Knee Tissue Lesions and Prediction of Incident Knee osteoarthritis over 7 Years In A Cohort of Persons At Higher Risk

Leena Sharma, MD[1], **Marc Hochberg, MD, MPH**[2], **Michael Nevitt, PhD**[3], **Ali Guermazi, MD, PhD**[4], **Frank Roemer, MD**[4,5], **Michel D. Crema, MD**[4,6], **Charles Eaton, MD, MS**[7], **Rebecca Jackson, MD**[8], **Kent Kwoh, MD**[9], **Jane Cauley, DrPH**[10], **Orit Almagor, MA**[1], and **Joan S. Chmiel, PhD**[1]

[1]Northwestern University, Chicago, IL

[2]University of Maryland, Baltimore, MD

[3]University of California at San Francisco, San Francisco, CA

[4]Boston University, Boston, MA

[5]University of Erlangen-Nuremberg, Erlangen, Germany

[6]Saint-Antoine Hospital, Paris VI University, Paris, France

[7]Brown University, Pawtucket, RI

[8]The Ohio State University, Columbus, OH

[9]University of Arizona, Tucson, AZ

[10]University of Pittsburgh, Pittsburgh, PA

Abstract

Objective—Among high risk individuals, whether knee lesions in tissues involved in osteoarthritis can improve prediction of knee osteoarthritis is unclear. We hypothesized that models predicting 1) incident osteophytes and 2) incident osteophytes and joint space narrowing can be improved by including symptoms or function, and further improved by lesion status.

Design—In Osteoarthritis Initiative participants with normal knee x-rays, we assessed cartilage damage, bone marrow lesions (BMLs), and menisci. Cox proportional hazards models were used to develop risk prediction models for risk of each outcome. Nested models (increasingly larger baseline covariable sets) were compared using likelihood ratio tests and Schwarz Bayesian

Corresponding Author: Leena Sharma, M.D. 240 E. Huron, M300, Division of Rheumatology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA, 312-908-5597, L-Sharma@northwestern.edu.

Information Criterion. Model discrimination used receiver operating characteristic curves and area under the curve (AUC).

Results—In 841 participants [age 59.6, BMI 26.7, 55.9% women] over up to 7 years follow-up, each larger set improved prediction (+hand osteoarthritis, injury, surgery, activities; +symptoms/ function). Prediction was further improved by including cartilage damage both compartments, BMLs both compartments, meniscal tear, meniscal extrusion, sum of lesion types, number of subregions with cartilage damage, number of subregions with BMLs, and (concurrently) subregion number with cartilage damage, subregion number with BMLs, and meniscal tear. AUCs were 0.80 for both outcomes for number of subregions with cartilage damage and the combined model.

Conclusions—Among persons at higher risk for knee osteoarthritis with normal x-rays, MRI tissue lesions improved prediction of mild as well as moderate disease. These findings support that disease onset is likely occurring during the "high-risk" period and encourage a reorientation of approach.

Introduction

Current treatments for knee osteoarthritis (OA) may help symptoms but do not affect disease progression. Disease modification requires tackling the multi-faceted, downward spiral of joint tissue events that is progressive knee OA. Efforts to delay disease development and early-stage intervention may be more cost-effective than treatment of established OA[1,2]. The widely established definition of knee OA (radiographic Kellgren/Lawrence (KL) grade 2) hinges on unequivocal osteophyte presence. However, in knees graded KL 0 (normal) or KL 1 (possible osteophyte), knee tissue lesions are not infrequent in MRI studies, whether they include only persons at higher risk[3,4] or not[5,6]. Evidence is accumulating to support the hypothesis that these lesions are not incidental[3,4,7-13]. These lesions occur in tissues typically involved by the whole-organ disease of OA, i.e., cartilage, subchondral bone, and menisci; given this, they likely are not risk factors per se but may represent signs of disease. Whether they represent early OA is a challenging question to address at a study population level. Risk prediction modeling in this context enables evaluation of whether tissue lesion status improves the prediction of incident knee OA using the established definition (after considering known predictors), as would be expected if these lesions represent early disease.

To our knowledge, previous studies have not examined whether tissue lesion status can improve prediction of incident knee OA risk. Furthermore, previous studies of risk prediction models, while groundbreaking, were not limited to radiographically normal (KL 0) knees, the status of the vast majority of the high risk population. Zhang et al developed a risk prediction model for incident knee OA (defined as KL 2), incorporating age, gender, body mass index (BMI), occupational kneeling/lifting, injury, and family history of OA; to validate the model, they used Osteoarthritis Initiative (OAI) and Genetics of Osteoarthritis and Lifestyle (GOAL) data[14]. In the Rotterdam Study (RS)-I (with validation in RS-II and the Chingford Study), Kerkhof et al developed a prediction model for incident KL 2; questionnaire variables, genetic score, or a urinary biomarker did not improve prediction [comparing the area under a receiver operating characteristic (ROC) curve (AUC)] vs. age,

gender, and BMI[15]. The AUC was improved by adding baseline KL 1 (possible osteophytes)[15].

We undertook MR image readings in a cohort of OAI participants who were KL 0 in both knees (since risk of knee OA is increased by contralateral knee OA). We hypothesized that, in persons KL 0 in both knees, models for prediction of 1) incident KL 2 and 2) incident KL 2 and joint space narrowing can be improved by including baseline symptoms or function, and further improved by inclusion of knee lesion status, as would be expected if these lesions represented an early stage of disease. The widely established definition of knee OA (KL 2) is the basis of a large literature documenting impact and burden of knee OA. However, because KL = 2 knees may or may not progress beyond a mild stage, we included a secondary outcome, requiring the additional presence of joint space narrowing, corresponding with moderate disease.

