Statistical Shape Modeling of Skeletal Anatomy for Sex Discrimination: Their Training Size, Sexual Dimorphism, and Asymmetry

E. A. Audenaert [1,2,3*], C. Pattyn [1], G. Steenackers [3], J. De Roeck [1], D. Vandermeulen [4,5] and P. Claes [4,5,6,7,8]

[1] Department of Orthopedic Surgery and Traumatology, Ghent University Hospital, Ghent, Belgium, [2] Department of Trauma and Orthopedics, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom, [3] Op3Mech Research Group, Department of Electromechanics, University of Antwerp, Antwerp, Belgium, [4] Medical Imaging Research Center, University Hospitals Leuven, Leuven, Belgium, [5] Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium, [6] Department of Human Genetics, KU Leuven, Leuven, Belgium, [7] Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, VIC, Australia, [8] Department of Biomedical Engineering, University of Oxford, Oxford, United Kingdom

Purpose: Statistical shape modeling provides a powerful tool for describing and analyzing human anatomy. By linearly combining the variance of the shape of a population of a given anatomical entity, statistical shape models (SSMs) identify its main modes of variation and may approximate the total variance of that population to a selected threshold, while reducing its dimensionality. Even though SSMs have been used for over two decades, they lack in characterization of their goodness of prediction, in particular when defining whether these models are actually representative for a given population.

Methods: The current paper presents, to the authors' knowledge, the most extent lower limb anatomy shape model considering the pelvis, femur, patella, tibia, fibula, talus, and calcaneum to date. The present study includes the segmented training shapes ($n = 542$) obtained from 271 lower limb CT scans. The different models were evaluated in terms of accuracy, compactness, generalizability as well as specificity.

Results: The size of training samples needed in each model so that it can be considered population covering was estimated to approximate around 200 samples, based on the generalizability properties of the different models. Simultaneously differences in gender and patterns in left-right asymmetry were identified and characterized. Size was found to be the most pronounced sexual discriminator whereas intra-individual variations in asymmetry were most pronounced at the insertion site of muscles.

Conclusion: For models aimed at population covering descriptive studies, the number of training samples required should amount a sizeable 200 samples. The geometric morphometric method for sex discrimination scored excellent, however, it did not largely outperformed traditional methods based on discrete measures.

Keywords: morphometric analysis, sex discrimination, shape modeling, PCA data analysis, validation and simulation

INTRODUCTION

The increasing use of and ease of access to 3D and 4D imaging technologies has had a tremendous impact on understanding the complexity of human anatomy by enabling detailed non-invasive exploration of the human body. With improved modalities, such as Multi-Detector/Multi-Slice Computed Tomography (MD/MSCT) and Magnetic Resonance Imaging (MRI) to visualize and describe skeletal anatomy, a growing interest in the description of the anatomical variation has emerged and scientific findings have been reported in numerous areas including anthropology, evolutionary biology, forensics, implant design, anatomy, epidemiology, and last but not least clinics, for the distinction of physiological vs. pathological anatomical variation (Audenaert, 2014).

This area of research has grown enormously since the description of geometric morphometrics (Slice, 2007). *Geometric morphometrics* uses homologous landmarks, defined as "a point of correspondence on an object that matches between and within populations" (Slice, 2007). This has resulted to the development of statistical shape models (SSM) that realistically describe anatomy and its variation in any population by conventional multivariate statistics of dense sets of homologous landmarks representing the shape of the underlying structures (Heimann and Meinzer, 2009; Audenaert, 2014). Unlike before, these techniques have the potential for accurate parameterization of complex data such as an individual's morphology or the description of the distribution of anatomy in the population.

SSMs have proved to be extremely valuable in all of the previously defined research areas, especially those involving the human skeleton. For instance, the models can be used to precisely describe skeletal maturity in children (Thodberg et al., 2009), forensic researchers can estimate age, gender, and many other features from skeletal findings (Gehring et al., 2001; Hauser et al., 2005), dysmorphism in rare clinical syndromes can be linked to genetic information and even used for identification of the specific condition (Claes et al., 2012a; Khanduja et al., 2016) automated detection and description of bones and organs in medical images can be performed (Seim et al., 2008; Almeida et al., 2016; Audenaert et al., 2019), detailed 3D images can be generated from sparse radiographical data (Fleute and Lavallee, 1999) reducing radiation exposure in surgical planning, and allowing for individualized clinical and mechanical models of high complexity to be generated from simple routine images.

While there is no doubt on the added value of SSMs in numerous application domains, several limitations in dense geometric morphometry studies of skeletal anatomy do exist. Probably the most relevant, and although applications of SSMs have been increasingly published in the literature, never a model has been effectively studied in terms of sample size properties and a statistical significance in terms of population coverage and generalization properties was similarly never obtained. Furthermore, and directly related to the difficulty in reaching statistically relevant sizes of training samples, issues like sexual dimorphism, or analysis of the left/right asymmetry have not previously been studied in detail, for one exception being the pelvis. The obstetric specialization of the female pelvis is reflected in an unquestioned interest in metrics based sex determination, particularly in forensic, and evolutionary sciences. Geometric morphometric studies are to be expected to increase the robustness of such analysis (Krishan et al., 2016).

Finally, an additional limitation in most studies has been the absence of cortical shape and medullar anatomy. The definition of cortical morphology is clinically and functionally important as not only it is an integral part of lower-limb computational musculoskeletal models, cortical thickness is as well a crucial parameter in stress simulations and the medullar anatomy is crucial for implant design optimization, surgical planning and fitting, and mechanical prosthesis testing and design (Zhang et al., 2016).

The aim of the present work is to develop and describe a detailed statistical shape model of the lower limb skeleton, based on a cohort of 542 samples obtained from 271 CT scans. Secondly, by analyzing such a significant data set, we aim to define the necessary standards for statistical work on human anatomy in terms of appropriate samples sizes to accurately describe a population. Thirdly, we aim to provide improved understanding of sexual dimorphism and asymmetry issues in skeletal anatomy of the lower limb bones.

