Towards developing diagnostic criteria for early knee osteoarthritis: data from the CHECK study

Runhaar, J.; Kloppenburg, M.; Boers, M.; Bijlsma, J.W.J.; Bierma-Zeinstra, S.M.A.; CREDO Expert Grp

RHEUMATOLOGY

Rheumatology 2021;60:2448–2455
doi:10.1093/rheumatology/keaa643
Advance Access publication 27 November 2020

Original article

Towards developing diagnostic criteria for early knee osteoarthritis: data from the CHECK study

J. Runhaar[1], M. Kloppenburg[2], M. Boers[3], J. W. J. Bijlsma[4] and S. M. A. Bierma-Zeinstra[1,5]; and the CREDO expert group

Abstract

Objectives. There is a general consensus that a shift in focus towards early diagnosis and treatment of knee OA is warranted. However, there are no validated and widely accepted diagnostic criteria for early knee OA available. The current study aimed to take the first steps towards developing diagnostic criteria for early knee OA.

Methods. Data of 761 individuals with 1185 symptomatic knees at baseline were selected from the CHECK study. For CHECK, individuals with pain/stiffness of the knee, aged 45–65 years, who had no prior consultation or a first consultation with the general practitioner for these symptoms in the past 6 months were recruited and followed for 10 years. A group of 36 experts (17 general practitioners and 19 secondary care physicians) evaluated the medical records in pairs to diagnose the presence of clinically relevant knee OA 5–10 years after enrolment. A backward selection methods was used to create predictive models based on pre-defined baseline factors from history taking, physical examination, radiography and blood testing, using the experts' diagnoses as gold standard outcome.

Results. Prevalence of clinically relevant knee OA during follow-up was 37%. Created models contained 7–11 baseline factors and obtained an area under the curve between 0.746 (0.002) and 0.764 (0.002).

Conclusion. The obtained diagnostic models for early knee OA had 'fair' predictive ability in individuals presenting with knee pain in primary care. Further modelling and validation of the identified predictive factors is required to obtain clinically feasible and relevant diagnostic criteria for early knee OA.

Key words: early diagnosis, knee osteoarthritis, diagnostic criteria, expert diagnosis

CLINICAL SCIENCE

Rheumatology key messages

- A shift towards early diagnosis and treatment of knee osteoarthritis is warranted.
- Herewith, first steps towards the development of diagnostic criteria for early knee osteoarthritis are presented.
- Clinical assessment and radiography had acceptable diagnostic abilities for the development of 'clinically relevant osteoarthritis'.

Introduction

Despite the high prevalence of knee OA and its burden to patients and healthcare, the diagnosis of knee OA is not straightforward. In 1986, Altman and colleagues developed the so-called ACR criteria for knee OA [1]. Although the ACR criteria were originally presented as classification criteria, they were soon also used as diagnostic outcome criteria in OA research. The ACR criteria include case criterion sets based on clinical features ('clinical ACR criteria') only and based on clinical features combined with radiography ('clinical + radiographic ACR criteria') or lab testing ('clinical + laboratory ACR criteria'). These ACR criteria were validated among knee pain patients in secondary care and modelled to distinguish OA from inflammatory arthritis (mainly RA) [2]. Accepted definitions for knee OA based on structural features, like Kellgren and Lawrence (KL) criteria [3] or OARSI scores [4], represent late-stage disease as structural features develop slowly over time [5, 6].

[1]Department of General Practice, Erasmus MC University Medical Center Rotterdam, Rotterdam, [2]Department of Rheumatology, Leiden University Medical Center, Leiden, [3]Department of Epidemiology & Biostatistics, Amsterdam UMC, Amsterdam, [4]Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht and [5]Department of Orthopaedics, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands

Submitted 9 April 2020; accepted 23 August 2020

Correspondence to: Jos Runhaar, Department of General Practice, Erasmus MC University Medical Center Rotterdam, Room NA1911, 3000 CA Rotterdam, PO-box 2040, the Netherlands.
E-mail: j.runhaar@erasmsumc.nl

