

**Document Version**

Final published version

**Licence**

CC BY

**Citation (APA)**

Gielis, W. P., Weinans, H., Welsing, P. M. J., van Spil, W. E., Agricola, R., Cootes, T. F., de Jong, P. A., & Lindner, C. (2020). An automated workflow based on hip shape improves personalized risk prediction for hip osteoarthritis in the CHECK study. *Osteoarthritis and Cartilage*, 28(1), 62-70. <https://doi.org/10.1016/j.joca.2019.09.005>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.  
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

**Sharing and reuse**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

## An automated workflow based on hip shape improves personalized risk prediction for hip osteoarthritis in the CHECK study



W.P. Gielis †\*, H. Weinans ‡§, P.M.J. Welsing ||, W.E. van Spil ||, R. Agricola ¶, T.F. Cootes #, P.A. de Jong ††, C. Lindner #

† UMC Utrecht, Department of Orthopedics and Department of Radiology, Utrecht, the Netherlands

‡ UMC Utrecht, Department of Orthopedics and Department of Rheumatology & Clinical Immunology, Utrecht, the Netherlands

§ TU Delft, Department of Biomechanical Engineering, Delft, the Netherlands

|| UMC Utrecht, Department of Rheumatology & Clinical Immunology, Utrecht, the Netherlands

¶ Erasmus MC, Department of Orthopedics, Rotterdam, the Netherlands

# University of Manchester, Division of Informatics, Imaging & Data Sciences, Manchester, United Kingdom

†† UMC Utrecht, Department of Radiology, Utrecht, the Netherlands

### ARTICLE INFO

#### Article history:

Received 3 January 2019

Accepted 22 September 2019

#### Keywords:

Hip osteoarthritis

Imaging

Statistical shape analysis

Epidemiology

### SUMMARY

**Objective:** To design an automated workflow for hip radiographs focused on joint shape and tests its prognostic value for future hip osteoarthritis.

**Design:** We used baseline and 8-year follow-up data from 1,002 participants of the CHECK-study. The primary outcome was definite radiographic hip osteoarthritis (rHOA) (Kellgren–Lawrence grade  $\geq 2$  or joint replacement) at 8-year follow-up. We designed a method to automatically segment the hip joint from radiographs. Subsequently, we applied machine learning algorithms (elastic net with automated parameter optimization) to provide the Shape-Score, a single value describing the risk for future rHOA based solely on joint shape. We built and internally validated prediction models using baseline demographics, physical examination, and radiologists scores and tested the added prognostic value of the Shape-Score using Area-Under-the-Curve (AUC). Missing data was imputed by multiple imputation by chained equations. Only hips with pain in the corresponding leg were included.

**Results:** 84% were female, mean age was 56 ( $\pm 5.1$ ) years, mean BMI 26.3 ( $\pm 4.2$ ). Of 1,044 hips with pain at baseline and complete follow-up, 143 showed radiographic osteoarthritis and 42 were replaced. 91.5% of the hips had follow-up data available. The Shape-Score was a significant predictor of rHOA (odds ratio per decimal increase 5.21, 95%-CI (3.74–7.24)). The prediction model using demographics, physical examination, and radiologists scores demonstrated an AUC of 0.795, 95%-CI (0.757–0.834). After addition of the Shape-Score the AUC rose to 0.864, 95%-CI (0.833–0.895).

**Conclusions:** Our Shape-Score, automatically derived from radiographs using a novel machine learning workflow, may strongly improve risk prediction in hip osteoarthritis.

© 2019 The Authors. Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

### Introduction

Hip osteoarthritis (HOA) is often diagnosed relatively late in the disease process and currently there are no drugs available to modify disease progression. Therefore, HOA treatment is necessarily restricted to education, exercise, weight loss and analgesics. Total hip replacement (THR) often follows when these do not suffice<sup>1</sup>.

To guide current care and develop interventions to modify the disease course, accurate prediction of HOA development in persons presenting with hip complaints is important. Many risk factors for

\* Address correspondence and reprint requests to: W.P. Gielis, UMC Utrecht, Huispostnummer G05.228, PO box 85500, 3508 GA, Utrecht, the Netherlands. Tel: 31 (0) 8875 – 59025; Fax: 31 (0) 8875 – 55493.

E-mail addresses: [w.p.gielis@umcutrecht.nl](mailto:w.p.gielis@umcutrecht.nl) (W.P. Gielis), [h.h.weinans@umcutrecht.nl](mailto:h.h.weinans@umcutrecht.nl) (H. Weinans), [p.m.j.welsing@umcutrecht.nl](mailto:p.m.j.welsing@umcutrecht.nl) (P.M.J. Welsing), [w.e.vanspil@umcutrecht.nl](mailto:w.e.vanspil@umcutrecht.nl) (W.E. van Spil), [r.agricola@erasmusmc.nl](mailto:r.agricola@erasmusmc.nl) (R. Agricola), [timothy.f.cootes@manchester.ac.uk](mailto:timothy.f.cootes@manchester.ac.uk) (T.F. Cootes), [p.dejong-8@umcutrecht.nl](mailto:p.dejong-8@umcutrecht.nl) (P.A. de Jong), [claudia.lindner@manchester.ac.uk](mailto:claudia.lindner@manchester.ac.uk) (C. Lindner).

HOA are reported in the literature. However, no established risk prediction tool for HOA is currently available. The rise of automated image processing techniques using artificial intelligence, offers the possibility to extract information from images beyond traditional visual interpretation. For example, deep neural networks can be used on computed tomography scans to classify arterial calcifications or pulmonary peri-fissural nodules<sup>2,3</sup>.

Shape variations in the hip play a role in the development of HOA<sup>4,5</sup>. Geometric measurements for assessing hip dysplasia or cam morphology are used as clinical tools<sup>6</sup>, but only describe particular components of the hip shape. Statistical Shape Models (SSMs) have the potential to quantify the overall shape variation of the bone, including more subtle variations<sup>7–10</sup>. However, SSMs require labor-intensive manual input to outline (i.e., segment) shapes with landmark points, hampering their use in large study populations. Therefore, we developed a segmentation software system to automatically extract hip shape from standard pelvic radiographs<sup>11,12</sup>.

This study describes the development and validation of a prediction tool for future HOA in a large cohort of persons with hip pain that had never or only recently (<6 months) consulted a physician for their complaints. Our prediction tool, Shape-Score, utilizes overall hip shape based on SSMs, using our segmentation software system on standard pelvic radiographs. Moreover, we quantified the added predictive value of our tool over clinically available predictors alone.

## Participants and methods

### Participants

Cohort Hip and Cohort Knee (CHECK) aimed to examine the course of early OA in the hip and/or knee<sup>13</sup>. Between October 2002 and September 2005, 1,002 participants were enrolled in 10 participating centers throughout the Netherlands. Possible candidates were approached by their general practitioner and/or recruited via local media. Participants were aged 45–65 years at the time of inclusion and had pain and/or stiffness in at least one knee and/or hip. They had not, or only recently (<6 months) consulted a physician for these complaints. Exclusion criteria were (i) pathology other than OA explaining symptoms, (ii) expected inability to complete 10-year follow-up, and (iii) inability to sufficiently understand Dutch. Radiographic knee OA (defined as KL 2 or higher) was not present in patients at baseline.