Methods

The OAI is a prospective, observational cohort study of persons with or at higher risk for knee OA, enrolled in Baltimore MD, Columbus OH, Pittsburgh PA, or Pawtucket RI, between February 2004 and May 2006. OAI incidence subcohort eligibility required characteristics associated with increased risk of developing knee OA [symptoms, overweight, injury, surgery, family history of total knee replacement (TKR), Heberden's nodes, repetitive knee bending, age 70-79 years] and absence of: symptomatic knee OA; inflammatory arthritis; severe bilateral joint space narrowing; TKR and severe contralateral narrowing; bilateral TKR or planned within 3 years; MRI contraindications; aide > 50% of ambulation (except 1 cane); severe comorbidity; double-blind trial participation[16,17]. An additional requirement for our study was bilateral KL = 0 (by centralized readings) at the 12-month visit, our ancillary study's baseline MRI assessment (per the timing of the award of the grant that added other assessments to the 12-month evaluation and funded the MRI readings). The Institutional Review Board at each site approved the study.

Tissue Lesions

MR image acquisition utilized 3.0T Siemens Trio scanners at each site. Sequences included coronal intermediate-weighted (IW) turbo spin echo (TSE), sagittal IW TSE with fat-suppression, and 3D Dual Echo Steady State water excitation, acquired in the sagittal plane with coronal and axial multiplanar reconstructions; acquisition parameters are described in detail in the publicly available OAI manual[18].

We undertook right knee (left knee, if right knee images technically unacceptable) MRI readings in persons determined by the coordinating center to meet the KL criterion at the 12-month visit. Three expert musculoskeletal radiologists (MC, AG, FWR) used the MRI OA Knee Score (MOAKS)[19], blinded to hypotheses, KL criterion, and all study data. In terms of lesion assessment relevant to this study, with the exception of inter-rater reliability for tibial cartilage surface area, all measures of reliability were very good (0.61-0.8) or reached near-perfect agreement (0.81-1.0) according to the criteria developed by Landis and Koch[19,20]. The low prevalence of certain features in certain sub-regions may have adversely affected the kappa results hence the percent agreement was also calculated. All features relevant to

this study were scored with overall percent agreement above 75% for both the intra- and inter-reader exercise (19). Paired images were read with chronology known[21]. Cartilage morphology was scored in 4 patellofemoral (PF) and 10 tibiofemoral (TF) subregions: **0**, normal; **1.0**, 1-10% area damaged, no full-thickness; **1.1**, 1-10% area, 1-10% full-thickness; **2.0**, 10-75% area, no full-thickness; **2.1**, 10-75% area, 1-10% full-thickness; **2.2**, 10-75% area, 10-75% full-thickness; **3.0**, > 75% area, no full-thickness; **3.1**, > 75% area, 1-10% full-thickness; **3.2**, > 75% area, 10-75% full-thickness; **3.3**, > 75% area, > 75% full-thickness. Bone marrow lesions (BMLs) were scored in the same subregions: **0**, none; **1**, < 25% of subregion; **2**, 25-50%; **3**, > 50%. For each meniscus, 3 subregions were scored: **0**, normal; **1**, signal abnormality; **2**, horizontal tear; **3**, vertical tear; **4**, complex tear; **5**, root tear; **6**, partial maceration; **7**, progressive partial maceration**; 8**, complete maceration. Extrusion was scored for each meniscus [22]: **0**, none; **1**, < 50% extruded; **2**, ≥ 50% extruded[19,22].

For cartilage damage and BMLs: "any" was defined as score > 0 in ≥ 1 tibiofemoral or patellofemoral subregion; "both tibiofemoral and patellofemoral" was defined as score > 0 in ≥ 1 tibiofemoral and ≥ 1 patellofemoral subregion; and "number of subregions" was number of tibiofemoral and patellofemoral subregions with score > 0. Meniscal tear was defined by any subregion score > 1, and extrusion by any score > 0. Sum of lesion types was the number of lesion types present (0-4). Maximum cartilage damage (surface area) was defined as most severely damaged cartilage across all knee subregions (0 = normal; 1 = MOAKS 1.0 or 1.1; 2 = MOAKS 2.0, 2.1, or 2.2; and 3 = MOAKS 3.0, 3.1, 3.2, or 3.3) and maximum cartilage damage (full thickness) as most severely damaged (0 = normal; 1 = 1.0, 2.0, or 3.0; 2 = 1.1, 2.1, or 3.1; 3 = 2.2 or 3.2; and 4 = 3.3). Maximum BML severity was defined as worst BML score across all subregions.

Predictors

Predictors were also assessed at the 12-month visit. BMI was analyzed as a categorical variable [normal (reference category), overweight (BMI ≥ 25 and < 30 kg/m^2), obese (BMI ≥ 30)]. Race was analyzed as African-American vs. other. Family history of TKR for OA was defined by self-report for biological parent or sibling. Hand OA was assessed by: self-reported hard bumps on joints nearest fingertips; and exam, as ≥ 2 bony enlargements of distal interphalangeal or thumb interphalangeal joints in ≥ 1 hand. In the study knee, injury was defined as ever injured enough to limit walking ≥ 2 days, and surgery as any previous surgery. Isometric knee extensor strength was measured using the Good Strength chair (Metitur, Jyvaskyla, Finland) at 60° from full extension. Frequent knee bending was defined as, on most days, climb total of ≥ 10 flights of stairs, kneel ≥ 30 minutes, squat/deep knee bend ≥ 30 minutes, or lift/move objects ≥ 25 lbs by hand. Individual activities were assessed as, during ≥ 1 in the past 30 days: kneel ≥ 30 minutes; squat ≥ 30 minutes (SQUAT); get in/out of squatting position ≥ 10 times; lift/move objects ≥ 25 lbs (LIFT); and climb total of ≥ 10 flights. Comorbidity was assessed using a questionnaire adaptation of the Charlson Index[23].

