Bone texture analysis for prediction of incident radiographic hip osteoarthritis using machine learning

data from the Cohort Hip and Cohort Knee (CHECK) study

Hirvasniemi, J.; Gielis, W. P.; Arbabi, S.; Agricola, R.; van Spil, W. E.; Arbabi, V.; Weinans, H.

DOI
10.1016/j.joca.2019.02.796

Publication date
2019

Document Version
Accepted author manuscript

Published in
Osteoarthritis and Cartilage

Citation (APA)
Hirvasniemi, J., Gielis, W. P., Arbabi, S., Agricola, R., van Spil, W. E., Arbabi, V., & Weinans, H. (2019). Bone texture analysis for prediction of incident radiographic hip osteoarthritis using machine learning: data from the Cohort Hip and Cohort Knee (CHECK) study. *Osteoarthritis and Cartilage*, *27*(6), 906-914. https://doi.org/10.1016/j.joca.2019.02.796

Bone Texture Analysis for Prediction of Incident Radiographic Hip Osteoarthritis Using Machine Learning: Data from the Cohort Hip and Cohort Knee (CHECK) study

Jukka Hirvasniemi[1,2], Willem Paul Gielis[2], Saeed Arbabi[3], Rintje Agricola[4], Willem Evert van Spil[5], Vahid Arbabi[2,6,7], Harrie Weinans[2,4,6]

[1]Center for Machine Vision and Signal Analysis, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland

[2]Department of Orthopedics, University Medical Center Utrecht, Utrecht, The Netherlands

[3]Department of Computer Engineering, Faculty of Engineering, University of Zabol, Zabol, Iran

[4]Department of Orthopaedics, Erasmus University Medical Center, Rotterdam, the Netherlands

[5]Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

[6]Department of Biomechanical Engineering, Delft University of Technology, Delft, The Netherlands

[7]Department of Mechanical Engineering, Faculty of Engineering, University of Birjand, Birjand, Iran

E-mail addresses

Jukka Hirvasniemi: jukka.hirvasniemi@oulu.fi

Willem Paul Gielis: w.p.gielis@umcutrecht.nl

Saeed Arbabi: gsarbabi@gmail.com

Rintje Agricola: r.agricola@erasmusmc.nl

Willem Evert van Spil: w.e.vanspil@umcutrecht.nl

Vahid Arbabi: v.arbabi@umcutrecht.nl

Harrie Weinans: h.h.weinans@umcutrecht.nl

Running title: Bone texture and incident hip OA

Address correspondence and reprint requests to: Jukka Hirvasniemi, Center for Machine Vision and Signal Analysis, Faculty of Information Technology and Engineering, University of Oulu, Oulu, Finland, PO Box 4500, FI-90014 Oulu, Finland. Tel: +358 294482778. E-mail address: jukka.hirvasniemi@oulu.fi.

Abstract

Objective: To assess the ability of radiography-based bone texture variables in proximal femur and acetabulum to predict incident radiographic hip osteoarthritis (rHOA) over a 10 years period.

Design: Pelvic radiographs from CHECK (Cohort Hip and Cohort Knee) at baseline (987 hips) were analyzed for bone texture using fractal signature analysis in proximal femur and acetabulum. Elastic net (machine learning) was used to predict the incidence of rHOA (including Kellgren-Lawrence grade (KL) ≥ 2 or total hip replacement (THR)), joint space narrowing score (JSN, range 0-3), and osteophyte score (OST, range 0-3) after 10 years. Performance of prediction models was assessed using the area under the receiver operating characteristic curve (ROC AUC).

Results: Of the 987 hips without rHOA at baseline, 435 (44%) had rHOA at 10-year follow-up. Of the 667 hips with JSN grade 0 at baseline, 471 (71%) had JSN grade ≥ 1 at 10-year follow-up. Of the 613 hips with OST grade 0 at baseline, 526 (86%) had OST grade ≥ 1 at 10-year follow-up. AUCs for the models including age, gender, and body mass index to predict incident rHOA, JSN, and OST were 0.59, 0.54, and 0.51, respectively. The inclusion of bone texture variables in the models improved the prediction of incident rHOA (ROC AUC 0.68 and 0.71 when baseline KL was also included in the model) and JSN (ROC AUC 0.62), but not incident OST (ROC AUC 0.52).

Conclusion: Bone texture analysis provides additional information for predicting incident rHOA or THR over 10 years.

Keywords: Radiography, hip osteoarthritis, prediction, bone texture, machine learning

1. Introduction

Plain radiography is a cheap, fast and widely available imaging method for osteoarthritis (OA). Bony changes can be clearly seen on plain radiographs and provide useful information about bone deformities, density, and structure. A plain radiograph is a projection (summation) through a three-dimensional structure and this is one main limitation of this imaging method. However, it has been shown that radiography-based bone texture is significantly related with the three-dimensional structure of bone[1-5].

Medical image analysis often involves interpretation of tissue appearance, *e.g.*, smooth, grainy, rough, or homogenous. These image properties are related to the spatial arrangement of pixel intensities in images, *i.e.*, image texture, and can be quantified using texture analysis[6]. Radiography-based texture analysis of the proximal femur has been applied for example in osteoporosis and in the assessment of femoral neck fracture risk[7, 8]. However, in OA research, the majority of studies analyzing bone texture are concentrated on the knee, using mostly fractal-based texture analysis methods[1, 9-13]. There is evidence that tibial trabecular bone texture can be used to predict both development and progression of OA as well as total knee replacement[14-21]. Only one study applied fractal signature analysis (FSA) on hip radiographs to quantify trabecular bone changes in subjects with prevalent hip OA and reported changes in fractal dimension of femoral head between baseline and 18 months follow-up[22]. However, the sample size of that study was relatively small (14 subjects) and the follow-up rather short.

Given the previous results showing that FSA can be applied on hip radiographs[22] and that bone density related variables from dual energy X-ray absorptiometry (DXA) contribute to the risk and progression of hip OA[23-26], we hypothesize that radiography-based bone texture gives additional information in predicting the development of radiographic hip OA (rHOA). Consequently, our aim was to create a method for automated assessment of bone texture in

proximal femur and acetabulum from plain hip radiographs and to assess the ability of these bone texture variables to predict incident rHOA.

2. Subjects and Methods

2.1. Study cohort

Data from CHECK (Cohort Hip and Cohort Knee) cohort was used in this study[27]. CHECK is a prospective cohort study of 1002 subjects initiated to study the course of early knee and hip OA. Data was collected in ten medical centers in The Netherlands. Subjects were recruited by general practitioners and via advertisements. At baseline, subjects were aged 45-65 years, had first onset of pain or stiffness in hip(s) and/or knee(s), and had never or not longer than 6 months ago consulted a physician for these complaints. Subjects with a pathological condition other than early OA that could explain symptoms were excluded. The study was approved by medical ethics committees of all ten participating centers and written informed consent was obtained from all participants.

Plain pelvic anterior-posterior radiographs and clinical data at baseline and 10-year follow-up were used in the current study. Subjects with missing data (radiographs, demographics, clinical examination), KL grade ≥ 2 at baseline, and/or with insufficient radiograph quality (artefacts or underexposed) were excluded (Figure 1). As such, the final subset for assessing incident rHOA (KL grade ≥ 2 or total hip replacement (THR)) within the period from baseline to 10 years included 987 hips from 601 subjects (Table 1).

2.2. Acquisition and grading of the radiographs

Weight-bearing, anterior-posterior pelvic radiographs were acquired according to a standardized protocol. A wedge was used to assure 15-degree internal rotation in feet. The

source – detector distance was 120 cm, and the X-ray beam was centered on the superior part of the pubic symphysis.

Hips were classified according to the KL grading scale[28] at baseline and 10-year follow-up. Superior and medial joint space narrowing (JSN) and superior and inferior osteophytes (OST) in acetabulum/femur were classified according to the Osteoarthritis Research Society International (OARSI) grading scale[29]. The scale for classifying the changes in JSN and OST was from 0 (normal) to 3 (severe change)[29]. Highest JSN and OST grades of the analyzed regions were used in the analyses. Table 1 summarizes the distribution of the KL, JSN, and OST grades of included hips at baseline and 10-year follow-up.

2.3. Selection of regions of interests

Prior to extraction of the regions of interests (ROIs), all images were resampled to have the same femoral head diameter in pixels as the smallest femoral head diameter on the data. Bicubic interpolation was used to ensure comparability of the structural variables, without producing as much artefacts as bilinear or nearest neighbor interpolation algorithms. The resampling was also needed because part of the baseline radiographs were digitized and saved in TIFF images (501 hip images were in DICOM and 486 in TIFF format) and the actual pixel size on the detector was not available. After resampling, to assess bone texture from the radiographs, 41 circular ROIs with 70 pixels diameter were extracted from femoral head and acetabulum (25 on femoral head and 16 on acetabulum) (Figure 2). Although previous studies have typically used rectangular ROIs, circular ROIs were used in the current study to better cover femoral head and acetabular area and to enable bone texture assessment in many different directions inside the ROI without losing pixels when rotating the ROI. ROI selection procedure was based on fourteen out of seventy-five landmarks, which were manually placed

on the proximal femur and pelvis in a previous study (Figure 2)[30]. Two circles were fitted in femoral head and acetabulum for ROI placement using Least Squares Optimization algorithm, which calculates the center and radius of best fitting circle in an iterative process. Locations of the ROIs were determined after a robustness assurance step which guarantees that the same ROI number selects the same corresponding pixels on images despite the size and rotation differences among the images. Locations of the ROI1, ROI17, ROI25, and ROI26 were defined based on the center of the femoral head and the second-most lateral landmark on the acetabular rim (Figure 2). Other ROIs were automatically placed based on the locations of those ROIs.

2.4. Bone texture analysis

Before texture analysis, images were median filtered with a 3x3 pixels filter to remove high-frequency noise and grayscale values were expanded to full dynamic range (0 – 255). Bone texture was assessed using the FSA method[1, 10, 11]. FSA produces fractal dimension values that are related to the roughness and complexity of the image. To calculate the fractal dimensions, the image was dilated and eroded with a rod-shaped, horizontally oriented, one-pixel wide structuring element. After that, the volume, *V*, between dilated and eroded images was calculated. Calculations were repeated by varying the element length *r* from 2 to 5 pixels. The surface area, *A(r)*, was obtained from the Equation 1:

$$A(r) = (V(r) - V(r-1))/2, \quad (1)$$

Subsequently, a log-log plot was constructed by plotting log of *A(r)* against log of *r*. Finally, the fractal dimension was estimated by the slope of regression line that fitted the three points in the log-log plot. High fractal dimension values are associated with high complexity of the image, whereas low complexity results in low fractal dimension values.

Because the orientation of the bone structures in femoral head and acetabulum varies, we assessed fractal dimensions in 18 different angles, *i.e.* from 0 to 170 degrees with 10 degrees increments, by rotating the ROI. To reduce the number of fractal dimension values (18 values for 41 ROIs = 738 features per hip), minimum (FD_{min}) and maximum (FD_{max}) fractal dimension values and their respective angles ($Angles_{min}$, $Angles_{max}$) per ROI were selected. Consequently, 164 texture features were used in the analyses.

2.5. Statistical analyses

Logistic regression with generalized estimating equation (GEE) was used to assess the association of each baseline covariate (age, gender, body mass index (BMI), and baseline KL grade (KL 0 or KL1)) as well as bone texture variable with incident rHOA (KL grade ≥ 2 or THR), incident JSN (JSN grade ≥ 1 or THR), and incident OST (OST grade ≥ 1 or THR). GEE was used to take into account the potential correlation between bilateral measurements. Odds ratios (OR) and 95% confidence interval of the univariable models were reported. It should be noted that the OR is not directly interpretable as a relative risk, because the OR most likely overestimates the relative risk. The analyses were performed using IBM SPSS Statistics (version 25, IBM Corp., Armonk, NY, USA).

