Bioarchaeology of the Near East, 10:1–25 (2016)

Metric sex estimation of ancient Egyptian skeletal remains
Part I: Testing of published methods

Emily J. Marlow
KNH Centre for Biomedical Egyptology,
The University of Manchester,
Stopford Building, Oxford Road, Manchester M13 9PT, UK
email: emily.marlow@alumni.manchester.ac.uk

Abstract: *This paper is the first of two that explore sex estimation based on metric measurements of ancient Egyptian human skeletons. Use of measurement-based sex estimation methods is often essential when skeletal remains are fragmentary; however, metric techniques are prone to error as a result of several biases, notably population differences in body size and skeletal proportions. In addition, many commonly used metric equations, created using "modern" (c. 19th and 20th century) population samples, have not been validated for use with ancient Egyptians, and few population-specific equations exist. The study sample consists of 318 adult individuals, each represented by either a complete skeleton (n=162) or an isolated cranium (n=156). The majority of individuals date to Old Kingdom (n=106) or Late Period (n=154) Giza. In addition, 43 individuals date to Predynastic Period Keneh, 13 individuals to Middle Kingdom Sheikh Farag, and two individuals to Rameside Period Thebes. The sex of each individual was estimated using standard morphological methods. A total of 63 skeletal dimensions, or as many as it was possible to obtain, were measured for each individual in the sample. Testing of 12 "modern" metric sex estimation methods revealed total weighted accuracy rates as low as 30–40%; many of the methods were exceptionally poor at estimating the sex of males. Population-specific metric equations created by other researchers produced total accuracy rates ranging from 78–100% when tested on the study sample. The results of this study, the first to test "modern" metric sex estimation methods on ancient Egyptian skeletons, demonstrate that three methods can be applied to this population. This finding is of importance for all researchers currently engaged in excavation projects in Egypt, who require sex estimation methods that have been tested and validated for use in ancient Egyptian samples.*

Key words: human skeleton; discriminant analysis; skeletal size and proportions; osteometrics

Received 11 December 2015; accepted 4 May 2016; published online 9 June 2016 on www.anthropology.uw.edu.pl

Introduction

Egypt is one of the oldest and best-known civilisations of the Ancient World, and continues to hold the fascination of scholars even after several centuries of study. Important insights into the population of this ancient culture can be gained from several different lines of evidence, including the analysis of human skeletal remains. Skeletal analysis usually always begins with the assessment of sex, given its importance to the identity of the individual under study, its impact on other skeletal identification parameters such as age-at-death and stature, and its relevance to studies of other biological factors such as pathology or diet. The sex of an individual is additionally an integral component of studies using human skeletal remains to explore how past societies functioned via analysis of mortuary practices and beliefs, social organisation, and hierarchies. This is because differences in body treatment, grave form, and cemetery or burial location can provide important information about the roles of males and females within a society, as well as the overall cultural implications of differential treatment of the sexes (Knudson & Stojanowski 2008).

Sex may be estimated by morphological (qualitative) and metric (quantitative) methods, both of which have advantages and disadvantages (Rogers 2005). Often, the methods available for use will be dictated by the state of preservation of the human remains. In instances of good preservation, morphological examination of the bony pelvis is widely considered to be the most accurate technique (Bruzek & Murail 2006: 227; Buikstra & Ubelaker 1994:16; Byers 2008:177; Ferembach et al. 1980; Klales et al. 2012; Rösing et al. 2007). This is because the shape and size of the bony pelvis is directly related to biological function. Females give birth; therefore, the pelvis must be sufficiently wide and voluminous to allow the safe passage of a large foetal head through the birth canal (Bruzek & Murail 2006:227). Archaeological skeletons, however, are sometimes damaged, fragmented, or represented by a few isolated bones only, necessitating that sex be estimated by other means. Metric methods of sex estimation therefore have considerable value in the analysis of the remains of individuals from past human populations. A number of different bioarchaeologists engaged in ongoing excavation projects in Egypt have discussed the requirement to use metric methods of sex estimation because of poor preservation of the skeletal remains at their sites (Kaiser 2008:51; Rose 2006; Zabecki & Dabbs 2010; Zabecki et al. 2012). The metric methods used are generally not specific to the ancient Egyptians, nor have they been validated for use in this population. However, their use is considered essential because poor preservation means that sex cannot be assessed using other methods (Kaiser 2008:51).

Metric methods of sex estimation also suffer from some important limitations. For example, some methods may be impossible to apply if the bones required for measurement are broken or parts are missing. In addition, the creation of new metric sex estimation techniques generally requires the availability of a known (documented) sex reference sample. While several large known-sex reference samples are available, those that have been used most often in physical anthropological research, namely the Terry Collection (Smithsonian Institution), the Hamann-Todd Collection (Cleveland Museum of Natural History) and the Spitalfields Collection (Natural History Museum, London), consist only of American and European individuals who lived and died in the 19th and 20th centuries.

Other issues with metric sex estimation methods relate to population differences in skeletal size and proportions. Within each population, the phenotypic expression of adult body size and shape results from a synergistic interaction between hereditary (genetic) factors and environmental conditions experienced during growth, including climatic conditions (temperature, altitude), diet, subsistence strategy, activity patterns, disease, and access to resources (Vercellotti et al. 2011). Metric methods of sex estimation, which rely on absolute differences in measured skeletal dimensions and population-specific sectioning points, are therefore prone to error when generalised on a global level (Rogers 2005). This has been demonstrated in previous research where testing of a metric sex estimation method on a population sample other than the one used to create it resulted in lower accuracy rates than were reported in the original investigation (Cowal & Pastor 2008; Marlow & Pastor 2011).

To address these issues, several researchers have suggested that population-specific standards are required for all estimates of sex based on osteometric data (e.g. Bidmos & Asala 2003; Bidmos & Dayal 2004; Çöloglu et al. 1998; King et al. 1998; Mall et al. 2000; Özer & Katayama 2008; Šlaus & Tomičić 2005; Steyn & Işcan 1999; Trancho et al. 1997). Unfortunately, few population-specific metric methods of sex estimation are available for the ancient Egyptians, and those that are available have not been tested for accuracy on independent and/or dissimilar population samples. It is not surprising, therefore, that well-established metric sex estimation methods created using "modern" (c. 19th and 20th century) population samples have been applied to ancient Egyptian skeletal remains, despite a lack of studies validating this practice. This paper seeks to rectify this gap in knowledge by addressing a number of key objectives:

- to test the accuracy of commonly-cited and well-established methods of metric sex estimation created using "modern" (c. 19th and 20th century) population samples,

- to test the accuracy of two previously created metric sex estimation methods that are specific to the ancient Egyptians. To date, these methods have not been tested on a different sample of skeletons from ancient Egypt.

It is intended that the results of this study will be of value to all bioarchaeologists currently engaged in excavation projects in Egypt who require a set of metric sex estimation equations of tested and acceptable accuracy.

Material and methods

The skeletal reference sample consists of 318 adult individuals, each represented by either a complete skeleton (n=162) or an isolated cranium (n=156). The majority of individuals were recovered from cemeteries in Giza, which date to the Old Kingdom (c. 2686–2160 BC; n=106) or the Late Period (c. 664–332 BC; n=154). In addition, 43 individuals date to Predynastic Period (up to c. 3000 BC) Keneh, 13 individuals to Middle Kingdom (c. 2055–1650 BC) Sheikh Farag, and two individuals to Ramesside Period (c. 1186–1069 BC) Thebes. The locations of the sites sampled are shown in **Figure 1**. The skeletal collections are held at the Peabody Museum at Harvard University, Boston, US; the Natural History Museum (NHM) Vienna, Austria; and the Leverhulme Centre for Human Evolutionary Biology, University of Cambridge, UK.

