Prediction models for the risk of total knee replacement: development and validation using data from multicentre cohort studies

Qiang Liu*, Hongling Chu*, Michael P LaValley, David J Hunter, Hua Zhang, Liyuan Tao, Siyan Zhan, Jianhao Lin, Yuqing Zhang

Summary

Background Few prognostic prediction models for total knee replacement are available, and the role of radiographic findings in predicting its use remains unclear. We aimed to develop and validate predictive models for total knee replacement and to assess whether adding radiographic findings improves predictive performance.

Methods We identified participants with recent knee pain (in the past 3 months) in the Multicenter Osteoarthritis Study (MOST) and the Osteoarthritis Initiative (OAI). The baseline visits of MOST were initiated in 2003 and of OAI were initiated in 2004. We developed two predictive models for the risk of total knee replacement within 60 months of follow-up by fitting Cox proportional hazard models among participants in MOST. The first model included sociodemographic and anthropometric factors, medical history, and clinical measures (referred to as the clinical model). The second model added radiographic findings into the predictive model (the radiographic model). We evaluated each model's discrimination and calibration performance and assessed the incremental value of radiographic findings using both category-free net reclassification improvement (NRI) and integrated discrimination improvement (IDI). We tuned the models and externally validated them among participants in OAI.

Findings We included 2658 participants from MOST (mean age 62·4 years [SD 8·1], 1646 [61·9%] women) in the training dataset and 4060 participants from OAI (mean age 60·9 years [9·1], 2379 [58·6%] women) in the validation dataset. 290 (10·9%) participants in the training dataset and 174 (4·3%) in the validation dataset had total knee replacement. The retained predictive variables included in the clinical model were age, sex, race, history of knee arthroscopy, frequent knee pain, current use of analgesics, current use of glucosamine, body-mass index, and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain score, and the most predictive factors were age, race, and WOMAC pain score. The retained predictive variables in the radiographic model were age, sex, race, frequent knee pain, current use of analgesics, WOMAC pain score, and Kellgren–Lawrence grade, and the most predictive factors were Kellgren–Lawrence grade, race, and age. The C-statistic was 0·79 (95% CI 0·76–0·81) for the clinical model and 0·87 (0·85–0·99) for the radiographic model in the training dataset. The calibration slope was 0·95 (95% CI 0·86–1·05) and 0·96 (0·87–1·04), respectively. Adding radiograph findings significantly improved predictive performance with an NRI of 0·43 (95% CI 0·38–0·50) and IDI of 0·14 (95% CI: 0·10–0·18). Both models, with tuned coefficients, showed a good predictive performance among participants in the validation dataset.

Interpretation The risk of total knee replacement can be predicted based on common risk factors with good discrimination and calibration. Additionally, adding radiographic findings of knee osteoarthritis into the model substantially improves its predictive performance.

Introduction

Total knee replacement is a common surgical procedure for the end stage of knee osteoarthritis.[1] Accompanied by an increasing prevalence of knee pain and knee osteoarthritis,[2] the incidence of total knee replacement has been steadily growing and is projected to substantially increase in the coming decades.[3,4] Predictive models that can accurately identify individuals at high risk for this surgery might allow both physicians and patients with knee pain or knee osteoarthritis to develop appropriate and efficient

non-surgical management strategies so that it can be postponed or avoided.

To date, a number of predictive models for total knee replacement have been developed.[5–9] Of them, three were developed based on a case-control study design,[5,8,9] whereas another study included only participants who received total knee replacement;[7] thus, the results from these studies cannot predict its absolute risk.[5,7–9] As a result, only one study served a prognostic purpose,[6] in which investigators developed a predictive model for the

Evidence before this study

Total knee replacement is a common surgical procedure for the end stage of knee osteoarthritis. Predictive models identifying individuals at high risk of total knee replacement can guide the development of appropriate and efficient non-surgical management strategies. We searched PubMed from inception to July 20, 2021, using the terms "predict" and "total knee replacement", with no language restrictions. A few predictive models for total knee replacement have been developed. However, most of them cannot be used to predict the absolute risk of total knee replacement and only one study served a prognostic purpose. The lack of data on knee-specific pain and function as well as radiographic findings could limit the clinical utility of the one previous predictive model.

Added value of this study

This study used data collected from two US multicentre cohorts to develop, tune, and externally validate two prognostic predictive models for the risk of total knee replacement among individuals with recent knee pain. Both models were more parsimonious but showed similar, if not superior, discrimination and calibration to the model

developed in the previous prognostic study. Adding radiographic findings of knee osteoarthritis into the predictive model not only reduced the number of predictors, but also improved the predictive performance.

Implications of all the available evidence

This study, along with all the previous studies, addresses an important clinical issue and our models show potential for clinical utility. To our knowledge, these are the first predictive models for total knee replacement that were developed and externally validated using data from multicentre cohort studies, taking advantage of well-defined, clinically relevant populations, standardised measurement of predictors and outcomes, and the use of validated instruments. This study also adds empirical evidence that radiographs, which are commonly collected in routine clinical care, significantly improved the performance of predicting future total knee replacement. Patients and clinicians should be informed by these findings that the risk of total knee replacement can be accurately estimated using several common risk factors, but radiographs, if available, provide added value and should be included.

risk of total knee replacement over 10 years using population-based electronic health records in the UK. However, because measures of knee-specific pain and function, as well as radiographic findings, were not available in the electronic health records, it is unclear whether adding these variables into the predictive model could improve the predictive performance. In addition, several variables included in the prediction model, such as occupational and leisure-time physical activities, were not assessed using validated questionnaires, making it difficult for others to adopt the model.[6]

Knee pain is a major factor leading to the decision to seek medical care among patients with knee osteoarthritis,[10] and treatment needs increase with pain severity among patients with osteoarthritis.[11] Physicians are likely to request a knee radiograph before they refer their patients to orthopaedic surgeons.[12] Radiographic findings are one of the key indications when orthopaedic surgeons decide to operate.[13,14] Despite the common use and clinical importance of radiographic findings, no study has evaluated to what extent adding radiographic findings improves the performance of a predictive model for total knee replacement among individuals who seek health care for knee pain. We aimed to develop predictive models for total knee replacement among individuals with knee pain using data from the Multicenter Osteoarthritis Study (MOST) and validate the models using data from the Osteoarthritis Initiative (OAI). We assessed whether adding radiographic findings to the predictive model could improve the model's performance.

