Heli Maijanen

TESTING OF PUBLISHED SEX ESTIMATION STANDARDS FOR FEMORAL AND HUMERAL HEAD DIAMETER IN FINNISH SKELETAL SAMPLES

Abstract

This is a pilot study for testing sex estimation metric standards in Finnish skeletal material. The focus of the study is to apply published standards of femoral and humeral head diameters to Finnish skeletal samples to investigate their applicability. These measurements are commonly used to support sex estimation but there are no standards based on Finnish population, and no actual testing of how standards from other populations perform in Finnish samples. Four standards of both diameters were tested on three samples including one anatomical sample from Helsinki, and two archaeological samples from Porvoo and Renko. The rates of correctly classified individuals were quite variable (60–100 %) depending on the standard, sex, and sample. The paper discusses the potential explanations for the differences in the classification accuracies, and makes some recommendations on procedures.

Keywords: Sex estimation, metrics, femur, humerus, classification rates, Finland

Heli Maijanen, Archaeology, P.O. Box 1000, 90014 University of Oulu, Finland: heli.maijanen@oulu.fi

Received: 30 April 2020; Revised: 9 February 2021; Accepted: 3 June 2021

INTRODUCTION

Sex estimation is one of the most foundational aspects of osteological analysis of human remains (see Ferembach et al. 1980 for more details on sex estimation). The most accurate and reliable method of estimating sex is from the morphology of the pelvis. However, it is quite common that the os coxa and especially the most sexually dimorphic portion, the pubic bone, is not fully preserved in archaeological contexts (Waldron 1987; Walker 2005). The greater sciatic notch usually preserves better, but it is not as good an indicator of sex as the pubic bone (Rogers & Saunders 1994).

Cranial morphology is also commonly used in sex estimation (Buikstra & Ubelaker 1994), but it encompasses several uncertainties. These uncertainties include the subjectivity of scoring the traits (Walrath et al. 2004; Walker 2008), and temporal and geographical differences in morphology (Walker 2008; Godde 2015). In addition to genes, these differences between populations are partly explained by different diets and activities influencing the muscle attachment sites that are scored as cranial traits, however these effects are difficult to quantify (Walker 2008; Krüger et al. 2015). In addition to non-metric traits, cranial measurements have been used in assessing whether the individual is male or female (for example, Kajanoja 1966; Franklin et al. 2013; Santos et al. 2014).

Sometimes the pelvis or cranium is not intact or present for sex estimation, especially in cases of commingled human remains (e.g. Niskanen 2006), and postcranial measurements may be used. Some studies have shown that long bone measurements, especially the articular areas, are more accurate for sex estimation than cranial measurements (Spradley & Jantz 2011; Bašić et al. 2017). Specifically, femoral and humeral head diameters have been utilized as indicators of sex since the early 20th century (Dwight 1905; Pearson & Bell 1919; Stewart 1979). Spradley and Jantz (2011) reported classification rates of 88 % and 83 % for femoral and humeral head metrics, respectively, in their sample of modern White Americans. These measurements were ranked as the fourth (femur) and tenth (humerus) best individual sex indicators among 44 postcranial and 34 cranial measurements including the mandible (Spradley & Jantz 2011).

Using measurements for sex estimation requires standards from a reference sample that is similar to the population under study (Gonçalves et al. 2014; Spradley et al. 2015). This can be problematic, especially in cases where the population is unknown, or appropriate reference samples are not available (Gonçalves et al. 2014). There are no postcranial measurement standards based on Finnish skeletal samples, thus the existing standards from other populations have been generally used in the osteological analyses. Most of these commonly used standards are based on American samples from the 19th–20th century (Stewart 1979; Spradley & Jantz 2011). However, their suitability for Finnish samples has not been investigated. This current study tests humeral and femoral head diameter standards from other populations against Finnish skeletal samples to examine their applicability for historic and archaeological Finnish samples. The study combines three different samples: one historic anatomical collection with known sexes to test the accuracy of the standards and two archaeological samples with estimated sexes to test the internal consistency between the estimates from long bones and other traits. Several European- or American-based standards are applied, and classification rates are calculated for the total sample, and for the three samples separately.

