

# Disregarding population specificity: its influence on the sex assessment methods from the tibia

Anežka Kotěrová<sup>1</sup> · Jana Velemínská<sup>1</sup> · Ján Dupej<sup>2</sup> · Hana Brzobohatá<sup>3</sup> · Aleš Pilný<sup>4</sup> · Jaroslav Brůžek<sup>1</sup>

Received: 27 January 2016 / Accepted: 30 June 2016 / Published online: 20 July 2016  
© Springer-Verlag Berlin Heidelberg 2016

**Abstract** Forensic anthropology has developed classification techniques for sex estimation of unknown skeletal remains, for example population-specific discriminant function analyses. These methods were designed for populations that lived mostly in the late nineteenth and twentieth centuries. Their level of reliability or misclassification is important for practical use in today's forensic practice; it is, however, unknown. We addressed the question of what the likelihood of errors would be if population specificity of discriminant functions of the tibia were disregarded. Moreover, five classification functions in a Czech sample were proposed (accuracies 82.1–87.5 %, sex bias ranged from –1.3 to –5.4 %). We measured ten variables traditionally used for sex assessment of the tibia on a sample of 30 male and 26 female models from recent Czech population. To estimate the classification accuracy and error (misclassification) rates ignoring population specificity, we selected published classification functions of tibia for the Portuguese, south European, and the North American populations. These functions were applied on the dimensions of the Czech population. Comparing the classification success of the

reference and the tested Czech sample showed that females from Czech population were significantly overestimated and mostly misclassified as males. Overall accuracy of sex assessment significantly decreased (53.6–69.7 %), sex bias –29.4–100 %, which is most probably caused by secular trend and the generally high variability of body size. Results indicate that the discriminant functions, developed for skeletal series representing geographically and chronologically diverse populations, are not applicable in current forensic investigations. Finally, implications and recommendations for future research are discussed.

**Keywords** Forensic anthropology population data · Sex determination · Tibia · Population specificity · Discriminant function analysis · GAME method

## Introduction

Reliable sex determination is of paramount importance in forensic practice [1–3]. It provides a component of an individual's biological profile complementing age at death estimation, stature, and population affinity assessment [1, 4]. It is widely accepted that the most accurate sex estimation from skeleton can be performed on pelvis and then skull [5, 6]. Unfortunately, they are not always preserved in forensic cases [7] and in such situations, one can use long bones. Especially robust ones like femur or tibia which are often better preserved can be used for sex estimation [8, 9]. Numerous studies were published that prove usefulness of tibia in forensic anthropology [8, 10–21]. Almost all of them agree that its proximal end is the most sexual dimorphic. Generally, the accuracies of sex determination are high (above 85 % when multiple measurements are

✉ Anežka Kotěrová  
koterova.a@seznam.cz; koterova@natur.cuni.cz

<sup>1</sup> Department of Anthropology and Human Genetics, Faculty of Science, Charles University in Prague, Viničná 7, Praha 2, 128 43 Prague, Czech Republic

<sup>2</sup> Department of Software and Computer Science, Faculty of Mathematics and Physics, Charles University in Prague, Malostranské nám. 25, 118 00 Prague, Czech Republic

<sup>3</sup> Department of Prehistory, Institute of Archaeology of the Academy of Science, Letenská 4, 118 01 Praha, Czech Republic

<sup>4</sup> Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague, Karlovo náměstí 13, 121 35 Prague 2, Czech Republic

used) especially with proximal and distal widths as well as circumference [22].

In sex assessment, discriminant function analysis (DFA) is the most often used classification tool among authors. Their advantages and shortcomings have been discussed recently [23, 24]. DFA tend to be samples-specific and their generalization needs to be assessed in different samples [25].

Any attempts to develop standards for sex estimation of human skeletal remains must take into account that the pattern of sexual dimorphism varies among human populations. Therefore, it is important to avoid the application of metric standards proposed in different populations from different periods of time than the studied sample [26]. Also, as several studies have shown, body size has changed over generations in the population as a consequence of secular trend [27–29]. These changes are especially related to the body height [30]. Secular trend in stature is widely observed in humans since the nineteenth century, and this trend directly impacts the dimensions of long bones [31–34].

Major problems with methods that use size-based variables are that standards can be influenced by secular trend and are usually population-specific. On the other hand, this is to a certain extent also true for shape-based characteristics [22]. Thus, it should be assumed that methods for sex determination, based on skeleton collections of known sex from the first half of the twentieth century, cannot guarantee the same reliability of results when they are used in attempts to identify unknown human remains from recent populations [35].

Application of metric data from one population into the DFA derived from different population groups results in high classification error, and the results are also affected by large sex bias [36, 37]. In a study designed to quantify the effect of applying Euro-American [38] and South African of European ancestry [39] standards to Australian population sample [40], classification accuracy was approximately the same in Australian sample as in target sample (e.g., 80–83 %). On the other hand, correct estimation of sex was distinctly distorted by unacceptable sex bias, i.e., the difference in correct sex determination of males compared to females (which was 31 and 36 %, respectively).

The effect of disregarding population specificity on the classification accuracy of DFA is often mentioned [25] but is rarely described for different parts of the skeleton. In recent publications, we found such studies only for crania from Western Australian and Indian populations [25, 30] and in European populations for the clavicle [41], the calcaneus

[37], and for the femur [36]. Recently, many authors have argued for the development and use of population-specific formulae for diverse parts of skeleton when metric data are used [25, 30, 35, 41–45].

The first objective of the present study is to propose classification functions (CFs) for sex estimation in recent Czech population based on CT imaging. The collection of osteometric data from CT images is a reliable and acceptable source and is utilized more and more often for sex assessment in forensic anthropology [46–51]. The other objective is to simulate ignoring population specificity and practically show the range of errors in sex classification.

## Material and methods

### Material

In this study, we used surface models which were constructed from anonymized CT scans from angiography (the slice increment was set at 0.5 mm). Tibial 3D models were reconstructed using specialized software Mimics (Materialise, Leuven, Belgium) and thresholding method. Our virtual material comes from Czech population of the twenty-first century. The age, sex, and race of all specimens are known. Sample consists of nonpathological 56 left human tibiae where 30 belong to male and 26 to female. Male individuals were born between the years 1943 and 1980, and the mean age was 56.1 years (ranging from 31 to 68). Female individuals were born between the years 1920 and 1978 when their mean age was 69.0 years (ranging from 33 to 91). Table 1 shows age distributions, mean ages, and standard deviations. The same material was originally utilized in the study of Brzobohatá et al. [52], where authors performed sex determination from tibiae using geometric morphometrics.

### Selection of measured dimensions

Following relevant published works, we chose ten dimensions commonly measured in anthropology. Their list together with used abbreviations, references where they were defined and landmarks used for computing these dimensions are given in Table 2. Most of the chosen dimensions were defined by Martin and Saller [53] marked by letter M; in other cases, the authors of the definitions is given in the same table.