### Measurements

#### Demographics

Age, gender, BMI and current smoking (yes or no) were registered. Highest education level was scored on a scale from 1 to 8, as a proxy for socio-economic status. The scale is described in Table 1.

#### Clinical examination

Trained health professionals registered hip pain, when pain was present around the groin/buttock/upper leg. Additionally, knee pain was registered if pain was present around the knee (possible referred pain). The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) total score was used to summarize pain, stiffness, and physical function. Analgesic use and morning stiffness were registered as present or absent. Active internal hip rotation was measured using a goniometer, according to Norkin and White<sup>14</sup>. Pain during internal rotation was registered (yes/no).

**Table 1**  
Baseline characteristics (n = 1,140 painful hips)\*

Demographics	(Data based on N = )	
Age in years, mean (SD)	55.9 (5.1)	1140
BMI in kg/m <sup>2</sup> , mean (SD)	26.3 (4.2)	1119
Highest education, median (IQR) <sup>†</sup>	5 (4–7)	1109
Female sex, N (%)	954 (87.3)	1140
Current smoking, N (%)	158 (14.2)	1115
Clinical examination		
WOMAC total score, mean (SD)	24.9 (16.3)	1113
Pain located around hip, N (%)	673 (59.4)	1133
Pain located around knee, N (%)	961 (84.5)	1137
Analgesic use, N (%)	387 (34.5)	1123
Morning stiffness, N (%)	423 (38.3)	1104
Range of motion hip, mean (SD)	30.6 (8.7)	1134
Pain on internal hip rotation, N (%)	342 (30.2)	1134
Basic radiographic parameters		
Hip Joint space narrowing, N (%)	367 (32.8)	1108
Hip Osteophytosis, N (%)	374 (33.8)	1118
Hip Buttressing, N (%)	64 (5.7)	1118
Hip KL grade 1, N(%) <sup>‡</sup>	434 (38.6)	1118
Knee KL grade 1, N(%) <sup>‡</sup>	266 (23.8)	1140
Shape-Score		
Shape-Score, mean (SD)	0.18 (0.066)	1140

\* As missing values are drawn from a distribution in multiple imputations, the number of included hip joints (range 1,140–1,143) differed per imputed dataset.

<sup>†</sup> 1 = no school, 2 = primary school, 3 = basic vocational education, 4 = secondary education, 5 = secondary vocational education, 6 = Higher and university preparatory education, 7 = higher professional education, 8 = university, IQR = interquartile range.

<sup>‡</sup> Kellgren and Lawrence grade for radiographic osteoarthritis, all hips and knees were graded 0 or 1 at baseline.

### Basic radiographic parameters

Standardized weight-bearing anteroposterior pelvic radiographs, with 15-degree internal hip rotation were made. Presence of joint space narrowing (JSN), osteophytes and thickening of the femoral calcar (buttressing) were scored (yes/no) on baseline radiographs by five trained observers as previously described<sup>15,16</sup>.

### Automatic quantification of subtle shape variations

SSMs provide a global representation of shape rather than reducing shape to a set of geometric measurements, enabling quantification and analysis of complex and subtle shape aspects. Using predefined (anatomical) landmark points, an object, such as the bones of the hip joint, can be outlined and segmented. Based on all landmark points across a set of images, a Statistical shape modelling (SSM) can be generated by applying principal component analysis to the aligned shapes<sup>17</sup>. The SSM then describes every shape by the combination of a mean shape and a linear combination of a number of shape modes. Each mode describes a distinct shape aspect. The first shape mode explains the highest proportion of variation across the dataset and each additional mode explains a smaller part of the total variation.

We developed a fully automatic segmentation system (FASS) to segment the hip using 75 landmark points (Supplementary text and S1 Figure and S2 Figure)<sup>11,12,18</sup>. We used all 1,373 baseline pelvic radiographs of sufficient quality with manual segmentations available from previous work, enabling to compare the predictive value of the data produced by the FASS to the data produced by manual segmentations<sup>18</sup>. Below, we give a concise overview of the development of the Shape-Score using FASS/SMM shape modes. More detailed information and a comparison between Shape-Scores from manual vs automatic segmentations is provided in the Supplementary text.

We used the first 26 SSM modes, explaining 90% of the overall shape variation across our dataset to develop the Shape-Score.

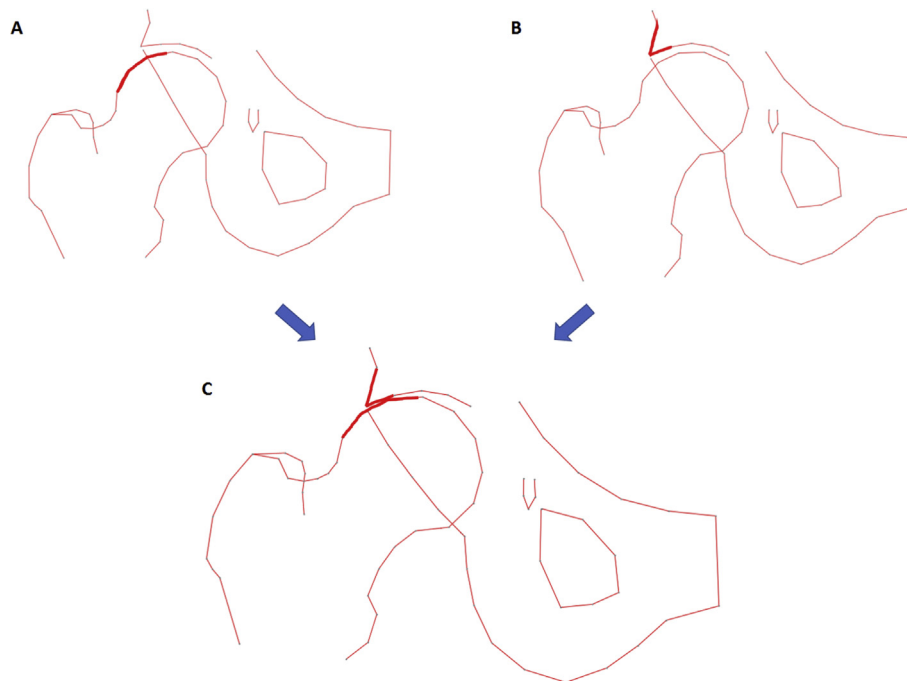
While individual shape modes are independent by the nature of SSMs, the simultaneous effect of two (or more) shape modes on the risk for OA may interact (e.g., be multiplicative instead of additive). For example, a mode describing cam morphology may strengthen the effect of a mode with increased acetabular coverage (i.e., pincer morphology). The theoretical explanation would be that an increased acetabular coverage causes cam morphology to impinge earlier against the acetabulum, which might increase the risk of labral damage and subsequent HOA (Fig. 1)<sup>19</sup>. Considering interactions between all 26 modes produces 325 combinations (the sum of the arithmetic row from 1 to 25 with a common difference of one). Adding the 26 separate modes would produce a total of 351 variables, a number too large for standard regression techniques. Therefore, we used a penalized regression technique (an elastic net) suited for high dimensional data to relate all these variables to the incidence of HOA and produce a single score representing HOA risk based on hip shape (Supplementary text). The resulting Shape-Score ranges between 0 and 1, and contains various clear and subtle aspects of hip geometry. Compared to the low-risk shape, the high-risk shape shows a cam morphology (an aspherical femoral head-neck junction), a narrower superior joint space, decreased acetabular depth in combination with lateralization of the femur. Additionally, the femoral shaft is narrower, while the femoral neck is wider. However, these shape variations do not have to coincide within a patient and a single high risk shape variation may increase the Shape-Score (Fig. 2).

