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

C. Lindner[1], S. Thiagarajah[2], J.M. Wilkinson[2], The arcOGEN Consortium*, G.A. Wallis[3], T.F. Cootes[1]

[1] Imaging Sciences, The University of Manchester, Manchester, U.K.

[2] Department of Human Metabolism, The University of Sheffield, Sheffield, U.K.

[3] Wellcome Trust Centre for Cell Matrix Research, The University of Manchester, Manchester, U.K.

*Members of The arcOGEN Consortium: N Arden, A Carr, K Chapman, P Deloukas, M Doherty, J Loughlin, A McCaskie, WER Ollier, SH Ralston, TD Spector, AM Valdes, GA Wallis, JM Wilkinson, E Zeggini

Corresponding author:

Claudia Lindner

Imaging Sciences

Institute of Population Health

The University of Manchester

Stopford Building

Manchester M13 9PT, U.K.

Email: claudia.lindner@postgrad.manchester.ac.uk

Tel: +44(0)161 2755173

Fax: +44(0)161 2715145

ABSTRACT

Objective: To evaluate the accuracy and sensitivity of a *fully automatic shape model matching (FASMM)* system to derive statistical shape models (SSMs) of the proximal femur from non-standardised anteroposterior (AP) pelvic radiographs.

Design: AP pelvic radiographs obtained with informed consent and appropriate ethical approval were available for 1105 subjects with unilateral hip osteoarthritis who had been recruited previously for The arcOGEN Study. The *FASMM* system was applied to capture the shape of the unaffected proximal femur from these radiographs. The accuracy and sensitivity of the *FASMM* system in calculating geometric measurements of the proximal femur and in shape representation were evaluated relative to validated manual methods.

Results: *De novo* application of the *FASMM* system had a mean point-to-curve error of less than 0.9mm in 99% of images (n=266). Geometric measurements generated by the *FASMM* system were as accurate as those obtained manually. The analysis of the SSMs generated by the *FASMM* system for male and female subject groups identified more significant differences in their global proximal femur shape than those obtained from the analysis of conventional geometric measurements.

Conclusions: The *FASMM* system rapidly and accurately generates a global SSM of the proximal femur from radiographs of varying quality and resolution. This system will facilitate complex morphometric analysis of global shape variation across large datasets. The *FASMM* system could be adapted to generate SSMs from the radiographs of other skeletal structures such as the hand, knee or pelvis.

Keywords: Osteoarthritis; statistical shape models; proximal femur; shape analysis; hip morphology

INTRODUCTION

The analysis of geometric measurements made of skeletal elements from standard two-dimensional radiographs is an established and routine clinical tool used in the diagnosis, prognosis and management of many skeletal disorders. Recently, detailed morphometric analyses have been conducted to identify key features of bone shape that contribute to disease incidence and progression and that may serve as biomarkers for pre-symptomatic diagnosis and treatment evaluation. Such studies have also been extended to identify potential genetic regulators of bone shape that may thereby contribute to disease susceptibility[1,2]. Specifically, with regards to the relationship between geometric measurements of proximal femur shape and hip osteoarthritis (OA), analyses have established, for example, that the radiographic shape of the femoral neck is associated with the onset and progression of hip OA[3], and that femoral neck width (FNW) and the ratio of femoral head diameter (FHD) to FNW can be used to identify individuals at risk of OA[4].

We have previously generated a semi-automatic system that digitally captures the shape of skeletal elements through the use of Active Shape Models (ASMs) and represents shape as a statistical shape model (SSM)[5]. An SSM describes every shape by the sum of a mean shape and a linear combination of a number of shape modes where the shape mode values vary between subjects. Hence, SSMs provide a *global* representation of shape rather than reducing shape to a series of linear measurements which enables the description and analysis of shape variation across datasets. Gregory *et al* first described the application of ASMs to generate an SSM of proximal femur shape in their analysis of osteoporotic hip fracture[6]. Thereafter, this approach has been used in a number of studies to investigate the relationship between features of proximal femur shape and the onset, incidence and progression of hip OA[1,7,8,9,10]. More recently, Agricola *et al* have shown that SSM variation can be used as a biomarker to predict the risk of total hip replacement in OA cases[11]. In the

above studies the standard, semi-automatic ASM methodology has been applied which requires the initial manual placement of landmark points to define the approximate shape and position of the skeletal element (the initialisation phase) followed by an automated search to outline the contour of the bone. The accuracy of the resultant SSM therefore depends on the precision and consistency of the manual initialisation phase.

As the current *semi-automatic* ASM methodology is time-consuming and vulnerable to intra- and inter-operator error, its application to large-scale morphometric analyses is limited. We have therefore developed a *fully automatic shape model matching* (*FASMM*) system that uses an SSM to capture and represent the shape of skeletal element/s within standard radiographs[12]. In this paper, we demonstrate the use of the *FASMM* system for segmentation of the proximal femur from anteroposterior (AP) radiographs via its rapid and accurate placement of 65 reference points along the contour of the proximal femur without any manual intervention. We show that these automatically placed reference points and the SSM derived from these points can reliably predict a number of standard hip geometric measurements. Further, we demonstrate the utility of this system in identifying differences in shape variation between large datasets through its application to an analysis of proximal femur shape variation between males and females.