**Table 1** Characteristics of the Czech sample by age and sex

| Sex    | Total | Mean age (year) | SD   | 30–40 years | 41–60 years | 61 years and over |
|--------|-------|-----------------|------|-------------|-------------|-------------------|
| Male   | 30    | 56.1            | 10.1 | 3           | 13          | 14                |
| Female | 26    | 69.0            | 11.6 | 1           | 3           | 22                |

SD standard deviation

**Table 2** Selected and measured dimensions with their name, references, and used landmarks

|    | Abbreviation | Main landmarks | Auxiliary landmarks     | Name                             | Reference                      |
|----|--------------|----------------|-------------------------|----------------------------------|--------------------------------|
| 1  | M3           | 11, 12         | 22, 23, 24 <sup>a</sup> | Proximal epiphyseal breadth      | Martin and Saller [53]         |
| 2  | BB           | 9, 10          |                         | Biarticular breadth              | Holland [11]                   |
| 3  | M3a          | 1, 2           |                         | Medial condyle articular width   | Martin and Saller [53]         |
| 4  | M3b          | 5, 6           |                         | Medial condyle articular length  | Martin and Saller [53]         |
| 5  | M4a          | 3, 4           |                         | Lateral condyle articular width  | Martin and Saller [53]         |
| 6  | M4b          | 7, 8           |                         | Lateral condyle articular length | Martin and Saller [53]         |
| 7  | M6           | 17, 21         | 18, 20 <sup>b</sup>     | Distal epiphysel breadth         | Martin and Saller [53]         |
| 8  | DB           | 17, 19         |                         | Distal breadth                   | Işcan and Miller-Shaivitz [10] |
| 9  | M8a          | 13, 15         |                         | Anteroposterior diameter         | Martin and Saller [53]         |
| 10 | M9a          | 14, 16         |                         | Transverse diameter              | Martin and Saller [53]         |

<sup>a</sup> Those landmarks were used to construct medial plane and landmarks 11 and 12 lay in the tangent plane which is parallel to defined medial plane

<sup>b</sup> Landmark 21 was computed as the middle point between landmarks 18 and 20

The selection was adjusted to several conditions:

1. Dimensions had to correspond with the selection of discriminant functions.
2. According to the related literature, the most dimorphic dimensions were selected (area of the knee joint mostly).
3. Circumferential dimensions were excluded because of the difficulty of their reproducibility in a virtual environment as opposed to measuring real dry bones.
4. Finally, length was also left out as the length of the bone does not perceptibly contribute to sex classification, compared to joint dimensions.

Measurements were carried out in the Morphome3cs software ([www.morphome3cs.com](http://www.morphome3cs.com)), which was developed in the Department of Software and Computer Science under the Charles University in Prague. In the environment of this program, we located 23 landmarks on surface models of each bone and one landmark (number 21) was computed by the software. For the dimension M3, we first had to construct the medial plane (defined by landmarks 22–24) and then we located landmarks 11 and 12 as the most laterally prominent points of proximal part. These two points lied on the tangent plane which is parallel to defined medial plane. Dimension M6 was measured by locating auxiliary landmarks 18 and 20 and defining landmark 21 as their midpoint. Dimension M6 is the distance between the landmarks 17 and 21. Other dimensions were measured simply as the distances between two relevant landmarks. The locations of all landmarks are documented in Fig. 1. All metric values were collected by one observer.

#### Intraobserver error

Intraobserver rate was calculated by relocating all main landmarks of seven randomly selected tibiae. Repeated

measurements were performed seven times with at least 1-day interval between them. Average error was calculated in Morphome3cs software and reached 0.4734 mm.

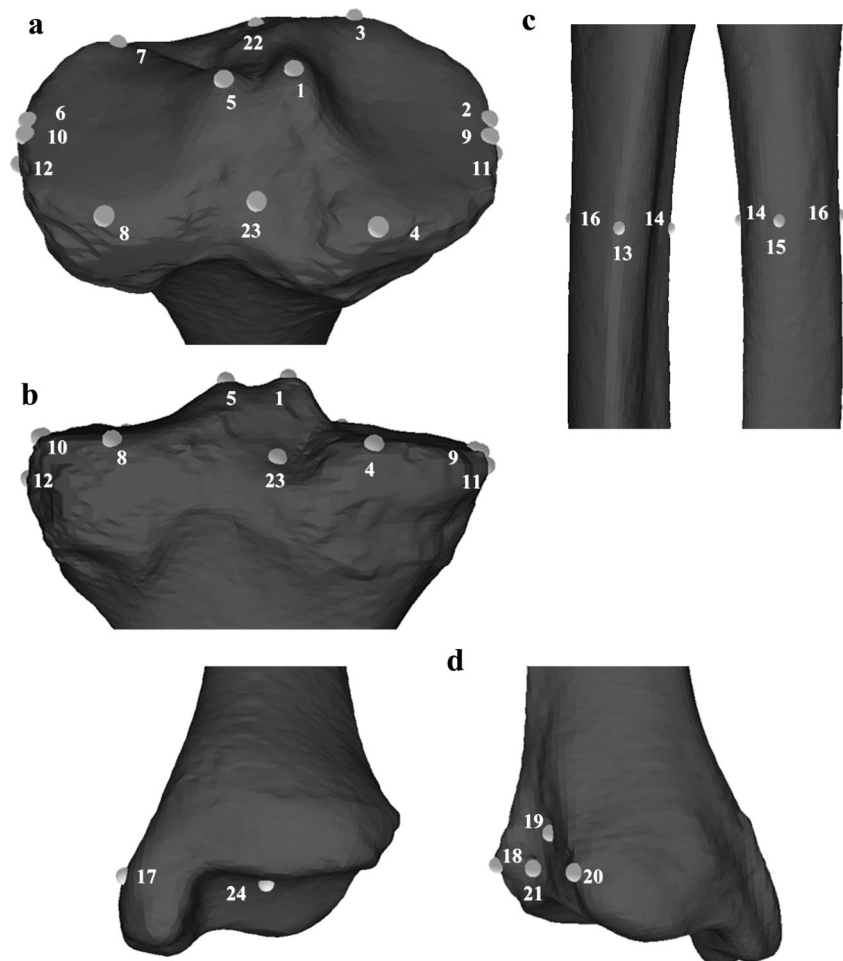
#### Selection of published discriminant functions

The process of selection of discriminant functions follows these requirements:

1. Only multidimensional discriminant functions were chosen, because one dimension cannot reliably determine the sex, due to overlapping features of sexual dimorphism between female and male [54].
2. Sex bias could not exceed 7.5 %.
3. The accuracy of sex estimation was greater than 75 %.
4. Selection was made with regard to the criteria and limits of the selected metric dimensions.
5. Selected functions were designed for both European and non-European populations from various periods of time.