Patients with symptoms of knee OA will mainly be treated in primary care. However, for the clinical diagnosis of established knee OA little guidance is available. It was shown that the clinical ACR criteria, although assessable in primary care, had no prognostic value for predicting persisting knee complaints or an increase of disability at 1 year of follow-up in patients with non-traumatic knee complaints in primary care [7], did not predict unfavourable outcome after 6 years of follow-up [8] and seemed to be indicative for late stage disease [9]. EULAR released evidence-based recommendations for the diagnosis of knee OA in 2009 that included three recommended symptoms (persistent knee pain, limited morning stiffness and reduced function) and three signs (crepitus, restricted movement and bony enlargement) that 'appeared to be most useful' for the clinical diagnosis of knee OA [2]. Also the National Institute for Health and Care Excellence (NICE) published diagnostic criteria for the clinical diagnosis of knee OA (age $\geq$45, activity-related joint pain and morning joint-related stiffness $\leq$30 min) [10].

With a lack of treatment options that can cure OA, there is general consensus that a shift in focus towards early diagnosis and treatment is warranted [11–13]. By identifying OA patients early in the disease, treatment can potentially start prior to the occurrence of irreversible joint damage, before pain becomes chronic and triggers sensitization of the central nervous system, and before severe decline in physical functioning has occurred [11–13]. In theory, in the early phase of OA, treatment options are applied when the disease is more amenable to modification [13]. As of today, there are no validated and widely accepted criteria for early knee OA available. Like for established OA (in primary care), validation of diagnostic criteria for early OA is challenging with the lack of a gold standard [2].

One way to overcome the lack of a gold standard in a research setting is to obtain a clinical expert based diagnosis to identify clinically relevant cases of knee OA. The first aim of the current study was to obtain a clinical expert based diagnosis of clinically relevant knee OA development 5–10 years after first presentation, in patients aged 45–65 years who consulted their general practitioner (GP) for knee complaints. Next, using the clinical expert based diagnosis as gold standard, we aimed to develop diagnostic criteria for early knee OA based on a set of pre-defined clinical and radiographic factors obtained at first consultation. As the set of predicting factors have been collected at first consultation, these should be seen as early diagnostic criteria.

Methods

Cohort

Individuals were eligible for the Cohort Hip and Cohort Knee (CHECK) cohort if they had pain or stiffness of the knee, were aged 45–65 years, and had no prior consultation (recruited via media campaign) or a first consultation with the GP for these symptoms no longer than 6 months before recruitment. Exclusion criteria were: presence of a clear pathological condition (assessed through history taking and/or physical examination) other than OA that could explain the existing complaints (e.g. other rheumatic disease, previous hip or knee joint replacement, congenital dysplasia, osteochondritis dissecans, intra-articular fractures, septic arthritis, Perthes' disease, ligament or meniscus damage, plica syndrome, Baker's cyst), comorbidity that did not allow physical evaluation and/or follow-up of at least 10 years, malignancy in the past 5 years, and inability to understand the Dutch language. Patients were followed for 10 years at regular intervals (see Wesseling *et al.* for details [14]). The CHECK study has been approved by the Medical Ethics Committee of the University Medical Center Utrecht (02/017-E) and all patients provided informed consent prior to data collection. For the present study, data of the 761 individuals with 1185 symptomatic knees at baseline who had any follow-up data available were selected from the CHECK cohort. At baseline there were 832 individuals with knee symptoms included in the CHECK cohort.

Baseline measures

At baseline, all knees were physically examined to evaluate range of motion and the presence of pain on flexion and on extension, warmth, crepitus (during squatting), bony swelling at the joint margins, joint line tenderness, patellofemoral (PF) grinding (presence of pain upon isometric contractions of the m. quadriceps while pressure is put onto the knee cap), effusion (refill test), Heberden and/or Bouchard nodes, and BMI were determined. All subjects completed one Western Ontario and McMaster Universities OA Index (WOMAC) questionnaire (not limb or joint specific) [15] and additional demographical questions to determine symptomatic joints (left/right knee or both), age, sex, ethnicity, menopausal status (women only), marital status, education level, chronic conditions, employment status, (former) occupation, number of days with $\geq$30 min of physical activity, smoking status and alcohol consumption. Standardized posterior–anterior (PA) and lateral radiographs were centrally read for the presence of medial and lateral osteophytes (0–3 scales) on the femur, tibia and patella, for joint space narrowing (JSN) in the medial, lateral and patellofemoral compartments (0–3 scales), and for KL grading (0–4 scale) [3]. With the Knee Images Digital Analysis (KIDA) [16], medial knee alignment angles were determined on the PA radiographs. Blood testing included high-sensitive CRP (hsCRP).

