

# Sex prediction from the femur and hip bone using a sample of CT images from a Spanish population

Ana Clavero · Miquel Salicrú · Daniel Turbón

Received: 27 February 2014 / Accepted: 28 August 2014 / Published online: 1 October 2014  
© Springer-Verlag Berlin Heidelberg 2014

**Abstract** Sex estimation and the analysis of sexual dimorphism is an essential part of forensic and archaeological studies of skeletons. However, osteologists often have to rely on single measurements, such as femoral head diameters, to estimate sex, especially when skeletons are incomplete. We have obtained a sex-prediction model based on CT images by applying the logistic regression technique to the measurements obtained for the proximal femoral epiphyses and coxal. Nine variables for 114 Spaniards (58 females and 56 males) of known age and sex from a region close to Madrid have been studied. The prediction equation obtained using these nine variables correctly classifies 99.1 % of these individuals. Reducing the equation to the three most explanatory variables (VDH, HDH and MIB) resulted in the correct classification of 98.3 %. These findings suggest that this procedure is highly effective for sex prediction. However, a lack of expertise may produce biases in the measurements obtained from CT images. Moreover, these equations are only most effective for the population for which they were calculated as human growth and body size are sensitive to nutritional variations, environmental stress and the so-called secular trend.

**Keywords** Sex prediction · Sexual dimorphism · CT images · Logistic regression · Femur bone · Os coxa

A. Clavero · D. Turbón  
Department of Animal Biology, Faculty of Biology, University of Barcelona, Av. Diagonal, 643, 08028 Barcelona, Spain

M. Salicrú  
Department of Statistics, Faculty of Biology, University of Barcelona, Av. Diagonal, 643, 08028 Barcelona, Spain

D. Turbón (✉)  
Department of Animal Biology, Faculty of Biology, University of Barcelona, 08020 Barcelona, Spain  
e-mail: turbon@ub.edu

## Introduction

Human growth and body size are sensitive to environmental quality. In addition, human populations living in the same region may experience changes in growth and development over successive generations, which has been termed the secular trend. As this can affect sexual development, regional differences in anthropometric standards have to be constantly renewed. In human skeletal biology this has been observed when sexing skulls [1], as discriminant functions that perform well in one population may produce high error rates and sex determination biases when applied to another population. The complex interaction between genes, culture and the environment results in a population-based variation, with several studies showing that the appropriate evaluation of this variation requires specific standards for each population [2–5].

Skeletal sexual dimorphism is of interest in physical anthropology and forensic science as it is critical for the identification of individuals and analysis of the lifestyles of ancient human populations. Moreover, sex diagnosis in adult individuals is often used to subsequently calculate height, weight or body structure [6]. As a result, numerous efforts have been directed towards achieving maximum reliability in sex prediction, thereby indirectly increasing the reliability when predicting other variables when sex-differentiated models are required.

When all information is available, the skull and pelvis are the best indicators of sexual dimorphism in the human skeleton. The obstetric specialisation of the female pelvis is reflected in a broad range of indices and angles, thereby justifying the unquestioned interest of the pelvis in metrics-based sex determination. In addition, the sexual dimorphism of the cranium has been highlighted by way of variables that characterise the mental eminence, orbital margin, glabellar area, nuchal area and mastoid process [7, 8]. Moreover, the diagnostic accuracy when postcranial skeletal bones are used

has improved markedly in recent years [9]. Thus, using radiological techniques, a wide range of postcranial bones have been used to improve the reliability of sex prediction, characterise dimensions or predict age, amongst others [10–16].

Bone remnants are occasionally poorly preserved (damaged, fragmented), thus meaning that bone information is incomplete. The femur, pelvis and other postcranial bones have been used as sex-diagnosis tools on such occasions, and their metrics have also been related to the vertical dimensions of the human body, mainly height. The femur presents high predictive accuracy and good post-mortem preservation in many populations because of its strength and toughness [17–20]. In contrast, the greater fragility of the pelvis explains its more marked deterioration and, consequently, the possibility of losing relevant information. Under these conditions, sex prediction and the estimation of vertical dimensions for the human body may inevitably be limited to relationships with one or only a few variables that provide little information.

When the bone is fragmented or damaged, the accuracy of sex diagnosis when using one or only a few pelvic variables may be lower than that obtained using other postcranial bones. This effect has been highlighted by comparing the predictive ability of the most informative variable for the pelvis (possibly damaged) with the most informative variable provided by other, more robust bones. In this regard, for the white population of South Africa, pelvic metrics have provided an 86 % correct sex classification using the ischium length variable [21], whereas femoral metrics have provided a 90.5 % correct classification using the distal breadth of the femur variable [22]. Similarly, in the white American population, the os coxa height variable has provided a reliability of 85 %, whereas the corresponding value for tibial proximal epiphyseal breadth was 90 %; in a sample of black Americans, the ischium length provided a reliability of 83 %, whereas the value for the femoral epicondylar breadth was 89 % [9].

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. Consequently, depending on the variables considered, the equations that relate pelvic metrics with the size or height of the body can differ markedly for males and females [23–25]. As a result, the use of pelvic metrics to predict the vertical dimensions of the human body from fragmented or damaged bones may not be particularly robust when errors in sex prediction are significant. This transmission of errors (from the sex to the vertical measurements) is less pronounced when femoral metrics and those for other postcranial bones, the characteristics of which do not present such divergent patterns, are used [13, 26]. When bones are broken or damaged, the combined use of pelvic variables with those for other postcranial bones may increase the predictive ability.

The study of the proximal epiphysis of the femur is of great value for sexual diagnosis when the pelvis is missing or damaged, which is useful in forensic and palaeoanthropological studies. The sexual dimorphism of the femoral head has been studied in several populations and, when the femur is well preserved, the widths of the femoral head, distal epiphysis and midshaft, respectively, have been found to be highly sexually dimorphic. One of the first studies of sexual dimorphism in a Spanish population, namely a series of 132 femora from an osteological collection deposited in Madrid, achieved an accuracy of between 84 and 97 % when each variable was treated independently and 99 % when two variables for the femoral epiphyses were combined [27]. The vertical and transverse diameters of this feature have been found to vary from one population to another [28]. Mall et al. have reported measurements for six femoral dimensions in a contemporary German population [29]. When applying a univariate treatment, only the femoral head diameters provided correct classification values higher than 86 %. A combination of the two least discriminatory variables, namely the diameter of the femoral diaphysis and the circumference of the femoral head, led to a value of 91.7 %. Srivastava et al. have reported measurements for eight standard parameters to establish osteometric standards for the femur of contemporary North Indian populations, adult femora of known sex ( $n=122$ ; M, 94; F, 28) [30]. Epicondylar breadth, proximal breadth, and antero-posterior diameter of the lateral condyle were found to be the most discriminating variables providing an accuracy of 90.2 %.