METHODS

Dataset

AP pelvic radiographs were available from hip OA subjects recruited in Stage 2 of The arcOGEN Consortium study[13]. Inclusion criteria and ethical approval for Stage 2 subject recruitment were as previously described[13]. For this study, radiographs were selected from hip OA subjects who had no evidence of radiographic OA in one of their hip joints (with the presence of OA being defined as Kellgren-Lawrence score ≥ 2). Demographic data for each subject included body mass index (BMI) and date of birth but the date of X-ray was not consistently available. Geometric measurements (derived using a previously validated open-access software tool for measuring hip morphological characteristics (SHIPs)[14], which comprised the *manual measurements* for this study) were available for the unaffected hip joint from radiographs of 786 of the above 1105 subjects. These manual measurements included FHD, FNW, femoral neck axis length (FNAL), femoral neck shaft angle (FNSA), and alpha angle (AA). From these measurements the FHD-FNW ratio, FNAL-FHD ratio, and FNAL-FNW ratio were calculated. The FHD, FNW and FNAL measurements were all given in pixels. Pixel size information was not however available for any of the radiographs. Measurements made by two independent observers for a subset of 25 radiographs were available for estimates of inter-observer variability[14]. The formula for calculating the coefficient of variation (CV%) for each measurement was: $CV\% = 100 \times ((\delta/\sqrt{2}) / \mu)$ where δ represents the standard deviation of the differences between the paired measurements, and μ is the mean of all the values for that measurement.

The *FASMM* system

We have previously described methodology for automatic femur segmentation in pelvic radiographs using regression voting[12], which improves and extends the standard ASM approach[5], and is the basis

of the *FASMM* system. Here, we focused on finding and segmenting the left proximal femur in AP radiographs where the left hip was OA-free. For cases where the right hip was OA-free we mirrored the image so that it appeared on the left. The *FASMM* system was trained to segment the proximal femur by first detecting it in the radiograph and then outlining its contour using 65 points (see Figure 1). All points were placed in consistent positions across images through using 9 anchor points that defined specific anatomical features with remaining points being evenly spaced relative to these points. As in Figure 1, the system uses a front-view femur model that excludes both the lesser and greater trochanters. The system was designed to accommodate radiographs of differing quality, femur position/orientation and area of the pelvis. This allowed inclusion of radiographs obtained retrospectively from several recruitment centres that varied in resolution (555-4723 pixels wide), exposure contrast, focus to film distance, and position of beam centre relative to the hip joint.

The *FASMM* system was developed using 839 of the 1105 images. For this study to avoid training and testing on the same dataset, we split the 839 images into two subsets of 419 and 420 images respectively, and trained a system on each subset. We trained a further system using all 839 images for *de novo* application to the remaining 266 images. The refined *FASMM* system now trained on all 1105 images is freely available for non-commercial research purposes from http://personalpages.manchester.ac.uk/postgrad/claudia.lindner/.

Calculation of geometric measurements from the points placed by the FASMM system

Manual measurements were available for 786 images. For comparison, these measurements (see Figure 2) were automatically calculated from a subset of the 65 points generated by the *FASMM* system for each of the 786 images and application of custom code developed in Matlab R2010a. As with the manual measurements, FHD, FNW and FNAL were pixel-based. Due to the extent of variation in image-size across our dataset, these measurements were utilised for the subject-specific

comparison of manual vs. automatic measurements and to derive ratios that were independent of image-size.

Based on the 65 points as generated by the *FASMM* system (see Figure 1), the geometric measurements were calculated on a subset of these points as follows. To obtain the diameter and centre of the femoral head a circle was fitted to points [20-35] so as to minimise the average distance between the circle and these points. The femoral neck axis was derived by firstly identifying the minimum path between the contours through point sets [10-17] and [36-44], and then fitting a line through the midpoint of this path (femoral neck centre) and the centre of the femoral head. The femoral neck axis length was given by the distance between the centre of the femoral head and the intersection point of the femoral neck axis and the contour through points [56-62]. The femoral neck width was calculated by firstly fitting a line through the centre of the femoral neck that is perpendicular to the femoral neck axis, and then secondly calculating the distance between the intersection points of this line with the contours through point sets [10-17] and [36-44]. To obtain the alpha angle we calculated the angle between the femoral neck axis and a line fitted to the centre of the femoral head as well as the intersection point of the femoral head circle and the contour through points [36-44]. The femoral shaft axis was obtained by calculating the midpoints of point pairs [1,63], [2,63], [1,64], [2,64], [1,65] and [2,65] and then fitting a line through these midpoints. We calculated the femoral neck shaft angle by defining the angle between this shaft axis and the femoral neck axis.

The automatically and manually calculated measurements were compared using the Bland and Altman method[15]. We also calculated the Pearson's correlation coefficient for each measurement. All plots and calculations were made using Matlab R2010a. As the *FASMM* system is a deterministic algorithm (i.e. for the same radiograph it always produces the same measurements), the CV% was

calculated between the automatic measurements and the average of the manual measurements made by two observers. This was then compared to the inter-observer variability of the manual measurements.