Selection of skeletons

Skeletons from the three institutional collections listed previously were selected for inclusion in the reference sample based on a number of criteria:

- adult individuals, as demonstrated by complete epiphyseal fusion of all long bones (excluding the clavicle),
- presence of at least 75% of the skeleton (Buikstra & Ubelaker 1994:7), or presence of intact pubic bones/fragments of the bony pelvis commonly used for morphological sex estimation, such as the greater sciatic notch or sacrum,
- ability to assess sex as unambiguously male or female based on pelvic and/or cranial morphology,
- no evidence of pathology or trauma affecting the metric proportions of the bones studied.

Complete skeletons were rejected from the reference sample if no fragments of the *ossa coxae* from which sex could be estimated were present, or if the sex assessment was ambiguous. In instances where pelvic and cranial morphology gave differing sex estimates, greater weight was given to the sex estimate using pelvic traits.

Importantly, two named skeletons from the Peabody Museum collection, Yi-Neferti and her son Khonsu (accession numbers 33-63-50/N847.0 and 33-63-50/N846.0, respectively), were included in the reference sample. These skeletons, whose

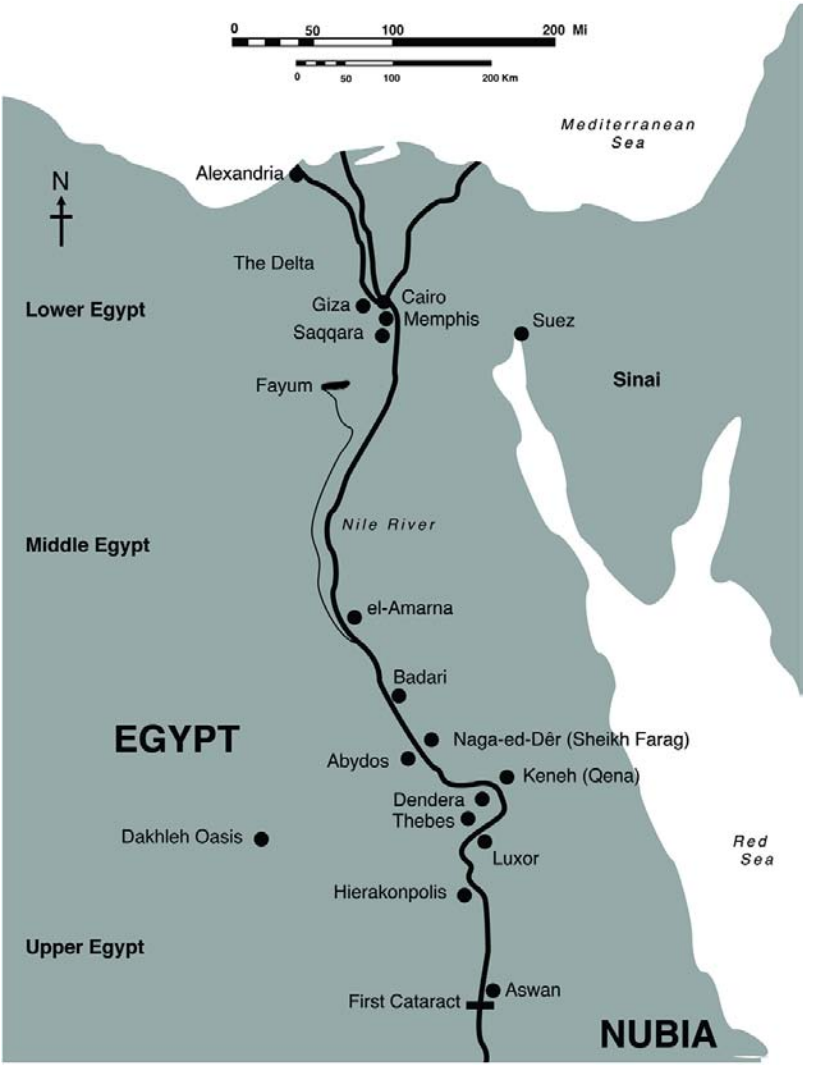

Figure 1. Map of ancient Egypt showing key cemetery sites and important cities and settlements. Adapted from The Oriental Institute, University of Chicago, Map Series.

names are known from coffin inscriptions (Gillet 1898), date to Ramesside Period (Twentieth Dynasty) Thebes and are important because they represent two individuals of known (documented) sex.

Morphological sex estimation

Ideally, sex estimation methods should be developed and tested using skeletal samples of known (documented) sex. Unfortunately, large series of ancient Egyptian skeletons of this type are extremely rare and/or currently unavailable for study; therefore, sex must be established by other means. For each individual in the reference sample, sex was therefore estimated using three standard morphological methods (Buikstra & Ubelaker 1994:16–19; Ferembach et al. 1980; Phenice 1969):

1. assessment of the Phenice characteristics (presence or absence of the ventral arc, the shape and appearance of the subpubic concavity and the appearance of the medial aspect of the ischiopubic ramus),
2. morphological assessment of sexually dimorphic features of the bony pelvis (*ossa coxae* and sacrum), notably the shape and form of the ilium, pelvic inlet, pubic bones, subpubic angle, obturator foramen, greater sciatic notch, preauricular sulcus, as well as the shape and level of curvature of the sacrum,
3. morphological assessment of sexually dimorphic features of the skull (occiput, supraorbital ridges, glabella, mastoid processes, frontal and parietal eminences, orbits, zygomatics, palate, occipital condyles, mandible, mental eminence, and gonial angle and flare).

Each individual sex indicator of both the bony pelvis and skull was scored after its level of masculinity and femininity in accordance with Buikstra and Ubelaker (1994: 21), where:

0 = Indeterminate sex: there are insufficient data available for sex estimation,
1 = Female: there is little doubt that the features represent a female,
2 = Probable female: the features are more likely to be female than male,
3 = Ambiguous sex: sexually diagnostic features are ambiguous,
4 = Probable male: the features are more likely to be male than female,
5 = Male: there is little doubt that the features represent a male.

Only two individuals in the study sample, Yi-Neferti and her son Khonsu, excavated from Ramesside Period Thebes, are of known sex. As a result, the testing of metric sex estimation equations within this study is based predominantly on an estimated sex reference sample. The extent to which use of an estimated sex reference sample is considered a limiting factor to the research design varies between investigators. The approach taken in the present study, to only include individuals in the study sample whose sex was considered to be unambiguous based on morphological indicators of the bony pelvis and skull, is supported by other researchers (Dabbs

2010; Dittrick & Suchey 1986; MacLaughlin & Bruce 1985; Murphy 2005; Šlaus & Tomičić 2005).

Inter-observer error test for morphological sex estimation

To reduce the impact that use of an estimated sex reference sample might have on the results of this study, an inter-observer error test for morphological sex estimation was performed. A total of 30 skeletal specimens (20 *ossa coxae* and 10 skulls) from the Tissue Bank located within the KNH Centre for Biomedical Egyptology at the University of Manchester, UK were independently sexed by two observers (the present author and a trained osteologist) using the Phenice traits and morphological indicators, as described previously. Each observer made their assessments independently and without reference to the other. In addition, each observer used the same recording form and estimated morphological sex using the same list of sexually dimorphic features and the same scoring system. Sex assessments made by the two observers were compared using the kappa statistic (K), which provides a measure of agreement among observers corrected for chance. Kappa values are scaled between 0–1, with 0 indicating the amount of agreement expected if scores were assigned randomly to specimens, and 1 indicating perfect agreement (Walker 2006).