Purkait developed a triangle of measurements in the posterior region of the proximal femur which achieved an accuracy of 85 % for sexual diagnosis in an Indian population. In addition, application of the Purkait triangle in a population from the USA provided an accuracy of more than 90 % [31]. Albanese et al. [32] hoped to improve upon the Purkait method by using new measurements from the femur using samples from the Terry Collection ( $n=324$ ; M, 154; F, 170) and the Grant Collection ( $n=40$ ; M, 19; F, 21). Using logistic regression techniques, these authors achieved a high accuracy of between 95 and 97 %.

Sex prediction and the quantification of bone size based on the human skeleton remains a key objective for solving both bioarchaeological and forensic problems. The recent and rapid technological evolution of the high-precision scanners used in medicine has led to an ability to readily obtain images in high definition. The potential and applicability of computed tomography (CT) in forensic science has been highlighted in a wide range of recent studies [33, 34]. Sex prediction or an estimation of vertical measurements has been performed using CT images of the scapula [14], pubic bone [16], pelvis [12], femur [35, 36] or to obtain anthropological measurements of human lower limbs and foot bones in different states of decomposition using Multi-Detector Computed Tomography [11]. This non-invasive technique allows high-resolution 3D images of

the current population to be rapidly obtained. The possibility of transferring the digital flow from acquisition of the radiographic material to a computer allows fast access to the images, thereby reducing handling errors and promoting the renewal and updating of collections. In addition, data treatment using appropriate software allows errors associated with the orientation and location of the points that define the variables to be reduced. In contrast to more physical settings, the widespread use of CT techniques in current populations facilitates the establishment and renewal of reference data collections with more representative samples of larger size.

Technological advances have improved the reliability of these measurements and therefore improved their predictive ability. Recent interest in this field has focused on reducing the errors associated with positioning of the bones, obtaining the measurements and the expertise of the observer. These so-called personal and technical measurement errors affect both accuracy (difference between the observed value and the true value) and precision (variability in repeat measurements) [37]. Moreover, various statistical techniques have been shown to significantly improve the accuracy depending on whether the analysis is uni- or multivariate. In our case, as the choice is between two categories (male/female), the use of logistic regression greatly improves the results [1, 38, 39].

In studies with poorly preserved skeletal remains, interest tends to focus on identifying the set of variables defined on robust parts of the bone containing the greatest amount of information. For local populations, the understanding acquired in previous studies of the same population, those concerning other populations and the accessibility of information allow an initial list of potentially informative variables to be prepared. The most informative variables are obtained by applying combinatorial programming algorithms, especially stepwise methods.

The aim of this paper is to test the variables measured for the femur and hip bone in CT images, using OsiriX software, for sex prediction in a sample of the Spanish population. This is the first study of this type in a Spanish population.

## Materials and methods

The database studied was prepared from abdominal/pelvic CT images from 114 individuals (58 females and 56 males) aged between 20 and 85 years from hospitals in the Autonomous Community of Castilla-La Mancha, to whose populational biological background the individuals studied belong. All cases were selected prior to the study by the head of the Radiology Dept. at the Hospital Universitario de Guadalajara on the basis of the medical records and anonymised by biocomputing technicians at the institution for the research project mentioned in the acknowledgements. Patients with a history of trauma, hip prosthesis, no age-related diseases, bone

tumors and growth disorders and severe osteoporosis induced by a tumor or infection were excluded from the study. Although specific information concerning the ethnicity of the subjects was unavailable, the sample was taken as being representative of a typical contemporary Spanish population.

All DICOM files for the femoral and pelvic variables selected (Table 1: description, Fig. 1: graphical representation) were obtained using a 64-slice Helical CT scanner manufactured by GE Medical Systems (model: Light Speed VCT) equipped with an hvscmul64 station (rotation time 0.5 s, 120 kV, 140 mA, pitch 0.98, collimation  $40 \times 0.625$  mm, soft kernel, matrix  $512 \times 512$ ). The slice thickness was 0.625 mm.

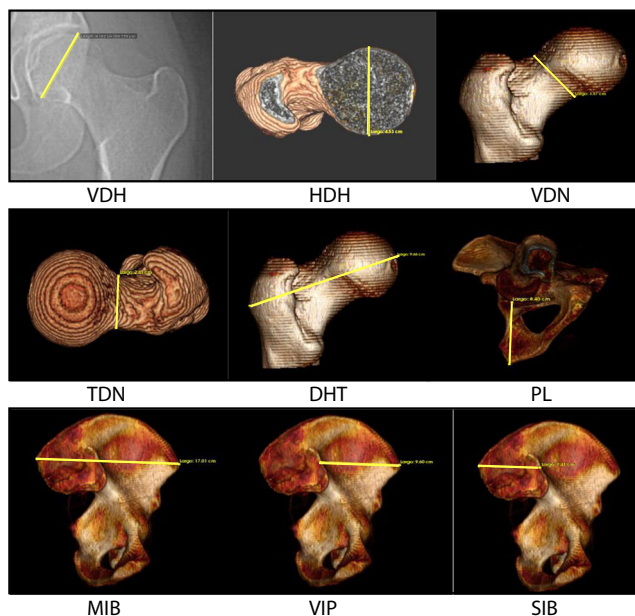
A standardised measuring procedure was developed to improve measurement precision and repeatability. After referencing the tools provided by OsiriX to eliminate soft tissue, the procedure specified the variables (definition), the appropriate orientation for the measurement (anterior, posterior, inferior, superior or lateral view), and the measurement of the variables (points that define each variable). Upon applying this protocol, the observer obtains a minimum of three consecutive measurements for each variable. For those variables defined as the maximum distance between end points (vertical diameter of the femoral head (VDH), horizontal diameter of the femoral head (HDH), maximum distance between the femoral head and the greater trochanter (DHT) and maximum iliac breadth (MIB)) or the minimum distance between end points (vertical diameter of the femoral neck (VDN) and pubic length (PL)), the measuring process finishes when an inflection in the sequence of values obtained is observed (due to a change in orientation of the image). In such cases, the reading is established as the maximum or minimum value of the measurements when the maximum difference between the last three observations is lower than the permitted value (2 mm). If the difference between the last three observations is higher than the permitted value, the measurement process is started again. The values for transverse diameter of the femoral neck (TDN) and ventral breadth of the ilium (VIB) are established as the mean value of the observations for the VDN (upper view) and MIB orientations. The value of sacral breadth of the ilium (SIB) is established as the difference between MIB and VIB.