MATERIAL AND METHODS

Three skeletal samples from Finland were used for this study: Helsinki (19–20th century), Porvoo (17–18th century), and Renko (17–19th century). The Helsinki sample is an autopsy sample housed at the Finnish Museum of Natural History in Helsinki. The individuals with known sex in this sample (N =134, 99 males, 35 females) were born between 1843 and 1911. The collection consists mainly of crania, pelvises, and long bones. Most of the individuals have only upper or lower limb bones and sometimes one side only (For more information, see for example Telkkä 1950; Kajanoja 1966; Söderholm 2002). The two other samples are both archaeological materials that have been housed at the National Board of Antiquities (now Finnish Heritage Agency). The Porvoo sample (N = 22, 11 males, 11 females) is from a coastal town in southern Finland. These individuals were possibly of Swedish descent (For more information, see Salo 2016; Niskanen et al. 2018). The Renko sample (N = 30, 16 males and 14 females) is from southern Finland as well but from a rural inland locality (For more information, see Salo 2016; Niskanen et al. 2018). Sex estimates for the archaeological materials were primarily based on the pelvis. Cranial and mandibular features were used if pelvis was not available, but only if they could be calibrated against individuals with both pelvis and cranium (Salo 2016; Ruff 2018a). Only adult individuals with fused long bone epiphyses were included in the study. Individuals with pathological or traumatic changes to the measured area were excluded.

Two measurements (vertical diameter of the humeral head and vertical diameter of the femoral head) were tested in this study. The Helsinki collection was measured by the author in 2007, except for seven female individuals that were taken from Ruff's (2018b) European dataset (this measurement data is available online: https://www.hopkinsmedicine.org/fae/cbr.html). The Porvoo and Renko measurements are from Ruff's study as well (2018b; and the same online dataset). For the Helsinki sample, both sides were measured, if present. The same applies to the humeral measurements in the archaeological samples, but their femoral diameters were mostly collected from the right side. The

Table 1. The applied standards with their sectioning points and sample populations. Values below the sectioning point are classified as females and values above it as males.

Reference	Sample	Humeral diameter	Humerus sectioning point (mm)	Femoral diameter	Femur sectioning point (mm)
Spradley & Jantz 2011	20th century White Americans	Vertical	46	Maximum	45
Stewart 1979*	19–20th century Americans	Vertical	F <43 F? 43–45 M? 45–47 M > 47	Maximum	F <42.5 F? 42.5–43.5 ? 43.5–46.5 M? 46.5–47.5 M > 47.5
Tomczyk et al. 2017	18–19th century Polish	Vertical	43	-	-
Kranioti & Michalodimitrakis 2009	20th century Cretans	Vertical	43.79	-	-
Slaus et al. 2003	20th century Croatians	-	-	Maximum	45.45
Pearson & Bell 1919*	17th century English	-	-	Vertical	F<41.5 F? 41.5–43.5 ? 43.5–44.5 M? 44.5–45.5 M > 45.5

*F = female, F? = probable female, ? = indeterminate, M? = probable male, M = male.

measurements were taken with sliding calipers to the nearest 0.1 mm.

Four sex estimation standards were used as references for the classification of humeral and femoral head diameters (Table 1). The standards and their sectioning points were chosen from the literature, and their inclusion in this study is mainly based on the fact that they are commonly used for sex estimation (Pearson & Bell 1919; Stewart 1979; Spradley & Jantz 2011) and whether they are based on European reference samples (Slaus et al. 2003; Kranioti & Michalodimitrakis 2009; Tomczyk et al. 2017).

The vertical diameter of the humeral head was used in this study (Buikstra & Ubelaker 1994). The measurement technique and definition of femoral head diameter varies in the literature. Three measurements are generally described in the literature: vertical diameter, transverse diameter, and maximum diameter (see Martin & Saller 1957 for the first two; Buikstra & Ubelaker 1994 for the last one). Three of the applied standards in this study use maximum diameter anywhere it is found (Stewart 1979; Slaus et al. 2003; Spradley & Jantz 2011), but Pearson and Bell's (1919) study is based on the vertical axis of the femoral head.