#### Prediction model development and performance testing

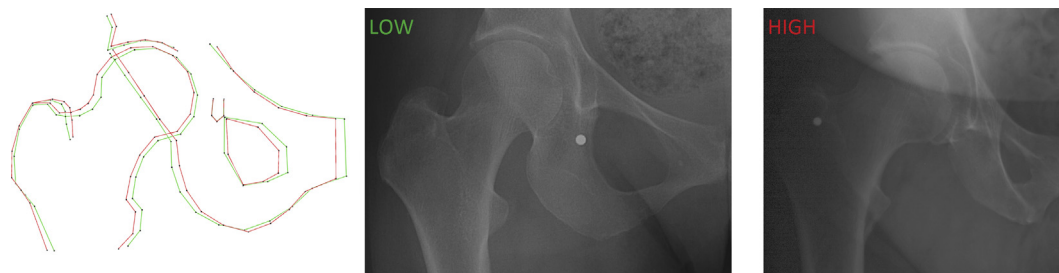
To develop the prediction model we only used data of hips with pain around the hip and/or the knee of the respective leg (possible referred pain) as these are the joints where this prediction will most likely be used for in clinical practice. This is in contrast to the development of the shape model itself, in which all available

baseline pelvic radiographs with data on HOA on follow-up were used. Furthermore, for a hip to be included in the prediction model a baseline radiograph of sufficient quality had to be available. Depending on these criteria, one or both hips of a participant were included in the analyses. Baseline predictors were used to predict the outcome, radiographic hip osteoarthritis (rHOA) at 8-year follow-up, defined as a Kellgren–Lawrence grade (KL-grade)  $\geq 2$ , or THR<sup>16,20</sup>. All predictors were measured at baseline and categorized as demographics, clinical examination, basic radiographic parameters, or Shape-Score. To account for missing predictor and outcome data, we imputed 15 datasets using predictive mean matching and logistic regression<sup>21</sup>. We performed a sensitivity analysis including only hips with complete data using logistic regression and generalized estimating equations (GEE).

To develop the prediction model, logistic regression was used and predictors were added per category. First demographics were added, secondly variables from the clinical examination, thirdly basic radiographic parameters, finally the Shape-Score. After each step (addition of a category of predictors) we simplified the model by removing redundant predictors from the added category only, using backwards selection with a pooled alpha-level of 0.15<sup>21,22</sup>. To optimize parameter estimates for predictors and avoid overfitting, we used logistic ridge regression on each imputed dataset separately. Optimal penalties were based on corrected Akaike's Information Criteria<sup>23</sup>. When using multiple imputation, the imputed values per imputed dataset may differ, as they are drawn from a distribution. We averaged the intercepts and parameter estimates of the ridge regression models from all imputed datasets, to obtain formulas to calculate individual risks for rHOA or THR in future research or clinical work (Supplementary text)<sup>23,24</sup>. We calculated predicted risks and stratified all hips into arbitrary risk categories of <20%, 20–50% and >50% risk for OA. We calculated positive and negative predictive value for the low (<20%) and high risk (>50%) categories. For the low-risk category the absence of OA at 8-year



**Fig. 1. Interaction between shape modes.** The risk for HOA produced by a shape mode may depend on the presence of other shape features. In this hypothetical example, shape mode A represents cam morphology (Aspherical femoral head-neck junction), the shape mode B represents increased acetabular coverage. Both shape modes have a risk factor for OA. When both features coincide in one hip, the risk for OA may be greater than the sum of two risk factors from shape modes A and B. In C we combine the femur of mode A with the pelvis of mode B and simulate hip motion by applying 15 degrees of abduction to the femur. The risk for femoroacetabular impingement becomes clear and is very plausible. Femoroacetabular impingement may increase the risk of labral damage and subsequent HOA.



**Fig. 2. High vs low risk Shape-Score.** The left shows a schematic representation of the mean shape of the 5% highest Shape-Score (high risk for future HOA, in red) and 5% lowest shape-score (low risk for future HOA, in green). Compared to the low-risk hip, the high-risk hip is characterized by a cam morphology (femoral head-neck asphericity), a narrower superior joint space, and decreased acetabular depth in combination with lateralization of the femur and a higher neck-shaft angle. Additionally, in the high-risk hip the femoral shaft is narrower, while the femoral neck is wider. These shape variations, however, do not have to coincide within a patient and a single high risk shape variation may increase the Shape-Score. In the middle and on the right, a real radiograph of a low risk and high risk hip are shown.

follow-up was considered a positive gold standard. For the high-risk category the presence of OA at 8-year follow-up was considered a positive gold standard.

Performance of the model was further assessed in terms of calibration, i.e., the agreement between predictions and observed outcomes, as well as discrimination, i.e., the ability of a model to differentiate hips with developing hip OA from those which will not<sup>25</sup>. To assess calibration of the model, we plotted the percentage of observed OA cases in groups of hips with increasing predicted risk ( $n = 23$  per group to create 50 data points in the plot). Using a loess smoothing function, we visualized the calibration of the model. We assessed discrimination by calculating Area Under the Curve (AUC) statistics with bootstrapped 95% confidence intervals pooled over the 15 imputed datasets. Any prediction model will perform better in the dataset used to train the model compared to a dataset containing new patients. Therefore, performance measures based on training data will be over-optimistic, also known as overfitted<sup>26</sup>. To internally validate the models and estimate their performance in new patients we used bootstrapping, a resampling method. We drew 1,000 bootstrap samples per imputed dataset and pooled the AUC and calibration plots to test for over-optimism of both calibration and discrimination<sup>26</sup>. All data analyses were performed using R v3.3.1. with MICE v2.30, caret v6.0–73, rms v5.1–0, and glmnet v2.0–5.

## Results

Baseline characteristics can be found in [Table I](#). Of 1,044 hips with data on KL-grade or THR at 8-year follow-up, 143 showed KL-grade 2 or higher and 42 had undergone THR. Among the demographics included as predictor in the initial model, smoking status was non-significant ( $p$ -value 0.74) and removed from the model. Among the predictors from clinical examination, morning stiffness ( $p$ -value 0.46) and pain on internal rotation ( $p$ -value 0.17) were removed. All basic radiographic parameters were significant (at the alpha level of 0.15) and were retained. Predictors that were retained in the models are shown in [S1 Table](#) with their respective unpenalized odds ratios. The formulas used to calculate the predicted risks are given in [Supplementary text](#).