Methods

Data sources and participants

We used data from MOST as the training dataset and from OAI as the validation dataset. MOST and OAI are two prospective, multicentre, longitudinal cohort studies of risk factors for the development and progression of knee osteoarthritis and knee pain among older adults (aged 50–79 years) with or at increased risk of knee osteoarthritis in the USA. Sociodemographic, lifestyle, and anthropometric factors, clinical assessments of knee pain and osteoarthritis, and radiological data (x-ray and MRI), as well as other measures and instruments, were longitudinally collected. The baseline visits of MOST were initiated in 2003 and of OAI were initiated in 2004. Both studies did a follow-up visit at 60 months. We included participants who reported knee pain in the past 3 months at their baseline visit (appendix p 5) and used follow-up data until 60 months so that participants of MOST and OAI had the same length of follow-up time. We selected only one knee from each participant based on worse severity of knee pain measured by either Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain questionnaire or other instruments (appendix p 2). Knees with a history of total knee replacement were excluded from the analyses.

The institutional review board at each of the sites of MOST and OAI approved the study. MOST and OAI obtained written informed consent from all participants.

	MOST training dataset (n=2658)	OAI validation dataset (n=4060)
Follow-up events		
Knee replacement	290 (10·9%)	174 (4·3%)
Followed up until 60 months	2211 (83·2%)	3409 (84·0%)
Death	65 (2·4%)	79 (1·9%)
Loss of follow-up	92 (3·5%)	398 (9·8%)
Follow-up time, mean (95% CI)	58·3 (58·1–58·6)	57·2 (56·9–57·5)
Age, years	62·4 (8·1)	60·9 (9·1)
<55	602 (22·6%)	1233 (30·4%)
55–59	436 (16·4%)	684 (16·8%)
60–64	542 (20·4%)	651 (16·0%)
65–69	487 (18·3%)	600 (14·8%)
70–74	361 (13·6%)	559 (13·8%)
≥75	230 (8·7%)	333 (8·2%)
Sex		
Male	1012 (38·1%)	1681 (41·4%)
Female	1646 (61·9%)	2379 (58·6%)
Race		
White	2191 (82·4%)	3179 (78·3%)
Non-White	467 (17·6%)	881 (21·7%)
Education		
High school and below	793/2658 (29·8%)	654/4023 (16·3%)
College	1228/2658 (46·2%)	1843/4023 (45·8%)
Graduate and above	637/2658 (24·0%)	1526/4023 (37·9%)
Married	1917/2644 (72·5%)	2672/4024 (66·4%)
Living alone	523 (19·7%)	933 (23·0%)
Smoker	1177/2657 (44·3%)	806/4007 (20·1%)
History of knee injury	900/2647 (34·0%)	1441/4015 (35·9%)
History of knee arthroscopy	348 (13·1%)	573 (14·1%)
History of hip fracture	33 (1·2%)	40 (1·0%)
History of spine or vertebrae fracture	131 (4·9%)	117 (2·9%)
Frequent knee pain	1906 (71·7%)	1806 (44·5%)
Use of medication		
Analgesics	1419 (53·4%)	862 (21·2%)
Glucosamine	701 (26·4%)	9 (0·2%)
Chondroitin	513 (19·3%)	8 (0·2%)
Hyaluronic acid	21 (0·8%)	1 (<0·1%)
Steroids	251 (9·4%)	2 (<0·1%)
		(Table 1 continues in next column)

	MOST training dataset (n=2658)	OAI validation dataset (n=4060)
(Continued from previous column)		
Kellgren-Lawrence grade		
0	960 (36·1%)	1306 (32·2%)
1	409 (15·4%)	677 (16·7%)
2	441 (16·6%)	1117 (27·5%)
3	534 (20·1%)	693 (17·1%)
4	314 (11·8%)	267 (6·6%)
Charlson comorbidity index		
0	1796/2658 (67·6%)	2999/4003 (74·9%)
1	490/2658 (18·4%)	626/4003 (15·6%)
2	244/2658 (9·2%)	253/4003 (6·3%)
3	77/2658 (2·9%)	78/4003 (1·9%)
≥4	51/2658 (1·9%)	47/4003 (1·2%)
Body-mass index, kg/m²	30·9 (6·1)	28·8 (4·9)
WOMAC pain score	5·0 (2·0–8·0)	3·0 (1·0–6·0)
WOMAC disability score	16·7 (12·9)	9·2 (11·6)
WOMAC stiffness score	2·6 (1·8)	2·1 (1·7)
CES-D score	6·0 (2·0–11·0)	5·0 (2·0–10·0)
PASE score	173·5 (88·0)	160·8 (82·4)
Number of other sites with frequent pain	3·0 (1·0–5·0)	1·0 (0·0–2·0)

Data are n (%), n/N (%), mean (SD), or median (IQR), unless stated otherwise. MOST=Multicenter Osteoarthritis Study. OAI=Osteoarthritis Initiative. WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index. CES-D=Center for Epidemiologic Studies Depression Scale. PASE=Physical Activity Scale for the Elderly.