To prevent overfitting, machine learning was used for dimensionality reduction and to assess the predictive ability of the bone texture variables and baseline covariates. For the dimensionality reduction and prediction, a regularized logistic regression method called elastic net was used[31, 32]. The elastic net linearly combines the L_1 and L_2 penalties of lasso and ridge regression methods[31, 32]. The samples were randomly divided into a training and validation set (790 hips, 80% of the data) and a hold-out test set (197 hips, 20% of the data) by stratifying the proportion of the controls and subjects with incident rHOA at follow-up in

each set. To optimize the ratio of the L_1 and L_2 penalties (α) and the strength of the penalty parameter (λ) of the elastic net, 10-fold cross-validation was performed. When α is close to zero, the elastic net approaches ridge regression, while when α is 1, lasso regression is performed. The combination of the α- and λ-parameters that had the highest area under the receiver operating characteristics curve (ROC AUC) in the cross-validation, was selected. After optimizing elastic net parameters, the predictive ability of 1) covariate model, 2) texture feature model, 3) covariate + baseline KL (KL0 or KL1) model, and 4) model with covariates, baseline KL, and texture features combined in the test set were assessed using ROC AUC value. The agreement between observed outcomes and predictions in the test set were assessed using calibration plots[33]. Analyses were repeated to predict incident rHOA among subjects with only KL0 or KL1 at baseline separately. Furthermore, incident JSN (JSN $\geq$ 1 or THR) among subjects with JSN grade 0 at baseline and incident OST (OST grade $\geq$ 1 or THR) among subjects with OST grade 0 at baseline were predicted with the elastic net. Distribution of the proportion of incident rHOA cases was different between imaging centers. To remove the potential effect of imaging center to fractal dimension values, the variables were standardized with mean and standard deviation values of the center where the imaging was performed ($z = (x - \mu)/SD$, where x is the value of each measurement, μ and SD are the average and standard deviation of the variable at the center where the imaging was performed). Statistical analyses were performed using R (version 3.4.1) software with Caret[34] (version 6.0-47), pROC[35] (version 1.13.0), glmnet[31] (version 2.0-16), and CalibrationCurves[36] (version 0.1.2) packages.

3. Results

Of the 987 hips without rHOA at baseline, 435 (44%) had developed incident rHOA (KL ≥ 2 or THR) at 10-year follow-up (Table 1). Of the 667 hips with JSN grade 0, 471 (71%) had JSN grade ≥ 1 or THR at 10-year follow-up. Of the 613 hips with OST grade 0, 526 (86%) had OST grade ≥ 1 or THR at 10-year follow-up.

The GEE models for covariates showed that higher age (OR: 1.06) and baseline KL grade (OR: 3.80) were associated with significantly (p<0.05) higher odds (Table 2). Female gender (OR: 0.54) was associated with lower odds of incident rHOA at the 10-year follow-up (Table 2). When looking at the univariable texture models, ORs for minimum fractal dimension (FD_{min}) variable models were statistically significant in 11/41 ROIs (OR range: 0.71 – 1.15), the maximum fractal dimension (FD_{max}) in 8/41 ROIs (OR range: 0.83 – 1.11), the angle associated with FD_{min} ($Angles_{min}$) in 5/41 ROIs (OR range: 0.96 – 1.02), and the angle associated with FD_{max} ($Angles_{max}$) in 12/41 ROIs (OR range: 0.93 – 1.05) (Table 3). For incident JSN grade ≥ 1, ORs for 9 texture variables (OR range: 0.64 – 1.43) were significant (Supplementary Tables 6 and 7). For incident OST grade ≥ 1, ORs for 5 texture variables (OR range: 0.78 – 1.08) were significant (Supplementary Tables 8 and 9).

The selected elastic net parameters and ROC AUC values for the covariate model, texture model, covariate + baseline KL model, and for the combined covariate, baseline KL, and texture feature model from 10-fold cross-validation are shown in Table 4. The model that included covariates, baseline KL, and texture features had the highest ROC AUC (0.73 [95% CI: 0.70 – 0.76]) in cross-validation. The variables that were selected based on the 10-fold cross-validation of the elastic net are listed in Supplementary Table 5 and visualized in Figure 3. BMI was not selected in any of the models by the algorithm.

When assessing the performance of the optimized elastic net models in test set, the combined covariate, baseline KL, and bone texture feature model had the highest AUC (0.71 [95% CI: 0.63 – 0.78]). ROC curves for the 1) covariate, 2) texture, 3) combined covariates and baseline KL, and 4) combined covariate, baseline KL, and texture feature models in the test set are shown in Figure 4. Calibration-in-the-large coefficients of the models varied from -0.42 to -0.31 and calibration slopes from 0.81 to 1.56 (Supplementary Figure 5).

When assessing subjects with KL0 or KL1 at baseline separately, adding the texture variables improved model performance in the test set. ROC AUC value increased from 0.51 [95% CI: 0.41 – 0.61] to 0.62 [95% CI: 0.53 – 0.72] among KL0 subjects (Supplementary Figure 6) and from 0.54 [95% CI: 0.37 – 0.70] to 0.65 [95% CI: 0.48 – 0.81] among KL1 subjects (Supplementary Figure 7). Calibration plots of texture and covariate + texture models among KL0 subjects showed an agreement between actual and predicted risks for low risk subjects, but failed for high risk subjects (data not shown). For KL1 subjects, calibration plots showed a reasonable agreement between actual and predicted risks for high risk subjects, but failed for low risk subjects (data not shown). For analyses in subjects with KL0 at baseline, ROC AUC values in cross-validation were 0.60 (95% CI: 0.55 – 0.66), 0.60 (95% CI: 0.57 – 0.62), and 0.62 (95% CI: 0.59 – 0.64) for covariate, texture, and covariate + texture models, respectively. For analyses in subjects with KL1 at baseline, ROC AUC values in cross-validation were 0.62 (95% CI: 0.56 – 0.69), 0.71 (95% CI: 0.65 – 0.76), and 0.71 (95% CI: 0.63 – 0.79) for covariate, texture, and covariate + texture models, respectively.

When predicting incident JSN in the test set, the combined texture and covariate model had the highest ROC AUC value of 0.62 (95% CI: 0.52 – 0.72) (Supplementary Figure 8). Calibration plots of texture and covariate + texture models showed a reasonable agreement between actual and predicted risks (data not shown). ROC AUC values in cross-validation

were 0.54 (95% CI: 0.50 – 0.58), 0.67 (95% CI: 0.63 – 0.71), and 0.67 (95% CI: 0.63 – 0.72) for covariate, texture, and covariate + texture models, respectively.

Selected models performed poorly when predicting incident OST in the test set (Supplementary Figure 9) and showed a poor calibration (data not shown). The highest ROC AUC value (0.52 [95% CI: 0.38 – 0.65]) was obtained with the combined texture and covariate model. ROC AUC values in cross-validation were 0.55 (95% CI: 0.45 – 0.65), 0.60 (95% CI: 0.52 – 0.68), and 0.60 (95% CI: 0.53 – 0.67) for covariate, texture, and covariate + texture models, respectively.

4. Discussion

In this study, we created a method for the assessment of bone texture in proximal femur and acetabulum from plain pelvic radiographs and assessed the ability of bone texture to predict incident rHOA (KL $\geq$ 2 or THR). Fractal dimension was measured from 41 ROIs that were placed on femoral head and acetabulum. Inclusion of bone texture variables in the prediction model increased the ROC AUC value of the model during cross-validation (from 0.69 to 0.73) and in the hold-out test set (from 0.68 to 0.71) as compared to the model with baseline patient characteristics and baseline KL grade.

As there were no previous data for the optimal location of ROIs for bone texture analysis in the hip, we decided to cover the whole femoral ROI and also incorporate the acetabulum in our analyses. As shown in a previous study for the knee, areas distal from the subchondral bone might also include relevant texture information[15]. Interestingly, majority of the relevant ROIs for the prediction of rHOA in our analyses were either at or next to the principal compressive trabeculae or close to the joint space (Figure 3). Depending on the ROI and variable, either higher or lower values were predictive for rHOA. For example, higher values in FD_{min} in ROIs 15, 24, 27, 32 and 33 and lower values in ROIs 1, 2, 6, 10, 19, 28, 29, 40, and 41 were predictive for rHOA. Lower fractal dimension values are likely associated with trabecular thickening or a reduction in trabecular number[22]. Changes in the angles associated with the fractal dimensions indicate the changes in the orientation of the trabeculae within the ROI. For example higher $Angles_{max}$ values indicate that the maximum fractal dimension value (FD_{max}) was detected from higher angles. One possible explanation for the differences between the directions of the predictive values in the ROIs may be the adaptation of bone according to the daily loading of the joint. Some areas may experience higher loads whereas

the loads may be reduced in other areas. Furthermore, subchondral bone sclerosis affect the values in areas that are near the joint space.

In a previous study assessing changes in FSA in OA hips, 14 subjects were followed 18 months. Fractal dimension of small and medium sized structures in the image were decreased during follow-up probably due to trabecular thickening or a reduction in trabecular number[22]. However, the relationship between FSA and development and/or progression of OA was not studied. Another study assessed texture on the hip joint space area and reported 88.9% classification accuracy between controls and OA subjects. However, that study was cross-sectional, did not assess femoral head or acetabular bone texture, and had a limited sample size (n = 64)[37].

In contrast to the scarce assessment of associations between bone texture and hip OA, the association between tibial bone texture and knee OA has been described in many papers[14-18, 20, 21]. ROC AUC values between 0.65 and 0.79 have been reported for models predicting progression or development of knee OA using bone texture and clinical covariates (*e.g.*, age, BMI and gender)[14-16, 18, 21]. Associations of covariates and bone texture with incident rHOA in our study are in line with these results. The ROC AUC to predict rHOA increased from 0.68 to 0.71 after including bone texture to the model that included covariates and KL grading. Relatively low increase in ROC AUC may be because baseline KL grade alone is already a quite strong predictor of rHOA. Bone texture does not directly provide information about JSN, whereas JSN affects directly to KL grade, as this is included in its definition. However, trabecular bone structure is not evaluated in KL grading whereas bone texture analysis provides information about that.

When the ability to predict incident rHOA was compared between subjects with KL0 and KL1 at baseline, ROC AUC values were higher for KL1 subjects (ROC AUC values for the

full model: 0.62 vs 0.65). These results suggest that in KL0 subjects there might not yet be changes that can be captured with bone texture analysis and predict incident rHOA.

In this study, the ability to predict incident JSN or OST in the hip was worse than that reported for the knee[14]. Our results showed that ROC AUC to predict incident JSN (0.62) was higher than incident OST (0.52). Distribution of controls and cases was more unbalanced for incident OST analyses, which may explain some of the difference. Another reason might be that OSTs were assessed from four different locations and combined into one variable, whereas JSN was assessed from two different locations and combined into one variable. Furthermore, JSN and OST are different characteristics of OA.

We decided to resample all images to same size (in pixels) based on the femoral head diameter, pre-process images with median filtering, and standardize FSA values within each center, because ten different medical centers with different proportions of incident rHOA cases (p<0.05) participated in the data collection and different X-ray machines were used for the imaging. We believe that the standardization lower the possibility that the prediction models recognize a center based on texture values and use that information to improve the prediction. It should be noted that the performance of the models were very similar without pre-processing and/or standardization (0.01 – 0.03 difference in ROC AUCs, data not shown). The resampling of the images and the lack of actual pixel size of some images complicates the assessment of the actual scale that the FSA quantified here. However, based on the images that had the pixel size was available, the scale was around 600 μm.

Assessment of calibration (agreement between observed outcomes and predictions) in the test set indicated that predicted risks were on average slightly overestimated when predicting rHOA (calibration-in-the-large coefficients < 0). The model using only FSA variables showed some overfitting (calibration slope < 1) and other models some underfitting (calibration slope > 1). For the elastic net models, the combination of the α- and λ-parameters that had the

highest ROC AUC in the cross-validation, were selected. We also tested selecting the largest λ-value such that error in cross-validation was within 1 standard error of the λ that had the highest ROC AUC, but calibration plots indicated that these models were in general underfitted (data not shown). Our sample size was too low for reliable assessment of the calibration intercept and slope when predicting other outcomes than incident rHOA (123 controls, 74 OA in the test set), because at least 100 events and 100 non-events have been recommended[36] and were not therefore calculated.

This study contains some limitations that need to be discussed. First, the relevance of bone changes to the OA disease process might differ between HOA phenotypes. In the current study, different phenotypes were mixed as there is no consensus on how to define OA phenotypes yet. Second, radiographic scoring is subjective, semi-quantitative, and a plain radiograph is a projection of 3-dimensional structure. Therefore, some OA changes may have been missed. Third, rHOA and THR were combined as an outcome, while they might be different. We decided to combine these outcomes due to low number of THR subjects. Fourth, ten different medical centers with different proportions of incident rHOA cases (p<0.05) participated in the data collection and different X-ray machines were used for the imaging, which may have affected texture analysis. However, FSA has been shown to be robust to the changes in imaging settings (*e.g.*, exposure and pixel size)[11] and FSA values were standardized within each center. We think that including data from multiple X-ray machines increases the generalizability of our results. Fifth, training, validation, and test sets were derived from CHECK and the model was not tested in another cohort. However, to reduce overfitting, the hold-out test set was not used in cross-validation and the optimal elastic net parameters were searched using 10-fold cross-validation. Sixth, some ROIs (*e.g.*, ROIs 32-38 in acetabulum) may contain both cortical and trabecular bone and should be considered when interpreting the results. Seventh, the potential correlation between bilateral

measurements was not taken into account in the elastic net analyses, which may have introduced some bias into our analyses. However, correlation between hips was taken into the account in the GEE analyses.