WOMAC Pain, Stiffness, and Function[24], adapted in the OAI to separately score each knee, were analyzed for the study knee, dichotomized by ≥ 1 item scored ≥ 2 (≥ "moderate"). KOOS Symptoms and Quality of Life scores [25] were each dichotomized by ≥ 1 item

"moderate". Medication use was defined as use for knee symptoms half the days of 1 month in the past 12 months.

Radiographic Outcome

The posteroanterior fixed-flexion weight-bearing protocol[16,26] with a SynaFlexer™ frame was used. In centralized readings, experts (weighted kappa between-reader agreement 0.79), blinded to other's reading, hypotheses, and all other data, assessed KL[27]. Adjudication for KL 0-1 vs. 2 included a third reader[28]. KL 2 was the primary outcome and KL 2 and joint space narrowing was a secondary outcome that is provided in the public data release as a more advanced OA stage, based on follow-up data through the 96-month visit. It is highly recommended by the OAI Coordinating Center to use the released calculated outcome data for statistical analyses, as opposed to applying raw radiographic feature data to define outcomes.

Statistical Analysis

All analyses used 1 knee/person (right; left, if image quality of right unacceptable). Baseline predictors were analyzed (12-month for lesions). Follow-up time for persons without an outcome occurrence was censored at the last visit where the outcome was assessed. Persons with an outcome contributed follow-up time until the outcome was first documented. Cox proportional hazards regression models were used to model risk (expressed as a hazard function) for developing each outcome. We constructed a base model using baseline predictors from univariate analyses of plausible predictors; to be selected, each predictor had to have a p-value < 0.10 for the univariate hazard ratio (HR). Since findings were similar in univariate analyses for each outcome, we used the same base model for both outcomes. We then added pain or function variables and finally lesion variables (meeting the p-value threshold). Results are summarized using adjusted HRs and 95% confidence intervals (CIs); nested proportional hazards models (with increasingly larger covariable sets) were compared using a likelihood ratio chi-square test (LR) with appropriate degrees of freedom (df) and Schwarz Bayesian Information Criterion (SBC). The LR test statistically compares the lesion model to the model without the lesion. More global measures of fit based on the log likelihood function, such as the SBC value, can be used as summary measures to compare the predictive value of the models vs. the base model. Lower SBC values suggest model improvement compared to a model with a larger SBC.

Although insensitive for the purpose of assessing the impact of adding a new predictor to a model, the AUC is useful to assess discrimination (probability that an individual with outcome will be assigned a higher risk than an individual without outcome). We calculated the AUC of the ROC for logistic regression models that included as predictors variable sets that significantly improved prediction in the corresponding proportional hazards model. To assess calibration (comparison of observed and predicted risks), we used Hosmer-Lemeshow X^2 statistics for goodness-of-fit; large p-values indicate good calibration. OAI files used in these analyses include: Subject (version 9.2.1); Enrollees (version 22); Physexam (version 8.2.1); medhist (version 9.2.1); Jointsx (version 9.2.1); outcomes99 (version 8). SAS software version 9.3 (SAS Institute Inc., Cary, NC) was used.

Results

Figure 1 delineates sample derivation. The 841 persons in the analysis sample [mean age 58.6 years (SD 8.8), BMI 26.7 kg/m^2 (4.2), 471 (56.0%) women] were similar to persons KL 0 in both knees and not in the sample [age 58.5 (9.2), BMI 27.1 (4.8), 62.6% women]. Of the 841 knees from 841 persons in the analysis sample, 53 developed incident KL >2 and 36 developed KL >2 and joint space narrowing by their final x-ray acquisition. Table 1 shows, for each time point, the number of persons experiencing each outcome and the number for which that was the time of final follow-up without the outcome (censoring time).

Multivariable Proportional Hazards Regression Models

The multivariable proportional hazards model that included age, gender, overweight, obesity, hand OA, injury, surgery, LIFT, and SQUAT (Model A) predicted risk of KL   2 significantly better than age, gender, overweight, and obesity (LR test 13.31, 5 df, p = 0.02). Compared with Model A, prediction was not further improved by KOOS Symptoms, KOOS QOL, or QOL items, but was improved by WOMAC Pain (LR test 5.30, 1 df, p = 0.02 vs. Model A) and WOMAC Function (LR test 6.09, 1 df, p = 0.01 vs. Model A). Several lesion variables improved prediction vs. Model A+WOMAC Pain (Table 2) and vs. Model A +WOMAC Function (Supplementary Table 1). The greatest improvement (lowest SBC) was seen for number of subregions with cartilage damage, sum of lesion types, meniscal tear, and for a model including number of subregions with cartilage damage, number of subregions with BMLs, and meniscal tear ("combined model") (Table 2).

