Accuracy and reliability of Southern European standards for the tibia
A test of two Mediterranean populations

Citation for published version:
García-Donas, JG, Ekizoglu, O, Er, A, Bozdag, M, Akcaoglu, M, Can, IO & Kranioti, E 2017, 'Accuracy and reliability of Southern European standards for the tibia: A test of two Mediterranean populations', Journal of Forensic Science and Criminology, vol. 2, no. 1, pp. 1-5. https://doi.org/10.15761/FSC.1000107

Digital Object Identifier (DOI):
10.15761/FSC.1000107

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Journal of Forensic Science and Criminology

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Accuracy and reliability of Southern European standards for the tibia: a test of two Mediterranean populations

Julieta G. García-Donas1,*, Oguzhan Ekizoglu2, Ali Er3, Mustafa Bozdag3, Mustafa Akcaoglu3, Ismail Ozgur Can4 and Elena F. Kranioti1,5,6
1Edinburgh Unit for Forensic Anthropology, School of History, Classics and Archaeology, University of Edinburgh, 4 Teviot Place, EH8 9AG, UK
2Department of Forensic Medicine, Tepecik Training and Research Hospital, Izmir, Turkey
3Department of Radiology, Tepecik Training and Research Hospital, Izmir, Turkey
4Department of Forensic Medicine, Dokuz Eylül University, Faculty of Medicine, Izmir, Turkey
5Forensic Pathology Division Crete, Hellenic Republic Ministry of Justice, Transparency and Human Rights, Heraklion, Crete, Greece
6Department of Medical Imaging, University Hospital of Heraklion, Heraklion, Crete 71110, Greece

Abstract
Sexual dimorphic variation between populations must be taken into consideration when applying existing methods on unrelated samples. Validation studies are extremely important to avoid misclassification and ensure high quality standards. This paper presents a test of a Southern European metric method on Greek-Cypriots (N=132) and Turkish (N=203). Three tibia measurements were taken, sex differences were explored using a Wilcoxon test and the parameters were applied to the original discriminant functions. The results showed accuracy rates ranging from 79 to 86% for Greek-Cypriots and from 80 to 88% for Turkish. Differences in the performance of the formulae applied were observed between the samples. Correct classification rates are very similar to the ones reported by the original method. This study demonstrates that the application of the Southern European method to estimate sex on these two Mediterranean populations is reliable. A larger and more diverse sample is required to verify our results.

Correspondence to: Elena F. Kranioti, Edinburgh Unit for Forensic Anthropology, School of History, Classics and Archaeology, University of Edinburgh, William Robertson Wing, Old Medical School, Teviot Place, Edinburgh, EH8 9AG, Tel. +44 (0)131 650 2368, Fax. +44 (0)131 650 2378, E-mail: elena.kranioti@ed.ac.uk

Key words: Tibia, sex estimation, Greek-Cypriot, Turkish, discriminant function analysis

Received: January 06, 2017; Accepted: January 19, 2017; Published: January 23, 2017
were selected at random from a cemetery population housed in the ossuary of the main cemetery in the city of Limassol in Cyprus. The sample consisted of individuals who died between 1976 and 2003. The mean age for males was 69.3 ± 12 years and 70 ± 17.8 years for females.

Turkish sample

Due to religious prohibitions there are no osteological collections available in Turkey; therefore biometric standards for this ethnic group mainly derive from examinations through medical imaging modalities. Two hundred and three CT scans of patients admitted to the Tepecik Training and Research Hospital in Izmir, Turkey, were used in this study. Patients with injuries, previous surgery, congenital or an acquired anomaly in the tibia were excluded from the study. The mean age for males (N=124) was 59.8 ± 12.2 years and 60.2 ± 14.5 years for females (N=79). CT scans were performed by a 64-slice CT scanner (Siemens Medical Solutions, Erlangen, Germany). A routine peripheral angiography multi-detector row computed tomography (MDCT) protocol was followed. The scanning parameters included 80 kV, 115 mAs, a slice thickness of 1mm and 512 x 512 matrix.

