Arthritis Care & Research
Vol. 0, No. 0, Month 2026, pp 1–12
DOI 10.1002/acr.25629

American College of Rheumatology
Empowering Rheumatology Professionals

Hip Morphology–Based Osteoarthritis Risk Prediction Models: Development and External Validation Using Individual Participant Data From the World COACH Consortium

Myrthe A. van den Berg,[1] Fleur Boel,[1] Michiel M. A. van Buuren,[1] Noortje S. Riedstra,[1] Jinchi Tang,[1] Harbeer Ahedi,[2] Nigel K. Arden,[3] Sita M. A. Bierma-Zeinstra,[1] Cindy G. Boer,[1] Flavia M. Cicuttini,[4] Timothy F. Cootes,[5] David T. Felson,[6] Willem Paul Gielis,[7] Joshua Heerey,[8] Graeme Jones,[9] Stefan Kluzek,[10] Nancy E. Lane,[11] Claudia Lindner,[5] Joyce B. J. van Meurs,[1] Andrea Mosler,[8] Amanda E. Nelson,[12] Michael C. Nevitt,[13] Edwin H. Oei,[1] Jos Runhaar,[1] Harrie Weinans,[14] Jesse H. Krijthe,[15] and Rintje Agricola[1]

Objective. This study aims to develop hip morphology-based radiographic hip osteoarthritis (RHOA) risk prediction models and investigates the added predictive value of hip morphology measurements and the generalizability to different populations.

Methods. We combined data from nine prospective cohort studies participating in the Worldwide Collaboration on OsteoArthritis prediCtion for the Hip (World COACH) consortium. RHOA grades were harmonized, and incident RHOA was defined as hips without definite RHOA at baseline that developed definite RHOA within four to eight years. Baseline hip morphology was quantified with automatically and uniformly determined lateral center edge angle and alpha angle measurements on anteroposterior radiographs. Discriminative performance of generalized linear mixed model (GLMM) definitions with and without hip morphology measurements was determined with stratified cross-validation. With leave-one-cohort-out cross-validation, the generalizability to unseen populations of hip morphology–based GLMMs and random forest (RF) models was evaluated.

Results. From the included 35,984 hips without definite RHOA at baseline, 4.7% developed incident RHOA within four to eight years. The GLMM with cohort-specific intercept, considering baseline demographics, RHOA grade, and hip morphology measurements, showed a mean area under the receiver operating characteristic curve (AUC) of 0.80 (±0.01) in stratified cross-validation. Using a marginal intercept decreased performance by 0.1 in AUC. Similar results were found for a GLMM without hip morphology measurements. Leave-one-cohort-out cross-validation showed comparable discrimination (AUC between 0.56–0.88) and calibration performance for hip morphology-based GLMMs and RF models.

Conclusion. In hips free of definite RHOA, our AUCs for the incident RHOA models showed good predictive performance in similar populations. However, the added predictive value of the morphology measurements was small, and model performance was heterogeneous in leave-one-cohort-out cross-validation.

Initial results of this work have been previously presented at the 2024 Osteoarthritis Research Society International (OARSI) World Congress on Osteoarthritis; Vienna, Austria; April 14–21, 2024 (https://doi.org/10.1016/j.joca.2024.02.105), Combined 61st NOF Congress and NOV Jaarcongres 2024; Rotterdam, June 12–14, 2024; and the 18th International Workshop on Osteoarthritis Imaging; Marrakesh, Morocco; June 25–28, 2024 (https://doi.org/10.1016/j.ostima.2024.100212).

The World COACH consortium has been funded through grants by the Dutch Arthritis Society (grant 21-1-205, LLP34), the Dutch Research Council (now; Veni: 09150161910071), and Erasmus MC University Medical Center Rotterdam. This publication is part of the Flagship Healthy Joints, which is (partly) financed by Convergence Health and Technology. Dr Lindner's work was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant 223267/Z/21/Z). This research was funded in whole, or in part, by the Wellcome Trust (grant 223267/Z/21/Z).

[1]Myrthe A. van den Berg, MSc, Fleur Boel, MSc, Michiel M. A. van Buuren, MD, Noortje S. Riedstra, MD, PhD, Jinchi Tang, MD, PhD, Sita M. A. Bierma-Zeinstra, PhD, Cindy G. Boer, PhD, Joyce B. J. van Meurs, PhD, Edwin H. Oei, PhD, Jos Runhaar, PhD, Rintje Agricola, MD, PhD: Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; [2]Harbeer Ahedi, PhD: University of Tasmania, Menzies Institute for Medical Research, Hobart, Tasmania, and Sydney Orthopaedic Research Institute, Sydney, New South Wales, Australia; [3]Nigel K. Arden, MD: University of Oxford, Botnar Research Centre, Oxford, United Kingdom; [4]Flavia M. Cicuttini, PhD: Monash University, Melbourne, Victoria, Australia; [5]Timothy F. Cootes, PhD, Claudia Lindner, PhD: The University of Manchester, Manchester, United Kingdom; [6]David T. Felson, MD: Boston University School of Medicine, Boston, Massachusetts; [7]Willem Paul Gielis, MD, PhD: University Medical Center Utrecht, Utrecht, The Netherlands; [8]Joshua Heerey PhD, Andrea Mosler, PhD: La Trobe University, Melbourne, Victoria, Australia; [9]Graeme Jones, MD: University of Tasmania,

SIGNIFICANCE & INNOVATIONS

- Our work on hip morphology–based osteoarthritis risk prediction modeling uses an individual participant data meta-analysis to overcome the limitations of small sample sizes in single-cohort designs in current epidemiologic research focused on the role of single-type (eg, cam or pincer) hip morphology on radiographic hip osteoarthritis risk.
- Combining data from 35,984 hips from all available prospective cohort studies with consecutive hip imaging data worldwide (via World COACH), our work uses advanced statistical models to evaluate the importance of continuous hip morphology measures on radiographic hip osteoarthritis (RHOA) risk.
- Although hip morphology was expected to enhance identification of individuals at risk for RHOA, based on its previously established association with RHOA development in isolation, our findings showed that these measures did not improve risk prediction beyond baseline demographics and RHOA grades, highlighting both the complexity of RHOA incidence and challenges in modeling across heterogeneous cohorts.
- Our research demonstrates the importance of large-scale meta-analyses for risk modeling and highlights the significance of properly validating risk models in various populations for future clinical application.

INTRODUCTION

Despite the growing burden of hip osteoarthritis (HOA) on society and health care systems, prevention and treatment approaches are still only slowly emerging.[1] Early identification of HOA is important for improving clinical decision-making and advancing our understanding of disease development and potential prevention and treatment options.[2] By gaining insights into risk factors and thereby identifying individuals with a high risk of developing HOA, more specific management strategies could be designed for specific subgroups.

Previous research has established that hip morphology plays a key role in the development of radiographic HOA (RHOA).[3,4] Hip morphology–related conditions that have been shown to increase the risk of developing RHOA include acetabular dysplasia, pincer morphology, and cam morphology. In research settings, these hip morphologies are assessed using cutoff values of continuous morphologic measurements.[5] However, it is currently recommended to avoid these cutoff values and use the continuous measurements when possible.[6,7]

Current RHOA risk prediction models based on morphology are limited in the level of flexibility and generalizability to unseen data. This problem is caused by the relatively small size of the available datasets and varying study designs.[8] Therefore, most studies have focused on the risk posed by one type of hip morphology determined by conventional statistical methods.