Prediction of geometric measurements from the SSM generated by the FASMM system

The SSM mode values were calculated by building an SSM based on the 65 points generated for each subject by the *FASMM* system. We then used linear regression to predict the morphometric measurements based on the SSM mode values. As our SSM did not include any absolute scales (as most images were not calibrated), we excluded FHD, FNW and FNAL from this analysis. The automatically predicted SSM based measurements and the manual measurements were compared using the Bland and Altman method[15]. We also calculated the Pearson's correlation coefficient for each measurement. All plots and calculations were made using Matlab R2010a.

Male vs. female proximal femur shape

We used the SSM mode values to investigate whether there was a significant difference in radiographic proximal femur shape between the 662 females and 443 males within the 1105 case cohort. The SSM mode values were calculated by building an SSM based on the 65 points for every image obtained by applying the *FASMM* system. The mean shape variation was calculated across all images and then individually for males and females. A univariate independent two-sample Welch's t-test on the SSM mode values was used to compare the two groups.

We then investigated the accuracy of the SSM and the conventional geometric measurements to gender-classify the data using Gaussian probability density functions (PDFs). Out of the 786 images for which manual measurements were available we used 598 images to train the system used for

classification. We tested the classification system on the remaining 188 images (108 females and 80 males) for which manual measurements were available. These 188 images were a subset of the 266 images that were not involved in developing the *FASMM* system. We analysed the gender-classification performance of *(i)* a combination of SSM mode values that explain 98% of shape variation using male and female multivariate PDFs; *(ii)* a combination of all five non-pixel-based geometric measurements (FHD-FNW ratio, FNAL-FHD ratio, FNAL-FNW ratio, FNSA, and AA) using male and female multivariate PDFs; *(iii)* individual non-pixel-based geometric measurements (FHD-FNW ratio, FNAL-FHD ratio, FNAL-FNW ratio, FNSA, and AA) using male and female univariate PDFs.

As the male and female groups did not differ significantly in terms of year of birth (YOB) or BMI [108 females (YOB: 1939.3 ±SD: 9.1; BMI: 29.5 ±SD: 5.3) and 80 males (YOB: 1940.2 ±SD: 9.5; BMI: 28.9 ±SD: 5.9)], findings were not adjusted according to these criteria.

RESULTS

Application of the FASSM system

The *FASMM* system was applied to the dataset of 1105 images in three stages: *(i)* We trained the system on 419 images and applied it to 420 different images; *(ii)* we switched the datasets and trained the system on the set of 420 images before applying it to the set of 419 images; *(iii)* we trained the system on all 839 images from the first two stages and tested it on the remaining 266 images. These 266 images had therefore not been included in any of the training or testing phases of the development of the system[12]. *De novo* testing of the system against these images, using previously described methods[12], demonstrated a mean point-to-curve error of less than 0.9mm on 99% of images and of less than 3.3mm on 100% of the images. This accuracy is consistent with our previous findings[12]. The latest version of the system can process one image on average in 10s (on a 3.3 GHz Intel Core Duo PC using 3.5GB RAM and running Windows XP SP3). An analysis of the resultant SSM demonstrated that 17 shape modes accounted for 98% of the overall shape variation given by the whole dataset.

Comparison of manual vs. automatically generated measurements

To determine whether conventional geometric measurements could be automatically calculated from the 65 points placed by the *FASMM* system and whether these measurements were as accurate as those generated manually, we compared the two sets of measurements obtained from 786 images. Figure 3 shows the Bland and Altman plots for the automatically generated points based measurements against the manual measurements[14], for each of the 786 images. As is evident from Figure 3, the automatic and manual measurements were similar across a broad range of values. The limits of agreement between the two sets of measurements were less than 10% across the range of

measurements, and with a bias approaching 0%. The only exception to this was the alpha angle which had poor limits of agreement with up to 50% difference in measurement over ±2SD.

We then calculated the CV% between the manual measurements obtained by two independent observers on 25 images. As shown in Table 1, the inter-observer CV% was between 0.7% and 1.5% for all measurements, with the exception of the alpha angle which was 3.2%. For comparison, we then calculated the CV% between the mean values for the manually obtained measurements and those obtained automatically for the same set of images. The automatic-manual CV% was between 1.0% and 1.9% for all measurements, with the exception of the alpha angle which was 19.2% (see Table 1).

Comparison of manual measurements vs. those predicted from the SSM

To determine whether geometric measurements could be accurately predicted from the SSM mode values generated by the *FASMM* system, we compared the SSM based measurements with those obtained manually. Figure 4 shows the Bland and Altman plots for the manual measurements against measurements predicted by applying linear regression to the SSM mode values. The plots indicate that the SSM mode values can predict the manual measurements. As above, the automatically predicted SSM based measurements and those obtained manually were in very good agreement across a broad range of values. The limits of the agreement between the methods were less than 10% across the range of measurements, and with a bias approaching 0%. Again, the only exception was the alpha angle which had poor limits of agreement with up to 30% difference in measurement over ±2SD.

Comparison of male vs. female proximal femur shape

SSM mode values for the 662 females and 443 males within the 1105 case cohort were calculated and the mean SSM values for each mode and each of the groups compared using a univariate independent two-sample Welch's t-test. We used the Kolmogorov-Smirnow test to verify that the data for each mode and every subject group were normally distributed. Eleven of the 17 modes had male *vs.* female mean values that were significantly different ($p<0.05$). After Bonferroni adjustment, five of the 17 modes had male *vs.* female mean values that were significantly different ($p<0.003$). Figure 5 provides a pictorial representation of the difference between male and female radiographic proximal femur shape showing the overall mean shape as well as indicating the mean male and female shapes.