Table 1: Characteristics of study populations for the knee replacement model in the training and validation cohorts

using Kellgren–Lawrence (KL) grade, for which all knee radiographs were read by both a musculoskeletal radiologist and a rheumatologist masked to clinical data. Disagreements between the two readings were resolved by a third reader.[15,16]

We identified 26 potential predictors, which were collected and assessed at baseline visit, based on an expert review of the current studies on the risk factors of total knee replacement by a consensus group of orthopaedic surgeons, led by the chief investigator (JL).[5–9,17,18] Candidate predictors included sociodemographic factors (age in years [<55, 55–59, 60–64, 65–69, 70–74, and ≥75], sex, race (White *vs* non-White), education attainment [high school and below, college, and graduate and above], married or not, living alone or not), lifestyle and anthropometric factors (body-mass index [BMI], smoking status [ever smoker *vs* non-smoker], Physical Activity Scale for the Elderly score), medical history (history of knee injury, frequent knee pain [ie, pain, aching, or stiffness on more than half of the past 30 days], history of knee arthroscopy, history of fracture in hip, history of fracture in spine or vertebrae, current use of analgesics [non-steroidal anti-inflammatory drugs, COX-2 inhibitors, and narcotics], glucosamine, or chondroitin, recent use [in the past 6 months] of hyaluronic acid or steroids, Charlson

Procedures

Total knee replacements that occurred during the 60-month follow-up period in both cohorts were identified and adjudicated by medical records, postoperative radiographs, or both. The date of surgery was obtained from medical records or was self-reported by participants when medical records were unavailable.

All participants in both studies underwent bilateral weight-bearing fixed flexion posteroanterior knee radiographs using the standardised protocol.[15] The severity of radiographic knee osteoarthritis was measured

	Clinical model		Radiographic model	
	Crude HR (95% CI)	Adjusted HR (95% CI)	Crude HR (95% CI)	Adjusted HR (95% CI)
Age, years				
<55	1 (ref)	1 (ref)	1 (ref)	1 (ref)
55–59	1·82 (1·13–2·93)	1·74 (1·08–2·8)	1·82 (1·13–2·93)	1·72 (1·06–2·78)
60–64	2·64 (1·72–4·05)	2·49 (1·62–3·84)	2·64 (1·72–4·05)	1·88 (1·22–2·90)
65–69	2·87 (1·87–4·42)	2·84 (1·83–4·40)	2·87 (1·87–4·42)	1·88 (1·21–2·9)
70–74	3·55 (2·28–5·51)	3·62 (2·31–5·67)	3·55 (2·28–5·51)	1·86 (1·19–2·91)
≥75	2·72 (1·62–4·54)	2·75 (1·63–4·66)	2·72 (1·62–4·54)	1·13 (0·67–1·9)
Sex				
Male	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Female	1·69 (1·3–2·19)	1·32 (1·01–1·73)	1·69 (1·3–2·19)	1·57 (1·2–2·05)
Race				
White	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Non-White	0·71 (0·51–1·00)	0·53 (0·36–0·77)	0·71 (0·51–1·00)	0·43 (0·30–0·61)
History of knee arthroscopy				
No	1 (ref)	1 (ref)	..	..
Yes	2·39 (1·84–3·11)	1·99 (1·52–2·61)	..	..
Frequent knee pain				
No	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Yes	5·18 (3·35–8·00)	2·70 (1·72–4·24)	5·18 (3·35–8·00)	1·98 (1·26–3·10)
Use of analgesics				
No	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Yes	2·43 (1·87–3·14)	1·67 (1·28–2·18)	2·43 (1·87–3·14)	1·66 (1·27–2·18)
Use of glucosamine				
No	1 (ref)	1 (ref)	..	..
Yes	1·97 (1·56–2·49)	1·67 (1·31–2·13)	..	..
Body-mass index	1·05 (1·03–1·07)	1·05 (1·03–1·06)	..	..
WOMAC pain score	1·16 (1·14–1·19)	1·12 (1·09–1·16)	1·16 (1·14–1·19)	1·08 (1·05–1·11)
Kellgren-Lawrence grade				
0	..	..	1 (ref)	1 (ref)
1	..	..	2·40 (0·84–6·84)	2·26 (0·79–6·46)
2	..	..	9·01 (3·94–20·6)	7·19 (3·14–16·5)
3	..	..	36·2 (16·9–77·4)	24·6 (11·4–53·0)
4	..	..	74·7 (34·9–160)	54·7 (25·3–118)

HR=hazard ratio. WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index. MOST=the Multicenter Osteoarthritis Study.

Table 2: Hazard ratios of knee replacement for predictors in the clinical model and radiographic model in the MOST training dataset (n=2658)

Comorbidity Index, Center for Epidemiologic Studies Depression Scale score), and pain and function assessment, (WOMAC pain score, WOMAC disability score, WOMAC stiffness score, and number of other sites [ie, neck, shoulders, elbows, wrists, hands and fingers, back, hips, and ankles and feet] with frequent pain [ie, pain, aching, or stiffness on more than half of the past 30 days]). We also included KL grade to reflect the radiographic severity of knee osteoarthritis. We combined similar predictors that measure similar constructs into single predictors, according to the recommendations from literature.[19] Although the willingness to receive total knee replacement is a strong risk factor affecting use of

the surgery,[20] it was not available in either study cohort and thus was not included.