### Model performance

The discriminative ability of the models improved each time an additional category of predictors was added, meaning that the models' ability to separate cases from non-cases increased ([Fig. 3](#), [Table II](#)). A model containing only the Shape-Score discriminated comparable to a model combining demographics, clinical examination and basic radiographic parameters (AUC 0.798 vs 0.795).

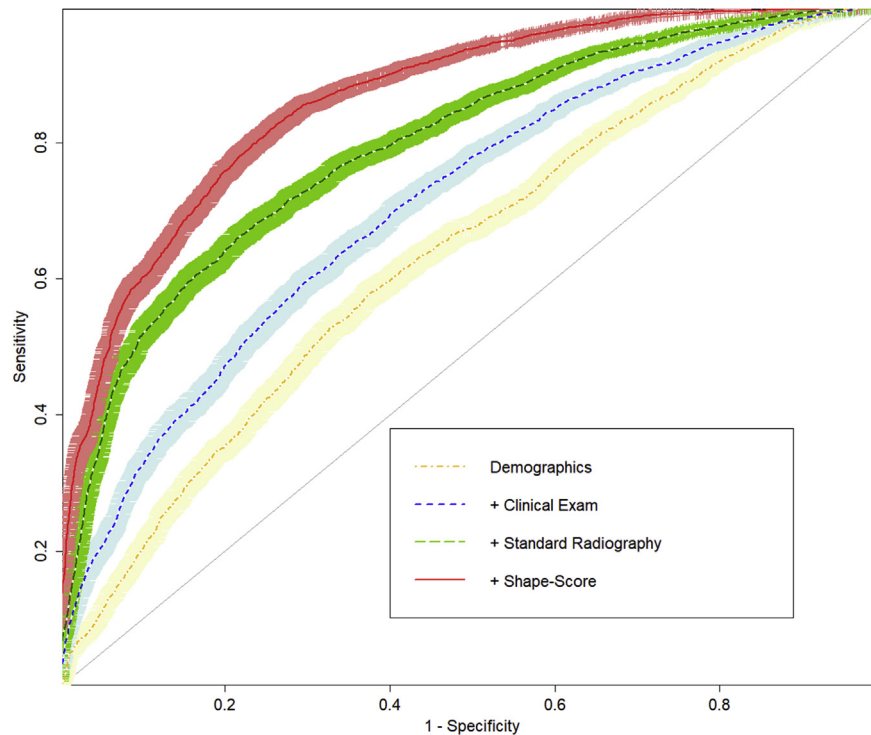
Adding the Shape-Score to the latter model improved the discriminative ability from an AUC of 0.795 to an AUC of 0.863. Adding the Shape-Score also improved the calibration of the prediction model ([Fig. 4](#)). The calibration curve is very close to the diagonal representing optimal fit, meaning that the predicted risk closely resembles the observed risk for rHOA or THR.

Calibration slopes and AUCs in internal-validation based on bootstrapped samples differed minimally from those in development, indicating that the predictive models are not overoptimistic (AUC 0.795 and 0.864 respectively ([Table II](#) and [Fig. 4](#))). The distributions of predicted risks resulting from each of the models show that adding the Shape-Score helps to stratify more medium-risk patients into low and high-risk categories ([Table III](#)). In the sensitivity analysis, using both logistic regression and GEE on hips with complete data only, AUC values were within 0.01 of the values found using the imputed datasets, and calibration plots were comparable.

## Discussion

We developed and internally validated models to predict incident rHOA or THR over 8 years in persons with first onset hip pain. Until now, no predictive model for HOA is widely used. We built a prediction model that combines innovative automated analysis of plain radiographs using machine learning, with clinical data that can easily be obtained ([Fig. 5](#)). The discriminative ability of the final model was high (AUC 0.863) given the relatively early stage of possible HOA at baseline and rHOA or THR as outcomes at 8-year follow-up.

In the literature, multiple prediction models are available for HOA. However, most are actually diagnostic models, aiming to diagnose HOA cross-sectionally, situated in a population or end-stage OA cohort. Saberi Hosnijeh *et al.* recently developed a prediction model for HOA in the Rotterdam cohort<sup>27</sup>. Their model uses demographics, urinary CTX-II levels and radiographic parameters including the Wiberg-angle and alpha-angle (to quantify acetabular coverage and cam morphology, respectively), but no parameters from the physical examination or SSM. Their model showed an AUC of 0.82 in the Rotterdam cohort and 0.71 when validated in CHECK. Furthermore, calibration in CHECK was far off the perfect slope, with observed risks being 2.5 times higher than predicted risks. Developing the model in a general population cohort and testing it in a target population with hip pain likely caused this. Our prediction model was developed in CHECK, which represents our target population, and includes parameters from physical examination and the Shape-Score. External validation was not performed as most cohorts focus on OA in later stages and/or do not have pelvic radiographs of sufficient quality available. Nevertheless,



**Fig. 3. Discrimination of the prediction models.** The ability to separate cases from non-cases, is visualized as area under the curve with the 95%-confidence interval. Sensitivity and specificity of the model improved for all cut-off points after adding the Shape-Score to the model.

internal bootstrap validation suggested that our model is not overoptimistic<sup>26</sup>. In the future, external validation should be performed, preferably in a cohort with symptomatic patients prone to HOA.

Strengths of this study include the use of a large prospective cohort with clinical complaints and inclusion criteria that allude early-stage knee and/or hip OA, with an adequate follow-up time and a sufficient number of incident rHOA or THR after 8 year (185) to test the 16 predictor candidates for the prediction models<sup>28</sup>. We used backward selection on clusters of predictors to mimic the flow of information in clinical care. While this may produce a slight reduction in absolute performance of the models compared to a fully data driven method, it improves the applicability of the models in clinical care and reduces the chance of overfitting. We tested the association between the Shape-Score and baseline clinical OA characteristics. The Shape-Score was related to hip OA characteristics but not knee OA characteristics ([Supplementary text](#)). Furthermore, follow-up data were rather complete and we used multiple imputation to reduce bias and increase precision of our analyses. Finally, we used optimism-adjustment methods throughout the development and validation of the models to reduce overfitting and overoptimistic results. By using bootstrap

validation instead of multi-fold cross-validation we used the data available more efficiently<sup>29</sup>.