We then investigated the accuracy of the subject-specific SSM mode values generated by the *FASMM* system to predict whether a radiograph had been obtained from a male or a female subject. We found that gender-classification based on a multivariate probability density function (PDF) using all 17 shape modes had a 76.3% success rate. In contrast, when using a multivariate PDF based on the five non-pixel-based conventional geometric measurements (combined ratios of FHD-FNW, FNAL-FHD, FNAL-FNW, FNSA and AA) to gender-classify the same data, the success rate was 71.8%. Further, when using a univariate PDF based on the individual ratios for gender-classification the success rates were: FHD-FNW ratio: 68.7%, FNAL-FHD ratio: 62.7%, FNAL-FNW ratio: 69.1%, FNSA: 63.6%, and AA: 67.9%.

DISCUSSION

We have demonstrated that the newly developed *FASMM* system can automatically, rapidly and accurately capture the shape of the proximal femur from non-standardised AP pelvic radiographs. Conventional hip morphologic measurements (with the exception of the alpha angle) can also be accurately and automatically calculated from the *FASMM* generated contour as well as from the *FASMM* generated SSM, despite it being an abstract representation of radiographic shape. Hence, the *FASSM* system will facilitate detailed analysis of global shape variation across large (either retrospective or prospective) datasets.

The *FASMM* system can be applied to the analysis of shape variation between subject groups, as demonstrated by our identification of significant differences between male and female proximal femur shape using the system. Further, we found that applying a multivariate approach to gender-classify images using combinations of geometric measurements increased classification accuracy in comparison to that of single measurements. Gender-classification accuracy was increased still further by incorporating global shape information (via using SSM mode values generated by the *FASMM* system). These findings suggest that using SSM mode values increases the power to detect shape differences between subject groups. The shape difference identified between males and females (Figure 5) is consistent with findings that the pistol-grip deformity is more prevalent in men[19]. In our analysis, only relative shape differences between males and females were analysed and not differences in size. Future extension of the *FASMM* system will explore inclusion of both specification and quantification of shape variation across subject groups.

The alpha angle is a measurement used to quantify cam-type deformity which is a risk factor highly associated with hip OA that might be modifiable via early surgical intervention[10,16]. In our analyses,

the alpha angle estimated using the *FASMM* system did not correlate well with the manual measurements. We assume the reason for this to be that there is no objective and reproducible way of measuring the alpha angle on AP pelvic radiographs but that its measurement includes the subjective positioning and fitting of a circle to the contour of the femoral head. When we attempted to implement an automatic (and hence objective) calculation of the alpha angle, we were not able to define a method of calculation that could be applied consistently across the range of femoral head shapes that also reflected cam-type deformity correctly. In particular, using the intersection point of the femoral head circle and the superior femoral neck contour given by points [35-44] (see Figure 1) to define the alpha angle resulted in a large number of cases where the alpha angle did not accurately reflect the presence/absence of cam-type deformity. The two main reasons for this are likely to be the range of skeletal shape variation in the area of the superior femoral head-neck junction, and because point 35 (which defines the superior end of the hemispheric femoral head shape) does not always lie inside the circle that best fits to points [20-35]. However, as the automatic fitting of a circle to the femoral head using points [20-35] as described above led to high agreement between manual and fully automatic values for all other geometric measurements, we did not change our approach to fitting the femoral head circle but excluded point 35 from the superior femoral neck contour for the calculation of the alpha angle. This appeared to improve the agreement between the calculated alpha angle and the observed presence/absence of cam-type deformity but showed poor alpha angle estimation when compared to manual measurements. In addition to inherent problems in its measurement, the use of the alpha angle obtained from AP pelvic radiographs to predict the presence/absence of cam-type deformity has been questioned because alpha angle measurements based on AP pelvic radiographs do not correlate very well with measurements derived from three-dimensional MRI data[17,18]. Doherty *et al* have, however, suggested that cam-type deformity can be identified by the presence of a pistol-grip deformity which in turn can be defined by the FHD-FNW ratio[19]. We have shown accurate prediction of the

FHD-FNW ratio using the *FASMM* system. In addition, as SSMs provide a global representation of shape, information about cam-type deformity will be contained in the SSM mode values and thereby provide a more accurate method for its detection.

A limitation of this study is that it relies on single two-dimensional radiographic views to quantify proximal femur shape. The two-dimensional radiographic shape of the proximal femur may vary due to the inclination of the pelvis and rotation of the legs during image acquisition. In previous work, we have however shown that the degree of position-related shape variation is small compared to the overall shape variation of the proximal femur as derived from AP pelvic radiographs[20], and hence will not significantly hamper analysis of global shape variation across large datasets.