In conclusion, bone texture analysis in proximal femur and acetabulum provides additional information when trying to predict incident rHOA or THR. Our results suggest that bone texture variables could be valuable when building prediction tools for OA. Given the current results and the previous results on knee, a similar analysis approach could be tested in other joints (*e.g.*, hands) as well in the future.

Author contributions

Conception and design: JH, WPG, VA, HW; acquisition of data: JH, SA, RA, VA; analysis and interpretation of the data: all authors; drafting of the article: JH; critical revision of the article for important intellectual content: all authors; final approval of the article: all authors.

Role of the funding source

The research leading to these results has received funding from the Academy of Finland (project 308165). The funding sources had no role in the study design, data collection or analysis, interpretation of data, writing of the manuscript, or in the decision to submit the manuscript for publication.

Conflict of interest

The authors report no conflicts of interests.

Acknowledgements

The CHECK-cohort study is funded by the Dutch Arthritis Foundation. Involved are: Erasmus Medical Center Rotterdam; Kennemer Gasthuis Haarlem; Leiden University Medical Center; Maastricht University Medical Center; Martini Hospital Groningen /Allied Health Care Center for Rheumatology and Rehabilitation Groningen; Medical Spectrum Twente Enschede /Ziekenhuisgroep Twente Almelo; Reade Center for Rehabilitation and

Rheumatology; St.Maartens-kliniek Nijmegen; University Medical Center Utrecht and Wilhelmina Hospital Assen.

References

1. Hirvasniemi J, Thevenot J, Kokkonen HT, Finnilä MA, Venäläinen MS, Jämsä T, et al. Correlation of Subchondral Bone Density and Structure from Plain Radiographs with Micro Computed Tomography Ex Vivo. Annals of Biomedical Engineering 2016; 44: 1698-709.
2. Le Corroller T, Pithioux M, Chaari F, Rosa B, Parratte S, Maurel B, et al. Bone texture analysis is correlated with three-dimensional microarchitecture and mechanical properties of trabecular bone in osteoporotic femurs. Journal of Bone and Mineral Metabolism 2013; 31: 82-8.
3. Ollivier M, Le Corroller T, Blanc G, Parratte S, Champsaur P, Chabrand P, et al. Radiographic bone texture analysis is correlated with 3D microarchitecture in the femoral head, and improves the estimation of the femoral neck fracture risk when combined with bone mineral density. European Journal of Radiology 2013; 82: 1494-8.
4. Ranjanomennahary P, Ghalila SS, Malouche D, Marchadier A, Rachidi M, Benhamou CL, et al. Comparison of radiograph-based texture analysis and bone mineral density with three-dimensional microarchitecture of trabecular bone. Medical Physics 2011; 38: 420-8.
5. Steines D, Liew SW, Arnaud C, Vargas-Voracek R, Nazarian A, Müller R, et al. Radiographic trabecular 2D and 3D parameters of proximal femoral bone cores correlate with each other and with yield stress. Osteoporosis International 2009; 20: 1929-38.

6. Bankman IN, Spisz TS, Pavlopoulos S. Two-Dimensional Shape and Texture Quantification. In: Handbook of Medical Image Processing and Analysis (Second Edition), Bankman IN Ed. Burlington: Academic Press 2009:261-77.
7. Pulkkinen P, Saarakkala S, Nieminen MT, Jämsä T. Standard radiography: untapped potential in the assessment of osteoporotic fracture risk. European radiology 2013; 23: 1375-82.
8. Thevenot J, Hirvasniemi J, Pulkkinen P, Maatta M, Korpelainen R, Saarakkala S, et al. Assessment of risk of femoral neck fracture with radiographic texture parameters: a retrospective study. Radiology 2014; 272: 184-91.
9. Buckland-Wright C. Subchondral bone changes in hand and knee osteoarthritis detected by radiography. Osteoarthritis and Cartilage 2004; 12: S10-9.
10. Lynch JA, Hawkes DJ, Buckland-Wright JC. Analysis of texture in macroradiographs of osteoarthritic knees using the fractal signature. Physics in Medicine & Biology 1991; 36: 709-22.
11. Lynch JA, Hawkes DJ, Buckland-Wright JC. A robust and accurate method for calculating the fractal signature of texture in macroradiographs of osteoarthritic knees. Medical Informatics 1991; 16: 241-51.
12. Hirvasniemi J, Thevenot J, Guermazi A, Podlipská J, Roemer FW, Nieminen MT, et al. Differences in tibial subchondral bone structure evaluated using plain radiographs between knees with and without cartilage damage or bone marrow lesions-the Oulu Knee Osteoarthritis study. European radiology 2017; 27: 4874-82.
13. Hirvasniemi J, Thevenot J, Multanen J, Haapea M, Heinonen A, Nieminen MT, et al. Association between radiography-based subchondral bone structure and MRI-based cartilage composition in postmenopausal women with mild osteoarthritis. Osteoarthritis and Cartilage 2017; 25: 2039-46.

14. Janvier T, Jennane R, Toumi H, Lespessailles E. Subchondral tibial bone texture predicts the incidence of radiographic knee osteoarthritis: data from the Osteoarthritis Initiative. Osteoarthritis and Cartilage 2017; 25: 2047-54.
15. Janvier T, Jennane R, Valery A, Harrar K, Delplanque M, Lelong C, et al. Subchondral tibial bone texture analysis predicts knee osteoarthritis progression: data from the Osteoarthritis Initiative: tibial bone texture & knee OA progression. Osteoarthritis and Cartilage 2017; 25: 259-66.
16. Kraus VB, Feng S, Wang S, White S, Ainslie M, Brett A, et al. Trabecular morphometry by fractal signature analysis is a novel marker of osteoarthritis progression. Arthritis and Rheumatism 2009; 60: 3711-22.
17. Kraus VB, Feng S, Wang S, White S, Ainslie M, Le Graverand MPH, et al. Subchondral bone trabecular integrity predicts and changes concurrently with radiographic and magnetic resonance imaging-determined knee osteoarthritis progression. Arthritis and Rheumatism 2013; 65: 1812-21.
18. Kraus VB, Collins JE, Charles HC, Pieper CF, Whitley L, Losina E, et al. Predictive validity of radiographic trabecular bone texture in knee osteoarthritis: the Osteoarthritis Research Society International/Foundation for the National Institutes of Health Osteoarthritis Biomarkers Consortium. Arthritis & Rheumatology 2018; 70: 80-7.
19. Podsiadlo P, Cicuttini FM, Wolski M, Stachowiak GW, Wluka AE. Trabecular bone texture detected by plain radiography is associated with an increased risk of knee replacement in patients with osteoarthritis: a 6 year prospective follow up study. Osteoarthritis and Cartilage 2014; 22: 71-5.
20. Podsiadlo P, Nevitt MC, Wolski M, Stachowiak GW, Lynch JA, Tolstykh I, et al. Baseline trabecular bone and its relation to incident radiographic knee osteoarthritis and

increase in joint space narrowing score: directional fractal signature analysis in the MOST study. Osteoarthritis and Cartilage 2016; 24: 1736-44.

21. Woloszynski T, Podsiadlo P, Stachowiak GW, Kurzynski M, Lohmander LS, Englund M. Prediction of progression of radiographic knee osteoarthritis using tibial trabecular bone texture. Arthritis and Rheumatism 2012; 64: 688-95.
22. Papaloucas CD, Ward RJ, Tonkin CJ, Buckland-Wright C. Cancellous bone changes in hip osteoarthritis: a short-term longitudinal study using fractal signature analysis. Osteoarthritis and Cartilage 2005; 13: 998-1003.
23. Barbour KE, Murphy LB, Helmick CG, Hootman JM, Renner JB, Jordan JM. Bone Mineral Density and the Risk of Hip and Knee Osteoarthritis: The Johnston County Osteoarthritis Project. Arthritis care & research 2017; 69: 1863-70.
24. Betancourt MCC, Van der Linden JC, Rivadeneira F, Rozendaal RM, Zeinstra SMB, Weinans H, et al. Dual energy x-ray absorptiometry analysis contributes to the prediction of hip osteoarthritis progression. Arthritis Research & Therapy 2009; 11: R162.
25. Hardcastle SA, Dieppe P, Gregson CL, Hunter D, Thomas GE, Arden NK, et al. Prevalence of radiographic hip osteoarthritis is increased in high bone mass. Osteoarthritis and Cartilage 2014; 22: 1120-8.
26. Javaid MK, Lane NE, Mackey DC, Lui LY, Arden NK, Beck TJ, et al. Changes in proximal femoral mineral geometry precede the onset of radiographic hip osteoarthritis: the study of osteoporotic fractures. Arthritis & Rheumatology 2009; 60: 2028-36.
27. Wesseling J, Boers M, Viergever MA, Hilberdink WKHA, Lafeber FPJG, Dekker J, et al. Cohort profile: cohort hip and cohort knee (CHECK) study. International journal of epidemiology 2014; 45: 36-44.

28. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Annals of the Rheumatic Diseases 1957; 16: 494-502.
29. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis and Cartilage 2007; 15: A1-A56.
30. Agricola R, Reijman M, Bierma-Zeinstra SM, Verhaar JA, Weinans H, Waarsing JH. Total hip replacement but not clinical osteoarthritis can be predicted by the shape of the hip: a prospective cohort study (CHECK). Osteoarthritis and Cartilage 2013; 21: 559-64.
31. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software 2010; 33: 1-22.
32. Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2005; 67: 301-20.
33. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass.) 2010; 21: 128.
34. Kuhn M. Building predictive models in R using the caret package. Journal of statistical software 2008; 28: 1-26.
35. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S to analyze and compare ROC curves. BMC bioinformatics 2011; 12: 77.
36. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. Journal of clinical epidemiology 2016; 74: 167-76.

37. Boniatis IS, Costaridou LI, Cavouras DA, Panagiotopoulos EC, Panayiotakis GS. Quantitative assessment of hip osteoarthritis based on image texture analysis. The British journal of radiology 2006; 79: 232-8.

Figure legends

Figure 1. Flowchart of the selection of CHECK subjects and hips for the current study.

Figure 2. Location of regions of interest (ROIs). Landmarks that were used when fitting circles to the femoral head and acetabulum are indicated with white "x". The white dashed line shows the centers of the femoral head, ROI1, ROI25 and ROI26. Other ROIs were automatically placed based on the locations of those ROIs.

Figure 3. Location of regions of interest (ROIs) selected to the final elastic net model.

Figure 4. Receiver operating characteristics (ROC) curves and respective area under the curve (AUC) values for predicting incident rHOA (KL ≥ 2 or THR) using 1) covariates (age, gender and body mass index), 2) texture variables from fractal signature analysis (FSA), 3) covariates and baseline KL grade, and 4) texture variables combined with covariates and KL grade.

Supplementary Figure 5. Calibration plots for models predicting incident rHOA (KL ≥ 2 or THR) using (**A**) covariates (age, gender and body mass index), (**B**) covariates and baseline KL grade, (**C**) texture variables from fractal signature analysis (FSA), and (**D**) texture variables combined with covariates and KL grade.

Supplementary Figure 6. Receiver operating characteristics (ROC) curves and their respective area under the curve (AUC) values for predicting incident rHOA (KL ≥ 2 or THR) among subjects with KL0 grade at baseline using 1) texture variables from fractal signature

analysis (FSA), 2) covariates (age, gender, and body mass index), and 3) texture variables combined with covariates.

Supplementary Figure 7. Receiver operating characteristics (ROC) curves and their respective area under the curve (AUC) values for predicting incident rHOA (KL ≥ 2 or THR) among subjects with KL1 grade at baseline using 1) texture variables from fractal signature analysis (FSA), 2) covariates (age, gender, body mass index), and 3) texture variables combined with covariates.

Supplementary Figure 8. Receiver operating characteristics (ROC) curves and their respective area under the curve (AUC) values for predicting incident joint space narrowing (JSN) among subjects with JSN grade 0 at baseline using 1) texture variables from fractal signature analysis (FSA), 2) covariates (age, gender, body mass index), and 3) texture variables combined with covariates.