MATERIALS AND METHODS

A total of 542 training data sets (left and right combined) were considered, originated from 271 CT scans. All scans were processed on a Dell Precision M6800 Laptop (Intel Core i7−4910MQ, 16 GB RAM, 64 bit). Each scan data set consisted of an average of 1864 slices with a pixel size between 0.575 and 0.975 mm. Every image domain included the full lower limb anatomy ranges from rib 12 to toes. The imaging database was constructed from living subjects receiving angio-CT scanning for vascular work-out between 2012 and 2016. CT data demonstrating metallic implants (e.g., hip and knee prosthesis) was excluded from the data base. The participating subjects were not exposed to additional radiation for the present study. This study involving human participants was reviewed and approved by the ethics committee of the Ghent University Hospital (under reference B670201111480). The patients provided their written informed consent to participate in this study.

In-vivo clinical imaging of the human skeleton, and more specifically the lower limb anatomy, through detailed tomographic imaging modalities such as X-ray Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) allows the most complete non-invasive depiction of the morphology of the structures involved in loco-motor function. However, accurate and robust extraction of these structures from a large image database requires automated procedures. The segmentation task for the different bones included in the present study was described in detail in the work by Audenaert et al. (2019). Comparing automatic with manual segmentations demonstrated rooted mean squared differences ranging from 0.53 to 0.76 mm with the largest differences found in the pelvic bones (Audenaert et al., 2019).

Starting from the segmented structures, a dense set of correspondences between homologous structures in the data set were automatically established by a non-rigid mapping of an anthropometric mask (quasi-landmarks) onto the original 3D reconstructions using a selection of readily available point/surface matching techniques (Claes et al., 2012b; Audenaert et al., 2019). Upon completion, all relevant structures in all images were represented as a homologous series of dense landmarks as required for the geometric morphometric analysis and the development of the final statistical analysis of shape of the lower limb anatomy. Homologous defined as each quasi-landmark occupying the same position on each structure relative to all the other quasi-landmarks. Following robust Least Squares (Procrustes) superimposition of these homologous series of dense landmarks to account for uninteresting positional, rotational and, possibly, scale differences, the variance/covariance of morphological differences within a body part over a population can be established (Claes et al., 2012a).

For each sample a right and a left morphology were available. Considering these are coupled data sets with similarities among bony anatomies, right and left morphologies were superimposed using a Procrustes analysis and a perfectly symmetrical consensus configuration was defined. Each single shape is then further decomposed into its bilaterally symmetric part (i.e., the mean consensus shape) and its asymmetric part (i.e. the residue) (Mardia et al., 2000). Model evaluation was than performed following Principal Component Analysis (PCA) of the symmetrical Euclidian data set ($n = 271$), whereas left-right asymmetry was studied following PCA of the asymmetry residuals.

PCA was originally described by Pearson and later adopted by $$Fisher and MacKenzie (Jolliffe and Cadima, 2016). It is one of the oldest and most widely used dimensionality reduction techniques, decomposing a multivariate data set into its mean and corresponding covariance matrix. The eigenvectors of the covariance matrix are usually referred to as principal components or eigenmodes, whereas the eigenvalues indicate there relative importance. The first principal component is usually called the main mode of variation, as it represents the direction of maximal variance within the data. In the particular case of anatomical data this component nearly always defines size differences between subjects. Accordingly, the second principal component represents the direction that maximizes the variance in the data under the constraint that it is orthogonal to the first principal component, and so on. As such, it is a descriptive tool that allows for a systematical exploration of shape variation in a model.

PCA was accordingly used to determine the (co-)variance of morphological differences within the symmetrical data set and for each structure a statistical shape model (Cootes et al., 1995) was generated, described as

$$S = \bar{S} + Pb \qquad (1)$$

with S the shape vector represented as the ordered list of vertex coordinates (following Generalized Procrustes Alignment). $\bar{S}$ defines the corresponding average shape, $P = (p_1, p_2, \ldots p_t)$ the

matrix of eigenvectors of the covariance matrix $(S - \bar{S})^T (S - \bar{S})$, and $b = (b_1, b_2, \ldots b_t)^T$ a vector of weights.

In order to quantitatively evaluate the obtained SSMs several properties were studied. First, for geometric representations of a person's anatomy made from a series of spatially sampled primitives, there needs to be sound correspondence of the primitives across training cases. Reduced correspondence can introduce noise to training samples that can mask existing variation between bone shapes, resulting in an untrustworthy estimation of the shape probability distribution (model specificity) (Styner et al., 2003). Second, geometric depictions need to be efficient and compact, so that they can describe the shape of a structure with a minimal number of parameters (model compactness). Finally, such a model needs to be able to accurately describe members of the population outside of the training sample (model generalization) (Styner et al., 2003; Gollmer and Buzug, 2010). More specifically, and following Styner et al., we implemented the following "goodness" measures: rooted mean squared distance to in-training-set landmarks (accuracy) for different amounts of explained model variance, accumulated variance (compactness), approximation error (RMSE) to each training sample in a leave-one-out setting (generalization) and the average (RMSE) distance of uniformly distributed, randomly generated objects in the model shape space to their nearest member in the training set (specificity) (Styner et al., 2003).

Model Accuracy

The first test analyzes how well-osteological entries within the set used to create the model, are described in terms of accuracy using models capturing different amounts of total variance expressed in percentage. The question to be answered is: How much percentage variance is sufficient or what is disregarded by not including the remaining variance? For clinical usage we defined accuracy as the number of principal components necessary to reach an accuracy for surface definition of 0.6 mm, which is basically below the average resolution of the scanned images. In a first instance, the number of principal components needed to describe a certain amount of variance is determined to construct model descriptions of the osteological entries within the training set. Subsequently, the accuracy of the shape is computed by calculating the Root-Mean-Squared-Error (RMSE) of the distances between corresponding points of the model description S_i' (M) and an original osteological entry of the training set (S_i). The accuracy is computed as the average absolute difference between the model description and the osteological entry. The property value accuracy is the absolute difference between both.

$$Accuracy\,(M) = \frac{1}{K} \sum_{i=1}^{K} \left\| S_i'\,(M) - S_i \right\|^2 \qquad (2)$$

Model Compactness and Size Dominance

A compact shape model is a model that can accurately reconstruct new shape instances with as little shape parameters as possible. Thus, the compactness is defined as the cumulative

explained variance of the Mth eigenmode obtained by the models covariance matrix decomposition (Wang and Shi, 2017).

$$Compactness\,(M) = \frac{1}{\sum \lambda} \sum_{m=1}^{M} \lambda_m \qquad (3)$$

where λ_m is the mth eigenvalue. For compactness, the higher the value is, the lesser variables are required for the constructed shape model to describe its population variance. Considering the variable impact of size of bones on the total population variance, in particular in long bones like femur and tibia, size dominance was reported as the percentual variance described by the first principal component.