Selection of metric sex estimation methods to be tested

The key aim of this study was to test the accuracy and precision of metric sex estimation methods that were created using modern population (c. 19th–20th century) reference samples of known sex (herein referred to as the "modern" methods) when applied to ancient Egyptian skeletal remains. However, given that the ancient Egyptian study sample is of estimated sex, accuracy in this context actually refers to 'consistency' with the morphological sex estimate. This aim was extended to include testing of two methods created by Raxter (2007) and Dabbs (2010) using ancient Egyptian population samples that are different and/or dissimilar to the present reference sample (the "population-specific methods"). Until now, the accuracy (or in this instance the ability of the techniques to produce a sex estimate that is consistent with the morphological assessment of sex) of these methods has never been tested.

A total of 12 "modern" methods were selected for inclusion and testing in this study. These methods were selected because they met a number of predefined criteria. These were that the method:

- had been cited or recommended in osteological textbooks, handbooks, or standards for data collection from human skeletal remains,

- had received a moderate (≥30 in the case of papers published in 1998 or later) or high (≥50 in the case of papers published in 1997 or earlier) number of citations on Google Scholar,
- was created using a modern population sample of known (documented) sex, such as the Terry or Hamann-Todd collections,
- included standard skeletal dimensions or skeletal dimensions that were novel but provided the opportunity to estimate sex even in highly fragmented remains,
- had demonstrated an accuracy of greater than or equal to 80%, both in total and in males and females separately, when tested on the original study sample. In forensic contexts, 80% is the cut-off point at which methods of metric sex estimation are generally considered useful (Rogers 1999).

Table 1 provides a summary of the 12 "modern" methods selected for testing in the study, including published accuracy rates and precision (as indicated by the level of intra-observer error, measured using percent error). In addition to the criteria given above, this group of methods was further selected because collectively it included a wide range of skeletal elements, as well as methods that only require a single measurement. These types of methods might prove to be highly valuable in instances where skeletal remains are fragmented or incomplete.

To date, only two studies have presented metric sex estimation methods that are specific to the ancient Egyptians. These methods, summarised in **Table 2**, were tested on the reference sample to establish the level of consistency of "population-specific methods" when applied to a different and/or dissimilar sample from the same general population.

Collection of metric data

A total of 63 skeletal dimensions were used in the present study. These dimensions were included because they are essential to the testing of the 14 metric methods summarised in **Tables 1** and **2**. For example, to test the equations presented within the Giles and Elliot (1963) "modern" method, a total of 11 specific dimensions of the cranium, listed in **Table 3**, are required. Similarly, testing of the equations created by Wescott (2000) required measurement of eight specific dimensions of the second cervical vertebra, and so on. The acronyms used in this table were taken from the publications in which the dimensions were described and defined and may not always correspond with the standard acronyms cited in Brothwell (1981) or Buikstra and Ubelaker (1994).

The dimensions used in the study were measured using digital sliding calipers, digital spreading (cranial) calipers, an osteometric board, and paper tape measure, as appropriate. Measurements were taken to the nearest millimetre, 0.1mm or 0.01mm, depending on the equipment used (the spreading calipers measured to one decimal

Table 1. Summary of "modern" metric methods tested, including original study populations and published accuracy rates.

No.	Bone/dimension	Study	Collection	Accuracy[1] %	Prec.[2] %
1	Cranial	Giles & Elliot 1963	Terry & HT[3]	82–86[4]	NR[6]
2	Second cervical vertebra	Wescott 2000	Terry & HT[3]	M[7]: 80–86[4,5] F[7]: 80–85[4,5]	1.5
3	Femoral head diameter	Krogman & Işcan 1986	Terry	90	NR[6]
4	Femoral neck diameter	Seidemann et al. 1998	HT[3]	93[5]	NR[6]
5	Femoral shaft circumference	Işcan & Miller-Shaivitz 1984a	Terry	84	NR[6]
6	MC1	Scheuer & Elkington 1993	UKMS[10]	94	NS[11]
7	Tibia (univariate)	Işcan & Miller-Shaivitz 1984b	Terry	White[8]: 84.8–87.3[4] Black[8]: 80.0–86.2[4]	NR[6]
	Tibia (multivariate)			White[8]: 84.8–86.1[4] Black[8]: 88.8–91.3[4]	
8	Humeral head diameter	Spradley & Jantz 2011	FADB[9]	White[8]: 83.0[5] Black[8]: 86.0[5]	NR[6]
9	Humerus, radius and ulna	Holman & Bennett 1991	Terry	M[7]: 80–86[4] F[7]: 80–86[4]	NR[6]
10	Radial head diameter	Berrizbeitia 1989	Terry	Maximum: 83[5] Minimum: 82[5]	NR[6]
11	MT1	Robling & Ubelaker 1997	Terry	M[7]: 91[5]; F[7]: 92[5]	1.0
12	Multiple bones	Stewart 1979:123	NR[6]	93–99[4]	NR[6]

[1] Accuracy and reliability rates are weighted (by sample size) or given for males and females separately if a weighted mean could not be calculated from the available data. [2] Mean precision. [3] Hamann-Todd Collection. [4] Depending on specific function used. [5] Cross-validated accuracy. [6] Not reported. [7] M – male; F – female. [8] 'Black' denotes equations developed in populations of African ancestry; 'White' denotes equations developed in populations of European ancestry. [9] Forensic Anthropology Data Bank. [10] UK-based medical schools. [11] Not significant.

Table 2. Summary of "population-specific methods" tested, including original study populations and published consistency rates.

No.	Bone/dimension	Study	Original population	Consistency rate[1], %
13	Long bones (FHD, CNF, HHD)[2]	Raxter 2007	Primarily Predynastic Period and Old Kingdom, Egyptian	89.0 for all three dimensions
14	Scapula	Dabbs 2010	New Kingdom, Egyptian	84–88[3,4]

[1] Consistency with morphological sex estimate. [2] FHD – femoral head diameter; CNF – circumference of tibia at nutrient foramen; HHD – humeral head diameter. [3] Depending on specific function used.
[4] Cross-validated accuracy.

Table 3. Definitions of skeletal dimensions recorded for metric analysis.