Three measurements [7,23] were taken from the left tibia: Maximum length (ML), Upper epiphyseal breadth (UB) and Lower epiphyseal breadth (LB).

Inter- and Intra-observer error was calculated in a random sample of 30 dry bones (Cypriot sample) and 30 virtual reconstructions (Turkish sample). For error quantification Technical measurement error (TEM), relative TEM (rTEM) and the coefficient of reliability (R) of the measurements were calculated as suggested [24].

Sex differences of the measurements were explored using a one-way ANOVA. Validation of published formulae for Southern Europeans (equations F1-F4) [25] were tested on this sample using one-way ANOVA. Test of the formulae for Southern Europeans

Univariate and multivariate equations of the three variables of the tibia were published by Kranioti and Apostol [25] in a pooled Southern European sample consisting of populations from Spain, Italy, and Greece. We tested these formulae in our two samples. The accuracy for Greek-Cypriots ranged from 78 to 84% and for Turkish from 71 to 87%.

Table 1. Inter and Intra-observer error for both osteometric and virtual measurements.

|                | Intra-Observer Error | Inter-Observer Error |
|----------------|----------------------|----------------------|
|                | TEM      | rTEM    | R   | TEM      | rTEM    | R   |
| Cypriot sample |          |         |      |          |         |      |
| ML             | 0.56     | 0.16    | 0.99 | 1.88     | 0.51    | 0.96 |
| UB             | 0.45     | 0.66    | 0.99 | 1.30     | 1.89    | 0.70 |
| LB             | 0.74     | 1.50    | 0.94 | 0.95     | 1.93    | 0.86 |
| Turkish sample |          |         |      |          |         |      |
| ML             | 1.49     | 0.42    | 1.00 | 2.41     | 0.69    | 0.99 |
| UB             | 1.51     | 2.07    | 0.89 | 2.36     | 3.22    | 0.73 |
| LB             | 0.58     | 1.15    | 0.97 | 1.08     | 2.14    | 0.90 |

Table 2. Descriptive statistics and Wilcoxon test results for mean differences between the sexes for the Greek-Cypriots and Turkish samples.

|                | Greek-Cypriots | Turkish |
|----------------|----------------|---------|
|                | Males           | Females | Males           | Females |
|                | N   | Mean | SD   | N   | Mean | SD   | N   | Mean | SD   |
| ML             | 67  | 374.2 | 21.3 | 58  | 341.6 | 21.65 | 2274.5 | 0.001 |
| UB             | 63  | 74.17 | 4.07 | 59  | 67.59 | 5.02 | 2379.0 | 0.001 |
| LB             | 66  | 44.7  | 2.92 | 58  | 39.68 | 3.01 | 2242.0 | 0.001 |
|                |          |       |      |     |       |      |       |      |
| ML             | 124 | 364.5 | 21.1 | 79  | 335.4 | 19.97 | 4717.0 | 0.001 |
| UB             | 124 | 77.39 | 3.89 | 79  | 69.08 | 3.752 | 3800.0 | 0.001 |
| LB             | 124 | 53.12 | 3.15 | 79  | 47.28 | 2.721 | 3890.5 | 0.001 |

Results

Intra-observer error

For the Greek-Cypriot skeletal sample, 30 randomly selected tibiae were measured by the same observer within 4 weeks of the first measurement. TEM, rTEM and R for each variable are presented in Table 1. rTEM was below 5% in all cases while R was consistently over 0.95 with the exception of tLB, which was slightly lower. This is in accordance with the acceptable human error (rTEM<5%, R>0.95) as suggested by Ulijaszek and Kerr [24]. For the Turkish sample, 30 randomly selected tibiae were measured following the same protocol as for the Greek-Cypriot sample. The results showed that rTEM is consistently below 5%, while for R only UB fell out of the limit of acceptance.