Based on these criteria, we chose nine published discriminant functions whose coefficients for each variable, constants, sectioning points, and levels of accuracy are summarized in Table 3. Discriminant functions 1–4 were proposed by Brůžek [14] for a Portuguese population whose birth falls between the early nineteenth and early twentieth centuries. DF 5–8 proposed by Kranioti and Apostol [21] for a sample of south European populations (Spanish, Italian, Greek, and pooled populations) whose death falls into the second half of twentieth century. Last discriminant function DF9 proposed by Işcan and Shaivitz [10] for the North American population whose was born between the early nineteenth century and first half of the twentieth century.

**Fig. 1** Location of landmarks on the **a** proximal part (1–12, 22, 23), **b** proximal part posterior view, **c** body (13–16), and **d** distal part (17–21, 24) of tibia



To evaluate classification performance, accuracy and sex bias were used. According to several authors [25, 55, 56] sex bias is computed as the difference between the classification accuracy of males and that of females, both expressed as percentages. A sex bias lower than 5 % is ideal in forensic anthropology [25]; however, to increase the number of functions available for comparison, a limit of 7.5 % was used.

#### Selection of dimensions used for proposing own CF

For classification purposes, the most sexual dimorphic dimensions of tibia were chosen. We thus preferred the dimensions on the proximal part of bone in the area of knee joint (BB, M3, M3a, M3b, M4a, M4b). The last dimension M6 was selected because several authors [9, 12, 14, 20] had used some of the dimensions which measured distal breadth of bone in combination with measurements on the proximal part. Then, using the software Statistica [57], we tested combinations of two or more dimensions and selected the ones with the greatest discriminatory power.

#### Statistical analyses

Basic statistical characteristics and the application of Czech dimensions into the selected discriminant functions were performed in MS Excel. Two-sample *t* test was used to quantify the level of dimorphism in the measured lengths. Descriptive statistics for both sexes of the Czech sample, including standard deviation and *t* test for each dimension are shown in Table 4. As expected, males have greater tibial dimensions than females. *t* Test indicates that all of these differences are statistically significant ( $\alpha = 0.05$ ). Only one dimension (DB) was found above the significance level. For all these statistical analyses and for computing our own discriminant functions as well, Statistica software [57] was used. Also tenfold cross-validation was done for linear discriminant functions in program R.

#### Classification techniques

For the classification, we used two different approaches. Linear DFA in software Statistica and one advanced data modeling method based on the Group of Adaptive Models

**Table 3** Selection of published discriminant functions with coefficients, sectioning points, and accuracy of sex determination

| Discriminant functions |              | DF1                                  | DF2         | DF3         | DF4         | DF5  | DF6                       | DF7                       | DF8                       | DF9   |
|------------------------|--------------|--------------------------------------|-------------|-------------|-------------|--|---------------------------|---------------------------|---------------------------|---|
|                        |              | Brůžek [14]<br>Portuguese population | Brůžek [14] | Brůžek [14] | Brůžek [14] | Kranioti and Apostol [21]<br>Spanish, Italian, Greek, and pooled populations | Kranioti and Apostol [21] | Kranioti and Apostol [21] | Kranioti and Apostol [21] | Işcan and Miller-Shaivitz [10]<br>North American population |
| Variable               | Coefficients |                                      |             |             |             |  |                           |                           |                           |   |
| M3                     |              |                                      |             |             |             | 0.2980   | 0.1379                    | 0.1724                    | 0.2255                    | 0.2346  |
| BB                     | 0.3209       | 0.2968                               |             |             |             |  |                           |                           |                           |   |
| M3a                    | -0.2450      | -0.2355                              |             |             |             |  |                           |                           |                           |   |
| M3b                    | 0.1736       | 0.1661                               | 0.2831      |             |             |  |                           |                           |                           |   |
| M4a                    |              | 0.0622                               |             |             |             |  |                           |                           |                           |   |
| M4b                    | 0.4658       | 0.4492                               | 0.5654      |             |             |  |                           |                           |                           |   |
| M6                     |              |                                      |             |             | 0.0066      | 0.1460   | 0.1527                    | 0.0543                    |                           |   |
| DB                     |              |                                      |             |             |             |  |                           |                           |                           | 0.0977  |
| M8a                    |              |                                      |             | 0.3654      |             |  |                           |                           |                           |   |
| M9a                    |              |                                      |             | 0.4289      |             |  |                           |                           |                           |   |
| Constant               | -36.965      | -37.485                              | -29.236     | -21.640     | -21.822     | -19.164  | -18.924                   | -18.760                   | -21.359                   |   |
| Sectioning point       | -0.0632      | -0.0632                              | -0.0632     | -0.7499     | 0           | 0  | 0                         | 0                         | 0                         |   |
| Accuracy M             | 84.8         | 84.8                                 | 82.6        | 80.4        | 95.2        | 80.4   | 89.4                      | 84.1                      | 85.0                      |   |
| F                      | 83.7         | 83.7                                 | 87.8        | 77.6        | 92.0        | 91.4   | 85.9                      | 87.6                      | 84.6                      |   |

Pop population, M male, F female

Evolution (GAME) [58, 59]. The GAME method was utilized to search for new classification functions. This inductive method is based on a feed-forward artificial neural network and consists of different types of transfer functions. Both, the structure of the network and the parameters of the transfer functions, are set automatically during a run of GAME. The final structure of the network represents the new classification function. An example of such a network structure is shown in figure (Fig. 2).

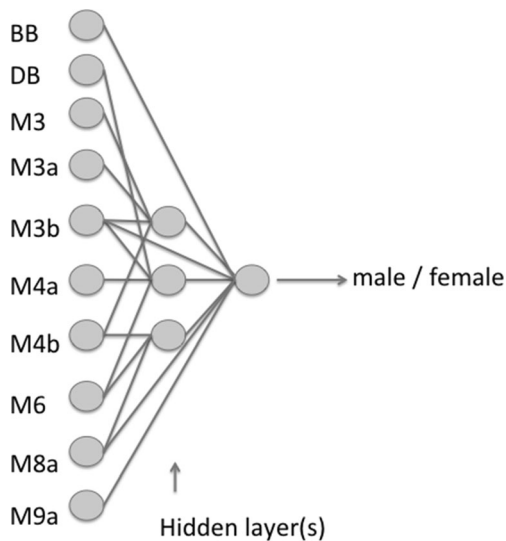
Because of the random initialization, the GAME method provides different discriminant functions in each run of the method. The result DF can be either linear or nonlinear. To obtain representative results from the GAME method, we used ten times repeated tenfold cross-validation.