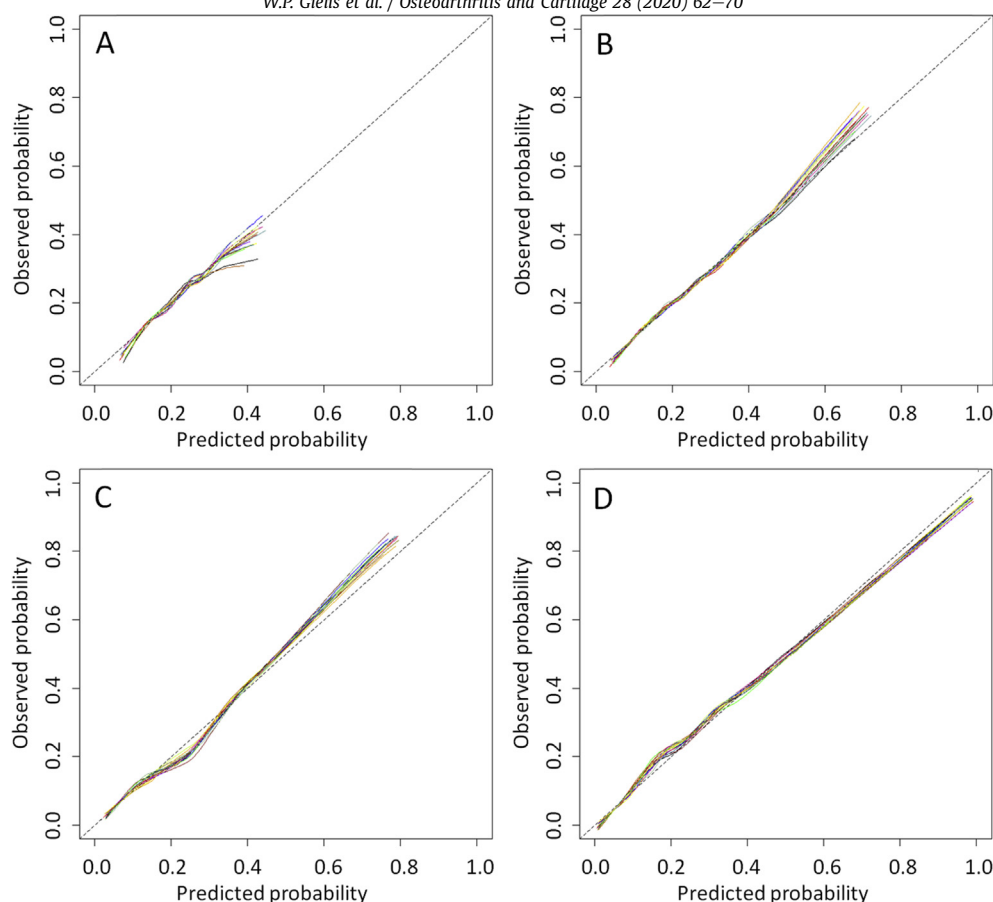
Combining rHOA (KL-grade  $\geq 2$ ) and THR as disease outcome may be debatable. The severity of clinical and radiographic symptoms correspond poorly in HOA, so that rHOA and THR may not always represent similar processes<sup>30</sup>. However, THR most often results from both clinical symptoms and radiographic signs. For a number of participants we included both hips. We used logistic regression analysis, which does not incorporate intra-participant correlation. However, GEE was not applicable in combination with the statistical packages used in the analysis, and mixed models regression had problems to converge when used on the available data. In the sensitivity analysis on the hips with complete data, the results between logistic regression analysis and GEE were very comparable.

The relationship between sex and HOA is less clear. In our models, female sex was initially associated with an increased risk of rHOA or THR, but with a decreased risk after adding the Shape-Score. This suggests that gender differences in OA risk may be related to hip shape. Higher BMI is a well-known risk factor for knee and hand OA, but its relationship with HOA is less clear<sup>31–33</sup>. In our models, a higher BMI had even a mild preventive effect for

**Table II**  
Discriminative ability

Model	AUC* in development	AUC* in validation	95%-confidence interval of AUC in validation
Demographics	0.634	0.635	(0.596–0.675)
Demographics and Clinical examination	0.710	0.710	(0.668–0.751)
Demographics, Clinical examination and Standard radiographic examination	0.795	0.795	(0.757–0.834)
Demographics, Clinical examination, Standard radiographic examination and Shape-Score	0.863	0.864	(0.833–0.895)
Shape-Score only	0.798	0.798	(0.762–0.833)

\* Area Under the Curve or C-statistic.



**Fig. 4. Calibration plots in validation.** Fig. 4 shows the predicted probabilities plotted against the observed outcomes in internal validation. This is used to assess the calibration for five different models. The striped black line represents a perfect match between predicted probabilities and observed outcomes, and thus perfect calibration. The dotted black line represents the calibration in training data. The colored lines each represent a different imputed dataset, and represents the mean calibration in validation, using a 1,000 bootstraps. A. Demographics. B. Demographic and clinical examination. C. Demographics, clinical and standard radiographic examination. D. Demographics, clinical, standard radiographic examination and Shape-Score produced using the fully automatic search model.

HOA. Education served as a proxy for socio-economic status in our study. Higher educated persons had a lower risk for HOA or THR, in line with literature<sup>34</sup>. Smoking did not predict rHOA or THR. Although some studies show a protective effect of smoking, this effect may be caused by selection bias<sup>35</sup>.

Pain in the hip area (groin/buttock/upper thigh) increased the risk for HOA, which is in correspondence with literature<sup>36</sup>. Pain around the knee sometimes directs a physician to search for a diagnosis in the knee only. However, hip OA should always be considered as a source of the pain<sup>37</sup>. Limited or painful internal rotation are clinical signs that suggest HOA and may predict THR<sup>36,38–40</sup>. In the present study, pain with/during internal rotation had a significant univariate relation with OA on follow-up (OR 1.7,

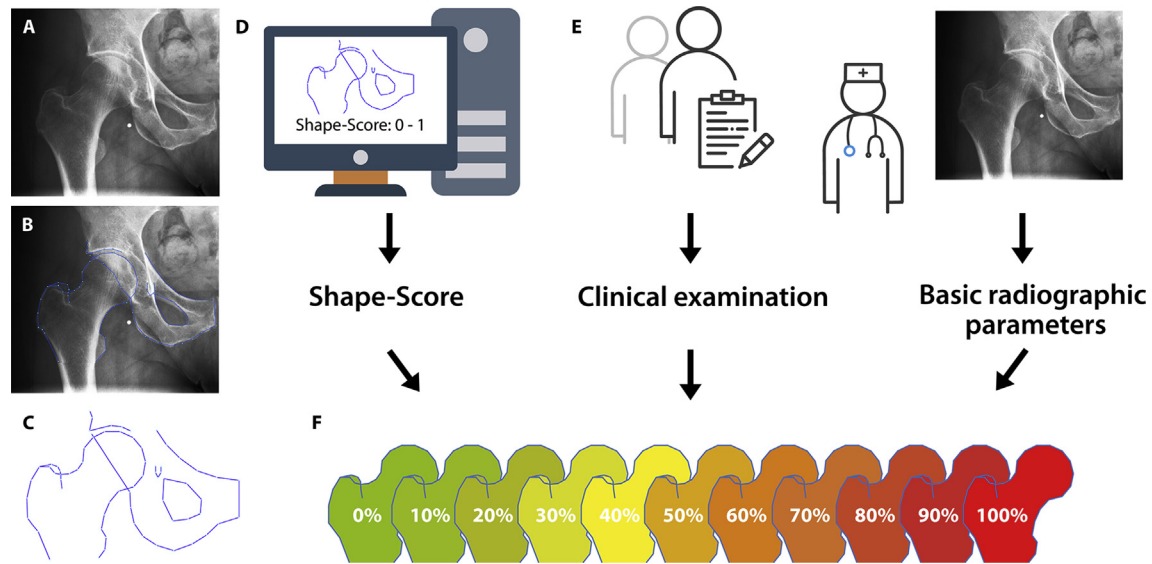
95% CI(1.3–2.3), but was eliminated from the prediction models as its  $p$ -value was 0.17. However, the range of internal rotation was included in the prediction models, perhaps overrunning the weaker predictive effect of pain on internal rotation. The WOMAC is a tool to measure pain, stiffness and physical functioning in patients with knee and/or hip OA. While it is widely used, its predictive value for incident rHOA is unknown<sup>41</sup>. In this study, baseline WOMAC score was associated with future rHOA or THR. Morning stiffness is included as a diagnostic criterion for HOA in the widely used Altman criteria for HOA, and showed a high sensitivity, but low specificity<sup>38</sup>. Its predictive value as a risk factor for HOA is doubtful and in the present study it did not add predictive value to the models, perhaps because morning stiffness is a non-specific