Model A predicted risk of KL   2 and joint space narrowing significantly better than age, gender, overweight, and obesity (LR test 12.98, 5 df, p = 0.02). Compared with Model A, prediction was improved by WOMAC Pain (LR test 5.39, 1 df, p = 0.02) or WOMAC Function (LR test 5.80, 1 df, p = 0.02). Several lesion variables improved prediction vs. Model A+WOMAC Pain (Table 3) and vs. Model A+WOMAC Function (Supplementary Table 2). The greatest improvement was seen with number of subregions with cartilage damage, meniscal tear, and the combined model (Table 3). In subsequent OAI x-ray readings, centralized readers changed (in hindsight) their assessment of earlier x-rays in a small number of knees, such that images initially graded KL 0 and selected for MR image assessment in our study were no longer considered KL 0 at baseline. Sensitivity analyses in which we excluded these persons as potentially having the "event" at baseline, i.e., 15 persons for KL   2 outcome and 9 persons for the KL   2 and joint space narrowing outcome (Table 1), revealed similar findings for all analyses.

Calibration and Discrimination

Table 4 shows AUCs for lesion variable sets that had significantly improved prediction based on proportional hazards models. For the KL   2 outcome, AUCs were   0.80 for Model A +WOMAC Pain and: sum of lesion types; number of subregions with cartilage damage; and the combined model. For the secondary outcome, AUCs were   0.80 for Model A+WOMAC Pain and: meniscal tear; number of subregions with cartilage damage; and the combined model. Calibration was good for nearly all models (Table 4). Figure 2 depicts ROC curves

for lesion models including number of subregions with cartilage damage (Figure 2A) and sum of lesion types (Figure 2B).

Discussion

In persons at higher risk for knee OA but KL 0 in both knees, tissue lesion status improved prediction of incident knee OA, defined by KL 2 as well as by KL 2 and joint space narrowing, over up to 7 years of follow-up, vs. age, gender, BMI, hand OA, injury, surgery, specific activities, and knee symptoms or function. For both outcomes, prediction was further improved by including tibiofemoral and patellofemoral cartilage damage, tibiofemoral and patellofemoral BMLs, meniscal tear, meniscal extrusion, sum of lesion types, number of subregions with cartilage damage, number of subregions with BMLs, and for the combined model. For both outcomes, number of subregions with cartilage damage, sum of lesion types, meniscal tear, and the combined model most strongly improved prediction based on proportional hazards models. AUCs 0.80 were found for both outcomes for number of subregions with cartilage damage and the combined model.

Notably, building on pain or function models, cartilage damage characterized in different ways, in both compartments, number of subregions, and maximum severity (both surface area and full-thickness, separately examined) all improved prediction of risk of KL 2. BMLs in both compartments, number of subregions, and maximum severity improved prediction of risk of KL 2. The severity variables did not improve prediction of the secondary outcome, perhaps signifying lower power. Borderline or non-significant findings for *any* cartilage damage and *any* BML may reflect too low a threshold of lesion status. Meniscal tear and extrusion improved prediction of risk of KL 2. Meniscal tear improved prediction of the secondary outcome; extrusion findings were significant for the pain model and borderline for the function model. Overall, these findings provide evidence that these lesions in cartilage, subchondral bone, and menisci represent early stages of OA.

In the Nottingham risk prediction study which had objectives differing from ours, pain was incorporated into the outcome and so was not evaluated as a predictor[14]. In the OAI validation sample, incident disease was defined by diagnosis, as centralized radiographic reading had not as yet occurred[14]. In RS-1 (validation in RS-II and Chingford Study), logistic regression models of knees KL < 2 found AUC improvement (to 0.70-0.84) by adding an indicator for KL = 1[15]. These AUCs cannot be compared to our study, since, having different goals, they included KL 1 knees, and KL 1 emerged as the AUC-improving predictor[15]. Case et al used symptom trajectories to estimate prodrome duration, the divergence point between case and control knees[29]. Their study and that by Hensor et al provided a compelling rationale to include symptoms and function in the base models[29,30].

To our knowledge, the predictive value of tissue lesions for development of knee OA has not been evaluated. Also, previously reported risk models have not been focused on persons at higher risk for knee OA but KL 0 in both knees, a stage at which prevention strategies may be most likely to be successful. A strength of our analyses is use of proportional hazards regression models that incorporate time to first documentation of outcome or time to final

study visit if no outcome, rather than simply dichotomizing outcomes as present or not at end of variable lengths of follow-up.

Our study has limitations. The OAI focuses on persons at higher risk for knee OA, a group of public health importance that will grow with expansion of the aging and overweight populations. However, the significance of tissue lesion status should be separately studied in populations not at higher risk. Our MRI readings did not include effusion/synovitis, which has also been associated with OA development[11]. We had considered using net reclassification improvement or integrated discrimination improvement to further assess improvement in model performance, but the number of outcomes was insufficient for these computationally complex approaches[31]. To date, we have not been able to identify a bilateral KL 0 sample with comparable MRI assessment in which to validate these findings (e.g., due to less sensitive 1.0T MRI acquisition protocol, too few participants KL 0 in both knees at baseline, or no response from the lead investigators); without validation, these estimates may be optimistic. It was not among our objectives to evaluate the clinical utility of MRI in persons at higher risk. The role of MRI in patients at this stage is uncertain, particularly since there are no established treatments for such lesions.

These findings help to advance risk prediction modeling for knee OA; risk scores will help to stratify persons, e.g., for trials and biomarker studies. In addition, the findings support a reorientation in how we approach knee OA. Current health care emphasizes symptom management until TKR. Action is rarely taken at the stage during which it may have greatest impact; these findings offer a window into events in high-risk knees over years before the established clinical definition of OA is reached. They support the existence of a potentially long pre-radiographic period of disease development, during which health-promoting behavior is more likely to be achievable and effective in preventing fully developed OA and its consequences on quality of life vs. customary intervention initiated after OA is established in the knee.