Apart from the classification functions, the GAME method encapsulates also a feature ranking method called FeRaNGA [60]. The FeRaNGA method provides

**Table 4** Descriptive statistics of the measured variables of the tibia from Czech population

| Dimension (mm) | Male—descriptive statistics, N= 30 |      |       |       | Female—descriptive statistics, N= 26 |      |       |       | t test |
|----------------|------------------------------------|------|-------|-------|--------------------------------------|------|-------|-------|--------|
|                | Mean                               | SD   | Min   | Max   | Mean                                 | SD   | Min   | Max   |        |
| M3             | 81.67                              | 4.01 | 72.51 | 90.02 | 76.85                                | 5.29 | 69.92 | 92.10 | 0.000  |
| BB             | 77.72                              | 3.99 | 69.48 | 85.72 | 72.47                                | 4.75 | 64.57 | 84.20 | 0.000  |
| M3a            | 34.13                              | 2.21 | 30.15 | 39.71 | 31.91                                | 2.39 | 27.66 | 36.49 | 0.001  |
| M3b            | 34.04                              | 2.91 | 29.05 | 40.43 | 31.79                                | 2.41 | 28.43 | 39.20 | 0.003  |
| M4a            | 49.40                              | 3.24 | 41.44 | 55.32 | 45.89                                | 3.74 | 40.69 | 56.22 | 0.001  |
| M4b            | 39.57                              | 3.43 | 31.69 | 45.61 | 36.42                                | 2.61 | 32.13 | 42.11 | 0.000  |
| M6             | 54.93                              | 3.52 | 48.91 | 63.74 | 50.43                                | 2.56 | 43.16 | 56.91 | 0.000  |
| DB             | 50.04                              | 2.93 | 45.25 | 55.59 | 48.29                                | 3.50 | 41.42 | 56.40 | 0.050  |
| M8a            | 37.74                              | 3.95 | 30.20 | 46.37 | 34.91                                | 3.19 | 29.69 | 43.22 | 0.006  |
| M9a            | 28.32                              | 2.18 | 24.91 | 33.32 | 26.09                                | 2.29 | 23.19 | 32.62 | 0.001  |

N number of individuals, SD standard deviation, Min minimum, Max maximum



**Fig. 2** Example structure of the GAME neural network with one input layer, one hidden layer, and one output layer for classification of males and females. The number of hidden layers can differ for each particular run of the algorithm

importance of particular input variables used during the ten times repeated tenfold cross-validation process while searching for classification functions.

## Results

### Discriminant analyses

#### *Computation of DFA in Czech sample*

We designed two linear discriminant function equations (Table 5) for modern Czech population. First of them (CF1) includes seven variables (BB, M3, M3a, M3b, M4a, M4b, and M6) and provides 85.7 % (83.3 % males, 88.5 % females) of correct determination (sex bias  $-5.2$  %). The second uses five variables (BB, M3a, M3b, M4a, and M4b) and determines sex with 82.1 % (80.0 % males, 84.6 % females) of correct assessment (sex bias  $-4.6$  %). Cross-validation is not identical to the original data; in the first case, value is equal to 75.0 % and in the second to 76.8 %. The sectioning point is equal to 0, decision value above 0 indicate a male, while values below zero indicate a female.

#### *Advanced classification technique (GAME) in Czech sample*

The GAME method creates automatically a structure of the artificial neural network. The structure is represented by a classification function. Due to a random initialisation of GAME at the beginning, the classification function is different for each run of the method. From this reason, we used ten times repeated tenfold cross-validation (thus, we created 100

classification functions) to get mean classification accuracy of the GAME method for representative comparison with another approaches. Each particular classification function has different level of complexity. It can be either linear, nonlinear or any combination of both types. Classification function with the highest classification accuracy does not necessarily means that it is also applicable in practice (e.g., an exponential function of a logarithmic function having as an argument a sigmoidal function is not the right candidate for practice). The overall average classification accuracy from 100 runs of the GAME method was 82.35 % with a standard deviation of 2.25 %. Based on above mentioned we selected three user friendly and practically applicable classification functions with high classification accuracy while having minimum complexity. One linear classification function (CF3) and two nonlinear classification functions (CF4, CF5) are shown in Table 5. The linear CF3 includes six variables (DB, M3a, M4a, M4b, M6, M9a) and provides classification accuracy of 83.9 % (males 83.3 %, females 84.6 %) with sex bias of  $-1.3$  %. The first of our nonlinear functions is CF4, and its type is exponential. The classification accuracy of CF4 is the same as for CF3 with the same sex bias, though it needs only four variables to reach such accuracy. The last classification function CF5 is of sigmoidal type and provides the highest classification accuracy (from CF1 to CF5) with 87.5 % (male 90 %, female 84.6 %) and sex bias 5.38 %.

The discriminatory value for CF3, CF4, and CF5 is equal to 0.5. Smaller value indicates male subject, while greater or equal values indicates female subject.

The importance of particular input variables determined by the FeRaNGA method during the search for best classification function are listed from the most important to the least important followed by their importance: M6 (28.2 %), DB (21.7 %), BB (16.4), M3 (14.9 %), M9a (8.7 %), M3a (4.8 %), M4a (2.5 %), M4b (1.8 %), M8a (0.6 %), and M3b (0.4 %). The four most important variables are M6, DB, BB, and M3.

#### *Application of dimensions from Czech sample into published DFA*

Table 6 compares classification success rates in the original reference sample and in target sample (Czech population) separately for males and females. When we used Czech dimensions in DF1, originally created for Portuguese sample, almost all of Czech males (90 %) were determined correctly, only three of them were by this function considered as female. On the other hand, females were overestimated and determined as male. Only 38.5 % of female sample was successfully classified. We got very similar results using DF2. Success of classification among males was 93.3 % and among females coincidentally 38.5 %. DF3 classified correctly 83.3 % of males and more than half (53.9 %) of females. The last tested function originally proposed for Portuguese population,

**Table 5** The proposed classification functions in the Czech population

|                                      | Classification functions |         |        |        |        |
|--------------------------------------|--------------------------|---------|--------|--------|--------|
|                                      | CF1                      | CF2     | CF3    | CF4    | CF5    |
| Variable/no. of used variables in CF | 7                        | 5       | 6      | 4      | 3      |
| M3                                   | -0.434                   |         |        |        | *      |
| BB                                   | 0.576                    | 0.199   |        | *      |        |
| M3a                                  | 0.005                    | 0.102   | 0.004  |        |        |
| M3b                                  | -0.162                   | -0.105  |        |        |        |
| M4a                                  | 0.056                    | 0.011   | -0.042 |        |        |
| M4b                                  | 0.169                    | 0.163   | -0.042 |        | *      |
| M6                                   | 0.395                    |         | -0.071 | *      |        |
| DB                                   |                          |         | 0.073  | *      | *      |
| M9a                                  |                          |         | -0.016 | *      |        |
| Constant                             | -33.724                  | -21.565 | 4.315  |        |        |
| Sectioning point                     | 0                        | 0       | 0.5    | 0.5    | 0.5    |
| Classification success (accuracy)    | Male                     | 83.3 %  | 80.0 % | –      | –      |
|                                      | Female                   | 88.5 %  | 84.6 % | –      | –      |
|                                      | Average                  | 85.7 %  | 82.1 % | –      | –      |
| Cross-validated                      | Male                     | –       | –      | 83.3 % | 83.3 % |
|                                      | Female                   | –       | –      | 84.6 % | 84.6 % |
|                                      | Average                  | 75.0 %  | 76.8 % | 83.9 % | 83.9 % |
| Sex bias                             | -5.2 %                   | -4.6 %  | -1.3 % | -1.3 % | -5.4 % |