Model Generalization

The generalization ability quantifies the capability of the constructed SSM to represent new shape instances of the same class, which are not part of the original training set. The generalization ability is evaluated by performing a series leave-one-out tests on the training data. Having enough training data we expect the model to be able of describing unseen structures quite accurately or to generalize well (Wang and Shi, 2017). The question then becomes: how many training samples are sufficient? This is evaluated by comparing the accuracy evolution to the in-training-set values. Specifically, a shape model is built by using an increasing number of randomly selected training shapes while excluding a target training shape S_i, and then the previously constructed model is than used to reconstruct the excluded shape S_i. The approximation error is consequently defined as the distance (RMSE) between the excluded shape S_i and its reconstructed duplicate $S_i'\,(M)$. The generalization metric is the average reconstruction error over all the performed K tests:

$$Generalization\,(M) = \frac{1}{K} \sum_{i=1}^{K} \left\| S_i - S_i'\,(M) \right\|^2 \qquad (4)$$

where the reconstructed shape $S_i'\,(M)$ is defined as a linear combination of the first M modes of variation:

$$S_i'\,(M) = S + \sum_{m=1}^{M} P_m b_m \qquad (5)$$

A smaller error in the generalization ability of the model designates a better constructed shape model (Wang and Shi, 2017). A model is considered population covering when by increasing the amount of training samples it does not improve the generalization ability of the model any longer.

It is important to note that fitting the excluded shape entries with the respective shape models comes with a number of particular challenges. Firstly, during the alignment of the "unknown" shapes, local dysmorphologies (e.g., a trochlear bump up to 6 mm), can lead to alignment errors during the ICP procedure. These abnormalities are typically localized in small areas and do not extend over the whole region of interest. To account for general disturbance during alignment of the data sets by these local abnormalities, the so-called "Pinocchio effect,"

iterative exclusion of outliers at the 0.05 level was performed while using bidirectional (target to source and back) vertex correspondences, to obtain a robust alignment prior to solving the principal component weights needed to describe the shapes (Besl and Mckay, 1992; Claes et al., 2012a; Audenaert et al., 2019). Secondly, fitting the excluded shapes into the models to define their respective PC weights is solved iteratively using a deterministic annealing scheme by progressively increasing the number of principal components used to describe the shape entry of interest. Similarly to the alignment problem, using bidirectional ICP information results into an overdetermined set of linear equations that can efficiently be solved in a least squares sense using the Moore-Penrose pseudoinverse matrix. Details of the different algorithms used can be found in Audenaert et al. (2019), further all coding used is open source and available at Matlab® file exchange (https://nl.mathworks.com/matlabcentral/profile/authors/4165925-manu).

Model Specificity

The specificity measures the realistic construct of new shape instances randomly generated by the developed shape model. It is measured by generating a large set (N) of virtual shape examples using the constructed model and calculate their difference from real samples available in the training set (Wang and Shi, 2017). The approximation error is defined as the distance between the generated shape instance S_j and its most similar sample in the training data S_j'. The specificity metric is than defined as the averaged approximation error of all the generated N shape instances:

$$Specificity\,(M) = \frac{1}{N} \sum_{j=1}^{N} \left\| S_j\,(M) - S_j' \right\|^2 \qquad (6)$$

where the shape instance S_j is generated by choosing random normal distributed values n for the first shape parameter b from the range $b_m \in \left[-n\sqrt{\lambda_m}, n\sqrt{\lambda_m} \right]$; with λ_m representing the mth eigenvalue.

$$S_j\,(M) = \bar{S} + \sum_{m=1}^{M} b_m P_m \qquad (7)$$

and the most similar sample S_j' is defined as:

$$S_j' = \arg \min_{k\,\in\,[1,k]} \left\| S_j\,(M) - S_j' \right\| \qquad (8)$$

A value of K = 10,000 was used to obtain the results reported in this study.

Sexual Dimorphism and Asymmetry

Sexual dimorphism implies sex interactions in patterns of underlying gene expression and function resulting in phenotypic differences between the sexes (Claes et al., 2012c). The gender-shape relationship was evaluated by means of canonical correlation analysis. In particular, the PC weights of the training data, serving as predictor variables for the observed male (+1)

either female (−1) gender, were used. Overall explained variance in the observed shape components by the factor gender was evaluated by means of partial least squares (PLS) regression. Gender differentiation potential of the models was established by Linear Discriminant Analysis and its accuracy in a k-fold leave-one-out simulation. As previously mentioned, PCA on the asymmetrical residuals was used to describe the most relevant asymmetry findings in the study population.

RESULTS

The obtained dataset of training data represents 181 male and 90 female cases with an average age of 67.8 (±10.8) and 69 (±13.3) years, respectively. Each case was represented by its averaged symmetrical consensus as well as the remaining asymmetry component. Model evaluation was than performed onto the consensus shapes ($n = 271$).

Segmentation of the data sets was performed fully automatically and based on a validated segmentation protocol for which details can be found in the work by Audenaert et al. (2019). The segmented structures used for construct and analysis of the respective SSMs, included the pelvis, femur, patella, fibula, tibia, talus, and calcaneum.

Model Accuracy
The following procedure was performed for every entry in the database and the accuracy results for the different information parts were averaged and presented in **Table 1**. Considering the relative dominance of the factor size in skeletal anatomy, in particular for long bones, we chose to define the level of aimed in-sample accuracy in terms of number of PCs required to reach a minimal and clinically relevant submillimeter (image resolution - size) distance error of 0.6 mm while including at least 95% of data variance. Hence, the amount of principal components needed for such accuracy, the variance and the accuracy are reported as well in **Table 1**.

Model Compactness
Figure 1 shows the cumulative compactness for the different models, for increasing number of modes of variation included. For all structures investigated the first and dominant PC, as evaluated by visual inspection, involved nearly exclusively

difference in size between subjects. The amount of variance described by this first PC is reported as well in **Table 1**.