In conclusion, among persons at higher risk for knee OA but KL 0 in both knees, lesions in tissues known to be involved in OA, improved prediction of development of mild and moderate OA disease, defined as incident KL 2 and as KL 2 and joint space narrowing, over up to 7 years of follow-up when added to models including age, gender, BMI, hand OA, injury, surgery, occupational activity, and knee symptoms or function. These findings support that disease onset is likely occurring during the "high-risk" period and support consideration of a reorientation of approach to knee OA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Nelson AE, Allen KD, Golightly YM, Goode AP, Jordan JM. A systematic review of recommendations and guidelines for the management of osteoarthritis: The chronic osteoarthritis management initiative of the U.S. bone and joint initiative. Semin Arthritis Rheum. 2014; 43(6): 701–12. [PubMed: 24387819]

2. Losina E, Burbine SA, Suter LG, et al. Pharmacologic regimens for knee osteoarthritis prevention: can they be cost-effective? Osteoarthritis Cartilage. 2014; 22(3):415–30. [PubMed: 24487044]

3. Javaid MK, Lynch JA, Tolstykh I, et al. Pre-radiographic MRI findings are associated with onset of knee symptoms: the most study. Osteoarthritis Cartilage. 2010; 18(3):323–8. [PubMed: 19919856]

4. Sharma L, Chmiel JS, Almagor O, et al. Significance of preradiographic magnetic resonance imaging lesions in persons at increased risk of knee osteoarthritis. Arthritis Rheumatol. 2014; 66(7): 1811–9. [PubMed: 24974824]

5. Guermazi A, Niu J, Hayashi D, et al. Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). BMJ. 2012; 345:e5339. [PubMed: 22932918]

6. Englund M, Guermazi A, Gale D, et al. Incidental meniscal findings on knee MRI in middle-aged and elderly persons. N Engl J Med. 2008; 359(11):1108–15. [PubMed: 18784100]

7. Davies-Tuck ML, Wluka AE, Wang Y, English DR, Giles GG, Cicuttini F. The natural history of bone marrow lesions in community-based adults with no clinical knee osteoarthritis. Ann Rheum Dis. 2009; 68(6):904–8. [PubMed: 18677011]

8. Dore D, Quinn S, Ding C, et al. Natural history and clinical significance of MRI-detected bone marrow lesions at the knee: a prospective study in community dwelling older adults. Arthritis Res Ther. 2010; 12(6):R223. [PubMed: 21190555]

9. Roemer FW, Guermazi A, Felson DT, et al. Presence of MRI-detected joint effusion and synovitis increases the risk of cartilage loss in knees without osteoarthritis at 30-month follow-up: the MOST study. Ann Rheum Dis. 2011; 70(10):1804–9. [PubMed: 21791448]

10. Sharma L, Nevitt M, Hochberg M, et al. Clinical significance of worsening versus stable preradiographic MRI lesions in a cohort study of persons at higher risk for knee osteoarthritis. Ann Rheum Dis. 2015

11. Roemer FW, Kwoh CK, Hannon MJ, et al. What Comes First? Multitissue Involvement Leading to Radiographic Osteoarthritis: Magnetic Resonance Imaging-Based Trajectory Analysis Over Four Years in the Osteoarthritis Initiative Arthritis Rheumatol. 2015; 67(8):2085–96. [PubMed: 25940308]

12. Niu J, Felson DT, Neogi T, et al. Patterns of Coexisting Lesions Detected on Magnetic Resonance Imaging and Relationship to Incident Knee Osteoarthritis: The Multicenter Osteoarthritis Study. Arthritis Rheumatol. 2015; 67(12):3158–65. [PubMed: 26414125]

13. Felson DT, Niu J, Neogi T, et al. Synovitis and the risk of knee osteoarthritis: the MOST Study. Osteoarthritis Cartilage. 2016; 24(3):458–64. [PubMed: 26432512]

14. Zhang W, McWilliams DF, Ingham SL, et al. Nottingham knee osteoarthritis risk prediction models. Ann Rheum Dis. 2011; 70(9):1599–604. [PubMed: 21613308]

15. Kerkhof HJ, Bierma-Zeinstra SM, Arden NK, et al. Prediction model for knee osteoarthritis incidence, including clinical, genetic and biochemical risk factors. Ann Rheum Dis. 2014; 73(12): 2116–21. [PubMed: 23962456]

16. Nevitt, MC., Felson, D., Lester, G. [accessed May 27 2015] http://oai.epi-ucsf.org/datarelease/docs/StudyDesignProtocol.pdf

17. http://oai.epi-ucsf.org/datarelease/docs/StudyDesignAppendices.pdf. (accessed May 27 2015)

18. http://oai.epi-ucsf.org/datarelease/operationsManuals/MRI_ManualRev.pdf. (accessed May 27 2015)

19. Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage. 2011; 19(8):990–1002. [PubMed: 21645627]

20. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1):159–74. [PubMed: 843571]

21. Felson DT, Nevitt MC. Blinding images to sequence in osteoarthritis: evidence from other diseases. Osteoarthritis Cartilage. 2009; 17(3):281–3. [PubMed: 18977156]

22. Roemer FW, Zhang Y, Niu J, et al. Tibiofemoral joint osteoarthritis: risk factors for MR-depicted fast cartilage loss over a 30-month period in the multicenter osteoarthritis study. Radiology. 2009; 252(3):772–80. [PubMed: 19635831]

23. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? Med Care. 1996; 34(1):73–84. [PubMed: 8551813]

24. 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(12):1833–40. [PubMed: 3068365]

25. 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 Orthop Sports Phys Ther. 1998; 28(2):88–96. [PubMed: 9699158]