A larger dataset would allow the investigation of linear and nonlinear interactions between RHOA risk and multiple continuous morphology measurements with flexible statistical and machine learning models. A promising solution to increase the sample size is to combine individual participant data (IPD) of various studies while taking differences between studies into account. These differences could also provide insight into the generalizability of the developed prediction models during internal and external model validation. This approach is expected to increase our understanding of the effect of identified hip morphologies as risk factors on the development of RHOA and could result in more robust prediction models.[9,10] Additionally, this approach could show if automatic and interpretable assessment of hip morphology on radiographs could enable scalable risk identification and facilitate implementation in everyday clinical settings.

This study aimed to develop a hip morphology–based RHOA risk prediction model by investigating whether including hip morphology measurements could improve the ability of a model to distinguish people with and without RHOA risk within four to eight years. With the use of a large multicohort dataset, we tested the generalizability of our results in both similar and unseen populations.

MATERIAL AND METHODS

Study design and participants. This study was designed and reported according to the transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster) checklist.[11] Study participants were selected from the Worldwide Collaboration on OsteoArthritis prediCtion for the Hip (World COACH) consortium. This consortium combines data from prospective cohort studies worldwide with sequential pelvic or hip imaging available. Details of this consortium are published elsewhere.[12] The current study first

Menzies Institute for Medical Research, Hobart, Tasmania, Australia; [10]Stefan Kluzek, PhD: University of Oxford, Botnar Research Centre, Oxford, and University of Nottingham, Nottingham, United Kingdom; [11]Nancy E. Lane, MD: University of California at Davis, Davis; [12]Amanda E. Nelson, MD: University of North Carolina at Chapel Hill, Chapel Hill; [13]Michael C. Nevitt, PhD: University of California, San Francisco; [14]Harrie Weinans, PhD: University Medical Center Utrecht, Utrecht, and Delft University of Technology, Delft, The Netherlands; [15]Jesse H. Krijthe, PhD: Delft University of Technology, Delft, The Netherlands.

Author disclosures and a graphical abstract are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25629.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25629).

Address correspondence via email to Rintje Agricola, MD, PhD, at r.agricola@erasmusmc.nl.

Submitted for publication January 8, 2025; accepted in revised form July 31, 2025.

considered cohorts with standardized anteroposterior (AP) pelvic, long-limb, and/or hip radiographs taken and graded for RHOA at baseline and four to eight years of follow-up. This resulted in the initial inclusion of 38,594 participants from nine cohorts (ie, Cohort Hip and Cohort Knee [CHECK], the Chingford Study, The Johnston County Project [JoCoOA], Multicenter Osteoarthritis Study [MOST], OsteoArthritis Initiative [OAI], Rotterdam Study [RS-I, RS-II, RS-III], and the Study of Osteoporotic Fractures [SOF]).

Additional inclusion criteria were applied to IPD at the hip level. Baseline AP pelvic radiographs were required to have sufficient quality to determine pelvic landmark points for automatic hip morphology assessment. Hips only shown in AP hip (rather than pelvic) radiographs were excluded because a horizontal reference line, needed for adjustment for pelvic tilt, requires both hips in view. Additionally, hips must have been graded for RHOA grade at baseline and follow-up visit by the original cohort. All hips with the outcomes (ie, definite RHOA or a total hip replacement [THR] at baseline) were also excluded. In total, 35,984 hips (47%, 18,854 participants) with complete data were included in this study.

Prediction model definition. *Outcome definition and considered risk factors.* RHOA grades were obtained from the original cohort data, all of which used either the Kellgren and Lawrence (KL) or modified Croft grading systems.[13,14] These scores were harmonized to a three-grade system: free of RHOA (KL or Croft = 0), doubtful RHOA (KL or Croft = 1), and definite RHOA (KL or Croft ≥2 or THR). Incident RHOA was defined as hips with no definite RHOA at baseline that developed definite RHOA or had a THR within four to eight years. Our baseline population, therefore, included hips with no signs of RHOA (RHOA grade = 0) or doubtful RHOA (RHOA grade = 1).

To assess hip morphology as a risk factor, the lateral center edge angle (LCEA) and alpha angle (AA) were determined uniformly on the baseline radiographs of all included cohorts. Using the BoneFinder software (www.bone-finder.com; The University of Manchester, UK), the outline of the proximal femur and acetabulum was automatically annotated with a custom-made search model.[15,16] Subsequently, this point-set facilitated the automated LCEA and AA measurements through a Python script. Details regarding the reliability and validity of this method have been published previously.[17] The intermethod intraclass correlation coefficients (ICCs) with 95% confidence interval (CI) between manual and automated measurements were 0.89 (95% CI 0.78–0.94) for the LCEA and 0.46 (95% CI 0.12–0.70) for the AA, similar to the found interobserver ICCs.[16]

The LCEA measurement is depicted in Figure 1A. This measurement quantifies the coverage of the femoral head by the acetabulum and is corrected for pelvic tilt. Both under coverage by the acetabulum (smaller LCEA) and over coverage by the acetabulum (larger LCEA) could be potential risk factors for incident RHOA.[3,18–25] Therefore, we expected a nonlinear relationship between the continuous LCEA measurement and incident RHOA risk.

The AA measurement is depicted in Figure 1B. This measurement describes the degree of sphericity of the lateral femoral head-neck junction and is used to detect the presence and quantify the degree of cam morphology. A linear relationship between the continuous AA and RHOA is expected, such that the higher the AA (lower sphericity), the higher the risk of incident RHOA.[3,22,23,26,27]

Four different models were defined. Model 1 included demographic information, hip side, and follow-up time. The demographic information included baseline age, body mass index, and sex assigned at birth, as they are previously proposed risk factors for incident RHOA.[28] Hip side (left or right hip) and follow-up time in years were included to account for our hip-level modeling approach and adjusted for the different follow-up times of the different cohorts, potentially imposing differences in RHOA risk. In model 2, baseline RHOA grade was added to model 1 to adjust for the expected risk difference for hips with no (RHOA grade 0) and doubtful (RHOA grade 1) RHOA. This risk difference is expected, as hips having an RHOA grade of 1 already show early radiographic changes that can lead to RHOA. After subsequently adding the LCEA and AA measurements in model 3, we considered RHOA grading approach (KL or Croft), cohort location (United States or Europe), and cohort type (open or closed population) as cohort-describing variables in model 4.

Risk prediction modeling approaches. The dataset used is a multilevel dataset, meaning that the data of participants is clustered within cohorts with different cohort-specific characteristics. In this IPD meta-analysis, we considered the correlation among hips from the same cohort in our modeling approaches. One of the conventional approaches to handling multilevel data in risk prediction is generalized linear mixed-effects models (GLMM) for a binary outcome.[29] For the present study, different GLMMs were built with a binomial distribution and a logit link function. We defined a nested random intercept to account for the correlation of measurements in hips within a participant within a cohort. Ultimately, the estimated marginal model intercept was shifted with a combination of a participant-specific and cohort-specific offset value for the participants the model was built on. When new data from the same cohort were used in model testing, the estimated intercept was adjusted with this cohort-specific adjustment. For participants from new cohorts, the estimated marginal model intercept was used. We considered all risk factors to have a common effect across cohorts and to have no interactions with each other.

As an alternative approach, to step away from the assumptions made in a GLMM model structure, we also built random forest (RF) models that considered cohort variability by including the cohort label as a risk factor. An RF model combines the output of multiple decision trees, making it a nonlinear modeling approach that considers all possible interactions between risk factors.