Statistical analysis

We developed two predictive models for the risk of total knee replacement using the data collected from MOST. Participants were censored if they experienced death or loss to follow-up or were followed up until 60 months. Because only 65 (2·4%) participants in MOST and 79 (1·9%) participants in OAI died during 60 months of follow-up, we did not consider death to be a competing risk. We fitted a Cox proportional hazard model to predict the risk of total knee replacement. In the first regression model, hereafter referred to as the clinical model, we assessed 25 predictors, including the aforementioned predictors except KL grade. Specifically, we first examined the association between each predictor and the risk of total knee replacement using univariate analysis. Second, we included all variables with p<0·20 in the univariate analysis in a multivariable Cox proportional regression model with backward stepwise elimination (p>0·20) to obtain the final predictive model.[19,21] A significance level of 0·20 was chosen so that important predictors relevant to total knee replacement would not be missed and to avoid deleting less significant ones that could have practical and clinical implications.[22] We performed a ten-fold cross-validation to minimise overfitting.[23] We evaluated the model's discriminative ability using the C-statistic[24] and assessed calibration accuracy using the calibration slope and visual inspection of the calibration curve.[25] There were 290 (10·9%) incident cases of total knee replacement in the training dataset; the sample size was adequate for the assessment of 18 potential predictors with 27 corresponding degrees of freedom for total knee replacement in the current study.[26]

We used the same approach to develop another predictive model, hereafter referred to as the radiographic model, that included KL grade as well as the other 25 potential predictors. To evaluate whether adding KL grade could improve predictive performance, we calculated the category-free net reclassification improvement (NRI) and integrated discrimination improvement (IDI).[27,28] Finally, we derived nomograms to graphically present the predictive models using RStudio (through the regplot package) following the steps described by Bonnett and colleagues (appendix p 3).[29] The effect of the predictors on the risk of total knee replacement is represented in the format of axes, and risk points were attributed according to the predictive importance of the predictor of interest. We also made the prediction models available through a website calculator that provides individualised risk estimates based on the user's inputted predictor values. To compare the contribution of each predictor for the risk of total knee replacement in the predictive model, we obtained the standard regression coefficients from the Cox proportional hazard regression model.

We externally validated both clinical model and radiographic model among the participants in OAI. Specifically, we applied the models developed from MOST, with their predictors and assigned weights, to the eligible individuals in OAI and assessed the models' predictive performance.[30] When both models, with originally assigned weights, showed a poor calibration, largely because the risk of total knee replacement varied greatly between MOST (n=290, 10·9%) and OAI (n=174, 4·3%), we performed model-tuning to adjust the coefficient of each predictor (appendix pp 4, 11).[31] The validation results from the originally assigned weights are shown in the appendix (pp 7, 12). We reported the predictive performance of tuned models as primary results from external validation.

We conducted several sensitivity analyses. First, we collapsed KL grades into two (KL grade <2 and KL grade ≥2) or three categories (KL grade <2, KL grade of 2, and KL grade >2), and assessed the predictive performance of the radiographic model. In addition, we restricted the analyses to knees with a KL grade of 0–2 to examine the incremental predictive value of radiographic findings among participants without moderate to severe radiographic osteoarthritis.

All statistical analyses were performed using R version 4.0.3 and Stata/SE 15.1. The current study was conducted and reported in line with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines (appendix p 1).[32] We reported the performance measures with 95% CIs for the prediction model. Knees with missing values in variables of interest were excluded in each step of the development, tuning, and external validation of the models.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

We included 2658 participants from MOST (mean age 62·4 years [SD 8·1], 1646 [61·9%] women) and 4060 participants from OAI (mean age 60·9 years [9·1], 2379 [58·6%] women; table 1). Many of the baseline characteristics of participants in MOST were different from those of participants in OAI. The time of follow-up, frequency of total knee replacement and censoring events, and cumulative incidence of total knee replacement are shown in table 1 and the appendix (p 6). The incidence of total knee replacement was higher in MOST (n=290, 10·9%) than in OAI (n=174, 4·3%).

When developing the clinical model, 17 of 25 predictors were associated with the risk of total knee replacement (p<0·20) in a univariate proportional hazard model (appendix pp 13–15). Nine of these 17 potential predictors remained in the clinical model after backward stepwise

	Clinical model		Radiographic model	
	β′	Standard HR (95% CI)	β′	Standard HR (95% CI)
Age, years				
<55	··	1 (ref)	··	1 (ref)
55–59	0·55	1·73 (1·08–2·80)	0·54	1·71 (1·06–2·77)
60–64	0·91	2·49 (1·62–3·83)	0·63	1·88 (1·22–2·89)
65–69	1·04	2·83 (1·83–4·39)	0·63	1·87 (1·21–2·89)
70–74	1·28	3·61 (2·30–5·66)	0·62	1·85 (1·19–2·90)
≥75	1·01	2·75 (1·62–4·65)	0·12	1·13 (0·67–1·90)
Sex				
Male	··	1 (ref)	··	1 (ref)
Female	0·14	1·14 (1·01–1·3)	0·22	1·24 (1·09–1·42)
Race				
White	··	1 (ref)	··	1 (ref)
Non-White	–0·64	0·53 (0·36–0·77)	–0·85	0·43 (0·3–0·61)
History of knee arthroscopy				
No	··	1 (ref)	··	··
Yes	0·23	1·26 (1·15–1·38)	··	··
Frequent knee pain				
No	··	1 (ref)	··	1 (ref)
Yes	0·45	1·56 (1·28–1·92)	0·31	1·36 (1·11–1·66)
Use of analgesics				
No	··	1 (ref)	··	1 (ref)
Yes	0·25	1·29 (1·13–1·47)	0·25	1·29 (1·13–1·47)
Use of glucosamine				
No	··	1 (ref)	··	··
Yes	0·23	1·25 (1·13–1·39)	··	··
Body-mass index	0·27	1·31 (1·18–1·46)	··	··
WOMAC pain score	0·47	1·59 (1·41–1·80)	0·30	1·35 (1·19–1·54)
Kellgren-Lawrence grade				
0	··	··	··	1 (ref)
1	··	··	0·82	2·26 (0·79–6·5)
2	··	··	1·97	7·19 (3·13–16·5)
3	··	··	3·20	24·5 (11·4–52·9)
4	··	··	3·99	54·3 (25·1–117·5)

β′ was estimated from standardised variables. The standard HR was calculated using Cox regression from standardised variables. β′=standard regression coefficient. HR=hazard ratio. WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index. MOST=Multicenter Osteoarthritis Study.