Model Generalization
Figure 2 depicts the shape accuracy evolution with increasing amount of training shapes for the different human bones investigated. It can be seen that the accuracy errors approximate their in-training-set values as more training samples are added. Despite the small difference at the end of the curve (using 270 training shapes) we can conclude that the amount of training shapes in our database seems sufficient to represented the population described by the model. When applying an arbitrarily measure of convergence of $<10^{-3}$ mm in improvement of RMS generalization error evolution with increasing number of training samples, convergence was obtained ranging from 160 samples for to calcaneum to 210 samples for the pelvic bone.

Model Specificity
Model specificity results are as well-presented in **Table 1** and range from 0.6 to 2.08 mm, the highest value being obtained for the pelvis.

Asymmetry
Pelvic asymmetry was predominantly located at the insertion sites of muscle groups (**Figure 3**), namely the demonstrating superior SIAS, hamstrings and adductor insertion regions. For femur and tibia asymmetry was mostly pronounced at the site of the joints, in particular hip and ankle. The smaller bones, patella, calcaneum, and talus all presented with submillimeter left/right differences. Asymmetry was only poorly gender pronounced, with 1.8% of variance being explained by gender. Asymmetry of the femur, although not pronounced, was most evident around the femur head. On overview of average asymmetry findings is presented in **Figure 3**.
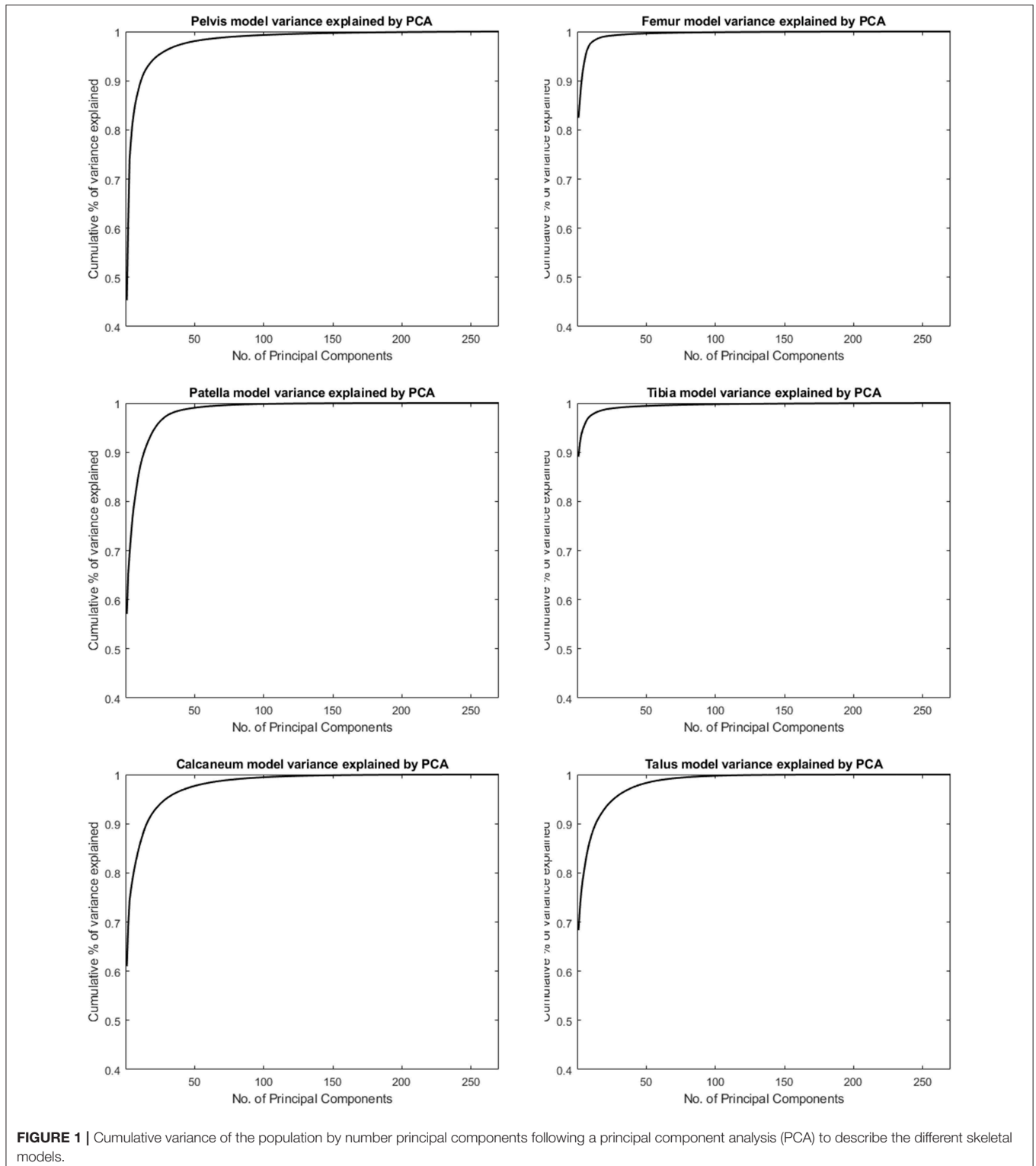
Sexual Dimorphism
Gender accounted from 12.69 up to 42.31% of anatomical variance and predominantly accounted for variation in the first two modes of variation. Sex discrimination was robustly performed in the pelvic bone anatomy (100% success rate), and could be estimated with reasonable probability by the long bones tibia and femur. On overview of gender differentiation capabilities is presented in **Table 2**. The patella bone provided

TABLE 1 | Accuracy, generalization, specificity, and compactness results for the different SSMs considered.

	Percent of variance explained by size (%)	Nr. of PCs included	Percent of variance explained by the included PCs (%)	Model accuracy (RMSE mm)	Model specificity (RMSE mm)
Pelvis	45.35	39	97.47	0.59 ± 0.08	2.05 ± 0.35
Femur	82.45	25	99.19	0.59 ± 0.06	1.91 ± 0.28
Tibia	89.09	21.	98.74	0.59 ± 0.06	1.36 ± 0.21
Patella	63.59	21	95.01	0.31 ± 0.07	0.71 ± 0.11
Calcaneum	61.02	29	95.03	0.36 ± 0.05	0.93 ± 0.15
Talus	68.37	27	95.16	0.25 ± 0.08	0.63 ± 0.08

The impact of the first component describing predominantly size in the models is, respectively, described in the first column.