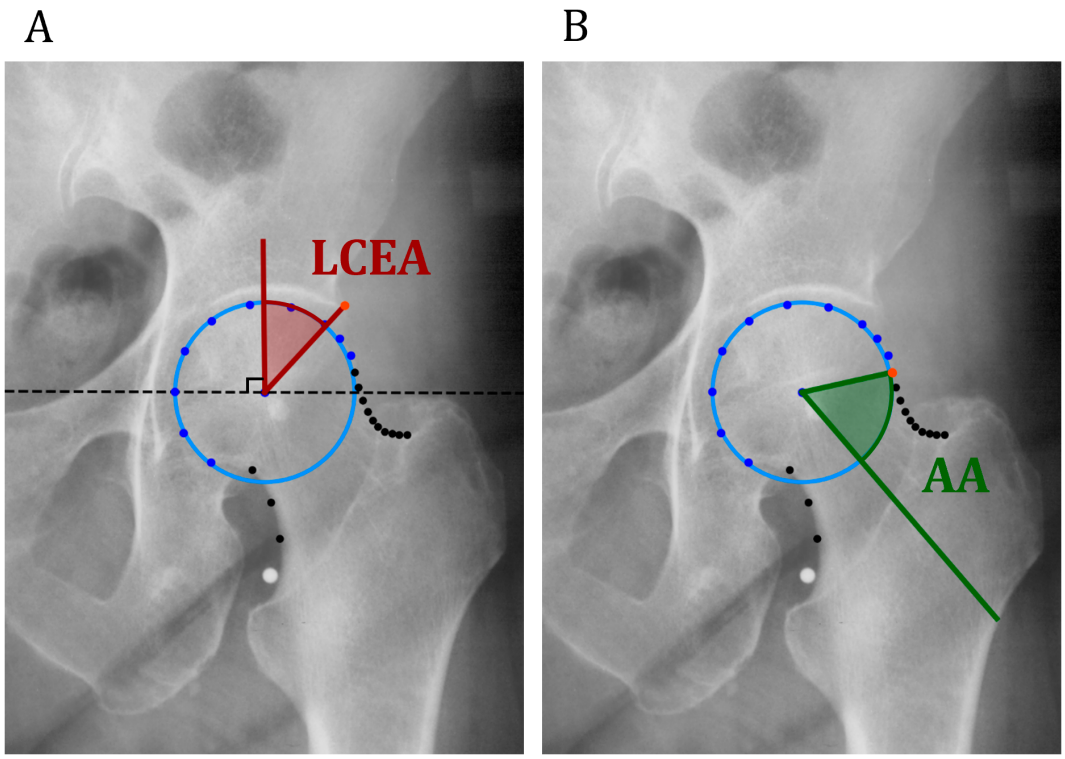

Figure 1. Anteroposterior pelvic radiographs focused on the left hip showing the automatically measured morphology measurements. These methods rely on automated landmark points. To determine the center of the femoral head, a best-fitting circle (light blue) is outlined around the femoral head based on a subset of femoral head landmark points (dark blue). (A) LCEA (in red): The angle is calculated between a line drawn vertically through the center of the femoral head, perpendicular to the horizontal reference line (black, dashed), and a line originating from the center of the femoral head to the most lateral part of the acetabulum (orange landmark). (B) AA (in green): The angle is calculated between a line through the femoral head center and the point where the femoral head or femoral neck is first outside of the best-fitting circle (orange landmark) and the femoral neck axis. AA, alpha angle; LCEA, lateral center edge angle.

It can also model a nonlinear relationship between the risk factors and incident RHOA.[30] For individuals from new cohorts, these variables indicating cohort participation would all be set to 0. After initial experiments, hyperparameters on the RF models were set to a maximum depth of 10, number of estimators of 200, and a minimum of eight samples per leaf.

Statistical analysis. Differences among baseline characteristics of the included and excluded hips and between cohorts were investigated with descriptive statistics. The continuous variables were standardized based on the mean and SD determined on the full dataset before fitting the models. Outcomes were reported based on the original unit scale. Two primary analyses were performed to (1) investigate the associations between the considered risk factors and incident RHOA and to determine the discriminative performance of multiple GLMM definitions; and (2) validate the model performance of both GLMMs and RF models on unseen datasets.

(1) Associations between risk factors and incident RHOA and determining the discriminative performance of GLMM definitions. We first compared the adjusted odds ratios (aOR) with 95% CI of the risk factors in four different GLMM specifications. All GLMMs were built with the lme4 package in R version 4.1.1.[31] The assumed nonlinear effect of the LCEA measurement was modeled with natural cubic splines with two degrees of freedom. Model performance was assessed using stratified five-fold cross-validation, in which the relative contribution to the dataset of a cohort and the percentage of hips at risk were kept constant. Therefore, information for cohort-specific adjustment was available. The discriminative performance of models with or without a cohort-specific intercept was compared. The area under the receiver operating characteristic curve (AUC) was reported as the mean and SD of the measurement on the five resulting test sets. The AUC classifies the predictive performance as fail ($0.5 \leq AUC < 0.6$), poor ($0.6 \leq AUC < 0.7$), fair ($0.7 \leq AUC < 0.8$), good ($0.8 \leq AUC < 0.9$), or excellent ($0.9 \leq AUC$).[32]

(2) Validation of GLMM and RF models on unseen datasets. To assess the discriminative and calibration performance of both the GLMMs and RF models including the hip morphology measurements to unseen populations, leave-one-cohort-out cross-validation was used. Each time, a model was built on a dataset of eight cohorts and evaluated on the left-out cohort. The GLMMs

were built with the lme4 package in R version 4.1.1.[31] The RF models were built with the RandomForestClassifier function in Scikit-learn version 1.2.0, within Python version 3.9.[33] The discriminative performance was evaluated by the mean AUC value and 95% CI in the nine left-out cohorts. Additionally, the calibration statistics of the GLMMs and RF models were compared in each left-out cohort.

Data sharing statement. Data are available upon reasonable request. Data may be obtained from a third party and are not publicly available. We encourage the use of data by third parties, although this is subject to approval by the steering committees of the World COACH consortium and the participating cohorts, as well as to legal boundaries regarding data ownership. A standardized data request form is available for which will be reviewed uniformly in order to consistently handle World COACH data requests.

Patient and public involvement. The World COACH consortium encourages the public to share ideas, stay up to date on our findings, and ask questions through the website (www.worldcoachconsortium.com). Additionally, we help organize annual conferences for patients with OA in the Netherlands (Artrose Gezond, https://artrosegezond.nl/). Patients and public were involved in the design, conduct, reporting, and/or dissemination plans of this research. OA representatives were also involved in setting up the consortium goals and research questions.

Ethical approval. This study involves human participants but was excepted from ethical approval (Erasmus MC Medical Ethics Review Committee) as it uses previously collected observational data for which the participants had originally given informed consent, and all cohort studies included in this consortium already had ethics approval from their respective committees. Participants gave informed consent to participate in the study before taking part.

RESULTS

Participants. A detailed overview of the flow of inclusion and exclusion of hips while pooling the data of the nine cohorts to the final included set is shown in Figure 2. The largest proportion (62%) of hips were excluded because these hips were not graded for RHOA at baseline by the cohort or had no radiograph and/or RHOA grade available at follow-up. Descriptive statistics of the pooled included and excluded dataset, stratified by cohort, are given in Table 1. The excluded population was older than the included population (67.5 ± 10.6 years compared to 63.5 ± 8.5 years). The other descriptive characteristics were similar.