26. [accessed May 27 2015] http://oai.epi-ucsf.org/datarelease/operationsManuals/RadiographicManual.pdf

27. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957; 16(4):494–502. [PubMed: 13498604]

28. [accessed May 27 2015] http://oai.epi-ucsf.org/datarelease/forms/kXR_SQ_BU_Descrip.pdf?V01XRKL

29. Case R, Thomas E, Clarke E, Peat G. Prodromal symptoms in knee osteoarthritis: a nested case-control study using data from the Osteoarthritis Initiative. Osteoarthritis Cartilage. 2015; 23(7):1083–9. [PubMed: 25843364]

30. Hensor EM, Dube B, Kingsbury SR, Tennant A, Conaghan PG. Toward a clinical definition of early osteoarthritis: onset of patient-reported knee pain begins on stairs. Data from the osteoarthritis initiative. Arthritis Care Res (Hoboken). 2015; 67(1):40–7. [PubMed: 25074673]

31. 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(2):157–72. discussion 207-12. [PubMed: 17569110]

4146 persons with centralized reading of knee x-rays of both knees at 12-month visit

↓

1114 persons identified by coordinating center as KL 0 in right and left knees at 12-month visit

↓

861 persons who underwent MRIs at 12- and 48-month follow-up

↓ → *12 persons with inadequate technical quality of MR images*

849 persons

↓ → *8 missing covariate data*

841 knees (from 841 persons)
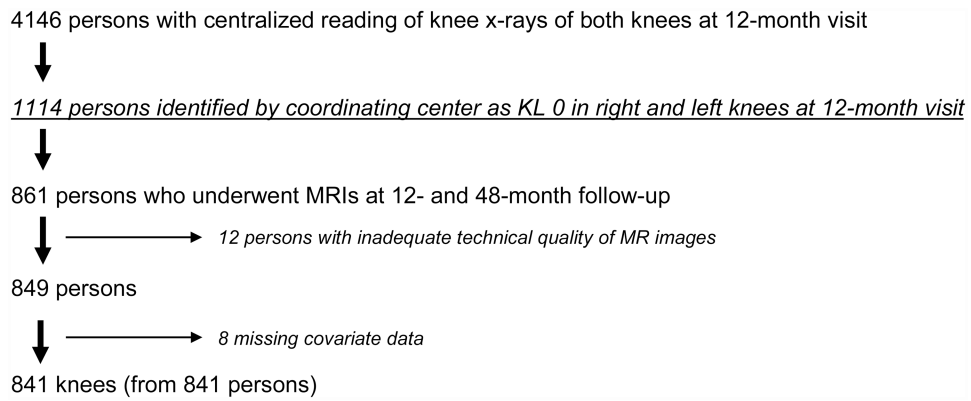
Figure 1. Derivation of Analysis Sample

Among 1114 eligible persons, 77 missed the 48-month evaluation [withdrew (28), difficulty scheduling (32), died (14), health problems (2), caregiving responsibilities (1)], 176 attended without MRI, and 12 had inadequate images. We assessed knee MR images (1 knee/person) in the remaining 849 individuals, of whom 8 were excluded for missing covariable data. Although this study utilized the 12-month MRI data, MRIs at 12- and 48-month follow-up were required. While MRI acquisition was included in the core funding of the OAI, MRI readings were not. We successfully applied for funding for an ancillary study that had additional goals which required the 48-month reading. Reading data are therefore not available in those missing the 48-month MRI. In any case, including these participants would only have added 8 additional cases of incident KL   2.
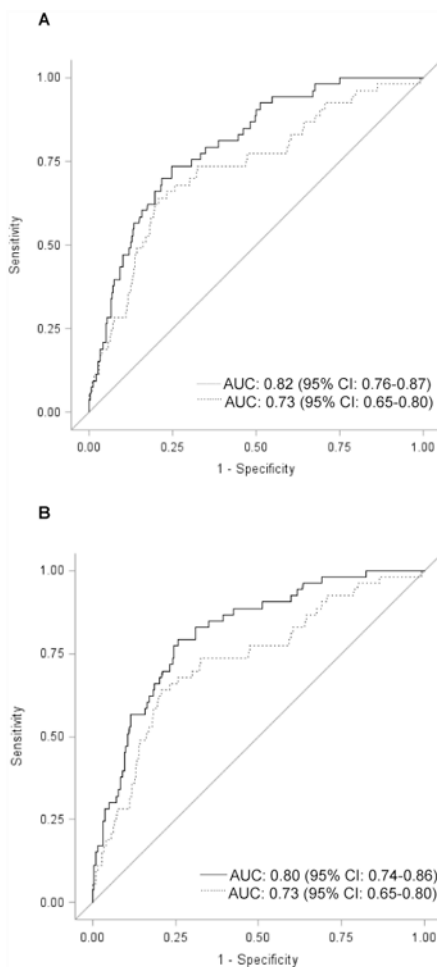
Figure 2. Receiver Operating Characteristic Curves for Models to Predict Radiographic Knee OA Defined as KL ≥ 2

The figure shows receiver operating characteristic (ROC) curves for two prediction models (to predict KL ≥ 2). In each figure, the dotted line represents the model including age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT, and WOMAC Pain. The solid line in Figure 2A represents the model including age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT, WOMAC Pain + number of subregions with cartilage damage. The solid line in Figure 2B represents the model including age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT, WOMAC Pain + sum of lesion types. (n=841 persons, 1 knee/person) (AUC=area under the curve; CI=confidence interval)

