



METHOD ARTICLE

**REVISED** **Deriving alpha angle from anterior-posterior dual-energy x-ray absorptiometry scans: an automated and validated approach [version 2; peer review: 2 approved]**

Benjamin G. Faber <sup>1,2</sup>, Raja Ebsim<sup>3</sup>, Fiona R. Saunders <sup>4</sup>, Monika Frysz <sup>1,2</sup>, George Davey Smith <sup>2</sup>, Timothy Cootes<sup>3</sup>, Jonathan H. Tobias <sup>1,2</sup>, Claudia Lindner <sup>3</sup>

<sup>1</sup>Musculoskeletal Research Unit, University of Bristol, Bristol, UK

<sup>2</sup>Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK

<sup>3</sup>Division of Informatics, Imaging and Data Sciences, University of Manchester, Manchester, UK

<sup>4</sup>Centre for Arthritis and Musculoskeletal Health, University of Aberdeen, Aberdeen, UK

**V2** First published: 15 Mar 2021, 6:60  
<https://doi.org/10.12688/wellcomeopenres.16656.1>  
 Latest published: 19 Jul 2022, 6:60  
<https://doi.org/10.12688/wellcomeopenres.16656.2>

### Abstract

**Introduction:** Alpha angle (AA) is a widely used imaging measure of hip shape that is commonly used to define cam morphology, a bulging of the lateral aspect of the femoral head. Cam morphology has shown strong associations with hip osteoarthritis (OA) making the AA a clinically relevant measure. In both clinical practice and research studies, AA tends to be measured manually which can be inconsistent and time-consuming.

**Objective:** We aimed to (i) develop an automated method of deriving AA from anterior-posterior dual-energy x-ray absorptiometry (DXA) scans; and (ii) validate this method against manual measures of AA.

**Methods:** 6,807 individuals with left hip DXAs were selected from UK Biobank. Outline points were manually placed around the femoral head on 1,930 images before training a Random Forest-based algorithm to place the points on a further 4,877 images. An automatic method for calculating AA was written in Python 3 utilising these outline points. An iterative approach was taken to developing and validating the method, testing the automated measures against independent batches of manually measured images in sequential experiments.

**Results:** Over the course of six experimental stages the concordance correlation coefficient, when comparing the automatic AA to manual measures of AA, improved from 0.28 [95% confidence interval 0.13-0.43] for the initial version to 0.88 [0.84-0.92] for the final version. The

### Open Peer Review

Approval Status

	1	2
<b>version 2</b> (revision) 19 Jul 2022	 view	
<b>version 1</b> 15 Mar 2021	 view	 view

1. **Pim van Klij** , Isala, Zwolle, The Netherlands

2. **James Masters** , University of Oxford, Oxford, UK

Any reports and responses or comments on the article can be found at the end of the article.

inter-rater kappa statistic comparing automatic versus manual measures of cam morphology, defined as  $AA^3 \geq 60^\circ$ , improved from 0.43 [80% agreement] for the initial version to 0.86 [94% agreement] for the final version.

**Conclusions:** We have developed and validated an automated measure of AA from DXA scans, showing high agreement with manually measuring AA. The proposed method is available to the wider research community from [Zenodo](#).

### Keywords

Alpha Angle, DXA, Automation, Cam, Femoro-acetabular impingement

**Corresponding author:** Benjamin G. Faber ([ben.faber@bristol.ac.uk](mailto:ben.faber@bristol.ac.uk))

**Author roles:** **Faber BG:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Ebsim R:** Data Curation, Investigation, Methodology, Software, Supervision, Writing – Review & Editing; **Saunders FR:** Data Curation, Formal Analysis, Project Administration, Validation, Writing – Review & Editing; **Frysz M:** Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Writing – Review & Editing; **Davey Smith G:** Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Supervision, Writing – Review & Editing; **Cootes T:** Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Software, Supervision, Writing – Review & Editing; **Tobias JH:** Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Supervision, Writing – Review & Editing; **Lindner C:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** RE, MF, FS are supported by, and this work is funded by a Wellcome Trust collaborative award (209233). BGF is supported by a Medical Research Council (MRC) clinical research training fellowship (MR/S021280/1). BGF, MF, JHT, GDS work in the MRC Integrative Epidemiology Unit at the University of Bristol, which is supported by the MRC (MC\_UU\_00011/1). CL was funded by the MRC, UK (MR/S00405X/1).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2022 Faber BG *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Faber BG, Ebsim R, Saunders FR *et al.* **Deriving alpha angle from anterior-posterior dual-energy x-ray absorptiometry scans: an automated and validated approach [version 2; peer review: 2 approved]** Wellcome Open Research 2022, 6:60 <https://doi.org/10.12688/wellcomeopenres.16656.2>

**First published:** 15 Mar 2021, 6:60 <https://doi.org/10.12688/wellcomeopenres.16656.1>

**REVISED Amendments from Version 1**

We thank the reviewers for their comments and suggestions. Taking these into account, we have revised the manuscript as follows. We have added further detail regarding how self-reported osteoarthritis status was defined and ethical approval gained. We have improved the labelling on [Figure 1](#) and [Figure 2](#). We have addressed in the discussion the finding that the automatically obtained AA measures are slightly higher AA than the manual ones. We have added further information in the limitations section regarding the inclusion of left hip DXAs only and the positioning of participants for these scans. Finally, we have added information and a new citation discussing work validating these methods on hip AP radiographs.

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

Alpha angle (AA) is a measure designed to examine the presence and severity of cam morphology at the hip joint<sup>1</sup>. Cam morphology describes a bulging of the lateral aspect of the femoral head that causes the femoral head to become aspherical leading to a pistol grip type appearance; it is a key component of femoro-acetabular impingement (FAI)<sup>2,3</sup>. AA is the angle measured between two lines, the first line from the mid-point of the femoral neck to the centre of the femoral head, and a second line from the centre of the femoral head to a point on the femoral head where the femoral head or neck leaves a circle of best fit placed over the femoral head ([Figure 1](#))<sup>1,4</sup>. The higher the AA the more indicative of cam morphology it is – with previously published thresholds of 50°, 55°, 60° and 83° all being used to define the presence of cam morphology<sup>1,4-7</sup>, and with a recent systematic review concluding a threshold of 60° is best<sup>8</sup>.

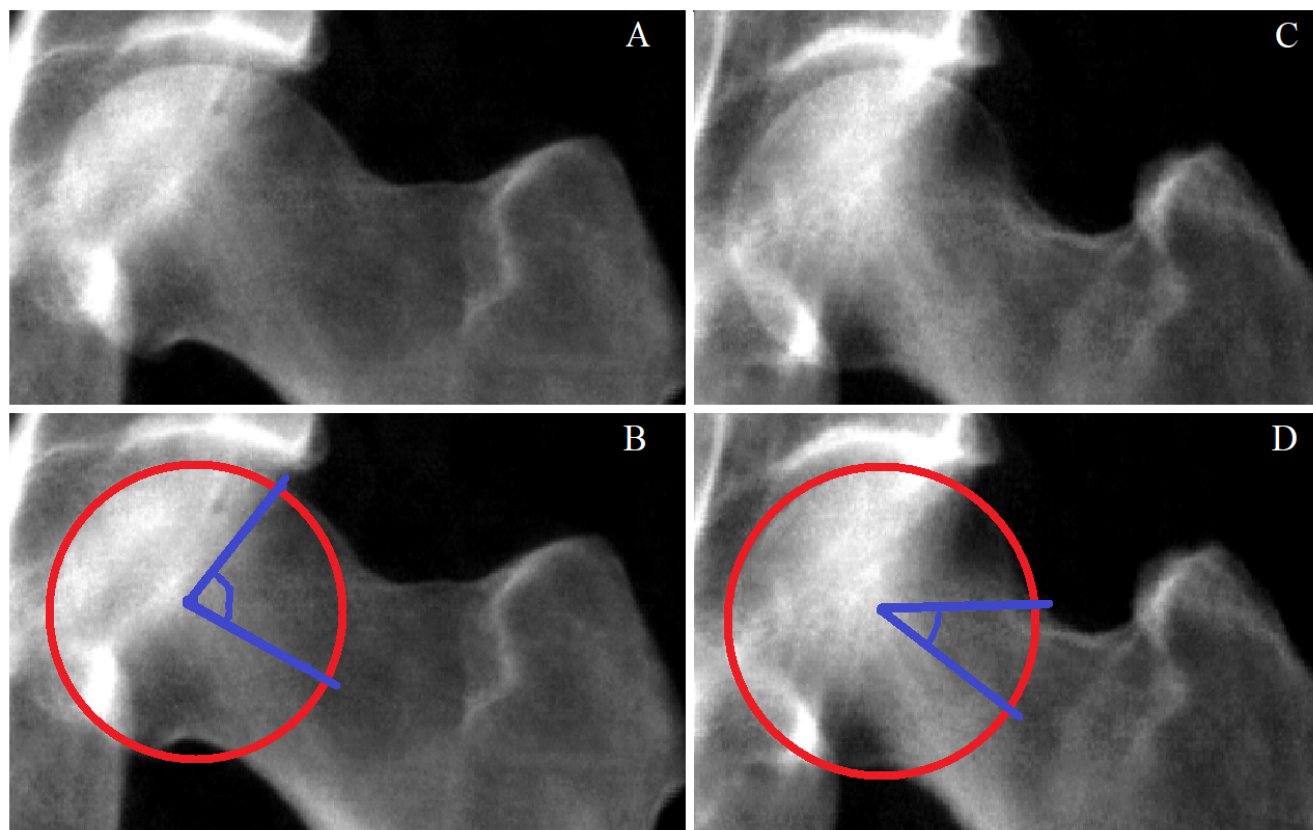
Cam morphology is an important shape variation of the femoral head. It has been associated with hip osteoarthritis (OA) and subsequent total hip replacement (THR), a proxy for end-stage OA<sup>4,9</sup>. Cam morphology is thought to lead to pain and limited function of the hip in FAI syndrome<sup>10</sup>. AA has been used in clinical trials as inclusion criteria to investigate the use of surgical procedures to treat FAI syndrome, but no agreed standardised approach of measuring AA exists<sup>7,11</sup>. AA was first developed utilising magnetic resonance imaging (MRI) before being applied to anterior-posterior (AP) and lateral radiographs for large-scale epidemiological analyses<sup>1,6</sup>. One approach to manually measuring AA on AP hip radiographs is to use software such as HipMorf or OxMorf<sup>12-14</sup>. These packages allow the user to manually fit a circle over the femoral head and then mark where the femoral head leaves this. Alternatively, AA has been calculated using MATLAB based software which utilises outline points that have been placed manually around the femoral head<sup>4,9</sup>. AA derived from outline points has been shown to be predictive of hip OA<sup>9</sup>. When using automatically placed points, concerns have been raised about the validity and reproducibility of such an outline points-based approach due to the difficulty in deciding where exactly the

