AP-HP Saint-Antoine Hospital, [7]INSERM, Centre de Recherche Saint-Antoine, Sorbonne University, Paris, France, [8]Nordic Bioscience A/S, Herlev, Denmark, [9]BioBone B.V., Amsterdam, The Netherlands, [10]Institut de Recherches Internationales Servier, Suresnes, France, [11]GlaxoSmithKline USA, Collegeville, PA, USA, [12]Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK, [13]Research Unit of Medical Imaging, Physics and Technology, Faculty of Medicine, University of Oulo, Oulo, Finland, [14]Department of Regenerative Medicine, State Research Institute Centre for Innovative Medicine, Vilnius, Lithuania, [15]Department of Joint Surgery, First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China, [16]World Health Organization Collaborating Centre for Public Health Aspects of Musculoskeletal Health and ging, Liege, Belgium, [17]Department of Orthopedics, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands and [18]School of Computing, Newcastle University, Newcastle upon Tyne, UK

Submitted 13 January 2022; accepted 4 May 2022

Correspondence to: E. M. van Helvoort, UMC Utrecht, Rheumatology and Clinical Immunology, Postbus 85500, internal mail no. G02.232, 3508 GA Utrecht, The Netherlands.
E-mail: E.M.vanHelvoort-3@umcutrecht.nl

Rheumatology key messages

- Machine learning (ML) models might improve patient selection by increasing the number of progressors.
- ML derived predicted-progression score for structural damage is poorly able to identify actual progressors.
- ML derived predicted-progression score for pain is able to distinguish progressors from non-progressors.

Introduction

One of the major challenges in knee OA clinical trials is the selection of patients. Because actual cure is not anticipated, patients who will sufficiently progress without intervention are needed to provide an opportunity to observe arrest or reduction of disease progression. Since progression in OA is on average (very) slow, without preselection of fast progressive patients, clinical trials require large group sizes and long follow-up [1]. The Innovative Medicines Initiative Applied Public-Private Research enabling OsteoArthritis Clinical Headway (IMI-APPROACH) consortium brings together European clinical centres, basic research institutes, business, pharmaceutical companies and patients [2]. As part of the consortium an international multicentre observational prospective clinical study was performed. A two-step selection method using machine learning (ML) models was used to include participants from five European observational OA cohorts (CHECK [3], HOSTAS [4], MUST [5], PROCOAC [6] and DIGICOD [7]) with an increased likelihood of structural and/or pain progression. Approximately 50% of the included patient were selected from the CHECK cohort. The selected participants were followed for 2 years. Conventional and new disease markers were combined to evaluate the innovative selection procedure and identify different OA phenotypes.

To select the most eligible patients after screening, a minimal data set including WOMAC pain and multiple radiographic features was used in the ML models to assign a structure (S) predicted-progression score and a pain (P) predicted-progression score to each individual. The S predicted-progression score (range 0–1) reflected the predicted structural progression (higher value means higher chance of fulfilling the progression criterion) during the follow-up period with minimum joint space width (minJSW), osteophyte size and varus angle being the most impactful features in determining the S score. Likewise, the P predicted-progression score (range 0–1) reflected the predicted pain progression (higher value means higher chance to fulfil one of the progression criteria) during the follow-up period. For the P score, the four WOMAC subscales were the most impactful features [8]. Both scores were normalized and combined into one score for ranking purposes, where the patient with the highest ranking score has the highest chance of showing actual structural and/or pain progression. The ~75% of patients with the highest ranking scores were selected. Details of the selection procedure and a cohort profile were described previously [9].

At baseline, IMI-APPROACH patients with a high S predicted-progression score were found to have a higher minJSW compared with patients with a low S predicted-progression score [3.6 mm (95% CI 3.4, 3.7 mm) and 1.6 mm (95% CI 1.4, 1.8 mm) for high *vs* low S predicted-progression score, respectively], providing an opportunity to actually show structural progression. Patients with a high P predicted-progression score reported more pain compared with patients with a low P predicted-progression score [Knee injury and Osteoarthritis Outcomes Score (KOOS) 51.2 (95% CI 48.2, 54.2) and 82.1 (95% CI 79.6, 84.6) for high *vs* low P predicted-progression score, respectively], and with that, a smaller window to actually show pain progression. This combination provides potentially a good patient selection for the evaluation of new treatment modalities intended to decrease pain and to prevent, stop or slow-down progression of structural damage [10].

This present study evaluates the actual 2-year structural and pain progression within the IMI-APPROACH prospective cohort in relation to the initially assigned S and P predicted-progression score. The final aim was to determine whether the originally designated S and P predicted-progression scores are able to distinguish between actual structural/pain progressors and non-progressors. If so, the selection procedure used for IMI-APPROACH might be a first step to enrich selection of structural and/or pain progressors in future clinical trials.

Methods

Patient selection

All patients of the IMI-APPROACH study [$n = 297$ at baseline; mean age 66.5 years (S.D. 7.1), mean BMI 28.1 kg/m^2 (S.D. 5.3), female/male ratio 230/67, mean minJSW 2.5 (95% CI 2.4, 2.7), mean KOOS pain 66.4 (95% CI 64.2, 68.5) [9]] of whom right follow-up data were available were included in the current analyses. The study is being conducted in compliance with the protocol, Good Clinical Practice (GCP), the Declaration of Helsinki, and the applicable ethical and legal regulatory requirements (for all countries involved). The study is registered under clinicaltrials.gov nr: NCT03883568. All patients have received oral and written information and provided written informed consent.

Evaluation of structural damage

For each participant an index knee was selected based on American College of Rheumatology clinical criteria

for knee OA [11], using history and physical examination. If both knees fulfilled these criteria, the most painful knee according to the participant was designated as index knee. If both knees were equally painful, the right knee was chosen as index knee. Posterior–anterior weight-bearing semi-flexed knee radiographs according to the protocol of Buckland-Wright *et al*. [12] were analysed by Knee Image Digital Analysis (KIDA) [13] to determine minJSW.