Follow-up measures

At 5, 8 and 10 years after baseline, the above procedure was repeated, and patients were also questioned for the occurrence of osteochondritis dissecans, intra-articular

fractures, bacterial arthritis, ligament or meniscal trauma, plica syndrome, and Baker's cyst.

Expert diagnosis

A group of 36 experts was recruited: 17 GPs and 19 secondary care physicians (nine rheumatologists, eight orthopaedic surgeons and two sport physicians). For all experts, number of years treating OA patients, number of OA patients treated per week and subjective importance of radiography for diagnosing OA ('not important', 'little importance', 'some importance', 'very important') were obtained prior to medical record evaluation.

All experts independently evaluated the medical records for 40–50 CHECK participants; of these, seven were evaluated by all experts. Pairs of experts read the same records: 17 pairs comprised a GP and a secondary care physician, and one pair comprised two secondary care physicians. Medical records presented patients' demographics and all follow-up measures, with in-house developed software to obtain optimal representation of the data. First, each expert was presented with all clinical data from the questionnaires and physical examination for one individual. The expert was asked to answer the question 'Is there clinically relevant knee OA present in this knee?' for each joint (Y/N) and to provide a certainty of the diagnosis, ranging from 1 ('definitely no clinically relevant OA is present') to 100 ('definitely clinically relevant OA is present'), into the software system. No formal definition of clinically relevant knee OA was provided to the experts; they were instructed to use their own clinical expertise to judge this.

As WOMAC scores are lower limb specific and not knee specific, clinical data from physical examination of the hips for the individual (e.g. pain upon flexion, ab-/adduction, internal/external rotation, presence of morning stiffness, and ROM in all directions) were available for the expert. The software recorded access to these data.

After the experts had provided their diagnoses and certainty scores, the radiographic data for the individual were made available. Experts were provided with the KL grading and the scores for JSN and osteophytes for each joint at each follow-up time point. The actual radiographs were also available for inspection. Then, the experts were again asked 'Is there clinically relevant OA present in this knee?' for each joint, and they had to provide a new certainty score. In this phase, the clinical data were still available to the expert on a read-only basis.

After completion, agreement on diagnosis and its certainty was assessed within each expert pair. All cases where experts disagreed on the diagnosis were re-evaluated, except those labelled 'uncertain': predefined, experts disagreeing on the presence of clinically relevant OA, but both scoring the certainty >30 and <70. In an online consensus meeting each expert pair re-evaluated the disagreed cases. They followed the same procedure as before, but now diagnoses and certainty scores from both experts were presented. Cases where still no consensus could be reached were also labelled as 'uncertain'.

Statistics

Baseline factors were limited to predefined factors described in literature as diagnostic or prognostic for knee OA. Factors were checked for completeness and multiple imputation replaced missing values (creating 50 data sets, as 42% of cases had incomplete data, but only two variables had $>10\%$ of missing values). Next, categorical factors were dichotomized based on literature and authors' expertise. For the selected separate WOMAC pain, function and morning stiffness questions, absence of pain/functional limitations/stiffness was defined by merging the 'none' and 'slight' categories. Presence of 'restricted or painful extension' was defined as an extension deficit ≥ 1 degree or pain at knee extension. Presence of 'restricted or painful flexion' was defined as maximal knee flexion ≤ 115 degrees or pain at knee flexion. Osteophytes and JSN were defined as a grade ≥ 2 (equals $\geq$'minimal'). Varus malalignment was defined as medial knee angle <2 degrees and valgus as >0 degrees [17].