In our included set of 35,984 hips free of definite RHOA at baseline, 1,690 hips developed incident RHOA within four to eight years. Whereas the pooled prevalence of incident RHOA was 4.7%, the prevalence among the cohorts ranged from 1.5% for OAI to 20.2% for CHECK. Additionally, the prevalence of doubtful RHOA ranged from 9% in RS-III to 76% in JoCoOA at baseline. Other differences between the cohort populations at baseline were their mean baseline age, ranging from 54 to 71 years, and sex distribution, with Chingford and SOF being female-only cohorts.

Associations of risk factors and incident RHOA and discriminative performance of GLMM definitions. The aORs and discriminative performance in terms of AUC for our four GLMM definitions are shown in Table 2. Summary statistics on the data used for each train and test dataset combination and model specifications can be found in the Supplementary Material. For the LCEA, the two values of the aOR represent the exponent of the first and second components of the natural cubic spline transformation, respectively. To interpret the effect of 1° of change in the LCEA value, the predictor effect plot for GLMM3 is shown in Figure 3. An aOR of 1.018 for the AA can be interpreted as that, for every 1° increase in the AA value, the odds of incident RHOA increase by 1.8%. Similarly, having an LCEA of 20° compared to 21°, the odds of incident RHOA decrease by 8% (aOR = 0.924).

When using the cohort-specific intercept, adding the baseline RHOA grade (ie, GLMM1 vs GLMM2) improved the discriminative performance of the model from a mean AUC of 0.72 to 0.80 (Table 2), whereas the morphology measurements (GLMM3) did not change this performance. Overall, the discriminative performance was ∼0.1 in AUC value higher when using a cohort-specific intercept to predict the risk of incident RHOA in GLMM1, GLMM2, and GLMM3. When adding the cohort descriptive variables as fixed-effects in the model, the AUC was reduced to 0.76 using the cohort-specific intercept but increased to 0.75 with the marginal intercept, compared to the other model definitions. In GLMM4, the cohort-specific adjustments to the intercept were found to be nonexistent. However, the cohort-specific variables did not show a statistically significant association with incident RHOA. Adding these variables resulted in different aOR values and larger CIs for the other considered risk factors. Specifically, the aOR for having doubtful RHOA at baseline was 17.8. Therefore, the results for GLMM4 should be interpreted with caution.

GLMM and RF model validation on unseen datasets. Table 3 shows the results of the leave-one-cohort-out validation for both the GLMMs and RF models. These models included the baseline patient characteristics, RHOA grade, and morphology measurements (as defined in GLMM3). The summary statistics on the data used for each train and test dataset combination, model specifications, calibration curves, and receiver operating characteristic curves can be found in the Supplementary Material.

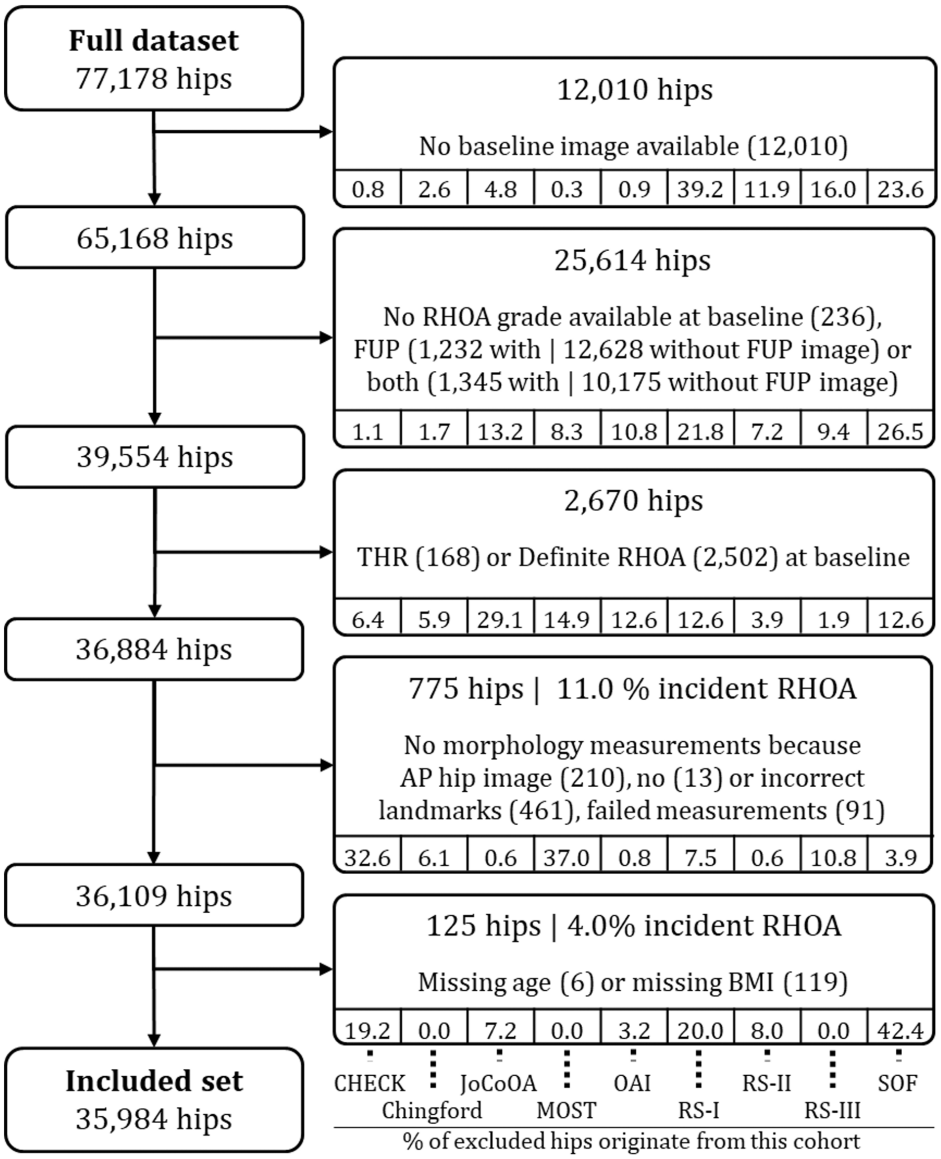

Figure 2. Flow of hips from the fully pooled dataset of the nine cohorts to the final included set. AP, anteroposterior; BMI, body mass index; CHECK, Cohort Hip and Cohort Knee; Chingford, Chingford Study; FUP, follow-up visit; JoCoOA, Johnston County Project; MOST, Multicenter Osteoarthritis Study; OAI, OsteoArthritis Initiative; RHOA, Radiographic hip osteoarthritis; RS-I, Rotterdam Study-I; RS-III, Rotterdam Study-III; SOF, Study of Osteoporotic Fractures; Study-II RS-II, Rotterdam; THR, total hip replacement.

The discriminative performance (AUC) ranged from 0.57 to 0.88 and 0.56 to 0.87 for the GLMMs and RF models, respectively. Validation on the OAI and RS-II cohorts had the highest AUC values, which were the cohorts with the lowest prevalence of incident RHOA. Validation on Chingford and JoCoOA showed the worst performance. The mean predicted values were mostly higher for the RF model, except for RS-I and SOF. When analyzing the model calibration, we observed that models overestimated the risk for the same two cohorts. For CHECK, JoCoOA, Chingford, and RS-III, both models underestimated the risk. For MOST, OAI, and RS-II, the GLMM underestimated and the RF model overestimated the risk. The RF models seemed better calibrated than the GLMM.