In the specific application of the *FASMM* system presented here, we have trained the system to capture the shape of the unaffected proximal femur from AP radiographs of subjects with unilateral OA. This provides a mechanism for the analysis of proximal femur shape variation that may increase susceptibility to OA and disease progression. Our preliminary experiments (unpublished) have also shown that the system as currently configured performs well on joints with mild OA where there is only minor deviation from the *unaffected* shape. In order to capture and analyse the range of shape variation of joints affected by OA related pathology, the *FASMM* system will be extended by including OA-affected images in the training phase of the system to capture features such as osteophytes, atrophic bone patterns and cysts. Thereby the system will fulfil a recognised need for better joint-specific phenotypic definition of OA as well as for fast and reliable methods to standardise radiographic scoring[21]. Further, we are currently extending the *FASSM* system to incorporate the lesser and greater trochanters into the model and we are applying the *FASMM* system to the segmentation of the pelvis, the knee joint and the joints of the hand. Application of

the *FASSM* system will therefore benefit morphological studies aimed at understanding the aetiology and progression of diseases of the major joint sites.

ROLE OF FUNDING SOURCE

C. Lindner is funded by the Medical Research Council, UK. arcOGEN was funded by a special purpose grant from Arthritis Research UK (18030). S. Thiagarajah was funded by Joint Action at the British Orthopaedic Association, and by the National Institutes of Health Research, UK.

AUTHOR CONTRIBUTIONS

All authors have made substantial contributions to data acquisition or study execution, manuscript preparation, and final approval of the submitted article. CL takes responsibility for the integrity of the work as a whole.

CONFLICT OF INTEREST

The authors declare not to have any conflict of interest.

REFERENCES

1. Baker-Lepain JC, Lynch JA, Parimi N, et al. Variant alleles of the Wnt antagonist FRZB are determinants of hip shape and modify the relationship between hip shape and osteoarthritis. Arthritis & Rheumatism 2012;64(5):1457–1465.
2. Waarsing JH, Kloppenburg M, Slagboom PE, et al. Osteoarthritis susceptibility genes influence the association between hip morphology and osteoarthritis. Arthritis & Rheumatism 2011;63(5):1349–54.
3. Javaid M, Lane N, Mackey D, et al. Changes in proximal femoral mineral geometry precede the onset of radiographic hip osteoarthritis. Arthritis & Rheumatism 2009;60(7):2028–2036.
4. 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. Arthritis & Rheumatism 2011;63(11):3392–3400.
5. Cootes T, Taylor C, Cooper D, Graham J. Active Shape Models - Their training and application. Computer Vision and Image Understanding 1995;61(1):38–59.
6. Gregory JS, Testi D, Stewart A, Undrill PE, Reid DM, Aspden RM. A method for assessment of the shape of the proximal femur and its relationship to osteoporotic hip fracture. Osteoporosis International 2004;15(1):5–11.
7. Lynch JA, Parimi N, Chaganti RK, Nevitt MC, Lane NE. The association of proximal femoral shape and incident radiographic hip OA in elderly women. Osteoarthritis and Cartilage 2009;17(10):1313–1318.
8. Barr RJ, Gregory JS, Reid DM, 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 2012;51(3):562–570.

9. Waarsing JH, Rozendaal RM, Verhaar JAN, Bierma-Zeinstra SMA, Weinans H. A statistical model of shape and density of the proximal femur in relation to radiological and clinical OA of the hip. Osteoarthritis and Cartilage 2010;18(6):787–7.
10. Agricola R, Heijboer MP, Bierma-Zeinstra SMA, Verhaar JAN, Weinans H, Waarsing JH. Cam impingement causes osteoarthritis of the hip: a nationwide prospective cohort study (CHECK). Annals of the Rheumatic Diseases 2013;72(6):918–923.
11. Agricola R, Bierma-Zeinstra S, Verhaar J, Weinans H, Waarsing J. Total hip replacement but not clinical osteoarthritis can be predicted by shape variations of the hip: a prospective cohort study (check). Osteoarthritis and Cartilage 2012;20(Suppl 1):S12–S13.
12. Lindner C, Thiagarajah S, Wilkinson JM, arcOGEN Consortium, Wallis GA, Cootes. Accurate Fully Automatic Femur Segmentation in Pelvic Radiographs Using Regression Voting. MICCAI 2012;LNCS#7512:353–360.
13. arcOGEN Consortium, arcOGEN Collaborators. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. The Lancet 2012;380(9844):815–823.
14. Thiagarajah S, Macinnes SJ, Yang L, Doherty M, Wilkinson JM. Quantifying the charachteristics of the acetabulum and proximal femur using a semi-automated hip morphology software programme (SHIPS). Hip International 2013;3(23):330–336.
15. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986;8476(1):307–310.
16. Ganz R, Parvizi J, Beck M, Leunig M, Nötzli H, Siebenrock KA. Femoroacetabular impingement: a cause for osteoarthritis of the hip. Clinical Orthopaedics and Related Research 2003;417:112–120.