Supplementary Figure 9. Receiver operating characteristics (ROC) curves and their respective area under the curve (AUC) values for predicting incident osteophytes (OST) among subjects with OST grade 0 at baseline using 1) texture variables from fractal signature analysis (FSA), 2) covariates (age, gender, body mass index), and 3) texture variables combined with covariates.

Tables

Table 1. Characteristics of the subjects that were included in the current study.

Variable	Baseline	10-year follow-up	Incident OA (KL ≥ 2)		Incident JSN (JSN grade ≥ 1)		Incident OST (OST grade ≥ 1)	
			Controls (*n* = 552)	OA (*n* = 435)	Controls (*n* = 196)	JSN (*n* = 471)	Controls (*n* = 87)	OST (*n* = 526)
Anthropometric variables								
Age [years]	55.7 (5.2)		55.0 (5.2)	56.5 (5.1)	55.1 (5.0)	56.0 (5.1)	55.3 (5.6)	55.2 (5.1)
Body mass index [kg/m^2]	25.9 (3.8)		25.9 (3.9)	25.9 (3.6)	25.5 (3.2)	25.9 (3.8)	25.5 (3.5)	26.0 (3.8)
Gender:								
Male	160 (16.2%)		71 (12.9%)	89 (20.5%)	23 (11.7%)	67 (14.2%)	9 (10.3%)	74 (14.1%)
Female	827 (83.8%)		481 (87.1%)	346 (79.5%)	173 (88.3%)	404 (85.8%)	78 (89.7%)	452 (85.9%)
KL grade distribution:								
KL0	720 (72.9%)	131 (13.3%)						
KL1	267 (27.1%)	421 (42.7%)						
KL2		389 (39.4%)						
KL3		15 (1.5%)						
KL4		-						
Total hip replacement		31 (3.1%)						
JSN grade distribution:								
0	667 (100%)	196 (29.4%)						
1		438 (65.7%)						
2		21 (3.1%)						
3		2 (0.3%)						
Total hip replacement		10 (1.5%)						
OST grade distribution:								
0	613 (100%)	87 (14.2%)						
1		348 (56.8%)						
2		167 (27.2%)						
3		5 (0.8%)						
Total hip replacement		6 (1.0%)						

KL = Kellgren-Lawrence, JSN = joint space narrowing, OST = osteophyte. All values are given as mean (standard deviation) or *n* (%).

Table 2. Odds ratios (95% confidence interval) of the univariable covariate models to assess incident rHOA (KL ≥ 2 or THR).

Predictor	Odds ratio
Age (years)	1.06 (1.03 – 1.09)*
Female gender	0.54 (0.37 – 0.80)*
Baseline KL grade	3.80 (2.79 – 5.16)*
Body mass index (kg/m^2)	1.00 (0.96 – 1.04)

*$p < 0.05$

Table 3. Odds ratios (95% confidence interval) of univariable texture models to assess incidence of rHOA (KL ≥ 2 or THR).

ROI	FD_{min}	FD_{max}	$Angles_{min}$	$Angles_{max}$
1	0.84 (0.74 – 0.96)*	0.88 (0.78 – 1.01)	1.02 (1.00 – 1.05)*	0.93 (0.90 – 0.97)*
2	0.84 (0.73 – 0.96)*	0.88 (0.76 – 1.01)	1.00 (0.97 – 1.04)	0.98 (0.95 – 1.01)
3	0.96 (0.85 – 1.09)	0.88 (0.78 – 0.99)*	0.98 (0.94 – 1.02)	1.02 (0.99 – 1.05)
4	1.01 (0.89 – 1.15)	0.97 (0.86 – 1.10)	0.97 (0.94 – 1.00)*	1.00 (0.97 – 1.03)
5	0.99 (0.87 – 1.13)	1.00 (0.89 – 1.13)	0.97 (0.94 – 0.99)*	0.99 (0.95 – 1.03)
6	0.83 (0.73 – 0.95)*	0.87 (0.77 – 0.99)*	0.99 (0.97 – 1.02)	0.99 (0.96 – 1.02)
7	0.95 (0.84 – 1.08)	1.00 (0.88 – 1.13)	0.99 (0.97 – 1.01)	1.00 (0.96 – 1.03)
8	1.02 (0.89 – 1.16)	1.04 (0.92 – 1.18)	1.00 (0.98 – 1.02)	0.98 (0.95 – 1.01)
9	1.01 (0.89 – 1.14)	1.05 (0.93 – 1.19)	0.99 (0.97 – 1.02)	0.98 (0.96 – 1.00)
10	0.81 (0.72 – 0.93)*	0.88 (0.78 – 1.01)	1.01 (0.98 – 1.05)	0.96 (0.94 – 0.98)*
11	0.89 (0.79 – 1.01)	0.95 (0.84 – 1.07)	1.01 (0.98 – 1.04)	0.95 (0.93 – 0.98)*
12	0.97 (0.85 – 1.09)	1.04 (0.92 – 1.18)	1.02 (0.99 – 1.04)	0.95 (0.92 – 0.98)*
13	0.94 (0.82 – 1.07)	1.01 (0.89 – 1.15)	1.01 (0.99 – 1.03)	0.98 (0.94 – 1.01)
14	1.03 (0.91 – 1.17)	1.11 (0.98 – 1.26)	1.02 (1.00 – 1.03)	0.96 (0.93 – 0.99)*
15	0.97 (0.85 – 1.10)	0.96 (0.85 – 1.09)	0.99 (0.97 – 1.01)	1.02 (0.98 – 1.05)
16	1.06 (0.93 – 1.20)	0.92 (0.81 – 1.03)	0.96 (0.93 – 0.98)*	0.97 (0.93 – 1.02)
17	1.03 (0.90 – 1.17)	1.03 (0.91 – 1.17)	1.00 (0.97 – 1.02)	0.93 (0.90 – 0.97)*
18	0.88 (0.77 – 1.00)	0.92 (0.81 – 1.04)	0.97 (0.94 – 1.01)	0.98 (0.96 – 1.01)
19	0.81 (0.71 – 0.93)*	0.87 (0.76 – 0.99)*	1.02 (0.99 – 1.06)	0.99 (0.96 – 1.02)
20	0.92 (0.81 – 1.04)	1.00 (0.88 – 1.12)	1.01 (0.98 – 1.04)	0.98 (0.96 – 1.00)
21	1.03 (0.91 – 1.17)	0.97 (0.85 – 1.10)	1.00 (0.96 – 1.04)	0.98 (0.96 – 1.01)
22	0.93 (0.83 – 1.05)	0.88 (0.77 – 0.99)*	0.97 (0.93 – 1.01)	0.97 (0.94 – 0.99)*
23	0.88 (0.78 – 1.00)	0.92 (0.82 – 1.04)	1.03 (0.99 – 1.06)	0.95 (0.92 – 0.98)*
24	1.05 (0.92 – 1.19)	1.09 (0.96 – 1.24)	1.01 (0.99 – 1.04)	0.93 (0.90 – 0.97)*
25	0.98 (0.87 – 1.12)	1.01 (0.90 – 1.14)	1.00 (0.94 – 1.05)	0.95 (0.92 – 0.99)*
26	0.96 (0.84 – 1.09)	1.01 (0.90 – 1.15)	1.03 (1.00 – 1.06)	1.03 (0.99 – 1.07)
27	0.79 (0.70 – 0.90)*	0.85 (0.75 – 0.96)*	1.00 (0.97 – 1.03)	0.99 (0.97 – 1.00)
28	0.78 (0.69 – 0.90)*	0.89 (0.78 – 1.02)	1.03 (1.00 – 1.06)	0.99 (0.98 – 1.01)
29	0.78 (0.68 – 0.89)*	0.95 (0.84 – 1.08)	0.98 (0.95 – 1.01)	1.01 (0.99 – 1.04)
30	0.84 (0.75 – 0.95)*	0.98 (0.87 – 1.11)	1.00 (0.97 – 1.03)	0.99 (0.96 – 1.02)
31	1.00 (0.88 – 1.13)	0.97 (0.85 – 1.10)	0.99 (0.96 – 1.01)	0.99 (0.96 – 1.02)
32	1.13 (0.99 – 1.30)	0.92 (0.81 – 1.04)	0.99 (0.97 – 1.01)	1.05 (1.00 – 1.09)*
33	1.14 (1.00 – 1.30)	1.07 (0.94 – 1.22)	0.98 (0.96 – 1.01)	0.96 (0.90 – 1.03)
34	1.09 (0.96 – 1.24)	1.12 (0.98 – 1.28)	1.00 (0.98 – 1.02)	1.00 (0.89 – 1.14)
35	1.00 (0.88 – 1.13)	1.02 (0.90 – 1.16)	1.02 (1.00 – 1.04)	1.01 (0.90 – 1.13)
36	1.01 (0.89 – 1.15)	0.95 (0.84 – 1.08)	1.02 (1.00 – 1.03)	0.93 (0.81 – 1.07)
37	1.02 (0.90 – 1.15)	1.03 (0.91 – 1.17)	1.01 (0.99 – 1.03)	0.92 (0.80 – 1.05)
38	1.00 (0.87 – 1.14)	1.09 (0.96 – 1.24)	0.99 (0.97 – 1.01)	0.99 (0.93 – 1.06)
39	1.15 (1.01 – 1.32)*	1.07 (0.94 – 1.22)	0.99 (0.97 – 1.01)	0.99 (0.95 – 1.03)
40	0.92 (0.81 – 1.04)	0.86 (0.76 – 0.98)*	0.98 (0.96 – 0.99)*	1.04 (1.01 – 1.07)*
41	0.71 (0.63 – 0.81)*	0.83 (0.73 – 0.94)*	0.98 (0.96 – 1.00)	1.01 (0.98 – 1.03)

*$p < 0.05$

Table 4. Selected λ and α parameters for the elastic net from 10-fold cross-validation and areas under the receiver operating characteristics curve (ROC AUC) (95% confidence interval) for predicting incidence of rHOA (KL ≥ 2 or THR) with the covariate, texture, covariate + baseline KL, and texture + covariate + baseline KL models.

Model	Selected λ	Selected α	ROC AUC in validation
Covariates (age, gender, body mass index)	0.096	0.15	0.61 (0.58 – 0.64)
Texture features	0.059	0.10	0.68 (0.65 – 0.71)
Covariates + baseline KL	0.001	0.30	0.69 (0.64 – 0.74)
Texture features + covariates + baseline KL	0.153	0.10	0.73 (0.70 – 0.76)

Figure 1.

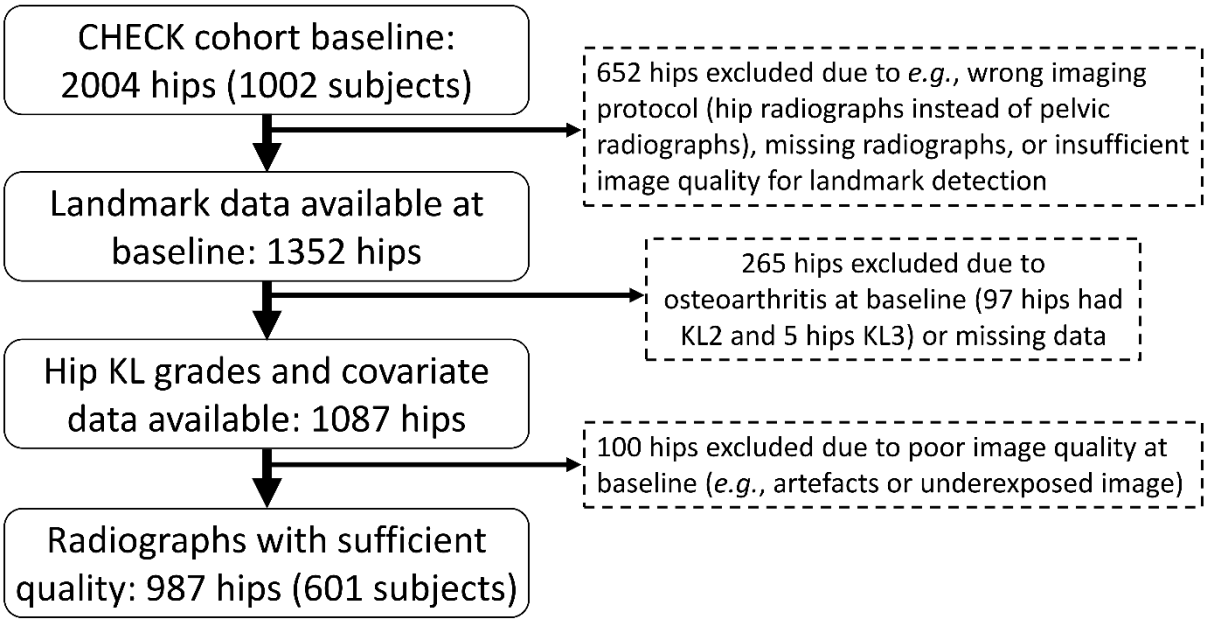

CHECK cohort baseline: 2004 hips (1002 subjects)
652 hips excluded due to *e.g.*, wrong imaging protocol (hip radiographs instead of pelvic radiographs), missing radiographs, or insufficient image quality for landmark detection
Landmark data available at baseline: 1352 hips
265 hips excluded due to osteoarthritis at baseline (97 hips had KL2 and 5 hips KL3) or missing data
Hip KL grades and covariate data available: 1087 hips
100 hips excluded due to poor image quality at baseline (*e.g.*, artefacts or underexposed image)
Radiographs with sufficient quality: 987 hips (601 subjects)

Figure 2.