Evaluation of pain

Pain was evaluated using the pain subscale of the KOOS questionnaire [14]. This score uses nine questions for pain, each scored on a 5-point Likert scale. A normalized score was calculated where 0 means maximal pain and 100 means no pain. Since the ML models used WOMAC pain [15] collected in the original cohorts and the subsequent prospective clinical study used KOOS pain, additionally an alternative analysis was performed using the WOMAC pain scores deduced from the corresponding questions from the KOOS questionnaire as outcome (see Supplementary Data S4, available at *Rheumatology* online).

Data acquisition

Radiographs and questionnaires were obtained at screening/baseline (BL) (M000, 0 year), after 6 months (M006, 0.5 year), after 12 months (M012, 1 year), and after 24 –2/+6 months (M024, 2 year). The larger time window for the M024 visit was allowed due to COVID-19 limitations that arose during conduct of the study. As the variation in scores due to variation in the actual timing of M000, M006, M012 and M024 were deemed negligible compared with other sources of variation (e.g. acquisition [16], pain perception) no corrections were made for this variation in visit time. All available KOOS pain data were used. For minJSW data, the change over time for all patients was visually checked by two observers as reference for quality of data acquisition; in case of doubtful data sets over time (e.g. unexpected variations or extreme changes over time), the radiographs from which the data were deduced were checked as time series per patient and individual patient time point acquisitions were removed in case of incorrect image acquisition ($n = 18$ images in total). Additionally, this check revealed one incorrect reading ($n = 1$ image in total), which was removed. One patient was completely removed for analysis of structural progression because of doubtful image acquisitions at multiple time points.

Progression criteria

Based on literature, in IMI-APPROACH actual structural progression was pre-defined as: a reduction of minJSW over time at a threshold of 0.3 mm narrowing per year (i.e. $\geq$0.6 mm decrease in the 2-year follow-up period) [17]. Actual pain progression (on a 0–100 scale; 100 means no pain and 0 means maximum pain) was pre-defined as fast/significant pain increase and/or stable significant pain, as follows. Fast pain increase: KOOS pain decrease between baseline and follow-up $\geq$10 points per year (i.e. $\geq$20 points decrease in the 2-year follow-up period) and final KOOS pain score $\leq$65 points. Significant pain increase: KOOS pain decrease between baseline and follow-up $\geq$5 points per year (i.e. $\geq$10 points decrease in the 2-year follow-up period) and final KOOS pain score $\leq$60 points. Stable significant pain: KOOS pain score $\leq$60 points during the whole study period [9, 18].

Clearly, actual progression can be defined by several different parameters including multiple imaging modalities for S progression, and different pain and physical function outcome measurements for P progression. In the present study, evaluation was limited to the parameters that were originally used to define progression (clinicaltrial.gov no.: NCT0388568). Performance of the S and P predicted-progression scores for actual progression based on other outcome parameters is of relevance as well, but beyond the scope of the present study.

Data are presented as a change score based on actual values [throughout the paper referred to as delta (Δ) minJSW and ΔKOOS pain; M024 minus M000] and as a change over 2 years based on a per patient fitted regression line over time using the observed M000, M006, M012 and M024 values. Regression coefficients (increase per year) of these regression lines were multiplied by 2 to determine regression per 2 years (throughout the paper referred to as regression minJSW and regression KOOS pain). For ΔminJSW, patients were included if M000 and M024 minJSW data were present ($n = 224$). For regression minJSW, patients with minJSW available at no less than three of the four time points were included ($n = 266$). Likewise, for ΔKOOS pain, patients of which M000 and M024 KOOS pain data were present were included ($n = 246$). For regression KOOS pain, only patients with KOOS pain data available at M000, M024 and at least one other time point were included ($n = 246$), since KOOS pain at M024 is included in the definition (different from the definition for radiographic progression). There was a good correlation between Δ and regression-based change for both minJSW and KOOS pain, shown in Supplementary Data S1 Fig. S1a and b, available at *Rheumatology* online.

Statistical analyses

First, the number of actual structural progressors and actual pain progressors was determined for Δ and regression based change separately and described using frequency and proportions. Secondly, the S predicted-progression scores of the actual structural progressors and structural non-progressors, and P predicted-progression scores of the actual pain progressors and pain non-progressors were compared using means (S.D.) and by plotting the histogram of the S and P predicted-progression scores of progressors and non-progressors. Differences in S and P predicted-progression scores between actual progressors and non-progressors were tested for statistical significance using

an independent samples *t*-test. Thirdly, receiver operating characteristic (ROC) curves were constructed and the area under the ROC curve (AUC) was calculated to evaluate the discriminatory ability of the S and P predicted-progression scores for the actual structural and pain progression, respectively. Youden's index (YI) was used to determine the optimal cut-offs for both predicted-progression scores. All analyses were performed using IBM SPSS Statistics version 26.0.0.1 (IBM Corp., Armonk, NY, USA). *P*-values <0.05 were considered statistically significant.

Results

Actual structural and pain progressors in the IMI-APPROACH cohort

Of the 297 included patients, 249 (84%) completed the study. The S and P predicted-progression score did not statistically significantly differ between patients that completed the study and patients that withdrew. Main reason for withdrawal was unwillingness to visit the hospital during the COVID-19 pandemic.

Table 1A shows the numbers of patients with $\geq$0.6 mm loss of minJSW during the 2-year follow-up period (defined as actual structural progressors) and those with $<$0.6 mm minJSW loss during the 2-year follow-up period (defined as structural non-progressors).

Table 1B shows the numbers of patients fulfilling the criterion for *pain increase* (fast pain increase and significant pain increase combined) or *stable significant pain*, as well as both progression groups combined (defined as actual pain progressors), and those not fulfilling the progression criteria (defined as pain non-progressors).

Table 1C combines the criteria for actual structural and pain progression to evaluate the final number of progressor types within the IMI-APPROACH cohort.

Table 2 shows the baseline characteristics of actual structural progressors and non-progressors (Table 2A), actual pain progressors and non-progressors (Table 2B), and combination of both (Table 2C).