Table 3: **Standard regression coefficient and hazard ratio for each predictor in the clinical model and the radiographic model in the MOST training dataset (n=2658)**

selection (table 2, appendix p 8). The retained predictive variables were age, sex, race, history of knee arthroscopy, frequent knee pain, current use of analgesics, current use of glucosamine, BMI, and WOMAC pain score. As indicated by the standard hazard ratios (table 3), the most predictive factors were age, race, and WOMAC pain score in the clinical model.

Seven predictors were retained in the radiographic model using the training dataset. These were age, sex, race, frequent knee pain, current use of analgesics,

	MOST training dataset (n=2658)		OAI validation dataset (n=2932)	
	Clinical model	Radiographic model	Clinical model	Radiographic model
C-statistic	0·79 (0·76 to 0·81)	0·87 (0·85 to 0·89)	0·78 (0·71 to 0·85)	0·88 (0·84 to 0·92)
Calibration slope	0·95 (0·86 to 1·05)	0·96 (0·87 to 1·04)	1·004 (0·81 to 1·19)	1·03 (0·95 to 1·10)
Calibration intercept	0·01 (–0·08 to 0·11)	0·01 (–0·04 to 0·05)	–0·004 (–0·18 to 0·18)	–0·02 (–0·09 to 0·04)
Net reclassification improvement	Ref	0·43 (0·38 to 0·50)	Ref	0·51 (0·39 to 0·59)
Integrated discrimination improvement	Ref	0·14 (0·10 to 0·18)	Ref	0·09 (0·06 to 0·11)

MOST=Multicenter Osteoarthritis Study. OAI=Osteoarthritis Initiative. Ref=reference.

Table 4: Performance in discrimination and calibration and measures of improvement of the radiographic model compared with the clinical model in the development and validation cohorts

WOMAC pain score, and KL grade (table 2, appendix p 8). The most predictive factors were KL grade, race, and age in the radiographic model (table 3).

The predictive performances of the clinical model and the radiographic model are shown in table 4. The C-statistic from the ten-fold cross-validation of the clinical model from the training dataset was 0·79 (95% CI 0·76 to 0·81). Visual inspection of the calibration plot indicated that the model predicted the risk of total knee replacement well (figure A). The calibration slope was 0·95 (95% CI 0·86 to 1·05) with an intercept of 0·01 (95% CI –0·08 to 0·11), suggesting the model predicted the risk of total knee replacement accurately. The C-statistic from the ten-fold cross-validation of the radiographic model from the training dataset was 0·87 (0·85 to 0·89). Visual inspection of the calibration plot indicated that the model predicted the risk of total knee replacement well (figure B). The calibration slope was 0·96 (0·87 to 1·04) with an intercept of 0·01 (–0·04 to 0·05), indicating that the radiographic model predicted the risk of total knee replacement accurately. Both NRI (0·43, 95% CI 0·38 to 0·50) and IDI (0·14, 95% CI 0·10 to 0·18) were statistically significant when KL grade was included in the predictive model, suggesting that adding radiographic findings improved the model's predictive performance.

The models were tuned with an adjustment size of 0·839 for the clinical model and 0·941 for the radiographic model, using the validation dataset. The C-statistic was 0·78 (95% CI 0·71 to 0·85) for the tuned clinical model and 0·88 (0·84 to 0·92) for the tuned radiographic model (table 4). The calibration slopes were 1·004 (95% CI 0·81 to 1·19) for the clinical model and 1·03 (0·95 to 1·10) for the radiographic model, with a corresponding intercept of –0·004 (95% CI –0·18 to 0·18) for the clinical model and –0·02 (–0·09 to 0·04) for the radiographic model. The calibration plots of both the clinical model (figure C) and the radiographic model (figure D) in the validation dataset showed similar patterns to those in the training dataset. Adding radiographic findings significantly improved predictive performance, with an NRI of 0·51 (95% CI 0·39 to 0·59) and an IDI of 0·09 (95% CI 0·06 to 0·11; table 4).

In the sensitivity analyses in which KL grade was collapsed into either two (<2 and ≥2) or three (<2, 2, and >2) categories, the radiographic model still showed a better predictive performance than the clinical model (appendix p 16). Both models also showed good generalisability when applied to validation datasets. When we restricted analyses to knees with a KL grade of 0–2, the radiographic model showed better discrimination and calibration than the clinical model (appendix p 17). The calibration slope of the radiographic model was closer to 1·00 than that of the clinical model, the calibration intercept was closer to zero, and the NRI was significantly positive, suggesting that the radiographic model predicts the risk of total knee replacment more accurately than the clinical model.

Discussion
We developed two prognostic predictive models for the risk of total knee replacement among individuals with recent knee pain in MOST. Although both the clinical model and the radiographic model showed good predictive ability for the risk of total knee replacement within 60 months, the radiographic model was more parsimonious (ie, seven *vs* nine predictors) and had a better predictive ability than the clinical model. Both models, with tuned coefficients, showed a good predictive performance among participants of OAI.

Although total knee replacement can greatly improve symptoms, physical function, and quality of life for patients with end-stage knee osteoarthritis, the surgical procedure itself is neither inexpensive nor risk-free. Previous studies have reported that around 0·4% of patients died within 90 days after arthroplasty surgery,[33] and the risk of venous thromboembolism increased by almost six times over 1 year after a surgical procedure on the knee.[34] Furthermore, approximately 20% of patients report dissatisfaction after primary total knee replacement.[35] Addressing modifiable predictors associated with an increased risk of total knee replacement might help to delay its need. For example, a history of knee arthroscopy has been associated with a higher risk of total knee replacement.[36] Guidelines make recommendations against knee arthroscopy among patients with degenerative knee osteoarthritis and meniscus tear.[37] Therefore, it is crucial to follow the guidelines for osteoarthritis management regarding knee arthroscopy. As another example, weight loss though diet and exercise could mitigate symptoms of knee osteoarthritis and slow down the progression of structure damage, which might postpone the need for total knee replacement.[38] By contrast, patients could experience poor outcomes due to
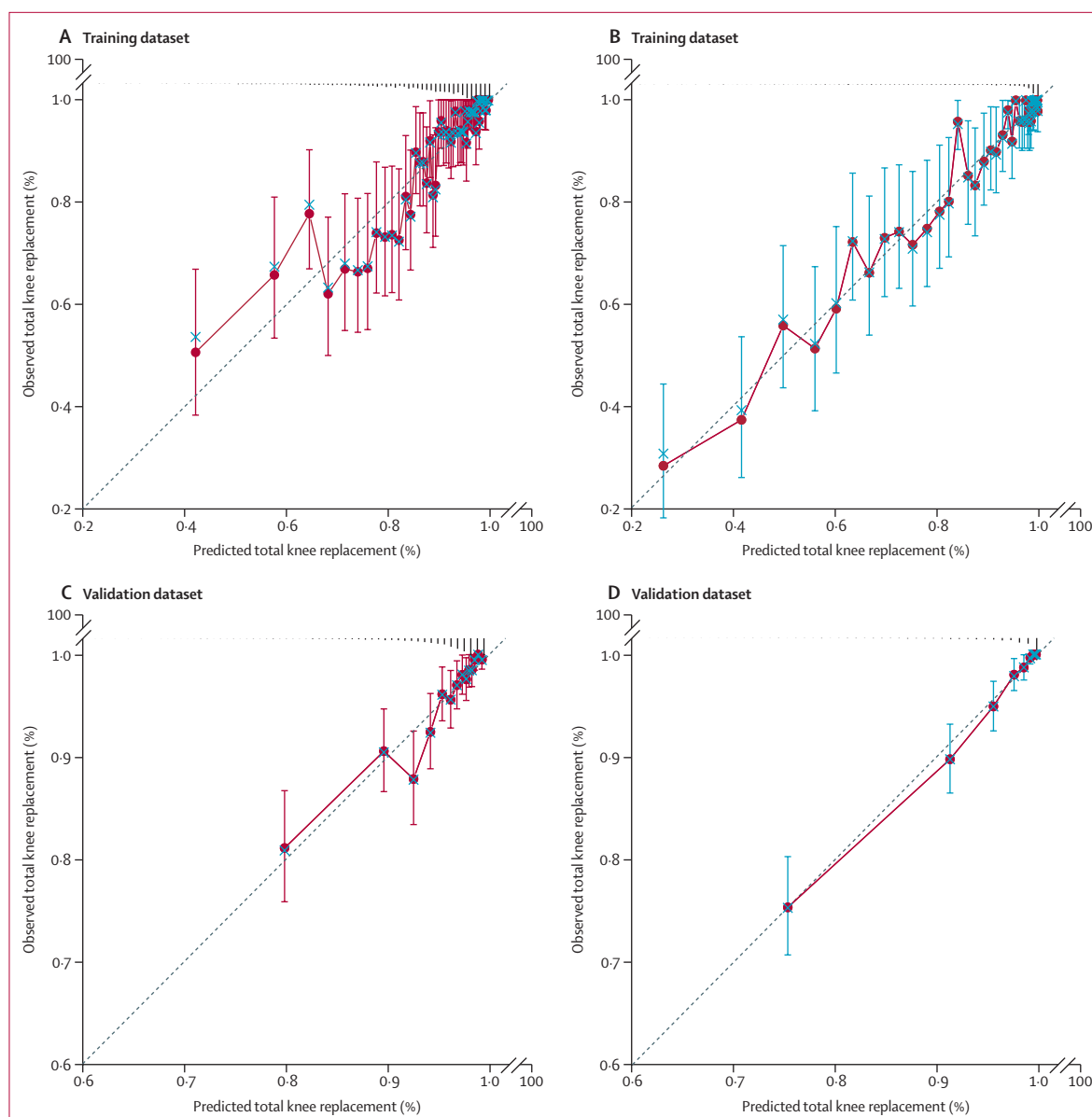
***Figure:* Calibration plots showing predictive performances of the clinical model and radiographic model**
Calibration plots are shown for the clinical model (A) and radiographic model (B) in the training dataset and for the clinical model (C) and radiographic model (D) in the validation dataset.

delayed total knee replacment.[39] Thus, it is imperative to delay total knee replacement for patients who do not reach the indication for it, while recommending it in a timely manner for those who do.[40] However, substantial numbers of patients have had either premature total knee replacement or delayed surgery when indicated, probably due to patients' preferences as well as clinicians' attitudes and personal experiences towards this surgery.[13] Using a model that is independent of clinicians' scientific knowledge and experience and patients' preference as a tool to aid clinical decision making might improve the outcome for patients.[41,42]

To our knowledge, only one study has previously reported a prognostic predictive model for the risk of total knee replacement.[6] The model consisted of 24 predictors, including demographic and socioeconomic factors and medical history, with a C-statistic of 0·79. The investigators noted that the lack of measures of pain and structural markers could jeopardise the model's predictive performance. In the current study, both models showed better discrimination, calibration ability, and external validity with fewer predictors included, supporting their potential clinical utility. Furthermore, the participants' baseline characteristics, especially in the validation

dataset, were very similar to those in the previous study that used population-based electronic health records (appendix p 18),[6] suggesting the models in the current study could be generalisable to patients with similar risk profiles in clinical practice. Of note, unlike making causal inference, a predictor is not necessarily the cause of an outcome event. For example, current use of analgesics and glucosamine were predictors for a higher risk of total knee replacement, but these medications are taken to address the symptoms of knee osteoarthritis.