FIGURE 1 | Cumulative variance of the population by number principal components following a principal component analysis (PCA) to describe the different skeletal models.

the least reliable predictive features. Male/female differences as obtained following the canonical correlation analysis are demonstrated in **Figure 3**, average male/female canonical scores were amplified with a factor 2. Clearly distinguishable features

include narrower pubis in males, narrower subpubic angle and overall narrower pelvic inlet as compared to males. For the femur, clinical relevant particularities were as well-observed. Although female subjects have overall smaller sizes, a relative
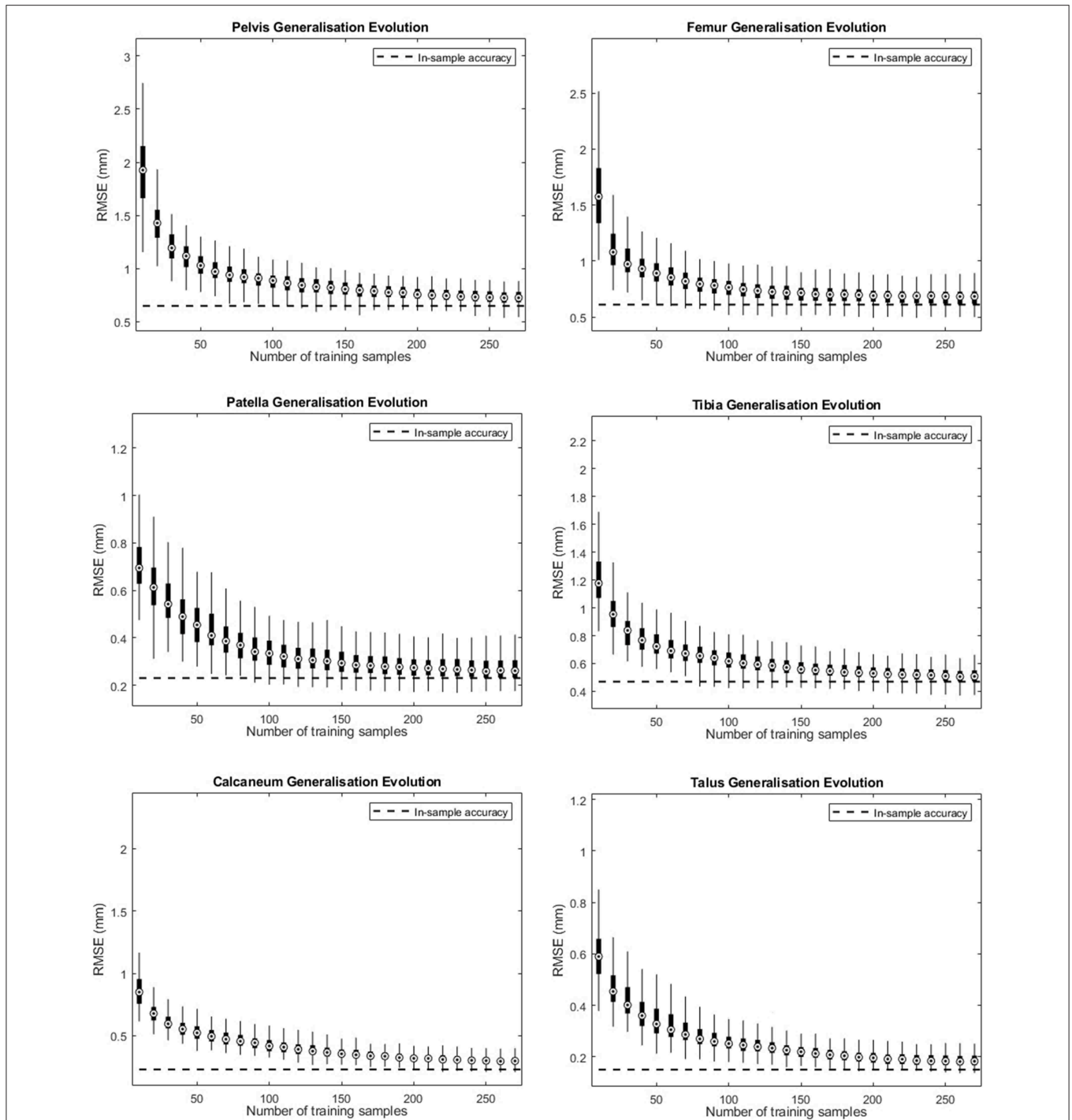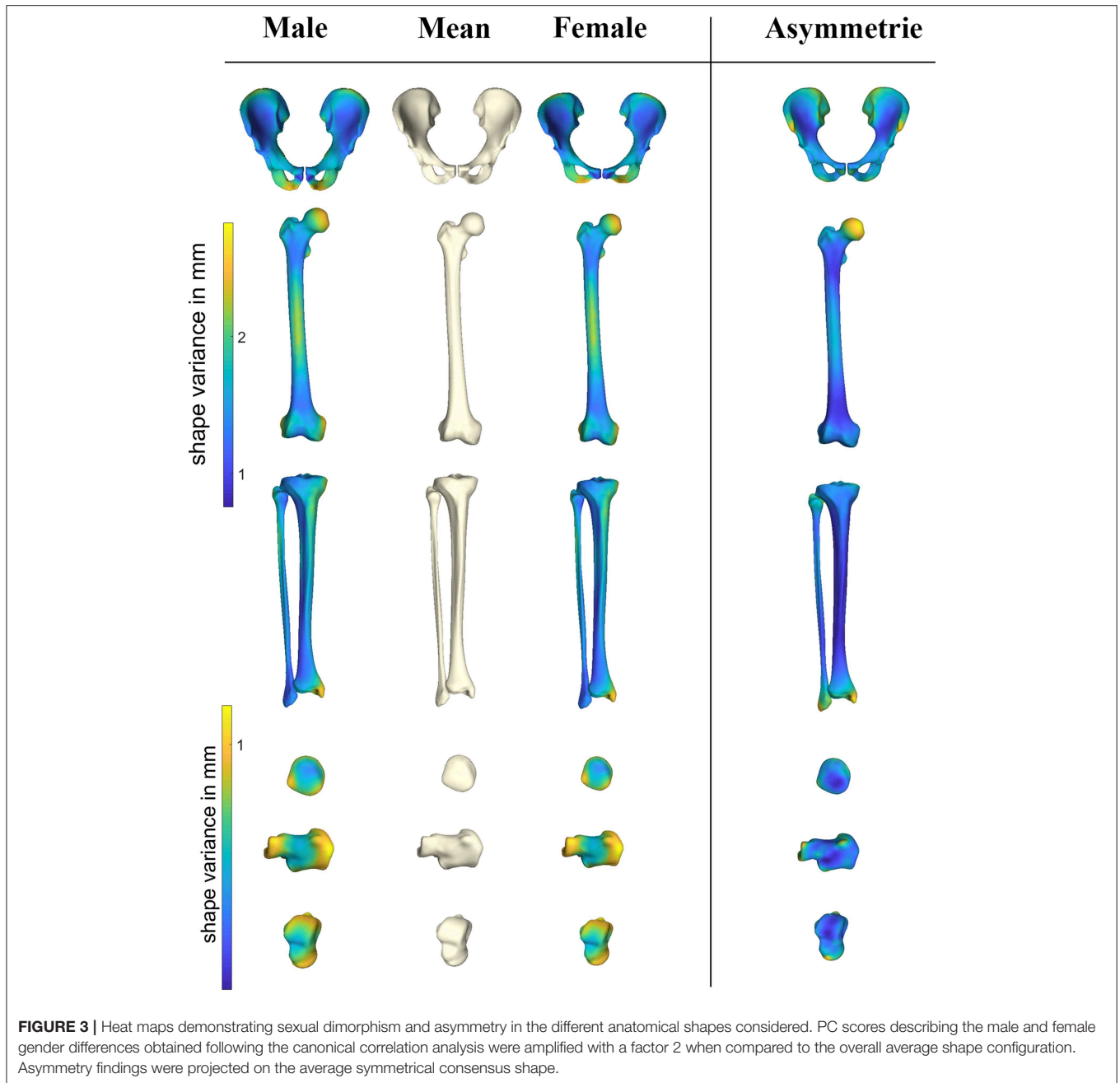
FIGURE 2 | Accuracy evolution of the SSM-based shape representation (solid curve) and in-sample target accuracy (dotted curve) for different levels of prior knowledge expressed as amounts of training data in the SSM.