femoral head deviates from the circle of best fit<sup>15</sup>. To date, no reproducibility studies on methods for automatically measuring AA have been published nor is any open source code to do this available.

Dual-energy x-ray absorptiometry (DXA) scans are commonly used to derive measures of bone mineral density, and are increasingly being used to look at hip shape<sup>16,17</sup>. In addition, a new generation of hip DXA scanner allows for higher resolution images which has led to the use of DXA scans in detecting radiographic OA for research purposes<sup>18,19</sup>. Compared to radiographs, DXA scans involve lower radiation doses and are available from larger population studies, such as the UK Biobank (UKB) enhanced imaging study<sup>20,21</sup>. This work aimed to (i) develop a method to automatically derive AA from outline points placed around the femoral head in DXA scans obtained in UKB; and (ii) validate the method against manual measures of AA. We also compare values obtained using our method to previously published population level statistics.

**Methods****Population**

UKB is a UK-based mixed sex cohort made up of 500,000 individuals aged between 40–69 years at recruitment (2006–2010)<sup>22</sup>. A full data catalogue is available [online](#). A subset of 100,000 individuals are due to have high resolution iDXA scans done on both hips (2013 – ongoing) with over 45,000 already completed<sup>20</sup>. An initial training sample of 2,000 individuals with a DXA scan was selected but 70 DXA scans were excluded due to poor image quality leaving a training sample of 1,930 individuals. A further extension sample of 5,000 individuals with a DXA scan was selected but 123 DXA scans were excluded due to poor image quality leaving an extension sample of 4,877 individuals. The training and extension samples were selected from an overall sample of 13,496 individuals with DXA scans available at the time (January 2019). The first 20% of both the training and extension samples were selected randomly from those with a self-reported diagnosis of OA based on a questionnaire completed at the same visit as the DXA scan. In the questionnaire, each participant was asked to record all their known diagnoses and a trained nurse subsequently coded the diagnosis. The OA code was not site specific. This was done to increase the number of pathological scans in the training sample as part of a wider research programme to automate the assessment of radiographic osteoarthritis. It was this wider research programme that guided the sample size selection as large samples are needed for machine learning. The remainder of the training and extension samples (80%) were selected randomly ensuring the sexes were equally weighted<sup>19</sup>. The combined sample is made up of both the training and extension sample. All demographic information was obtained on the same day as the DXA scan. This study was approved by UKB (application number 17295) which is overseen by its own Ethics Advisory Committee. In addition, UKB received approval from the National Information Governance Board for Health and Social Care and Northwest Multi-Centre Research Ethics Committee (11/NW/0382). All participants provided informed consent.



**Figure 1. Alpha angle depicted on UK Biobank hip dual-energy absorptiometry scans.** **A** A representative UK Biobank (UKB) hip dual-energy x-ray absorptiometry (DXA) scan depicting a femoral head with cam morphology. **B** The same DXA as in A marked with a red circle of best fit plotted over the femoral head. The two blue lines illustrate the lines from which the alpha angle (AA) is calculated; one line goes from the middle of the femoral neck to the centre of the circle and the other goes from the centre of the circle to the point at which the femoral head leaves the circle of best fit. **C** A UKB hip DXA without cam morphology. **D** The same DXA as in C marked with a red circle of best fit and blue lines from which the AA is calculated. Bottom left image: The same DXA as above marked with a red circle of best fit plotted over the femoral head. The two blue lines illustrate the lines from which the alpha angle (AA) is calculated; one line goes from the middle of the femoral neck to the centre of the circle and the other goes from the centre of the circle to the point at which the femoral head leaves the circle of best fit. Top right image: A UKB hip DXA without cam morphology. Bottom right image: The same DXA as above marked with a red circle of best fit and blue lines from which the AA is calculated.

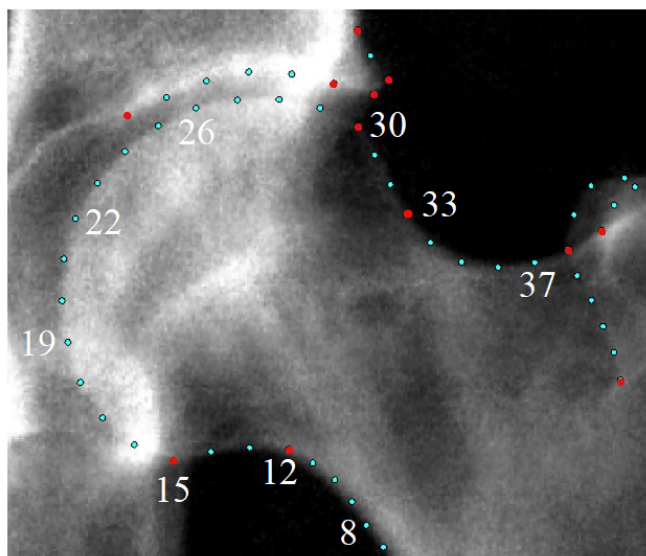
### DXA images and outline points

As part of UKB, DXA scans of both hips (iDXA GE-Lunar, Madison, WI) were obtained from participants positioned supine with 15–25° internal rotation using a standardised protocol<sup>21</sup>. In this study, we only examined the left hip DXA scans. All DXAs in the training sample had 85 outline points positioned around the femoral head, metaphysis, lesser and greater trochanters, and the superior acetabulum by four manual annotators. Of the 85 points, 18 points were placed on anatomical landmarks (key points) and the remaining points were placed equidistant apart along the edge of the bone. A Random Forest-based machine learning algorithm was then trained on these images and used to automatically annotate the extension sample with the 85 outline points<sup>23,24</sup>. All automatically placed points were checked and manually corrected where necessary. The mean correction distance was 0.7mm (movement orthogonal to bone boundary: 0.1mm) with the majority of points remaining unchanged. When osteophytes were present the outline points were moved manually inside of the osteophyte (if not already correct)

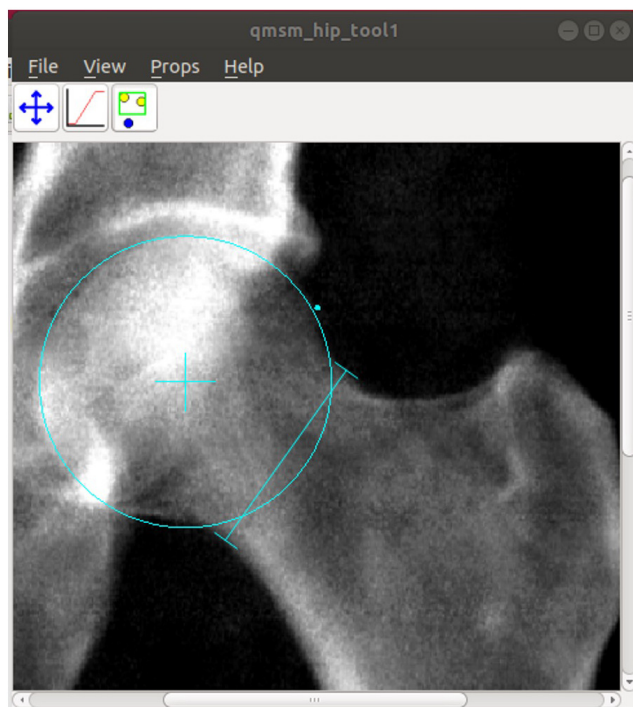
to avoid including osteophytes in our AA. Of the 85 outline points only points 8 to 39 along the femoral head and neck were used in this study to derive the AA measurements (Figure 2). For each image, all point positions were stored as x, y coordinates in a text file.