Inter-observer error

Inter-observer error was also quantified for both osteometric and digital data. TEM, rTEM and R for each variable are presented in Table 1. rTEM was below 5% in all cases while R ranged between 0.70 and 0.99. Interestingly the lowest value for R was noted for UB for both data acquisition modalities.

Sexual dimorphism

A Shapiro-Wilk’s test (p<0.05) and a visual inspection of the histograms, Q-Q plots and box plots, were used to assess normal distribution. On some occasions data was not normally distributed for females in both sexes. Wilcoxon test confirmed the mean differences between the two sexes (p<0.001) for all variables in both samples (Table 2).
The accuracy of the multivariate formulae are very close to the cross-validated accuracy reported in the original study (83-88%) [25]. Sex bias was noted in the multivariate formulae for both samples. F3 misclassified 38% of the Turkish females while F4 misclassified 24% of the Turkish females and 30% of the Greek-Cypriot males. Sex bias for F3 and F4 are notably higher compared to the ones from the original study [25] as seen in Table 3.

### Discussion

Christensen and Crowder [22] stressed the necessity of testing and re-evaluating existing forensic anthropological methods both for routine practise and for court cases. They stated that even if a methodology does produce high error rates, there is no need to exclude it or omit reporting the unreliability; instead, the fundamental concern is to be aware of the errors and to use the proper approach to measure them. One of the means for testing reliability and applicability of anthropological methods is the performance of validation studies. Sex estimation methods were tested previously due to rising methodological concerns, like the importance of the observer experience to ensure accurate estimates, or anthropological issues biasing the results like ancestry differences between sexes [26,27].

In this study, we tested the accuracy of the Kranioti and Apostol formulae [25] on threeibia measurements taken from two populations (Greek-Cypriots and Turkish) to verify whether the method can be reliably applied to those two samples. The results showed that correct classification achieved in both Greek-Cypriots and Turkish populations is very similar to the one reported by the original study for cross-validated data. However, differences in the performance of the four original formulae applied to the samples were observed. For Greek-Cypriots, the highest accuracy and lowest sex bias was achieved by the Formula including length and lower breadth measurements (F3). Actually, the rates of correct classification are 4% higher for the Greek-Cypriot sample than for the original Southern European sample. Concerning the Turkish sample, the most accurate formula was the one including all the measurements (F1) although it performed better for males than for females. The poorest correct classification was achieved by upper and lower breadth formula for Greek-Cypriots, and length and lower breadth formula for Turkish. It is worth noting that the correct classification accuracies are still not lower than 80% as reported by Kranioti and Apostol [25]. Regarding sex bias, one of the equations for each sample reached as high as 20% of misclassification; a slightly similar sex bias value was also reported in the original study when only lower breadth was used for sex estimation. Caution should be taken before using these functions in order to avoid misclassification. It can be concluded that the error rates reported by the Southern Europeans method are comparable to the ones observed in this study. A recent study undertaken by Kranioti et al. [28] used seven measurements from the tibia to develop population specific standards for Greeks and Greek-Cypriots achieving correct classification rates ranging from 78 to 87% for the latter group using univariate and multivariate functions. Interestingly, the highest accuracy rate was obtained by a formula that also included tibia lower breadth, a parameter that was also included in the best function for the validation study. Although a larger sample might verify our results, it could be suggested that the Southern European formulae F1-F3 are accurate on this population, and therefore, can be applied for sex estimation.

On the other hand, the Turkish sample was previously used to develop its own population specific standards using the tibia with a range of 66 to 86% of correct classification rates for univariate and multivariate approaches [29]. The highest correct classification for single sex indicators was achieved by upper breadth, and by length, and upper and lower breadth for multivariate functions (86% classification accuracy for both formulae). When the Turkish sample was tested against the Southern European method in our study, the best performance was achieved by the same three measurements confirming their power for discriminating between sexes. Although the classification accuracies are very close, it seems that the Southern European formula (F1) performs slightly better than the original Turkish formula that includes the same parameters.