17. Barton C, Salineros MJ, Rakhra KS, Beaule PE. Validity of the Alpha Angle Measurement on Plain Radiographs in the Evaluation of Cam-type Femoroacetabular Impingement. Clinical Orthopaedics and Related Research 2011;469(2):464–469.
18. Dudda M, Albers C, Mamisch TC, Werlen S, Beck M. Do normal radiographs exclude asphericity of the femoral head-neck junction? Clinical Orthopaedics and Related Research 2009;467(3):651–659.
19. Doherty M, Courtney P, Doherty S, et al. Nonspherical femoral head shape (pistol grip deformity), neck shaft angle, and risk of hip osteoarthritis: a case-control study. Arthritis & Rheumatism 2008;58(10):3172–3182.
20. Lindner C, Thiagarajah S, Wilkinson JM, arcOGEN Consortium, Wallis GA, Cootes TF. Short-term variability of proximal femur shape in anteroposterior pelvic radiographs. Proceedings of the 15th Conference on Medical Image Understanding and Analysis: BMVA Press 2011:69–73.
21. Bijlsma JWJ, Berenbaum F, Lafeber FPJG. Osteoarthritis: an update with relevance for clinical practice. The Lancet 2011;377(9783):2115–2126.

FIGURE LEGENDS

Figure 1 Fully automatic segmentation of the proximal femur: Detecting the proximal femur in the image (left) and outlining its contour using 65 points (right). All anchor points are highlighted in red and were chosen to mark the following features: the beginning (point 3) and the end (point 10) of the lesser trochanter projected to the nearby femoral shaft contour, inflexion points at the inferior (point 17) and superior (point 38) femoral head-neck junction, the inferior (point 20) and superior (point 35) end of the hemispheric femoral head shape, the projected beginning of the trochanteric fossa (point 44) as well as local maxima at the superior contour of the greater trochanter (point 48) and at the inferior end of the greater trochanter (point 55). Note that these features are based on the visual appearance of the proximal femur in AP pelvic radiographs; this does not necessarily correspond with three-dimensional anatomical features.

Figure 2 Geometric measurements based on output of the *FASMM* system: FHD = femoral head diameter, FNAL = femoral neck axis length, FNW = femoral neck width, FNSA = femoral neck shaft angle, AA = alpha angle. Note that fully automatic *calculation* of these measurements is based on the blue contour points only whereas the fully automatic *prediction* uses a statistical shape model that includes all 65 contour points.

Figure 3 Bland-Altman plots for manual measurements vs. fully automatic measurements (FHD = femoral head diameter, FNW = femoral neck width, FNAL = femoral neck axis length, FNS = femoral neck shaft, cc = Pearson's correlation coefficient, SD = standard deviation): These data show high correlation and less than 10% difference between the methods (red dotted line labelled ±2SD, indicating limits of agreement between methods), and minimal bias (red dotted lines indicating 95% confidence interval of mean) for all but the alpha angle.

Figure 4 Bland-Altman plots for manual measurements vs. prediction using linear regression based on shape model mode values (FHD = femoral head diameter, FNW = femoral neck width, FNAL = femoral neck axis length, FNS = femoral neck shaft, cc = Pearson's correlation coefficient, SD = standard deviation): These data show high correlation and less than 10% difference between the methods (red dotted line labelled ±2SD, indicating limits of agreement between methods), and minimal bias (red dotted lines indicating 95% confidence interval of mean) for all but the alpha angle.

Figure 5 Shape differences between male (- - -) and female (**...**) proximal femur shape based on fully automatic search results for 1105 images, where differences from the mean (–) have been exaggerated by a factor of 3 to aid visualisation.

Table 1 Manual inter-observer and fully automatic vs. manual repeatability (SD = standard deviation, CI = confidence interval, CV = coefficient of variation): FHD = femoral head diameter, FNW = femoral neck width, FNAL = femoral neck axis length, FNSA = femoral neck shaft angle, AA = alpha angle. The automatic measurements were calculated from the 65 points for each image generated using the *FASMM* system.

	FHD	**FNW**	**FNAL**	**FNSA**	**AA**	**FHD-FNW ratio**	**FNAL-FHD ratio**	**FNAL-FNW ratio**
Inter-observer variation based on two independent sets of manual measurements								
Mean	376.5	262.6	602.4	128.5	49.4	1.44	1.60	2.31
SD	6.5	4.7	6.5	1.2	2.2	0.03	0.03	0.04
95% CI	373.9 – 379.0	260.8 – 264.4	599.8 – 604.9	128.0 – 129.0	48.6 – 50.3	1.43 – 1.45	1.59 – 1.62	2.30 – 2.33
CV%	1.2	1.3	0.8	0.7	3.2	1.5	1.5	1.3
Automatic-manual variation based on automatic search results and averaged manual measurements								
Mean	373.9	265.2	601.1	128.4	49.9	1.41	1.60	2.27
SD	5.2	6.4	10.1	3.2	13.6	0.03	0.04	0.05
95% CI	371.8 – 375.9	262.7 – 267.7	597.1 – 605.1	127.1 – 129.7	44.5 – 55.2	1.40 – 1.42	1.59 – 1.62	2.26 – 2.30
CV%	1.0	1.7	1.2	1.8	19.2	1.6	1.9	1.6