Table 1

Frequency of Outcome (KL ≥ 2) and Censoring at Each Timepoint

Frequency of outcome (KL ≥ 2) and censoring at each timepoint							
Status	Baseline	12 month	24 month	36 month	48 month	72 month	96 month
At risk	841	826	825	818	809	677	649
KL ≥ 2 at this visit	15	1	5	4	4	14	10
KL < 2, time of last follow-up	0	0	2	5	128	14	639

Frequency of outcome (KL ≥ 2 and joint space narrowing) and censoring at each timepoint							
Status	Baseline	12 month	24 month	36 month	48 month	72 month	96 month
At risk	841	832	832	827	819	681	659
KL ≥ 2 and JSN at this visit	9	0	3	3	4	8	9
KL < 2 or no JSN, time of last follow-up	0	0	2	5	134	14	650

The rows labelled "at risk" represent the number of persons considered at risk for this outcome at the time of the designated visit. The second row in each section represents frequency of persons with the outcome based on x-rays acquired at the designated timepoint (and not at any prior timepoint). The last row shows the frequency of persons who had their final follow-up at the designated timepoint and did not experience the outcome at this or prior timepoints (censoring time). Of note, in this study, we undertook MRI readings in persons KL 0 in both knees at baseline by centralized reading. However, in subsequent centralized x-ray readings, some baseline KL = 0 readings were revised because they were felt in hindsight to show evidence of KL ≥ 2 (n = 15) or KL ≥ 2 and JSN (n = 9); this explains the knees in column 1 with the "event" at baseline. Findings for all analyses excluding these knees were similar.

Table 2

Tissue Lesion Contributions to Prediction of Knee OA, <u>Defined as KL ≥ 2</u>, by Model A plus WOMAC Pain, Cox Proportional Hazards Models using the Likelihood Ratio (LR) Test and Schwarz Bayesian Information Criterion (SBC) (n=841 persons, 1 knee/person)

Lesion status	Number (%) with outcome 53/841 (6.3%)	K/L ≥ 2 (outcome)		
		Adjusted HR (95% CI)	LR test (p-value)	SBC for base model[1] + lesion
Any cartilage damage (TF or PF)	48/634 (7.6)	2.51 (0.99, 6.38)	4.76 (0.29)	702.77
Cartilage damage, both TF and PF	29/262 (11.1)	*2.68 (1.54, 4.68)*	*11.96 (0.0005)*	695.57
Any BML (TF or PF)	40/508 (7.9)	1.70 (0.90, 3.21)	2.87 (0.90)	704.66
BML, both TF and PF	20/145 (13.8)	*2.77 (1.58, 4.86)*	*11.37 (0.0007)*	696.16
Meniscal tear	26/177 (14.7)	*4.31 (2.43, 7.64)*	*23.56 (<0.0001)*	683.96
Meniscal extrusion	17/115 (14.8)	*2.94 (1.61, 5.39)*	*10.66 (0.001)*	696.86
Sum of lesion types[2] (adjusted HR/additional lesion type)	—	*1.97 (1.49, 2.59)*	*23.78 (<0.0001)*	683.75
Number of subregions with cartilage damage (including TF and PF)[2] (adjusted HR/additional subregion)	—	*1.56 (1.36, 1.79)*	*32.30 (<0.0001)*	675.23
Maximum cartilage damage severity (surface area, among all TF and PFsubregions)[2] (adjusted HR/additional grade)	—	*1.87 (1.23, 2.85)*	*9.98 (0.002)*	697.55
Maximum cartilage damage severity (full thickness score, among all TF and PFsubregions)[2] (adjusted HR/additional grade)	—	*1.45 (1.11, 1.90)*	*7.12 (0.008)*	700.40
Number of subregions with BML(including TF and PF)[2] (adjusted HR/additional subregion)	—	*1.45 (1.20, 1.76)*	*12.94 (0.0003)*	694.58
Maximum BML severity (among all TFand PF subregions)[2] (adjusted HR/additional grade)	—	*1.45 (1.04, 2.04)*	*4.42 (0.04)*	703.10
Number of subregions with cartilage damage, number of subregions with BML, meniscal tear (concurrently included in model)	—	—	*47.05 (<0.0001)*	668.42

[1]Base model included age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT, and WOMAC Pain (SBC for base model = 703.56)

[2]continuous

Each row represents a separate model showing the impact (LR test) on the adjusted HR for incident OA (defined as KL ≥ 2) of adding the specific lesion(s) to the base model with age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT, and WOMAC Pain. Column 2 represents % of knees with the outcome among all knees with the specific (row) lesion. Obesity, LIFT, and WOMAC Pain were associated with the outcome in all or nearly all Model A+WOMAC Pain+lesion models. (HR=hazard ratio; CI=confidence interval; TF=tibiofemoral; PF=patellofemoral)

Table 3

Tissue Lesion Contributions to Prediction of Knee OA, <u>Defined as KL ≥ 2 and Joint Space Narrowing</u>, by Model A plus WOMAC Pain, Cox Proportional Hazards Models using the Likelihood Ratio (LR) Test and Schwarz Bayesian Information Criterion (SBC) (n = 841 persons, 1 knee/person)