The CT images were visualised and edited using the OsiriX software (Figs. 1 and 2). The accuracy and reliability of OsiriX for the reconstruction of surgical 3D images, especially for the identification, evaluation and planning of surgical and therapeutic procedures, has been demonstrated [40]. For the pelvis, OsiriX allows reconstructed CT images to be visualised with the ideal spatial orientation and magnification, and also allows the points labelled to be modified a posteriori [41]. The measurement precision and repeatability obtained using OsiriX is similar to that obtained with other commercial software packages and much higher than that obtained when performing measurements directly with bones [42]. The

**Table 1** Definition of the femoral and hip bone variables studied

Variable	Definition
Vertical diameter of the femoral head (VDH)	In the frontal view of the topogram, the 2D maximum vertical distance between the upper and lower ends of the femoral head
Horizontal diameter of the femoral head (HDH)	In the upper view, the maximum horizontal distance between the opposite ends of the femoral head. This variable is determined by displacing the slices comprising the 3D image until the image that gives the maximum diameter is found
Vertical diameter of the femoral neck (VDN)	In the right lateral/posterior view, the minimum distance between the opposite ends of the femoral neck (upper–lower)
Transverse diameter of the femoral neck (TDN)	In the upper view, and at the same height of the femoral neck at which the vertical diameter is measured, the distance between one side of the neck and the other (anterior–posterior)
Maximum distance between femoral neck and greater trochanter (DHT)	In the right lateral/posterior view, the maximum distance between the point immediately above the fovea of the femoral head and the outermost point of the greater trochanter
Maximum iliac breadth (MIB)	The maximum distance between the outermost point of the anterior superior iliac spine and the outermost point of the posterior superior iliac spine, as measured through the ventral face (right lateral/anterior view)
Ventral iliac breadth (VIB)	Following the line marked by the maximum iliac breadth, the distance from the outermost point of the anterior superior iliac spine to the point on that line which touches the iliac tuberosity and where the sacral joint starts. It is measured through the ventral face (right lateral/anterior view)
Sacral iliac breadth (SIB)	Following the line marked by the maximum iliac breadth, the distance from the outermost point of the posterior superior iliac spine to the point on that line which touches the sacral joint. It is measured through the ventral face (right lateral/anterior view)
Pubic length (PL)	In the lower view, the minimum distance between the most antero-superior point of the pubic spine and the anterior-most edge of the acetabulum

sources of error have been related to incorrect orientation of the 3D image and cleaning of the viscera and muscles (too much or too little). The main drawback of OsiriX is the inability to freely manipulate the three planes of the 3D image in a similar manner to multiplanar reconstruction (MCR).

**Fig. 1** Femoral and hip bone variables studied

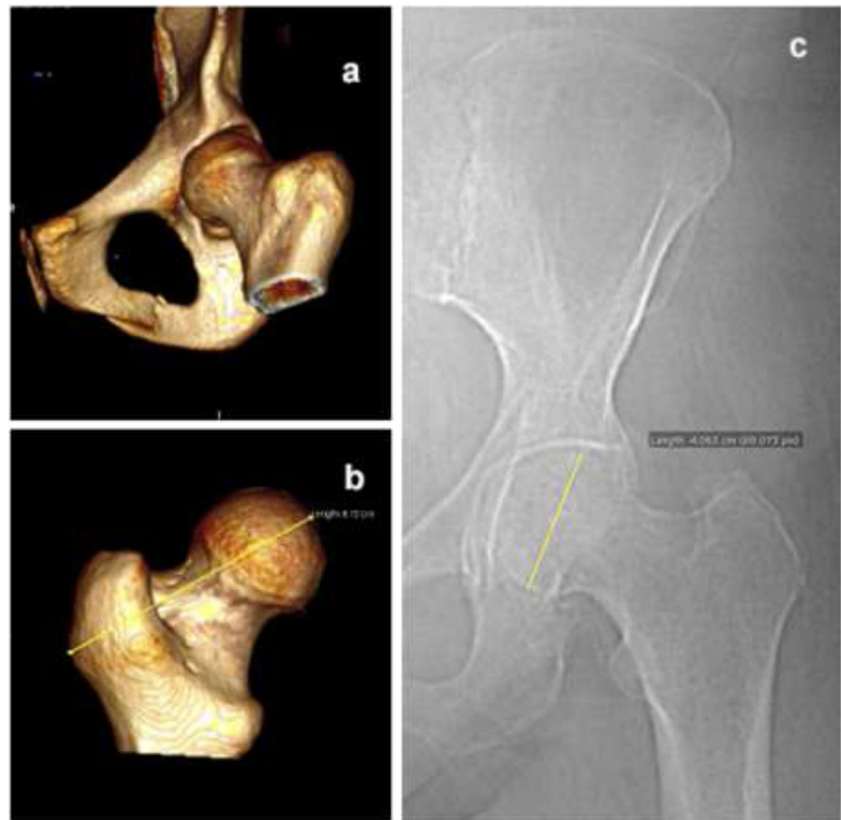
Measurements were obtained by one of us (AC) experienced observer measuring bone on CT. All images were studied in 3D on the basis of the series of slices for subsequent rendering and measurement of the variables in 2D. Only VDH was measured from the topogram (scout view), which is a digitalised radiograph of the region to be studied to which the different parameters are fitted, due to the difficulty and limited reliability of measuring this parameter from the CT image. All measurements were performed for the left femur and hip bone, except for those cases in which they were deformed or had been manipulated, in which case the right femur or hip bone was used.

Lin's concordance correlation coefficient [43] has previously been calculated to assess the reproducibility of the measurement technique. The value of this coefficient was calculated for males and females and for all variables considered:

$$\rho_c = 1 - \frac{1}{n} \cdot \frac{\sum_{i=1}^n (y_{1i} - y_{2i})^2}{s_{y_1}^2 + s_{y_2}^2 + (\bar{y}_1 - \bar{y}_2)^2} \quad (1)$$

where  $n$  is the sample size,  $y_{1i}$  is the first set of measurements,  $y_{2i}$  the second set of measurements,  $\bar{y}_1$ ,  $s_{y_1}^2$  and  $\bar{y}_2$ ,  $s_{y_2}^2$  are the mean and variance of the first and second sets of measurements.