Bone	Dimension	Reference
Cranium	Glabello-occipital length (GO); Maximum width (MW); Basion-bregma height (BB); Maximum bizygomatic diameter (DB); Prosthion-nasion height (PN); Basion-nasion (BN); Basion-prosthion (BP); Nasal breadth (NB); Palate external breadth (PB); Opisthion-forehead length (OF); Mastoid length (ML)	Giles & Elliot 1963
Second cervical vertebra (C2)	Maximum sagittal length (XSL); Maximum height of dens (XDH); Dens sagittal diameter (DSD); Dens transverse diameter (DTD); Length of vertebral foramen (LVF); Maximum breadth across superior facets (SFB); Superior facet sagittal diameter (SFS); Superior facet transverse diameter (SFT)	Wescott 2000
Femur	Maximum diameter of femoral head (FHD)	Krogman & Işcan 1986
	Supero-inferior femoral neck diameter (FND)	Seidemann et al. 1998
	Femoral shaft circumference (FSC)	Işcan & Miller-Shaivitz 1984a
	Maximum femoral length (XFL); Minimum femoral transverse diameter (FTD); Epicondylar breadth of femur (EBF)	Stewart 1979
Tibia	Tibial length (TL); Circumference at nutrient foramen (CNF); Minimum shaft circumference (MSC); Antero-posterior diameter (APD); Transverse breadth (TB); Proximal epiphyseal breadth (PEB); Distal epiphyseal breadth (DEB)	Işcan & Miller-Shaivitz 1984b
Humerus	Vertical (maximum) humeral head diameter (HHD)	Spradley & Jantz 2011
	Maximum length of humerus (XHL); Maximum epicondylar width of humerus (EWH)	Stewart 1979
Radius	Maximum length of radius (XRL); Radius semibistyloid breadth (RSBB)	Holman & Bennett 1991
	Maximum head diameter (MAXD); Minimum head diameter (MIND)	Berrizbeitia 1989
Ulna	Maximum length of ulna (XUL); Ulna semibistyloid breadth (USBB)	Holman & Bennett 1991
Metacarpal 1 (MC1)	Interarticular length (IAL); Medio-lateral breadth of base (BML); Antero-posterior breadth of base (BAP); Medio-lateral breadth of head (HML); Antero-posterior breadth of head (HAP); Maximum midshaft diameter (MS)	Scheuer & Elkington 1993
Metatarsal 1 (MT1)	Length (L); Supero-inferior head height (SIH); Medio-lateral head width (MLH); Supero-inferior base height (SIB); Medio-lateral base width (MLB); Midshaft diameter (MSD)	Robling & Ubelaker 1997
Bony pelvis	Ischial length (IL); Pubic length (PL); Height of sciatic notch (HSN); Acetabulo-sciatic breadth (ASB)	Stewart 1979
Clavicle	Maximum length of clavicle (XCL)	Stewart 1979
Scapula	Maximum length of scapula (XHS); Maximum length of spine (XLS); Breadth of infraspinous body (BXB); Height of glenoid prominence (HAX); Breadth of glenoid prominence (BCB)	Dabbs 2010

place, whereas the digital sliding calipers gave measurements to two decimal places). For bilateral elements, the left side was recorded with substitution for the right side in instances where the left could not be measured. Alternatively, both sides were measured if it was required in order to test a particular method. The complete set of 63 dimensions, or as many as it was possible to record, was measured for each individual in the skeletal reference sample. Given that the purpose of this study was to test published methods of metric sex estimation, the procedures for measuring the required bones as stated by the author of the published study were followed. This ensured that the test was fair, having followed the methodology that the creator of the method had intended. Any special instructions for application of the metric sex estimation equations were also closely followed.

Each equation was applied in turn to all individuals in the reference sample for whom the required measurements were available. Where several different equations were presented as part of the same method, each equation was tested separately. Given differences in skeletal morphology and sexual dimorphism of populations of African ('Black') or European ('White') ancestry, several researchers developed different equations for populations pertaining to different ancestral groups. Where possible, equations developed using populations of unknown ancestry, that is, those consisting of pooled 'Black' and 'White' individuals, were preferentially selected for testing. In instances where pooled functions were not available, the equations for 'Black' and 'White' populations were tested separately. Consistency (accuracy) rates in percent were calculated for males and females separately by dividing the total number of consistent sex estimates for the equation by the number of individuals to whom the equation or method could be applied, and multiplying the result by 100. A weighted total consistency rate for males and females combined was obtained by adding the counts of consistent sex estimation across the two sexes and dividing by the total number of cases across the sexes, then multiplying the result by 100.

Results

Inter-observer error test for morphological sex estimation

The results of the Kappa test, performed to determine the level of agreement between morphological sex estimates made by two observers, are shown in **Table 4**.

As can be seen in **Table 4**, there was a substantial to near perfect level of agreement between the two observers in estimating sex using features of the *ossa coxae* (including the Phenice traits). The level of agreement for sex estimates using the skull was lower, but still in excess of what would have been expected from chance alone, suggesting a good level of consistency and repeatability in the morphological sex estimation technique employed in this study. Given that this test does not indicate the presence of systematic bias in the author's ability to assign sex to unknown individuals using the methodology described, it might be fair to assume that the morphological sex assessments of the study sample are reliable.

Descriptive demographics of study sample

The sex distribution of the sampled skeletons, broken down by geographic location and time period, is shown in **Table 5**. It is assumed that the study sample represents the general population from ancient Egypt, although in reality it might not. The civilisation of ancient Egypt spanned a period of around 3,000 years, and not all time periods are represented in this sample.

Table 4. Kappa values and level of agreement for assessment of inter-observer morphological sex estimation.

	Kappa[1]	**P-value**	**Agreement**
Phenice traits	0.792	<0.0004	Substantial
Ossa coxae morphological indicators	0.903	<0.00002	Almost perfect
Skull morphological indicators	0.583	0.065	Moderate

[1] Levels of agreement: K=0.0, no agreement; K=0.01 to 0.20, slight agreement; K=0.21 to 0.40, fair agreement; K=0.41 to 0.60, moderate agreement; K=0.61 to 0.80, substantial agreement; and K=0.81 to 1.0, almost perfect to perfect agreement (Landis & Koch 1977).

Table 5. The sex distribution of the Egyptian samples included in the study. Only unambiguously male or female skeletons were included.

Collection	**Region**	**Period**	**Number of skeletons**		
			M	**F**	**Total**
Peabody Museum	Keneh	Predynastic Period	21	22	43
Peabody Museum & Natural History Museum, Vienna	Giza	Old Kingdom	69	37	106
Peabody Museum	Sheikh Farag	Middle Kingdom	7	6	13
	Thebes	Ramesside Period	1	1	2
Duckworth Lab.	Giza	Late Period	85	69	154
Total			183	135	318
			57.5%	42.5%	100%

Sex estimates were based on pelvic and/or cranial morphology, as described previously. **Table 6** provides a summary of the number of individuals whose skeletal remains included at least one *os pubis*, pelvic material (at least one *os coxa* plus the sacrum, at least one *os coxa* only, or the sacrum only), or cranial material (the skull, the cranium only, or the mandible only).

Table 6. Frequency of skeletons with pelvic or cranial material in the study sample.

Site[1]	**N**	**Pubis**	***Os coxa* and sacrum**	***Os coxa* only**	**Sacrum only**	**Skull**	**Cranium only**	**Mandible only**
PD Keneh	43	21	24	4	6	18	14	3
OK Giza	106	66	71	24	0	90	14	0
MK Sheikh Farag	13	12	13	0	0	10	3	0
RP Thebes	2	2	2	0	0	2	0	0
LP Giza	154					45	109	0
Total	318	101	110	28	6	165	140	3

[1] PD – Predynastic Period; OK – Old Kingdom; MK – Middle Kingdom; RP – Ramesside Period; LP – Late Period.

Consistency of "modern" metric sex estimation methods

Table 7 provides a summary of the number and percent of metric sex estimates that are consistent with the morphological sex estimates. In this table, N is used to denote the number of individuals to which the method or individual equation could be applied, n the number of consistent sex estimates, and % the consistency rate associated with the method or equation. The number of consistent sex estimates and percent consistency rates are presented for males and females separately. The total consistency rate in percent is the weighted mean for males and females combined (weighted by sample size).

As can be seen in **Table 7**, only three of the 12 "modern" methods tested produced male and female consistency rates that reached or exceeded the 80% cut-off mark at which metric sex estimation methods are considered useful. Of these three methods, the highest total weighted sex consistency rate was obtained using Function 1 of the 'multiple bones' method developed by Stewart (1979; 90.1%), followed by Function 16 developed by Giles and Elliot (1963; 89.9%) and the first metacarpal method of Scheuer and Elkington (1993; 84.2%). Based on these results, it is possible to suggest that these three methods or specific functions might be of value to other researchers requiring metric methods of tested 'accuracy' to estimate sex in ancient Egyptian skeletal samples.