DISCUSSION

We thoroughly investigated the effect of two continuous hip morphology measurements on incident RHOA within four to eight years in a large international IPD analysis of nine cohorts and nearly 36,000 hips. During stratified cross-validation, similar discriminative performance was found for models with and without the LCEA and AA measurements (mean AUC 0.80 ± 0.01).

Table 1. Descriptive statistics of the included cohorts, the pooled included set, and pooled excluded set*

	CHECK	Chingford	JoCoOA	MOST	OAI	RS-I	RS-II	RS-III	SOF	Included set	Excluded set[a]
Participants											
Excluded participants, n	361	408	2,212	1,310	1,494	5,260	1,671	2,207	4,812		19,735
Included participants, n	641	595	1,772	1,716	3,302	2,860	1,341	1,732	4,895	18,854	
Sample size											
Hips, n	1,189	1,060	3,209	3,197	6,372	5,542	2,624	3,408	9,383	35,984	41,194
Right hips, %	49.9	47.0	51.6	50.4	49.9	49.7	49.6	50.1	49.6	49.9	50.1
Baseline factors											
Age at baseline, mean ± SD, y	55.8 ± 5.22	53.5 ± 5.71	60.1 ± 8.68	61.7 ± 7.72	60.7 ± 9.03	65.3 ± 6.52	63.2 ± 6.56	56.3 ± 5.01	70.6 ± 4.50	63.4 ± 8.47	67.5 ± 10.58
Male hips, n (%)	210 (17.7)	0 (0.0)	1,232 (38.4)	1,096 (34.3)	2,704 (42.4)	2,412 (43.5)	1,159 (44.2)	1,507 (44.2)	0 (0.0)	10,320 (28.7)	11,750 (28.5)
BMI at baseline, mean ± SD, kg/m^2	26.1 ± 4.00	25.4 ± 3.99	29.5 ± 5.88	29.5 ± 4.85	28.2 ± 4.57	26.3 ± 3.47	27.1 ± 3.90	27.5 ± 4.13	26.5 ± 4.36	27.4 ± 4.57	27.7 ± 5.29
Hips with RHOA grade 1, n (%)	293 (24.6)	71 (6.7)	2,445 (76.2)	1,193 (37.3)	809 (12.7)	2.203 (39.8)	372 (14.2)	322 (9.4)	4,583 (48.8)	12,291 (34.2)	5,784 (27.7)
AA, mean ± SD, degrees	46.4 ± 9.92	51.3 ± 12.63	44.7 ± 8.64	47.1 ± 9.64	48.1 ± 11.45	46.4 ± 10.13	45.5 ± 9.52	50.8 ± 12.84	45.1 ± 9.57	46.7 ± 10.57	47.7 ± 11.75
LCEA, mean ± SD, degrees	34.7 ± 5.59	36.1 ± 6.04	36.3 ± 5.86	33.4 ± 6.05	35.3 ± 5.64	36.6 ± 5.81	35.5 ± 5.68	34.7 ± 5.62	38.2 ± 5.89	36.1 ± 5.99	36.7 ± 6.35
Cohort-level covariates											
Follow-up time, y	8	7	6	5	4	8	4	6	8	4–8	4–8
RHOA grading system	KL	KL	KL	KL	Croft	KL	KL	KL	Croft	KL, Croft	KL, Croft
Location of cohort	Europe	Europe	United States	United States	United States	Europe	Europe	Europe	United States	Europe, United States	Europe, United States
Included population type	Closed	Open	Open	Closed	Closed	Open	Open	Open	Open	Open, closed	Open, closed
Outcome of interest											
Hips with incident RHOA, n (%)	240 (20.2)	110 (10.4)	386 (12.0)	133 (4.2)	95 (1.5)	173 (3.1)	44 (1.7)	146 (4.3)	363 (3.9)	1,690 (4.7)	94 (10.4)

* AA, alpha angle; BMI, body mass index; CHECK, Cohort Hip and Cohort Knee; Chingford, Chingford Study; JoCoOA, Johnston County; KL, Kellgren and Lawrence grade; LCEA, lateral center edge angle; MOST, Multicenter Osteoarthritis Study; OAI, OsteoArthritis Initiative; RHOA, radiographic hip osteoarthritis; RS-I, Rotterdam Study-I; RS-II, Rotterdam Study-II; RS-III, Rotterdam Study-III; SOF, Study of Osteoporotic Fractures.

[a] Sample sizes vary for the descriptive statistics of the excluded set as they are based on only the available data for this variable.

Table 2. Risk factor contribution in terms of aORs of four different multivariate GLMM definitions on the pooled included dataset and their discriminative performance based on stratified cross-validation*

	GLMM1	GLMM2	GLMM3	GLMM4
Age, in years	**1.051 (1.041–1.062)**	**1.043 (1.033–1.053)**	**1.043 (1.033–1.054)**	1.011 (0.947–1.079)
Sex at birth				
Female [reference]				
Male	1.118 (0.954–1.310)	0.979 (0.830–1.155)	**0.828 (0.695–0.987)**	0.389 (0.113–1.336)
BMI, in kg/m^2	1.012 (0.998–1.026)	1.012 (0.998–1.027)	1.009 (0.994–1.024)	1.023 (0.920–1.138)
Hip side				
Left [reference]				
Right	1.101 (0.987–1.229)	**1.149 (1.025–1.290)**	**1.158 (1.031–1.301)**	**1.476 (1.206–1.806)**
Follow-up time, in years	**1.511 (1.124–2.032)**	1.410 (0.928–2.144)	1.409 (0.942–2.106)	1.212 (0.764–1.923)
RHOA grade				
Free of RHOA [reference]				
Doubtful RHOA		**9.000 (7.602–10.655)**	**8.751 (7.378–10.38)**	**17.820 (11.768–26.985)**
AA, in degrees			**1.018 (1.012–1.024)**	1.018 (0.999–1.038)
LCEA,[a] in degrees				
First spline component			**0.062 (0.019–0.197)**	0.009 (0.000–0.390)
Second spline component			1.771 (0.856–3.662)	4.931 (0.528–46.029)
Scoring approach				
KL [reference]				
Croft				0.353 (0.074–1.677)
Cohort type				
Open [reference]				
Closed				2.593 (0.775–8.672)
Location				
Europe [reference]				0.483 (0.115–2.029)
United States				
AUC				
With cohort-specific intercept	0.72 ± 0.01	0.80 ± 0.01	0.80 ± 0.01	0.76 ± 0.02
With marginal intercept	0.56 ± 0.01	0.70 ± 0.01	0.71 ± 0.01	0.75 ± 0.01

* Values are in aOR (95% confidence interval) or mean ± SD. Bold numbers indicate a significant association ($P < 0.05$). AA, alpha angle; aOR, adjusted odds ratio; AUC, area under the receiver operating characteristic curve; BMI, body mass index; GLMM, generalized linear mixed model; KL, Kellgren and Lawrence; LCEA, lateral center edge angle; RHOA, radiographic hip osteoarthritis.
[a] Variable is modeled with natural cubic splines (*df* = 2), displayed aORs are the exponent of the beta values of the two spline components.

When comparing GLMM and RF modeling strategies and their generalizability to unseen populations, similar performances were found with AUC values ranging between 0.56 and 0.88 for each left-out cohort. This implies that using a more flexible model definition that includes possible risk factor interactions or nonlinear relationships would not improve the models. The RF models were better calibrated, which implies that the estimated probabilities are more consistent with the observed RHOA risk distribution within the population.