### Manual measure of alpha angle

To provide a manual ‘gold-standard,’ against which to test the automated method, AA was measured manually by BF, a rheumatology doctor, for a random selection of images divided into 5 batches of 100 (n=400 from the training sample and n=100 from the expansion sample). This was done using custom software (University of Manchester) that allows the user to manually (i) place and scale a circle to best fit the femoral head; (ii) place a point where the femoral head leaves the circle; and (iii) position callipers across the narrowest section of the femoral neck (Figure 3). The software saves the centre point of the circle, the midpoint of the narrowest section of the femoral neck, and the point at which the femoral head leaves the



**Figure 2.** A representative dual-energy x-ray absorptiometry (DXA) scan from UK Biobank (UKB) with all femoral head and acetabular outline points. The points marked in red are key anatomical landmarks. Only points 8–39 were used in this study. The acetabular points and the remaining proximal femur points were not necessary.



**Figure 3.** A screenshot of customised software from the University of Manchester. A circle is manually scaled and moved to fit the circle of best fit to the femoral head. A point is placed where the femoral head leaves the circle (cyan dot). Callipers are positioned across the narrowest section of the femoral neck. The software saves the centre point of the circle (cyan cross), the narrowest point of the femoral neck, and the point at which the femoral head deviates from the circle of best fit. The manual AA is then calculated from these points.

circle. The manual AA was then calculated from these points using a custom Python 3 script. Intra-rater variability was assessed on a subset of 100 scans, repeating the measurements 9 months after they were initially obtained.

#### Automated measure of alpha angle - model refinement approach

We followed a sequential experimental design to test and iteratively optimise our automatic AA calculation based on comparisons with manually derived measures. Initially, a model was designed to derive automatic AA (*model 1*) and was tested on 100 images randomly selected from the training sample with manually placed outline points. Refinements were made to the way AA was automatically measured (*models 2–4*) and are detailed in the following methods. To test each model iteration a subset of 100 images was analysed, with AA measured both manually and automatically in a blinded fashion. The method to automatically measure AA was finalised using subsets of the training sample with manually placed outline points resulting in *model 4*. Following this, a final evaluation was done to analyse the performance of the method when using automatically placed but manually corrected outline points using the extension sample.

#### Model 1 - defining a circle of best fit

All point position text files were read into Python 3 using [Jupyter Notebook](#). Within Python 3, a freely available least-squares regression model package (*circle\_fit*) was used to place a circle of best fit to points 15 and 28 on the femoral head ([Figure 2](#))<sup>25</sup>. The *circle\_fit* output gives the *x* and *y* coordinates of the centre of the circle, the radius of the circle and the variance (the variance was not used). Point 15 marked the inferior medial curvature of the femoral head and was chosen to be the starting point of the circle fitting. Point 30 marked the superior lateral curvature of the femoral head. However, instead of point 30, point 28 (i.e. two points medial of point 30) was chosen to be the end point of the circle fitting to avoid overfitting the circle of best fit to cam-type femoral heads. This method of circle fitting was manually qualitatively assessed on 500 DXA scans and deemed appropriate.

#### Model 1 - defining the femoral neck mid-point

Finding the narrowest point of the femoral neck was done using a line-segment approach. The femoral neck was demarcated by points 8–12 for the medial side and points 32–36 on the lateral side ([Figure 2](#)). For these two sets of points, a straight line was constructed between each pair of consecutive points. For each straight line segment, the shortest distance was measured between that line and a point on the opposing side of the femoral neck. For example, a line would be drawn between points 8 & 9 and the shortest distance may be found between this line and point 35. The shortest distance across all line segments defined the narrowest width of the femoral neck; the mid-point on this line is calculated and saved as the femoral neck point.

#### Model 1 - defining the index point

The index point is referred to as the first outline point judged to be truly 'outside' of the circle of best fit. It is critical to

defining the intersection position, the coordinates at which the femoral head or neck leaves the circle which is the key element for calculating AA. For identifying the index point, we defined the residual as the distance between each point (from points 15–28) and the centre of the circle (distance to centre) minus the radius of the circle. For each image, the maximum residual is the index point threshold for that image. For spherical femoral heads the index point threshold will be lower compared to aspherical femoral heads which are harder to fit a circle to. The index point was defined to be the first point after point 28 which deviated from the circle by more than the index point threshold (i.e. the maximum residual) with the point afterwards (in clockwise direction) also having a residual greater than that of the index point. For example, if point 30 deviated from the circle by more than the maximum residual and point 31 also deviated from the circle by more than point 30's residual then point 30 would be the index point. In contrast, if the residual of point 31 was smaller than or equal to that of point 30 then point 30 would not be considered to have left the circle and therefore not be the index point.

#### Model 1 - defining the intersection position

Once the index point has been identified, the intersection position can be calculated. The calculation of the intersection position depends on whether the outline point preceding the index point lies inside or outside of the circle of best fit. In the former case the intersection position is defined by the coordinates at which a line between the index point and the preceding point crosses the circle (Figure 4). If the outline point preceding the index point lies outside of the circle, but within the index point threshold, then there is no clear intersection position (Figure 5). In this case, the intersection position is approximated to be the outline point before the index point. This is also the case if the line between the index point and

preceding point is a tangent to the circle where again there is no clear intersection position.

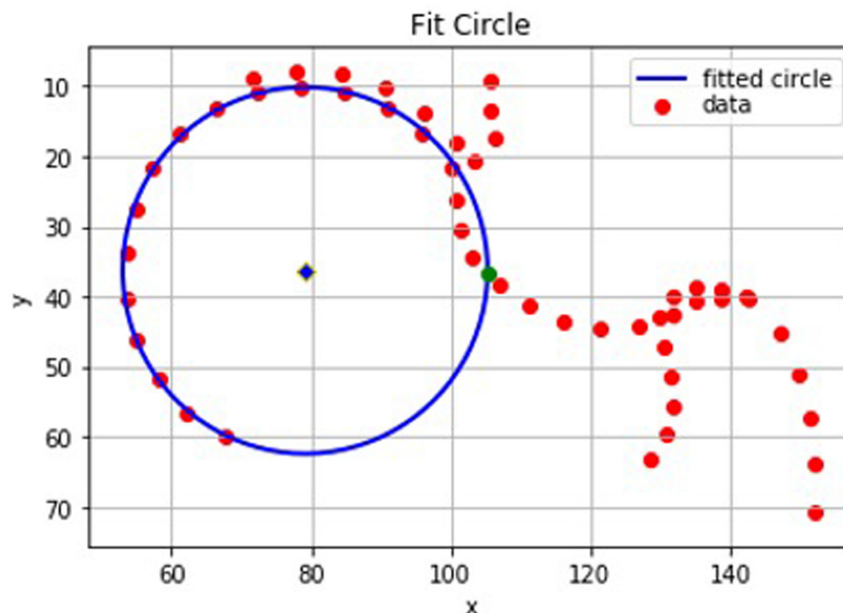
#### Model 2 – index point definition change

Initial testing showed that the index point definition was too sensitive in *model 1*, identifying outline points as having left the circle which by manual visual inspection appeared to be on the circle. To improve performance and decrease the ‘false positive rate’ of cam detection, the index point definition was changed in *model 2*, to now require three consecutive points leaving the circle by increasing residual values above the index point threshold (Figure 6). *Model 2* was no different to *model 1* with regards to the other key elements: circle fitting, femoral neck midpoint and intersection position.