Differences in S/P predicted-progression score at baseline between the actual structural/pain progressors and non-progressors after 2 years

The S predicted-progression scores of actual structural progressors and non-progressors are shown in Fig. 1. As anticipated, actual structural progressors were assigned on average a statistically significantly higher S predicted-progression score at inclusion, as compared with non-progressors [0.426 (0.075) *vs* 0.396 (0.076) for ΔminJSW, $P = 0.023$; 0.427 (0.085) *vs* 0.397 (0.075) for regression minJSW, $P = 0.013$].

The P predicted-progression scores of patients with actual pain increase, patients with stable significant pain, total progressors and non-progressors are shown in Fig. 2. Mean P predicted-progression scores for actual pain progressors compared with pain non-progressors were 0.605 (0.179) *vs* 0.359 (0.201) for ΔKOOS pain and 0.612 (0.175) *vs* 0.354 (0.198) for regression KOOS pain, respectively (both $P < 0.001$).

Table 1 Actual progressors in the IMI-APPROACH cohort

A. Actual structural progressors

	Total	Non-progressors	Progressors
Δ minJSW	224 (75)	183 (81.7)	41 (18.3)
Regression minJSW	266 (90)	203 (76.3)	63 (23.7)

B. Actual pain progressors

	Total	Non-progressors	Progressors	Pain increase	Stable significant pain
Δ KOOS pain	246 (83)	181 (73.5)	65 (26.5)	25 (38.5)	40 (61.5)
Regression KOOS pain	246 (83)	179 (72.7)	67 (27.3)	28 (41.8)	39 (58.2)

C. Actual structural and/or pain progressors

	Total	Non-progressors	Radiographic progressors	Pain progressors	Radiographic + pain progressors
Δ	221 (74)	127 (57.5)	31 (14.0)	54 (24.4)	9 (4.1)
Regression	242 (81)	130 (53.7)	45 (18.6)	57 (23.6)	10 (4.1)

Values are given as *n* (%). Actual structural and pain progressors according to the definition described in the study protocol and above. The total cohort consisted of 297 patients; because of the COVID-19 pandemic a relatively large number of M024 visits were missed. KOOS: Knee injury and Osteoarthritis Outcomes Score; minJSW: minimum joint space width.

Table 2 Baseline characteristics of actual progressors and non-progressors in the IMI-APPROACH cohort

	n	Age, mean (s.d.), years	Female, *n* (%)	BMI, mean (s.d.), kg/m^2	minJSW, mean (s.d.)	KOOS pain, mean (s.d.)
A. Baseline characteristics of actual structural progressors						
Δ minJSW						
Determined	224	66.3 (7.0)	172 (77)	27.7 (5.0)	2.6 (1.2)	67.7 (19.0)
Non-progressors	183	66.2 (7.2)	138 (75)	27.5 (4.9)	2.5 (1.2)	67.7 (18.8)
Progressors	41	66.8 (6.5)	34 (83)	28.6 (5.2)	3.1 (1.0)	67.4 (20.2)
Regression minJSW						
Determined	266	66.3 (7.1)	205 (77)	28.0 (5.3)	2.6 (1.2)	67.2 (18.6)
Non-progressors	203	66.0 (7.2)	154 (76)	27.8 (5.2)	2.5 (1.2)	67.3 (18.7)
Progressor	63	67.5 (6.7)	51 (81)	28.6 (5.8)	2.9 (1.1)	66.9 (18.7)
B. Baseline characteristics of actual pain progressors						
Δ KOOS pain						
Determined	246	66.5 (7.0)	192 (78)	27.7 (5.1)	2.6 (1.2)	67.0 (19.0)
Non-progressors	181	66.7 (7.0)	145 (80)	27.1 (4.7)	2.7 (1.2)	72.2 (17.3)
Progressors	65	65.8 (7.0)	47 (72)	29.5 (5.8)	2.2 (1.3)	52.6 (16.2)
Pain increase	25	66.5 (7.9)	17 (68)	29.3 (5.5)	2.1 (1.1)	67.1 (12.6)
Stable significant pain	40	65.4 (6.4)	30 (75)	29.6 (5.9)	2.3 (1.4)	43.5 (10.5)
Regression KOOS pain						
Determined	246	66.5 (7.0)	192 (78)	27.7 (5.1)	2.6 (1.2)	67.0 (19.0)
Non-progressors	179	66.8 (7.0)	142 (79)	27.0 (4.7)	2.7 (1.2)	72.4 (17.2)
Progressors	67	65.8 (6.9)	50 (75)	29.6 (5.6)	2.3 (1.3)	52.5 (15.9)
Pain increase	28	66.0 (7.8)	21 (75)	29.7 (5.2)	2.1 (1.1)	64.6 (14.2)
Stable significant pain	39	65.6 (6.4)	29 (74)	29.5 (5.9)	2.4 (1.4)	43.8 (10.5)
C. Baseline characteristics of actual structural and/or pain progressors						
Δ						
Determined	221	66.3 (6.9)	171 (77)	27.7 (5.0)	2.6 (1.2)	67.7 (19.0)
Non-progressors	127	66.5 (7.1)	99 (78)	27.0 (4.4)	2.6 (1.2)	73.5 (17.2)
Radiographic progressors	31	66.9 (6.5)	27 (87)	27.6 (5.2)	3.1 (0.9)	73.0 (16.1)
Pain progressors	54	65.7 (7.0)	38 (70)	28.8 (5.8)	2.1 (1.3)	54.1 (15.1)
Radiographic + pain progressors	9	65.3 (5.8)	7 (78)	31.9 (4.3)	3.1 (1.3)	47.8 (21.5)
Regression						
Determined	242	66.5 (7.0)	188 (78)	27.7 (5.1)	2.6 (1.2)	67.3 (19.0)
Non-progressors	130	66.4 (7.1)	102 (79)	26.9 (4.3)	2.6 (1.2)	73.3 (17.3)
Radiographic progressors	45	67.6 (6.6)	36 (80)	27.5 (5.7)	3.1 (1.0)	71.8 (16.5)
Pain progressors	57	65.6 (7.1)	42 (74)	29.3 (5.8)	2.2 (1.2)	53.0 (15.0)
Radiographic + pain progressors	10	66.4 (6.6)	8 (80)	30.9 (3.9)	2.6 (1.5)	50.3 (20.5)

Baseline characteristics of actual structural and pain progressors. The total cohort consisted of 297 patients; because of the COVID-19 pandemic a relatively large number of M024 visits were missed. KOOS: Knee injury and Osteoarthritis Outcomes Score; minJSW: minimum joint space width.