**Fig. 2** Example of image quality used in our study (a) measuring example variable DHT with the Osirix software (b) or 2D image scout on in most cases the VDH measured variable (c)



Descriptive statistics of the variables observed, including means, standard deviations and Pearson's correlation coefficients (overlap of information between variables), was performed to describe the characteristics of the femoral and pelvic measurements in the study population. The means for each population were compared using the Student's *t* test, applying the Bonferroni correction to the *p* value, to identify those variables whose behaviour differed in the male and female populations. Wilks' lambda statistic ( $\Lambda$ ) was used to determine the discriminatory power of the variables. Wilks' lambda statistic varies between 0 and 1, with low values indicating a high discriminatory power and values close to 1 a low discriminatory power. To favour reading as a percentage, the discriminatory power was expressed as  $(1-\Lambda) \cdot 100$ .

For a binary response (female/male), logistic regression discriminant analysis appears to be the technique best suited to these data owing to its ability to allow sex determinations with both low misclassification and low sex bias rates. Logistic regression has several additional desirable characteristics in the present context: (1) it provides predicted values that can be interpreted as probabilities of group membership, (2) it does not assume a linear relationship between the dependent and independent variables, (3) it does not require multivariate

normality of independent variables, (4) it does not assume homoscedasticity, and (5) in general, it has less stringent requirements than linear discriminant analysis [38]. The predictive ability of the classification function was estimated using a leave-one-out method for cross-validation. With significant correlations between the explanatory variables and no significance for the coefficients in the model, stepwise discriminant function analysis (forward selection criterion) is used to simplify the model. Factor analysis with extraction of the principal components and Varimax rotation (with standardisation of the variables) was used to understand the importance of the variables observed and the simplification criterion of the predictive model. The Kaiser-Guttman criterion was used to extract the number of factors involved in the analysis [44, 45]. Predictive models for one and two variables were also obtained. A loss in predictive ability is inevitable when bone remnants are fragmented and there is no possibility of measuring any of the more explanatory variables.

Three observers with limited experience in CT measurements were selected to check the robustness of the measuring technique in terms of the predictive ability of the model. Lin's concordance correlation coefficient (with respect to the measurements obtained by the trained observer), the mean percentage bias in the

measurement (Bias  $\bar{X}$ ):

$$\text{Bias } \bar{X} = \frac{\bar{X}_i - \bar{X}_r}{\bar{X}_r} \cdot 100. \tag{2}$$

and the percentage variation in the standard deviation (Bias  $s$ ),

$$\text{Bias } s = \frac{s_i - s_r}{s_r} \cdot 100 \tag{3}$$

where  $\bar{X}_i$ ,  $s_i$  and  $\bar{X}_r$ ,  $s_r$  are the means and standard deviations for the values obtained by observer  $i$  and by the trained observer, were evaluated for each observer.

**Results**

The reproducibility of the measuring technique was determined for all variables with a sample of 40 individuals (20 males and 20 females). The values for Lin’s concordance correlation coefficient can be found in Table 2.

The results obtained for the 114 abdominal/pelvic CT scans corresponding to 58 females and 56 males from the Spanish population studied show that the variables with the highest discriminatory ability are as follows: VDH, HDH, VDN and DHT. The variables TDN, MIB and SIB are also discriminatory, but less so than the other variables. No significant differences were observed between the means for each group (males and females) as regards the variables PL and VIB (Table 3).

The logistic regression model provided the predictive equation:

$$PF = \frac{e^\eta}{1 + e^\eta}$$

**Table 2** Lin’s concordance correlation coefficient for each of the parameters measured

Variable	$\rho_c$ value
VDH	0.96
HDH	0.97
VDN	0.93
TDN	0.89
DHT	0.98
MIB	0.99
VIB	0.93
SIB	0.94
PL	0.97

where

$$\eta = -48.1171 + 0.8497 \cdot \text{VDH} + 1.17694 \cdot \text{HDH} + 0.554869 \cdot \text{VDN} - 0.3858 \cdot \text{TDN} + 0.1420 \cdot \text{DHT} + 24.0634 \cdot \text{MIB} - 24.3493 \cdot \text{VIB} - 24.1715 \cdot \text{SIB} - 0.3886 \cdot \text{PL}. \tag{4}$$

The decision criterion for this predictive equation was based on a value of 0.5 (which allows the greatest balance in the probability of an erroneous classification between the two groups). Thus, the abdominal/pelvic remnants were associated with a female for a PF value of less than 0.5 and a male for a PF value of more than 0.5. Application of this predictive model to the abdominal/pelvic CT database corresponding to 58 females and 56 males provided a correct classification in 99.1 % of cases (100 % for males and 98.3 % for females).

As the coefficients for some variables were not significant, the predictive model was simplified. Thus, applying stepwise discriminant function analysis (forward selection criterion), the  $\eta$  equation was simplified using only the variables VDH, HDH and MIB,

$$\eta = -92.901 + 2.6389 \cdot \text{VDH} + 1.2925 \cdot \text{HDH} - 0.4763 \cdot \text{MIB}. \tag{5}$$

With the simplified model (5) and the decision criterion defined for the full model, a correct classification of 98.3 % was obtained (98.2 % for males and 98.3 % for females).

This simplification of the model can be explained on the basis of the high correlation (overlap of information) between the original variables. For the group of variables with greatest discriminatory ability (VDH, HDH, VDN and DHT), the correlations between the variables are very high: the minimum correlation observed is 0.77, the mean of the correlations is 0.84 and the maximum correlation observed is 0.92. The transverse diameter of the femoral neck (TDN) is also highly correlated with the most informative group: 0.60, 0.65 and 0.73 are the minimum, mean and maximum correlations observed. The maximum iliac breadth (MIB) is highly correlated with the sacral breadth of the ilium (SBI) and the ventral iliac breadth, and to a lesser extent with the group of more informative variables (Table 4).