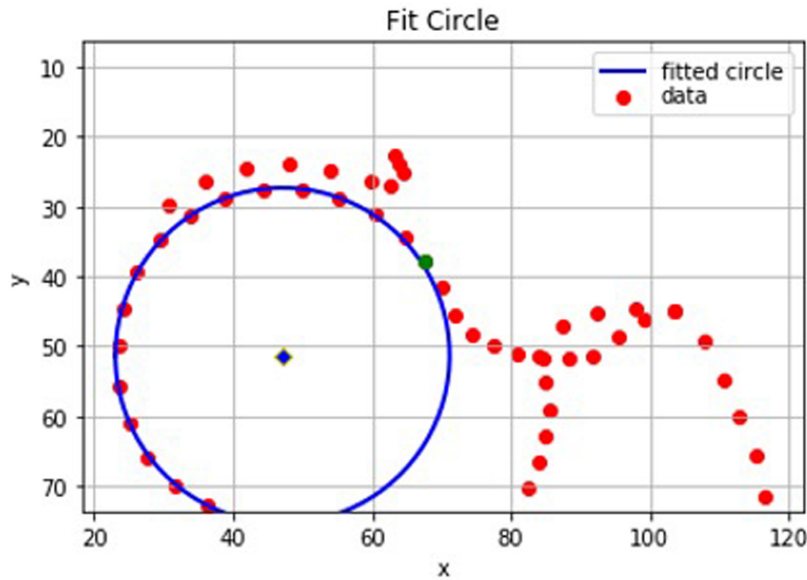
Testing the performance of *model 2* against manual measures in 100 randomly selected images showed good agreement but there were only two images with manually classified cam morphology in this sample. This led to the suspicion that the high percentage agreement (see *Results*) achieved in this experiment may be due to the sample being a poor testing set. *Model 2* was then tested again using a weighted random sample such that one third (33%) of the images had an automatic AA  $\geq 60^\circ$ . This improved the discriminatory performance for cam morphology compared to *model 1*. However, it was still deemed that there were too many ‘false positive’ results where AA was too high as compared to manual assessment. From here on all testing subsamples were weighted to include one third (33%) of images with an automatic AA  $\geq 60^\circ$ .

#### Model 3 – refined index and intersection position definitions

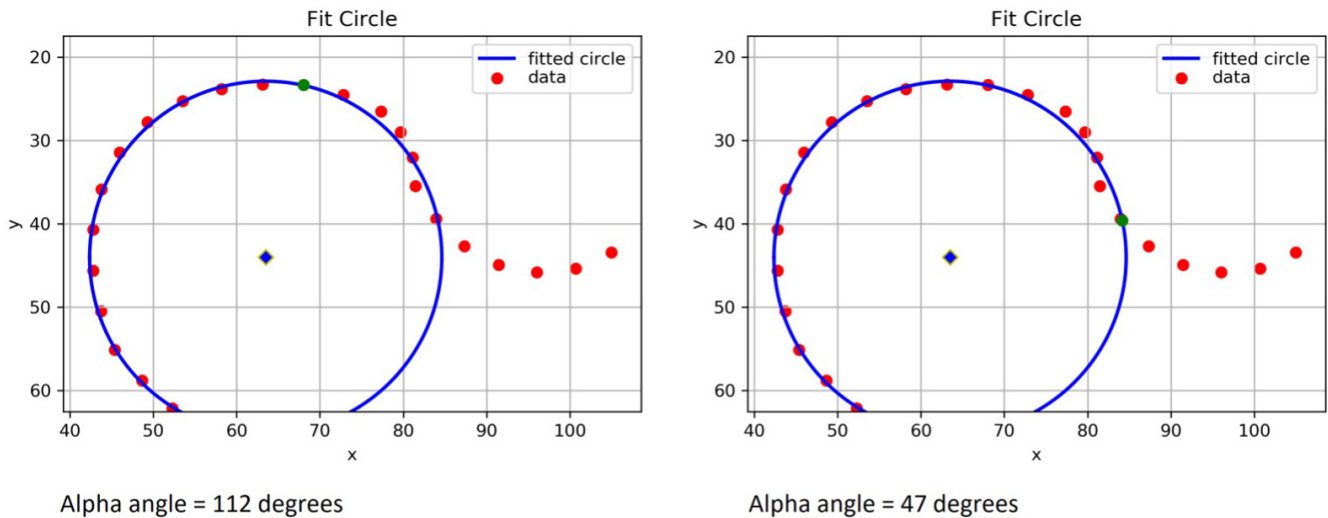
To further improve the index point threshold for *model 3*, negative residuals were included for the first time (after multiplication by -1) and could now inflate the index point



**Figure 4.** A graphical representation from Python of the points on a ‘straight forward’ dual-energy x-ray absorptiometry (DXA) scan. In this case the intersection position (green) is easily defined between two outline points which lie either side of the circle of best fit.



**Figure 5.** A graphical representation from Python of a set of points where the line between the index point and its preceding point does not intersect the circle. In this case, the preceding point (green) is the approximate intersection position as it has a smaller residual than the index point.



**Figure 6.** A graphical representation from Python of the same set of points for Experiments 1 and 2. The blue circle shows the circle of best fit plotted from point 15 to 28. The left image shows the results of Experiment 1 with the green interception point marked at around 12 o'clock giving an automatic alpha angle (AA) of 112°. This was viewed as a falsely elevated AA as on visual inspection of this hip dual-energy x-ray absorptiometry (DXA) scan. The slight deviation seen on this graphical illustration of the points from 1 to 2 o'clock was not visible by eye. The right image shows the same set of points for Experiment 2 with the green interception point marked close to 3 o'clock, giving an automatic AA of 47°. The manual AA for this image was 46°. Experiment 2 yielded improved results because of the optimised definition of the index point.

threshold; if points 15–28 lay inside the circle then their negative residual might show greater deviation from the circle than those outline points which lay outside of the circle with positive residuals. Moreover, a minimum index point threshold of 1mm was included for cases of near perfect circle fit (i.e. cases

where the femoral head was deemed by the automatic method to have left the circle but this was not discernible manually), aiming to reduce ‘false positives’. A value of 1mm was selected as the minimum threshold to represent an approximation of what was detectable by eye on the images. Prior to the

minimum index point threshold, a point lying a fraction of a millimetre outside of the circle could be incorrectly registered as the index point (e.g. for very spherical femoral heads). In addition, if there was no intersection position on the circle then the approximated intersection position was moved to the outline point before, on or after the index point depending on which outline point had the smallest residual. Otherwise, *model 3* remained the same as *model 2* in terms of circle fitting and the femoral neck mid-point.

Qualitative analysis of *model 3* results and in particular outliers indicated the index point definition had become too stringent with several ‘false negative’ scans (i.e. non-detected cam morphologies).

#### Model 4 – final model

For the fourth model iteration (*model 4*), the index point definition was modified so that the residual for three consecutive points following the index point had to be greater than the index point threshold but they did not need to be ever increasing. That is, the point after the index point could have a smaller residual than the index point as long as it remained above the index point threshold. Otherwise *model 4* remained the same as *model 3* with regards to circle fitting, femoral neck mid-point and the intersection position.

#### Statistical analysis

We report the mean AA along with the AA range for both manual and automatic measures. To assess agreement between manual and automatic AA measures, the mean absolute difference is presented along with its standard deviation (SD) and a concordance correlation coefficient was calculated, presented with its 95% confidence interval (CI)<sup>26</sup>. Bland-Altman plots were used to visualise this agreement and to identify outliers. Cam morphology was defined as  $AA \geq 60^\circ$  based on a recent systematic review of previous studies<sup>8</sup>. To give a broad perspective on the automatic classification of cam, the prevalence of cam morphology, derived from the automatically calculated AA, is given for each model for either the training sample

( $n=1,930$ ) or the extension sample ( $n=4,877$ ) depending on the test population. Due to the known AA differences between sexes (i.e. females tend to have a lower AA than males) the summary results of our final model were stratified by sex<sup>6</sup>. In addition, inter-rater kappa statistics, presented with percentage agreement, were used to compare the cam morphology classifications based on the manual versus automatic AA measurements. Following the described iterative approach, we aimed to achieve strong agreement between the manual and automatic cam classifications as defined by a target kappa of 0.8<sup>27</sup>. All statistical analysis was performed using [Stata](#) version 15 (StataCorp, College Station, TX, USA).

## Results

A description of basic demographic variables is provided for the training ( $n=1930$ ), extension ( $n=4877$ ) and combined samples ( $n=6807$ ) in [Table 1](#). Each model 1–4 derived AA from the manually placed points in the training sample ( $n=1930$ ) and these measures were tested against manually derived AA in four subsets of 100 images. Mean manual and automatic AA including ranges, mean absolute difference, concordance correlation coefficients and cam-based inter-rater kappa's for each model iteration are given in [Table 2](#).