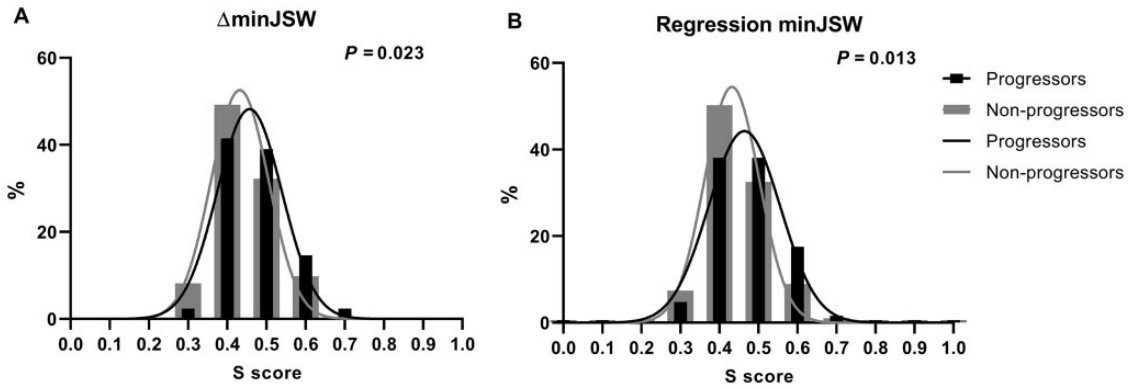
For patients with pain increase, P predicted-progression scores were 0.529 (0.197) for ΔKOOS pain and 0.555 (0.192) for regression KOOS pain. For patients with stable significant pain, mean P predicted-progression scores were 0.653 (0.150) for ΔKOOS pain and 0.652 (0.152) for regression KOOS pain (all $P < 0.001$ as compared with non-progressors).

Ability of S/P predicted-progression score to identify actual structural/pain progressors

ROC curves for the discrimination of actual progressors *vs* non-progressors by the S and P predicted-progression score are shown in Figs 3 and 4, respectively. The AUC of 0.612 and 0.599 for ΔminJSW and regression minJSW, respectively, indicate that the S predicted-progression score is poorly able to distinguish actual structural progressors from structural non-progressors. At the optimal cut-off according to the YI, the sensitivity and specificity were found to be 43.9% and 80.9% for ΔminJSW and 34.9% and 85.7% for regression minJSW, respectively.

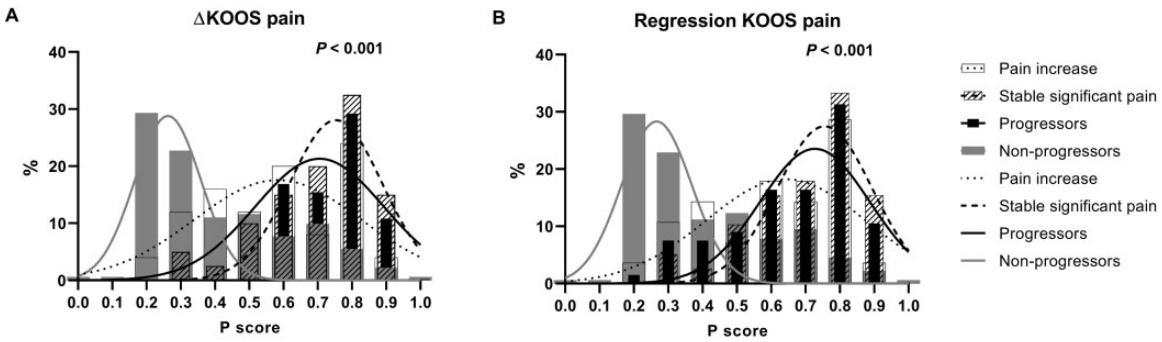
For the total group of actual pain progressors, the AUC was 0.817 and 0.830 for ΔKOOS pain and regression KOOS pain, respectively, indicating that the P predicted-progression score is much better able to distinguish between actual pain progressors and pain non-progressors then the S predicted-progression score is for actual structural progressors and non-progressors. When separately analysing patients with pain increase and patients with stable significant pain, the AUC was 0.664 and 0.689 for patients with pain increase, and 0.843 and 0.841 for patients with stable significant pain for ΔKOOS pain and regression KOOS pain,

Fig. 1 S predicted-progression score of actual radiographic progressors and non-progressors



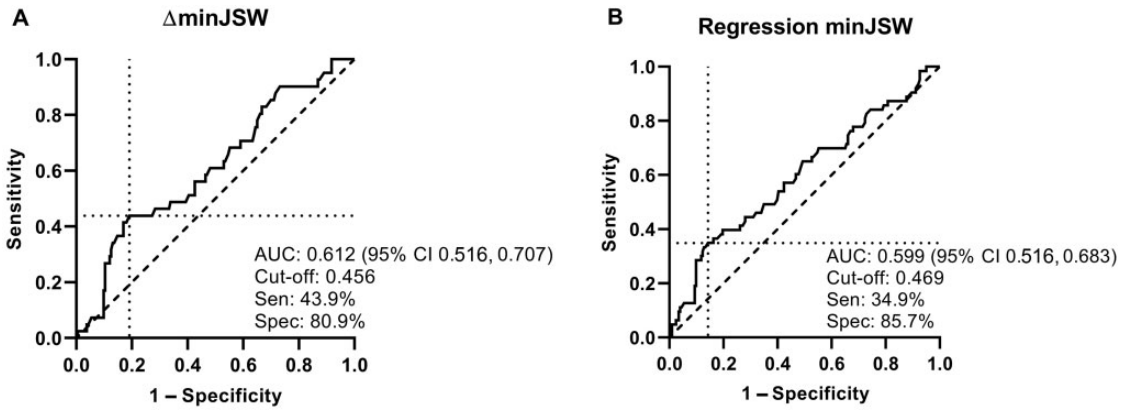
(**A**) S predicted-progression scores for actual radiographic progressors (absolute decrease in 2 years $\geq$0.6 mm, $n = 41$) and non-progressors ($n = 183$). (**B**) S predicted-progression scores for actual radiographic progressors (regression of each patient $\geq$0.6 mm/2 years, $n = 63$) and non-progressors ($n = 203$). minJSW: minimum joint space width.