FIGURES

Figure 1

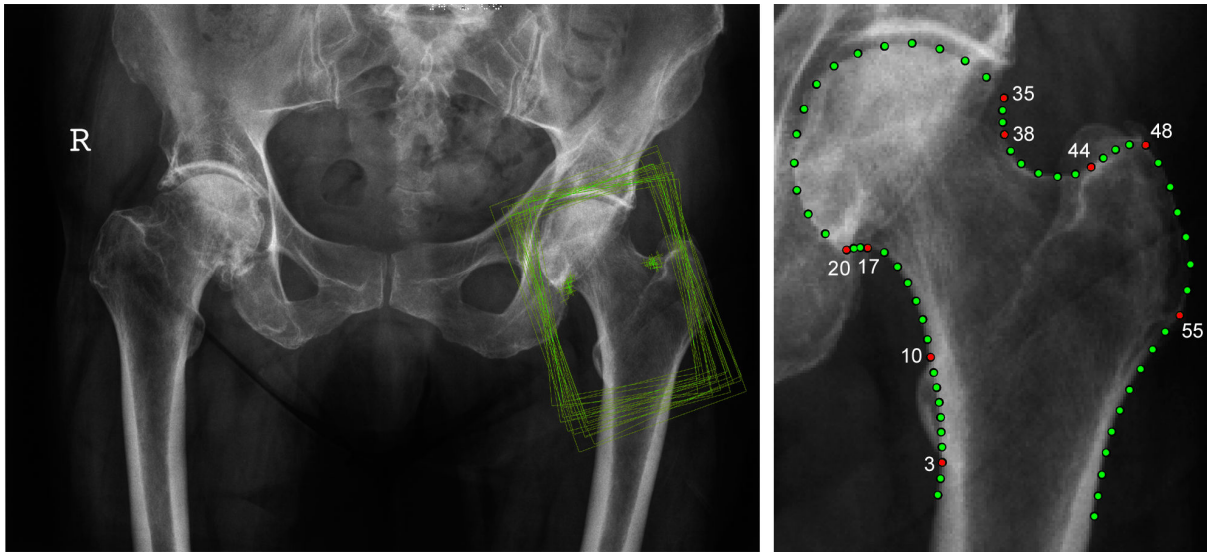

Figure 2

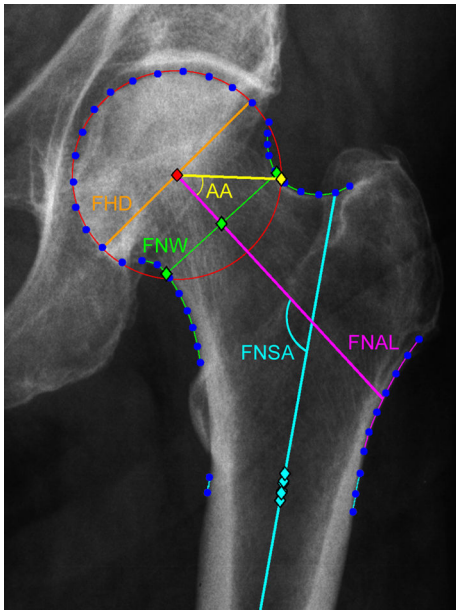

Figure 3

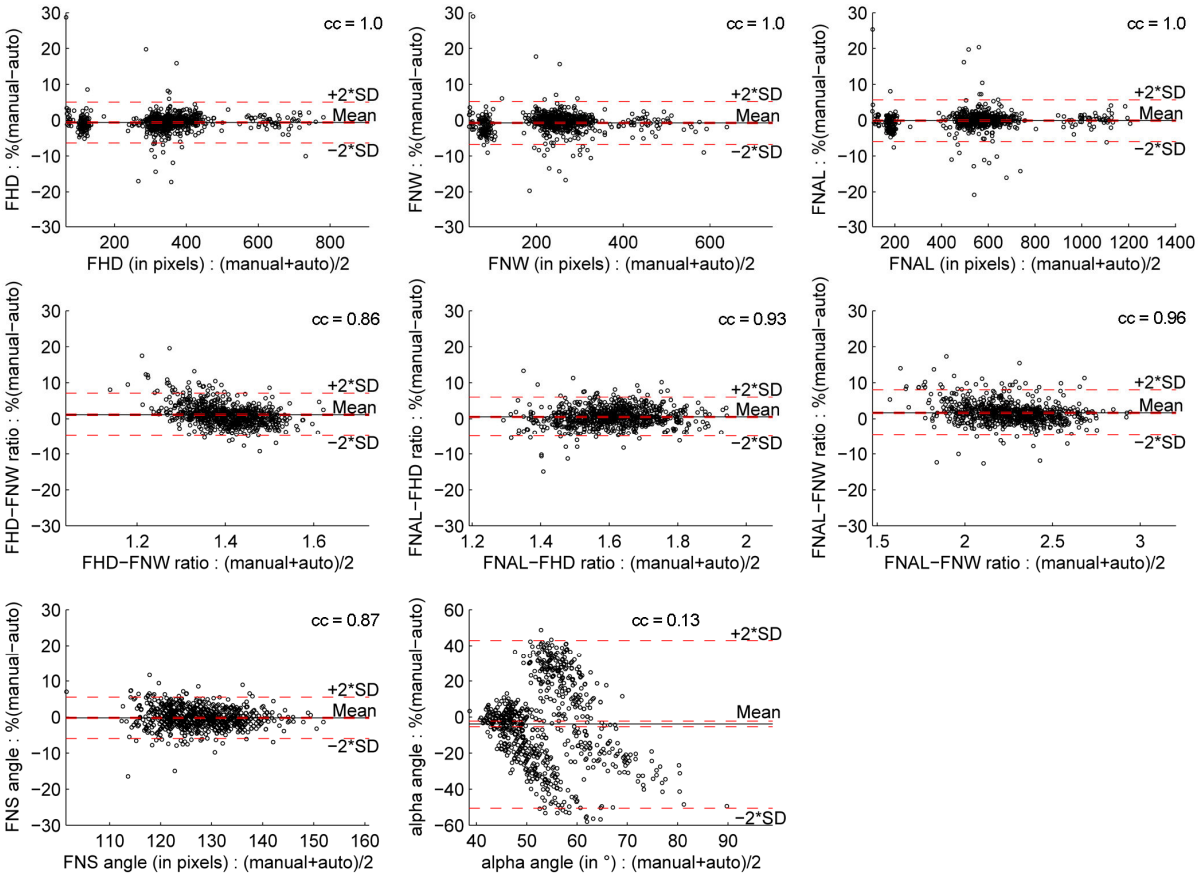
cc = 1.0
+2*SD
Mean
−2*SD
FHD : %(manual−auto)
FHD (in pixels) : (manual+auto)/2
cc = 1.0
FNW : %(manual−auto)
FNW (in pixels) : (manual+auto)/2
cc = 1.0
FNAL : %(manual−auto)
FNAL (in pixels) : (manual+auto)/2
cc = 0.86
FHD−FNW ratio : %(manual−auto)
FHD−FNW ratio : (manual+auto)/2
cc = 0.93
FNAL−FHD ratio : %(manual−auto)
FNAL−FHD ratio : (manual+auto)/2
cc = 0.96
FNAL−FNW ratio : %(manual−auto)
FNAL−FNW ratio : (manual+auto)/2
cc = 0.87
FNS angle : %(manual−auto)
FNS angle (in pixels) : (manual+auto)/2
cc = 0.13
alpha angle : %(manual−auto)
alpha angle (in °) : (manual+auto)/2

