Abstract: The assessment of sex is crucial to the establishment of a biological profile of an unidentified skeletal individual. The best methods currently available for the sexual diagnosis of human skeletal remains generally rely on the presence of well-preserved pelvic bones, which is not always the case. Postcranial elements, including the femur, have been used to accurately estimate sex in skeletal remains from forensic and bioarchaeological settings. In this study, we present an approach to estimate sex using two measurements (femoral neck width [FNW] and femoral neck axis length [FNAL]) of the proximal femur. FNW and FNAL were obtained in a training sample (114 females and 138 males) from the Luis Lopes Collection (National History Museum of Lisbon). Logistic regression was used to develop a model to predict sex in unknown individuals. The logistic regression model correctly predicted sex in 85.3% to 85.7% of the cases. The model was also evaluated in a test sample (96 females and 96 males) from the Coimbra Identified Skeletal Collection (University of Coimbra), resulting in a sex allocation accuracy of 80.1% to 86.2%. This study supports the relative value of the proximal femur to estimate sex in skeletal remains, especially when other exceedingly dimorphic skeletal elements are not accessible for analysis.
ACKNOWLEDGMENTS

Fundação para a Ciência e Tecnologia (SFRH/BPD/74015/2010 [FC], SFRH/BPD/84268/2012 [DG], SFRH/BD/99676/2014 [DN]), Gerda Henkel Foundation (AZ 09/F/15 [MTF]), University of Coimbra (CC). We are grateful to Dr. Susana Garcia and Dr. Diana Carvalho for the access to the Luís Lopes Collection (National History Museum of Lisbon), and to Osteomics (http://osteomics.com/) for hosting the web-based app for sex estimation.

The authors state that they do not have any conflict of interest to declare.
Dear Editor of the Forensic Science International,

With this paper – “A METHOD FOR SEX ESTIMATION USING THE PROXIMAL FEMUR” – we aim to present a technique for sex estimation using two measurements of the proximal femur, as an alternative for integration in the forensic anthropologist toolkit, particularly when more dimorphic skeletal elements are not accessible for study. Results of this study support previous research that emphasized the value of the proximal femur to estimate sex in unidentified skeletal individuals and allow for the estimation of sex in a typical binary approach or, rather, in a probabilistic assessment – more appropriate within the reliability standards required by the Daubert criteria.

Yours sincerely,
A METHOD FOR SEX ESTIMATION USING THE PROXIMAL FEMUR

Francisco Curate⁠¹,²,³,a, João Coelho³,⁴; David Gonçalves¹,³,⁴,⁵; Catarina Coelho³,⁴; Maria Teresa Ferreira¹,³,⁴; David Navega³,⁴; Eugénia Cunha³,⁴

¹ Research Centre for Anthropology and Health, Department of Life Sciences, University of Coimbra, Coimbra, Portugal
² Interdisciplinary Center for Archaeology and Evolution of Human Behavior, University of Algarve, Faro, Portugal
³ Laboratory of Forensic Anthropology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal
⁴ Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal
⁵ Archaeosciences Laboratory, Directorate General for Cultural Heritage and LARC/CIBIO/InBIO, Lisboa, Portugal.

a Corresponding author at: CIAS – Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Apartado 3046, 3001-401 Coimbra, Portugal

E-mail addresses: fcurate@uc.pt, franciscocurate@gmail.com.
A METHOD FOR SEX ESTIMATION USING THE PROXIMAL FEMUR

ABSTRACT

The assessment of sex is crucial to the establishment of a biological profile of an unidentified skeletal individual. The best methods currently available for the sexual diagnosis of human skeletal remains generally rely on the presence of well-preserved pelvic bones, which is not always the case. Postcranial elements, including the femur, have been used to accurately estimate sex in skeletal remains from forensic and bioarcheological settings. In this study, we present an approach to estimate sex using two measurements (femoral neck width [FNW] and femoral neck axis length [FNAL]) of the proximal femur. FNW and FNAL were obtained in a training sample (114 females and 138 males) from the Luís Lopes Collection (National History Museum of Lisbon). Logistic regression was used to develop a model to predict sex in unknown individuals. The logistic regression model correctly predicted sex in 85.3% to 85.7% of the cases. The model was also evaluated in a test sample (96 females and 96 males) from the Coimbra Identified Skeletal Collection (University of Coimbra), resulting in a sex allocation accuracy of 80.1% to 86.2%. This study supports the relative value of the proximal femur to estimate sex in skeletal remains, especially when other exceedingly dimorphic skeletal elements are not accessible for analysis.