CF4 is exponential function and CF5 is sigmoidal, for that reason they are presented out off table directly beneath. Used variables in those cases are marked by “\*”

$$CF4 = 2.058 * EXP(-0.041 * M6 - 0.052 * M9a + 3.516) - 0.017 * BB + 0.065 * DB - 3.452$$

$$CF5 = \frac{1}{1 + \exp(-23.18 * DB + 70.676 * M3 + 3.642 * M4b - 4489.232)}$$

DF4, estimated males flawlessly (100 %) but totally overestimated females of our sample (0 %). DF5 was fitted to Spanish population and separated properly 100 % of males and just 11.5 % of females. Discriminant function 6 derived from Italian population was not eventually incorporated (see below). DF7 was proposed on Greek population, and this function estimated absolutely correctly in our male sample,

and largely overestimated female sample (3.9 %). DF8 was derived from mixed south European populations, and this function gave 100 % of correct assessment of males and 7.7 % of correct determination of females. The last tested discriminant function DF9 proposed for Euro-American population successfully classified all Czech males but only 7.7 % in female sample.

**Table 6** Application of DFA proposed in different populations in recent Czech population: simulation of disregarding DFA population specificity

| Discriminant function | Classification success in reference sample (%) |        |             |          | Classification success in Czech sample (%) |        |             |          |
|-----------------------|--|--------|-------------|----------|--|--------|-------------|----------|
|                       | Male   | Female | Pooled mean | Sex bias | Male                                       | Female | Pooled mean | Sex bias |
| DF1 (Portuguese)      | 84.8   | 83.7   | 84.2        | 1.1      | 90.0                                       | 38.5   | 66.1        | 51.5     |
| DF2 (Portuguese)      | 84.8   | 83.7   | 84.2        | 1.1      | 93.3                                       | 38.5   | 67.9        | 54.8     |
| DF3 (Portuguese)      | 82.6   | 87.8   | 85.3        | -5.2     | 83.3                                       | 53.9   | 69.7        | 29.4     |
| DF4 (Portuguese)      | 80.4   | 77.6   | 79.0        | 2.8      | 100.0                                      | 0.0    | 53.6        | 100      |
| DF5 (Spanish)         | 95.2   | 92.0   | 93.4        | 3.2      | 100.0                                      | 11.5   | 58.9        | 88.5     |
| DF7 (Greek)           | 89.4   | 85.9   | 87.8        | 3.5      | 100.0                                      | 3.9    | 55.4        | 96.1     |
| DF8 (pooled)          | 84.1   | 87.6   | 86.0        | -3.5     | 100.0                                      | 7.7    | 57.1        | 92.3     |
| DF9 (Euro-American)   | 85.0   | 84.6   | 84.8        | 0.4      | 100.0                                      | 7.7    | 57.1        | 92.3     |

The same table compares computed sex biases of reference and target sample. While values of published DFA are low (<7.5 %), sex biases of tested functions are unacceptably high (above 29.4 %).

## Discussion

Forensic anthropology benefits from accurate and reliable methods for estimating parameters of an individual's biological profile. The Daubert standard requires forensic anthropologists to confirm the validity of their methodologies through empirical testing, peer review, publication, and calculating error rates. External validity must be tested, and if the proposed procedures are based on one population, application to other population is required [1, 61, 62]. In cases of natural mass disasters, accidents, and mutilated human remains, sex estimation is often based on single skeletal element, most frequently on the long bones of extremities, vertebrae, or skull fragments [63, 64]. The success of determining sex is dependent not only on the method and degree of sexual dimorphism of the reference population in which the method was proposed, but also on the degree of sexual dimorphism of the target population in which the method is applied. Moreover the reliability of the sex allocation is influenced by secular trend [27, 28]. However, for these reasons, as we demonstrated in the **Introduction**, most of the methods for sex estimation are population-specific. At the same time, the tibia may be more susceptible to short-time secular changes and thus, more research is required to test this hypothesis [21]. These short-term changes in dimensions of the tibia are also affected by the increasing prevalence of obesity [65]. A number of methods have been proposed on the identified skeleton collections that come from populations that do not longer exist. A number of such studies were published on historical skeletal collection and their methods are offered and available to users [8–15, 66]. As noted by Ousley and Jantz (2012), “Data from modern American show that standards derived from the nineteenth century are not appropriate for the assessment of twentieth-century groups.” and “Also, postcranial samples from the Hamann-Todd Collection used by Iscan and Cotton (1990) perform very poorly when applied to modern cases” [67]. However, a major problem with use of the sex classification tools is that the population is not stable and is changed continuously under the influence of modifications in socioeconomic factors. Therefore, it is important to know the risk of errors, when the population specificity of the methods is completely ignored.

Altogether, we proposed five classification functions in Czech sample to determine sex (two discriminant functions, one linear, exponential, and sigmoidal). Maximum classification accuracy using a linear DF analysis was 85.7 % and –5.2 % sex bias with seven

variables included (CF1). However, the highest classification accuracy of 87.5 % with –5.4 % sex bias and comprising three variables (CF5) was achieved using feed-forward artificial neural network–GAME method. This confirms that other classification functions than classic linear DFA are suitable as classification methods [24, 68]. However, with all five functions, we reach an acceptable sex bias according to Franklin et al. [25], ranging from –1.3 to –5.4 %, only two of them slightly exceed the 5 % limit. More importantly, we have achieved comparable success rate and sex bias (Table 5) as other mentioned authors did in other population samples. Similar results for samples from twentieth and twenty-first centuries reached also Brzobohatá et al. [52] using geometric morphometry. Although CF5 (sigmoidal function) provides the best determination of sex, it has the greatest sex bias (5.4 %), which is still acceptable. It should be noticed that CF3 (linear) and CF4 (exponential) have the same accuracy and even sex bias, but CF3 includes six variables and CF4 only four variables. Generally, nonlinear classification functions (CF4 and CF5) tend to require less input variables. This is advantageous, at least in cases of incomplete or damaged remains, where it is not possible to measure all dimensions.

In this study, we tested accuracy of discriminant functions fitted to various populations with dimensions measured on our Czech. We did it in the same manner as a potential user when estimating the sex from the tibia. We used only multivariable methods because a single variable has little practical value for sex assessment according to Peckmann et al. [37] due to the large overlapping area of both sexes. Results reveal strong sexual dimorphism of the tibial dimensions in the Czech population which significantly contributes to sex discrimination.