Figure 4

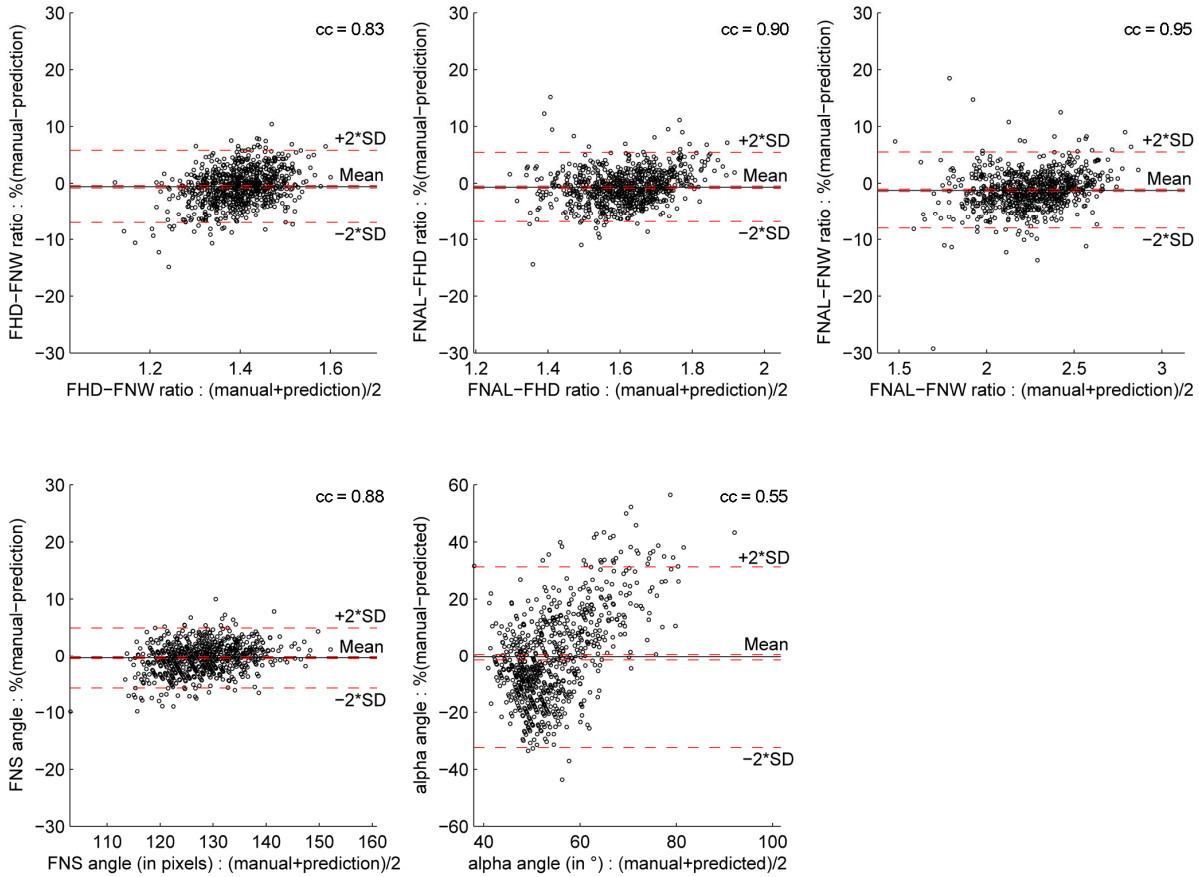

Figure 5