Regardless of the slight difference in the femoral head measurements, this study used the vertical diameter as an equivalent to the maximum diameter. The Helsinki data collected by the author had both vertical and transverse femoral head diameters, and the vertical diameter was generally found to be greater in comparison of the two. When the transverse diameter was subtracted from the vertical, the mean differences were positive and less than one millimeter (males (N=45) 0.76 mm and females (N=13) 0.93 mm). Thus, it is assumed that the vertical diameter is closer to maximum diameter than the transverse diameter. This is in accordance with studies such as Curate et al. (2017) and Asala (2001), but disagrees with studies that report greater transverse diameter (Caiaffo et al. 2019; Cuzzullin et al. 2020). Mall et al.'s (2000) study showed almost equal values of transverse and vertical diameters.

Since the study includes two different data sets, inter-observer error was calculated for a Helsinki subsample of 22 and 23 individuals using femoral head and humeral head diameters respectively. TEM (technical error of measurement) was used to calculate the difference between the author's and Ruff et al.'s (2018b)

Table 2. Descriptive statistics for the vertical diameter of the humeral head (mm). N = sample size, SD = standard deviation, R = right, L= left.

	Helsinki males					Helsinki females					
	N	Mean	Median	SD	Min–max	N	Mean	Median	SD	Min–max	Effect size, r
Humeral head R	58	48.37	48.66	2.51	39.84–52.41	19	41.65	41.88	1.75	38.80–44.40	
Humeral head L	64	47.76	47.77	2.42	39.55–52.24	20	41.72	41.94	2.05	38.28–45.27	
Humeral head*	70	47.63	47.71	2.53	39.55–52.24	22	41.72	41.94	2.14	38.28–45.27	0.66
	Porvoo males					Porvoo females					
	N	Mean	Median	SD	Min–max	N	Mean	Median	SD	Min–max	Effect size, r
Humeral head R	9	46.93	46.50	1.88	44.60–51.10	7	41.08	41.30	1.38	39.30–42.70	
Humeral head L	10	46.95	46.65	1.36	45.80–50.30	8	41.93	41.60	1.84	39.70–45.00	
Humeral head*	10	46.95	46.65	1.36	45.80–50.30	9	42.02	41.90	1.74	39.70–45.00	0.84
	Renko males					Renko females					
	N	Mean	Median	SD	Min–max	N	Mean	Median	SD	Min–max	Effect size, r
Humeral head R	13	46.06	45.50	2.43	42.60–50.90	9	39.70	40.30	2.09	35.00–41.60	
Humeral head L	11	45.70	46.30	2.35	41.30–48.40	9	40.23	39.90	1.87	36.50–43.00	
Humeral head*	15	45.92	45.50	2.44	41.30–50.90	9	40.23	39.90	1.87	36.50–43.00	0.79

*Only left side per individual, unless left is missing then right side is used.

All the differences between sexes were statistically significant at the level of 0.05. Effect sizes, r, for Humeral head* listed on the table.

Differences in Humeral head* between samples were statistically non-significant, except between Helsinki and Renko males at the level of 0.05. Effect sizes, r, as follows:

Males: Helsinki–Porvoo 0.17, Helsinki–Renko 0.26, Porvoo–Renko 0.23.

Females: Helsinki–Porvoo 0.07, Helsinki–Renko 0.29, Porvoo–Renko 0.44.

measurements (See more about TEM; Ulijaszek & Kerr 1999; Goto & Mascie-Taylor 2007). Mann-Whitney U test for independent samples was used to compare the diameter measurements between the male and female samples from all three materials separately and to compare the diameters between these three materials. Non-parametric test was applied due to small sample sizes. IBM SPSS Statistics 24 was used for the statistical analyses. Effect sizes for the differences were also calculated to quantify the size of the differences (see Tomczak & Tomczak 2014). Classification rates were calculated as the percentages of correctly classified males and females in the samples, both separately and combined (total).

RESULTS

The inter-observer error was assessed using TEM, which was found to be small in both humeral head (0.36 mm with relative TEM 0.77 %) and femoral head (0.19 mm with relative TEM 0.49 %) measurements. Thus, combining measurements from different datasets seems to be acceptable in this study.

The descriptive statistics of the humeral and femoral diameters of the Finnish samples are presented in Table 2 and 3. They show that the male diameter means are greater in all samples. The differences between female and male measurements were all statistically significant (p-value level 0.05 with large effect sizes). The

Table 3. Descriptive statistics for the vertical diameter of the femoral head (mm). N = sample size, SD = standard deviation, R = right, L= left.