The need to use the information provided by the femur and hip bone together can be explained by the two-dimensional factorial representation. Projection of the dispersion graph for factors F1 and F2 shows that the diagonal line best separates the regions delimiting the two sexes (Fig. 3). With an explained variability of 83.6 % (69.4 % F1 and 13.2 % F2) and high communalities (the lowest communality (0.64) is obtained with the variable TDN and the highest (0.92) with the variable HDH), F1 would be related to the characteristics of

**Table 3** Descriptive statistics and comparison of means for the groups of males and females

	Females		Males		<i>t</i> test		Discrimination	
	$\bar{X}_F$	$S_F$	$\bar{X}_M$	$S_M$	$t_{exp}$	<i>p</i> value	$\Lambda$	$(1-\Lambda) \cdot 100$
VDH	40.21	1.99	46.94	2.69	15.22	0.0000 <sup>a</sup>	0.33	67.41
HDH	40.49	1.99	47.16	2.86	14.48	0.0000 <sup>a</sup>	0.35	65.19
VDN	32.57	1.92	37.99	2.34	13.55	0.0000 <sup>a</sup>	0.38	62.10
TDN	24.76	2.29	28.42	2.92	7.45	0.0000 <sup>a</sup>	0.67	33.13
DHT	86.36	5.71	99.61	4.97	13.21	0.0000 <sup>a</sup>	0.39	60.89
MIB	160.03	8.35	166.63	8.83	4.10	0.0007 <sup>a</sup>	0.87	13.05
VIB	100.45	7.01	101.98	6.33	1.22	1.0000	0.99	1.30
SIB	59.58	5.55	64.65	6.78	4.38	0.0002 <sup>a</sup>	0.85	14.60
PL	70.29	3.86	69.10	3.79	1.66	0.8885	0.98	2.42

Mean and standard deviation (measured in mm),  $t_{exp}$  statistic and *p* value corresponding to the Student’s *t* test with the Bonferroni correction, and Wilks’ lambda statistic

<sup>a</sup> Denotes statistical significance

the femur and factor 2 to the characteristics of the hip bone (Table 5).

To make prediction easier when the bone remnants are fragmented, the equations were calculated for one and two variables, which, although less accurate, may nevertheless still be useful (Table 6). The highest accuracy was obtained when using the three most explanatory variables in pairs (VDH, HDH; VDH, MIB; and HDH, MIB). It should be noted that in the event of erosion of the femoral head in archaeological samples, VDH and HDH can reasonably be substituted by VDN. In this context, no differences in predictive ability have been observed when considering the model obtained with the variable VDH and that obtained with the variables VDH and MIB. The greater weight of the femur-related variables (VDH, HDH VDN, DHT, TDN) is again found for the single-variable equations. Finally, the iliac breadth (MIB) isolated from the femur-related variables is not as effective for sex prediction although, in our series, it is preferable to PL.

The simplified predictive model (2) with VDH, HDH and MIB was tested by three observers not trained in the measurement protocol. The correct classification percentages obtained

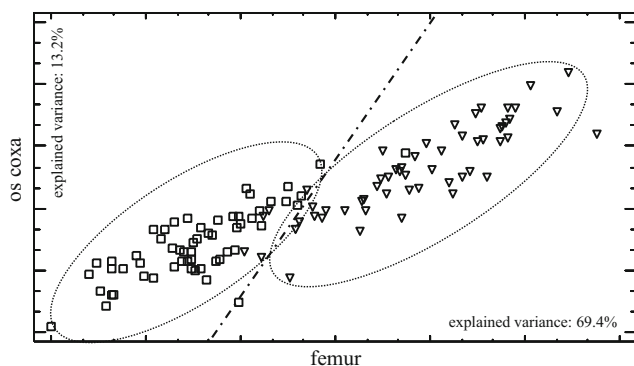
by these observers with the images used in this study (114 individuals, 58 females and 56 males) were as follows: 92.1 % for observer 1, 93.0 % for observer 2 and 76.3 % for observer 3. The loss of predictive capacity was related to a lack of concordance in the measurements (Lin’s coefficient), which in turn is related to measurement bias and its variability (Table 7). The lower levels of concordance due to the higher measurement bias and variability markedly reduce the classification accuracy for observer 3.

**Discussion**

With all the bone-related information available, it is widely accepted that the cranium and pelvis are the best indicators of sexual dimorphism as regards both discontinuous and metric parameters. The effects of preservation, damage and recovery factors on the composition of archaeological assemblages often results in damaged, fragmented or incomplete skeletal remains. The results obtained herein show that sex prediction can also be very accurate when considering variables related

**Table 4** Pearson’s correlation coefficients between for the original variables (correlations of 0.6 or higher are indicated in italics)

	VDH	HDH	VDN	TDN	DHT	MIB	VIB	SIB	PL
VDH									
HDH	<i>0.92</i>								
VDN	<i>0.82</i>	<i>0.82</i>							
TDN	<i>0.64</i>	<i>0.73</i>	<i>0.65</i>						
DHT	<i>0.86</i>	<i>0.86</i>	<i>0.77</i>	<i>0.60</i>					
MIB	0.54	0.56	0.51	0.44	<i>0.60</i>				
VIB	0.25	0.26	0.30	0.22	0.33	<i>0.69</i>			
SIB	0.50	0.51	0.40	0.39	0.49	<i>0.69</i>	−0.06		
PL	−0.02	0.05	−0.05	0.19	0.04	0.28	0.23	0.15	



**Fig. 3** 2D factorial representation of the sample of 58 women (*square*) and 56 men (*triangle*). The oblique *diagonal line* separates the two sexes

to the femur and pelvis which, due to their robustness, are usually preserved over time in cemeteries or common graves. These findings are in agreement with previous studies which have shown that the length of certain long bones may be just as discriminatory as the cranium and pelvis [9, 46, 47] and especially with literature references in which highly accurate sex determination has been achieved using various femoral dimensions [27, 29, 32, 48]. As far as preservation is concerned, the greater robustness of the epiphysis with respect to the shaft of long bones and the common fracture of the most informative pelvic dimensions (pubic and ischial dimensions) has been found in archaeological records [49–51].