narrowing of the condylar width, a decreased cortical thickness and a relatively smaller diameters of the femoral head were noticed. The tibia plateau in accordance with the femoral condyles, presented smaller sizes in females as compared to males. Similarly, the tibia pilon about the ankle presented smaller in width in females as compared to males. The smaller bones, patella, calcaneum, and talus all presented on average compacter in all dimensions in females as compared to males. An overview of the gender differences is provided in **Figure 3**.

FIGURE 3 | Heat maps demonstrating sexual dimorphism and asymmetry in the different anatomical shapes considered. PC scores describing the male and female gender differences obtained following the canonical correlation analysis were amplified with a factor 2 when compared to the overall average shape configuration. Asymmetry findings were projected on the average symmetrical consensus shape.

DISCUSSION

It is generally accepted that skeletal anatomy is defined by a person's genetic background, his environment and his functional level. The genetic background of morphogenesis, however, lacks a clear Mendelian pattern of inheritance and is associated with interactions of different genes, each conferring small effects (i.e., a polygenetic complex trait). Environmental and functional variables in combination with a person's genetic background affect growth during adolescence, whereas morphology impacts importantly on the stress distribution within biological tissue,

which in turn can be the signal or trigger for a specific gene expression (Bastow et al., 2005; Waarsing et al., 2011; Baker-Lepain et al., 2012; Rolfe et al., 2014). Recently, a number of authors have successfully started to explore the potential of genetic association analysis of multivariate/multiple phenotypes using in depth profiling such as statistical models of shape (Lindner et al., 2015). Clearly, such analysis requires a valid representation of a subject's anatomical phenotype and its positioning within a given population.

Where any structured *a priori* knowledge is valuable in applications such as medical image segmentation, claiming a

TABLE 2 | Canonical correlation analysis and partial least squares regression results relating variation in shape with gender and sexual discriminative features of the different SSMs described.

	Correlation r gender-shape	Percent of variance explained by gender (%)	Percent of correct male gender discrimination (%)	Percent of correct female gender discrimination (%)
Pelvis	0.97 ($p = 4.1e\text{-}117$)	12.69	100	100
Femur	0.88 ($p = 3.9e\text{-}67$)	27.28	97.24	95.56
Tibia	0.89 ($p = 9.8e\text{-}73$)	31.71	98.34	96.67
Patella	0.75 ($p = 1.9e\text{-}32$)	18.96	91.71	83.33
Calcaneum	0.82 ($p = 4.4e\text{-}44$)	33.21	92.82	85.56
Talus	0.84 ($p = 3.34e\text{-}50$)	42.31	93.92	86.67

shape model is valid for representation of a given population in other medical applications (e.g., genetic studies and implant design) is not that obvious. In the present study we aimed to provide the boundaries in terms of sample sizes for the description of human osteology in population covering studies. As a rule of thumb we can conclude from our data that a minimum of 200 training samples are required to sufficiently cover population variance. Do note that this rule of thumb applies for the population that was investigated, which is a single homogenous population of European Descent. Conversely, these numbers can further increase for heterogeneous and admixed populations.

The secondary aim of this study was to evaluate gender differences, as well as to describe common patters of asymmetry within subjects. While sex prediction is a common problem in anthropology and forensic sciences, these areas have mainly focused on discrete measures, lines and distances between landmarks, in particular of pelvis and proximal femur, whereas the exact description of shape variation has rarely been the focus. In particular, the pelvis, given its obstetric relation, has been the focus in several studies. Studies of sexual dimorphism in the human pelvis show that while, in general, many pelvic characteristics reflect full body size, and are therefore larger in men than in women, other dimensions of the pelvic canal follow the inverse model (Tague, 2000; Kurki, 2011). The use of geometric morphometrics provides an unmet means to investigate and describe sexual dimorphism as to differentiate based on relevant shape predictors between genders. Interestingly, from our results it appeared that the geometric morphometric method for sex discrimination did not largely outperformed traditional methods based on discrete measures. It appears that size dominance in the data variation, an excellent discriminator on its own, importantly weakens the magic of geometric morphometrics. Traditional methods, however, are fast, applicable in the field and can be used when specimens are incomplete or even fractured. Based on their ease to use and the results found in this work we can only conclude they still stand as valid alternatives. Novel techniques are in the phase of development that might be readily applicable in the field with the same advantages of geometric morphometrics including

promising neural network applications such as geometric deep learning (Bronstein et al., 2017).