#### Final model – model 4

In 100 randomly selected images from the training sample, weighted to include one third (33%) with an automatic  $AA \geq 60^\circ$ , *model 4* gave an automatic mean AA of  $56.0^\circ$  [range  $35.5$ – $106.2^\circ$ ] compared to a manual mean AA of  $53.8^\circ$  [ $35.3$ – $100.6^\circ$ ]. The mean absolute difference was  $2.2^\circ$  [SD 10.8]. The concordance correlation coefficient was 0.83 [95% CI: 0.77–0.89] and the cam-based inter-rater kappa was 0.84 [93% agreement] ([Table 2](#)). A Bland-Altman plot ([Figure 7](#)) showed only five of the 100 images lay outside of the 95% confidence interval; all of the five outliers had higher than average AAs. On review of the five outlier images, four showed errors in the manual AA measurement with poor manual circle fitting. The remaining image showed the automatic method had failed to recognise a visually noticeable deviation of the femoral head

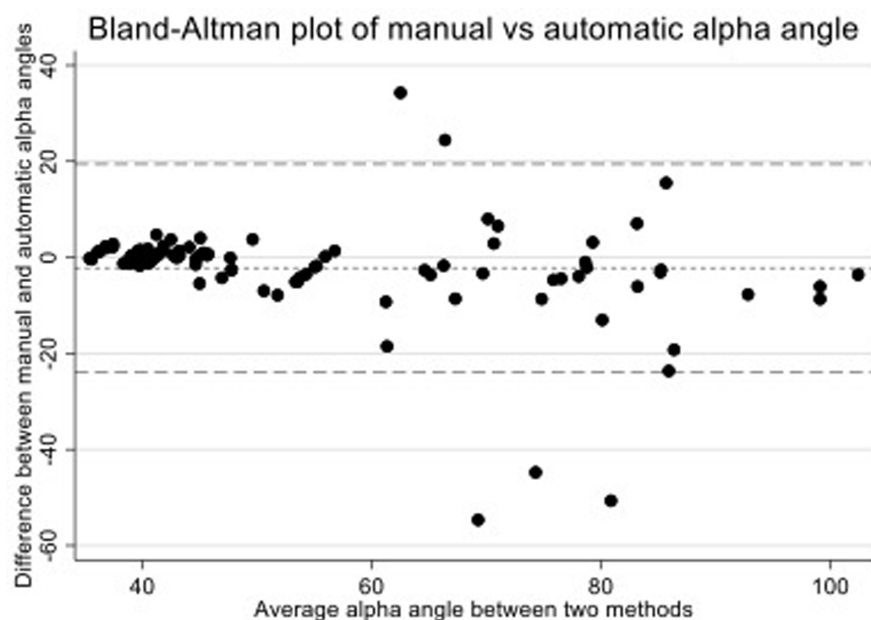
**Table 1. Sample demographics.** This table shows the age, height, weight and sex of the individuals in the training, extension and combined samples. The combined sample is the training and extension sample together.

Demographic variables	Training sample	Extension sample	Combined sample
<b>Continuous</b>	<b>Mean [range]</b>	<b>Mean [range]</b>	<b>Mean [range]</b>
Age (years)	62.3 [45-78]	62.3 [46-80]	62.7 [45-80]
Height (cm)	169.8 [145-198]	170.2 [137-203]	170.1 [137-203]
Weight (kg)	76.8 [42-154]	76.0 [36-160]	76.2 [36-160]
<b>Demographic variables</b>	<b>Training sample</b>	<b>Extension sample</b>	<b>Combined sample</b>
<b>Binary</b>	<b>Count [percentage]</b>	<b>Count [percentage]</b>	<b>Count [percentage]</b>
Sex (male)	949 [49.2]	2433 [49.9]	3382 [49.7]
Sex (female)	981 [51.8]	2444 [50.1]	3425 [50.3]
Total (n)	1930	4877	6807



**Table 2. Comparing of the performance of each automatic model iteration versus manual measurement of alpha angle (AA).** The manual and automatic means are given for each set of images (n=100) compared in the respective testing samples. The mean absolute difference calculated between the manual and automatic means is presented with its standard deviation (SD). The inter-rater kappa was calculated between manual and automatic cam classifications. The percentage agreement is presented alongside the kappa statistic. The concordance correlation coefficient compares the continuous alpha angle measures and is presented with a 95% confidence interval (CI). The prevalence of cam morphology defined as automatic AA  $\geq 60^\circ$  is given for all 1930 participants in the training sample for model iteration 1-4 and all 4,877 participants from the extension sample in the final model test.

Model iteration	Test sample	Manual AA Mean [Range]	Automatic AA Mean [Range]	Mean absolute difference [SD]	Kappa [percentage agreement]	Concordance correlation coefficient [95% CI]	Automatic cam prevalence
1	Random training sample	49.5 [36.4-115.8]	60.2 [38.0-117.9]	10.7 [25]	0.43 [80%]	0.28 [0.13-0.43]	28.7%
2	Random training sample	44.1 [34.2-77.3]	46.9 [36.3-103.3]	2.8 [6.9]	0.66 [98%]	0.60 [0.52-0.69]	9.6%
2	Random and weighted training sample	56.9 [36.7-103.7]	66.7 [39.8-113.2]	9.8 [16]	0.59 [80%]	0.62 [0.52-0.72]	9.6%
3	Random and weighted training sample	56.9 [36.7-103.7]	59.1 [39.1-106.2]	2.1 [11.0]	0.70 [87%]	0.81 [0.74-0.87]	4.7%
4	Random and weighted training sample	53.8 [35.3-100.6]	56.0 [35.5-106.2]	2.2 [10.8]	0.84 [93%]	0.83 [0.77-0.89]	7.2%
Final model	Random and weighted extension sample	54.7 [32.4-99.0]	58.1 [33.2-102.0]	3.4 [8.6]	0.86 [94%]	0.88 [0.84-0.92]	9.1%



**Figure 7. A Bland-Altman plot for model 4 testing in the training sample to compare manual and automatic alpha angles.** The figure shows that all measures were within a 95% confidence interval (dashed lines) apart from 5 outliers. All outliers were examined and visually inspected: four outliers showed errors in the manual AA measure where the automatic measure was able to better fit a circle to the femoral head; and one outlier where the automatic method ignored a slight bulging of the femoral head as one outline point had a residual of 0.96mm, 0.04mm below the threshold of 1mm.

from the circle of best fit; the residual for one of the three outline points encompassing this deviation was 0.96mm (0.04mm beneath the automatic minimum index point threshold of 1mm) meaning the algorithm did not measure the AA from this area. The kappa statistic for *model 4* was above the target threshold of 0.8 meaning it was selected as the final model to test in the extension sample.

### Testing the final model in the extension sample

Our final model (*model 4*) was applied to the extension sample of 4,877 individuals with automatically placed outline points which were subsequently manually corrected. From the extension sample, a randomly selected subsample of 100 DXA scans, weighted to include one third (33%) of images with an automatic AA  $\geq 60^\circ$ , had manual AA derived. Comparison between automatic AA based on points that were automatically derived but manually corrected and manual AA showed an automatic mean AA of  $58.1^\circ$  [33.2–102.0°] compared to a manual mean AA of  $54.7^\circ$  [32.4–99.0°]. The mean absolute difference was  $3.4^\circ$  [SD 8.6]. The concordance correlation coefficient was 0.88 [95% CI: 0.84–0.92] and the cam-based inter-rater kappa was 0.86 [94% agreement].

### Manual intra-rater comparison

The subsample of 100 images which was used to test *model 4* in the training sample was re-assessed for manually derived AA after 9 months. The two sets of manual AA measures were compared, giving a concordance correlation coefficient of 0.83 [95% CI: 0.76–0.89] and the cam-based intra-rater kappa was 0.80 (91% agreement).

### Automatic alpha angle in the combined sample

Our final model (*model 4*) was then applied to all the individuals in the combined sample ( $n=6807$ ), the mean AA was  $47.8^\circ$  [33.2–115.0°], in males the mean AA was  $51.6^\circ$  [35.8–106.2°] and in females the mean AA was  $44.2^\circ$  [35.2–115.0°].

## Discussion

We propose a method to automatically derive AA from AP hip DXA scans. The method is based on outline points and has been validated against manual AA measures. We have described how the method was developed, providing our Python code for the final version of the method for wider use by the research community<sup>28</sup>.