Fig. 2 P predicted-progression score of actual pain progressors and non-progressors



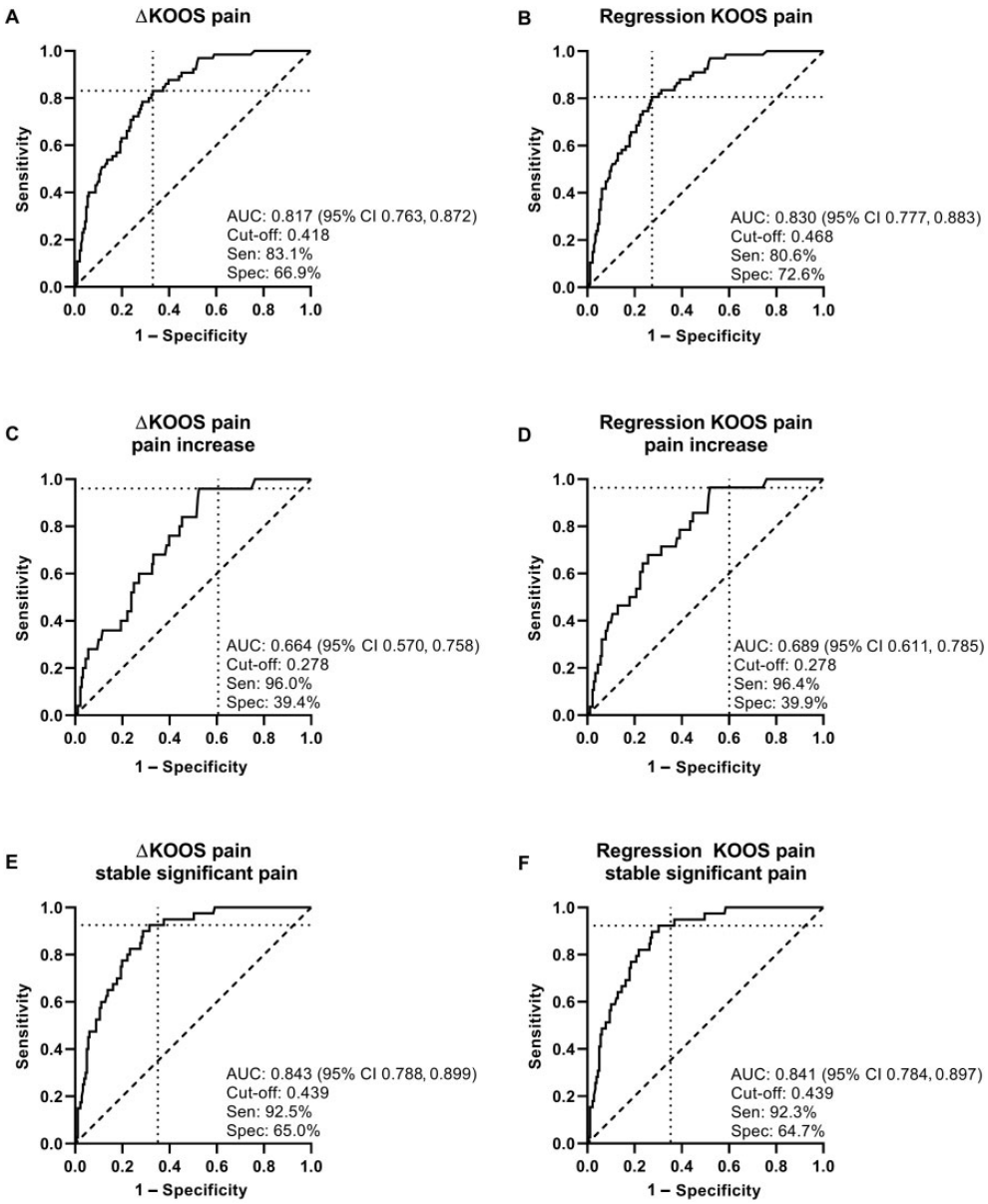
(**A**) P predicted-progression scores for actual pain progressors ($n = 65$; black) and non-progressors ($n = 181$; grey), as well as for patients with pain increase ($n = 25$; dotted) and patients with stable significant pain ($n = 40$; dashed) using the absolute decrease during the 2-year follow-up period. (**B**) P predicted-progression scores for actual pain progressors ($n = 67$; black) and non-progressors ($n = 179$; grey), as well as for patients with pain increase ($n = 28$; dotted line) and patients with stable significant pain ($n = 39$; dashed line) using the regression over 2 years of each individual patient. KOOS: Knee injury and Osteoarthritis Outcomes Score.

Fig. 3 ROC-curves S predicted-progression score



ROC curves for ΔminJSW (**A**) and regression minJSW (**B**). AUC: area under the curve; minJSW: minimum joint space width; ROC: receiver operating characteristic; Sen: sensitivity; Spec: specificity.

FIG. 4 ROC curves P predicted-progression score



ROC curves for ΔKOOS pain and regression KOOS pain for total progressors (**A**, **B**), patients with pain increase (**C**, **D**), and patients with stable significant pain (**E**, **F**). AUC: area under the curve; KOOS: Knee injury and Osteoarthritis Outcomes Score; ROC: receiver operating characteristic; Sen: sensitivity; Spec: specificity.

respectively, indicating that the P predicted-progression score is better at identifying patients with stable significant pain compared with patients with pain increase. At the optimal cut-off according to the YI, sensitivity and specificity were found to be 83.1% and 66.9% for ΔKOOS pain, and 80.6% and 72.6% for regression KOOS pain, respectively.