Our findings also seem to support some relevant clinical findings. In particular, gender differences around the knee have been a recent subject of interest in knee arthroplasty. A significant difference in knee width was demonstrated between male and female samples and presented the most pronounced component of variation (excluding size). This is in agreement with previous clinical reports (Chin et al., 2002; Hitt et al., 2003) and the industry has even adopted this concept for the development of gender specific implants. Although interesting from a commercial point of view, no evidence of any outcome advantage in the gender-specific design could be demonstrated in randomized clinical trials (Thomsen et al., 2012).

Yet another interesting finding that asymmetry in subjects coincides with the attachment site of muscles, clearly present in the pelvis for example, might be attributed to differences in left-right dominance. Obviously further studies are required to substantiate such claims. Probably of more clinical importance are the pronounced asymmetry findings around the femoral head and by extension the position and orientation of the natural center of rotation. While for several years it has been common clinical practice to use the contralateral hip for surgical planning and templating in case of hip arthroplasty for reason of severe deformity by trauma or bone necrosis, recently several authors have warned for important asymmetry to exist in the proximal femur. Our findings strongly support this finding and confirm that asymmetry may lead to inaccurate anatomical restoration of the hip if the templated sizes are routinely implanted (Dimitriou et al., 2016; Laumonerie et al., 2018).

An important limitation of the present works relates to the population under investigation, namely Belgian people and the unknown extent of which findings can be extrapolated to other populations. The complex interaction between environment, culture, and the genes, results in a population-based variation, with numerous studies demonstrating that the appropriate evaluation of this variation necessitates specific standards for each population (Rissech et al., 2013; San-Millán et al., 2017). Nevertheless, in general terms we expect our results to be representative by extension for a Western European population.

CONCLUSION

In conclusion and based on the quantitative results of the employed model valuation metrics, we dare to claim that the overall quality of the constructed shape models is high and that therefore the presented models can be employed for effective clinical, forensic, and population wide applications.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

This study involving human participants was reviewed and approved by the ethics committee of the Ghent University Hospital (under reference B670201111480). The patients provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EA, DV, and PC designed the algorithms. GS, CP, and JD assisted in the data collection and manipulation. All assisted in manuscript drafting.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Almeida, D. F., Ruben, R. B., Folgado, J., Fernandes, P. R., Audenaert, E., Verhegghe, B., et al. (2016). Fully automatic segmentation of femurs with medullary canal definition in high and in low resolution CT scans. *Med. Eng. Phys.* 38, 1474–1480. doi: 10.1016/j.medengphy.2016.09.019

Audenaert, E. (2014). "Development of a multi-modal, multi-component, articulated model of the lower limb," in *A Laborious Work in Progress*, ed G. Székely (Delémont: Shape; SICAS). p. 38.

Audenaert, E. A., Van Houcke, J., Almeida, D. F., Paelinck, L., Peiffer, M., Steenackers, G., et al. (2019). Cascaded statistical shape model based segmentation of the full lower limb in CT. *Comput. Methods Biomech. Biomed. Eng.* 22, 644–657. doi: 10.1080/10255842.2019.1577828

Baker-Lepain, J. C., Lynch, J. A., Parimi, N., McCulloch, C. E., Nevitt, M. C., Corr, M., et al. (2012). Variant alleles of the Wnt antagonist FRZB are determinants of hip shape and modify the relationship between hip shape and osteoarthritis. *Arthritis Rheumatism* 64, 1457–1465. doi: 10.1002/art.34526

Bastow, E. R., Lamb, K. J., Lewthwaite, J. C., Osborne, A. C., Kavanagh, E., Wheeler-Jones, C. P., et al. (2005). Selective activation of the MEK-ERK pathway is regulated by mechanical stimuli in forming joints and promotes pericellular matrix formation. *J. Biol. Chem.* 280, 11749–11758. doi: 10.1074/jbc.M414495200

Besl, P. J., and Mckay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* 14, 239–256. doi: 10.1109/34.121791

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. (2017). Geometric deep learning going beyond euclidean data. *IEEE Signal. Process. Mag.* 34, 18–42. doi: 10.1109/MSP.2017.2693418

Chin, K. R., Dalury, D. F., Zurakowski, D., and Scott, R. D. (2002). Intraoperative measurements of male and female distal femurs during primary total knee arthroplasty. *J. Knee Surg.* 15, 213–217.

Claes, P., Daniels, K., Walters, M., Clement, J., Vandermeulen, D., and Suetens, P. (2012a). Dysmorphometrics: the modelling of morphological abnormalities. *Theor. Biol. Med. Model.* 9:5. doi: 10.1186/1742-4682-9-5

Claes, P., Walters, M., and Clement, J. (2012b). Improved facial outcome assessment using a 3D anthropometric mask. *Int. J. Oral Maxillofac. Surg.* 41, 324–330. doi: 10.1016/j.ijom.2011.10.019

Claes, P., Walters, M., Shriver, M. D., Puts, D., Gibson, G., Clement, J., et al. (2012c). Sexual dimorphism in multiple aspects of 3D facial symmetry and asymmetry defined by spatially dense geometric morphometrics. *J. Anat.* 221, 97–114. doi: 10.1111/j.1469-7580.2012.01528.x

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models - their training and application. *Comput. Vision Image Understand.* 61, 38–59. doi: 10.1006/cviu.1995.1004

Dimitriou, D., Tsai, T. Y., Yue, B., Rubash, H. E., Kwon, Y. M., and Li, G. (2016). Side-to-side variation in normal femoral morphology: 3D CT analysis of 122 femurs. *Orthopaed. Traumatol. Surg. Res.* 102, 91–97. doi: 10.1016/j.otsr.2015.11.004

Fleute, M., and Lavallee, S. (1999). "Nonrigid 3-D/2-D registration of images using statistical models," in *Medical Image Computing and Computer-Assisted Intervention, Miccai'99, Proceedings* (Cambridge).