**Table III**

Distribution in risk categories with negative (NPV) and positive predictive value (PPV) for low risk (0–20%) and high risk (50–100%) categories

Model	0–20%	Low-Risk		20–50%	High-Risk		50–100%
	N (%)	NPV	PPV	N (%)	NPV	PPV	N (%)
Demographics	682 (60)	0.27	0.86	459 (40)	–	–	1 (0)
Demographics + Clinical examination	719 (63)	0.31	0.88	395 (34)	0.82	0.66	27 (2)
Demographics + Clinical examination + Basic radiographic parameters	768 (67)	0.40	0.91	288 (25)	0.85	0.67	86 (8)
Demographics + Clinical examination + Basic radiographic parameters + Shape-Score	800 (70)	0.48	0.93	215 (19)	0.87	0.71	126 (11)
Shape-Score	810 (71)	0.42	0.90	243 (21)	0.85	0.68	88 (8)

The numbers given are averages over 15 imputed datasets. For the low-risk category, absence of OA at 8-year follow-up was seen as a positive gold standard. PPV should be interpreted as the probability of not developing OA when being classified as low-risk (<20% chance). For the high-risk category, presence of OA at 8-year follow-up was seen as a positive gold standard. PPV should be interpreted as the probability of developing OA when being classified as high-risk (>50% chance).



**Fig. 5. Workflow to calculate personalized risk for future hip osteoarthritis.** A. A standard weight-bearing pelvic radiograph is made in the anteroposterior direction with 15-degrees internal rotation. B. The fully automatic segmentation system (FASS) annotates the anatomical landmarks on the radiograph. C. Statistical shape modelling (SSM) quantifies hip shape. D. The machine learning algorithm produces the Shape-Score, a single value representing the risk for incident HOA based on hip shape. E. Demographics, questionnaires, clinical examination, and basic radiographic parameters are assessed by trained personnel. F. Data from demographics, questionnaires, and physical examination and basic radiographic parameters are combined with the Shape-Score in easy to calculate formulas to produce accurate personalized risk for future hip osteoarthritis.

symptom<sup>27</sup>. Surprisingly, the use of analgesics was negatively associated with the future risk for rHOA or THR. Possibly, this is the case because analgesics are used more often in acute pain caused by transient disorders compared to the more elongated pain trajectory in OA.

JSN and osteophyte formation may be present before definite rHOA (defined as KL-grade  $\geq 2$ ) can be confirmed. These radiographic signs are known risk factors for progression to definite rHOA and indeed were strong risk factors in the present study<sup>27</sup>. Buttressing, thickening of the medial femoral neck, is a radiographic sign associated with rHOA<sup>42</sup>. The predictive value of buttressing has not been described before but it showed to be a significant risk factor for future rHOA or THR in this study. The Shape-Score may also include JSN and osteophyte formation, characteristics used to define KL-grades. However, the current study shows that the Shape-Score has added predictive values on top of traditional radiographic characters alone.

Our Shape-Score incorporates (i) cam morphology (Aspherical femoral head-neck junction) (ii) decreased acetabular depth, and (iii) a higher neck-shaft angle as risk factors for OA. Cam morphology, decreased acetabular depth and a higher neck-shaft angle have been shown to increase risk for OA in large cohorts before<sup>7,43–48</sup>.

Currently, it is challenging for clinicians to predict future hip OA in patients with early-stage joint pain that cannot be explained otherwise, especially for patients who don't have definite signs of OA on the radiograph. Some of these patients will develop OA, while other might have (had) hip pain for other reasons and will not develop OA. The proposed prediction model could help clinicians to optimally inform patients about their personalized future risk for disease and to choose appropriate treatment (intensity) and may boost treatment adherence.

In the future, the proposed FASS and prediction model could be integrated into a software package that can be linked to the electronic patient record (including PACS for radiographic images). This way, the Shape-Score could be derived fully automatically to assist clinicians in estimating the risk for future HOA. Of note is that the implementation of the proposed prediction model is not

dependent on time-consuming visual methods that may be subject to inter-/intra-observer variations. However, variation in positioning during image acquisition may cause differences in Shape-Score values and a standardized acquisition protocol, as used in the current cohort, is necessary. Unfortunately no data was available to assess the effect of.

Clinical trial inefficiency plays a major role in the current absence of disease-modifying OA drugs (DMOADs). By specifically selecting participants at an increased risk of incident OA, potential DMOADs have more potential to demonstrate detectable effects in a clinical trial. By adding the Shape-score or by using the Shape-Score only, we were able to stratify people for the risk of future rHOA or THR. When the Shape-Score is added to a screening with demographics, physical examination and basic radiographic parameters, 47% more patients could be stratified into the high-risk category (>50% risk), potentially improving screening efficiency.

We have developed an automatic Shape-Score tool, using machine learning algorithms, to optimally predict the risk for incident rHOA or THR based on hip shape as given by a pelvic radiograph. We demonstrated the added value of our Shape-Score in prediction models using easily obtainable parameters in persons with hip pain. Models including the Shape-Score had superior discriminative ability over models without and showed very good calibration. The Shape-Score may therefore prove to be a valuable tool for both patient care and research.

## Contributions

WG, WvS, RA, and CL contributed to the data collection. WG, PW, and CL contributed to the statistical analyses. All authors contributed to the study design and interpretation of the data. All authors contributed to the preparation of the manuscript and gave final approval of the current version for submission.

## Competing interests

Drs. Cootes and Lindner have a patent US 9928443, EP 2893491 issued.