To identify early diagnostic factors for the presence of clinically relevant knee OA 5–10 years later, predictive models (using the expert diagnoses as outcome) were created with a stepped approach; first all factors obtained from questionnaires and physical examination were used (model 1). Next, all radiographic factors were added (model 2) and finally hsCRP (model 3). For all models a backward selection method ($P > 0.1$ for removal) was used. To correct for correlated measures within subjects due to possible bilateral complaints, generalized estimating equations were used. For each model, area under the receiver operator curve (AUC) was calculated and odds ratios plus 95% CIs for each factor within the models were presented.

In sensitivity analyses, first continuous measures for age, BMI, duration of complaints and hsCRP were dichotomized. For age and duration of complaints, the upper tertile was compared with the lower two tertiles. BMI was dichotomized at two different cut-offs; <25 *vs* ≥ 25 kg/m^2 and <30 *vs* ≥ 30 kg/m^2. hsCRP was dichotomized at ≤ 3 mg/l *vs* >3 mg/l [18]. Second, to evaluate the potential of overestimation of the obtained models, AUC values for the final models were calculated after excluding all knees with mild to moderate radiographic knee OA (KL grade ≥ 2) at baseline, as obtained by single reading of baseline radiographs, blinded for clinical data [14].

Results

Seventy-nine per cent of the 761 selected individuals were female, baseline age was mean (S.D.) 56 (5) years, BMI was 26.3 (4.2) kg/m^2, median duration of complaints was 11.8 months (interquartile range 25.5 months) and average WOMAC pain and function (0–100) scores

Table 1 Baseline pooled prevalence, presented as percentage of knees with the factor out of 1185 knees

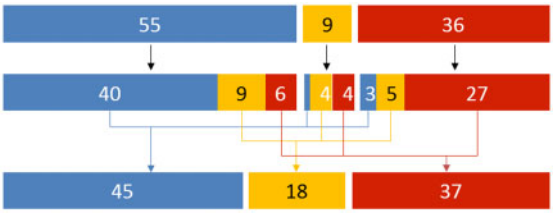
	Pooled prevalence (%)
Questionnaire and physical examination items	
Sex (female)	80
Bilateral pain	72
Heberden/Bouchard nodes	52
Joint line tenderness	45
Bony swelling	4
Warmth	4
Effusion	7
Crepitus	45
PF grinding	31
Painful/restricted extension[a]	51
Painful/restricted flexion[b]	30
WOMAC function[c]	
Descending	30
Ascending	41
Rising	37
Standing	19
Walking	14
Sitting	15
WOMAC pain[d]	
Walking	17
Standing	21
Stairs	48
Night	32
Rest	26
WOMAC stiffness[e]	
Morning stiffness	50
Radiography items	
Medial TF osteophytes[f]	11
Lateral TF osteophytes[f]	16
Medial JSN[f]	6
Lateral JSN[f]	1
Varus malalignment[g]	41
Valgus malalignment[h]	12
PF osteophytes[f]	11
PF JSN[f]	3

[a]Presence defined as extension deficit ≥1 degree or pain at knee extension. [b]Presence defined as flexion deficit ≥1 degree or pain at knee flexion. [c]Presence defined as ≥moderate difficulties. [d]Presence defined as ≥moderate pain. [e]Presence defined as ≥moderate stiffness. [f]Presence defined as ≥'minimal'. [g]Presence defined as medial knee angle <2 degrees. [h]Presence defined as medial knee angle >0 degrees. JSN: joint space narrowing; PF: patellofemoral; TF: tibiofemoral; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index.

were 25.7 (17.3) and 24.4 (17.1), respectively. Fifty-six per cent of individuals had bilateral pain. Baseline prevalence of the other selected factors, pooled after multiple imputation, are presented in Table 1. Mean baseline concentration for hsCRP was 3.2 (7.7) mg/l.