Keywords: forensic anthropology population data; forensic science, human identification, biological profile, sex diagnosis
INTRODUCTION

The estimation of sex is a fundamental component in the establishment of a biological profile and a critical step for the identification of skeletal remains in forensic contexts [1-3]. The pelvis is consensually regarded as the most reliable skeletal element for the attribution of sex in human remains [1-4]. Sexual dimorphism of the human pelvis is intimately associated with the selective forces of obstetrics and bipedal locomotion. Sexual selection also contributed to pelvic adaptative differences between sexes [2,5]. Although the skull has been traditionally considered the second best indicator of skeletal sex, recent research indicates that postcranial elements should be favored instead of the cranium for assessing sex when the pelvis is absent or fragmented [1].

The femur is the heaviest and strongest bone in the skeleton; as such, it is frequently recovered in forensic and archeological contexts [5,6]. It is also dimorphic within the same population [5], and very useful in sexing skeletal remains. Several dimensions of the femur, including femoral head diameter, femoral length, and bicondylar breadth have been utilized for the allocation of sex in unknown skeletal individuals [1, 7-13]. The head diameter is probably the single best femoral measurement for the attribution of sex [7], but previous studies have also demonstrated the capacity of other proximal femur dimensions, such as the femoral neck axis length (FNAL) or the femoral neck width (FNW), for sex [14-16] and ancestry attribution [15, 17]. Some geometric parameters of the proximal femur are associated with the risk of hip fractures [18-20], and sex differences in FNAL and FNW have long been noted in epidemiological studies [18-22]. Furthermore, the structural demands of the unrelated but sometimes conflictual functions of parturition and locomotion affected not only the pelvis but also the angle and length of the femoral neck [10].

The primary goal of this study is to create predictive models of sex based on two measurements of the proximal femur, the femoral neck axis length and the femoral neck width, in a Portuguese reference sample, that can be used as an alternative technique for sex estimation when other exceptionally dimorphic skeletal elements are not available for study. Another objective is to test the cross-sample reliability of the new sexing technique by evaluating the models in another Portuguese identified skeletal sample. The performance of the technique is also compared with the ones developed by Seidemann et al. [14] and Meeseun and colleagues [15], who also use proximal femur dimensions (FNW and FNAL, respectively) for sex estimation.
MATERIALS AND METHODS

Two samples from Portuguese reference skeletal collections were observed in this study [23,24]. A sample from the Luís Lopes Collection (LLC, National History Museum of Lisbon, Portugal) was used as a training assembly to fit the sex prediction models. The training set included 252 individuals (114 females and 138 males) with an age at death that ranged from 20 to 94 years old. All individuals died between 1891 and 1959, with the majority of deaths occurring between 1930 and 1959. One other sample from the Coimbra Identified Skeletal Collection (CISC, University of Coimbra, Portugal) was used to validate the predictive models created from the LLC assemblage. The test sample comprised 196 individuals (98 females and 98 males) with ages at death from 20 to 96 years. Dates of death spanned from 1910 to 1936.

Measurements of each individual included the femoral neck axis length and femoral neck width. FNAL was defined as the linear distance measured in the anterior plane from the base of the greater trochanter to the apex of the femoral head [20] (Figure 1). This measurement is occasionally mentioned in the biomedical literature as hip axis length [18]. FNW, also known as the supero-inferior femoral neck diameter, was typified as the narrowest distance across the femoral neck, perpendicular to the neck axis [14, 18] (Figure 1). All measurements were taken on the left femur with a digital caliper. A subgroup of 20 individuals was randomly selected to evaluate intra- and interobserver measurement error. Measurement error was assessed with the Technical Error of Measurement (TEM), the relative Technical Error of Measurement (rTEM), and the coefficient of reliability (R). TEM is an estimate of absolute precision, similar to the standard deviation of the magnitude of the error in the original measurement units (i.e., in mm). The coefficient of reliability represents the variance proportion exempt of measurement error [25,26].