Previous studies examining the role of hip shape in RHOA risk primarily converted the measurements into a binary risk factor using specific thresholds. Considering hip morphology measurements as a continuous risk factor allows for a more data-driven estimation of the effect. Despite recognizing several hip morphologies as RHOA risk factors, inconsistencies in the literature regarding the thresholds used complicate the interpretation of reported associations.[34] In agreement with previous reports, we found increased odds of incident RHOA when the morphologies were present.[21–23,35]

In contrast, we found comparable AUC values when evaluating the discriminative performance of the models with and without hip morphology measurements (GLMM3 vs GLMM2). This indicates that inclusion of the AA and LCEA did not improve the ability of the model to distinguish between hips that developed or did not develop incident RHOA. Possible explanations are that these geometric measurements alone are not descriptive enough to capture the structural differences between individuals at and not at risk for RHOA development. The low to moderate interrater and intermethod ICCs for the AA might also play a role, potentially reducing the sensitivity of the measurement to identify individuals at risk. Alternatively, abnormal hip morphology may be a true risk factor for early-onset hip OA, but the high-risk participants are already excluded from this analytic population.[23]

The largest increase in AUC values was observed when adding the baseline RHOA status to our model. This indicates that doubtful or mild radiographic changes (ie, RHOA grade 1) identified hips at risk for incident RHOA within four to eight years. Our findings are comparable to those in Saberi Hosnijeh et al, in which an increase of 0.11 in AUC on the training set (RS-I) was found when the imaging variables baseline KL score, thumb OA, hip dysplasia, and cam morphology were added to their basic model; a similar increase in AUC values was found in external validation on CHECK and RS-II data.[36] Based on these findings, in combination with our results, baseline RHOA grade was expectedly the largest contributor to this increase in AUC. Additionally, where Saberi Hosnijeh et al showed an aOR of 4.43 (3.14–5.77) for

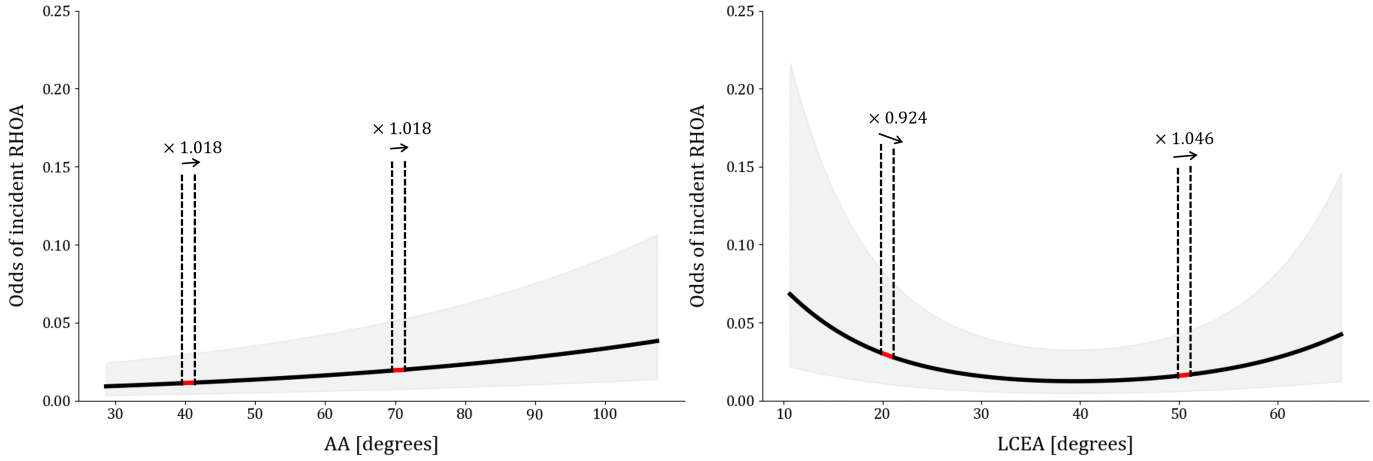

Figure 3. Predictor effect plots of the morphology measurements included in the multivariable risk-prediction model GLMM3. Odds are given for a female participant after eight years of follow-up, with the mean values of the continuous factors and the reference values for the categorical factors. The annotated multiplication factors describe the change in odds for an increase of 1° at that location. AA, alpha angle; LCEA, lateral center edge angle; GLMM, generalized linear mixed model; RHOA, radiographic hip osteoarthritis.

having KL grade 1, we found an aOR of 8.75 (7.38–10.38) for having either KL or Croft grade 1 in GLMM3. Although various opinions exist on considering no versus doubtful RHOA within models for incident RHOA to better understand the specific role of risk factors, including baseline RHOA status in prediction models better identifies hips at risk.

Additionally, differences in performance were found when a cohort-specific intercept instead of a marginal intercept was used. This implies that the currently considered variables do not explain the heterogeneity in incident RHOA prevalence between the cohorts. When cohort-level covariates were added to account for variability in inclusion criteria and cohort setup, the marginal and cohort-specific intercepts were identical. However, because the number of cohorts is always going to be limited, it is inherently difficult to estimate the cohort-level parameters accurately. We hypothesize that the heterogeneity in incident RHOA prevalence is mainly the result of the differences in cohort setup (ie, follow-up time and specific inclusion of people at risk of OA), the semiobjective nature of the RHOA scoring systems, and the different scoring approaches used by the cohorts.[37] Future work in the World COACH consortium aims to include standardized RHOA scoring across cohorts to ensure that observed heterogeneity reflects actual population differences rather than inconsistencies in scoring methods.

The heterogeneity between cohorts contributed to heterogeneous model performance for leave-one-cohort-out validation. Overall, we observed that when the prevalence in the training dataset was higher than the prevalence in the test dataset, discriminative performance improved. As noted by Riley et al, even when the found predictor effects are consistent, variations in characteristics or prevalence across different settings or populations can lead to genuine differences in model calibration due to

Table 3. Discriminative performance and mean predicted probabilities of the GLMM and RF models based on leave-one-cohort-out cross-validation*

	Risk of incident RHOA		GLMM		RF	
Cohort	OA/No-OA	Prevalence, %	Mean predicted values, %	AUC (95% CI)	Mean predicted values, %	AUC (95% CI)
CHECK	240/949	20.2	0.01	0.69 (0.65–0.73)	3.24	0.68 (0.64–0.72)
Chingford	110/950	10.4	0.91	0.57 (0.51–0.63)	2.26	0.57 (0.51–0.63)
JoCoOA	386/2,823	12.0	4.53	0.57 (0.54–0.60)	7.02	0.56 (0.53–0.59)
MOST	133/3,064	4.2	2.25	0.80 (0.77–0.83)	5.33	0.80 (0.76–0.83)
OAI	95/6,277	1.5	0.88	0.86 (0.82–0.90)	3.60	0.82 (0.77–0.87)
RS-I	173/5,369	3.1	9.72	0.78 (0.75–0.80)	5.99	0.76 (0.73–0.79)
RS-II	44/2,580	1.7	0.97	0.88 (0.83–0.94)	3.01	0.87 (0.80–0.93)
RS-III	146/3,262	4.3	1.04	0.75 (0.70–0.80)	2.44	0.74 (0.69–0.79)
SOF	363/9,020	3.9	14.61	0.69 (0.66–0.71)	7.16	0.69 (0.67–0.72)