The YI was used to determine the optimal cut-off points of both predicted-progression scores, equally weighing false positives and false negatives, thereby giving the optimal combination of sensitivity and specificity. Table 3A for structure and Table 3A for pain provide alternative cut-off points, with corresponding sensitivity and specificity. See Supplementary Data S2 Tables S1a and b, available at *Rheumatology* online, for cut-off points for pain increase and stable significant pain separately.

Discussion

IMI-APPROACH participants fulfilling the progression criterion for structural progression showed statistically significantly higher S predicted-progression scores at

Table 3 Possible cut-off points for the S and P predicted-progression score

	Δ		Regression	
Cut-off	Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %
A. Possible cut-off point for the S predicted-progression score, minJSW				
0.318	95.1	10.9	90.5	10.3
0.346	90.2	24.6	84.1	23.2
0.361	80.5	33.9	76.2	33.0
0.374	68.3	44.8	69.8	41.9
0.391	58.5	53.0	60.3	52.2
0.407	48.8	62.3	49.2	61.1
0.432	43.9	72.7	41.3	72.9
0.462	36.6	83.1	34.9	83.7
0.520	7.3	90.7	12.7	91.1
B. Possible cut-off points for the P predicted-progression score, KOOS pain				
0.158	100.0	14.9	100.0	15.1
0.198	98.5	28.7	98.5	29.1
0.240	96.9	41.4	97.0	41.9
0.300	90.8	51.9	91.0	52.5
0.407	83.1	63.5	83.6	64.2
0.492	73.8	72.9	76.1	74.3
0.580	56.9	80.7	59.7	82.1
0.663	49.2	89.0	50.7	89.9
0.742	26.2	95.6	25.4	95.5

Cut-offs are based on percentile of the predicted-progression score. With a cut-off for the S predicted-progression score of 0.318 (10th percentile), 90% will be classified as progressor, and with a cut-off of 0.520 (90th percentile), 10% will be classified as progressor, etc. With a cut-off for the P predicted-progression score of 0.158 (10th percentile), 90% will be classified as progressor, and with a cut-off of 0.742 (90th percentile), 10% will be classified as progressor, etc.

inclusion, as compared with those who did not fulfil the criterion. Likewise, actual pain progressors showed statistically significantly higher P predicted-progression scores at inclusion compared with pain non-progressors, as did the two separate categories for pain progression: patients with pain increase and patients with stable significant pain. However, the AUC for the S predicted-progression score was poor, showing that this score may not sufficiently predict actual structural progression. In contrast, the P predicted-progression score was found to be able to predict actual pain progression or non-progression. However, the majority of actual pain progressors belonged to the subgroup with stable significant pain. When analysing the separate subgroups of pain progression, the AUC for stable significant pain is better than the AUC for pain increase. The ability of the P predicted-progression score to predict actual *worsening* of pain is comparable to the ability of the S predicted-progression score to predict worsening in structural damage.

For the selection procedure used in the IMI-APPROACH cohort, the S and P predicted-progression scores were combined into one ranking score to include patients of both progression types. As such, the present study is not a validation of the original ranking. The difference in performance between both predicted-progression scores, and the required improvements for the ML models, made evaluation of overall prediction in our opinion not useful. Besides, for clinical trials one might like to select primarily structural progressors, or primarily pain progressors, depending on treatment modality. Therefore, this study was limited to evaluation of the separate predicted-progression scores instead of evaluating the combined (ranking) score.

The S and P predicted-progression scores are continuous variables, but cut-offs can be chosen to use for the selection of patients (e.g. in clinical trials). In this study, the YI was used to determine the optimal cut-off points of both progression scores, equally weighing false positives and false negatives, thereby giving the optimal combination of sensitivity and specificity. The high specificity found for the S predicted-progression score means that most of the actual structural non-progressors indeed were assigned an S score below the cut-off value. However, the low sensitivity indicates that a minority of the actual structural progressors were assigned an S predicted-progression score above the cut-off value and the majority will be missed.

Whereas the S predicted-progression score showed high specificity, but low sensitivity for actual structural progression, the P predicted-progression score showed high sensitivity and specificity for actual pain progression. However, this mainly results from patients with stable significant pain. For the identification of patients with pain increase, specificity is low (<40%). As demonstrated by Table 3A and B and Supplementary Tables S1a and b and Data S2 (available at *Rheumatology* online), alternative cut-off points can be chosen to increase

sensitivity or specificity (always at the expense of the other one), and with that increase the usefulness of both predicted-progression scores to select patients for clinical trials, depending on the goal of the study.

It is anticipated that the outcome of this study is generalizable as patients were selected from multiple different European OA cohorts. Nonetheless, the main limitation of the approach is that the primary ML model used for the selection was built using historical data from CHECK. The other cohorts (HOSTAS, MUST, PROCOAC and DIGICOD) used different data collection protocols, leading to different types of historical data for each cohort. As a second step a screening visit was performed to collect the same up-to-date information of each individual (irrespective of the original cohort). Clearly, this second step only provided cross-sectional data, and not longitudinal data from which the ML models were generated. To improve future trials, a uniform data collection protocol, describing which data to collect at which time point could support implementation of the here-described selection procedure, selecting participants from multiple cohorts.

As said, the ML models constructing an S and P predicted-progression score for each individual, used for the initial selection of IMI-APPROACH patients, were trained on historical data from the Cohort Hip and Cohort Knee (CHECK) [3]. Approximately 50% of IMI-APPROACH were recruited from the original CHECK cohort. For that reason, we also performed analyses including only patients recruited from CHECK ($n = 153$) or excluding patients recruited from CHECK ($n = 144$) to evaluate whether this dominance influences (improves) our results (see Supplementary Data S3, Tables S2a–f, Figs S2a–d, S3a and b and S6a and b, available at *Rheumatology* online). Unexpectedly, using only patients recruited from CHECK, the portion of actual non-progressors was slightly higher (66.4% for Δ and 61.5% for regression) compared with the full IMI-APPROACH cohort (57.5% for Δ and 53.7% for regression). Excluding patients recruited from CHECK led to 46.5% actual non-progressors for Δ and 44.6% actual non-progressors for regression. For both groups, the ROC curves were comparable to the ROC curves including all IMI-APPROACH participants. AUC for ΔminJSW was 0.612 for all IMI-APPROACH patients, 0.583 when including only patients recruited from CHECK, and 0.640 when excluding patients from CHECK. For regression minJSW AUC was 0.599 for IMI-APPROACH, 0.612 for CHECK, and 0.592 when excluding CHECK patients. AUC for ΔKOOS pain was 0.817 for all IMI-APPROACH patients, 0.818 when including only patients recruited from CHECK, and 0.783 when excluding patients recruited from CHECK, and for regression KOOS pain AUC values were 0.830, 0.828 and 0.801 for IMI-APPROACH, including only patients recruited from CHECK, and excluding CHECK patients, respectively. This indicates that the ML models, built on historical data of CHECK, can be used for other OA cohorts as well.