Gulhan et al. [30] sexual dimorphism study performed on CT Turkish femora reported accuracy rates ranging from 63 to 91% for univariate analysis and multivariate discriminant function analysis. Moreover, intra-observer error was within the limits of acceptance, which is in accordance with our results. This Turkish population was tested against standards from other populations reporting overall high misclassification rates [30]. By contrast, our results suggest that for our Turkish sample the Southern European sex estimation method is reliable. Further tests using a larger and more diverse sample will provide insight into different sexual dimorphism degrees of expression for various geographical areas within the same country. CT-scans have been used successfully for sex estimation –both morphological and metric analysis– on other bones [31,32] and for disaster victim identification [33] demonstrating their value as a forensic tool.

The percentages of classification accuracy obtained in our study are comparable with other studies using metric analysis on the tibia for other populations [34,35]. An anthropometric investigation comparing cranial and postcranial elements for sex estimation revealed that postcranial skeleton performs better than cranial when using multivariate analysis [36]. Skull measurements were applied on a validation study conducted by Ramsthaler et al. [37]. The authors tested USA discriminant functions on a German sample to compare the use of Fordisc data versus morphological assessment. It was recommended the application of both methods due to low average accuracy obtained from the metric approach. Low classification accuracy was also noted in another study [38] when standards produced for Portuguese, Southern

### Table 3. Accuracy of sex estimation for the test sample in comparison with the original study [25].

|                | Greek-Cypriots | Turkish | Total (N=203) | sex bias | Southern Europeans-original sample |
|----------------|----------------|---------|---------------|----------|------------------------------------|
|                | Males (N=77)  | Females (N=59) | Total (N=126) | sex bias | Males (N=212) Females (N=232) Total (N=444) | sex bias |
| ML             | 92.5          | 60.3    | 77.6          | 32.2     | 78.2                                | 73.4         | 76.4         | 4.8        | 76.5      | 81.2      | 79        | -4.7       |
| UB             | 69.8          | 89.3    | 79.7          | -20.0    | 89.5                                | 82.3         | 86.7         | 7.2        | 83.3      | 87.1      | 85.3      | -3.8       |
| LB             | 67.0          | 93.1    | 68.5          | -46.1    | 99.2                                | 24.1         | 70.0         | 75.1       | 64.3      | 78.1      | 71.5      | -13.8      |
| F1             | 84.7          | 83.9    | 84.3          | 0.8      | 91.2                                | 81.0         | 87.7         | -6.7       | 87.0      | 88.5      | 87.8      | -1.5       |
| F2             | 86.7          | 83.9    | 85.3          | 2.7      | 89.5                                | 83.5         | 87.2         | -3.7       | 87.5      | 85.7      | 86.6      | 1.9        |
| F3             | 86.2          | 86.4    | 86.3          | 0.3      | 92.7                                | 62.0         | 80.8         | -18.8      | 78.8      | 86.6      | 82.8      | -7.8       |
| F4             | 69.5          | 89.3    | 79.1          | -19.8    | 91.1                                | 75.9         | 85.2         | -9.3       | 84.1      | 87.6      | 85.9      | -3.5       |
European and North Americans were applied to a Czech sample. This study reported high misclassification of females and up to 100% sex bias. In fact, it should be expected to have low classification rates when standards are applied to such diverse populations. In contrast, our validation study includes populations that share similar Mediterranean diet, climate and to some extent genetic markers with the reference populations. For example, genetic studies showed that the DNA in Cypriots contains 23% Greek markers and 20% Italian markers while Turkish DNA contains 11% of Greek and 7% West Sicilian genetic markers [39]. This could explain the higher classification accuracy/ lower sex bias reported here. In addition, the sampling effect that can bias classification accuracy of an unknown sample is diminished as the reference sample increases and thus captures a higher percentage of morphological variability.