The experts had 14 (8) years of experience treating OA patients and treated five OA patients (median, interquartile range 2–20) per week. Prior to evaluating the medical records, 11% of the experts deemed

Fig. 1 Percentages for expert diagnoses



Expert diagnoses are based on clinical data (upper row), on clinical and radiographic data (middle row) and summed (bottom row). Blue bars represent knees without OA, yellow bars represent knees with an uncertain diagnosis and red bars represent knees with OA.

radiographs 'not important' for OA diagnoses, 22% of 'little importance', 36% 'some importance' and 31% 'very important'. Both knees of seven individuals were evaluated by all 18 expert pairs (252 diagnoses). There was agreement between expert pairs in 74% of these diagnoses (intraclass correlation coefficient 0.908; 95% CI: 0.821, 0.965).

Figure 1 presents the consensus-based diagnoses for all 1185 selected knee joints. Finally, in 37% of all knees, clinically relevant OA had developed, based on the expert diagnoses. In 18% of the knees, the final diagnosis was uncertain. For optimal contrast, first the cases diagnosed without clinically relevant knee OA were compared with the cases diagnosed with clinically relevant knee OA, ignoring all uncertain cases. First, the diagnoses obtained after experts' evaluations of the clinical data only were used as outcome. Next, the diagnoses obtained after experts' evaluations of the clinical + radiographic data were used as outcome. Baseline factors ending up in the final models are presented in Tables 2 and 3. As hsCRP did not end up in any model when the diagnoses obtained after experts' evaluations of the clinical + radiographic data were used as outcome, only models based on clinical factors (model 1) and on clinical + radiographic factors (model 2) are presented in Table 3. Categorizing the continuous variables in the sensitivity analyses resulted in minor changes in the predictive abilities, with AUCs ranging from 0.735 to 0.743 for model 1 and 0.744 to 0.752 for models 2 and 3. Excluding knees with radiographic knee OA at baseline did not affect the AUCs of the obtained models (see Supplementary Table S1, available at *Rheumatology* online).

To evaluate the contribution of individual factors, the backward selection method for the obtained models (as presented in Tables 2 and 3) was continued by removing the least significant factor step by step. After each removal, AUC was calculated and presented in Supplementary Figs S1–S4, available at *Rheumatology* online.

When incorporating the uncertain cases into the analyses, once as incident OA cases and once as cases of no OA, predicting ability of the obtained models

TABLE 2 Final models for developing clinically relevant knee OA based on the evaluation of clinical data only

Item	Value
Questionnaire and physical examination items at baseline	
Odds ratio (95% CI)	
WOMAC pain—stairs	1.99 (1.36, 2.92)
WOMAC pain—night	1.52 (1.06, 2.20)
WOMAC function—rising	1.61 (1.08, 2.39)
Sex (female)	1.87 (1.20, 2.92)
Joint line tenderness	2.36 (1.73, 3.22)
Effusion	1.86 (1.09, 3.16)
BMI	1.07 (1.03, 1.12)
Pooled AUC (pooled S.D.)	0.746 (0.002)
Questionnaire, physical examination and radiographic items at baseline	
Odds ratio (95% CI)	
WOMAC pain—stairs	1.98 (1.34, 2.90)
WOMAC pain—night	1.53 (1.06, 2.20)
WOMAC function—rising	1.58 (1.06, 2.35)
Sex (female)	1.81 (1.16, 2.83)
Joint line tenderness	2.29 (1.68, 3.13)
Effusion	1.85 (1.09, 3.15)
BMI	1.07 (1.03, 1.12)
Crepitus	1.32 (0.96, 1.81)
Lateral JSN	5.32 (1.14, 24.88)
Pooled AUC (pooled S.D.)	0.749 (0.002)
Questionnaire, physical examination and radiographic items and hsCRP at baseline	
Odds ratio (95% CI)	
WOMAC pain—stairs	2.05 (1.39, 3.03)
WOMAC pain—night	1.55 (1.07, 2.24)
WOMAC function—rising	1.67 (1.11, 2.49)
Sex (female)	1.90 (1.22, 2.97)
Joint line tenderness	2.43 (1.78, 3.32)
Effusion	1.84 (1.08, 3.13)
BMI	1.08 (1.03, 1.13)
Lateral JSN	4.66 (1.01, 21.65)
hsCRP	0.95 (0.90, 0.99)
Pooled AUC (pooled S.D.)	0.756 (0.002)

AUC: Area under the curve; hsCRP: high-sensitive CRP; JSN: joint space narrowing; PF: patellofemoral; WOMAC: Western Ontario and McMaster Universities OA Index.

presented in Tables 2 and 3 was only slightly reduced with AUCs ranging from 0.725 to 0.764, corresponding to 1.6–6.0% lower AUCs (Supplementary Table S2, available at *Rheumatology* online).