Logistic regression model provided a predictive equation that correctly classifies 99.1 % of the individuals studied using nine variables. This value highlights the effectiveness of the variables considered in the classification between males and females. Simplification of the model using stepwise methods resulted in the correct classification of 98.3 % of subjects by combining the vertical and horizontal diameters of the femoral head (VDH, HDH) and the maximum iliac breadth (MIB). Similarly to the reports of other authors [9, 27, 29], the femoral variables VDH and HDH have the highest discriminatory ability, whereas MIB has the highest discriminatory ability for those variables defined for the robust part of the pelvis. All variables for the proximal femoral epiphysis are closely related. The high correlation of VDN, TDN and DHT with VDH and HDH reflects the high degree of overlap in the information they provide and explains the minimal loss of

discriminatory ability upon simplification of the model. The complementary nature of the information provided by the variables defined for the robust parts of the femur and pelvis was seen from the 2D representation. In quantitative terms, a gain in predictive ability was observed when the model was broadened or reduced (98.3 % with VDH, HDH and MIB; 95.6 % with VDH and MIB; 93.0 % with VDH; and 65.8 % with MIB). Anthropometric studies demonstrated that there are differences between the two lower arms and that this difference is reflected in lower neck-shaft angles [52, 53]. Although this differentiation affects measurement of the DHT variable, we believe that it does not affect the other measurements considered in this study as they are not affected by the neck-shaft angle. Consequently, the predictive ability of the model is ensured when measurement of the variables is performed on the left side and a minimum loss of precision, if any, is to be expected when the reduced model (VDH, HDH and MIB) is considered and the variables are measured for the right side.

The complementary information provided by the femur and pelvis can be explained by the adaptation in two directions. Differentiation in the femur and pelvis has been related to functional adaptation, the transmission of body weight via the proximal femoral epiphysis and the mechanics of bipedalism, whereas an additional significant differentiation of the pelvis has been explained in obstetric terms [23–25, 54, 55].

In terms of classification accuracy, the model obtained using the variables VDH, HDH and MIB is essentially identical to that obtained by Trancho et al. in a study of both epiphyses from 132 femurs, in an osteological collection deposited at the Universidad Complutense de Madrid [27]. In contrast, the predictive ability of the model presented herein is somewhat higher than that reported by Albanese for the Terry and Coimbra collections [39] and Albanese et al. for the Terry and Grant collections [32], respectively (when considering the leave-one-out method for cross-validation with a single sample). A univariate comparison of the most informative variables also highlighted the concordance with the results obtained in other populations. The predictive ability obtained for the variable VDH is 93.0 % in this study, 91.23 % in another Spanish population [27] and 86.8 % in a German population [29]. The variable HDH provided an accuracy of 91.2 % in this study, 93.86 % in another Spanish population [27] and 89.60 % in a German population [29]. Finally, the variable MIB gave a correct sex classification of 65.8 % in this study, 64 % in a population of black Americans and 56 % in a population of white Americans [9].

The variables defined for robust parts of the femur and pelvis (VDH, HDH and MIB) have provided a sex-prediction model. The high predictive ability of this model and possibility of using it even with poorly preserved skeletal remains provide value to it and differentiate it from others that make use of less robust variables. In addition, an ability to estimate

**Table 5** Components of factors F1 and F2 (the coefficients of greatest magnitude are in italics)

Variable	F1	F2
VDH	<i>0.89</i>	0.31
HDH	<i>0.90</i>	0.32
VDN	<i>0.88</i>	0.23
TDN	<i>0.77</i>	0.21
DHT	<i>0.83</i>	0.37
MIB	0.35	<i>0.84</i>
SIB	0.23	<i>0.90</i>



**Table 6** Sex-prediction equations for the Spanish population using variables form from the hip bone and proximal femoral epiphysis

Model	Coefficients for logistic regression equations										Percent correct classification			
	Constant	VDH	HDH	VDN	TDN	DHT	MIB	VIB	SIB	PL	Total	Male	Female	Bias
Complete	-48.1171	0.8497	1.1769	0.5549	-0.3858	0.1420	24.0634	-24.3493	-24.1715	-0.3886	99.1	100.0	98.3	1.7
VDH, HDH, MIB	-92.9010	2.6389	1.2925				-0.4763				98.3	98.2	98.3	0.1
VDH, HDH	-68.1159	0.9606	0.6153								93.0	94.6	91.4	3.3
VDH, MIB	-64.7188	2.3959					-0.2370				95.6	94.6	96.6	1.9
HDH, MIB	-53.6810		1.9098				-0.1806				93.9	94.6	93.1	1.5
VDN, MIB	-37.5048			1.2159			-0.0315				90.4	87.5	93.1	5.6
VDH	-60.3275	1.3989									93.0	92.9	93.1	0.2
HDH	-55.2896		1.2724								91.2	91.1	91.4	0.3
VDN	-41.0160			1.1698							90.4	87.5	93.1	5.6
TDN	-13.8791				0.5219						73.7	67.9	79.3	11.5
DHT	-40.1642					0.4309					89.5	89.3	89.7	0.4
MIB	-14.7909						0.0904				65.8	62.5	69.0	6.5
VIB	-3.5632							0.0349			57.0	51.8	62.1	10.3
SIB	-8.6890								0.1394		65.8	55.4	75.9	20.5
PL	5.7693									-0.0833	59.7	50.0	69.0	19.0

vertical measurements (or age) is also a subject of interest in forensic science. The transmission of errors from sex to vertical measurements due to use of the wrong equation (that corresponding to the opposite sex) occurs to a lesser extent when metrics in which the variables considered to do not present divergence patterns are considered. In the model obtained herein, the variables considered (VDH, HDH and MIB) present the same pattern. The mean for all three variables is higher for males than for females.

There is evidence that sex-related differences in a population depend on the geographic region to which that population belongs and therefore that specific formulae are probably required for each such population [17, 27, 28, 56]. Our

findings only appear to achieve maximum accuracy in the Spanish population. However, it would be of interest to check whether this accuracy is maintained in other geographically related populational samples of similar sizes. In any case, the limited number of literature references to the Spanish population and the excellent predictive ability with variables defined for robust parts of the bone justify this study and the discriminatory functions obtained.