Of the remaining nine "modern" methods tested, three produced total weighted consistency rates that were worse than what would have been achieved using simple

Table 7. Consistent sex estimates and percent consistency rates associated with 12 "modern" metric sex estimation methods when tested on ancient Egyptian skeletal remains.

Function	**Consistent sex estimates[1]**				
	Male, n/N	**Male, %**	**Female, n/N**	**Female, %**	**Total, %**
1. Cranium (Giles & Elliot 1963)					
Function 3	83/87	95.4	59/74	79.7	88.2
Function 6	97/101	96.0	72/89	80.9	88.9
Function 9	100/105	95.2	72/89	80.9	88.7
Function 10	72/88	81.8	68/74	91.9	86.4
Function 13	84/88	95.5	60/74	81.1	88.9
Function 16	101/108	93.5	77/90	85.6	89.9
Function 17	85/103	82.5	81/89	91.0	86.5
Function 18	74/87	85.1	67/74	90.5	87.6
Function 21	90/103	87.4	79/88	89.8	88.5
2. Second cervical vertebra (Wescott 2000)					
Function 1	14/46	30.4	34/37	91.9	57.8
Function 2	13/45	28.9	31/34	91.2	55.7
Function 3	15/44	34.1	31/34	91.2	59.0
Function 4	17/44	38.6	31/34	91.2	61.5
Function 5	17/44	38.6	31/32	96.9	63.2
3. Femoral head diameter (Krogman & Işcan 1986)					
Function 1	37/78	47.4	57/58	98.3	69.1
4. Femoral neck diameter (Seidemann et al. 1998)					
Function 1	64/85	75.3	51/58	87.9	80.4
5. Femoral shaft circumference (Işcan & Miller-Shaivitz 1984a)					
Function 1	50/85	58.9	53/56	94.6	73.0
	ind. 1/85	1.2	ind. 1/56	1.8	ind. 1.4
6. Metacarpal 1 (Scheuer & Elkington 1993)					
MC1 function	48/57	84.2	32/38	84.2	84.2
7. Tibia (Işcan & Miller-Shaivitz 1984b) – Univariate					
PEB[2] – White	17/56	30.4	34/36	94.4	55.4
	ind. 5/56	8.9	-	-	ind. 5.4
PEB[2] – Black	6/56	10.7	35/36	97.2	44.6
	ind. 6/56	10.7	-	-	ind. 6.5
DEB[3] – White	22/78	28.2	44/47	93.6	52.8
	ind. 13/78	16.7	-	-	ind. 10.4
DEB[3] – Black	17/78	21.8	44/47	93.6	48.8
	ind. 4/78	5.1	ind. 1/47	2.1	ind. 4.0

[1] N – the number of individuals to which the method or equation could be applied; n – the number of consistent sex estimates; ind. – indeterminate sex.
[2] PEB – proximal epiphyseal breadth. [3] DEB – distal epiphyseal breadth.

Table 7. (continued)

Function	Consistent sex estimates[1]				
	Male, n/N	**Male, %**	**Female, n/N**	**Female, %**	**Total, %**
7. Tibia (Işcan & Miller-Shaivitz 1984b) – Multivariate					
Function 4 – White	23/55	41.8	33/35	94.3	62.2
Function 4 – Black	24/55	43.6	33/35	94.3	63.3
Function 6 – White	32/51	62.7	32/34	94.1	75.3
Function 6 – Black	20/51	39.2	32/34	94.1	61.2
Function 7 – White	34/56	60.7	34/36	94.4	73.9
Function 7 – Black	26/56	46.4	34/36	94.4	65.2
Function 8 – White	21/51	41.2	32/34	94.1	62.4
Function 8 – Black	12/51	23.5	32/34	94.1	51.8
Function 9 – White	18/51	35.3	32/34	94.1	58.8
Function 9 – Black	3/51	5.9	34/34	100	43.5
8. Humeral head diameter (Spradley & Jantz 2011)					
White	14/76	18.4	54/56	96.4	51.5
Black	33/76	43.4	54/56	96.4	65.0
9. Humerus, radius and ulna (Holman & Bennett 1991)					
Function 1	10/52	19.2	36/36	100.0	52.3
Function 2	9/56	16.1	38/38	100.0	50.0
Function 3	37/63	58.7	40/41	97.6	74.0
Function 4	26/52	50.0	36/36	100.0	70.5
Function 5	7/54	13.0	38/38	100.0	48.9
Function 6	5/48	10.4	36/36	100.0	48.8
Function 7	3/48	6.3	39/39	100.0	48.3
10. Radial head diameter (Berrizbeitia 1989)					
Maximum	8/42	19.0	23/26	88.5	45.6
	ind. 33/42	78.6	ind. 2/26	7.7	ind. 51.5
Minimum	4/37	10.8	24/26	92.3	44.4
	ind. 30/37	81.1	ind. 2/26	7.7	ind. 50.8
11. Metatarsal 1 (Robling & Ubelaker 1997)					
MT1 function	7/58	12.1	29/29	100.0	41.4
12. Multiple bones (Stewart 1979)					
Function 1	39/41	95.1	25/30	83.3	90.1
Function 2	43/45	95.6	26/32	81.3	89.6
Function 3	2/19	10.5	9/9	100.0	39.3
Function 4	1/24	4.2	13/13	100.0	37.8
Function 5	1/27	3.7	15/15	100.0	38.1
Function 6	0/32	0.0	17/17	100.0	34.7

[1] N – the number of individuals to which the method or equation could be applied; n – the number of consistent sex estimates; ind. – indeterminate sex.

guesswork. These were Function 6 of the multiple bones method developed by Stewart (1979); the minimum and maximum radial head diameter methods of Berrizbeitia (1989); and the first metatarsal method of Robling and Ubelaker (1997). These consistency rates are the result of the tendency of the methods to classify individuals as female, which suggests that ancient Egyptian males were smaller on average than modern European or American males, the populations used to create the “modern” techniques. As such, female consistency rates of 100%, as can be seen in **Table** 7, do not indicate that the method in question was highly accurate in correctly classifying the sex of females if it is associated with a very low male consistency rate. Rather, it simply indicates that there is a systematic bias in the classification of the sex of males. To put it another way, these methods have no discriminatory power when applied to ancient Egyptian skeletal remains and should not be used to estimate sex in samples from this population.

Three of the 12 “modern” metric sex estimation methods tested were unable to classify the sex of a number of individuals in the study sample, resulting in estimates of indeterminate sex. The proportion of individuals who could not be assigned a sex using these methods was generally low, with the exception of those produced using the radial head diameter method developed by Berrizbeitia (1989). This method assigned 78.6% and 81.1% of male individuals to the category of indeterminate sex using the maximum and minimum radial head diameters, respectively. All of these three methods are univariate and the indeterminate cases are the result of non-overlapping sectioning points associated with the method. For example, considering the minimum radial head diameter method of Berrizbeiria (1989), measurements of greater than or equal to 23mm indicate a male and less than or equal to 20mm indicate a female (Berrizbeitia 1989); therefore, measurements that fall in the range 20.01 to 22.99mm cannot be assigned to either group.