Although the differences are only marginal, a possible explanation may be the inclusion criteria for CHECK, which resulted in a significant number of slow progressors in this cohort. With stable significant pain as one of the definitions for pain progression, several of these patients ended up being included in APPROACH.

Since the ML models used WOMAC pain scores from the original cohorts and the subsequent clinical prospective study used KOOS pain instead, additionally an alternative analysis was performed using WOMAC pain scores deduced from the corresponding questions of the KOOS pain questionnaire (see Supplementary Data S4, Tables S3a-b and S4, Figs S4 and S5, available at *Rheumatology* online). This analysis revealed that the number of actual pain progressors remained approximately the same (70 *vs* 65 for Δ, and 64 *vs* 67 for regression). However, when using KOOS pain the majority of the actual pain progressors showed stable significant pain. In contrast, when using WOMAC pain the majority of the actual pain progressors showed pain increase. The explanation most likely lies in the number of questions used for the pain score. WOMAC pain is constructed out of five questions [15], while KOOS pain is constructed out of nine questions [14]. As a result, a higher score on one question will have more weight in the WOMAC pain score compared with the KOOS pain score, and with that a patient is more likely to fulfil the pain increase criterion. The AUC for actual pain progressors based on WOMAC pain scores were comparable to AUC for actual pain progressors based on KOOS pain scores (0.821 for ΔWOMAC pain *vs* 0.817 for ΔKOOS pain and 0.817 for regression WOMAC pain *vs* 0.830 for regression KOOS pain). For patients with pain increase AUC values were slightly better for WOMAC pain based progression (0.756 for ΔWOMAC pain and 0.731 for regression WOMAC pain) compared with KOOS pain based progression (AUC 0.664 for ΔKOOS pain and 0.689 for regression KOOS pain). For patients with stable significant pain, AUC values were slightly worse for WOMAC pain based progression compared with KOOS pain based progression (AUC values were 0.790 and 0.823 for ΔWOMAC pain and regression WOMAC pain and 0.843 and 0.841 for ΔKOOS pain and regression KOOS pain, respectively); see Supplementary Data S4, available at *Rheumatology* online.

Because pain and performance are intertwined, also the change in performance-based tests could have been used as measurement for actual pain progression. However, as the progression criteria were predefined, only progression in KOOS pain was evaluated.

The recruitment procedure of IMI-APPROACH used a two-step approach with the aim to limit the number of included patients in the cohort that show neither structural nor pain progression [8]. In an uninformed selection, not using this multi-step selection procedure, 61% was non-progressor. Evaluating the actual progression indicated that in the final selection, the number of non-progressors was 57.5% (for absolute progression) or 53.7% (for regression). From this, in combination with

the fact that the 25% of patients with the lowest ranking were already excluded beforehand, it may be concluded that the IMI-APPROACH selection process, based on ML models, indeed enriched the selection with OA progressors.

In conclusion, the S and P predicted-progression scores as provided by the ML models developed and used for the selection of IMI-APPROACH patients were to some degree able to distinguish between actual progressors and non-progressors over the subsequent 2-year follow-up period. At present, it would be too speculative to suggest to what level the present model would add to patient reduction in future trials. More uniformly acquired data are needed to adjust the models to improve the accuracy of the S and P predicted-progression scores so that, in future trials, the use of ML models might improve patient selection by increasing the number of actual structural and/or pain progressors and with that reduce the trial sample size. Depending on the goal of the trial and the nature of the study intervention, one can use an S or P predicted-progression score (or both) and adjust the cut-off point to select the most appropriate study population.

Acknowledgements

We thank M. Melief for performing the KIDA analysis on all radiographs. E.M.H., M.P.J., A.C.A.M., P.M.J.W. and F.P.J.G.L. contributed to the conception and design of the study, or acquisition of data, or analysis and interpretation of data. E.M.H., M.P.J., A.C.A.M., P.M.J.W., F.P.J.G.L., M.K., F.J.B., I.K.H., F.B., A.C.B.-J., C.L., A.L., J.La., J.Lo., A.M., P.W., J.B., and H.W. contributed to drafting the article or revising it critically for important intellectual content. All authors gave final approval of the version to be submitted.

Funding: This work was supported by the Innovative Medicines Initiative Joint Undertaking under Grant Agreement no. 115770, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in kind contribution. This communication reflects the views of the authors and neither IMI nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. See www.imi.europa.eu and www.approachproject.eu.

Disclosure statement: The IMI-APPROACH project received a grant from Innovative Medicines Institute, Grant Agreement 115770. Outside the submitted work: M.K. reports grants from IMI-APPROACH and Dutch Arthritis Association, during the conduct of the study; other from GlaxoSmithKline, Pfizer, Merck-Serono, Kiniksa, Abbvie, outside the submitted work. F.J.B. reports grants from Gebro Pharma, BIOIBERICA, AB Science, Abbvie, Ablynx N.V., Amgen, Archigen Biotech Limited, Boehringer, Bristol-Myers, Celgene Int., Eli Lilly and Company, F. Hoffmann- La Roche Ltd, Galapagos, Gedeon, Genentech Inc., Gideal Sciences, NC, Glaxosmithkline, Hospira, grants from INC Research UK Ltd, Inventiv Health Clinical, Janssen, Lilly, Nichi-IKO Pharmaceutical, Novartis, ONO Pharma, Pfizer, Pharmaceutical Research, Regeneron, Roche, SA UCB Pharma, Sanofi, TRB Chemedica and UCB Biosciences GMBH, outside the submitted work; in addition, F.J.B. has a patent Molecular block-matching method for gel image analysis issued, a patent Targeting A Specific Receptor On Cells With A Specific Compound For Use In The Treatment And/Or The Prevention Of Osteoarthritis And Rheumatoid Arthritis pending, a patent Genetic markers for osteoarthritis issued, a patent Method for the diagnosis of osteoarthritis issued, a patent Genetic markers for osteoarthritis pending, a patent Method for the diagnosing Arthrosis pending, a patent Method for diagnosing Arthrosis pending, a patent Method for the diagnosis of osteoarthritis pending, and a patent Anti-connexin compounds for use in the prevention and/or treatment of degenerative joint diseases pending. I.K.H. reports personal fees from AbbVie, grants from Pfizer, outside the submitted work; F.B. reports personal fees from Boehringer, Bone Therapeutics, Expanscience, Galapagos, Gilead, GSK, Merck Sereno, MSD, Nordic, Novartis, Pfizer, Regulaxis, Roche, Sandoz, Sanofi, Servier, UCB, Peptinov, TRB Chemedica, 4P Pharma. A.C.B.-J. reports non-financial support from Nordic Bioscience, personal fees from Nordic Bioscience, during the conduct of the study. C.H.L. reports other from Merck KGaA, during the conduct of the study. A.L. is an employee of Institut de Recherches Internationales Servier. J.La. reports personal fees and other from GlaxoSmithKline, outside the submitted work. E.M.H., M.P.J., A.C.A.M., J.Lo., A.M., H.W., P.W., J.B., P.M.J.W. and F.P.J.G.L. have nothing to disclose.

Ethical approval: The study is being conducted in compliance with the protocol, Good Clinical Practice (GCP), the Declaration of Helsinki, and applicable ethical and legal regulatory requirements (for all countries involved), and is registered under clinicaltrials.gov nr: NCT03883568. All participants have received oral and written information and provided written informed consent.

Data availability statement

In order to gain and govern access to the central APPROACH databases, tranSMART and XNAT, access has to be approved by the APPROACH Steering Committee.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Felson D, Niu J, Sack B *et al.* Progression of osteoarthritis as a state of inertia. Ann Rheum Dis 2013; 72:924–9.

2 Taylor J, Dekker S, Jurg D *et al.*; APPROACH research consortium and APPROACH Principal Investigators. Making the patient voice heard in a research consortium: experiences from an EU project (IMI-APPROACH). Res Involv Engagem 2021;7:24.

3 Wesseling J, Boers M, Viergever MA *et al.* Cohort profile: Cohort Hip and Cohort Knee (CHECK) study. Int J Epidemiol 2016;45:36–44.

4 Damman W, Liu R, Kroon FPB *et al.* Do comorbidities play a role in hand osteoarthritis disease burden? Data from the hand osteoarthritis in secondary care cohort. J Rheumatol 2017;44:1659–66.

5 Magnusson K, Hagen KB, Osteras N *et al.* Diabetes is associated with increased hand pain in erosive hand osteoarthritis: data from a population-based study. Arthritis Care Res 2015;67:187–95.

6 Oreiro-Villar N, Raga AC, Rego-Pérez I *et al.* PROCOAC (PROspective COhort of A Coruña) description: Spanish prospective cohort to study osteoarthritis. Reumatol Clin (Engl Ed) 2020, doi: 10.1016/j.reuma.2020.08.010.

7 Sellam J, Maheu E, Crema MD *et al.* The DIGICOD cohort: a hospital-based observational prospective cohort of patients with hand osteoarthritis – methodology and baseline characteristics of the population. Joint Bone Spine 2021;88:105171.

8 Widera P, Welsing PMJ, Ladel C *et al.* Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data. Sci Rep 2020; 10:8427.

9 van Helvoort EM, van Spil WE, Jansen MP *et al.* Cohort profile: the Applied Public-Private Research enabling OsteoArthritis Clinical Headway (IMI-APPROACH) study: a 2-year, European, cohort study to describe, validate and predict phenotypes of osteoarthritis using clinical, imaging and biochemical markers. BMJ Open 2020;10: e035101.

10 van Helvoort EM, Ladel C, Mastbergen S *et al.* Baseline clinical characteristics of predicted structural and pain progressors in the IMI-APPROACH knee OA cohort. RMD Open 2021;7:e001759.

11 Altman R, Asch E, Bloch D *et al.* Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.

12 Buckland-Wright JC, Ward RJ, Peterfy C, Mojcik CF, Leff RL. Reproducibility of the semiflexed (metatarsophalangeal) radiographic knee position and automated measurements of medial tibiofemoral joint space width in a multicenter clinical trial of knee osteoarthritis. J Rheumatol 2004;31:1588–97.

13 Marijnissen AC, Vincken KL, Vos PA *et al.* Knee Images Digital Analysis (KIDA): a novel method to quantify individual radiographic features of knee osteoarthritis in detail. Osteoarthritis Cartilage 2008;16:234–43.

14 Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. J Orthopaedic Sports Phys Ther 1998;28: 88–96.

15 Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.

16 Kinds MB, Vincken KL, Hoppinga TN *et al.* Influence of variation in semiflexed knee positioning during image acquisition on separate quantitative radiographic parameters of osteoarthritis, measured by Knee Images Digital Analysis. Osteoarthritis Cartilage 2012;20: 997–1003.

17 Ornetti P, Brandt K, Hellio-Le Graverand MP *et al.* OARSI-OMERACT definition of relevant radiological progression in hip/knee osteoarthritis. Osteoarthritis Cartilage 2009;17:856–63.

18 Angst F, Benz T, Lehmann S, Aeschlimann A, Angst J. Multidimensional minimal clinically important differences in knee osteoarthritis after comprehensive rehabilitation: a prospective evaluation from the Bad Zurzach Osteoarthritis Study. RMD Open 2018;4:e000685.