When classifying the Czech sample using seven DF from three different Mediterranean European populations [14, 21] or DF from American Caucasians [10], sex estimation was done correctly only in 53.6–69.7 % of cases. However, regardless of the rate of misclassification that is unacceptable for practical use in the forensic sciences, the sex bias ranged from 29.4 to 100 %. The results repeatedly demonstrated on the one hand, almost absolutely correct determination of males (DF1-5, DF7-9) and on the contrary, failure of classification of female individuals. DF6 is incorrect, and there is a mistake in coefficients or constant computed for function IF4 in the original paper of Kranioti and Apostol [21]. It can be verified by solving discriminant equation with mean values; in that case, Italians of both sexes have negative values despite sectioning point of zero.

The low classification rates in females were observed also in sexing of the metacarpals [69], and results in this study indicate that the standards developed from the continental Greece are not proper for application in forensic cases in Crete because they do not represent the local Cretan population. The original



sectioning point sometimes completely misses the distribution of the sexes. Low accuracy together with high sex bias is unacceptable for forensic anthropology. Similar results, proving that population differences cause that DF that perform well on one population may produce high error rates and sex biases when applied to other populations, were noted by several authors in crania [55, 70, 71].

The most likely explanations of such results are the variability of individual populations [72], different degree of sexual dimorphism of tibia among populations [73] and thus different sex-specific means of dimensions of each population (variance of values dimensions). We should take into consideration that unlike chosen studies, where DF1-DF9 were proposed, we used 3D models of bones. According to previous studies [74, 75], where authors support their comparability on skulls, we believe that measurements derived from 3D models are comparable to those from dry bones. Therefore, the general effect of applying nonpopulation standards reduces the accuracy of classification, the magnitude of which is proportionately related to the degree of divergence in size between the original reference sample and the individual to which those standards will be applied. It follows that the main limitation of using published methods for estimating sex using the metric data for tibia (and other bones) is their population specificity (e.g., [21, 42]). Evidence follows from contradiction of comparing the results of reliability and accuracy of classification methods. Population specificity is also related to the trend of globalization [76]. Problem is that the individual whose sex we want to estimate does not necessarily originally come from the area in which their remains were found (e.g., [4, 77]). The population affinity or ancestry of an unknown specimen in the worldwide increase in admixture of human populations is difficult (e.g., [78–80]).

What options are there for dealing with population specificity of methods which is characteristic not only of sex estimation methods but also of all methods estimating the biological profile of the individual? Forensic anthropology is constrained by a paucity of population-specific standards as number of repositories of documented skeletons, traditionally the main source of population-specific data, is limited [62]. However, as Albanese et al. [81] mentioned, group specificity should increase the precision of the estimation methods, but on the other hand, in many cases, the parameters of a given group are based on assumptions which make assigning an unknown to a group under the best conditions problematic and at worst impossible. Methods developed in pooled samples acquired, contrary to traditional assumptions, certain degree of robustness. For example, pooled-ancestry stature equation can be more appropriate than population-specific equations obtained from a single sample [82]. It should be kept in mind that it is not possible to respect the population specificity in bioarchaeology. The only possible solution

could be the creation of regional standards in Europe, which correspond to variability of body size, and on such a basis, a robust method for sex estimation should be developed. It was similarly designed in the study concerning stature in bioarchaeology [83], where the authors created northern and southern European formulae.

Medical imaging (CT, MRI, ultrasonography) of living individuals offer appropriate and reliable source of contemporary population data from selected geographic populations. These data can be pooled in a metapopulation sample from which skeletal standards for the estimation of age, sex, and stature could be developed. The combination of recent data from the CT images and historical collections of identified individuals could lead to suggestions of highly robust methods, whose reliability would provide a reliable estimation of the biological profile in forensic anthropology and bioarchaeology. For these reasons, one of the goals of the international scientific community is to develop a synergistic, well-coordinated activity to create a database of 3D imaging from clinical computed tomographic scans of the skeleton to formulate new nonpopulation-specific methods.

## Conclusion

Sexual dimorphism of the tibia is well established in the literature and standards for the Czech have been proposed. However, all the DFA proposed in Czech sample and all of already published functions in other populations were population-specific with classification accuracy ranging between 80 and 90 %. Disregarding or ignoring this fact causes complete failure of sex estimation. Direct application of published population-specific DFA in Czech sample leads to unacceptably low accuracy with sex bias ranging from –29.4 to 100 %. This failure of sex assessment is caused by a variability of skeletal dimensions among populations. For these reasons, the misclassification rate does not correspond to the Daubert criteria, and therefore, the DFA from tibia cannot be recommended also for the estimation of sex of recent skeletal remains in Europe.

**Acknowledgments** We would like to thank anonymous reviewer since he revealed that in the original article of Kranioti and Apostol must be numerical mistake in one proposed discriminant function.

## References

1. Cattaneo C (2007) Forensic anthropology: developments of a classical discipline in the new millenium. *Forensic Sci Int* 165:185–193
2. Dawson C, Ross D, Mallet X (2011) Sex determination. In: Black S, Ferguson E (eds) *Forensic Anthropology: 2000 to 2010*. CRC Press, Boca Raton, pp 61–94