Similar methods utilising outline points to calculate AA have been reported previously but these studies do not include details on method synthesis, validation, nor open source code to allow for replication. In addition, some do not incorporate automatically placed points and require full manual annotation of the outline points which is time-consuming<sup>4,9</sup> and those methods including automated point placement failed to achieve consistent results as compared with manual annotators<sup>15</sup>. In contrast, the work presented here allows for replication of our methods and details our comprehensive validation using

500 blinded manual AA measures. In terms of validation, previously reported studies investigating cam morphology defined by AA have reported inter-rater kappas of 0.73<sup>4</sup> and 0.83<sup>29</sup>; our method showed an inter-rater kappa of 0.84–0.86 which compares favourably to these studies. In addition, our automatically derived AA from DXA scans in 6,807 UKB participants (males: mean AA  $51.6^\circ$ , range 35.8–106.2°; females: mean AA  $44.2^\circ$ , range 35.2–115.0°) were closely aligned to AA derived manually from radiographs in a large Danish cohort (males: mean  $52.6^\circ$ , range 30–108°; females: mean  $45^\circ$ , range 26–92°) providing further indication that our methods work as expected<sup>6</sup>. Interestingly, the final automatic model did record, on average, a  $3.4^\circ$  higher AA than manual annotation. This likely reflects the automated methods capturing deviation of the femoral head from the circle of best fit sooner than the manual annotator.

There are limitations to our work. Firstly, although the outline point placement is automated it requires manual checking to make sure it is correct, and osteophytes are excluded from the outline. However, this requires much less time than manually placing outline points and makes it feasible to obtain the large sample sizes required for genome-wide association studies<sup>30</sup>. Further work is being undertaken to improve automated outline point placement, and to develop a flagging system to highlight images requiring manual inspection where the point placement is suboptimal. Secondly, when validating our automated method, we compared these measures to one highly trained manual operator only. Another limitation is the DXA scan positioning used in this study included a broad range of internal rotation tolerated (15–25°). That said, this probably makes our study more representative of clinical imaging where positioning can vary considerably. In addition, this code was developed using DXAs of only left hips which might be systematically different in shape to right hips. In an independent study, we have applied these methods to both right and left hip AP radiographs (the right hip images were mirrored before applying the outline points and code), validating the results with manual measures from two clinical annotators<sup>31</sup>. The automated methods performed well against manual measures in this study but more work is needed to assess the reproducibility of these results in external cohorts and clinical environments. A main contribution of this work is that this is the first paper to set out a detailed method of how to automatically derive measures of AA on AP hip images. More work is needed to see if and how the proposed method can be successfully repurposed for lateral hip radiographs.

To conclude, we have described the development and validation of a method to derive AA on AP hip DXA images. We have made the proposed method available to other researchers in the field, allowing for AA to be derived in a standardised way across studies and in particular large population cohorts. This will enable the analysis of AA against clinically relevant outcomes such as OA, hip pain and THR, paving the way for this technology to be integrated into clinical care.

## Data availability

### Source data

The outline points for the UKB hip DXA scans used in this study are developed by the Wellcome Collaborative Grant AUGMENT (project application 17295). The points files used in this study will be made available from UK Biobank and they will be contained in a subsequent data release. UK Biobank control the image specific data developed as part of this research (i.e. points files) and hence they cannot be uploaded to a separate repository. UK Biobank resources are open to all researchers which will allow for replication.

## Software availability

Source code available from: <https://github.com/benfaber20/Automatic-alpha-angle/tree/v1.2>. Archived source code at time of publication: <https://doi.org/10.5281/zenodo.4462770><sup>28</sup>.

License: GNU General Public License

## Acknowledgements

This work has been conducted using the UK Biobank resource, access application 17295.

## References

- Nötzli HP, Wyss TF, Stoecklin CH, *et al.*: **The contour of the femoral head-neck junction as a predictor for the risk of anterior impingement.** *J Bone Joint Surg Br.* 2002; **84**(4): 556–60.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Murphy NJ, Eyles JP, Hunter DJ: **Hip Osteoarthritis: Etiopathogenesis and Implications for Management.** *Adv Ther.* 2016; **33**(11): 1921–46.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ganz R, Parvizi J, Beck M, *et al.*: **Femoroacetabular Impingement: A Cause for Osteoarthritis of the Hip.** *Clin Orthop Relat Res.* 2003; (417): 112–20.  
[PubMed Abstract](#)
- Agricola R, Heijboer MP, Bierma-Zeinstra SMA, *et al.*: **Cam impingement causes osteoarthritis of the hip: a nationwide prospective cohort study (CHECK).** *Ann Rheum Dis.* 2013; **72**(6): 918–23.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Tannast M, Siebenrock KA, Anderson SE: **Femoroacetabular impingement: radiographic diagnosis—what the radiologist should know.** *AJR Am J Roentgenol.* 2007; **188**(6): 1540–52.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gosvig KK, Jacobsen S, Sonne-Holm S, *et al.*: **The prevalence of cam-type deformity of the hip joint: a survey of 4151 subjects of the Copenhagen Osteoarthritis Study.** *Acta Radiol.* 2008; **49**(4): 436–41.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Griffin DR, Dickenson EJ, Wall PDH, *et al.*: **Hip arthroscopy versus best conservative care for the treatment of femoroacetabular impingement syndrome (UK FASHIoN): a multicentre randomised controlled trial.** *Lancet.* 2018; **391**(10136): 2225–35.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- van Klij P, Reiman MP, Waarsing JH, *et al.*: **Classifying Cam Morphology by the Alpha Angle: A Systematic Review on Threshold Values.** *Orthop J Sports Med.* 2020; **8**(8): 2325967120938312.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Saberi Hosnijeh F, Zuidenwijk ME, Versteeg M, *et al.*: **Cam Deformity and Acetabular Dysplasia as Risk Factors for Hip Osteoarthritis.** *Arthritis Rheumatol.* 2017; **69**(1): 86–93.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Griffin DR, Dickenson EJ, O'Donnell J, *et al.*: **The Warwick Agreement on femoroacetabular impingement syndrome (FAI syndrome): an international consensus statement.** *Br J Sports Med.* 2016; **50**(19): 1169–76.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Palmer AJR, Gupta VA, Fernquest S, *et al.*: **Arthroscopic hip surgery compared with physiotherapy and activity modification for the treatment of symptomatic femoroacetabular impingement: multicentre randomised controlled trial.** *BMJ.* 2019; **364**: 1185.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nicholls AS, Kiran A, Pollard TCB, *et al.*: **The association between hip morphology parameters and nineteen-year risk of end-stage osteoarthritis of the hip: a nested case-control study.** *Arthritis Rheum.* 2011; **63**(11): 3392–400.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Thomas GE, Palmer AJ, Batra RN, *et al.*: **Subclinical deformities of the hip are significant predictors of radiographic osteoarthritis and joint replacement in women. A 20 year longitudinal cohort study.** *Osteoarthritis Cartilage.* 2014; **22**(10): 1504–10.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Edwards K, Leyland KM, Sanchez-Santos MT, *et al.*: **Differences between race and sex in measures of hip morphology: a population-based comparative study.** *Osteoarthritis Cartilage.* 2020; **28**(2): 189–200.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lindner C, Thiagarajah S, Wilkinson JM, *et al.*: **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.** *Osteoarthritis Cartilage.* 2013; **21**(10): 1537–44.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Faber BG, Baird D, Gregson CL, *et al.*: **DXA-derived hip shape is related to osteoarthritis: findings from the MrOS cohort.** *Osteoarthritis Cartilage.* 2017; **25**(12): 2031–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Faber BG, Bredbenner TL, Baird D, *et al.*: **Subregional statistical shape modelling identifies lesser trochanter size as a possible risk factor for radiographic hip osteoarthritis, a cross-sectional analysis from the Osteoporotic Fractures in Men Study.** *Osteoarthritis Cartilage.* 2020; **28**(8): 1071–1078.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Yoshida K, Barr RJ, Galea-Soler S, *et al.*: **Reproducibility and Diagnostic Accuracy of Kellgren-Lawrence Grading for Osteoarthritis Using Radiographs and Dual-Energy X-ray Absorptiometry Images.** *J Clin Densitom.* 2015; **18**(2): 239–44.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Faber BG, Ebsim R, Saunders FR, *et al.*: **Osteophyte size and location on hip DXA scans are associated with hip pain: findings from a cross sectional study in UK Biobank.** *Bone.* 2021; **153**: 116146.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Harvey NC, Matthews P, Collins R, *et al.*: **Osteoporosis epidemiology in UK Biobank: a unique opportunity for international researchers.** *Osteoporos Int.* 2013; **24**(12): 2903–5.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Littlejohns TJ, Holliday J, Gibson LM, *et al.*: **The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions.** *Nat Commun.* 2020; **11**(1): 2624.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bycroft C, Freeman C, Petkova D, *et al.*: **The UK Biobank resource with deep phenotyping and genomic data.** *Nature.* 2018; **562**(7726): 203–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ebsim R, Lindner C, Faber B, *et al.*: **Development of a machine learning-based fully automated hip annotation system for DXA scans.** *Proceedings of the Bone Research Society Annual Meeting 2020.* 2020.  
[Reference Source](#)
- Lindner C, Thiagarajah S, Wilkinson JM, *et al.*: **Fully automatic segmentation of the proximal femur using random forest regression voting.** *IEEE Trans Med Imaging.* 2013; **32**(8): 1462–72.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kanatani AK, Rangarajan AP: **Hyper least squares fitting of circles and ellipses.** *Comput Stat Data Anal.* 2011; **55**(6): 2197–208.  
[Publisher Full Text](#)
- Steichen TJ, Cox NJ: **A note on the concordance correlation coefficient.** *The Stata Journal.* 2002; **2**(2): 183–9.  
[Publisher Full Text](#)
- McHugh ML: **Interrater reliability: the kappa statistic.** *Biochem Med (Zagreb).* 2012; **22**(3): 276–82.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Faber B: **benfaber20/Automatic-alpha-angle: Alpha Angle from DXA v1.2**

(Version v1.2). Zenodo. 2021.