	Helsinki males					Helsinki females				
	N	**Mean**	**Median**	**SD**	**Min–max**	**N**	**Mean**	**Median**	**SD**	**Min–max**
Femoral head R	39	48.03	48.67	2.40	41.39–51.30	10	43.14	43.32	2.19	40.30–46.71
Femoral head L	45	48.04	48.46	2.39	41.09–52.74	17	42.40	42.00	2.87	37.80–49.10
Femoral head*	49	48.06	48.46	2.30	41.09–52.74	19	42.67	42.43	2.86	37.80–49.10
	Porvoo males					Porvoo females				
	N	**Mean**	**Median**	**SD**	**Min–max**	**N**	**Mean**	**Median**	**SD**	**Min–max**
Femoral head R	10	46.95	46.30	2.08	44.50–51.10	7	43.01	42.90	1.30	40.90–44.70
Femoral head L	-	-		-		3	42.53	43.20	3.35	38.90–45.50
Femoral head*	10	46.95	46.30	2.08	44.50–51.10	10	42.87	43.05	1.91	38.90–45.50
	Renko males					Renko females				
	N	**Mean**	**Median**	**SD**	**Min–max**	**N**	**Mean**	**Median**	**SD**	**Min–max**
Femoral head R	16	46.93	47.25	2.78	40.40–52.40	12	41.49	42.10	3.35	35.10–45.50
Femoral head L	-	-		-		1	46.60	46.60	-	-
Femoral head*	16	46.93	47.25	2.78	40.40–52.40	13	41.88	42.40	3.51	35.10–46.60

*Only left side per individual, unless left is missing then right side is used.

All the differences between sexes were statistically significant at the level of 0.05. Effect sizes, r, for Femoral head* listed on the table.

Differences in Femoral head* between samples were not statistically significant at the level of 0.05. Effect sizes, r, as follows:

Males: Helsinki–Porvoo 0.22, Helsinko–Renko 0.22, Porvoo–Renko 0.03.

Females: Helsinki–Porvoo 0.08, Helsinki–Renko 0.05, Porvoo–Renko 0.11.

Helsinki males exhibit the greatest diameters in both the humerus and femur in comparison to other materials. The difference is, however, statistically significant only in the humeral head between Helsinki and Renko (p-value 0.01 with small effect sizes) when using one side. In females, the Porvoo sample shows slightly higher means for both bones in comparison to Helsinki and Renko, but none of the comparisons between female groups were statistically significant.

Table 4 shows the percentages of correctly classified males and females combining all three samples and left and right sides separately. For humeral head diameter, the standard from Kranioti and Michalodimitrakis (2009) seems to work best for the Finnish sample when the total classification rate (91–96 %) is taken into account. Spradley and Jantz (2011) seems to classify females (100 %) better than other standards, but classification rates for males (~76 %) are lower than in other standards. When femoral head diameters are examined, Spradley and Jantz (2011) and Slaus et al. (2003) provide the highest total classification rates (83–88 %) and they classify females better than other standards. Pearson and Bell (1919) classifies males (~90 %) better than other standards. Stewart (1979) leaves about 15–28 % of the femoral diameters indeterminate which is a much higher percentage (vs. 5–7 %) than in Pearson and Bell (1919) which is the other standard with an indeterminate category.

In order to examine whether there were differences between the samples, percentages of

Table 4. Humeral and femoral head diameters and percentages (%) of correctly classified individuals with all samples together. Both measurements are vertical diameters.

	Humerus			Femur			
Reference	**Males**	**Females**	**Total**	**Males**	**Females**	**Total**	**Indeterminate**
Spradley & Jantz 2011							
Right side	76.3	100.0	83.5	87.7	89.7	88.3	-
Left side	76.5	100.0	83.6	86.7	76.2	83.3	-
Stewart 1979*							
Right side	85.0	100.0	89.6	69.2	62.1	67.0	27.7
Left side	85.9	91.9	87.7	77.8	76.2	77.3	15.2
Tomczyk et al. 2017							
Right side	96.3	82.9	92.2	-	-	-	-
Left side	95.3	73.0	88.5	-	-	-	-
Kranioti & Michalodimitrakis 2009							
Right side	95.0	97.1	95.7	-	-	-	-
Left side	92.9	86.5	91.0	-	-	-	-
Slaus et al. 2003							
Right side	-	-	-	81.5	89.7	84.0	-
Left side	-	-	-	86.7	76.2	83.3	-
Pearson & Bell 1919*							
Right side	-	-	-	90.8	62.1	81.9	7.4
Left side	-	-	-	88.9	76.1	84.8	4.5

*Combined groups: females and probably females, males and probably males.

correctly classified males and females were calculated for all the samples separately. The Helsinki sample's humeral head diameter is classified in a similar manner by all the standards, since the classification rates range from 85.9 to 90.2 % (Table 5). The Porvoo sample exhibits a larger range of classification rates from 84.2–94.7 % with Stewart (1979) being the most accurate standard. The classification rates for Renko based on Spradley and Jantz (2011) and Stewart (1979) are quite low (67–71 %) compared to the other two standards (92 %). These low total rates are a result of low classification rates of males, indicating that the humeral head diameters in Renko are smaller than in other samples.

The classification rates are lower for femoral head diameters than for humeral head diameters in all samples (Table 6). Other methods, but Stewart (1979), classify approximately 85 % of the total Helsinki sample correctly (Stewart classifies 77.9 %). The Porvoo sample has the largest range (60.0–90.0 %) for classification rates, whereas the Renko sample's rates vary between 62.1 % and 82.8 %. In all samples, Stewart (1979) provides the lowest rates. These results indicate that the male femoral heads tend to be smaller in Finnish samples than in Stewart's sample and particularly in these two archaeological samples.

DISCUSSION AND CONCLUSIONS

This pilot study tested humeral and femoral head diameter standards in Finnish skeletal samples. Four sectioning points of both measurements reported in the literature were applied to this study. Both diameter means of the more recent Helsinki sample are somewhat bigger in males than the ones in the archaeological samples. The difference might be due to the sample size, since the archaeological samples are very small in comparison to the modern sample. Another explanation could be secular change, in which femoral head size in the population would have increased gradually over centuries, and thus more recent 20th century samples exhibit larger diameters when compared to the archaeological samples. Although, the studies concentrating on temporal changes of femoral head diameter

Table 5. Humeral head diameter and percentages (%) of correctly classified individual by sample (using left side, unless only right side available).

	Helsinki			Porvoo			Renko		
	Males	**Females**	**Total**	**Males**	**Females**	**Total**	**Males**	**Females**	**Total**
Spradley & Jantz 2011	81.4	100.0	85.9	70.0	100.0	84.2	46.7	100.0	66.7
Tomczyk et al. 2017	94.3	63.6	87.0	100.0	77.8	89.5	93.3	88.8	91.7
Kranioti & Michal-odimitrakis 2009	92.9	81.8	90.2	100.0	77.8	89.5	86.7	100.0	91.7
Stewart 1979*	88.6	90.9	89.1	100.0	88.9	94.7	53.3	100.0	70.8

* Combined groups: females and probably females, males and probably males.

Table 6. Femoral head diameter and percentages (%) of correctly classified individuals by sample (using left side, unless only right side available).

	Helsinki			Porvoo			Renko		
	Males	**Females**	**Total**	**Males**	**Females**	**Total**	**Males**	**Females**	**Total**
Spradley & Jantz 2011	87.8	78.9	85.3	90.0	90.0	90.0	81.3	84.6	82.8
Slaus et al. 2003	87.8	78.9	85.3	80.0	90.0	85.0	68.8	84.6	75.9
Stewart* 1979	79.6	73.7	77.9	50.0	70.0	60.0	62.5	61.5	62.1
Pearson & Bell 1919**	89.8	73.7	85.3	100.0	70.0	85.0	87.5	61.5	65.5

* Combined groups: females and probably females, males and probably males. Indeterminate: Helsinki 16.2 %, Porvoo 40.0 %, Renko 31.0 %.

** Combined groups: females and probably females, males and probably males. Indeterminate: Helsinki 5.1 %, Porvoo 5.0 %, Renko 6.9 %.

show varying results. Jantz et al. (2017) found that femoral and humeral head dimensions were not significantly affected by secular trends in American Whites, whereas Gonçalves (2014) reported a significant increase in both humeral and femoral head diameters between two Portuguese samples dating from 19th–20th century and late 20th century. Cridlin (2016) noticed a significant decrease in femoral head diameter after 1920 in American Whites. This finding is different from Duthie et al.'s (1998) findings, which showed an increase in femoral head diameter in Scottish samples that consisted of individuals who had died between 1900–20 and the 1980s.

The accuracy for the Helsinki sample and internal consistency of the sex estimation for the two archaeological samples vary depending on the applied standards. When the total sample is used, the standards from Tomzyk et al. (2017) and Kranioti and Michalodimitrakis (2009) for humeral head seem to correctly classify the individuals at a greater rate than the other two (Spradley & Jantz 2011 and Stewart 1979). The original populations of these two standards are European (Polish and Cretan), which might explain their slightly better performance over the American standards. At least, the diameter means are closer to the Finnish means in these two samples, except for Polish males, which is smaller than the Finnish and Cretan means. However, the performance of the European standards differs for males and females, and the American standards seem to work better for the Finnish females.

The total accuracy and internal consistency for femoral head diameter are very similar in all standards, except Stewart (1979) which performs clearly worse than others. The total femoral classification rates are lower than for the humeral head. This could possibly be explained by the slightly dissimilar measurements. Three of these standards use maximum diameter, whereas the current study uses vertical diameter. Nevertheless, Pearson and Bell (1919) utilized vertical diameter, and only the male classification rates are higher in comparison to other standards.

The standards show some variation in their accuracy and internal consistency depending on the sample. Humeral head standards perform at similar rates in the Helsinki and Porvoo samples, but for the Renko sample European methods work better than American methods. This is because the Renko humeral heads are smaller than in other samples, as are the European standards. In addition, Tomczyk et al.'s (2017) standard is based on samples from the 18–19th century, so it is contemporaneous to the Finnish archaeological samples. In the case of femoral diameter, Stewart's (1979) standard seems to classify the Finnish samples poorly compared to others. This may be a result of the standard's greater sectioning points for males, and that diameters between 43.5–46.5 mm are considered indeterminate. Pearson and Bell (1919) also have a range for indeterminate (43.5–44.5 mm), but with a smaller sectioning point for males, and thus the males are classified more accurately.

It has to be pointed out that the overall internal consistency between the diameter measurements and previously estimated sex can also be influenced by the uncertainty of sex estimates of the archaeological samples. Ruff (2018b) and his colleagues used standard sex estimation methods from the pelvis and cranium, but it is possible that some of these archaeological individuals may have an incorrect sex assigned to them. Studies have shown that the morphology of the pelvis and cranium are generally in good agreement with the sex estimates based on DNA, but the results may depend on the traits used in the morphological sex estimation (see Bauer et al. 2013; Inskip et al. 2019).

This paper has demonstrated how variable the classification rates can be when applying various standards onto Finnish samples. Nevertheless, direct recommendations cannot be made at this point. For the total sample, it seems that the European-based samples provide the most suitable standards for humeral head, whereas Stewart's sectioning points of femoral head would not be recommended based on its performance in this study. However, the success in classification may depend on the sex of the unknown individual.

Future studies require a bigger sample and different measurements to be tested or even used to create population-specific sectioning points for the Finnish skeletal samples. The current sample has more males than females, and it should be balanced with more female measurements. In addition, more archaeological individuals should be added to the sample, and thus a more robust analysis of potential secular change could be performed. At this point, it is recommended, that caution should be used when sex is estimated from femoral and humeral head metrics in Finnish skeletal samples.

ACKNOWLEDGEMENTS

The author would like to thank C. Ruff and his colleagues for sharing the measurement data freely online for research. The European data set is available at https://www.hopkinsmedicine.org/fae/cbr.html.

REFERENCES

Unpublished sources

Söderholm, N. 2002. Den Anatomiska Bensamlingen vid Helsingfors Universitet. M.A. Thesis, Archaeology, University of Helsinki.

Literature

Asala, S. A. 2001. Sex determination from the head of the femur of South African Whites and Blacks. *Forensic Science International* 117(1–2): 15–22.

Bašić, Ž., Kružić, I., Jerković, I., Anđelinović, D., & Anđelinović, Š. 2017. Sex estimation standards for medieval and contemporary Croats. *Croatian Medical Journal* 58(3): 222–30.

Bauer, C., Niederstätter, H., McGlynn, G., Stadler, H. & Parson, W. 2013. Comparison of morphological and molecular genetic sex-typing on mediaeval human skeletal remains. *Forensic Science International: Genetics* 7(6): 581–6.

Buikstra, J. & Ubelaker, D. 1994. *Standards for Data Collection from Human Skeletal Remains*. Arkansas Archaeological Survey Research Series No. 44. Fayetteville, AR: Arkansas Archaeological Survey.

Caiaffo, V., Albuquerque, P., Albuquerque, P. & Oliveira, B. 2019. Sexual diagnosis through morphometric evaluation of the proximal femur. *International Journal of Morphology* 37(2): 391–6.

Cridlin, S. 2016. Piecewise regression analysis of secular change in the maximum femoral vertical head diameter of American White males and females. *Human Biology* 88(1): 47–55.

Curate, F., Umbelino, C., Perinha, A., Nogueira, C., Silva, A. M. & Cunha, E. 2017. Sex determination from the femur in Portuguese populations with classical and machine-learning classifiers. *Journal of Forensic and Legal Medicine* 52: 75–81.

Cuzzullin, M., Curate, F., Rodrigues Freire, A., Teixeira Costa, S., Bevilacqua Prado, F., Daruge Junior, E., Cunha, E. & Rossi, A. 2020. Validation of anthropological measures of the human femur for sex estimation in Brazilians. *Australian Journal of Forensic Sciences* (early view) DOI10.1080/00450618.2020.1729411.

Duthie, R. A., Bruce, M. F. & Hutchison, J. D. 1998. Changing proximal femoral geometry in north east Scotland: an osteometric study. *BMJ (Clinical research ed.)* 316: 1498.

Dwight, T. 1905. The size of the articular surfaces of the long bones as characteristic of sex: An anthropological study. *American Journal of Anatomy* 4: 19–31.

Ferembach, D., Schwindezky, I. & Stoukal, M. 1980. Recommendations for age and sex diagnoses of skeletons. *Journal of Human Evolution* 9: 517–49.

Franklin D., Cardini, A., Flavel, A. & Kuliukas, A. 2013. Estimation of sex from cranial measurements in a Western Australian population. *Forensic Science International* 229(1–3): 158.e1–158.e1588.

Godde, K. 2015. Secular trends in cranial morphological traits: A socioeconomic perspective of change and sexual dimorphism in North Americans 1849–1960. *Annals of Human Biology* 42: 253–9.

Gonçalves, D. 2014. Evaluation of the effect of secular changes in the reliability of osteometric methods for the sex estimation of Portuguese individuals. *Cadernos do GEEvH* 3(1): 53–65.

Gonçalves, D., Granja, R., Alves Cardoso, F. & Faustino de Carvalho, A. 2014. Sample-specific sex estimation in archaeological contexts with commingled human remains: a case study from the Middle Neolithic cave of Bom Santo in Portugal. *Journal of Archaeological Science* 49: 185–91.

Goto, R. & Mascie-Taylor, N. 2007. Precision of measurement as a component of human variation. *Journal of Physiological Anthropology* 26(2): 253–6.

Inskip, S., Scheib, C., Wohns, A., Ge, X., Kivisild, T. & Robb, J. 2019. Evaluating macroscopic sex estimation methods using genetically sexed archaeological material: The medieval skeletal collection from St John's Divinity School, Cambridge. *American Journal of Physical Anthropol*ogy 168: 340–51.

Jantz, R. Jantz, L. & Devlin, J. 2017. Secular changes in the postcranial skeleton of American Whites. *Human Biology* 88(1): 65–75.

Kajanoja, P. 1966. Sex determination of Finnish crania by discriminant function analysis. *American Journal of Physical Anthropology* 24: 29–33.

Kranioti, E. & Michalodimitrakis, M. 2009. Sexual dimorphism of the humerus in contemporary Cretans – A population-specific study and a review of the literature. *Journal of Forensic Sciences* 54(5): 996–1000.

Krüger, G., L'Abbé, E., Stull, K. & Kenyhercz, M. 2015. Sexual dimorphism in cranial morphology among modern South Africans. *International Journal of Legal Medicine* 129: 869–75.

Mall, G., Graw, M., Gehring, K. & Hubig, M. 2000. Determination of sex from femora.

Forensic Science International 113(1–3): 315–21.
Martin, R. & Saller, K. 1957. *Lehrbuch der Anthropologie in Systematischer Darstellung mit Besonderer Berücksichtigung der anthropologischen Methoden*. Stuttgart: Gustav Fischer Verlag.
Niskanen, M. 2006. Stature of the Merovingian-period inhabitants from Levänluhta, Finland. *Fennoscandia Archaeologica* XXIII: 24–36.
Niskanen, M., Maijanen, H., Junno, J.-A., Niinimäki, S., Salmi, A.-K., Vilkama, R., Väre, T., Salo, K., Kjellström, A. & Molnar, P. 2018. Scandinavia and Finland. In C. Ruff (ed.) *Skeletal Variation and Adaptation in Europeans: Upper Paleolithic to the Twentieth Century*: 355–96. Hoboken: Wiley Blackwell.
Pearson, K. & Bell, J. 1919. *A Study of the Long Bones of the English Skeleton. Text. Part I. Femur.* Cambridge: Cambridge University Press.
Rogers, T. & Saunders, S. 1994. Accuracy of sex determination using morphological traits of the human pelvis. *Journal of Forensic Sciences* 39(4): 1047–56.
Ruff, C. 2018a. Introduction. In C. Ruff (ed.) *Skeletal Variation and Adaptation in Europeans: Upper Paleolithic to the Twentieth Century*: 1–13. Hoboken: Wiley Blackwell.
Ruff, C. (ed.). 2018b. *Skeletal Variation and Adaptation in Europeans: Upper Paleolithic to the Twentieth Century.* Hoboken: Wiley Blackwell.
Salo, K. 2016. *Health in Southern Finland – Bioarchaeological Analysis of 555 Skeletons Excavated from nine Cemeteries (11th–19th century AD)*. PhD thesis, Archaeology, University of Helsinki.
Santos, F., Guyomarc'h, P. & Bruzek J. 2014. Statistical sex determination from craniometrics: Comparison of linear discriminant analysis, logistic regression, and support vector machines. *Forensic Science International* 245: 204.e1–204.e8.
Slaus, M., Strinović, D., Skavić, J. & Petrovecki, V. 2003. Discriminant function sexing of fragmentary and complete femora: standards for contemporary Croatia. *Journal of Forensic Sciences* 48(3): 509–12.
Spradley, K. & Jantz, R. 2011. Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences* 56: 289–96.
Spradley, K., Anderson, B. & Tise, M. 2015. Postcranial sex estimation criteria for Mexican Hispanics. *Journal of Forensic Sciences* 60: S27–S31.
Stewart, T. 1979. *Essentials of Forensic Anthropology*. Charles C. Thomas: Springfield.
Telkkä, A. 1950. On the prediction of human stature from the long bones. *Acta Anatomica* 9: 103–17.
Tomczak, M. & Tomczak, E. 2014. The need to report effect size estimates revisited: An overview of some recommended measures of effect size. *Trends in Sport Sciences* 1(21): 19–25.
Tomczyk, J., Nieczuja-Dwojacka, J., Zalewska, M., Niemiro, W. & Olczyk, W. 2017. Sex estimation of upper long bones by selected measurements in a Radom (Poland) population from the 18th and 19th centuries AD. *Anthropological Review 80*(3): 287–300.
Ulijaszek, S. & Kerr, D. 1999. Anthropometric measurement error and the assessment of nutritional status. *British Journal of Nutrition* 82: 165–77.
Waldron, T. 1987. The relative survival of the human skeleton: Implication for palaeopathology. In A. Boddington, A. Garland & R. Janaway (eds) *Death, Decay and Reconstruction*: 55–64. Manchester: Manchester University Press.
Walker, P. 2005. Greater sciatic notch morphology: Sex, age, and population differences. *American Journal of Physical Anthropology* 127: 385–91.
Walker, P. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology* 136: 39–50.
Walrath, D., Turner, P. & Bruzek, J. 2004. Reliability test of the visual Assessment of cranial traits for sex determination. *American Journal of Physical Anthropology* 125: 132–7.