The reproducibility of the measurement technique ensures the reliability of the information used to obtain these discriminatory functions. In contrast, the inter-observer error as regards measuring the variables and its consequences in terms of loss of predictive ability (first observer,

**Table 7** Measurement accuracy of three observers not trained in the measurement protocol. Bias  $\bar{X}$  (2), Bias *s* (3) and Lin's concordance correlation coefficient (1) (the main differences in italics)

	VHD		HDH		MIB	
	F	M	F	M	F	M
Observer 1						
Bias $\bar{X}$	2.87 %	3.94 %	-3.19 %	-1.81 %	-1.90 %	-0.73 %
Bias <i>s</i>	0.48 %	8.22 %	10.48 %	1.28 %	19.01 %	-5.84 %
Lin's coef.	0.70	0.74	0.61	0.79	0.86	0.85
Observer 2						
Bias $\bar{X}$	2.54 %	3.10 %	2.52 %	2.24 %	0.63 %	0.84 %
Bias <i>s</i>	3.98 %	9.90 %	-1.28 %	-0.33 %	-2.16 %	1.77 %
Lin's coef.	0.76	0.80	0.83	0.88	0.95	0.97
Observer 3						
Bias $\bar{X}$	2.58 %	3.01 %	2.76 %	<i>10.89 %</i>	0.61 %	-1.24 %
Bias <i>s</i>	<i>38.95 %</i>	7.39 %	<i>227.40 %</i>	<i>160.98 %</i>	<i>36.51 %</i>	4.21 %
Lin's coef.	<i>0.42</i>	0.81	<i>0.17</i>	<i>0.20</i>	0.71	0.89

92.1 %; second observer, 93.0 %; third observer, 76.3 %) have highlighted the need to protocolise the measurement system. The lower predictive ability of the third observer in this study has been attributed to measurement bias and higher measurement variability, which in turn have been related to the way in which soft parts are removed, the correct orientation for assessing each variable and the exact location of the points that defined said variables. In accordance with the principles established in standard ISO/TS 21748: 2010 [57], the accuracy of the predictive equation can be related to the informative ability of the variables used, the suitability and estimation of the predictive model and measurement bias and variability. The informative ability of the variables and the suitability and estimation of the model have been shown to be sufficient to achieve a high predictive ability for the target population. In contrast, the loss of predictive ability is very significant when the inter-observer bias or measurement variability are high. This is why the effective application of such models requires the availability of accurate measurement protocols and reference standards with which comparisons can be made.

## Conclusions

Logistic regression discriminant analysis has provided a highly accurate sex-prediction model based on variables defined for robust parts of the femur and pelvis (VDH, HDH and MIB) (correct classification, 98.3 %). The complementary nature of the information provided by femoral and pelvic variables has been highlighted in the 2D factorial representation of the sample of 58 women and 56 men and the higher predictive ability of those models that use variables for both bones.

A wide range of recent studies have shown the potential and applicability of CT in forensic science. This non-invasive technique allows high-resolution 3D images of the current population to be rapidly obtained and, consequently, favours the renewal and updating of collections. The loss of predictive ability of the models has been related to the measurement accuracy, which in turn is related to observer training. As a result, effective application requires the drafting of measurement protocols, consensus standards and the correct training of observers.

**Acknowledgments** This work was funded by the projects CGL2006-02170/BTE from the Ministry of Science and Innovation (DT) and 2009SGR-884 from the General Research Directorate of the Generalitat de Catalunya. We would like to thank JM Sevilla, Head of the Radiology Department at the Hospital Universitario de Guadalajara (Castilla-La Mancha health service-SESCAM) for allowing us to obtain the scientific data used in this study. We also thank C. Rissech and G. Tranco for advise on some aspects of this paper.

## References

- Walker PL (2008) Sexing skulls using discriminant function of visually assessed traits. *Am J Phys Anthropol* 136:39–50
- Rissech C, Marquez-Grant N, 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:S163–S168
- Cameriere R, Ferrante L (2008) Age estimation in children by measurement of carpals and epiphyses of radius and ulna and open apices in teeth: a pilot study. *Forensic Sci Int* 174:59–62
- Boccome S, Cremasco MM, Bortoluzzi S et al (2010) Age estimation in subadult Egyptian remains. *Homo* 61:337–358
- Charisi D, Eliopoulos C, Vanna V et al (2011) Sexual dimorphism of the arm bones in a Modern Greek population. *J Forensic Sci* 56:10–18
- Ruff C (2010) Body size and body shape in early hominins. Implications of the Gona pelvis. *J Hum Evol* 58:166–178
- White TD, Black MT, Folken PA (2012) *Human osteology*, 3rd Ed. Academic
- Moore MK (2013) Sex estimation and assessment. In: DiGangi EA, Moore MK (Eds) *Research methods in human skeletal biology*. Academic, pp 91–116
- Spradley MK, Jantz RL (2011) Sex estimation in forensic anthropology: skull versus postcranial elements. *J Forensic Sci* 56:289–296
- Kranioti EF, Vorniotakis N, Galiatsou C, Iscan MY, Michalodimitraki M (2009) Sex identification and software development using digital femoral head radiographs. *Forensic Anthropology Population Data*. *Forensic Sci Int* 189:113.e1–113.e7
- Robinson C, Eisma R, Morgan B, Jeffery A, Graham EA, Black S, Ruttly GN (2008) Anthropological measurement of lower limb and foot bones using multi-detector computed tomography. *J Forensic Sci* 53(6):1289–1294
- Franklin D, Cardini A, Flaver A, Marks MK (2014) Morphometric analysis of pelvic sexual dimorphism in a contemporary Western Australian population. *Int J Legal Med*. doi:10.1007/s00414-014-0999-8
- Giurazza F, Vescovo R, Schena E, Battisti S, Cazatto RL, Grasso FR, Silvestri S, Denaro V, Zobel B (2012) Determination of stature from skeletal and skull measurements by CT scan evaluation. *Forensic Sci Int* 222:398e1–398.e9. doi:10.1016/j.forsciint.2012.06.008
- Giurazza F, Vescovo R, Schena E, Cazatto RL, D’Agostino F, Grasso FR, Silvestri S, Zobel BB (2013) Stature estimation from scapular measurements by CT scan evaluation in an Italian population. *Legal Med* 15:202–208
- Macaluso PJ, Lucena J (2014) Estimation of sex from sternal dimensions derives from chest plate radiographs in contemporary Spaniards. *Int J Legal Med* 128:389–395
- López-Alcaraz M, Garamendi PM, Alemán I, Botella LM (2013) Image analysis of pubic bone for sex determination in a computed tomography sample. *Int J Legal Med* 127:1145–1155
- Asala SA (2001) Sex determination from the head of the femur of South African whites and blacks. *Forensic Sci Int* 117:15–22
- Iscan MY, Miller-Shaivitz P (1986) Sexual dimorphism in the femur and tibia. In: Reichs K (ed) *Forensic osteology: advances in identification of human remains*. Thomas, Springfield, pp 102–111
- Asala SA, Bidmos MA, Dayal MR (2004) Discriminant function sexing of fragmentary femur of South African blacks. *Forensic Sci Int* 145:25–29
- Hauser R, Smoliński J, Gos T (2005) The estimation of stature on the basis of measurements of the femur. *Forensic Sci Int* 147(2–3):85–190
- Patriquin ML, Stein M, Loth SR (2005) Metric analysis of sex differences in South African black and white pelvis. *Forensic Sci Int* 147:119–127