Conclusions

Past and recent conflicts provoked either the death of persons due to violent confrontations or dangerous forced migrations. The identification of missing persons from the Greek-Cypriots and Turkish due to the Turkish invasion of Cyprus and the 20th century Turkish conflict, respectively, is still a challenge [21,40]. In summary, sex can be estimated accurately on these two Mediterranean samples using the standards developed by Kranioti and Apostol [25]. Most of the formulae tested here produced similar discriminatory power. The accuracy rates were over 80% even if the remains were fragmented and only two variables could be used. The forensic community needs either to create new standards for assisting in the identification of specific groups or to validate the existing methodologies to ensure more accurate results. This validation study is one step further in this process of standardization and re-evaluation of forensic anthropology methods.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

The authors would like to thank X.P. Kyriakou and the Orthodox Church in Limassol (Cyprus) for permitting access to the skeletal material of the ossuary and C.E. Lill for the linguistic review and editing.

References

1. İçcan M, Loth S, Wright R (1984a) Age Estimation from the Rib by Phase Analysis: White Females. J Forensic Sci 30: 853-863. [Crossref]
2. İçcan M, Loth S, Wright R (1984b) Age Estimation from the Rib by Phase Analysis: White Males. J Forensic Sci 29: 1094-1104. [Crossref]
3. Todd TW (1920) Age changes in the pubic bone: I. The white male pubis. Am J Phys Anthropol 3: 467-470.
4. Phenicke TW (1969) A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 30: 297-301.
5. Rogers T, Saunders S (1994) Accuracy of Sex Determination Using Morphological Traits of the Human Pelvis. J Forensic Sci 39: 1047-1056. [Crossref]
6. Walker PL (2008) Sexing skulls using discriminant function analysis of visual assessed traits. Am J Phys Anthropol 136:39-50. [Crossref]
7. İçcan MV, Miller-Shaivitz P (1998) Sexual dimorphism in the femur and tibia. In: Reichs KJ (ed.) Forensic osteology: advances in the identification of human remains, Springfield: Charles C Thomas.
8. Albanese J (2013) A method for estimating sex using the clavicle, humerus, radius, and ulna. J Forensic Sci 58:1413-1419. [Crossref]
9. Pietruseswky M (2008) Metric analysis of skeletal remains: methods and applications (3rd ed) In Katzenberg MA, Saunders SR (eds.) Biological anthropology of the human skeleton, New York: Wiley-Liss.
10. Kralik M, Urbanova P, Wagenknechtova M (2013) Sex assessment using clavicle measurements: Inter- and intra-population comparisons. Forensic Sci Int 234: 181.e1-181.e5. [Crossref]
11. İçcan MY (2005) Forensic anthropology of sex and body size. Forensic Sci Int 147: 107-112.
12. Song HW, Lin ZQ, Jia JT (1992) Sex diagnosis of Chinese skulls using multiple stepwise discriminant function analysis. Forensic Sci Int 54: 135-40. [Crossref]
13. Saini V, Srivastava R, Rai RK, Shamal SD, Singh TB, et al. (2011) An Osteometric Study of Northern Indian Populations for Sexual Dimorphism in Craniofacial Region. J Forensic SciChb: 700-705. [Crossref]
14. Kranioti EF, Michalodimitrakis M (2009) Sexual dimorphism of the humerus in Contemporary Cretans-A population specific study and a review of the literature. J Forensic Sci 54:996-1000. [Crossref]
15. Asala SA (2003) Sex determination from the head of the femur of South African whites and blacks. Forensic Sci Int 117:15-22. [Crossref]