<http://www.doi.org/10.5281/zenodo.4462770>

29. Gosvig KK, Jacobsen S, Palm H, *et al.*: **A new radiological index for assessing asphericity of the femoral head in cam impingement.** *J Bone Joint Surg Br.* 2007; **89**(10): 1309–16.  
[PubMed Abstract](#) | [Publisher Full Text](#)
30. Gibson G: **Population genetics and GWAS: A primer.** *PLoS Biol.* 2018; **16**(3): e2005485.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Zucker B, Ebsim R, Lindner C, *et al.*: **High bone mass and cam morphology are independently related to hip osteoarthritis: findings from the High Bone Mass Cohort.** *BMC Musculoskeletal Disorders.* [in press].

# Open Peer Review

Current Peer Review Status:  

---

## Version 2

Reviewer Report 29 July 2022

<https://doi.org/10.21956/wellcomeopenres.20024.r51583>

© 2022 van Klij P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

 **Pim van Klij** 

Sports Medicine, Isala , Zwolle, The Netherlands

I am happy with the way my comments are addressed. I have no additional comments. Great to be involved in this paper. Happy to be involved to help with other manuscripts in the future.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Hip and groin problems in athletes with specific interest for cam morphology development.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Version 1

Reviewer Report 15 June 2022

<https://doi.org/10.21956/wellcomeopenres.18367.r50806>

© 2022 Masters J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

 **James Masters** 

Oxford Trauma, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

Summary: A technical validation piece of work on automated measurement of the Alpha angle

from DEXA scans.

Strengths: Large numbers of images used, robust processes appropriately comparing to gold standard diagnostic techniques.

Well reported with sufficient detail to reproduce work.

Considerations: Only left hips are used. There is a school of thought that considers FAI part of a pathoanatomic spectrum with DDH. DDH is recognised to be more prevalent on the left hip due to intrauterine moulding. I therefore wonder whether some additional validation on a right hip data set would be valuable or of interest if available.

For the manual measure of the alpha angle, there may be value in an additional manual rater, particularly as there is a consistent difference between the manual and automated methods.

Overall a well executed piece of work which will add to the field.

**Is the rationale for developing the new method (or application) clearly explained?**

Yes

**Is the description of the method technically sound?**

Yes

**Are sufficient details provided to allow replication of the method development and its use by others?**

Yes

**If any results are presented, are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions about the method and its performance adequately supported by the findings presented in the article?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Clinical effectiveness research with particular interest in randomised trials in trauma and orthopaedic surgery

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 13 Jul 2022

**Ben Faber**, University of Bristol, Bristol, UK

*Summary: A technical validation piece of work on automated measurement of the Alpha angle*

*from DEXA scans.*

*Strengths: Large numbers of images used, robust processes appropriately comparing to gold standard diagnostic techniques.*

*Well reported with sufficient detail to reproduce work.*

*Considerations: Only left hips are used. There is a school of thought that considers FAI part of a pathoanatomic spectrum with DDH. DDH is recognised to be more prevalent on the left hip due to intrauterine moulding. I therefore wonder whether some additional validation on a right hip data set would be valuable or of interest if available.*

Author response:

We now mention this point in our limitations. "In addition, this code was developed using DXAs of only left hips which might be systematically different in shape to right hips." Since submitting this paper to *WOR* we have applied this method to a cohort of left and right AP hip radiographs and published the results. We are now referring to the latter in the paper as follows: "In an independent study, we have applied these methods to both right and left hip AP radiographs (the right hip images were mirrored before applying the outline points and code), validating the results with manual measures from two clinical annotators<sup>31</sup>." In this study we examined left and right hips with no discernible side-based systematic errors encountered when comparing our automated readings with manual annotation.

*For the manual measure of the alpha angle, there may be value in an additional manual rater, particularly as there is a consistent difference between the manual and automated methods.*

Author response:

We acknowledge this limitation and mention it as such. In addition, we now include a reference to a study where this code was applied to radiographs and validated against two manual annotators. This now reads "In an independent study, we have applied these methods to both right and left hip AP radiographs (the right hip images were mirrored before applying the outline points and code), validating the results with manual measures from two clinical annotators<sup>31</sup>."

**Competing Interests:** We have no competing interests to declare.

Reviewer Report 07 April 2022

<https://doi.org/10.21956/wellcomeopenres.18367.r48970>

© 2022 van Klij P. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

? Pim van Klij 

Sports Medicine, Isala , Zwolle, The Netherlands

Thank you for the invitation to review this manuscript. It is a well-organized manuscript with some new information regarding the most used measurement for cam morphology. It might introduce some new insights in how we can automate these measurements in bigger cohorts. Despite being quite familiar with hip/groin problems and the alpha angle measurement, the extensive described technique behind it is still a bit outside my expertise.

**Title:** no suggestions.

**Abstract:**

- Introduction: would suggest to write down 'imaging measure' to make sure that the reader understands it is about an imaging measurement.

**Introduction**

Alinea 1:

- A recent systematic review (van Klij et al. 2020) tried to clear up the mess around the alpha angle threshold for cam morphology, might be good to refer to this paper early in the manuscript.<sup>1</sup>

Alinea 2:

- In FAI syndrome, it is not only pain/symptoms, but also clinical signs that got involved, please introduce this here as well.
- Other software used for AA measurement can be MATLAB, might be good to add this here as it is also widely used.

*Figure 1:* I would suggest to not use capitals for 'Morphology' and 'Angle'. And maybe the quality of the circle and lines could be a bit better, as lines are crossing the red circle or even gaps between.

**Methods**

General: I believe the Methods section is your most important section, but however due to the very extensive explanation of the process, it makes it hard to read for a reader who is not that familiar with these measures. You might think of reducing it a bit to keep the reader onboard.

*Population*

- Is it possible to get more insight in what questionnaire is used to define the OA diagnosis?
- I do not find any information about informed consent for these participants and by which University this study is approved. Please introduce some medical ethical insight for the reader.

*DXA images and outline point*

- Supine position I suppose? Please add.
- Why the 15 to 25 degrees difference/margin? And why only the left hips, this is simply loss of power? For analysis, you could simply mirror right side images, right?

*Figure 2:* My suggestion would be to increase font size for the numbers in the figure to help guide the reader through the process with all numbered dots.



*Statistical analysis*

- By 'the known AA differences between sexes', please explain briefly for the more unexperienced reader that you mean that AA in general are lower in females.

**Results**

- Please introduced the numbers (n) also in Table 1 to make it more readable.

*Final model – model 4*

- Was the difference between manually and automatic measurement statistically significant? It might be interesting for the reader to give more insight in this and to objectively decide what is best in their opinion.

*Testing the final model in extension sample*

- It seems that the automatic model measures higher AA's than the manually placed ones, it seems that the outliers outside the 95% confidence interval are investigated further but how about outliers that are close to that range?

**Discussion**

- I still do not read an explanation for the differences found in general between automatic and manually placed dots for AA measurement. In general higher AAs were found in the automatic group, how can this be explained? What is most reliable and why? What can explain this difference?
- How can this measurement method be used for lateral radiographs? Is it one-on-one reproducible?
- Can this method also be used by a more unexperienced clinician in the field? If not, why not? We need to work towards a system which fits for (nearly) all clinicians who want to perform a reliable measurement.
- Just a question, do you believe this method can be used for the younger audience? Did not find information about this. As younger athletes mostly develop cam morphology in their adolescence with an open growth plate, would this measurement method system be useful for all ages?

In general, great manuscript with surely additional information for the reader. Thank you for inviting me to review it. Looking forward to correction of the manuscript.