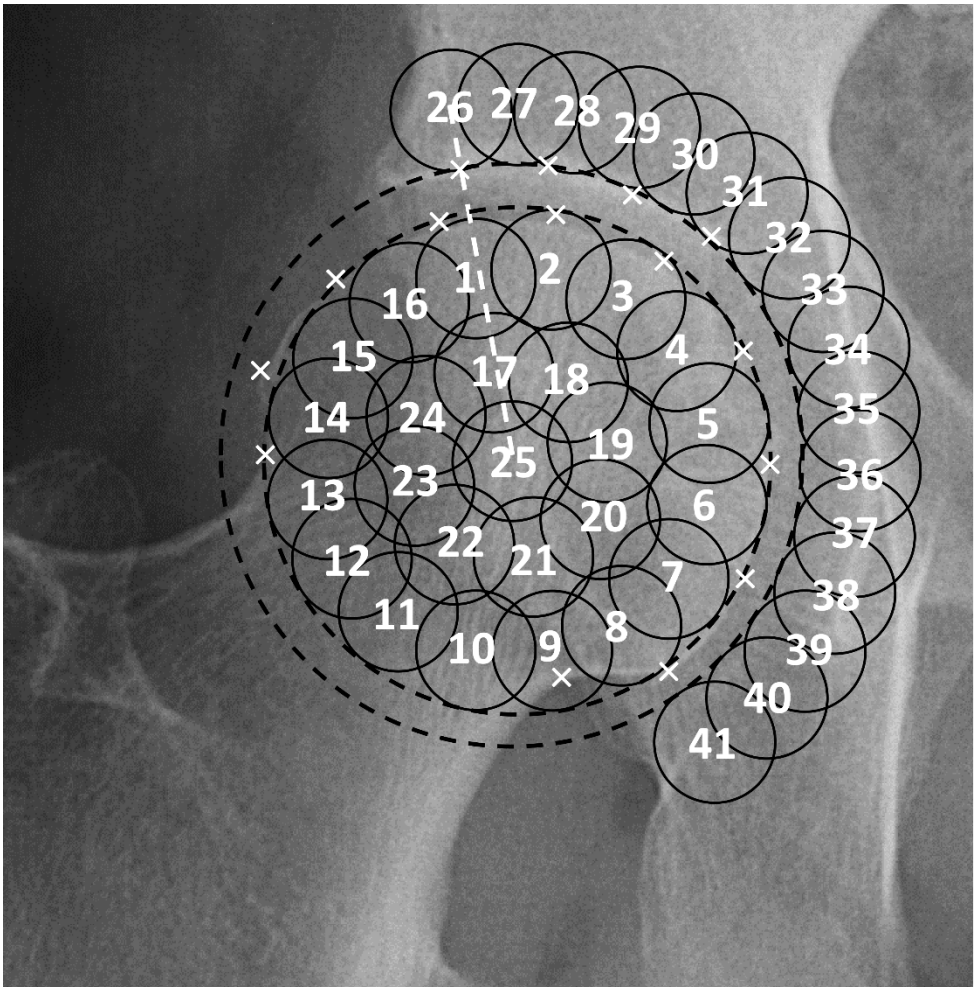

26
27
28
29
30
31
32
1
2
3
33
16
15
4
34
17
18
14
24
5
35
19
25
13
23
36
6
20
22
37
12
21
7
38
11
8
10
9
39
40
41

Figure 3.

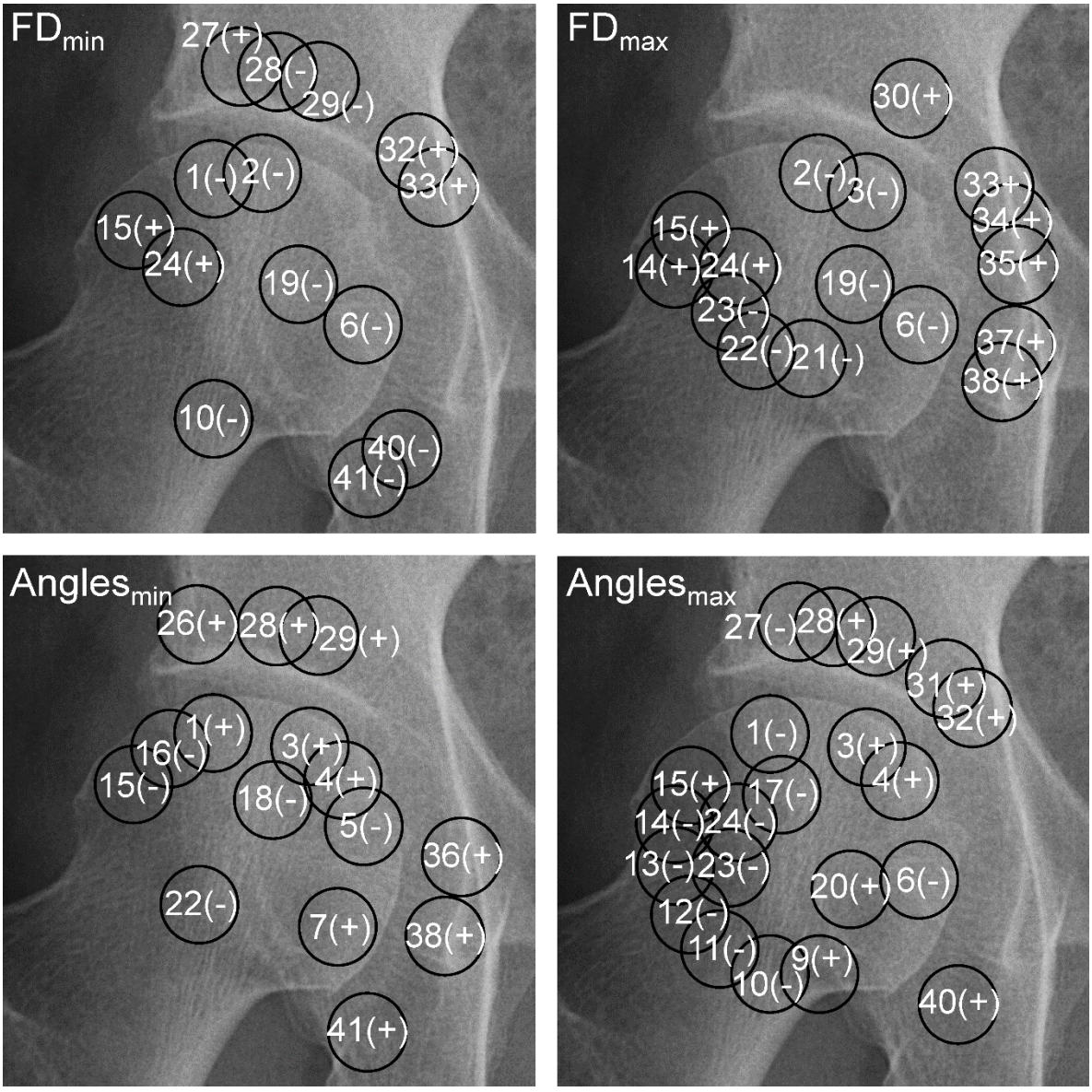
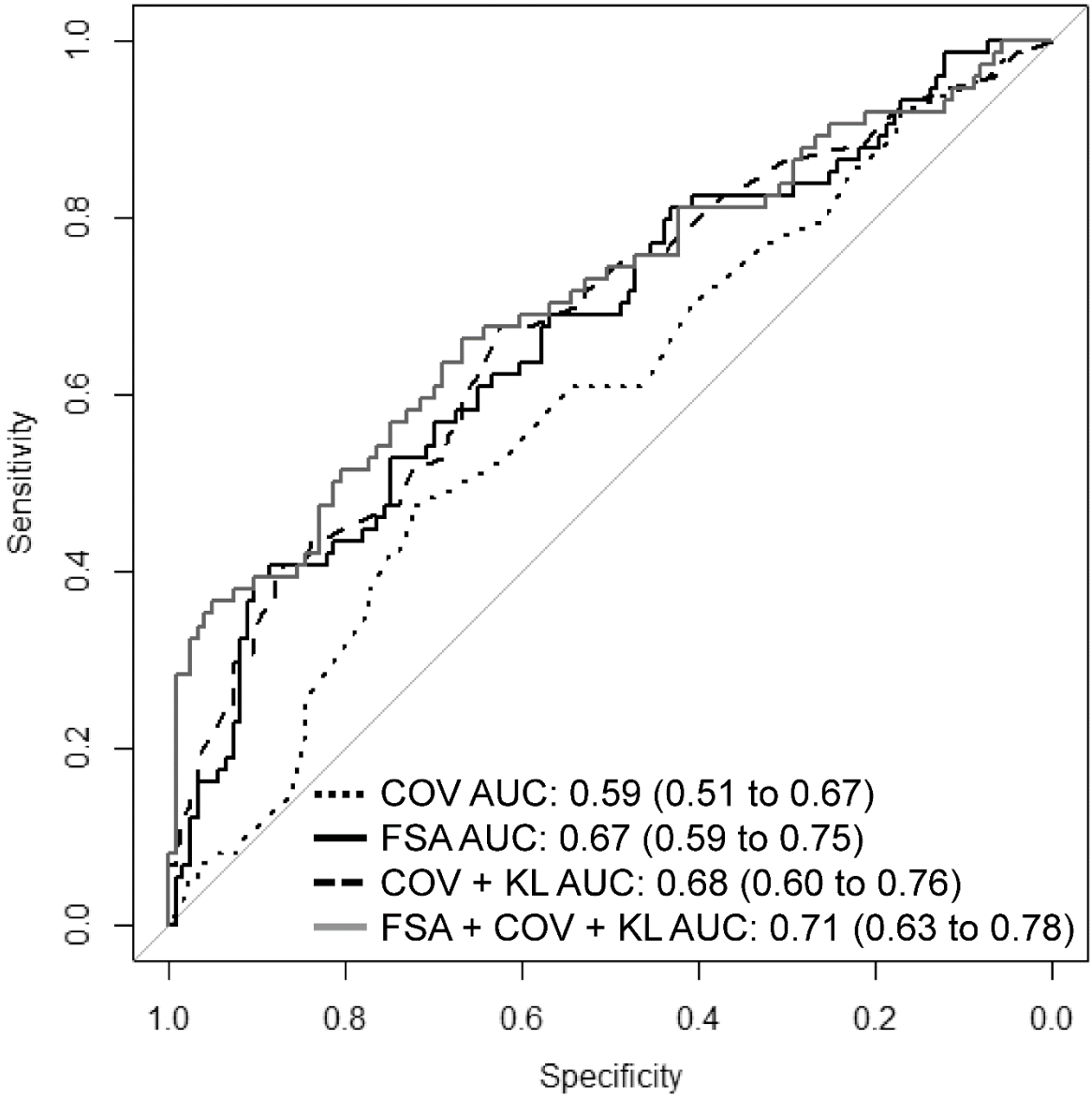
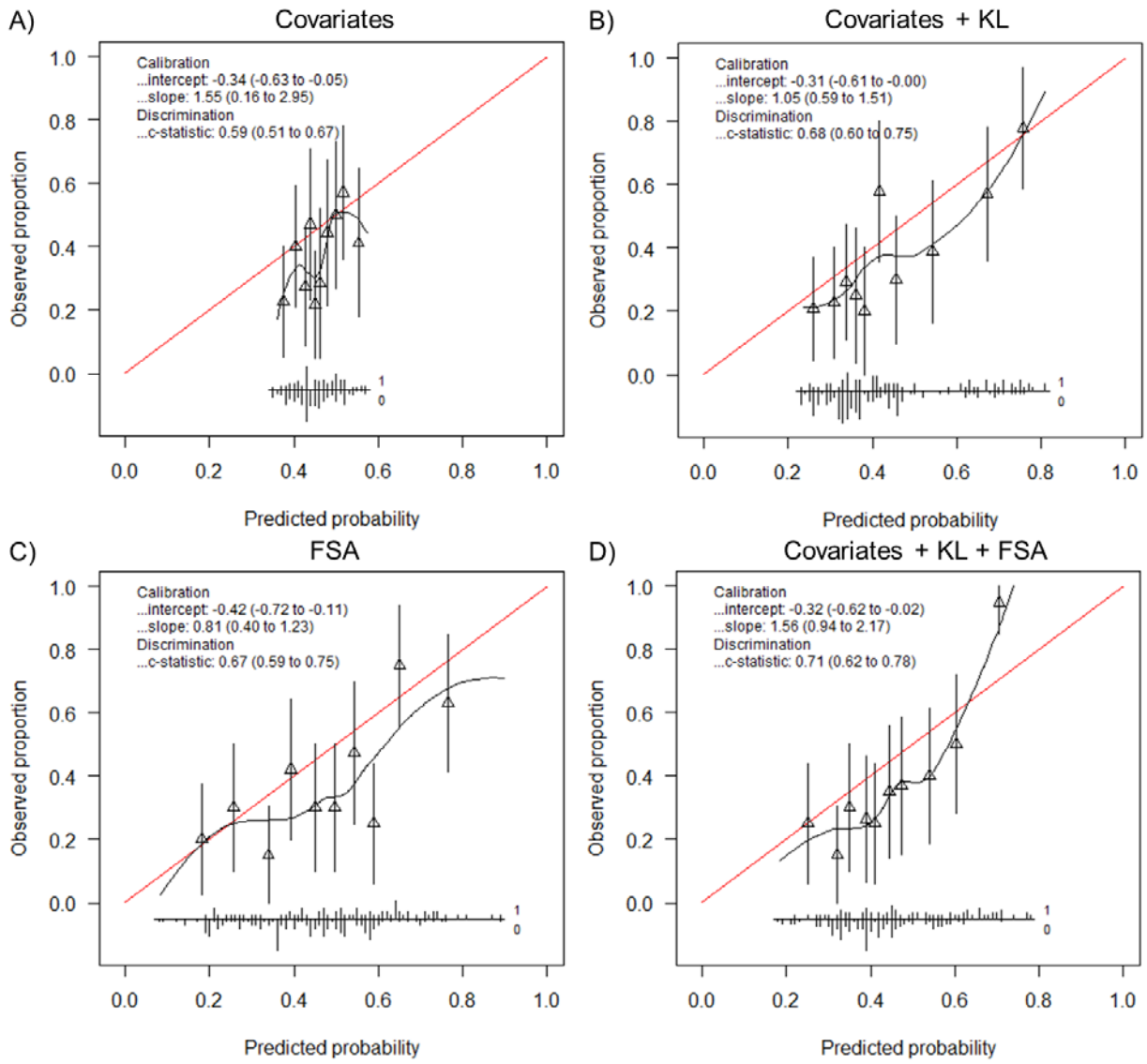