Gehring, K. D., Haffner, H. T., Weber, D., and Graw, M. (2001). Investigations on the reliability of determining an individual's age from the proximal femur. *Homo J. Compar. Hum. Biol.* 52, 214–220. doi: 10.1078/0018-442X-00029

Gollmer, S. T., and Buzug, T. M. (2010). "A method for quantitative evaluation of statistical shape models using morphometry," in *2010 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (Melbourne, VIC).

Hauser, R., Smolinski, J., and Gos, T. (2005). The estimation of stature on the basis of measurements of the femur. *Forensic Sci. Int.* 147, 185–190. doi: 10.1016/j.forsciint.2004.09.070

Heimann, T., and Meinzer, H. P. (2009). Statistical shape models for 3D medical image segmentation: a review. *Med. Image Anal.* 13, 543–563. doi: 10.1016/j.media.2009.05.004

Hitt, K., Shurman, J. R II, Greene, K., McCarthy, J., Moskal, J., Hoeman, T., et al. (2003). Anthropometric measurements of the human knee: correlation to the sizing of current knee arthroplasty systems. *J. Bone Joint Surg. Am.* 85-A(Suppl. 4), 115–122. doi: 10.2106/00004623-200300004-00015

Jolliffe, I. T., and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* 374. doi: 10.1098/rsta.2015.0202

Khanduja, V., Baelde, N., Dobbelaere, A., Van Houcke, J., Li, H., Pattyn, C., et al. (2016). Patient-specific assessment of dysmorphism of the femoral head-neck junction: a statistical shape model approach. *Int. J. Med. Robot.* 12, 765–772 doi: 10.1002/rcs.1726

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., and Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Sci. Int.* 261, 165.e1–165.e8. doi: 10.1016/j.forsciint.2016.02.007

Kurki, H. K. (2011). Pelvic dimorphism in relation to body size and body size dimorphism in humans. *J. Hum. Evol.* 61, 631–643. doi: 10.1016/j.jhevol.2011.07.006

Laumonerie, P., Ollivier, M., LiArno, S., Faizan, A., Cavaignac, E., and Argenson, J. N. (2018). Which factors influence proximal femoral asymmetry? *Bone Joint J.* 100-B, 839–844. doi: 10.1302/0301-620X.100B7.BJJ-2017-1601.R1

Lindner, C., Thiagarajah, S., Wilkinson, J. M., Panoutsopoulou, K., Day-Williams, A. G., arcOGEN Consortium, et al. (2015). Investigation of association between

hip osteoarthritis susceptibility loci and radiographic proximal femur shape. *Arthritis Rheumatol.* 67, 2076–2084. doi: 10.1002/art.39186

Mardia, K. V., Bookstein, F. L., and Moreton, I. J. (2000). Statistical assessment of bilateral symmetry of shapes. *Biometrika* 87, 285–300. doi: 10.1093/biomet/87.2.285

Rissech, C., Márquez-Grant, N., and Turbón, D. (2013). A collation of recently published Western European formulae for age estimation of subadult skeletal remains: recommendations for forensic anthropology and osteoarchaeology. *J. Forensic Sci.* 58(Suppl. 1), S163–S168. doi: 10.1111/1556-4029.12011

Rolfe, R. A., Nowlan, N. C., Kenny, E. M., Cormican, P., Morris, D. W., Prendergast, P. J., et al. (2014). Identification of mechanosensitive genes during skeletal development: alteration of genes associated with cytoskeletal rearrangement and cell signalling pathways. *BMC Genomics* 15:48. doi: 10.1186/1471-2164-15-48

San-Millán, M., Rissech, C., and Turbón, D. (2017). Shape variability of the adult human acetabulum and acetabular fossa related to sex and age by geometric morphometrics. Implications for adult age estimation. *Forensic Sci. Int.* 272, 50–63. doi: 10.1016/j.forsciint.2017.01.005

Seim, H., Kainmueller, D., Heller, M., Lamecker, H., Zachow, S., and Hege, H.-C. (2008). "Automatic segmentation of the pelvic bones from ct data based on a statistical shape model," in *Eurographics Workshop on Visual Computing for Biomedicine (VCBM)* (Leipzig), 93–100.

Slice, D. E. (2007). Geometric morphometrics. *Annu. Rev. Anthropol.* 36, 261–281. doi: 10.1146/annurev.anthro.34.081804.120613

Styner, M. A., Rajamani, K. T., Nolte, L. P., Zsemlye, G., Székely, G., Taylor, C. J., et al. (2003). Evaluation of 3D correspondence methods for model building. *Inf. Proc. Med. Imaging Proc.* 2732, 63–75. doi: 10.1007/978-3-540-45087-0_6

Tague, R. G. (2000). Do big females have big pelves? *Am. J. Phys. Anthropol.* 112, 377–393. doi: 10.1002/1096-8644(200007)112:3<377::AID-AJPA8>3.0.CO;2-O

Thodberg, H. H., Kreiborg, S., Juul, A., and Pedersen, K. D. (2009). The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans. Med. Imaging* 28, 52–66. doi: 10.1109/TMI.2008.926067

Thomsen, M. G., Husted, H., Bencke, J., Curtis, D., Holm, G., and Troelsen, A. (2012). Do we need a gender-specific total knee replacement? A randomised controlled trial comparing a high-flex and a gender-specific posterior design. *J. Bone Joint Surg. Br.* 94, 787–792. doi: 10.1302/0301-620X.94B6.28781

Waarsing, J. H., Kloppenburg, M., Slagboom, P. E., Kroon, H. M., Houwing-Duistermaat, J. J., Weinans, H., et al. (2011). Osteoarthritis susceptibility genes influence the association between hip morphology and osteoarthritis. *Arthritis Rheumatism* 63, 1349–1354. doi: 10.1002/art.30288

Wang, J., and Shi, C. (2017). Automatic construction of statistical shape models using deformable simplex meshes with vector field convolution energy. *Biomed. Eng. Online* 16:49. doi: 10.1186/s12938-017-0340-0

Zhang, J., Hislop-Jambrich, J., and Besier, T. F. (2016). Predictive statistical models of baseline variations in 3-D femoral cortex morphology. *Med. Eng. Phys.* 38, 450–457. doi: 10.1016/j.medengphy.2016.02.003