3. Moore MK (2013) Sex Estimation and Assessment. In: DiGangi EA, Moore MK (eds) *Research Methods in Human Skeletal Biology*. Academic Press, London, pp 91–116
4. Spradley MK, Jantz RL, Robinson A, Peccerelli F (2008) Demographic Change and Forensic Identification: Problems in Metric Identification of Hispanic Skeletons. *J Forensic Sci* 53:21–28
5. Özer BK, Özer I, Sağır M, Güleç E (2014) Sex determination using the tibia in an ancient anatolian population. *Med Archaeol Archaeom* 14:329–336
6. Novotný V, Işcan MY, Loth SR (1993) Morphologic and osteometric assesment of age, sex and race from the skull. In: Işcan MY, Helmer RP (eds) *Forensic analyses of the skull*. Wiley-Liss Inc., New York, pp 71–88
7. Dangar KP, Pandya AM, Rathod SP, Tank KC, Akbari VJ, Solanki SV (2012) Sexual dimorphism of proximal epiphyseal breadth of tibia. *Int J Biol Med Res* 3:1331–1334
8. Işcan MY, Miller-Shaivitz P (1984a) Determination of sex from the tibia. *Am J Phys Anthropol* 64:53–57
9. 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
10. Işcan MY, Miller-Shaivitz P (1984b) Discriminant function sexing of the tibia. *J Forensic Sci* 29:1087–1093
11. Holland TD (1991) Sex assessment using the proximal tibia. *Am J Phys Anthropol* 85:221–227
12. Kieser JA, Moggi-Cecchi J, Groeneveld HT (1992) Sex allocation of skeletal material by analysis of the proximal tibia. *Forensic Sci Int* 56:29–36
13. Işcan MY, Yoshino M, Kato S (1994) Sex determination from the tibia: standards from contemporary Japan. *J Forensic Sci* 39:785–792
14. Brůžek J (1995) Diagnose sexuelle à l'aide de l'analyse discriminante appliquée au tibia. *Anthropol Port* 13:93–106
15. Steyn M, Işcan MY (1997) Sex determination from the femur and tibia in South African whites. *Forensic Sci Int* 90:111–119
16. Gonzales-Reimers E, Velasco-Vázquez J, Arnan-de-la-Rosa M, Santolaria Fernandez F (2000) Sex determination by discriminant function analysis of the right tibia in the prehistoric population of the Canary Islands. *Forensic Sci Int* 108:165–172
17. Sakaue K (2004) Sexual determination of long bones in recent Japanese. *Anthropol Sci* 112:75–81
18. Šlaus M, Tomičić Ž (2005) Discriminant function sexing of fragmentary and complete tibiae from medieval Croatian sites. *Forensic Sci Int* 147:147–152
19. Garcia S (2010) Is the circumference at the nutrient foramen of the tibia of value to sex determination on human osteological collections? Testing a new method. *Int J Osteoarchaeol* 22:361–365
20. Šlaus M, Bedić Ž, Strinović D, Petrovečki V (2013) Sex determination by discriminant function analysis of the tibia for contemporary Croats. *Forensic Sci Int* 226:302e1–e4
21. Kranioti EF, Apostol MA (2015) Sexual dimorphism of the tibia in contemporary Greeks, Italians, and Spanish: forensic implications. *Int J Legal Med* 129:357–363
22. Işcan MY, Steyn M (2013) *The Human Skeleton in Forensic Medicine*. Charles C. Thomas Publisher, Springfield, p 493
23. Robinson MS, Bidmos MA (2011) An assessment of the accuracy of discriminant function equations for sex determination of the femur and tibia from a South African population. *Forensic Sci Int* 206:212.e1–212.e5
24. Santos F, Guyomarc'h P, Brůžek J (2014) Statistical sex determination from craniometrics: Comparison of Linear Discriminant Analysis, Logistic Regression, and Support Vector Machines. *Forensic Sci Int* 245:204.e1–204.e8
25. Franklin D, Cardini A, Flavel A, Kuliukas A (2013) Estimation of sex from cranial measurements in a Western Australian population. *Forensic Sci Int* 229:158.e1–158.e8
26. Buikstra JE, Ubelaker DH (1994) Standards for data collection from human skeletal remains. *Arkansas Archaeological Survey Research, Fayetteville*
27. Jantz RL (2001) Cranial change in Americans: 1850–1975. *J Forensic Sci* 46:784–787
28. Meadows Jantz L, Jantz RL (1999) Secular change in long bone length and proportion in the United States, 1800–1970. *Am J Phys Anthropol* 110:57–67
29. Meadows L, Jantz RL (1995) Allometric secular change in the long bones from the 1800s to the present. *J Forensic Sci* 40:762–767
30. Saini V (2014) Significance of temporal changes on sexual dimorphism of cranial measurements of Indian population. *Forensic Sci Int* 242:300e1–300e8
31. Mascie-Taylor CGN, Bogin B (1995) *Human variability and plasticity*. Cambridge University Press, Cambridge
32. Ishak NI, Hemy N, Franklin D (2012) Estimation of stature from hand and handprint dimensions in a Western Australian population. *Forensic Sci Int* 216:199.e1–199.e7
33. Danubio ME, Sanna E (2008) Secular changes in human biological variables in Western Countries: An updated review and synthesis. *J Anthropol Sci* 86:91–112
34. Sanna E, Milia N, Martella P, Danubio M (2015) Body and head dimensions of adults in Sardinia (Italy) support different intensities of relative secular trends. *J Anthropol Sci* 93:157–162
35. Bigoni L, Velemínska J, Brůžek J (2010) Three-dimensional geometric morphometric analysis of cranio-facial sexual dimorphism in a Central European sample of known sex. *Homo* 61:16–32
36. Timonov P, Fasová A, Radoniova D, Alexandrov A, Delev D (2014) A study of sexual dimorphism in the femur among contemporary Bulgarian population. *Euras J Anthropol* 5:46–53
37. Peckmann TR, Orr K, Meek S, Manolis SK (2015) Sex determination from the calcaneus in a 20th century Greek population using discriminant function analysis. *Sci Justice* 55:377–382
38. Giles E, Elliot O (1963) Sex determination by discriminant function analysis of crania. *Am J Phys Anthropol* 21:53–68
39. Steyn M, Işcan MY (1998) Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Sci Int* 98:9–16
40. Franklin D, Flavel A, Kuliukas A, Oxnard CE, Marks MK (2012) Cranial sexual dimorphism and anthropological standards: Preliminary investigations in a Western Australian population. 64th Annual Meeting of the American Academy of Forensic Sciences, Atlanta
41. Králík M, Urbanová P, Wagenknechtová M (2014) Sex assessment using clavicle measurements: Inter- and intra-population comparisons. *Forensic Sci Int* 234:181e1–181e15
42. Bidmos MA, Dayal MR (2004) Further evidence to show population specificity of discriminant function equations for sex determination using the talus of South African blacks. *J Forensic Sci* 49:1165–1170
43. Ross AH, Ubelaker DH, Kimmerle EH (2011) Implications of dimorphism, population variation, and secular change in estimating population affinity in the Iberian Peninsula. *Forensic Sci Int* 206:214.e1–214.e5
44. Steyn M, Patriquin ML (2009) Osteometric sex determination from the pelvis - Does population specificity matter? *Forensic Sci Int* 191:113e1–113e5
45. Guyomarc'h P, Velemínská J, Sedlak P, Dobisíková M, Švenkrťová I, Brůžek J (2016) Impact of secular trend on sex assessment evaluated through femoral dimensions in a Czech population. *Forensic Sci Int Sci Int*. doi:10.1016/j.forsciint.2016.02.042
46. Dedouit F, Telmon N, Costagliola R, Otal P, Florence LL, Joffre F, Rougé D (2007) New identification possibilities with postmortem multisllice computed tomography. *Int J Legal Med* 121:507–510
47. Ramsthaler F, Kettner M, Gehl A, Verhoff MA (2010) Digital forensic osteology: Morphological sexing of skeletal remains using volume-rendered cranial CT scans. *Forensic Sci Int* 195:148–152