FDmin
27(+)
28(-)
29(-)
32(+)
1(-)
2(-)
33(+)
15(+)
24(+)
19(-)
6(-)
10(-)
40(-)
41(-)
FDmax
30(+)
2(-)
3(-)
33+)
34(+)
15(+)
14(+)
24(+)
19(-)
35(+)
23(-)
6(-)
37(+)
22(-)
21(-)
38(+)
Anglesmin
26(+)
28(+)
29(+)
1(+)
16(-)
3(+)
15(-)
4(+)
18(-)
5(-)
36(+)
22(-)
7(+)
38(+)
41(+)
Anglesmax
27(-)
28(+)
29(+)
31(+)
32(+)
1(-)
3(+)
15(+)
17(-)
4(+)
14(-)
24(-)
13(-)
23(-)
20(+)
6(-)
12(-)
11(-)
9(+)
10(-)
40(+)

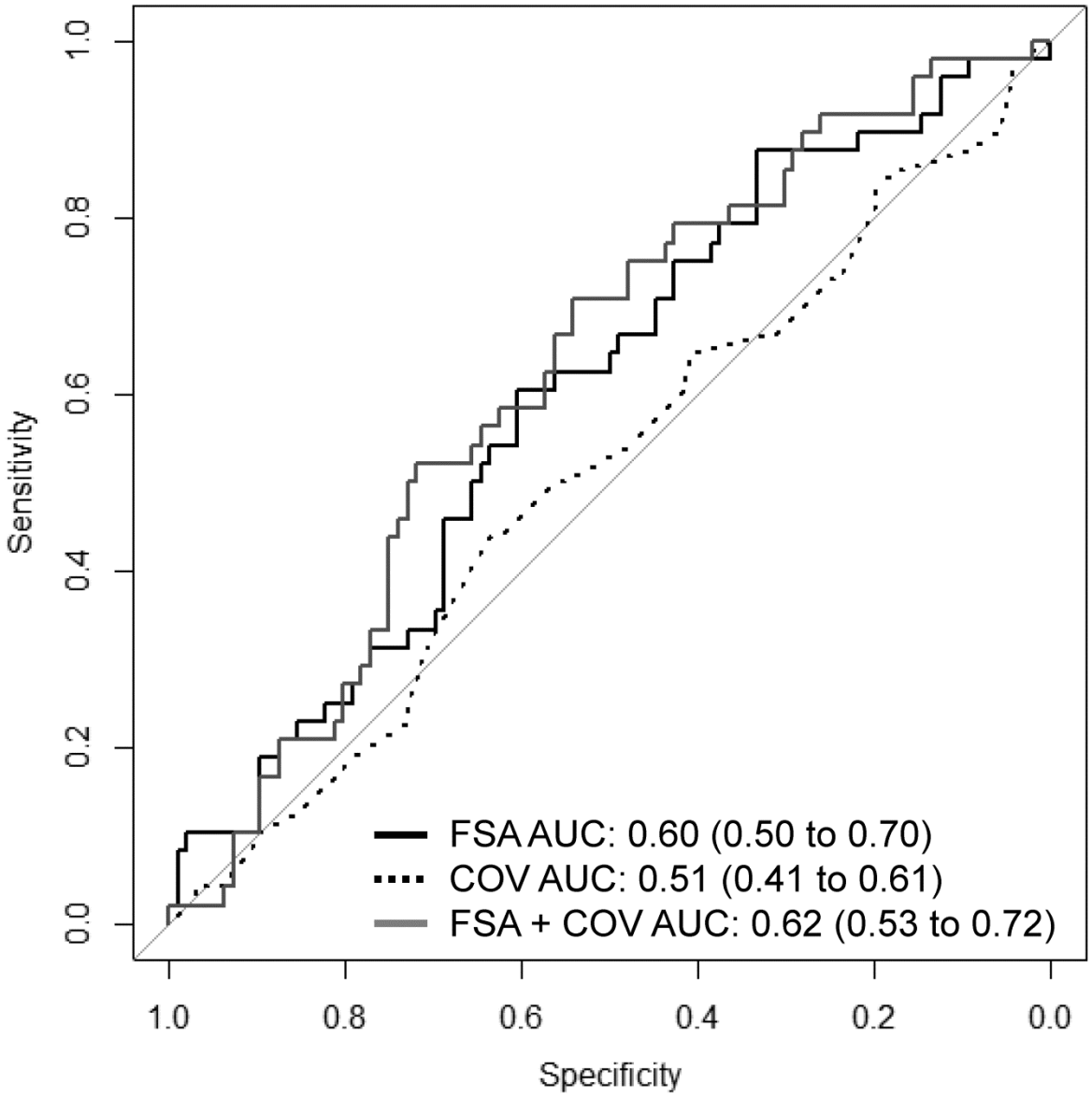
Figure 4.


Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
1.0
0.8
0.6
0.4
0.2
0.0
Specificity
COV AUC: 0.59 (0.51 to 0.67)
FSA AUC: 0.67 (0.59 to 0.75)
COV + KL AUC: 0.68 (0.60 to 0.76)
FSA + COV + KL AUC: 0.71 (0.63 to 0.78)

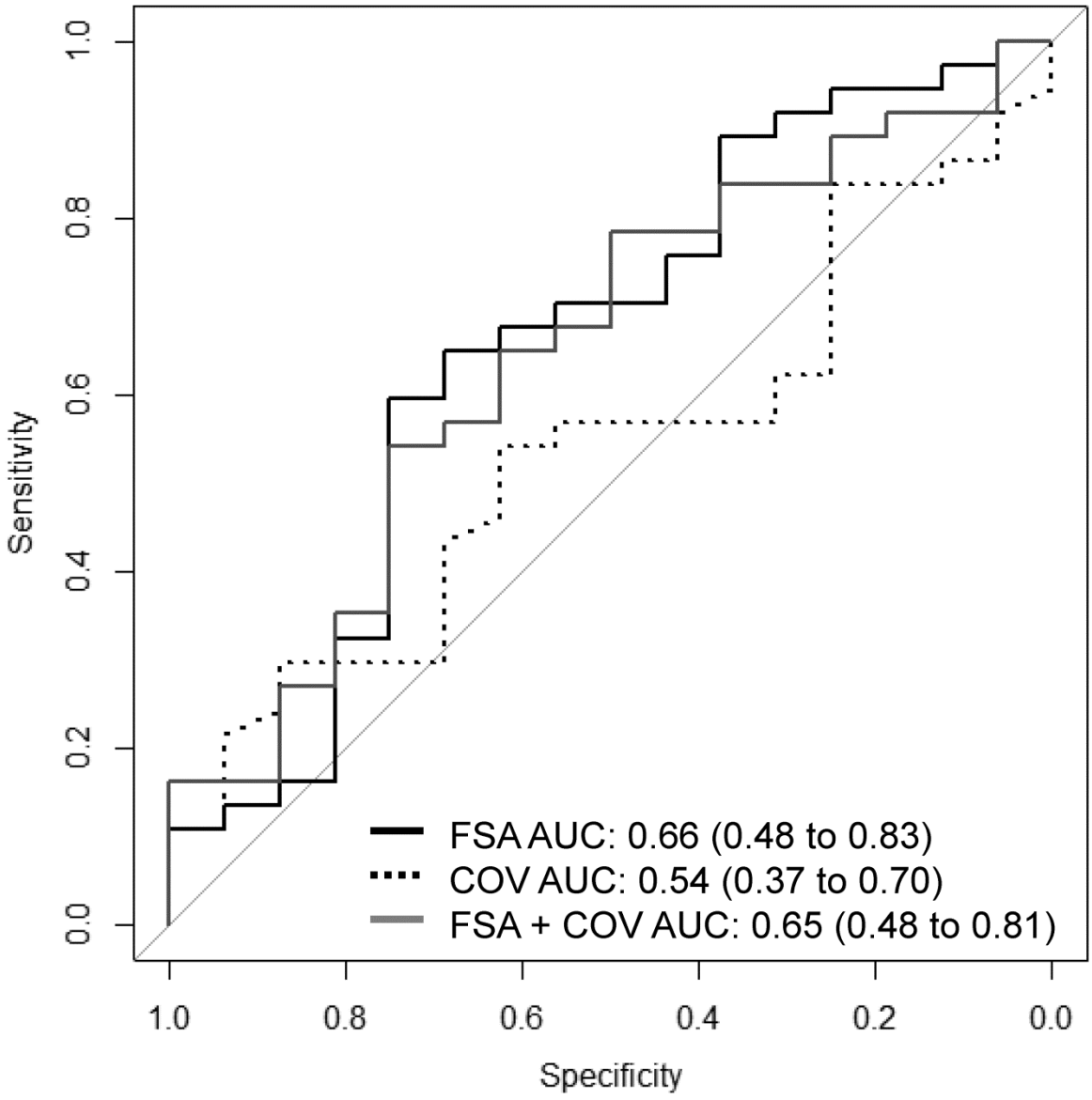
Supplementary Figure 5.



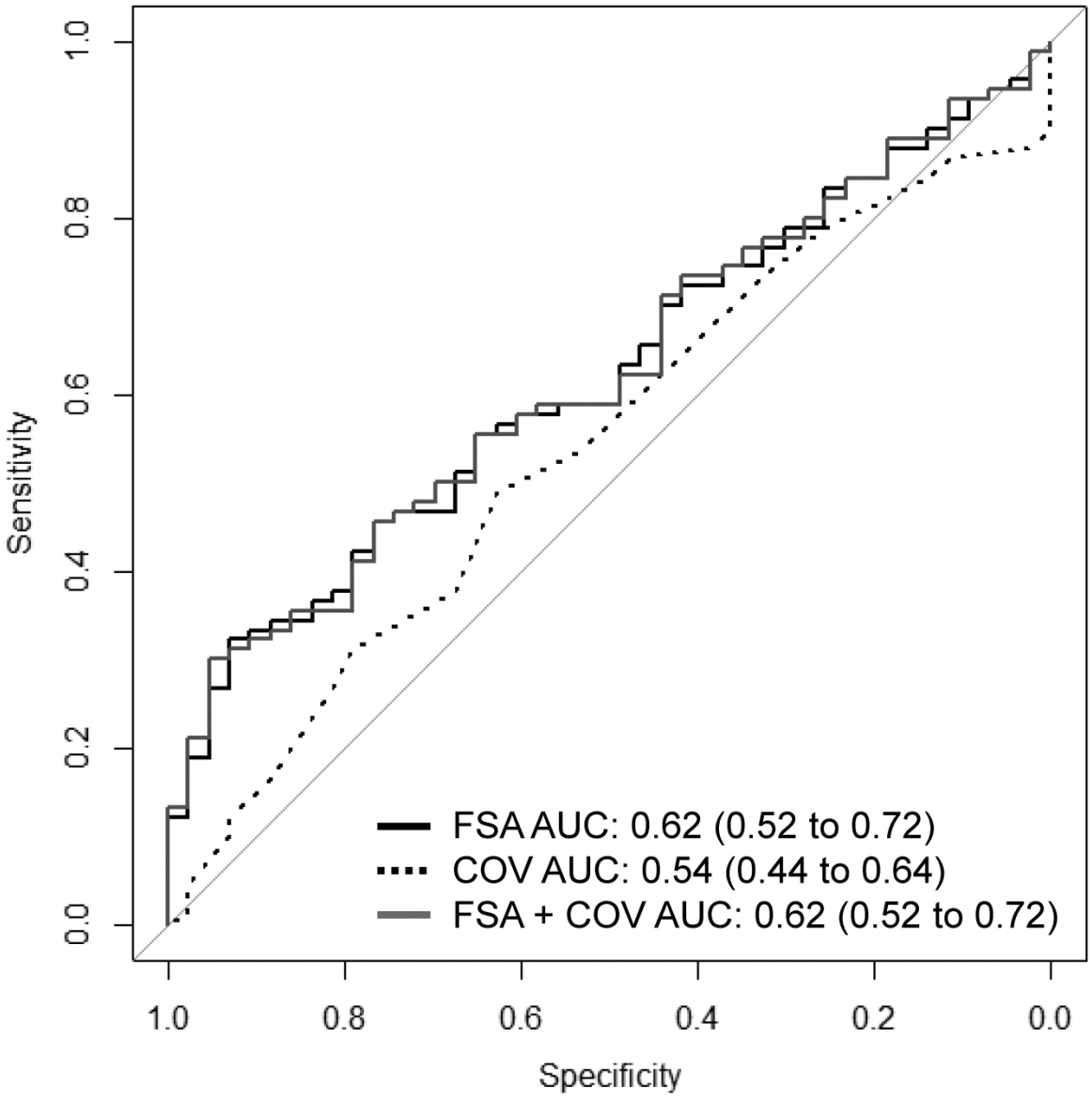
Supplementary Figure 6.



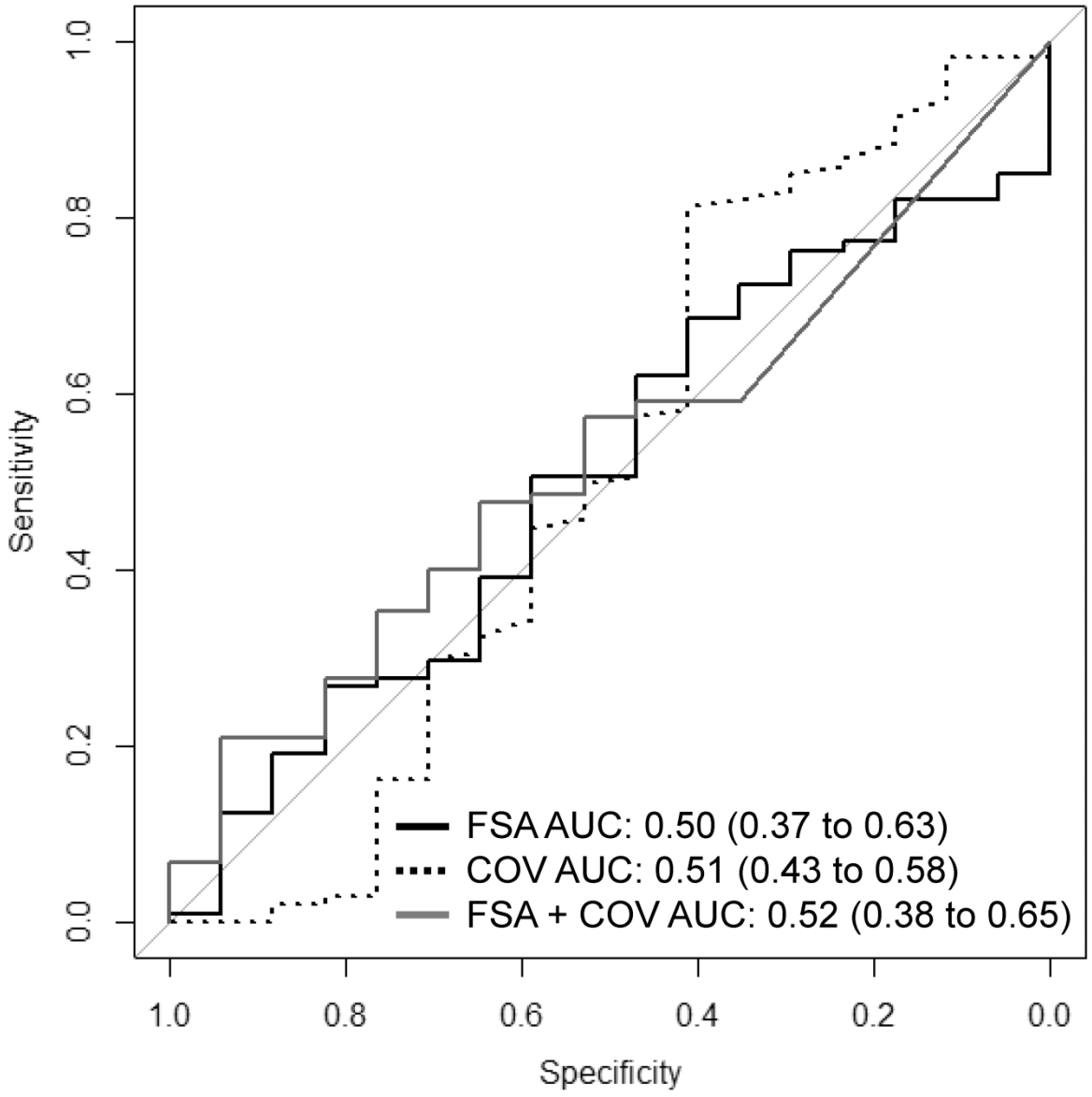
Supplementary Figure 7.



Supplementary Figure 8.



Supplementary Figure 9.



Supplementary Tables

Supplementary Table 5. Variables in the final elastic net model to predict incident rHOA (KL ≥ 2 or THR).

Variable	Coefficient	Variable	Coefficient
Intercept	-0.47	$Angles_{min}$ ROI7	0.00
FD_{min} ROI1	-0.04	$Angles_{min}$ ROI15	-0.01
FD_{min} ROI2	-0.04	$Angles_{min}$ ROI16	-0.01
FD_{min} ROI6	-0.05	$Angles_{min}$ ROI18	-0.01
FD_{min} ROI10	-0.03	$Angles_{min}$ ROI22	-0.01
FD_{min} ROI15	0.01	$Angles_{min}$ ROI26	0.00
FD_{min} ROI19	-0.08	$Angles_{min}$ ROI28	0.01
FD_{min} ROI24	0.01	$Angles_{min}$ ROI29	0.00
FD_{min} ROI27	0.00	$Angles_{min}$ ROI36	0.00
FD_{min} ROI28	-0.03	$Angles_{min}$ ROI38	0.00
FD_{min} ROI29	-0.05	$Angles_{min}$ ROI41	0.00
FD_{min} ROI32	0.03	$Angles_{max}$ ROI1	-0.02
FD_{min} ROI33	0.01	$Angles_{max}$ ROI3	0.01
FD_{min} ROI40	-0.02	$Angles_{max}$ ROI4	0.00
FD_{min} ROI41	-0.08	$Angles_{max}$ ROI6	-0.01
FD_{max} ROI2	-0.05	$Angles_{max}$ ROI9	0.00
FD_{max} ROI3	-0.04	$Angles_{max}$ ROI10	-0.01
FD_{max} ROI6	-0.03	$Angles_{max}$ ROI11	-0.01
FD_{max} ROI14	0.05	$Angles_{max}$ ROI12	-0.02
FD_{max} ROI15	0.01	$Angles_{max}$ ROI14	-0.01
FD_{max} ROI19	-0.03	$Angles_{max}$ ROI15	0.00
FD_{max} ROI21	-0.01	$Angles_{max}$ ROI17	-0.02
FD_{max} ROI22	-0.09	$Angles_{max}$ ROI20	0.00
FD_{max} ROI23	-0.04	$Angles_{max}$ ROI23	-0.01
FD_{max} ROI24	0.04	$Angles_{max}$ ROI24	-0.01
FD_{max} ROI30	0.01	$Angles_{max}$ ROI27	-0.01
FD_{max} ROI33	0.05	$Angles_{max}$ ROI28	0.00
FD_{max} ROI34	0.06	$Angles_{max}$ ROI29	0.00
FD_{max} ROI35	0.04	$Angles_{max}$ ROI31	0.00
FD_{max} ROI37	0.04	$Angles_{max}$ ROI32	0.00
FD_{max} ROI38	0.07	$Angles_{max}$ ROI40	0.01
$Angles_{min}$ ROI1	0.01	Age	0.03
$Angles_{min}$ ROI3	0.00	Baseline KL	0.69
$Angles_{min}$ ROI4	0.00	Gender	-0.22
$Angles_{min}$ ROI5	-0.01		

Supplementary Table 6. Odds ratios (95% confidence interval) of univariable covariate models to predict incident joint space narrowing (JSN grade ≥ 1 or THR).

Predictor	Odds ratio
Age (years)	1.04 (0.99 – 1.08)
Female gender	0.60 (0.18 – 2.04)
Body mass index (kg/m^2)	1.03 (0.98 – 1.08)

Supplementary Table 7. Odds ratios (95% confidence interval) of univariable texture models to predict incident joint space narrowing (JSN grade ≥ 1 or THR).

ROI	FD_{min}	FD_{max}	$Angles_{min}$	$Angles_{max}$
1	0.99 (0.71 – 1.38)	0.95 (0.72 – 1.27)	1.03 (0.98 – 1.09)	1.04 (0.93 – 1.17)
2	0.83 (0.60 – 1.16)	0.88 (0.68 – 1.14)	0.96 (0.90 – 1.02)	1.00 (0.92 – 1.09)
3	0.83 (0.60 – 1.16)	0.64 (0.52 – 0.79)*	0.93 (0.86 – 1.02)	1.02 (0.95 – 1.10)
4	1.05 (0.74 – 1.50)	0.87 (0.62 – 1.21)	1.02 (0.92 – 1.12)	0.96 (0.88 – 1.04)
5	1.00 (0.74 – 1.35)	0.81 (0.63 – 1.04)	0.96 (0.87 – 1.04)	1.01 (0.91 – 1.12)
6	0.78 (0.59 – 1.02)	0.74 (0.57 – 0.98)*	1.01 (0.95 – 1.06)	1.01 (0.91 – 1.13)
7	0.93 (0.65 – 1.33)	0.83 (0.62 – 1.10)	1.02 (0.97 – 1.06)	0.98 (0.89 – 1.07)
8	1.27 (0.91 – 1.78)	0.97 (0.66 – 1.41)	0.98 (0.94 – 1.02)	0.95 (0.87 – 1.04)
9	0.95 (0.70 – 1.30)	1.20 (0.88 – 1.65)	0.99 (0.93 – 1.06)	1.03 (0.97 – 1.08)
10	0.92 (0.65 – 1.29)	0.84 (0.59 – 1.20)	1.05 (0.96 – 1.15)	0.98 (0.93 – 1.02)
11	1.03 (0.69 – 1.54)	1.01 (0.75 – 1.36)	0.97 (0.90 – 1.05)	0.96 (0.89 – 1.05)
12	0.97 (0.72 – 1.30)	1.04 (0.81 – 1.33)	0.97 (0.91 – 1.03)	0.98 (0.90 – 1.08)
13	0.87 (0.68 – 1.1)	0.95 (0.73 – 1.25)	1.00 (0.95 – 1.04)	1.16 (1.02 – 1.32)*
14	1.19 (0.81 – 1.73)	1.29 (0.92 – 1.82)	1.04 (0.98 – 1.10)	1.01 (0.91 – 1.11)
15	1.43 (1.03 – 2.00)*	1.05 (0.78 – 1.41)	0.94 (0.90 – 0.99)*	0.94 (0.84 – 1.05)
16	1.06 (0.73 – 1.52)	0.85 (0.63 – 1.15)	1.01 (0.95 – 1.07)	0.88 (0.81 – 0.96)*
17	0.99 (0.67 – 1.45)	0.81 (0.57 – 1.14)	1.05 (0.98 – 1.13)	0.95 (0.89 – 1.01)
18	1.03 (0.72 – 1.48)	0.92 (0.70 – 1.19)	0.97 (0.89 – 1.07)	1.03 (0.97 – 1.09)
19	1.02 (0.71 – 1.45)	0.86 (0.67 – 1.09)	0.93 (0.86 – 1.02)	1.00 (0.91 – 1.09)
20	1.18 (0.87 – 1.59)	1.13 (0.7 – 1.46)	1.01 (0.91 – 1.11)	1.04 (0.98 – 1.10)
21	0.81 (0.61 – 1.08)	0.96 (0.66 – 1.38)	0.96 (0.87 – 1.06)	0.97 (0.91 – 1.04)
22	0.85 (0.57 – 1.27)	0.85 (0.55 – 1.31)	1.00 (0.91 – 1.10)	0.99 (0.91 – 1.08)
23	0.86 (0.60 – 1.24)	0.83 (0.62 – 1.12)	1.01 (0.94 – 1.09)	0.96 (0.88 – 1.05)
24	0.97 (0.67 – 1.40)	0.83 (0.73 – 1.29)	0.97 (0.91 – 1.03)	0.96 (0.88 – 1.05)
25	0.93 (0.67 – 1.30)	0.74 (0.58 – 0.93)*	0.94 (0.78 – 1.13)	0.98 (0.87 – 1.11)
26	1.22 (0.91 – 1.64)	1.06 (0.78 – 1.45)	1.03 (0.97 – 1.10)	1.07 (0.98 – 1.17)
27	0.98 (0.70 – 1.39)	0.93 (0.73 – 1.41)	0.96 (0.87 – 1.05)	1.02 (0.97 – 1.07)
28	0.82 (0.60 – 1.12)	0.69 (0.52 – 0.92)*	1.00 (0.92 – 1.09)	1.01 (0.97 – 1.06)
29	0.84 (0.64 – 1.10)	0.80 (0.61 – 1.05)	0.95 (0.87 – 1.03)	1.00 (0.92 – 1.09)
30	0.85 (0.64 – 1.13)	0.87 (0.63 – 1.21)	0.99 (0.92 – 1.06)	1.01 (0.95 – 1.08)
31	0.67 (0.46 – 0.98)*	0.78 (0.57 – 1.08)	1.02 (0.95 – 1.10)	1.01 (0.94 – 1.10)
32	0.93 (0.68 – 1.26)	0.95 (0.69 – 1.30)	1.01 (0.95 – 1.08)	1.08 (0.96 – 1.22)
33	1.01 (0.74 – 1.39)	1.06 (0.76 – 1.47)	0.97 (0.92 – 1.03)	1.07 (0.90 – 1.27)
34	1.01 (0.68 – 1.52)	1.30 (0.97 – 1.76)	0.96 (0.91 – 1.01)	1.16 (0.81 – 1.67)
35	0.81 (0.54 – 1.22)	1.08 (0.74 – 1.58)	0.97 (0.93 – 1.02)	0.94 (0.70 – 1.27)
36	0.91 (0.61 – 1.35)	0.99 (0.71 – 1.39)	1.02 (0.97 – 1.08)	0.65 (0.42 – 1.01)
37	1.07 (0.74 – 1.54)	1.07 (0.70 – 1.65)	0.98 (0.93 – 1.03)	1.10 (0.75 – 1.61)
38	1.08 (0.75 – 1.57)	1.12 (0.75 – 1.68)	0.98 (0.92 – 1.03)	0.94 (0.82 – 1.06)
39	1.19 (0.85 – 1.66)	1.32 (0.94 – 1.85)	0.97 (0.91 – 1.04)	1.02 (0.90 – 1.15)
40	0.81 (0.58 – 1.12)	0.84 (0.58 – 1.22)	0.97 (0.92 – 1.03)	0.99 (0.92 – 1.08)
41	0.80 (0.57 – 1.14)	0.98 (0.67 – 1.43)	1.00 (0.95 – 1.06)	1.05 (0.99 – 1.11)

*$p < 0.05$

Supplementary Table 8. Odds ratios (95% confidence interval) of the univariable covariate models to assess incident osteophytes (OST grade ≥ 1 or THR).

Predictor	Odds ratio
Age (years)	1.00 (0.95 – 1.05)
Female gender	0.71 (0.32 – 1.55)
Body mass index (kg/m^2)	1.03 (0.97 – 1.11)

Supplementary Table 9. Odds ratios (95% confidence interval) of univariable texture models to assess incident osteophytes (OST grade ≥ 1 or THR).

ROI	FD_{min}	FD_{max}	$Angles_{min}$	$Angles_{max}$
1	1.02 (0.80 – 1.28)	0.88 (0.66 – 1.15)	0.97 (0.93 – 1.01)	1.04 (0.97 – 1.11)
2	1.09 (0.88 – 1.36)	0.81 (0.62 – 1.05)	1.01 (0.95 – 1.07)	1.02 (0.97 – 1.08)
3	1.06 (0.80 – 1.40)	0.90 (0.70 – 1.16)	0.97 (0.90 – 1.04)	1.00 (0.95 – 1.06)
4	1.10 (0.85 – 1.42)	0.99 (0.77 – 1.27)	1.00 (0.94 – 1.07)	1.03 (0.97 – 1.09)
5	0.94 (0.74 – 1.21)	1.00 (0.80 – 1.25)	0.99 (0.94 – 1.04)	1.02 (0.94 – 1.10)
6	0.84 (0.66 – 1.08)	0.82 (0.66 – 1.01)	1.02 (0.98 – 1.06)	1.03 (0.97 – 1.10)
7	0.99 (0.77 – 1.26)	0.87 (0.58 – 1.11)	0.97 (0.94 – 1.01)	1.04 (0.99 – 1.09)
8	0.88 (0.69 – 1.11)	0.87 (0.67 – 1.11)	1.01 (0.97 – 1.05)	0.98 (0.92 – 1.05)
9	1.00 (0.80 – 1.26)	1.03 (0.80 – 1.31)	0.98 (0.93 – 1.03)	0.99 (0.95 – 1.03)
10	0.80 (0.64 – 1.01)	0.93 (0.75 – 1.15)	0.99 (0.93 – 1.06)	0.97 (0.93 – 1.02)
11	0.90 (0.70 – 1.16)	0.89 (0.70 – 1.12)	1.05 (0.99 – 1.11)	0.99 (0.95 – 1.03)
12	0.96 (0.76 – 1.21)	0.99 (0.79 – 1.24)	0.98 (0.93 – 1.03)	1.00 (0.94 – 1.06)
13	0.92 (0.72 – 1.19)	1.12 (0.90 – 1.40)	1.00 (0.95 – 1.04)	0.99 (0.93 – 1.05)
14	0.97 (0.78 – 1.22)	1.00 (0.80 – 1.24)	1.02 (0.99 – 1.06)	1.03 (0.95 – 1.11)
15	0.92 (0.74 – 1.14)	0.95 (0.74 – 1.22)	1.00 (0.96 – 1.03)	1.07 (0.99 – 1.16)
16	0.96 (0.76 – 1.21)	0.97 (0.75 – 1.24)	0.99 (0.95 – 1.03)	1.05 (0.95 – 1.16)
17	0.83 (0.66 – 1.04)	0.84 (0.69 – 1.04)	1.00 (0.95 – 1.05)	1.00 (0.94 – 1.06)
18	0.97 (0.76 – 1.23)	0.86 (0.69 – 1.08)	0.99 (0.92 – 1.07)	1.02 (0.96 – 1.08)
19	0.87 (0.68 – 1.11)	0.87 (0.69 – 1.09)	1.08 (1.02 – 1.16)*	1.01 (0.96 – 1.07)
20	0.84 (0.66 – 1.06)	0.88 (0.69 – 1.10)	1.00 (0.94 – 1.06)	1.04 (0.99 – 1.09)
21	0.82 (0.66 – 1.04)	0.95 (0.75 – 1.20)	0.93 (0.85 – 1.01)	1.02 (0.96 – 1.08)
22	0.85 (0.67 – 1.08)	0.84 (0.65 – 1.10)	0.97 (0.89 – 1.06)	0.98 (0.93 – 1.04)
23	0.78 (0.62 – 0.99)*	0.89 (0.67 – 1.19)	1.02 (0.97 – 1.08)	1.03 (0.97 – 1.10)
24	0.98 (0.78 – 1.22)	1.08 (0.85 – 1.38)	1.00 (0.95 – 1.05)	1.03 (0.96 – 1.10)
25	1.01 (0.81 – 1.25)	1.04 (0.83 – 1.31)	0.89 (0.79 – 1.00)*	0.99 (0.93 – 1.06)
26	0.99 (0.78 – 1.25)	0.90 (0.71 – 1.15)	1.04 (0.97 – 1.11)	0.98 (0.92 – 1.05)
27	0.93 (0.74 – 1.16)	1.02 (0.81 – 1.28)	1.05 (0.97 – 1.14)	0.98 (0.95 – 1.02)
28	0.86 (0.69 – 1.07)	0.87 (0.70 – 1.10)	1.01 (0.95 – 1.07)	1.01 (0.98 – 1.05)
29	0.91 (0.73 – 1.15)	0.95 (0.76 – 1.20)	0.98 (0.93 – 1.03)	0.99 (0.95 – 1.04)
30	0.95 (0.74 – 1.20)	1.05 (0.82 – 1.34)	0.99 (0.94 – 1.04)	0.99 (0.94 – 1.05)
31	0.88 (0.71 – 1.11)	0.83 (0.65 – 1.06)	1.00 (0.96 – 1.04)	1.01 (0.95 – 1.06)
32	1.05 (0.82 – 1.35)	0.88 (0.71 – 1.08)	0.99 (0.94 – 1.04)	0.99 (0.91 – 1.08)
33	1.01 (0.79 – 1.29)	1.10 (0.86 – 1.42)	1.02 (0.99 – 1.06)	0.99 (0.80 – 1.22)
34	1.10 (0.84 – 1.45)	1.10 (0.87 – 1.41)	0.99 (0.96 – 1.03)	1.08 (0.92 – 1.27)
35	1.00 (0.79 – 1.27)	1.10 (0.84 – 1.45)	0.99 (0.96 – 1.03)	0.96 (0.77 – 1.19)
36	1.01 (0.82 – 1.24)	0.95 (0.75 – 1.21)	1.06 (1.02 – 1.11)*	0.91 (0.73 – 1.13)
37	1.09 (0.86 – 1.39)	0.79 (0.62 – 0.99)*	1.02 (0.99 – 1.06)	0.85 (0.58– 1.23)
38	1.13 (0.86 – 1.48)	1.05 (0.82 – 1.34)	1.01 (0.98 – 1.05)	0.87 (0.71 – 1.05)
39	1.37 (0.97 – 1.93)	1.11 (0.87 – 1.42)	0.99 (0.95 – 1.04)	0.92 (0.83 – 1.01)
40	1.17 (0.90 – 1.52)	1.03 (0.82 – 1.28)	0.97 (0.92 – 1.03)	1.01 (0.96 – 1.06)
41	0.87 (0.67 – 1.13)	0.89 (0.69 – 1.15)	1.01 (0.96 – 1.05)	1.00 (0.95 – 1.04)

*$p < 0.05$