Two methods, using the tibia (Işcan & Miller-Shaivitz 1984b) and humeral head diameter (Spradley & Jantz 2011), present separate metric sex estimation equations for ‘Black’ and ‘White’ populations. When applied to ancient Egyptian skeletal remains, the ‘White’ univariate equations of the former method were found to produce higher consistency rates than the ‘Black’ equations, both in total and in males. In females, the ‘Black’ equations for proximal epiphyseal breadth (PEB) but not distal epiphyseal breadth (DEB) were found to be more accurate; the equation for the latter produced equal consistency rates in females. A similar pattern was observed using the multivariate equations of the tibia method. The ‘White’ equations were more accurate than the ‘Black’ equations, both in total and in males, for all functions excluding Function 4. In females, the consistency rates obtained using the equations for ‘Black’ and ‘White’ populations were equal for all functions excluding Function 9, where the consistency rates using the ‘Black’ equations were greater than those using the ‘White’ equations. In comparison, the consistency rates obtained from the humeral head diameter (HHD) method were greater using the 'Black' compared with the 'White' equation, both in total and in males separately. The female consistency rates using the two different equations were equal.

Consistency of "population-specific" methods

Table 8 shows the number of consistent sex estimates and consistency rates associated with metric sex estimation equations that were specifically created for use in ancient Egyptian populations when applied to a different Egyptian population sample.

The results presented in **Table 8** suggest that the femoral head diameter (FHD) sectioning point of Raxter (2007) and Functions 3–5 developed by Dabbs (2010) may be of value to other researchers working with ancient Egyptian skeletal remains, given that these methods produced acceptable levels of consistency in males and females separately in this independent test using a different population sample.

Table 8. Consistent sex estimates and percent consistency rates associated with two "population-specific" metric sex estimation methods when tested on the study sample.

Function	**Consistent sex estimates[1]**				
	Male, n/N	**Male, %**	**Female, n/N**	**Female, %**	**Total, %**
13. Long bones (Raxter 2007)					
FHD[2]	27/28	96.4	12/14	85.7	92.9
CNF[3]	34/37	91.9	5/13	38.5	78.0
	ind. 1/37	2.7	ind. 1/13	7.7	ind. 4.0
HHD[4]	30/32	93.8	10/13	76.9	88.9
	ind. 1/32	13.1	-	-	ind. 2.2
14. Scapula (Dabbs 2010)					
Function 1	68/73	93.2	34/45	75.6	86.4
Function 2	70/72	97.2	32/45	71.1	87.2
Function 3	15/15	100.0	12/12	100.0	100.0
Function 4	13/14	92.9	10/10	100.0	95.8
Function 5	13/14	92.9	10/10	100.0	95.8

[1] N – the number of individuals to which the method or equation could be applied; n – the number of consistent sex estimates; ind. – indeterminate sex.
[2] FHD – femoral head diameter. [3] CNF – circumference of tibia at nutrient foramen. [4] HHD – humeral head diameter.

Discussion

The primary aim of this research was to test the 'accuracy' (consistency) of "modern" metric sex estimation methods (c. 19^{th} and 20^{th} century population samples) when applied to human skeletal remains from ancient Egypt. Publications and reports written by bioarchaeologists contributing to ongoing excavation projects in Egypt highlight the use of such methods as part of the standard skeletal analysis procedures (Kaiser 2008:51), despite a lack of studies validating this practice. The findings of the present research indicate that many of the commonly cited "modern" metric sex estimation methods produce unacceptably low consistency rates (<80%), or have no real sex discriminatory power, when applied to ancient Egyptian skeletal remains. As such, they should not be considered for use in this population, even if no other methods are available.

A total of 12 "modern" methods were tested in the present study. Of these methods, only the functions that form part of three of them were able to meet or exceed the required 80% accuracy cut-off point in both males and females separately and hence in total. These were the functions (excluding Function 3) created by Giles and Elliot (1963), the Scheuer and Elkington (1993) MC1 method, and Functions 1 and 2 of the Stewart (1979) multiple bones method. The results of other tests of the Giles and Elliot (1963) cranial functions suggest that they are well-calibrated for ancient Egyptians but not necessarily for other populations (Franklin et al. 2005; Kajanoja 1966). Similarly, the metacarpals sex estimation method developed by Scheuer and Elkington (1993) was tested by Burrows et al. (2003) on a modern, known-sex dissection room sample held at Slippery Rock University School of Physical Therapy, Pennsylvania. They reported an accuracy rate of 65.9% using the multiple regression sex estimation equation for MC1. The present study is the first to test the Giles and Elliot (1963), Scheuer and Elkington (1993), and Stewart (1979; Functions 1 and 2) methods on a sample of ancient Egyptian skeletal remains. Based on the findings presented herein, it is possible to recommend these methods to other researchers requiring equations of high and tested 'accuracy' to estimate sex in fragmentary human skeletal remains from ancient Egypt.

Of the nine "modern" methods tested that failed to reach the 80% accuracy cut-off point in the present skeletal sample, the lowest consistency rates were obtained using the radial head diameter method of Berrizbeitia (1989), the MT1 method of Robling and Ubelaker (1997), and Function 6 of Stewart's (1979) multiple bones method. In fact, the weighted total accuracy rates obtained with these methods are worse than what would have been expected using simple guesswork (45.6% [maximum radial head diameter], 44.4% [minimum radial head diameter], 41.4% [MT1], and 34.7% [Function 6]). In each case, these findings are the result of very low consistency rates in males (19.0%, 10.8%, 12.1%, and 0%, respectively). In comparison, the female consistency rate associated with these methods ranged from 88.5–100%. However, this does not indicate that the methods are very accurate at classifying the sex of females. Given that to be classified as male the specific sectioning point for a function or equation must be exceeded, 'female' may be viewed as the "default" sex assignment for all individuals failing to reach the sectioning point, regardless of their actual sex. In other words, these methods have no actual power of discrimination between the sexes and should not be applied to ancient Egyptian skeletal remains.

It is a common trend among physical anthropologists to test a newly developed metric sex estimation method on a different population sample (Alunni-Perret et al. 2003; Cowal & Pastor 2008; Marlow & Pastor 2011). However, for the most part, this has been restricted to modern population samples used in forensic contexts, where there is a strict requirement for testing and peer review of all techniques and theories presented evidentially in a court of law (Rogers 2005). Typically, tests of this type result in lower rates of accuracy than were reported in the original investigation (Alunni-Perret et al. 2003; Burrows et al. 2003; Cowal & Pastor 2008; Marlow & Pastor 2011; Ríos Frutos 2003). Such a finding is, for the most part, clearly corroborated by the findings of the present research. There are several explanations that could account for the difference in original and tested sex estimation accuracy rates observed in these studies. These include high rates of intra- or inter-observer error (Liu 1988; Weinberg et al. 2005), bias in sampling methods during the creation of documented skeletal reference collections (Ericksen 1982; Hunt & Albanese 2005; Komar & Grivas 2008), or secular trends in growth (Jantz & Meadows Jantz 2000; Meadows & Jantz 1995). However, the most likely explanation for this lack of agreement is population differences in sexual dimorphism (Asala 2001; Çöloglu et al. 1998; Kemkes & Göbel 2006; MacLaughlin & Bruce 1986; Patriquin et al. 2003; Ruff 2002; Tanner 1976; Walker 2006), which has been found to vary considerably among extant human populations (Ruff 2002).