48. Stull KE, Tise ML, Ali Z, Fowler DR (2014) Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. *Forensic Sci Int* 238:133–140
49. Gulhan O, Harrison K, Kiris A (2015) A new computer-tomography-based method of sex estimation: Development of Turkish population-specific standards. *Forensic Sci Int* 255:2–8
50. Mesteková Š, Brůžek J, Velemínská J, Chaumoitre K (2015) A Test of the DSP Sexing Method on CT Images from a Modern French Sample. *J Forensic Sci* 60:1295–1299
51. Clavero A, Salicrú M, Turbón D (2015) Sex prediction from the femur and hip bone using a sample of CT images from a Spanish population. *Int J Legal Med* 129:373–383
52. Brzobohatá H, Krajčiček V, Horák Z, Velemínská J (2015) Sex Classification Using the Three-Dimensional Tibia Form or Shape Including Population Specificity Approach. *J Forensic Sci* 60:29–40
53. Martin R, Saller K (1957) *Lehrbuch der Anthropologie in systematischen darstellung mit besonderer berücksichtigung der anthropologischen methoden*. Band I, Gustav Fischer Verlag, Stuttgart
54. Sjøvold T (1988) Geschlechtsdiagnose am Skelett. In: Martin R, Knussmann R (eds) *Anthropologie, Handbuch des vergleichenden Biologie des Menschen*. Gustav Fischer Verlag, Stuttgart, pp 444–480
55. Walker PL (2008) Sexing skulls using discriminant function analysis of visually assessed traits. *Am J Phys Anthropol* 136:39–50
56. Franklin D, Cardini A, Flavel A, Kuliukas A, Marks MK, Oxnard CE, O'Higgins P (2012) Estimation of sex from sternal measurements in a Western Australian population. *Forensic Sci Int* 217: 230.e1–230.e5
57. Statistica [software]. version 12 ([www.statsoft.cz](http://www.statsoft.cz)).
58. Kordík P (2009) GAME-hybrid self-organizing modeling system based on GMDH. In: Onwubolu GC (ed) *Hybrid Self-Organizing Modeling Systems*. Springer-Verlag, Heidelberg, pp 233–280
59. Kordík P, Koutník J, Drchal J, Kovářik O, Čepek M, Šnorek M (2010) Meta-learning approach to neural network optimization. *Neural Netw* 23:568–582
60. Pilný A, Kordík P, Šnorek M (2008) Feature Ranking Derived from Data Mining Process. In: *Proceedings of the 18th international conference on Artificial Neural Networks, Part II*. Springer-Verlag, Heidelberg, pp 889–898
61. Grivas CR, Komar DA (2008) Kumho, Daubert, and the Nature of Scientific Inquiry: Implications for Forensic Anthropology. *J Forensic Sci* 53:771–776
62. Franklin D, Cardini A, Flavel A, Kuliukas A (2012) The application of traditional and geometric morphometric analyses for forensic quantification of sexual dimorphism: preliminary investigations in a Western Australian population. *Int J Legal Med* 126:549–558
63. Gama I, Navega D, Cunha E (2015) Sex estimation using the second cervical vertebra: a morphometric analysis in a documented Portuguese skeletal sample. *Int J Legal Med* 129:365–372
64. Masotti S, Succi-Leonelli E, Gualdi-Russo E (2013) Cremated human remains: is measurement of the lateral angle of the meatus acusticus internus a reliable method of sex determination? *Int J Legal Med* 127:1039–1044
65. Harrington KI, Wescott DJ (2015) Size and Shape Differences in the Distal Femur and Proximal Tibia between Normal Weight and Obese American Whites. *J Forensic Sci* 60:S32–S38
66. Albanese J (2013) A method for estimating sex using the clavicle, humerus, radius, and ulna. *J Forensic Sci* 58:1413–1419
67. Ousley D, Jantz RL (2012) Fordisc 3 and statistical methods for estimating sex and ancestry. In: Dirkmaat DC (ed) *A Companion to Forensic Anthropology*. Wiley-Blackwell, Malden
68. Navega D, Vicente R, Vieira DN, Ross AH, Cunha E (2015) Sex estimation from the tarsal bones in a Portuguese sample: a machine learning approach. *Int J Legal Med* 129:651–659
69. Nathena D, Gambaro L, Tzanakis N, Michalodimitrakis M, Kranioti EF (2015) Sexual dimorphism of the metacarpals in contemporary Cretans: Are there differences with mainland Greeks? *Forensic Sci Int* 257:515.e1–515.e2
70. Messer DL, Ousley SD, Tuamsuk P (2013) Pattern in worldwide craniometric sexual dimorphism and its importance in forensic anthropology. in: *Proceedings of the American Academy of Forensic Sciences XIX*, pp 394–395
71. Franklin D, Flavel A, Kuliukas A, Hart R, Marks MK (2012) The development of forensic anthropological standards in Western Australia. in: *Proceedings of the American Academy of Forensic Sciences XVIII*, pp 361–362
72. White TD, Folkens PA (2005) *The human bone manual*. Academic Press, San Diego
73. Rösing FW, Graw M, Marré B, Ritz-Timme S, Rotschild MA, Rotzschker K, Schmeling A, Schroder I, Geserick G (2007) Recommendations for the forensic diagnosis of sex and age from skeletons. *Homo* 58:75–89
74. Citardi MJ, Herrmann B, Hollenbeak CS, Stack BC, Cooper M, Bucholz RD (2001) Comparison of scientific calipers and computer-enabled CT review for the measurement of skull base and craniomaxillofacial dimensions. *Skull Base* 11:5–11
75. Verhoff MA, Ramsthaler F, Krähahn J, Deml U, Gille RJ, Grabherr S, Thali MJ, Kreutz K (2008) Digital forensic osteology—possibilities in cooperation with the Virtopsy project. *Forensic Sci Int* 174: 152–156
76. L'Abbé EN, Steyn M (2012) The Establishment and Advancement of Forensic Anthropology South Africa. In: Dirkmaat DC (ed) *A Companion to Forensic Anthropology*. Wiley-Blackwell, Malden, pp 626–638
77. Brůžek J, Murail P (2006) Methodology and reliability of sex determination from the skeleton. In: Schmitt A, Cunha E, Pinheiro J (eds) *Forensic Anthropology and Medicine: Complementary Sciences From Recovery to Cause of Death*. Humana Press Inc., Totowa, pp 225–242
78. Ramsthaler F, Kreutz K, Verhoff MA (2007) Accuracy of metric sex analysis of skeletal remains using Fordisc® based on a recent skull collection. *Int J Legal Med* 121:477–482
79. Urbanová P, Ross AH, Jurda M, Nogueira MI (2014) Testing the reliability of software tools in sex and ancestry estimation in a multi-ancestral Brazilian sample. *Leg Med* 16:264–273
80. Sierp I, Henneberg M (2015) Can ancestry be consistently determined from the skeleton? *Anthropolo Rev* 78:21–31
81. Albanese J, Tuck A, Gomes J, Cardoso HFV (2016) An alternative approach for estimating stature from long bones that is not population- or group-specific. *Forensic Sci Int* 259:59–68
82. Hatza AN, Ousley SD, Cabo LL (2015) Estimating stature when ancestry is unknown: What statistical methods work best? In: *Proceedings of the American Academy of Forensic Sciences XXI*, pp 152–153
83. Ruff CB, Holt BM, Niskanen M, Sladek V, Berner M, Garofalo E, Garvin HM, Hora M, Maijanen H, Niinimäki S, Salo K, Schuplerova E, Tompkins D (2012) Stature and body mass estimation from skeletal remains in the European Holocene. *Am J Phys Anthropol* 148:601–617