### Role of funding sources

The CHECK-cohort study is funded by Reuma Nederland. 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. C. Lindner was funded by the Engineering and Physical Sciences Research Council, UK (EP/M012611/1) and by the Medical Research Council, UK (MR/S00405X/1). The current analysis was funded by Reuma Nederland (LLP-22) and the APPROACH project. APPROACH has received support from the Innovative Medicines Initiative Joint Undertaking under Grant Agreement n° 115770, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in kind contribution. See [www.imi.europa.eu](http://www.imi.europa.eu).

### Data availability

Data from CHECK is available from <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:63523>.

The FASS and SSM are available via <http://bone-finder.com/>  
FASS and SSM are available via <http://bone-finder.com/>

### Acknowledgements

The authors like to thank Chris van Kesteren for his work on the graphic design of Fig. 5.

### Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.joca.2019.09.005>.

### References

- Gay C, Chabaud A, Guilley E, Coudeyre E. Educating patients about the benefits of physical activity and exercise for their hip and knee osteoarthritis. Systematic literature review. *Ann Phys Rehabil Med* 2016;59(3):174–83, <https://doi.org/10.1016/j.rehab.2016.02.005>.
- Lessmann N, van Ginneken B, Zreik M, de Jong PA, de Vos BD, Viergever MA, et al. Automatic calcium scoring in low-dose chest CT using deep neural networks with dilated convolutions. *IEEE Trans Med Imaging* 2018;37(2):615–25, <https://doi.org/10.1109/TMI.2017.2769839>.
- Ciampi F, de Hoop B, van Riel SJ, Chung K, Scholten ET, Oudkerk M, et al. Automatic classification of pulmonary periferfissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. *Med Image Anal* 2015;26(1):195–202, <https://doi.org/10.1016/j.media.2015.08.001>.
- Ganz R, Leunig M, Leunig-Ganz K, Harris WH. The etiology of osteoarthritis of the hip: an integrated mechanical concept. *Clin Orthop Relat Res* 2008;466(2):264–72, <https://doi.org/10.1007/s11999-007-0060-z>.
- Agricola R, Heijboer MP, Bierma-Zeinstra SMA, Verhaar JA, Weinans H, Waarsing JH. Cam impingement causes osteoarthritis of the hip: a nationwide prospective cohort study (CHECK). *Ann Rheum Dis* 2013;72(6):918–23, <https://doi.org/10.1136/annrheumdis-2012-201643>.
- Agricola R, Waarsing JH, Arden NK, Carr AJ, Bierma-Zeinstra SMA, Thomas GE, et al. Cam impingement of the hip—a risk factor for hip osteoarthritis. *Nat Rev Rheumatol* 2013;9(10):630–4, <https://doi.org/10.1038/nrrheum.2013.114>.
- Agricola R, Leyland KM, Bierma-Zeinstra SMA, Thomas GE, Emans PJ, Spector TD, et al. Validation of statistical shape modelling to predict hip osteoarthritis in females: data from two prospective cohort studies (Cohort Hip and Cohort Knee and Chingford). *Rheumatology* 2015;54(11):2033–41, <https://doi.org/10.1093/rheumatology/kev232>.
- Ahedi HG, Aspden RM, Blizzard LC, Saunders FR, Cicuttini FM, Aitken DA, et al. Hip shape as a predictor of osteoarthritis progression in a prospective population cohort. *Arthritis Care Res (Hoboken)*. 2016;69(10):1566–73, <https://doi.org/10.1002/acr.23166>.
- Barr RJ, Gregory JS, Reid DM, Aspden RM, Yoshida K, Hosie G, et al. Predicting OA progression to total hip replacement: can we do better than risk factors alone using active shape modelling as an imaging biomarker? *Rheumatology (Oxford)* 2012;51(3):562–70, <https://doi.org/10.1093/rheumatology/ker382>.
- Castañó-Betancourt MC, Van Meurs JBJ, Bierma-Zeinstra S, Rivadeneira F, Hofman A, Weinans H, et al. The contribution of hip geometry to the prediction of hip osteoarthritis. *Osteoarthr Cartil* 2013;21(10):1530–6, <https://doi.org/10.1016/j.joca.2013.06.012>.
- Lindner C, Thiagarajah S, Wilkinson JM, Wallis Ga, Cootes TF. Accurate bone segmentation in 2D radiographs using fully automatic shape model matching based on regression-voting. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 2013;8150 LNCS(PART 2):181–9, [https://doi.org/10.1007/978-3-642-40763-5\\_23](https://doi.org/10.1007/978-3-642-40763-5_23).
- Lindner C, Thiagarajah S, Wilkinson JM, Wallis Ga, Cootes TF. Development of a fully automatic shape model matching (FASMM) system to derive statistical shape models from radiographs: application to the accurate capture and global representation of proximal femur shape. *Osteoarthr Cartil* 2013;21(10):1537–44, <https://doi.org/10.1016/j.joca.2013.08.008>.
- Wesseling J, Boers M, Viergever MA, Hilberdink WKHA, Lafeber FPJG, Dekker J, et al. Cohort profile: cohort hip and cohort knee (CHECK) study. *Int J Epidemiol* 2016;45(1):36–44, <https://doi.org/10.1093/ije/dyu177>.
- Norkin Cynthia C, Joyce White D. Measurement of Joint Motion A Guide to Goniometry. 4th edn. 2009 <https://www.fadavis.com/product/physical-therapy-measurement-joint-motion-goniometry-norkin-white-5>.
- Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthr Cartil* 2007;15(Suppl A):A1–A56, <https://doi.org/10.1016/j.joca.2006.11.009>.
- Damen J, Schiphof D, Wolde STen, Cats HA, Bierma-Zeinstra SMA, Oei EHG. Inter-observer reliability for radiographic assessment of early osteoarthritis features: the CHECK (cohort hip and cohort knee) study. *Osteoarthr Cartil* 2014;22(7):969–74, <https://doi.org/10.1016/j.joca.2014.05.007>.
- Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models—their training and application. *Comput Vis Image Understand* 1995;61(1):38–59, <https://doi.org/10.1006/cviu.1995.1004>.
- Agricola R, Reijman M, Bierma-Zeinstra SMA, 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). *Osteoarthr Cartil* 2013;21(4):1–6, <https://doi.org/10.1016/j.joca.2013.01.005>.
- Ganz R, Parvizi J, Beck M, Leunig M, Nötzli H, Siebenrock K a. Femoroacetabular impingement: a cause for osteoarthritis of the hip. *Clin Orthop Relat Res* 2003;417:112–20, <https://doi.org/10.1097/01.blo.0000096804.78689.c2>.

20. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis* 1957;16(4):494–502, <http://www.scopus.com/inward/record.url?eid=2-s2.0-62949246527&partnerID=tZOtx3y1>.
21. Van Buuren S, Groothuis-Oudshoorn K. Multivariate imputation by chained equations. *J Stat Softw* 2011;45(3):1–67, <https://doi.org/10.1177/0962280206074463>.
22. Heymans MW, van Buuren S, Knol DL, van Mechelen W, de Vet HCW. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Med Res Methodol* 2007;7:33, <https://doi.org/10.1186/1471-2288-7-33>.
23. Harrell FE. In: *Regression Modeling Strategies* 2015;vol. 64, <https://doi.org/10.1007/978-1-4757-3462-1>.
24. Rubin DB. Inference and missing data. *Biometrika* 1976;63(3):581–92, <https://doi.org/10.1093/biomet/63.3.581>.
25. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21(1):128–38, <https://doi.org/10.1097/EDE.0b013e3181c30fb2>.
26. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003;56(5):441–7, [https://doi.org/10.1016/S0895-4356\(03\)00047-7](https://doi.org/10.1016/S0895-4356(03)00047-7).
27. Saberi Hosnijeh F, Kavousi M, Boer CG, Uitterlinden AG, Hofman A, Reijman M, et al. Development of a prediction model for future risk of radiographic hip osteoarthritis. *Osteoarthr Cartil* 2018;26(4):540–6, <https://doi.org/10.1016/j.joca.2018.01.015>.
28. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49(12):1373–9, [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3).
29. Steyerberg EW, Harrell FE, Borsboom GJJ, Eijkemans MJ, Vergouwe Y, Habbema JDF. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54(8):774–81, [https://doi.org/10.1016/S0895-4356\(01\)00341-9](https://doi.org/10.1016/S0895-4356(01)00341-9).
30. Kinds MB, Welsing PMJ, Vignon EP, Bijlsma JWJ, Viergever MA, Marijnissen ACA, et al. A systematic review of the association between radiographic and clinical osteoarthritis of hip and knee. *Osteoarthr Cartil* 2011;19(7):768–78, <https://doi.org/10.1016/j.joca.2011.01.015>.
31. Reyes C, Leyland KM, Peat G, Cooper C, Arden NK, Prieto-Alhambra D. Association between overweight and obesity and risk of clinically diagnosed knee, hip, and hand osteoarthritis: a population-based cohort study. *Arthritis Rheum* 2016;68(8):1869–75, <https://doi.org/10.1002/art.39707>.
32. Jiang L, Rong J, Wang Y, Hu F, Bao C, Li X, et al. The relationship between body mass index and hip osteoarthritis: a systematic review and meta-analysis. *Jt Bone Spine* 2011;78(2):150–5, <https://doi.org/10.1016/j.jbspin.2010.04.011>.
33. Grotle M, Hagen KB, Natvig B, Dahl FA, Kvien TK. Obesity and osteoarthritis in knee, hip and/or hand: an epidemiological study in the general population with 10 years follow-up. *BMC Musculoskelet Disord* 2008;9:132, <https://doi.org/10.1186/1471-2474-9-132>.
34. Reyes C, Garcia-Gil M, Elorza JM, Mendez-Boo L, Hermosilla E, Javaid MK, et al. Socio-economic status and the risk of developing hand, hip or knee osteoarthritis: a region-wide ecological study. *Osteoarthr Cartil* 2015;23(8):1323–9, <https://doi.org/10.1016/j.joca.2015.03.020>.
35. Hui M, Doherty M, Zhang W. Does smoking protect against osteoarthritis? Meta-analysis of observational studies. *Ann Rheum Dis* 2011;70(7):1231–7, <https://doi.org/10.1136/ard.2010.142323>.
36. Lieveense AM, Koes BW, Verhaar J a N, Bohnen AM, Bierma-Zeinstra SM a. Prognosis of hip pain in general practice: a prospective followup study. *Arthritis Rheum* 2007;57(8):1368–74, <https://doi.org/10.1002/art.23094>.
37. Dibra FF, Prieto HA, Gray CF, Parvataneni HK. Don't forget the hip! Hip arthritis masquerading as knee pain. *Arthroplast Today* 2018;4(1):118–24, <https://doi.org/10.1016/j.ARTD.2017.06.008>.
38. Sutlive TG, Lopez HP, Schnitker DE, Yawn SE, Halle RJ, Mansfield LT, et al. Development of a clinical prediction rule for diagnosing hip osteoarthritis in individuals with unilateral hip pain. *J Orthop Sport Phys Ther* 2008;38(9):542–50, <https://doi.org/10.2519/jospt.2008.2753>.
39. Birrell F, Croft P, Cooper C, Hosie G, Macfarlane G, Silman A. Predicting radiographic hip osteoarthritis from range of movement. *Rheumatology* 2001;40(5):506–12, <https://doi.org/10.1093/rheumatology/40.5.506>.
40. Holla JFM, Van Der Leeden M, Roorda LD, Bierma-Zeinstra SM, Damen J, Dekker J, et al. Diagnostic accuracy of range of motion measurements in early symptomatic hip and/or knee osteoarthritis. *Arthritis Care Res* 2012;64(1):59–65, <https://doi.org/10.1002/acr.20645>.
41. Bellamy N. *WOMAC Osteoarthritis Index User Guide*. Brisbane, Aust: Version V; 2002.
42. Dixon T, Benjamin J, Lund P, Graham A, Krupinski E. Femoral neck buttressing: a radiographic and histologic analysis. *Skelet Radiol* 2000;29(10):587–92, <https://doi.org/10.1007/s002560000260>.
43. Thomas GER, Palmer AJR, Batra RN, Kiran A, Hart D, Spector T, et al. Subclinical deformities of the hip are significant predictors of radiographic osteoarthritis and joint replacement in women. A 20 year longitudinal cohort study. *Osteoarthr Cartil* 2014;22(10):1504–10, <https://doi.org/10.1016/j.joca.2014.06.038>.
44. Lane NE, Lin P, Christiansen L, Gore LR, Williams EN, Hochberg MC, 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, [https://doi.org/10.1002/1529-0131\(200002\)43:2<400::AID-ANR21>3.0.CO;2-D](https://doi.org/10.1002/1529-0131(200002)43:2<400::AID-ANR21>3.0.CO;2-D).
45. Murphy NJ, Eyles JP, Hunter DJ. Hip osteoarthritis: etiopathogenesis and implications for management. *Adv Ther* 2016;33(11):1921–46, <https://doi.org/10.1007/s12325-016-0409-3>.
46. Doherty M, Courtney P, Doherty S, Jenkins W, Maciewicz RA, Muir K, et al. Nonspherical femoral head shape (pistol grip deformity), neck shaft angle, and risk of hip osteoarthritis: a case-control study. *Arthritis Rheum* 2008;58(10):3172–82, <https://doi.org/10.1002/art.23939>.
47. Ranawat AS, Schulz B, Baumbach SF, Meftah M, Ganz R, Leunig M. Radiographic predictors of hip pain in femoroacetabular impingement. *HSS J* 2011;7(2):115–9, <https://doi.org/10.1007/s11420-010-9192-x>.
48. Geoffry Ng KC, Lamontagne M, Beaulé PE. Differences in anatomical parameters between the affected and unaffected hip in patients with bilateral cam-type deformities. *Clin Biomech* 2016;33:13–9, <https://doi.org/10.1016/j.clinbiomech.2016.01.007>.