Lesion status	Number (%) with outcome 36/841 (4.3%)	KL ≥ 2 and joint space narrowing (outcome)		
		Adjusted HR (95% CI)	LR test (p-value)	SBC for base model[1] + lesion
Any cartilage damage (TF or PF)	31/634 (4.9)	1.65 (0.63, 4.30)	1.15 (0.28)	482.26
Cartilage damage, both TF and PF	19/262 (7.3)	*2.41 (1.23, 4.74)*	*6.44 (0.01)*	480.97
Any BML (TF or PF)	25/508 (4.9)	1.22 (0.59, 2.52)	0.29 (0.59)	487.12
BML, both TF and PF	15/145 (10.3)	*3.28 (1.68, 6.42)*	*10.82 (0.001)*	476.60
Meniscal tear	19/177 (10.7)	*4.94 (2.50, 9.78)*	*19.96 (<0.0001)*	467.45
Meniscal extrusion	10/115 (8.7)	*2.45 (1.14, 5.26)*	*4.62 (0.03)*	482.79
Sum of lesion types[2] (adjusted HR/additional lesion type)	—	*1.77 (1.27, 2.48)*	*11.60 (0.0007)*	475.81
Number of subregions with cartilage damage (including TF and PF)[2] (adjusted HR/additional subregion)	—	*1.57 (1.33, 1.87)*	*22.81 (<0.0001)*	464.60
Maximum cartilage damage severity (among all TF and PF subregions)[2] (adjusted HR/additional grade)	—	1.53 (0.95, 2.46)	3.43 (0.06)	483.98
Maximum cartilage damage severity (full thickness score, among all TF and PFsubregions)[2] (adjusted HR/additional grade)	—	1.32 (0.95, 1.84)	2.61 (0.11)	484.80
Number of subregions with BML(including TF and PF)[2] (adjusted HR/additional subregion)	—	*1.47 (1.17, 1.85)*	*9.81 (0.002)*	477.60
Maximum BML severity (among all TFand PF subregions)[2] (adjusted HR/additional grade)	—	1.28 (0.84, 1.94)	1.27 (0.26)	486.14
Number of subregions with cartilage damage, number of subregions with BML, meniscal tear (concurrently included in model)	—	—	*36.47 (<0.0001)*	458.11

[1] Base model included age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT, and WOMAC Pain (SBC for base model=483.83)

[2] continuous

Each row represents a separate model showing the impact (LR test) on the adjusted HR for incident OA (KL ≥ 2 and joint space narrowing) of adding the specific lesion(s) to the base model with age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT, and WOMAC Pain. Column 2 represents % of knees with the outcome among all knees with the specific (row) lesion. Obesity, hand OA, SQUAT, and WOMAC Pain were associated with the outcome in all or nearly all Model A+WOMAC Pain+lesion models. (HR=hazard ratio, CI=confidence interval; TF=tibiofemoral; PF=patellofemoral)

Table 4
Risk Prediction Models: Discrimination and Calibration

KL 2 outcome		
Predictors	**Discrimination: AUC (95% CI)**	**Calibration: Hosmer-Lemeshow p-value**
Age, gender, BMI	0.66 (0.58, 0.74)	0.66
Model A (age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT)	0.71 (0.63, 0.79)	0.80
Model A + WOMAC Pain	0.73 (0.65, 0.80)	0.13
Model A, WOMAC Pain + Cartilage damage, both TF and PF	0.77 (0.71, 0.84)	0.16
Model A, WOMAC Pain + BML, both TF and PF	0.76 (0.69, 0.83)	0.16
Model A, WOMAC Pain + Meniscal tear	0.79 (0.73, 0.86)	0.63
Model A, WOMAC Pain + Meniscal extrusion	0.77 (0.71, 0.83)	0.14
Model A, WOMAC Pain + Sum of lesion types	0.80 (0.74, 0.86)	0.79
Model A, WOMAC Pain + Number of subregions with cartilage damage (TF and PF)	0.82 (0.76, 0.87)	0.35
Model A, WOMAC Pain + Maximum cartilage damage (surface area, TF and PF)	0.77 (0.71, 0.83)	0.69
Model A, WOMAC Pain + Maximum cartilage damage (full thickness, TF and PF)	0.76 (0.70, 0.83)	0.58
Model A, WOMAC Pain + Number of subregions with BML (TF and PF)	0.76 (0.70, 0.83)	0.59
Model A, WOMAC Pain + Maximum BML severity (TF and PF)	0.74 (0.67, 0.81)	0.95
Model A, WOMAC Pain + Number of subregions with cartilage damage, number of subregions with BML, meniscal tear concurrently included in model	0.84 (0.78, 0.89)	0.95
KL 2 and joint space narrowing outcome		
Predictors	**Discrimination: AUC (95% CI)**	**Calibration: Hosmer-Lemeshow p-value**
Age, gender, BMI	0.64 (0.54, 0.74)	0.58
Model A (age, gender, overweight, obesity, hand OA, knee injury, knee surgery, LIFT, SQUAT)	0.73 (0.65, 0.81)	0.79
Model A + WOMAC Pain	0.74 (0.65, 0.82)	0.50
Model A, WOMAC Pain + Cartilage damage, both TF and PF	0.77 (0.69, 0.84)	0.81
Model A, WOMAC Pain + BML, both TF and PF	0.78 (0.70, 0.85)	0.74
Model A, WOMAC Pain + Meniscal tear	0.80 (0.73, 0.87)	0.60
Model A, WOMAC Pain + Meniscal extrusion	0.76 (0.69, 0.83)	0.11
Model A, WOMAC Pain + Sum of lesion types	0.78 (0.72, 0.85)	0.04
Model A, WOMAC Pain + Number of subregions with cartilage damage (TF and PF)	0.81 (0.73, 0.88)	0.09
Model A, WOMAC Pain + Number of subregions with BML (TF and PF)	0.77 (0.69, 0.85)	0.68
Model A, WOMAC Pain + Number of subregions with cartilage damage, number of subregions with BML, meniscal tear concurrently included in model	0.83 (0.75, 0.90)	0.70

(AUC = area under a receiver operating characteristic curve; CI=confidence interval)