22. Steyn M, Iscan MY (1997) Sex determination from the femur and tibia in South African whites. *Forensic Sci Int* 90:111–119
23. Tague RG (2000) Do big females have big pelvis? *Am J Phys Anthropol* 112:377–393
24. Tague RG (2005) Big-bodied males help us recognize that females have big pelvis. *Am J Phys Anthropol* 127:392–405
25. Kurki HK (2011) Pelvic dimorphism in relation to body size and body size dimorphism in humans. *J Hum Evol* 61:631–643
26. Manisha R, Dayala MR, Steynb M, Kuykendall KL (2008) Stature estimation from bones of South African whites. *S Afr J Sci* 104:124–128
27. Trancho GJ, Robledo B, López-Bueis I, Sánchez JA (1997) Sexual determination of the femur using discriminant functions. Analysis of a Spanish population of known sex and age. *J Forensic Sci* 42:181–185
28. Slaus M (1997) Discriminant function sexing of fragmentary and complete femora from medieval sites in continental Croatia. *Opuscula Archaeol* 21:167–175
29. Mall G, Graw M, Gehring KD, Hubig M (2000) Determination of sex from femora. *Forensic Sci Int* 113:315–321
30. Srivastava R, Saini V, Rai KR, Pandey S, Tripathi SL (2012) A study of sexual dimorphism in the femur along North Indians. *J Forensic Sci* 57:19–23
31. Purkait R (2005) Triangle identified at the proximal end of femur: a new sex determinant. *Forensic Sci Int* 147:135–139
32. Albanese J, Eklics G, Tuck A (2008) A metric method for sex determination using the proximal femur and fragmentary hipbone. *J Forensic Sci* 53:1283–1288
33. Christensen AM, Hatch GM, Brogdon BG (2014) A current perspective on forensic radiology. *J Forensic Radiol Imaging*. doi:10.1016/j.jofri.2014.05.001
34. Hatch GM, Dedouit F, Christensen AM, Thali MJ, Ruder TD (2014) RADid: a pictorial review of radiologic identification using postmortem CT. *J Forensic Radiol Imaging* 2:52–59. doi:10.1016/j.jofri.2014.02.039
35. Lorkiewicz-Muszniska D, Przystanska A, Kociemba W, Sroka A, Rewekant A, Zaba C, Paprzycki W (2013) Body mass estimation in modern population using anthropometric measurements from computed tomography. *Forensic Sci Int* 231:405.e1–405.e6
36. Grabherr S, Cooper C, Ulrich-Bochsler S, Uldin T, Ross S, Oesterhelweg L, Bolliger S, Christe A, Schnyder P, Mangin P, Thali MJ (2009) Estimation of sex and age of “virtual skeletons”—a feasibility study. *Eur Radiol* 19:419–429
37. Christensen AM, Crowder CM, Ousley SD, Houck MM (2014) Error and its meaning in forensic science. *J Forensic Sci* 59(1):123–126. doi:10.1111/1556-4029.12275
38. Hosmer D, Lemeshow S (2004) *Applied logistic regression*, 2nd edn. Wiley, New York
39. Albanese J (2003) A metric method for sex determination using the hipbone and the femur. *J Forensic Sci* 48:263–273
40. Yamauchi T, Yamazaki M, Okawa A, Furuya T, Hayashi K, Sakuma T, Takahashi H, Yanagawa N, Koda M (2010) Efficacy and reliability of highly functional open source DICOM software (OsiriX) in spine surgery. *J Clin Neurosci* 17:756–759
41. Vandebussche E, Saffarini M, Hansen U, Taillieu F, Mutschler C (2008) The asymmetric profile of the acetabulum. *Clin Orthop Relat Res* 466:417–423
42. Kim G, Jung HJ, Lee HJ, Lee JS, Koo S, Chang SH (2012) Accuracy and reliability of length measurements on three-dimensional computed tomography using open-source Osirix software. *J Digit Imaging* 25:486–491
43. Lin LI (1989) A concordance coefficient to evaluate reproducibility. *Biometrics* 45:255–268
44. Guttman L (1954) A new approach to factor analysis: the Radex. In: Lazarsfeld PF (ed) *Mathematical thinking in the social sciences*. Free Press, New York, pp 258–348
45. Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ Psychol Meas* 20:141–151
46. Kranjoti EF, Bastir M, Sanches-Messeguer A, Rosas A (2009) A geometric morphometric study of the Cretan humerus for the sex identification. *Forensic Sci Int* 189:113–121
47. Sakaue K (2004) Sexual determination of long bones in recent Japanese. *Anthropol Sci* 112:75–81
48. Slaus M, Strinovic D, Skavic J, Petrovec V (2003) Discriminant function sexing of fragmentary and complete femora: standards for contemporary Croatia. *J Forensic Sci* 48:509–512
49. Dittrick J, Suchey JM (1986) Sex determination of prehistoric central California skeletal remains using discriminant analysis of the femur and humerus. *Am J Phys Anthropol* 70:3–9
50. Chaudhary AK, Jain SK (2014) Osteoscopic assessment of sexual dimorphism in hip bone. *Acta Med Int* 1(1):28
51. Ali RS, MacLaughlin SM (1991) Sex identification from the auricular surface of the adult human ilium. *Int J Osteoarchaeol* 1(1):57–61
52. Anderson JY, Trinkaus E (1998) Patterns of sexual, bilateral and interpopulational variation in human femoral neck-shaft angles. *J Anat* 192:279–285
53. Pujol A, Rissech C, Ventura J, Badosa J, Turbón D (2014) Ontogeny of the female femur. Geometric morphometric analysis applied on current living individuals of a Spanish population. *J Anat*. doi:10.1111/joa.12209
54. Aiello L, Dean C (1990) *Human evolutionary anatomy*. Academic
55. Trevathan WR (1987) *Human birth. An evolutionary perspective*. de Gruyter, New York
56. Iscan MY, Loth SR, King CA, Shihai D, Yoshino M (1998) Sexual dimorphism in the humerus: a comparative analysis of Chinese, Japanese and Thais. *Forensic Sci Int* 98:17–29
57. ISO/TS 21748: 2010. Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation