Title
Tool for osteoarthritis risk prediction (TOARP) over 8 years using baseline clinical data, X-ray, and MRI: Data from the osteoarthritis initiative.

Permalink
https://escholarship.org/uc/item/54r1h966

Authors
Joseph, GB
McCulloch, CE
Nevitt, MC
et al.

Publication Date
2017-11-16

DOI
10.1002/jmri.25892

Peer reviewed
Tool for Osteoarthritis Risk Prediction (TOARP) Over 8 Years Using Baseline Clinical Data, X-ray, and MRI: Data From the Osteoarthritis Initiative

Gabby B. Joseph, PhD,1* Charles E. McCulloch, PhD,2 Michael C. Nevitt, PhD,2 Jan Neumann, MD,1 Alexandra S. Gersing, MD,1 Martin Kretzschmar, MD,1 Benedikt J. Schwaiger, MD,1 John A. Lynch, PhD,2 Ursula Heilmeier, MD,1 Nancy E. Lane, MD,3 and Thomas M. Link, MD, PhD1

Background: Osteoarthritis (OA), a multifactorial disease causing joint degeneration, often leads to severe disability. The rising rates of disability highlight the need for implementing preventative measures at early stages of the disease, which would especially benefit subjects at high risk for OA development.

Purpose: To develop a risk prediction tool for moderate-severe OA (TOARP) over 8 years based on subject characteristics, knee radiographs, and MRI data at baseline using data from the Osteoarthritis Initiative (OAI).

Study Type: Retrospective.

Subjects: 641 subjects with no/mild radiographic OA (Kellgren–Lawrence [KL] 0–2) and no clinically significant symptoms (Western Ontario and McMaster Universities Arthritis Index [WOMAC] 0–1) were selected from the OAI.

Field Strength/Sequence: MR images were obtained using 3.0T.

Assessment: Compartment-specific cartilage and meniscus morphology and cartilage T2 were assessed. Baseline subject demographics, risk factors, KL score, cartilage WORMS score, presence of meniscus tear, and cartilage T2 were used to predict the development of moderate/severe OA (KL = 3–4 or WOMAC pain ≥5 or total knee replacement [TKR]) over 8 years.

Statistical Tests: Best subsets variable selection followed by cross-validation were used to assess which combinations of variables best predict moderate/severe OA.

Results: Model 1 included KL score, previous knee injury in the last 12 months, age, gender, and BMI. Model 2 included all variables in Model 1 plus presence of cartilage defects in the lateral femur and patella, and presence of a meniscal tear. Model 3 included all variables in Models 1 and 2, plus cartilage T2 in the medial tibia and medial femur. Compared to Model 1 (cross-validated AUC = 0.67), Model 3 performed significantly better (AUC = 0.72, P = 0.04), while Model 2 showed a statistical trend (AUC = 0.71, P = 0.08).

Data Conclusion: We established a risk calculator for the development of moderate/severe knee OA over 8 years that includes radiographic and MRI data. The inclusion of MRI-based morphological abnormalities and cartilage T2 significantly improved model performance.

Level of Evidence: 2
Technical Efficacy: Stage 3

Osteoarthritis (OA), a multifactorial disease that causes joint degeneration, affects 14 million U.S. adults, and often leads to severe disability1 and total knee replacement (TKR). The rising rates of TKR2 and secondary revision surgeries3 highlight the need for implementing preventative measures, such as lifestyle modifications, at early stages of OA.

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.25892
Received Aug 8, 2017, Accepted for publication Oct 24, 2017.

*Address reprint requests to: G.B.J., Department of Radiology and Biomedical Imaging, University of California, San Francisco, 185 Berry St., Ste. 350, San Francisco, CA 94158. E-mail: gabby.joseph@ucsf.edu

From the 1Department of Radiology and Biomedical Imaging, University of California, San Francisco, California, USA; 2Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA; and 3Department of Rheumatology, University of California, Davis, California, USA

Additional supporting information may be found in the online version of this article.

© 2017 International Society for Magnetic Resonance in Medicine
the disease. Such preventative measures would be particularly beneficial for subjects at high risk for OA development.

In addition to the known clinical risk factors for OA (including obesity and previous injury\(^9\)), imaging of the knee joint may assist in the identification of subjects at high risk. Kellgren–Lawrence (KL) knee radiographic scores are positively associated with knee pain,\(^6\) and varus knee alignment at baseline is associated with a 4-fold increase in medial OA progression.\(^6\) In addition, cartilage damage and meniscal tears as assessed with magnetic resonance imaging (MRI) have been shown to be associated with incident radiographic knee OA 2–4 years later.\(^7\) More recently, MRI T\(_2\) mapping, which identifies biochemical changes in cartilage including abnormalities of collagen fiber orientation,\(^8\) has been shown to predict radiographic and symptomatic knee OA.\(^9\) MRI T\(_2\) measures early degenerative changes in knee cartilage that occur prior to macroscopic cartilage defects and thinning. Thus, a composite model consisting of clinical risk factors and imaging data may help identify subjects at high risk for OA.

Clinical and imaging prediction tools are used by clinicians to identify patients who may benefit from an intervention, either medical or surgical, to prevent an outcome. One such model, the WHo fracture risk assessment tool (FRAX), is used to identify subjects at risk of hip and major osteoporotic fractures using clinical risk factors (including age, sex, body mass index [BMI], prior fracture, and parental history of hip fracture) and a bone density measurement.\(^10\) Similarly, this study aimed to develop a risk prediction tool for moderate/severe OA. The purpose of this study was to develop a Tool for Osteoarthritis Risk Prediction (TOARP) over 8 years based on subject characteristics, knee radiographs, and MRI data at baseline using data from the Osteoarthritis Initiative (OAI).

### Materials and Methods

#### Subject Selection

This study utilized data from the OAI (http://www.oai.ucsf.edu/),\(^11\) a multicenter, longitudinal study of persons aged 45–79 years at enrollment, aimed at assessing biomarkers in OA including those derived from MRI. The OAI dataset includes both MRI and radiographic images of subjects scanned over 8 years. This database can be used to evaluate MRI biomarkers for the development and progression of OA. The study protocol, amendments, and informed consent documentation were reviewed and approved by the local Institutional Review Boards of all participating centers.

Participants for the present study were selected from the OAI, which excluded individuals with inflammatory arthropathies (ie, rheumatoid arthritis), MRI contraindications, positive pregnancy test, bilateral total knee joint replacement, and comorbid conditions that may affect the ability to participate in the study. Specific inclusion criteria for the present study were a baseline radiographic KL score ≤2 in the right knee and no symptoms in the right knee (Western Ontario and McMaster Universities Arthritis Index [WOMAC] of 0/1). A WOMAC pain score threshold of ≤1 was chosen based on a previous study\(^12\) that reported a minimal perceptible change in pain required a 10% difference (using the VAS scale), which equates to a change greater than 2 utilizing the WOMAC Likert scale. The sample of knees was selected from MRI scans that had both T\(_2\) and WORMS score assessments in the right knee.\(^13–17\) and also had complete data for known knee OA risk factors including family history of knee replacement and previous injury. The prior studies analyzing T\(_2\)/WORMS had vastly different goals from the current study. There are no selection biases since the distribution of subject characteristics in the current study is similar to that of the OAI database. Exclusion criteria included baseline knee deformity of the knee joint, total joint replacements in the right knee, MRI evidence of subchondral or stress fractures of the knee, or abnormalities that did not fit into the spectrum of OA and indicated other severe disease, such as tumor or inflammation. Based on our inclusion and exclusion criteria, right knees from 641 participants were selected and analyzed (Fig. 1).

#### MRI

MR images were obtained using four identical 3.0T (Siemens Magnetom Trio, Erlangen, Germany) scanners in Columbus, Ohio; Baltimore, Maryland; Pittsburgh, Pennsylvania; Pawtucket, Rhode Island. The following sequences were acquired: sagittal 2D intermediate-weighted fast spin-echo sequence (TR/TE = 3200/30 msec, spatial resolution = 0.357 × 0.511 mm, slice thickness = 3.0 mm), coronal 2D proton density fast spin-echo sequence (TR/TE = 3700/29 msec, spatial resolution = 0.365 × 0.456 mm, slice thickness = 3.0 mm), and sagittal 3D dual-echo in steady state sequence (TR/TE = 16.3/4.7 msec, spatial resolution = 0.365 × 0.456 mm, slice thickness = 0.7 mm). A sagittal 2D multislice multiecho sequence (MSME; TR = 2700 msec, TE\(_1\)–TE\(_7\) = 10–70 msec, spatial resolution = 0.313 × 0.446 mm, slice...
thickness = 3.0 mm, and 0.5 mm gap) was used for cartilage T2 measurements.\textsuperscript{18}

**Image Analysis**

**RADIOGRAPHY-BASED KL GRADE AND KNEE ALIGNMENT.** Baseline and annual radiographic KL grades\textsuperscript{19} over 8 years were provided in the OAI dataset. Subjects with baseline KL grades of 0–2 were selected, and worsening was defined as developing a KL grade of 3–4 over 8 years. Baseline knee alignment (femur-tibia angle) was measured based on a method developed by Iranpour-Boroujen et al.\textsuperscript{20} This method had high reproducibility with an intraclass correlation coefficient (ICC) of intra- and interreader reproducibility of 0.96 and 0.98, respectively.\textsuperscript{20}

**WORMS SCORING.** MR images of the right knee obtained at the baseline visit were reviewed on picture archiving communication system (PACS) workstations (Agfa, Ridgefield Park, NJ). Three radiologists with 7, 5, and 5 years of experience graded cartilage lesions. In equivocal cases, a consensus reading was performed with a musculoskeletal radiologist with 23 years of experience. Baseline cartilage lesions were assessed in five regions (patella, medial femur, medial tibia, lateral femur, and lateral tibia) using a modified semiquantitative whole-organ MRI score (WORMS).\textsuperscript{21} The highest score of any lesion was recorded for each region. For calibration purposes, 20 cases were read simultaneously by the four readers in consensus. A binary variable representing a cartilage defect was defined as positive if WORMS grade was ≥2 in a) each region individually (patella, medial femur, medial tibia, lateral femur, and lateral tibia) and b) in any region overall. Meniscal lesions were graded separately in six regions (medial/lateral and anterior/body/posterior) using the following 5-point scale: 0: normal; 1: intrasubstance signal; 2: nondisplaced tear; 3: displaced or complex tear; 4: complete destruction/maceration. A binary variable representing a meniscus tear was defined as positive if WORMS grade was ≥2 in any meniscus region.

**T2 MEASUREMENTS.** All baseline images were analyzed using a Sun Workstation (Sun Microsystems, Palo Alto, CA). Semiautomatic cartilage segmentation of lateral/medial femur, lateral/medial tibia, and patella regions was performed as previously described, using an in-house, spline-based software based on MatLab (MathWorks, Natick, MA).\textsuperscript{22} Our segmentations covered cartilage slices that did not contain partial volume effects; we also excluded sections with any compromised image quality such as artifacts.

Validated methods for obtaining a T2\textsubscript{MEASUREMENTS} map of the cartilage have been published.\textsuperscript{22,23} T2 maps were computed from the MSME images on a pixel-by-pixel basis using six echoes (TE = 20–70 msec) and three parameter fittings accounting for noise,\textsuperscript{24,25} and averaged over all of the slices in each cartilage region, accounting for the number of pixels in each slice. The first echo (TE = 10 msec) was not included in the T2 fitting procedure in order to reduce potential errors resulting from stimulated echoes, and a noise-corrected algorithm was implemented.\textsuperscript{24,25} The cartilage T2\textsubscript{REPRODUCIBILITY} results have been described previously.\textsuperscript{22,23} The intrareader reproducibility of mean T2 was determined by segmenting the cartilage in 15 subjects, three times by one operator. The interreader reproducibility was determined by segmenting five subjects, three times each by two operators. The mean T2 values had root mean square (RMS) coefficients of variation (CV) ranging from 0.83% in the medial femur to 3.21% in the patella for intrareader reproducibility, and from 1.22% in the patella to 1.86% in the lateral tibia for interreader reproducibility.

**Statistical Analysis**

Statistical analysis was performed using STATA v. 14 software (StataCorp, College Station, TX). The outcome variable was development of moderate to severe radiographic or symptomatic knee OA, defined by any of the following during 8 years of follow-up: worsening to KL grade 3–4 OA, a WOMAC pain score of ≥5 at any two follow-up timepoints, or an incident total knee replacement (TKR) in the right knee, confirmed by medical records and/or knee radiographs. We included KL score as an outcome for this study, as radiographic OA is considered a standard outcome definition for OA.\textsuperscript{26} The predictors for this study were: subject characteristics (age, gender, and BMI), locked to remain in the model, baseline risk factors, KL score, knee alignment, presence of a cartilage defect, presence of a meniscus tear, and cartilage T2\textsubscript{REPRODUCIBILITY} as listed in Fig. 2. Logistic regression was used for analysis to be consistent with the published literature on OA. A STATA algorithm\textsuperscript{27} that utilized “leaps-and-bounds”\textsuperscript{28} was used to perform best subsets variable selection with logistic regression to assess which combinations of the above-listed variables best predicted moderate to severe radiographic OA (outcome). Akaike’s information criteria (AIC) for each combination of predictors (the lowest being the most desirable) were assessed to compare each model’s goodness of fit relative to one another. Next, 10-fold crossvalidation was performed for the models with the lowest AICs (within 10),\textsuperscript{29} and the Discrimination Index (DI, defined as the average of the predicted probability of an event among the individuals with an event minus the average of the predicted probability of an event among individuals without an event) was quantified.\textsuperscript{30} We did not use a training and testing dataset in the study due to the small sample size with low outcome percentage (12.48%). Thus, we chose 10-fold crossvalidation.\textsuperscript{31}

First, the analysis was performed for a base model, which included clinical data only (age, gender, and BMI). Next, the analyses were performed three times independently to develop three models: Model 1 being the least sophisticated and included the base model and radiography, Model 2 being more sophisticated and included the base model, radiography, and MRI WORMS scoring and Model 3 being the most sophisticated and included the base model, radiography, MRI WORMS scoring, and MRI T2\textsubscript{REPRODUCIBILITY} (Fig. 2). Ten-fold cross-validated receiver operating characteristic (ROC) analyses were used to obtain an unbiased assessment of each model; then cross-validated area under the curves (AUCs) were compared using the test of DeLong et al.\textsuperscript{32}

**Results**

**Subject Characteristics**

The 641 participants in this study had a mean age of 56.4 ± 7.5 years and a mean BMI of 27.0 ± 4.3 kg/m\textsuperscript{2} at baseline. The distribution of KL grades and other participant characteristics are listed in Table 1. Eighty subjects
had a positive outcome (either an incident TKR \[n = 8, 1.25\%\], worsening to KL 3 or 4 \[n = 34, 5.31\%\], or progression to a WOMAC pain score of \(\geq 5\) \[n = 53, 8.27\%\]). In all, 381 subjects (59.44\%) had cartilage defects in any cartilage region, while 190 (29.64\%) had a meniscus tear in any meniscus region.

### OA Risk Models

The following models best predicted the development of knee OA over 8 years (Table 2): The base model included age, gender, and BMI. Model 1 added the radiography-based KL score, knee alignment, and previous knee injury in the last 12 months (cross-validated DI = 0.048). Model 2 included all variables in Model 1 plus presence of cartilage defects in the lateral femur and patella, and presence of a meniscal tear (cross-validated DI = 0.084). Model 3 included all variables in Models 1 and 2, plus mean cartilage T2 in the medial tibia and medial femur (cross-validated DI = 0.11). Compared to Model 1 with a cross-validated area under the ROC curve (AUC) = 0.67, Model 3 performed significantly better (AUC = 0.72, \(P = 0.04\)), while Model 2 showed a statistical trend (AUC = 0.71, \(P = 0.08\)). However, there was no difference in performance between Models 2 and 3. All models had significantly (\(P < 0.05\)) greater AUCs compared to a base model consisting of age, gender, and BMI (Fig. 3), demonstrating the added value of imaging for risk prediction. These results demonstrate that including both cartilage T2 and WORMS significantly improves model performance compared to a model with risk factors and KL-score alone. The AUC results were similar after a sensitivity analysis excluding African American subjects \(n = 56, 8\%,\ AUC\ for\ Model\ 3 = 0.74\). Figure 4 illustrates the improvement of model risk stratification from Model 1 to Model 3, especially in subjects who develop OA over 8 years.

### Tool for Individualized OA Risk Prediction (TOARP)

Figure 5a illustrates a risk calculator graphic designed for use in the clinic; Fig. 5b illustrates the isolated effects of low, medium, and high medial femur cartilage T2 on OA

---

**TABLE 2: Subject Demographics and Risk Factors**

<table>
<thead>
<tr>
<th>Subject Demographics</th>
<th>Model 1: Risk Factors + X-ray</th>
<th>Model 2: Risk Factors + X-ray + MRI WORMS</th>
<th>Model 3: Risk Factors + X-ray + MRI WORMS + MRI cartilage T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, gender, BMI</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Risk Factors:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee Injury ever</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Knee Injury in last 12 months</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Fam. History of TKR</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>X-ray:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KL Score</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Alignment Angle</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MRI WORMS:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cartilage defect: MF, MT, LF, LT, PAT, any</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Meniscus Tear: Med Ant, Body, Post; Lat Ant, Body, Post; Any</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>MRI Cartilage T2:</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**FIGURE 2: Illustration of the development and validation of three risk prediction models.**
risk probability, while keeping the other subject characteristics (risk factors, KL score, and WORMS scores) constant. For example, a 69-year-old female with a BMI of 25.8 kg/m², a previous injury, KL score of 2, lateral femur cartilage defect, medial meniscus tear, and a medial femur cartilage T2 of 43 msec (~98th percentile based on a reference database of subjects without cartilage degeneration) would have ~75% risk for progression of OA development, while a female with the same characteristics and a medial femur cartilage T2 of 31 msec (~2nd percentile) would have a risk of ~34%.

### Discussion

This study created a composite subject-specific risk assessment model for OA development over 8 years that includes clinical and advanced MRI data. The three models established in this study range from least sophisticated (including subject characteristics, risk factors, and radiography) to most sophisticated with WORMS and cartilage T2. Compared to the least sophisticated model, the addition of MRI parameters increased the AUC values, demonstrating significance for the model with WORMS and T2 and a statistical trend for WORMS alone. Overall, the three models provide versatility for OA risk prediction and they could be used by clinicians to provide individualized assessments to patients, and could motivate lifestyle changes to lower risk of OA progression.

While previous studies have developed risk prediction models for OA, our study is different, as it includes MRI and assessment of known risk factors for OA development over 8 years in subjects without, or with only mild, radiographic OA and no symptoms of OA at baseline. A variety of OA risk calculators have been developed that

### TABLE 1. Participant Characteristics

<table>
<thead>
<tr>
<th>All participants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>n</strong></td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
</tr>
<tr>
<td><strong>BMI (kg/m²)</strong></td>
</tr>
<tr>
<td><strong>Gender (male)</strong></td>
</tr>
<tr>
<td><em><em>WOMAC</em> pain</em>*</td>
</tr>
<tr>
<td><strong>Family history of knee replacement</strong></td>
</tr>
<tr>
<td><strong>Previous injury anytime</strong></td>
</tr>
<tr>
<td><strong>Previous injury in the last 12 months</strong></td>
</tr>
<tr>
<td><strong>KL</strong></td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

*WOMAC: Western Ontario and McMaster Universities.

### TABLE 2. Model Characteristics for Model 1 (Risk Factors + Radiography), Model 2 (Model 1 + WORMS), and Model 3 (Model 2 + Cartilage T2)

<table>
<thead>
<tr>
<th>Model</th>
<th>Variables included</th>
<th>Cross-validated AIC</th>
<th>Cross-validated DI</th>
<th>Cross-validated AUC (P value compared to Model 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>Age + gender + BMI</td>
<td>475</td>
<td>0.014</td>
<td>0.60 [0.53–0.67] (P = 0.02)</td>
</tr>
<tr>
<td>1</td>
<td>Base + KL grade + Previous knee injury in the last 12 months</td>
<td>457</td>
<td>0.048</td>
<td>0.67 [0.61–0.73] (reference)</td>
</tr>
<tr>
<td>2</td>
<td>Model 1 + Lateral femur cartilage lesion + Patella cartilage lesion + Meniscal tear</td>
<td>442</td>
<td>0.084</td>
<td>0.71 [0.65–0.77] (P = 0.08)</td>
</tr>
<tr>
<td>3</td>
<td>Model 2 + Medial femur T2 + Medial tibia T2 +</td>
<td>437</td>
<td>0.11</td>
<td>0.72 [0.66–0.78] (P = 0.04)</td>
</tr>
</tbody>
</table>

The values for the Akaike information criterion (AIC), Discrimination Index (DI), and Area under the ROC curve (AUC) were obtained using 10-fold crossvalidation. The AUCs for all models are compared to Model 1 to understand the importance of MR imaging in for prediction of OA development. The results demonstrate that the addition of WORMS and cartilage T2 improves classification over radiography alone.
range in complexity: Some included only subject demographics, clinical factors, and risk factors without imaging, while others integrate biochemical markers and radiography-based KL scores in their OA prediction model. Kerkhoff et al reported similar accuracy, when including clinical variables in addition to genetic scores and biochemical markers (AUC $\sim 0.66$, outcomes spanning 4 to 10 years); however, the inclusion of baseline radiographic KL score increased the AUC to 0.79, demonstrating the importance of imaging for risk prediction. The different AUCs Kerkhoff et al’s study compared to the current study may be due to the differences in the subject inclusion criteria (Kerkhoff et al included subjects with KL 0-1, while we included KL 0-2) and outcome definitions (Kerkhoff et al used incident OA defined by KL $\geq 2$, while our outcome was composite). In addition to KL score, we further investigated the role of advanced MRI morphology and T2 values (indicative of cartilage extracellular matrix [ECM] composition) for OA prediction, and demonstrated that the addition of these advanced imaging techniques improves model discrimination.

Two key features of the models developed in this study are 1) individualized assessments and 2) inclusion of advanced MRI. Individualized risk profiles are essential for developing personalized prevention strategies for OA, and for motivating subjects to adhere to recommendations. In addition to subject characteristics, individualized assessments that incorporate advanced MRI allow clinicians to consider joint morphology and cartilage biochemical composition when formulating their treatment plans. While the model with radiography findings alone provided fair prediction of OA risk probability, cartilage T2 relaxation time measurements improved prediction. Studies have shown that cartilage T2 can predict morphologic OA development (with outcomes of radiographic OA and changes in cartilage morphology) and symptomatic progression, highlighting the importance of T2 as a risk factor for OA development. Given that cartilage T2 can detect the earliest stages of cartilage ECM degeneration prior to irreversible cartilage defects, T2 is a distinctive feature of our risk prediction, and the novelty of this study stems from the development of a model that incorporates both standard radiographic assessment and MRI. Thus, an individualized risk assessment that includes cartilage T2 may be used to identify subjects at high risk for the development of OA but at early stages of cartilage biochemical degeneration, at which point preventative efforts may be most effective.

**FIGURE 3:** Illustration of the ROC curves for the risk prediction models. The base model includes age, gender, and BMI. Model 1 = Base + radiography + previous injury (12 months); Model 2 = Model 1 + WORMS; Model 3 = Model 2 + T2.

**FIGURE 4:** The model classification improves from Model 1 (radiography + Risk factors) to Model 3 (Model 1 + WORMS + T2), as shown by the increasing spread of the data. The higher the risk probability, the higher the likelihood for progression; this phenomenon is especially pronounced when comparing Models 1 to Model 3.
The model that included lateral femur and patella cartilage lesions had the best performance, possibly due to the fact that lateral femur (18%) and patella lesions (66%) were found most prevalent out of all cartilage compartments in subjects with risk factors for OA.\textsuperscript{39} The medial femur and tibia T2 had the best model performance compared to other T2 compartments. Several reasons could account for this: medial OA occurs more frequently than lateral OA,\textsuperscript{40,41} data from the OAI show that decreases in cartilage thickness over 1 year were greater in the medial than the lateral compartment,\textsuperscript{42} meniscus and cartilage lesions are more prevalent on the medial side of the joint,\textsuperscript{41} and the medial femur is a concentrated region of weight-bearing.\textsuperscript{41} Also, medial femur and tibia T2 has been shown to be associated with progression of OA.\textsuperscript{4}

We performed a sensitivity analysis examining the performance of the models excluding subjects with KL $\geq$ 2. Out of the 519 subjects with KL = 0 or 1, only 48 had a positive outcome. In the models excluding KL = 2, the cross-validated AUC of the model with T2 was 0.67 and the model with radiography was 0.62. These results may be affected by the lower sample size and the lower occurrence of a positive outcome in the subset of subjects with KL < 2. Also, since KL grade is an important predictor in the models, restricting the range to KL 0/1 will worsen the ability to risk stratify, and thus the AUC will be reduced. The positive aspect of including KL 2 is that it improves the prediction ability of the models.

Preventative efforts such as weight reduction\textsuperscript{17,43} and various levels of exercise\textsuperscript{16,44} may decrease risk for OA progression and may be advised after assessing an individual’s long-term risk for OA. One study found that a weight loss of 5% body weight over 30 months decreased the risk of incident radiographic knee OA\textsuperscript{45}; another study suggested that moderate exercise may be a “good treatment” for subjects at high risk for OA,\textsuperscript{46} and a meta-analysis showed that long-term weight loss is increased when diet and physical activity are combined.\textsuperscript{47} Weight loss also improves joint health and is associated with reduced medial cartilage volume,\textsuperscript{48} and with improvement in the cartilage quality (increase proteoglycan content) and reduced thickness in the medial cartilage.\textsuperscript{49} Thus, BMI is a modifiable risk factor for OA, and weight loss could be recommended if a subject is obese and at high risk for OA. In addition, subjects who

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{risk_score_calculator}
\caption{(a) A graphic of the Risk Score calculator. (b) An illustration of the effects of cartilage T2 on OA risk prediction, while keeping the subject characteristics including KL and WORMS scores constant. As cartilage T2 increases, the risk for OA development increases, as illustrated by the red areas in the “high risk” T2 map.}
\end{figure}
play sports may benefit from injury prevention programs that have been shown to decrease the rate of injury,\textsuperscript{50,51} and consequently decrease the rate of incident OA. Thus, the risk prediction models developed in this study could motivate individuals to adhere to tailored disease-modification strategies, and consequently decrease their risk for OA.

While model performance was significantly improved when comparing Model 1 (radiography) vs. Model 3 (radiography+WORMS+T2), Model 2, which included radiography+WORMS, was not significantly different from Model 3. Based on their AUCs, Models 2 and 3 were characterized as having “fair” performance; however, similar performance values were found for the FRAX score.\textsuperscript{52} Overall, in a clinical environment, a model with MR-based WORMS may be sufficient if T2 is not available. However, if T2 is available, the additional information could aid in risk stratification by providing information on early biochemical cartilage changes, which cannot be detected using WORMS or radiographic findings. A model that includes T2 may be particularly beneficial for research trials targeting therapeutic interventions for early stages of disease. It should be noted that efforts are in place to standardize T2 mapping through the Quantitative Imaging Biomarker Alliance and automatic segmentation algorithm are being developed.\textsuperscript{53,54} Based on these developments, it is likely that reproducible techniques and automated analysis algorithms will be available in the near future. In the meantime, two alternative models (Models 1 and 2) are available that are clinically applicable using standard imaging technologies.

Several limitations are pertinent to this study, including the use of a composite outcome, inclusion of only cartilage T2, and no other compositional measures, the challenges and costs to obtain standardized MRI and perform T2/WORMS analysis, and the lack of external validation in other cohorts. While singular outcomes would have been ideal, a composite outcome was chosen to obtain a broader clinical significance/applicability. In addition to cartilage T2, it would be beneficial to study other quantitative cartilage assessments such as T1rho mapping; however, only T2 was available in the OAI. We did not specifically assess chondrocalcinosis (CPPD) in this cohort, but we did assess this in a different study where we analyzed 2122 subjects and found CPPD only in 99 subjects (4.7%);\textsuperscript{55} given the small number of subjects with CPPD we do not expect a significant impact of CPPD in this relatively young cohort with no or only mild degenerative changes. Bone marrow edema pattern and effusion were not included, which may be considered another limitation; however, these are less frequent in early stages of degenerative joint disease and are not stable (often appearing and resolving). Thus, we decided not to include them in this model. We did not assess the secondary knee in the models due to the fact that cartilage T2 was only available in the right knee in the OAI. We did not account for potential between-knee interactions such as malalignment in the secondary knee (in this cohort only n = 23 subjects had contralateral KL ≥3 at baseline), which may have altered loading patterns in the primary knee, or knee injury in the secondary knee. In addition, we were unable to perform external validation as, to the best of our knowledge, no large longitudinal databases with cartilage T2 exist. Also, Since T2 values are known to vary based on acquisition methods, vendors, coils, and postprocessing techniques, a standardized imaging protocol would be necessary for these models to be utilized clinically. In addition, implementing a model with cartilage T2 may be complicated in the clinic due the required cartilage segmentation and post-processing, which require a significant amount of manpower and time. However, we believe that the ongoing work to standardize T2 mapping through the Quantitative Imaging Biomarker Alliance and to implement automatic segmentation techniques using Artificial Intelligence Algorithms will facilitate the translation of cartilage T2 mapping clinically. Currently, a model including WORMS is clinically more feasible, as MRI sequences are routinely acquired, and a radiologist can detect the presence of focal cartilage lesions or meniscal tears without difficulty, although reproducibility may vary. Other concerns that may be raised are that radiography and clinical data are not always routinely collected and may therefore be challenging to implement in a risk prediction model; standardized questionnaires and patient management, however, would facilitate these issues. Despite these limitations, we believe this study is the first step in the development of a risk prediction model that includes advanced MRI.

In conclusion, this study showed that a risk prediction model that includes advanced MRI has a higher performance than a model with only subject demographics, risk factors, and radiography. Since the difference between Models 1 and 3 reached statistical significance and the difference between 1 and 2 did not, perhaps a larger study should be undertaken to assess if Model 2 is sufficient for risk prediction. Overall, information about an individual’s risk for OA would be critical for the development of personalized treatment plans and preventative lifestyle interventions such as weight loss or exercise modification to improve long-term symptoms and overall knee degeneration.

**Role of Funding Source**

The OAI is a public-private partnership comprised of five contracts (N01-AR-2-2258; N01-AR-2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services, and conducted by the OAI Study Investigators. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals; GlaxoSmithKline; and
References