Discussion

Based on an extensive set of baseline factors that are thought to be related to knee OA development, the obtained models had a 'fair' distinction between cases with and without clinically relevant knee OA. As the follow-up time between baseline and the evaluation of incident knee OA was 5–10 years, the models should be interpreted as diagnostic for early stage knee OA, among middle-aged subjects who first present with knee symptoms suggestive for early stage knee OA in primary care. Our purpose was to develop diagnostic criteria, not to improve on classification criteria. ACR (and other) classification criteria are intended for use in research, where, after initial selection of patients based on clinical diagnosis, a homogeneous subset can be selected through application of the criteria. Classification criteria are not the gold standard for clinical diagnosis and should not be used for diagnostic purposes (although this is frequently done).

The obtained models had AUC ranging from 0.75 to 0.76, depending on the set of baseline factors and the data available to the experts when obtaining their diagnoses. Predictive ability of the obtained models mainly relied on factors from history taking and physical examination (AUC 0.75), with factors from radiography and hsCRP not adding much more (AUC ranging from 0.75 to 0.76). This is confirmed in the presented Supplementary Figures, available at *Rheumatology* online, where radiographic factors and hsCRP were among the first factors to be removed (weakest association with the outcome), without large reductions in AUC. This highlights the importance of history taking and physical examination in early diagnosis of knee

TABLE 3 Final models for developing clinically relevant knee OA based on the evaluation of clinical and radiographic data

Item	Value
Questionnaire and physical examination items at baseline	
Odds ratio (95% CI)	
WOMAC function—descending	1.69 (1.09, 2.61)
WOMAC function—rising	1.51 (0.98, 2.32)
WOMAC morning stiffness	1.59 (1.06, 2.37)
Sex (female)	1.79 (1.14, 2.82)
Joint line tenderness	1.68 (1.21, 2.33)
Effusion	4.17 (2.29, 7.59)
Crepitus	1.55 (1.11, 2.11)
Age	1.03 (1.00, 1.07)
BMI	1.12 (1.07, 1.17)
Pooled AUC (pooled S.D.)	0.752 (0.002)
Questionnaire, physical examination and radiographic items at baseline	
Odds ratio (95% CI)	
WOMAC function—descending	1.52 (0.96, 2.41)
WOMAC function—rising	1.55 (1.00, 2.40)
WOMAC morning stiffness	1.62 (1.08, 2.41)
Sex (female)	1.77 (1.14, 2.76)
Joint line tenderness	1.72 (1.23, 2.41)
Effusion	3.51 (1.93, 6.40)
Crepitus	1.43 (1.02, 2.02)
BMI	1.12 (1.06, 1.17)
Bony swelling	2.06 (0.88, 4.82)
Medial JSN	3.56 (1.56, 8.11)
PF JSN	3.37 (0.94, 12.10)
Pooled AUC (pooled S.D.)	0.764 (0.002)

AUC: area under the curve; JSN: joint space narrowing; PF: patellofemoral; WOMAC: Western Ontario and McMaster Universities OA Index.

OA. Unfortunately, physical examinations can be observer-dependent, so detailed descriptions of the examinations will be required when implementing current diagnostic models.