16. 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.
17. Dangar KP, Pandya AM, Rashid SP, Tank KC, Akbari VI, et al. (2012) Sexual dimorphism of proximal epiphyseal breadth of tibia. Int J Biol Med Res 1331-1334.
18. Holland TD (1991) Sex assessment using the proximal tibia. Am J Phys Anthropol 85: 221-227.
19. Seidman RM, Stojanoski CM, Doran GH (1998) The use of the supero-inferior femoral neck diameter as a sex assessor. Am J Phys Anthropol 107: 305-13. [Crossref]
20. 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-e5. [Crossref]
21. http://www.cmp-cyprus.org/. (Accessed January 2017).
22. Christensen AM, Crowder CM (2009) Evidentiary Standards for Forensic Anthropology. J Forensic Sci 54: 1211-1216. [Crossref]
23. Krogman WM, İçcan MY (1986) Human skeleton in forensic medicine, USA: Charles C. Thomas.
24. Ulijaszek SA, Kerr DA (1999) Anthropometric measurement error and the assessment of nutritional status. Br J Nutr 82: 165-77. [Crossref]
25. Kranioti EF, Apostol M (2015) Sexual dimorphism of the tibia in contemporary Greeks, Italians and Spanish: forensic implications. Int J Legal Med 36: 357-63. [Crossref]
26. Ubelaker D, Volk C (2002) Test of the Phenice Method for the Estimation of Sex. J Forensic Sci 47: 19-24. [Crossref]
27. Robling A, Ubelaker D (1997) Sex Estimation from the Metatarsals. J Forensic Sci 42:1062-9. [Crossref]
28. Kranioti EK, Garcia-Donas JG, Prado PSA, Langstaff HC (2016) Sexual Dimorphism of the Tibia in Contemporary Greek-Cypriots and Cretans: Forensic Applications. Forensic Sci Int. [Epub ahead of print][Crossref]
29. Ekizoglu O, Ali Er, Bozdag M, Akcaoğlu M, Ozgur Can I, et al. (2016) Sex estimation of the tibia in modern Turkish: A computed tomography study. Leg Med23:89-94. [Crossref]
30. 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.
31. Zhang K, Cui JH, Luo YZ, Fan F, Yang M, Li XH, Zhang W, Deng ZH (2016) Estimation of stature and sex from scapular measurements by three-dimensional volume-rendering technique using in Chinese. Leg Med 21: 58-63. [Crossref]
32. Biwaskha H, Aoki Y, Sato K, Tanjirri T, Fujita S, et al. (2012) Analysis of sexual dimorphism of reconstructed pelvic computed tomography images of contemporary Japanese using curvature of the greater sciatic notch, pubic arch and greater pelvis. Forensic Sci Int 219: 288.e1-e8. [Crossref]
33. Blau S, Robertson S, Johnstone M (2008) Disaster victim identification: new applications for postmortem computed tomography. *J Forensic Sci* 53: 956-61. [Crossref]

34. 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. [Crossref]

35. Šlaus M, Bediča Z, Strinović D, Petrovečki V (2013) Sex determination by discriminant function analysis of the tibia for contemporary Croats. *Forensic Sci Int* 226: 302.e1-e4. [Crossref]

36. Spradley MK, Jantz RL (2011) Sex estimation in forensic anthropology: skull versus postcranial elements. *J Forensic Sci* 56: 289-96.

37. Ramsthaler F, Kretz K, Verhoff M (2007) Accuracy of metric sex analysis of skeletal remains using Fordisc® based on a recent skull collection. *International Journal of Legal Medicine* 121: 477-82.

38. Kotěrová A, Velenimanská J, Dupej J, Brzobohatá H, Plňák A, et al. (2016) Disregarding Population Specificity: Its Influence on the Sex Assessment Methods From the Tibia. *Int J Legal Med* 131: 251-61.

39. Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falah D, Myers S (2014) Genetic atlas of human admixture history. *Science* 6172: 747-51.

40. www.thenational.ae/arts-lifestyle/the-review/the-long-read-where-are-turkeys-missing (Accessed January 2017)