Descriptive statistics, including group means, standard deviation (SD) and 95% confidence intervals (95% CI) for the mean were estimated for FNAL and FNW. Normal distribution of the variables was assessed through skewness and kurtosis [26]. Homoscedasticity was assessed with a Levene’s test. An independent samples t-test was used to evaluate the null hypothesis that FNAL and FNW means in males and females were equal. The models of statistical prediction of sex were created through logistic regression (LR), in order to find the most parsimonious models to describe the relationship between the outcome variable and the predictor variables. The logit model is mathematically stated as:
\[
L = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \cdots + \beta_p X_{pj} \quad (1),
\]

where \( L \) is the logit or the log-odd, \( \beta_0 \) is a constant, \( \beta_p \) are the regression coefficients and \( X_{pj} \) are the measurement values of the predictor variables. A negative logit value is associated with a female and a positive value with a male individual. The model can also be expressed to describe a probability, between 0 and 1:

\[
P(L) = \frac{1}{1 + e^{-L}} \quad (2),
\]

where \( L \) is the logit value computed from (1) and \( e \) is the Euler constant. For this study, the \( P(L) \) for a particular set of measurements estimates the probability of the individual being a male. The probability of an individual being a female is given by \( 1 - P(L) \).

Sensitivity and specificity, McFadden pseudo-R\(^2\) (\( R_{MF}^2 \)), and Area Under the Curve (AUC) were calculated to assess the goodness of fit of the models [28]. Sectioning points for each variable were calculated according to Spradley and Jantz [1].

Statistical analyses and graphical depictions were accomplished with IBM® SPSS® (version 21.0) and R programming language [29,30].

RESULTS

Measurement error is summarized in Table 1. Results indicate that FNAL and FNW were executed within proper levels of measurement error, being thus repeatable and reproducible.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>TEM</th>
<th>rTEM</th>
<th>R</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>FNAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intraobserver</td>
<td>0.22</td>
<td>0.24%</td>
<td>1.00</td>
<td>20</td>
</tr>
<tr>
<td>Interobserver</td>
<td>0.43</td>
<td>0.47%</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>FNW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intraobserver</td>
<td>0.39</td>
<td>1.21%</td>
<td>0.99</td>
<td>20</td>
</tr>
<tr>
<td>Interobserver</td>
<td>0.49</td>
<td>1.53%</td>
<td>0.98</td>
<td></td>
</tr>
</tbody>
</table>

Descriptive statistics for both the training and testing samples are summarized in Tables 2 and 3. Both FNAL and FNW were statistically different between sexes in the training sample (FNAL: \( t = -16.265 \); \( df = 244 \); \( p < 0.001 \) / FNW: \( t_{corrected} = -15.831 \); \( df = 249.204 \); \( p < 0.001 \)). The density distributions of FNAL and FNW are depicted in Figures 2 and 3.
Table 2: Descriptive statistics for FNAL and FNW in both sexes, Luis Lopes Collection.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>95% CI</th>
<th>N</th>
<th>Sectioning point</th>
</tr>
</thead>
<tbody>
<tr>
<td>♂ FNAL</td>
<td>86.39</td>
<td>4.65</td>
<td>85.53 – 87.26</td>
<td>114</td>
<td>91.29</td>
</tr>
<tr>
<td>♂ FNW</td>
<td>29.43</td>
<td>2.10</td>
<td>29.04 – 29.82</td>
<td>114</td>
<td>31.87</td>
</tr>
<tr>
<td>♀ FNAL</td>
<td>96.19</td>
<td>4.85</td>
<td>95.37 – 97.00</td>
<td>138</td>
<td>97.00</td>
</tr>
<tr>
<td>♀ FNW</td>
<td>34.31</td>
<td>2.69</td>
<td>33.86 – 34.77</td>
<td>138</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Descriptive statistics for FNAL and FNW in both sexes, Coimbra Identified Skeletal Collection.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>95% CI</th>
<th>N</th>
<th>Sectioning point</th>
</tr>
</thead>
<tbody>
<tr>
<td>♂ FNAL</td>
<td>87.91</td>
<td>6.22</td>
<td>86.66 – 89.16</td>
<td>98</td>
<td>92.70</td>
</tr>
<tr>
<td>♂ FNW</td>
<td>30.06</td>
<td>2.21</td>
<td>29.62 – 30.51</td>
<td>98</td>
<td>32.46</td>
</tr>
<tr>
<td>♀ FNAL</td>
<td>97.48</td>
<td>4.80</td>
<td>96.52 – 98.44</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>♀ FNW</td>
<td>34.85</td>
<td>2.37</td>
<td>34.38 – 35.33</td>
<td>98</td>
<td></td>
</tr>
</tbody>
</table>

The LR models fitting are epitomized in Table 4. A model using only FNAL correctly classified the sex of 85.3% of all individuals (sensitivity: 87.7%; specificity: 82.5%), providing an effective discriminant capacity (AUC = 0.923; $R^2_{MF} = 0.496$). This model is described by the following equation (females classified with negative values, whereas males are classified with positive values):

$$\text{Sex} = -37.156 + 0.410 \times \text{FNAL} \ (3).$$

The second LR model, with FNW as the only predictor variable, correctly allocated the sex of 85.3% of all individuals (sensitivity: 87.7%; specificity: 82.5%), with a good discriminant capability (AUC = 0.932; $R^2_{MF} = 0.525$). This model is depicted by the ensuing equation (females classified with negative values, whereas males are classified with positive values):

$$\text{Sex} = -30.445 + 0.968 \times \text{FNW} \ (4).$$

A model including both FNAL and FNW correctly predicted the sex of 85.7% of all individuals, with nearly equivalent sensitivity (86.2%) and specificity (85.1%). This step presented an excellent discriminative capacity (AUC = 0.959; $R^2_{MF} = 0.630$). For this model, the following LR equation is applicable (females classified with negative values, whereas males are classified with positive values):

$$\text{Sex} = -48.587 + 0.279 \times \text{FNAL} + 0.737 \times \text{FNW} \ (5).$$
Table 4: Logistic regression models fitting.

<table>
<thead>
<tr>
<th></th>
<th>Variable</th>
<th>β</th>
<th>SE</th>
<th>Wald</th>
<th>Sig.</th>
<th>Exp (β)</th>
<th>95% CI for Exp (β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>FNAL</td>
<td>0.410</td>
<td>0.049</td>
<td>69.812</td>
<td>&lt;0.001</td>
<td>1.506</td>
<td>1.368 – 1.658</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>–37.156</td>
<td>4.463</td>
<td>69.314</td>
<td>&lt;0.001</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>FNW</td>
<td>0.968</td>
<td>0.123</td>
<td>62.230</td>
<td>&lt;0.001</td>
<td>2.632</td>
<td>2.069 – 3.347</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>–30.445</td>
<td>3.866</td>
<td>62.015</td>
<td>&lt;0.001</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>FNAL</td>
<td>0.279</td>
<td>0.053</td>
<td>27.241</td>
<td>&lt;0.001</td>
<td>1.321</td>
<td>1.190 – 1.467</td>
</tr>
<tr>
<td></td>
<td>FNW</td>
<td>0.737</td>
<td>0.138</td>
<td>28.494</td>
<td>&lt;0.001</td>
<td>2.090</td>
<td>1.594 – 2.739</td>
</tr>
<tr>
<td></td>
<td>Constant</td>
<td>–48.587</td>
<td>6.444</td>
<td>56.852</td>
<td>&lt;0.001</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

In order to simplify the calculations, an online app that estimates sex and posterior probabilities from FNAL and FNW measurements is available (apps.osteomics.com/SeuPF).

In the test sample (CISC), sex was correctly assessed in 80.1%, 82.1% and 86.2% of the cases, for Model 1 (predictor variable: FNAL), Model 2 (predictor variable: FNW), and Model 3 (predictor variables: FNAL and FNW), respectively. The first model correctly identified 67.3% of females and 92.9% of males, the second model correctly classified 72.4% of females and 91.8% of males, and the third model correctly assigned 75.5% of females and 96.9% of males.

Seidemann et al. [14] developed predictive models for the estimation of sex using FNW. The linear discrimination function for an American White sample (Hamann-Todd skeletal collection) applied in the testing sample yielded a sex distribution accuracy of 70.4%, with 43.8% of females and 98.0% of males properly assigned. The LR equation proposed by Meuseen et al. [15], fitted after a pooled sample of Native Americans (Averbuch Site Skeletal Collection), and American Blacks and Whites (Robert J. Terry Anatomical Skeletal Collection and William M. Bass Donated Skeletal Collection), with FNAL as the only independent variable, was also tested in the CISC sample. The overall sex allocation accuracy for this equation was 80.1%, with 72.4% of females and 87.8% of males correctly identified.