As a result, several researchers have proposed that population-specific standards are required for all estimates of sex based on osteometric data (e.g. Bidmos & Asala 2003; Bidmos & Dayal 2004; Çöloglu et al. 1998; King et al. 1998; Mall et al. 2000; Özer & Katayama 2008; Šlaus & Tomičić 2005; Steyn & Işcan 1999; Trancho et al. 1997). To date, only two studies have attempted to create metric sex estimation methods that are specific to the ancient Egyptians (Dabbs 2010; Raxter 2007), and neither of these studies have been tested on a different Egyptian sample. The results of this study largely support the hypothesis that metric sex estimation equations that are specific to the ancient Egyptians can produce high consistency rates when applied to a different sample from ancient Egypt. Of the three sectioning points developed by Raxter (2007), all were found to produce high rates of sex estimation consistency in males; however, the CNF and HHD methods produced low consistency rates in females, and therefore cannot be recommended to other researchers working with ancient Egyptian skeletal remains. Of the five discriminant functions created by Dabbs (2010), two, Functions 1 and 2, also produced unacceptably low consistency rates in females (75.6% and 71.1%, respectively).

One explanation that may account for this finding relates to differences in the composition of samples used by Raxter (2007), Dabbs (2010), and the present author. Previous research has demonstrated that skeletal size and/or proportions of ancient Egyptians changed over time (Masali 1972; Raxter 2011; Zakrzewski 2003). The method created by Dabbs (2010) was based on a sample of skeletons from New Kingdom Tell El-Amarna, whereas the present sample contains only two individuals dated to the New Kingdom (Ramesside Period). As such, the reference samples used in these three research studies may not be comparable in terms of skeletal size and proportions despite being derived from the same general population. Furthermore, the reference sample used in the present study, though a reasonable size, is not representative of the entire population from ancient Egypt because it does not include individuals dating to the complete historical period.

Conclusion

The present research is unique in that it represents the first attempt to validate the use of a range of "modern" metric sex estimation methods in ancient Egyptian skeletal samples. The tests of consistency of previously created metric sex estimation methods performed in this study demonstrated that the following techniques, functions, or equations may be used by other researchers studying ancient Egyptian skeletal remains:

- pooled ancestry cranial functions (excluding Function 3) of Giles and Elliot (1963),
- MC1 method of Scheuer and Elkington (1993),
- Functions 1 and 2 of the multiple bones method of Stewart (1979).

Testing further revealed that a number of components of two previously created population-specific methods may also be applied to Egyptian population samples that differ from those used in the original investigations:

- the FHD sectioning point of Raxter (2007),
- Functions 3–5 of Dabbs (2010).

The failure of other metric sex estimation methods to accurately separate the sexes, including the nine remaining "modern" methods, the CNF and HHD sectioning points of Raxter (2007), and Functions 1 and 2 of Dabbs (2010), is most reasonably explained by differences in the skeletal size and proportions of the reference and target samples.

The results of this study are relevant to all researchers working with ancient Egyptian skeletal remains who require metric methods of sex estimation of high and tested consistency (accuracy) that may be applied to even highly fragmented skeletons or isolated bones. The ability to accurately estimate sex has a key role in all studies examining health and disease, stature, body treatment, diet and social status/organisation, given that many of the conclusions reached would be meaningless were it not possible to establish the demographic profile of the samples used. It is hoped that in the future, researchers will be more selective about the metric equations they use to estimate sex, relying only on those equations that have been tested and are accurate in skeletal remains from the populations in question.

Acknowledgements

I would like to extend my thanks to the following curators for allowing access to skeletal collections: M. Morgan, The Peabody Museum of Archaeology and Ethnology; M. Teschler-Nicola, The Natural History Museum of Vienna; and M. Lahr, the Duckworth Laboratory at the University of Cambridge, as well as their members of staff. I would also like to thank Iwona Kozieradzka-Ogunmakin and the Editor and reviewers of *Bioarchaeology of the Near East* journal for reviewing this manuscript. Finally, I would like to thank Professor Andrew Chamberlain of the University of Manchester for his supervision of my PhD project, of which the current study was a component. The raw data analysed in this paper are available from the author upon request.

References

Alunni-Perret V., Staccini P., Quatrehomme G. (2003), *Reexamination of a measurement for sexual determination using the supero-inferior femoral neck diameter in a modern European population*, Journal of Forensic Sciences 48:517–520.

Asala S.A. (2001), *Sex determination from the head of the femur of South African whites and blacks*, Forensic Science International 117:16–22.

Berrizbeitia E.L. (1989), *Sex determination with the head of the radius*, Journal of Forensic Sciences 34:1206–1213.

Bidmos M.A., Asala S.A. (2003), *Discriminant function sexing of the calcaneus of the South African whites*, Journal of Forensic Sciences 48:1213–1218.

Bidmos M.A., Dayal M.R. (2004), *Further evidence to show population specificity of discriminant function equations for sex determination using the talus of South African blacks*, Journal of Forensic Sciences 49:1165–1170.

Brothwell D.R. (1981), *Digging up bones*, Ithaca, New York: Cornell University Press.

Bruzek J., Murail P. (2006), *Methodology and reliability of sex determination from the skeleton* [in:] "Forensic anthropology and medicine. Complementary sciences from recovery to cause of death", A. Schmitt, E. Cunha, J. Pinheiro (eds.), Totowa, New Jersey: Humana Press, pp. 225–242.

Buikstra J.E., Ubelaker D.H. (eds.) (1994), *Standards for data collection from human skeletal remains*, Fayetteville: Arkansas Archaeological Survey Research Series No. 44.

Burrows A.M., Zanella V.P., Brown T.M. (2003), *Testing the validity of metacarpal use in sex assessment of human skeletal remains*, Journal of Forensic Sciences 48:17–20.

Byers S.N. (2008), *Introduction to forensic anthropology*, Boston: Pearson Education Inc.

Çöloglu A.S., Işcan M.Y., Yavuz M.F., Sari H. (1998), *Sex determination from the ribs of contemporary Turks*, Journal of Forensic Sciences 43:273–276.

Cowal L.S., Pastor R.F. (2008), *Dimensional variation in the proximal ulna: evaluation of a metric method of sex assessment*, American Journal of Physical Anthropology 135:469–478.

Dabbs G. (2010), *Sex determination using the scapula in New Kingdom skeletons from Tell El-Amarna*, HOMO – Journal of Comparative Human Biology 61:413–420.

Dittrick J., Suchey J.M. (1986), *Sex determination of prehistoric central Californian skeletal remains using discriminant analysis of the femur and humerus*, American Journal of Physical Anthropology 70:3–9.

Ericksen M.F. (1982), *How "representative" is the Terry Collection? Evidence from the proximal femur*, American Journal of Physical Anthropology 59:345–350.

Ferembach D., Schwidetzky I., Stloukal M. (1980), *Recommendations for age and sex diagnoses of skeletons*, Journal of Human Evolution 9:517–549.

Franklin D., Freedman L., Milne N. (2005), *Sexual dimorphism and discriminant function sexing in indigenous South African crania*, HOMO – Journal of Comparative Human Biology 55:213–228.

Giles E., Elliot O. (1963), *Sex determination by discriminant function analysis of crania*, American Journal of Physical Anthropology 21:53–68.

Gillet C.R. (1898), *The Egyptian Antiquities in Halls 3 & 4. The Metropolitan Museum of Art. Hand-Book No. 4*, New York: Metropolitan Museum of Art.

Holman D.J., Bennett K.A. (1991), *Determination of sex from arm bone measurements*, American Journal of Physical Anthropology 84:421–426.

Hunt D.R., Albanese J. (2005), *History and demographic composition of the Robert J. Terry anatomical collection*, American Journal of Physical Anthropology 127: 406–417.

Işcan M.Y., Miller-Shaivitz P. (1984a), *Determination of sex from the femur in blacks and whites*, Collegium Antropologicum 8:169–175.