It is not surprising that adding radiographic findings into the predictive model improves the performance for the risk of total knee replacement. Previous studies have shown that the severity of radiographic osteoarthritis is strongly associated with pain severity.[43,44] In addition, many risk factors, such as knee injury, kneeling, or habitual physical activity, are strong risk factors for knee radiographic osteoarthritis; however, these risk factors are difficult to assess accurately.[45,46] As a result, adding these variables into a predictive model for total knee replacement could jeopardise the predictive performance, owing to random misclassification of these variables. KL grade, a measure of structural lesions of knee osteoarthritis, is likely to be part of the causal pathway between these risk factors and total knee replacement,[47] and thus is likely to partially represent their effects on total knee replacement. Finally, radiographic evidence and severity are deemed as important evidence when orthopaedic surgeons make the decision to do this surgery in their clinical practice.[13,14] Nonetheless, the clinical model, which relies on common risk factors, showed good discrimination and calibration, supporting the guidelines that recommend against routinely prescribing radiographs in the management of knee osteoarthritis.[48] However, our findings also suggest that when radiographs are already available at the time of patient consultation, clinicians should use them in predicting the risk of total knee replacement.

Our study has several strengths. First, our predictive models, especially the radiographic model, are more parsimonious than the one previously reported, potentially making them easier to use in the clinical setting. Second, all predictors included in the two models can be collected when patients seek clinical care by either validated questionnaire or commonly collected knee radiographs, which can be graded using the standard grading scheme. Thus, the quality of data would be satisfactory for clinical use. Finally, our models were developed for a well-defined and clinically relevant population, (ie, individuals with recent knee pain who are likely to seek clinical care), supporting their utility in clinical settings.

Several limitations of our study are worth noting. First, the approaches to obtain radiographic views and their rigorous assessment in research settings could differ from those in routine clinical practice. For example, the standing anteroposterior view, instead of the semi-flexed

view, is common in clinical practice. However, the use of standing anteroposterior views was found to be similar to semi-flexed views in detecting changes in KL grade in analyses using a grade of 2 or higher as the knee osteoarthritis definition.[49] In addition, the accuracy of knee radiographic reading by clinical physicians might not be as high as that by the investigators in MOST and OAI. Nevertheless, appropriate training of either radiologists or physicians can improve radiographic reading.[50] Furthermore, we found that the model still showed an excellent predictive ability when KL grades were collapsed into two (<2 and ≥2) or three (<2, 2, and >2) categories. Therefore, different radiographic views and some degree of misclassification might not materially affect the predictive performance. Second, willingness to receive total knee replacement, a strong contributor to racial disparity in its use,[20] was not assessed at the baseline visit in either MOST or OAI, and thus was not included in the model development. Future studies are needed to assess to the extent to which willingness to receive total knee replacement affects the performance of the predictive model. Finally, the generalisability of our findings should be tested in populations outside of the USA.

In conclusion, the risk of total knee replacement within 60 months for individuals with recent knee pain can be predicted based on common risk factors with good discrimination and calibration. Adding radiographic findings of knee osteoarthritis into the model substantially improved predictive performance.

by the NIH and conducted by the Osteoarthritis Initiative Study Investigators. Private-sector funding for the Osteoarthritis Initiative is managed by the Foundation for the NIH.

References

1   Robertsson O, Bizjajeva S, Fenstad AM, et al. Knee arthroplasty in Denmark, Norway and Sweden. A pilot study from the Nordic Arthroplasty Register Association. *Acta Orthop* 2010; **81:** 82–89.

2   Nguyen US, Zhang Y, Zhu Y, Niu J, Zhang B, Felson DT. Increasing prevalence of knee pain and symptomatic knee osteoarthritis: survey and cohort data. *Ann Intern Med* 2011; **155:** 725–32.

3   Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK. Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink. *Osteoarthritis Cartilage* 2015; **23:** 594–600.

4   Ackerman IN, Bohensky MA, Zomer E, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. *BMC Musculoskelet Disord* 2019; **20:** 90.

5   Chan WP, Hsu SM, Huang GS, Yao MS, Chang YC, Ho WP. Creation of a reflecting formula to determine a patient's indication for undergoing total knee arthroplasty. *J Orthop Sci* 2010; **15:** 44–50.

6   Yu D, Jordan KP, Snell KIE, et al. Development and validation of prediction models to estimate risk of primary total hip and knee replacements using data from the UK: two prospective open cohorts using the UK Clinical Practice Research Datalink. *Ann Rheum Dis* 2019; **78:** 91–99.

7   Heisinger S, Hitzl W, Hobusch GM, Windhager R, Cotofana S. Predicting total knee replacement from symptomology and radiographic structural change using artificial neural networks—data from the Osteoarthritis Initiative (OAI). *J Clin Med* 2020; **9:** E1298.

8   Leung K, Zhang B, Tan J, et al. Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: data from the Osteoarthritis Initiative. *Radiology* 2020; **296:** 584–93.

9   Tolpadi AA, Lee JJ, Pedoia V, Majumdar S. Deep learning predicts total knee replacement from magnetic resonance images. *Sci Rep* 2020; **10:** 6371.

10 Hadler NM. Knee pain is the malady—not osteoarthritis. *Ann Intern Med* 1992; **116:** 598–99.

11 Jackson J, Iyer R, Mellor J, Wei W. The burden of pain associated with osteoarthritis in the hip or knee from the patient's perspective: a multinational cross-sectional study. *Adv Ther* 2020; **37:** 3985–99.

12 Chen A, Balogun-Lynch J, Aggarwal K, Dick E, Gupte CM. Should all elective knee radiographs requested by general practitioners be performed weight-bearing? *Springerplus* 2014; **3:** 707.

13 Gademan MG, Hofstede SN, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Indication criteria for total hip or knee arthroplasty in osteoarthritis: a state-of-the-science overview. *BMC Musculoskelet Disord* 2016; **17:** 463.

14 Schmitt J, Lange T, Günther KP, et al. Indication criteria for total knee arthroplasty in patients with osteoarthritis—a multi-perspective consensus study. *Z Orthop Unfall* 2017; **155:** 539–48.