* AUC, area under the receiver operating characteristic curve; CHECK, Cohort Hip and Cohort Knee; Chingford, Chingford Study; CI, confidence interval; GLMM, generalized linear mixed model; JoCoOA, Johnston County Project; MOST, Multicenter Osteoarthritis Study; OAI, OsteoArthritis Initiative; RF, random forest; RHOA, radiographic hip osteoarthritis; RS-I, Rotterdam Study-I; RS-II, Rotterdam Study-II; RS-III, RS-III, Rotterdam Study-III; SOF, Study of Osteoporotic Fractures.

incorrect estimation of the intercept.[38] In general, an underestimation of incident RHOA risk was found, except for the two older cohorts RS-I and SOF. This could indicate that the considered risk factors have a different effect on RHOA risk for this older population than in the younger training population. More efforts in modeling and understanding the underlying factors for incident RHOA prevalence in these different cohorts could therefore improve RHOA risk estimates. Despite our efforts to combine data from multiple cohorts via IPD to consider a more versatile population to better estimate the intercept and risk factors in new populations, we observed spectrum bias.[39] In addition to creating a larger dataset, the target population should be considered when selecting cohorts to combine, and more risk factors should be added to explain any variability between the cohorts.

The major strength of this study is that it is the first to use data from all available prospective cohort studies on HOA to create risk prediction models for incident RHOA within four to eight years. By combining these IPD, we can include a large and heterogeneous study population. This covers the general population of interest better than individual single-cohort studies. By uniformly measuring our baseline factors, automatically determining our morphology measurements, and taking the cohort differences into account in our modeling approaches, we could provide a more precise estimation of the performance of our prediction models.[40] Through evaluation of models via both internal and external validation, we assessed the generalizability of our predictions, which, although highly encouraged, can be challenging for clinical risk prediction models.

The limitations of the study are mainly caused by the differences in setup of the nine included cohorts, such as varying imaging protocols, inclusion criteria, and follow-up times. Firstly, our defined follow-up time of four to eight years might have been too short and too broad to determine reliable RHOA risk estimates due to the low prevalence of incident RHOA, and this variability in follow-up duration could reduce the apparent strength of associations. When considering an extended follow-up for cohorts with this data available, we see that the prevalence of incident RHOA can double when looking, for example, at 11 years of follow-up for RS-I and RS-II. Secondly, by only defining hip morphology on AP radiographs, we might underestimate its presence and severity, which is better captured in three dimensions. Nevertheless, these automated measurements on two-dimensional imaging allow for easier screening of large populations. Additionally, as the inclusion criteria were applied at the hip level, one potential unadjusted and unmeasured confounder was the increased risk of incident RHOA in individuals with contralateral RHOA. However, this was not the focus of the study. Moreover, including THR in our definition of incident RHOA might have led to some misclassification, as individuals could have undergone THR for other reasons than hip OA. Nonetheless, this group of people accounts for less than 5% of all incident RHOA cases in our study. Lastly, our outcome definition of incident RHOA was dependent on semiobjective scoring systems with high intra- and interrater variability,[41] and Fan et al have shown that prevalence estimates with the KL and Croft grading systems were statistically different.[1] Although the defined risk of interest aligns with the current clinical practice, it does limit the interpretability of the results.

Based on the current study, using more automated RHOA grading systems should be encouraged in the future to reduce variability in scoring for multicohort analyses. In addition, automated and uniform RHOA grading methods could reduce the time needed to reliably grade radiographs already collected in epidemiologic research and could therefore increase the sample size. To aid clinical decision-making and treatment options, multimodal prediction modeling with risk factors from different domains, such as genetics and clinical examination, is recommended.[42] Although the set of considered risk factors in this study already provided valuable insights into incident RHOA, creating artificial intelligence models that can analyze radiographs and consider many more than only demographic and hip morphology–based risk factors could improve our understanding of HOA development. This might also reduce the present heterogeneity in model performance when applied to new populations.

In conclusion, the hip morphology–based risk prediction models built with multicohort data had good predictive performance for incident RHOA within four to eight years for hips free of definite RHOA at baseline. However, we found relatively high heterogeneity in model performance in leave-one-cohort-out cross-validation. The added value of hip morphology measurements on the discriminative performance was small compared to a model with baseline patient characteristics and RHOA grade alone. More efforts are recommended to reduce the heterogeneity between cohorts and increase model calibration. Ultimately, risk prediction models built with multicohort data could lead to an improved understanding of HOA and be a first step toward individualized HOA care.

ACKNOWLEDGMENTS

We would like to thank all participants of each of the cohort studies that are involved in World COACH. We gratefully acknowledge all international organizations that collaborated with the cohort studies in World COACH, as well as the OARSI for endorsing the World COACH consortium. The CHECK study was initiated by the Dutch Arthritis Society and performed within Erasmus Medical Center Rotterdam; Kennemer Gasthuis Haarlem; Leiden University Medical Center; Maastricht University Medical Center; Martini Hospital Groningen/Allied Health Care Center for Rheum. and Rehabilitation Groningen; Medical Spectrum Twente Enschede/Ziekenhuisgroep Twente Almelo; Reade, formerly Jan van Breemen Institute/VU Medical Center Amsterdam; St.Maartens-kliniek Nijmegen; University Medical Center Utrecht; and Wilhelmina Hospital Assen. We would like to thank all the participants of the Chingford Women Study, Professor Nigel Arden, Professor Tim Spector, Dr Deborah Hart, Gem Lawson, Maxine Daniels, and Alison Turner for their time and dedication and Arthritis Research UK for their funding support to the study and the Oxford NIHR Musculoskeletal Biomedical Research Unit for funding contributions. The Femoroacetabular impingement and

hip OsteoaRthritis Cohort was supported by the Australian National Health and Medical Research Council; Early Career Fellowship grant 1119971; Australian Postgraduate Scholarship. Support for data from the Johnston County Osteoarthritis Project was provided in part by the Center for Disease Control and Prevention (CDC) (grants U01DP006266 and U01DP003206); Association of Schools of Public Health/CDC (grants S043, S1734, S3486); and National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (P60AR30701, P60AR049465, P60AR064166, and P30AR072580). The MOST study was funded by the National Institute on Aging, NIH (grants AG19069 [Michael Nevitt, University of California, San Francisco], AG18820 [David Felson, Boston University], AG18947 [Cora Lewis, University of Alabama at Birmingham], and AG18832 [James Torner, University of Iowa]). The cohort, clinical data and image acquisitions used in these analyses were funded as the Osteoarthritis Initiative by the NIH through a Foundation for NIH public private partnership with GlaxoSmithKline, Merck & Co., Inc., Novartis Pharmaceuticals Corporation, and Pfizer. The Rotterdam Study is funded by Erasmus University Medical Center and Erasmus University, Rotterdam, The Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The Study of Osteoporotic Fractures is supported by NIH funding (grants R01-AG-005407, R01-AR-35582, R01-AR-35583, R01-AR-35584, R01-AG-005394, R01-AG-027574, and R01-AG-027576. The TASOAC study was supported by the National Health and Medical Research Council of Australia, Tasmanian Community Fund, Masonic Centenary Medical Research Foundation, Royal Hobart Hospital Research Foundation, and Arthritis Foundation of Australia.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Agricola confirms that all authors have provided the final approval of the version to be published, and takes responsibility for the affirmations regarding article submission (eg, not under consideration by another journal), the integrity of the data presented, and the statements regarding compliance with institutional review board/Declaration of Helsinki requirements.