In many studies, incident knee OA is defined using either radiographic features (e.g. incident KL grade $\geq$2) or criteria based on symptoms (e.g. clinical/combined ACR criteria). Incidence of radiographic knee OA in the selected CHECK knees was 53%. This is somewhat higher than the 37% of clinically relevant cases diagnosed by the experts. However, as the results in Supplementary Table S2 (available at *Rheumatology* online) suggest, the 'uncertain cases' (18%) were more likely to be OA cases than non-OA cases, and the incidence rates for radiographic knee OA and the expert based diagnoses were both around 50% after 10 years. Nevertheless, overlap between incident radiographic knee OA and the expert based diagnoses (when considering 'uncertain cases' as OA cases) was only 59%, with 18% already showing KL grade $\geq$2 at baseline. Also at baseline, already 43% of knees fulfilled the clinical ACR criteria in CHECK and incidence of the clinical ACR criteria in those not fulfilling the criteria at baseline was 21% after 10 years [19]. The number of patients fulfilling the clinical ACR criteria strongly fluctuated over time, with few patients fulfilling the criteria at multiple time points [19], despite stable or progressing KL scores. These numbers confirm the mismatch between radiographic features of knee OA, available clinical criteria and clinically relevant knee OA.

The current initiative resulted in different sets of factors predictive for clinically relevant knee OA after 5–10 years, in individuals presenting with knee pain suggestive for knee OA. However, essential next steps include truncating models to fewer items (now containing 7–11 factors) or transforming the models into a scoring scheme, to become feasible in clinical practice, after which sensitivity, specificity, and positive and negative predictive values need to be reassessed. For the reduction of items and to increase the feasibility of the diagnostic models, one could consider, in analogy to the ACR criteria, evaluating the predictive abilities of models containing '*x* out of *y* factors'. Of course, all created models need to be externally validated before a final conclusion regarding the clinical relevance of the models can be made. Also, evaluating the additive value of other known risk factors for knee OA not assessed within the CHECK study, such as occupational overload, could be considered and the performance of the final models needs to be compared with existing criteria to assess their superiority.

One important factor not available within the CHECK study was a history of knee joint injury. As knee trauma has a strong link to knee OA development [6], this is a

limitation to the present study. The use of a clinical expert based diagnosis as gold standard could be seen as a strength and as a limitation to the current study. The expert based diagnosis should be seen as a strength as it is deemed a highly clinically relevant outcome measure; it does not rely on structural features only (like KL grading) and fluctuation in symptoms was also covered by presenting the experts' data over a 5-year interval. On the other hand, this outcome measure might be country specific, as only Dutch experts were involved, and as it will not be available in other cohort studies, external validation of the current diagnostic criteria will be challenging. Since the experts did not evaluate the medical files of the participants for the first 5 years of follow-up, we cannot rule out that some would already have been diagnosed with clinically relevant knee OA at enrolment. However, as population screening for early stage knee OA is not a realistic option, the first moment to diagnose early stage knee OA or clinically relevant knee OA is at the first consultation in primary care for knee complaints, which is exactly what was done in our cohort. Potential treatment options for those fulfilling the presented set of criteria at their first consultation (education, simple analgesics, and when applicable weight loss and lifestyle changes) will be identical for individuals with early stage knee OA and those with clinically relevant knee OA. Finally, the CHECK cohort included patients presenting not only with knee symptoms, but also with concurrent hip symptoms, limiting the generalizability. However, within the current study prediction models were only run among symptomatic knee joints and the experts also had access to the clinical hip data to confirm the presence of hip symptoms within a person. Noteworthy, also patients with clear pathologies that could explain their symptoms at study enrolment were excluded. Therefore, the obtained models will only be applicable to middle-aged patients suspected of (early stage) knee OA.

In conclusion, the current study created several diagnostic models for early knee OA with 'fair' predictive ability in individuals presenting with knee pain, suggestive for (early stage) knee OA, in primary care. As the included population is the ideal population to undergo early treatment, these criteria should be seen as a first step in shifting the focus of the clinical treatment of knee OA towards the early stage of the disease. Further modelling and validation of the identified predictive factors is required to obtain clinically feasible and relevant diagnostic criteria.