**DISCUSSION**

The results of this investigation suggest that sex estimation with measurements of the proximal femur is fairly accurate and valid across populations, in agreement with previously published studies [1, 7-16, 31-33].
The multivariable LR model (i.e., Model 3, with FNAL and FNW as predictor variables) shows the highest accuracy in the assessment of sex in skeletal remains, both in the training and testing samples. Interestingly, the model slightly improves its performance in the CISC sample, but the percentage of correctly allocated females is lower in the testing sample. The cross-sample classification percentage is similar to the accuracy provided by univariate sectioning points of femoral dimensions, but lower than the accuracy obtained with the multiple variable classification function of the femur, reported by Spradley and Jantz [1]. Notwithstanding, the classification percentages were obtained in a holdout sample resulting from the same skeletal collection, the Forensic Data Bank. Sex allocation accuracy for other methods that use proximal femur dimensions spans from 77.7% to 95.0% [10, 13-16, 31-33], with the majority of methods showing an accuracy in the 80–88% range.

Model 2 (FNW as the only predictor variable) correctly predicted the sex in 82.1% of individuals from the testing sample. This model is especially useful when other highly dimorphic characteristics of the proximal femur, such as the head diameter, are not available. The femoral neck is frequently very well preserved, unlike the head [14], which makes this equation especially useful in fragmentary and/or incomplete remains. The linear model by Seidemann et al. [14], also using FNW as the only predictor variable, performed much worse, particularly with reference to the extensive misclassification of females (>50%).

The LR equation that uses only FNAL as a predictor variable correctly classified 80.1% of individuals in the testing sample, precisely the same classification percentage as that obtained with the Meeusen et al. [15] equation, supporting the usefulness of the sexing methods with the proximal femur across independent populations, in spite of the ancestry differences in this parameter [15, 17]. Also, sectioning points for this variable are very similar across samples. As the tests using the CISC sample suggest, the methods perform almost as well [with FNAL or FNW only] or even better [with FNAL and FNW] when applied to a different sample.

The LR models are slightly biased towards the correct estimation of sex in males, a pattern commonly reported in other studies [8, 15, 31,32,34] – although not always [9]. In the testing sample, this bias was more pronounced, especially when only one of the measurements from the proximal femur was used as a predictor variable. As suggested by other researchers [34-36], sex-specific accuracy possibly relates with secular change in bone dimensions, usually associated with a higher misclassification of females when employing a method fitted in a chronologically older sample that has,
in comparison, been affected by a positive secular trend. FNAL, at least, is known to display secular change in both sexes, with a more pronounced increase in women [37]. Another study in a pooled sample of Portuguese reference collections (including LLC and CISC) found an inverse trend in women, with a weak negative association between FNAL and year of birth [38]. It was not possible to fully determine the influence of secular trend in our results. However, it is important to note that the Luis Lopes Collection, whose sample was used to develop the sexing technique, and the Coimbra Identified Skeletal Collection, used to test it, show some definite similarities. All individuals were Portuguese nationals (with the birthplace in various regions of the country) and most had a low socioeconomic status. Also, the samples considerably overlap in chronological terms, even if LLC is on average more recent. However, there are also slight differences between the samples and the collections, namely in the sexual composition (females are overrepresented in the LLC) and the mortality pattern [23, 24]. The first issue was solved by using the arithmetic mean instead of the weighted mean as a sectioning point. The second issue may indeed have led to some differences between the two samples due to secular changes but it is difficult to say if those can explain the sex-biased correct estimation of sex in the test sample.

CONCLUSIONS

Postcranial sex assessment typically depends on metric data, being less subjective than the visual evaluation of morphological traits and contributing to increased evidentiary standards [1]. The results of this study support previous research that highlighted the value of the proximal femur to estimate sex in unidentified skeletal individuals when other highly dimorphic skeletal elements, such as the pelvis or complete long bones, are not available for study.

Differences between sexes in FNAL and FNW are significant, and sex was correctly predicted in 85.3% to 85.7% of the cases. The model reliably estimated sex in an independent reference skeletal sample (accuracy percentages between 80.1% and 86.2%). Also, this method can be used to estimate sex in a classical binary approach (male or female) or, preferably, in a probabilistic assessment that is more appropriate within the framework of the “Daubert guidelines” [2]. The proposed model must endure further verification in independent skeletal material to validate its reliability in both forensic and bioarcheological contexts, especially because reliable classification may depend on the sex of the individual.
REFERENCES


Figure 1: FNAL (A-B), and FNW (C-D)
Click here to download high resolution image
Figure 2: Density distribution of FNAL (mm) by sex (LLC sample)
Click here to download high resolution image
Figure 3: Density distribution of FNW (mm) by sex (LLC sample)