**References**

1. van Klij P, Reiman MP, Waarsing JH, Reijman M, et al.: Classifying Cam Morphology by the Alpha Angle: A Systematic Review on Threshold Values. *Orthop J Sports Med.* 2020; **8** (8): 2325967120938312 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the rationale for developing the new method (or application) clearly explained?**

Yes

**Is the description of the method technically sound?**

Yes

**Are sufficient details provided to allow replication of the method development and its use by others?**

Yes

**If any results are presented, are all the source data underlying the results available to**

**ensure full reproducibility?**

Partly

**Are the conclusions about the method and its performance adequately supported by the findings presented in the article?**

Yes

**Competing Interests:** No competing interests were disclosed.**Reviewer Expertise:** Hip and groin problems in athletes with specific interest for cam morphology development.**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 13 Jul 2022

**Ben Faber**, University of Bristol, Bristol, UK

We thank the reviewers for their feedback and constructive comments.

**Reviewer 1:**

*Thank you for the invitation to review this manuscript. It is a well-organized manuscript with some new information regarding the most used measurement for cam morphology. It might introduce some new insights in how we can automatize these measurements in bigger cohorts. Despite being quite familiar with hip/groin problems and the alpha angle measurement, the extensive described technique behind it is still a bit outside my expertise.*

**Title:** no suggestions.**Abstract:**

- *Introduction: would suggest to write down 'imaging measure' to make sure that the reader understands it is about an imaging measurement.*

Author response:

We have made this change.

**Introduction**

- *A recent systematic review (van Klij et al. 2020) tried to clear up the mess around the alpha angle threshold for cam morphology, might be good to refer to this paper early in the manuscript.<sup>1</sup>*

Author response:

We have now written the following which includes the reference you suggest "The higher the AA the more indicative of cam morphology it is – with previously published thresholds of 50°, 55°, 60° and 83° all being used to define the presence of cam morphology<sup>1, 4-7</sup>, and

with a recent systematic review concluding a threshold of 60° is best <sup>8</sup>.”

- *In FAI syndrome, it is not only pain/symptoms, but also clinical signs that got involved, please introduce this here as well.*

Author response:

We have amended that sentence to now read “Cam morphology is thought to lead to pain and limited function of the hip in FAI syndrome.”

- *Other software used for AA measurement can be MATLAB, might be good to add this here as it is also widely used.*

Author response:

We now clarify that MATLAB is used in the references we give. “Alternatively, AA has been calculated using MATLAB based software which utilises outline points that have been placed manually around the femoral head.”

- *Figure 1: I would suggest to not use capitals for ‘Morphology’ and ‘Angle’. And maybe the quality of the circle and lines could be a bit better, as lines are crossing the red circle or even gaps between.*

Author response:

We have removed the text from the image as it is in the legend. In addition, all the lines now intersect the circle of best fit.

### **Methods**

- *General: I believe the Methods section is your most important section, but however due to the very extensive explanation of the process, it makes it hard to read for a reader who is not that familiar with these measures. You might think of reducing it a bit to keep the reader onboard.*

Author response:

We have tried to keep things as brief as possible. We believe it is necessary to include sufficient detail for the reader to be able to assess the resultant code.

### **Population**

- *Is it possible to get more insight in what questionnaire is used to define the OA diagnosis?*

Author response:

We have now included more information on the questionnaire used to define the OA diagnosis. The text now reads “The first 20% of both the training and extension samples were selected randomly from those with a self-reported diagnosis of OA based on a questionnaire completed at the same visit as the DXA scan. In the questionnaire, each participant was asked to record all their known diagnoses and a trained nurse subsequently coded the diagnosis. The OA code was not site specific.”

- *I do not find any information about informed consent for these participants and by which University this study is approved. Please introduce some medical ethical insight for the reader.*

Author response:

We have now included this information. The text reads “This study was approved by UKB (application number 17295) which is overseen by its own Ethics Advisory Committee. In addition, UKB received approval from the National Information Governance Board for Health and Social Care and Northwest Multi-Centre Research Ethics Committee

(11/NW/0382). All participants provided informed consent."

*DXA images and outline point*

- *Supine position I suppose? Please add.*

Author response:

We have now added this.

- *Why the 15 to 25 degrees difference/margin? And why only the left hips, this is simply loss of power? For analysis, you could simply mirror right side images, right?*

Author response:

Thank you. The agreed UKB DXA standard operating procedure dictated the positioning of the hips. This was out of the control of the authors. Aside from this, left hip images were chosen as part of a wider research programme to automatically assess hip shape in UKB with a view to then doing the same on the right hips. As yet only the left hips have been examined due to time constraints but the right hips will be examined in the future. In an independent study, we have now used this code on left and right *radiographs* with good repeatability.

We now mention both of these issues in the limitations "Another limitation is the DXA scan positioning used in this study included a broad range of internal rotation tolerated (15-25°). That said, this probably makes our study more representative of clinical imaging where positioning can vary considerably. In addition, this code was developed using DXAs of only left hips which might be systematically different in shape to right hips. In an independent study, we have applied these methods to both right and left hip AP radiographs (the right hip images were mirrored before applying the outline points and code), validating the results with manual measures from two clinical annotators<sup>31</sup>."

- *Figure 2: My suggestion would be to increase font size for the numbers in the figure to help guide the reader through the process with all numbered dots.*

Author response:

We have now increased the font size.

*Statistical analysis*

- *By 'the known AA differences between sexes', please explain briefly for the more unexperienced reader that you mean that AA in general are lower in females.*

Author response:

We have now included the following "Due to the known AA differences between sexes (i.e. females tend to have a lower AA than males) the summary results of our final model were stratified by sex"

**Results**

- *Please introduced the numbers (n) also in Table 1 to make it more readable*

Author response:

The sample numbers are already included in Table 1 in the last row. We have now modified the description of the row and columns to express this more clearly. The totals in each

sample are listed in the text "A description of basic demographic variables is provided for the training (n=1930), extension (n=4877) and combined samples (n=6807) in Table 1."

#### *Final model – model 4*

- *Was the difference between manually and automatic measurement statistically significant? It might be interesting for the reader to give more insight in this and to objectively decide what is best in their opinion.*

Author response:

We did not look at p-values nor p-value thresholds for "statistical significance" for the difference between the manual and automatic AA values as it is more appropriate to assess two types of measurements looking at agreement. The kappa values, concordance correlation coefficients and absolute mean differences provide an objective view of the data and allow assessment of the agreement of the two methods.

#### *Testing the final model in extension sample*

- *It seems that the automatic model measures higher AA's than the manually placed ones, it seems that the outliers outside the 95% confidence interval are investigated further but how about outliers that are close to that range?*

Author response:

These images were inspected and generally it is hard to see much difference between where the femoral head is adjudged to have left the circle. A small difference in distance (e.g. <1mm) between where annotators (e.g. manual or automated) think the femoral head leaves the circle can result in an AA that is several degrees higher. The automatic method recording a higher AA than the manual annotator suggests it declares the femoral head leaving the circle systematically earlier than the manual annotator. This is now mentioned in the discussion "Interestingly, the final automatic model did record, on average, a 3.4° higher AA than manual annotation. This likely reflects the automated methods capturing deviation of the femoral head from the circle of best fit sooner than the manual annotator."

#### **Discussion**

- *I still do not read an explanation for the differences found in general between automatic and manually placed dots for AA measurement. In general higher AAs were found in the automatic group, how can this be explained? What is most reliable and why? What can explain this difference?*

Author response:

Please see the response to the previous point. In terms of reliability, more work is needed to assess these methods in a clinical environment to assess reliability. However, given the automatic measures are more objective one would expect them to be more reliable. We have added the following limitation "The automated methods performed well against manual measures in this study but more work is needed to assess the reproducibility of these results in external cohorts and clinical environments."

- *How can this measurement method be used for lateral radiographs? Is it one-on-one reproducible?*

Author response:

These methods should be applicable to lateral radiographs but we have not validated this and it is beyond the scope of this study. We have included this point in the following sentence "More work is needed to see if and how the proposed method can be successfully repurposed for lateral hip radiographs."

- *Can this method also be used by a more inexperienced clinician in the field? If not, why not? We need to work towards a system which fits for (nearly) all clinicians who want to perform a reliable measurement.*

Author response:

Currently, the method has been developed as a research tool and has not been used in a clinical environment. The motivation for publishing this paper and code is for others to try. We would welcome collaborations to apply these methods clinically and we have plans to do so ourselves in future.

- *Just a question, do you believe this method can be used for the younger audience? Did not find information about this. As younger athletes mostly develop cam morphology in their adolescence with an open growth plate, would this measurement method system be useful for all ages?*

Author response:

We believe this method will work as long as the outline points are placed robustly. Our semi-automated approach facilitates this. For severely deformed hips some manual correction of points may be required. Young adults are unlikely to have joint space narrowing, osteophytes, enthesophytes etc, features which make it more difficult to automatically place the outline points. Thus, we would expect the methods to perform equally well or better in young adults. We would welcome collaborations to apply this method in other populations (although with the code being available the method could be used without our direct involvement).

**Competing Interests:** We have no competing interests to declare.