Acknowledgements

We would like to acknowledge the CREDO experts group (N.E. Aerts-Lankhorst, R. Agricola, A.N. Bastick, R.D.W. van Bentveld, P.J. van den Berg, J. Bijsterbosch, A. de Boer, M. Boers, A.M. Bohnen, A.E.R.C.H. Boonen, P.K. Bos, T.A.E.J. Boymans, H.P. Breedveldt-Boer, R.W. Brouwer, J.W. Colaris, J. Damen, G. Elshout, P.J. Emans, W.T.M. Enthoven, E.J.M. Frölke, R. Glijsteen, H.J.L. van der Heide, A.M. Huisman, R.D. van Ingen, M.L. Jacobs, R.P.A. Janssen, P.M. Kevenaar, M. Kloppenburg, M.A. van Koningsbrugge, P. Krastman, N.O. Kuchuk, M.L.A. Landsmeer, W.F. Lems, H.M.J. van der Linden, R. van Linschoten, E.A.M. Mahler, B.L. van Meer, D.E. Meuffels, W.H. Noort-van der Laan, J.M. van Ochten, J. van Oldenrijk, D.H.J. Pols, T.M. Piscaer, J.B.M. Rijkels-Otters, N. Riyazi, J.M. Schellingerhout, H.J. Schers, B.W.V. Schouten, G.F. Snijders, W.E. van Spil, S.A.G. Stitzinger, J.J. Tolk, Y.D.M. van Trier, M. Vis, V.M.I. Voorbrood, B.C. de Vos, and A. de Vries) for evaluating the medical files and their feedback on the manuscript.

Funding: This work was supported by the Dutch Arthritis Society (Project ID 15-1-301).

Disclosure statement: The authors have declared no conflicts of interest.

Data availability statement

The data underlying this article will be shared on reasonable request to the corresponding author.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 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.

2 Zhang W, Doherty M, Peat G *et al.* EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. Ann Rheum Dis 2010;69:483–9.

3 Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.

4 Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15:A1–56.

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

6 Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019;393:1745–59.

7 Belo JN, Berger MY, Koes BW, Bierma-Zeinstra SM. The prognostic value of the clinical ACR classification criteria of knee osteoarthritis for persisting knee complaints and increase of disability in general practice. Osteoarthritis Cartilage 2009;17:1288–92.

8 Kastelein M, Luijsterburg PA, Belo JN *et al.* Six-year course and prognosis of nontraumatic knee symptoms in adults in general practice: a prospective cohort study. Arthritis Care Res (Hoboken) 2011;63:1287–94.

9 Peat G, Thomas E, Duncan R *et al.* Clinical classification criteria for knee osteoarthritis: performance in the

general population and primary care. Ann Rheum Dis 2006;65:1363–7.

10 National Institute for Health and Clinical Excellence. NICE guideline on osteoarthritis: The care and management of osteoarthritis in adults, NICE clinical guideline 177 2014. Online document at: http://guidance.nice.org.uk/CG177.

11 Luyten FP, Bierma-Zeinstra S, Dell'Accio F *et al.* Toward classification criteria for early osteoarthritis of the knee. Semin Arthritis Rheum 2018;47:457–63.

12 Roos EM, Arden NK. Strategies for the prevention of knee osteoarthritis. Nat Rev Rheumatol 2016;12:92–101.

13 Emery CA, Whittaker JL, Mahmoudian A *et al.* Establishing outcome measures in early knee osteoarthritis. Nat Rev Rheumatol 2019;15:438–48.

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

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 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.

17 Runhaar J, van Middelkoop M, Reijman M *et al.* Malalignment: a possible target for prevention of incident knee osteoarthritis in overweight and obese women. Rheumatology (Oxford) 2014;53:1618–24.

18 Sanchez-Ramirez DC, van der Leeden M, van der Esch M *et al.* Elevated C-reactive protein is associated with lower increase in knee muscle strength in patients with knee osteoarthritis: a 2-year follow-up study in the Amsterdam Osteoarthritis (AMS-OA) cohort. Arthritis Res Ther 2014;16:R123.

19 Schiphof D, Runhaar J, Waarsing JH, van Spil WE van Middelkoop M, Bierma-Zeinstra SMA. The clinical and radiographic course of early knee and hip osteoarthritis over 10 years in CHECK (Cohort Hip and Cohort Knee). Osteoarthritis Cartilage 2019;27:1491–500.