REFERENCES

1. Fan Z, Yan L, Liu H, et al. The prevalence of hip osteoarthritis: a systematic review and meta-analysis. Arthritis Res Ther 2023;25(1):51.
2. Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. Bone 2012; 51(2):278–288.
3. Shapira J, Chen JW, Bheem R, et al. Radiographic factors associated with hip osteoarthritis: a systematic review. J Hip Preserv Surg 2020; 7(1):4–13.
4. Agricola R, van Buuren MMA, Kemp JL, et al. Femoroacetabular impingement syndrome in middle-aged individuals is strongly associated with the development of hip osteoarthritis within 10-year follow-up: a prospective cohort study (CHECK). Br J Sports Med 2024; 58(18):1061–1067.
5. Melugin HP, Hale RF, Lee DR, et al. Risk factors for long-term hip osteoarthritis in patients with hip dysplasia without surgical intervention. J Hip Preserv Surg 2022;9(1):18–21.
6. Fischer CS, Kühn JP, Ittermann T, et al. What are the reference values and associated factors for center-edge angle and alpha angle? A population-based study. Clin Orthop Relat Res 2018;476(11):2249–2259.
7. Dijkstra HP, Mc Auliffe S, Ardern CL, et al; Young Athlete's Hip Research (YAHiR) Collaborative. Oxford consensus on primary cam morphology and femoroacetabular impingement syndrome: part 1-definitions, terminology, taxonomy and imaging outcomes. Br J Sports Med. 2022;57(6):325–341.
8. Appleyard T, Thomas MJ, Antcliff D, et al. Prediction models to estimate the future risk of osteoarthritis in the general population: a systematic review. Arthritis Care Res (Hoboken) 2023;75(7):1481–1493.
9. Steyerberg EW, Nieboer D, Debray TPA, et al. Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: an overview and illustration. Stat Med 2019;38(22):4290–4309.
10. Wiemken TL, Kelley RR. Machine learning in epidemiology and health outcomes research. Annu Rev Public Health 2020;41(1):21–36.
11. Debray TPA, Collins GS, Riley RD, et al. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. BMJ 2023;380:e071018.
12. van Buuren MMA, Riedstra NS, van den Berg MA, et al. Cohort profile: worldwide Collaboration on OsteoArthritis prediCtion for the Hip (World COACH) - an international consortium of prospective cohort studies with individual participant data on hip osteoarthritis. BMJ Open 2024;14(4):e077907.
13. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16(4):494–502.
14. Croft P, Cooper C, Wickham C, et al. Defining osteoarthritis of the hip for epidemiologic studies. Am J Epidemiol 1990;132(3):514–522.
15. Lindner C, Thiagarajah S, Wilkinson JM, et al; arcOGEN Consortium. Fully automatic segmentation of the proximal femur using random forest regression voting. IEEE Trans Med Imaging 2013;32(8):1462–1472.
16. Boel F, Riedstra NS, Tang J, et al. Reliability and agreement of manual and automated morphological radiographic hip measurements. Osteoarthr Cartil Open 2024;6(3):100510.
17. Boel F, de Vos-Jakobs S, Riedstra NS, et al. Automated radiographic hip morphology measurements: an open-access method. Osteoarthr Imaging 2024;4(2):100181.
18. Jacobsen S, Sonne-Holm S. Hip dysplasia: a significant risk factor for the development of hip osteoarthritis. A cross-sectional survey. Rheumatology (Oxford) 2005;44(2):211–218.
19. Lane NE, Lin P, Christiansen L, et al. Association of mild acetabular dysplasia with an increased risk of incident hip osteoarthritis in elderly white women: the study of osteoporotic fractures. Arthritis Rheum 2000;43(2):400–404.
20. Reid GD, Reid CG, Widmer N, et al. Femoroacetabular impingement syndrome: an underrecognized cause of hip pain and premature osteoarthritis? J Rheumatol 2010;37(7):1395–1404.
21. Reijman M, Hazes JM, Pols HA, et al. Acetabular dysplasia predicts incident osteoarthritis of the hip: the Rotterdam study. Arthritis Rheum 2005;52(3):787–793.
22. Runhaar J, van Berkel AC, Agricola R, et al. Risk factors and population-attributable fractions for incident hip osteoarthritis. HSS J 2023;19(4):407–412.
23. Saberi Hosnijeh F, Zuiderwijk ME, Versteeg M, et al. Cam deformity and acetabular dysplasia as risk factors for hip osteoarthritis. Arthritis Rheumatol 2017;69(1):86–93.
24. Ganz R, Parvizi J, Beck M, et al. Femoroacetabular impingement: a cause for osteoarthritis of the hip. Clin Orthop Relat Res 2003;417: 112–120.

25. Riedstra NS, Boel F, van Buuren MMA, et al. Acetabular dysplasia and the risk of developing hip osteoarthritis within 4-8 years; an individual participant data meta-analysis of 18,807 hips from the World COACH consortium. Osteoarthritis Cartilage. 2025;33(3):373–382.
26. Agricola R, Heijboer MP, Bierma-Zeinstra SM, et al. Cam impingement causes osteoarthritis of the hip: a nationwide prospective cohort study (CHECK). Ann Rheum Dis 2013;72(6):918–923.
27. van Klij P, Heerey J, Waarsing JH, et al. The prevalence of cam and pincer morphology and its association with development of hip osteoarthritis. J Orthop Sports Phys Ther 2018;48(4):230–238.
28. Chaganti RK, Lane NE. Risk factors for incident osteoarthritis of the hip and knee. Curr Rev Musculoskelet Med 2011;4(3):99–104.
29. Xiong C, Zhu K, Yu K, et al. Statistical modeling in biomedical research: longitudinal data analysis. In: Rao CR, Miller JP, Rao DC, eds. Essential Statistical Methods for Medical Statistics. Elsevier; 2011:235–268.
30. Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
31. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1).
32. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol 2022;75(1):25–36.
33. Pedregosa FVG, Gramfort A, Michel V, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–2830.
34. Casartelli NC, Maffiuletti NA, Valenzuela PL, et al. Is hip morphology a risk factor for developing hip osteoarthritis? A systematic review with meta-analysis. Osteoarthritis Cartilage 2021;29(9):1252–1264.
35. Riedstra NS, Vinge R, Herfkens J, et al. Acetabular dysplasia and the risk of developing hip osteoarthritis at 2,5,8, and 10 years follow-up in a prospective nationwide cohort study (CHECK). Semin Arthritis Rheum 2023;60:152194.
36. Saberi Hosnijeh F, Kavousi M, Boer CG, et al. Development of a prediction model for future risk of radiographic hip osteoarthritis. Osteoarthritis Cartilage 2018;26(4):540–546.
37. Damen J, Schiphof D, Wolde ST, et al. Inter-observer reliability for radiographic assessment of early osteoarthritis features: the CHECK (cohort hip and cohort knee) study. Osteoarthritis Cartilage 2014;22(7):969–974.
38. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016;353:i3140.
39. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med 2002;137(7):598–602.
40. Costa ALF, Lopes SLPC. Highlighting the benefits and disadvantages of individual participant data meta-analysis in radiology. Radiol Imaging Cancer 2024;6(2):e240018.
41. Kohn MD, Sassoon AA, Fernando ND. Classifications in brief: Kellgren-Lawrence classification of osteoarthritis. Clin Orthop Relat Res 2016;474(8):1886–1893.
42. Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: a scoping review. NPJ Digit Med 2022;5(1):171.