15 Felson DT, Nevitt MC, Yang M, et al. A new approach yields high rates of radiographic progression in knee osteoarthritis. *J Rheumatol* 2008; **35:** 2047–54.

16 Hunter DJ, Niu J, Zhang Y, et al. Change in cartilage morphometry: a sample of the progression cohort of the Osteoarthritis Initiative. *Ann Rheum Dis* 2009; **68:** 349–56.

17 Apold H, Meyer HE, Nordsletten L, Furnes O, Baste V, Flugsrud GB. Risk factors for knee replacement due to primary osteoarthritis, a population based, prospective cohort study of 315,495 individuals. *BMC Musculoskelet Disord* 2014; **15:** 217.

18 Nielsen FK, Egund N, Jørgensen A, Jurik AG. Risk factors for joint replacement in knee osteoarthritis; a 15-year follow-up study. *BMC Musculoskelet Disord* 2017; **18:** 510.

19 Moons KG, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012; **98:** 683–90.

20 Ibrahim SA, Siminoff LA, Burant CJ, Kwoh CK. Differences in expectations of outcome mediate African American/white patient differences in "willingness" to consider joint replacement. *Arthritis Rheum* 2002; **46:** 2429–35.

21 Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003; **56:** 441–47.

22 Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. *Fam Med Community Health* 2020; **8:** e000262.

23 Lee M, Cronin KA, Gail MH, Feuer EJ. Predicting the absolute risk of dying from colorectal cancer and from other causes using population-based cancer registry data. *Stat Med* 2012; **31:** 489–500.

24 Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982; **247:** 2543–46.

25 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21:** 128–38.

26 Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995; **48:** 1503–10.

27 Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; **27:** 157–72, discussion 207–12.

28 Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011; **30:** 11–21.

29 Bonnett LJ, Snell KIE, Collins GS, Riley RD. Guide to presenting clinical prediction models for use in clinical settings. *BMJ* 2019; **365:** l737.

30 Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012; **98:** 691–98.

31 Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; **17:** 230.

32 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015; **162:** 55–63.

33 Berstock JR, Beswick AD, López-López JA, Whitehouse MR, Blom AW. Mortality after total knee arthroplasty: a systematic review of incidence, temporal trends, and risk factors. *J Bone Joint Surg Am* 2018; **100:** 1064–70.

34 Lu N, Misra D, Neogi T, Choi HK, Zhang Y. Total joint arthroplasty and the risk of myocardial infarction: a general population, propensity score-matched cohort study. *Arthritis Rheumatol* 2015; **67:** 2771–79.

35 Gunaratne R, Pratt DN, Banda J, Fick DP, Khan RJK, Robertson BW. Patient dissatisfaction following total knee arthroplasty: a systematic review of the literature. *J Arthroplasty* 2017; **32:** 3854–60.

36 Rongen JJ, Rovers MM, van Tienen TG, Buma P, Hannink G. Increased risk for knee replacement surgery after arthroscopic surgery for degenerative meniscal tears: a multi-center longitudinal observational study using data from the osteoarthritis initiative. *Osteoarthritis Cartilage* 2017; **25:** 23–29.

37 Siemieniuk RAC, Harris IA, Agoritsas T, et al. Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. *BMJ* 2017; **357:** j1982.

38 Messier SP, Mihalko SL, Legault C, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. *JAMA* 2013; **310:** 1263–73.

39 Ravi B, Croxford R, Austin PC, et al. The relation between total joint arthroplasty and risk for serious cardiovascular events in patients with moderate-severe osteoarthritis: propensity score matched landmark analysis. *BMJ* 2013; **347:** f6187.

40 Ghomrawi HMK, Mushlin AI, Kang R, et al. Examining timeliness of total knee replacement among patients with knee osteoarthritis in the U.S.: results from the OAI and MOST longitudinal cohorts. *J Bone Joint Surg Am* 2020; **102:** 468–76.

41 Bullock GS, Hughes T, Sergeant JC, Callaghan MJ, Riley R, Collins G. Methods matter: clinical prediction models will benefit sports medicine practice, but only if they are properly developed and validated. *Br J Sports Med* 2021; bjsports-2021-104329.

42   Kattan MW, Yu C, Stephenson AJ, Sartor O, Tombal B. Clinicians versus nomogram: predicting future technetium-99m bone scan positivity in patients with rising prostate-specific antigen after radical prostatectomy for prostate cancer. *Urology* 2013; **81:** 956–61.

43   Neogi T, Felson D, Niu J, et al. Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. *BMJ* 2009; **339:** b2844.

44   Riddle DL, Jiranek WA. Knee osteoarthritis radiographic progression and associations with pain and function prior to knee arthroplasty: a multicenter comparative cohort study. *Osteoarthritis Cartilage* 2015; **23:** 391–96.

45   Briggs KK, Kocher MS, Rodkey WG, Steadman JR. Reliability, validity, and responsiveness of the Lysholm knee score and Tegner activity scale for patients with meniscal injury of the knee. *J Bone Joint Surg Am* 2006; **88:** 698–705.

46   Lee IM, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *Br J Sports Med* 2014; **48:** 197–201.

47   Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002; **155:** 176–84.

48   Sakellariou G, Conaghan PG, Zhang W, et al. EULAR recommendations for the use of imaging in the clinical management of peripheral joint osteoarthritis. *Ann Rheum Dis* 2017; **76:** 1484–94.

49   Emrani PS, Katz JN, Kessler CL, et al. Joint space narrowing and Kellgren-Lawrence progression in knee osteoarthritis: an analytic literature synthesis. *Osteoarthritis Cartilage* 2008; **16:** 873–82.

50   Riddle DL, Jiranek WA, Hull JR. Validity and reliability of radiographic knee osteoarthritis measures by arthroplasty surgeons. *Orthopedics* 2013; **36:** e25–32.