Işcan M.Y., Miller-Shaivitz P. (1984b), *Discriminant function sexing of the tibia*, Journal of Forensic Sciences 29:1087–1093.

Jantz R.L., Meadows Jantz L. (2000), *Secular change in craniofacial morphology*, American Journal of Human Biology 12:327–338.

Kaiser J. (2008), *2008 KKT human osteology* [in:] "Giza Plateau mapping project. Season 2008 preliminary report", M. Lehner, M. Kamel, A Tavares (eds.), Giza Occasional Papers 4, Boston: Ancient Egypt Research Associates, Inc, pp. 49–62.

Kajanoja P. (1966), *Sex determination of Finnish crania by discriminant function analysis*, American Journal of Physical Anthropology 24:29–34.

Kemkes A., Göbel T. (2006), *Metric assessment of the "mastoid triangle" for sex determination: a validation study*, Journal of Forensic Sciences 51:985–989.

King C.A., Işcan M.Y., Loth S.R. (1998), *Metric and comparative analysis of sexual dimorphism in the Thai femur*, Journal of Forensic Sciences 43:954–958.

Klales A.R., Ousley S.D., Vollner J.M. (2012), *A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods*, American Journal of Physical Anthropology 149:104–114.

Knudson K.J., Stojanowski C.M. (2008), *New directions in bioarchaeology: recent contributions to the study of human social identities*, Journal of Archaeological Research 16:397–432.

Komar D.A., Grivas C. (2008), *Manufactured populations: what do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection*, American Journal of Physical Anthropology 137:224–233.

Krogman W.M., Işcan M.Y. (1986), *The human skeleton in forensic medicine*, Springfield, IL: Charles Thomas.

Landis J.R., Koch G.G. (1977), *The measurement of observer agreement for categorical data*, Biometrics 33:159–174.

Liu K. (1988), *Measurement error and its impact on partial correlation and multiple linear regression analyses*, American Journal of Epidemiology 127:864–874.

MacLaughlin S.M., Bruce M.F. (1985), *A simple univariate technique for determining sex from fragmentary femora: its application to a Scottish short cist population*, American Journal of Physical Anthropology 67:413–417.

MacLaughlin S.M., Bruce M.F. (1986), *Population variation in sexual dimorphism in the human innominate*, Human Evolution 1:221–231.

Mall G., Graw M., Gehring K., Hubig M. (2000), *Determination of sex from femora*, Forensic Science International 113:315–321.

Marlow E.J., Pastor R.F. (2011), *Sex determination using the second cervical vertebra – a test of the method*, Journal of Forensic Sciences 56:165–169.

Masali M. (1972), *Body size and proportions as revealed by bone measurements and their meaning in environmental adaptation*, Journal of Human Evolution 1:187–197.

Meadows L., Jantz R.L. (1995), *Allometric secular change in the long bones from the 1800s to the present*, Journal of Forensic Sciences 40:762–767.
Murphy A.M. (2005), *The femoral head: sex assessment of prehistoric New Zealand Polynesian skeletal remains*, Forensic Science International 154:210–213.
Özer I., Katayama K. (2008), *Sex determination using the femur in an ancient Japanese population*, Collegium Antropologicum 32:67–72.
Patriquin M.L., Loth S.R., Steyn M. (2003), *Sexually dimorphic pelvic morphology in South African whites and blacks*, HOMO – Journal of Comparative Human Biology 53:255–262.
Phenice T.W. (1969), *A newly developed visual method of sexing the os pubis*, American Journal of Physical Anthropology 30:297–302.
Raxter M.H. (2007), *Metric sex estimation in an ancient Egyptian skeletal sample*, SAS Bulletin: Newsletter of the Society for Archaeological Sciences 30(4):9–12.
Raxter M.H. (2011), *Egyptian body size: A regional and worldwide perspective*, unpublished PhD dissertation, University of South Florida, USA.
Ríos Frutos L. (2003), *Brief communication: sex determination accuracy of the minimum superoinferior femoral neck diameter in a contemporary rural Guatemala population*, American Journal of Physical Anthropology 122:123–126.
Robling A.G., Ubelaker D.H. (1997), *Sex estimation from the metatarsals*, Journal of Forensic Sciences 42:1062–1069.
Rogers T.L. (1999), *A visual method of determining the sex of skeletal remains using the distal humerus*, Journal of Forensic Sciences 44:57–60.
Rogers T.L. (2005), *Determining the sex of human remains through cranial morphology*, Journal of Forensic Sciences 50:493–500.
Rose J.C. (2006), *Paleopathology of the commoners at Tell Amarna, Egypt, Akhenaten's capital city*, Memórias do Instituto Oswaldo Cruz 101(S2):73–76.
Rösing F.W., Graw M., Marré B., Ritz-Timme S., Rothschild M.A., Rötzscher K., Schmeling A., Schröder I., Geserick G. (2007), *Recommendations for the forensic diagnosis of sex and age from skeletons*, HOMO – Journal of Comparative Human Biology 58:75–89.
Ruff C. (2002), *Variation in human body size and shape*, Annual Review of Anthropology 31:211–232.
Scheuer J.L., Elkington N.M. (1993), *Sex determination from metacarpals and the first proximal phalanx*, Journal of Forensic Sciences 38:769–778.
Seidemann R.M., Stojanowski C.M., Doran G.H. (1998), *The use of the supero-inferior femoral neck diameter as a sex assessor*, American Journal of Physical Anthropology 107:305–313.
Šlaus M., Tomičić Ž. (2005), *Discriminant function sexing of fragmentary and complete tibiae from Medieval Croatian sites*, Forensic Science International 147:147–152.

Spradley M.K., Jantz R.L. (2011), *Sex estimation in forensic anthropology: skull versus postcranial elements*, Journal of Forensic Sciences 56:289–296.

Stewart T.D. (1979), *Essentials of forensic anthropology*, Springfield, Illinois: Charles C. Thomas.

Steyn M., Işcan M.Y. (1999), *Osteometric variation in the humerus: sexual dimorphism in South Africans*, Forensic Science International 106:77–85.

Tanner J.M. (1976), *Population differences in body size, shape, and growth rate*, Archives of Disease in Childhood 51:1–3.

Trancho G.J., Robledo B., López-Bueis I., Sánchez J.A. (1997), *Sexual determination of the femur using discriminant functions. Analysis of a Spanish population of known sex and age*, Journal of Forensic Sciences 42:181–185.

Vercellotti G., Stout S.D., Boano R., Sciulli P.W. (2011), *Intrapopulation variation in stature and body proportions: social status and sex differences in an Italian Medieval population (Trino Vercellese, VC)*, American Journal of Physical Anthropology 145:203–214.

Walker P.L. (2006), *Greater sciatic notch morphology: sex, age, and population differences*, American Journal of Physical Anthropology 127:385–391.

Weinberg S.M., Scott N.M., Neiswanger K., Marazita M.L. (2005), *Intraobserver error associated with measurements of the hand*, American Journal of Human Biology 17:368–371.

Wescott D.J. (2000), *Sex variation in the second cervical vertebra*, Journal of Forensic Sciences 45:462–466.

Zabecki M., Dabbs G. (2010), *Bioarchaeological report for the 2009 Amarna South Tombs Cemetery skeletal analysis*, unpublished report available online, accessed November 2015.

Zabecki M., Dabbs G., Montgomery T. (2012), *Report on the 2011 skeletal analysis of the South Tombs Cemetery Project*, unpublished report available online, accessed November 2015.

Zakrzewski S.R. (2003), *Variation in ancient Egyptian stature and body proportions*, American Journal of Physical Anthropology 121:219–229.