Bioarchaeology of the Near East, 10:27–46 (2016)

Metric sex estimation of ancient Egyptian skeletal remains

Part II: Testing of new population-specific methods

Emily J. Marlow*, Iwona Kozieradzka-Ogunmakin
KNH Centre for Biomedical Egyptology,
The University of Manchester,
Stopford Building, Oxford Road, Manchester M13 9PT, UK
email: emily.marlow@alumni.manchester.ac.uk (corresponding author)

Abstract: *This paper is the second of two that explore metric sex estimation of ancient Egyptian human skeletons. The purpose of the study is to create metric sex estimation methods that are specific to ancient Egyptians, and to ensure the methods will be of value to other researchers by testing them on a dissimilar sample from the same population. The population-specific methods were created using a reference sample consisting of 318 adult individuals. The majority of individuals were recovered from cemeteries in Giza, which date to the Old Kingdom (n=106) or the Late Period (n=154). In addition, 43 individuals date to Predynastic Period Keneh, 13 individuals to Middle Kingdom Sheikh Farag, and two individuals to Ramesside Period Thebes. Sex was estimated using standard morphological techniques. Discriminant function analysis with a stepwise approach was used to create the metric methods. The test sample consists of the skeletal remains of 119 (81 male, 38 female) adult individuals recovered from the Saqqara-West cemetery site. This site consists of burials dated to both the late Old Kingdom (n=28) and the Ptolemaic Period (n=91). The results of this test demonstrate that a number of the population-specific methods presented herein might be of value to other researchers working in Egypt, provided their sample derives from the same time period(s) and geographic locations as the reference and test samples used in this study.*

Key words: human skeletons; discriminant analysis; Saqqara

Introduction

Sex estimation is critical to the identification of human skeletal remains, and is integral to our understanding of past human societies and how they functioned. For example, studies of past mortuary practices and causes of mortality, which are important because of their presumed relationship to social organisation, behaviour and functioning, would be severely limited were it not possible to estimate the sex of individuals with an acceptable degree of accuracy and reliability (Milner et al. 2000:467–497).

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

Sex is most accurately assessed by morphological examination of the bony pelvis; however, this is not always possible when skeletons are badly damaged or fragmented. A number of bioarchaeologists currently engaged in ongoing excavation projects in Egypt have described the importance of, and their reliance on, metric sex estimation methods because of issues related to poor preservation of the skeletal remains they excavate (Kaiser 2008:51; Rose 2006; Zabecki & Dabbs 2010; Zabecki et al. 2012). While metric sex estimation methods have many advantages over morphological methods—they are easy to apply, result in fewer indeterminate cases, and more readily lend themselves to statistical testing and data manipulation—they are also subject to a number of important limitations, including their reliance on differences in measured skeletal dimensions. Human body size and proportions have been found to vary considerably among populations (Ruff 2002). Male and female body size and skeletal proportions often overlap, and in extreme cases, the males from one population may actually be smaller than the females from another population (Migliano et al. 2007). As a result, metric methods for estimating sex are particularly prone to error because they are based on absolute differences in measured skeletal dimensions and the sectioning points used to separate the sexes are only reliably applied to the population used to create the technique (Rogers 2005). This was clearly demonstrated in Part I (Marlow 2016), as well as in previous research (Cowal & Pastor 2008; Marlow & Pastor 2011). A number of commonly-cited and popularly-used metric sex estimation methods created using "modern" (c. 19th and 20th century) reference collections of American and European individuals resulted in accuracy rates as low as 35% when applied to skeletal remains from ancient Egypt (Marlow 2016).

A very well explored solution to this issue is the creation of population-specific metric sex estimation methods. Such methods have so far been created for numerous archaeological and modern populations, including prehistoric Scottish (MacLaughlin & Bruce 1985), prehistoric Central Californian (Dittrick & Suchey 1986), New Zealand Polynesian (Murphy 2005), medieval Croatian (Šlaus & Tomičić 2005), ancient Japanese (Özer & Katayama 2008), modern South African (Barrier & L'Abbé 2008), Guatemalan (Ríos Frutos 2005), and Greek (Steyn & Işcan 2008). Two studies have additionally addressed the need to create population-specific metric sex estimation equations for ancient Egyptian populations (Dabbs 2010; Raxter 2007), and the accuracy of these methods was tested on a different Egyptian sample in Part I (Marlow 2016).

Testing of all newly-created, population-specific metric sex estimation equations should be a routine component of the methodology of studies that aim to create new metric techniques. This is due to the nature of the statistical procedures used to create the new methods, in which classification of sex is based on classification coefficients that are derived from the sample under study (the skeletal reference sample). As a result, the coefficients usually work too well for the sample from which they were derived (Tabachnick & Fidell 1996:545), and the accuracy rate produced may not necessarily provide a realistic estimate of the ability of the metric method to separate males and females. Tests of accuracy, where a method is applied to a different sample from the population used to create the method, are therefore strongly encouraged before new population-specific metric sex estimation methods can be presented for use by other researchers.

The purpose of this research is two-fold: to create new metric sex estimation methods using discriminant function analysis that are specific to the ancient Egyptians and to test the functions on a different skeletal sample from ancient Egypt to validate their use in this population. The overall goal of this research is to create a suite of metric sex estimation methods of high and tested accuracy that might be of value to other bioarchaeologists conducting research into the culture and civilisation of ancient Egypt.

Materials and methods

Population-specific metric sex estimation methods were created using a skeletal reference sample consisting of 318 adult individuals. These individuals are 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.

The sex of each skeleton included in the reference sample was estimated using three standard 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).

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.

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.

The use of this methodology necessitates that the reference sample used to create the population-specific methods was an estimated sex sample due to the current lack of large series of ancient Egyptian skeletons of known sex. There are, however, a number of important limitations associated with this methodology, notably because it creates a circular argument. An estimated-sex reference sample is used to generate metric sex estimation equations; these equations are used to estimate sex in an unknown-sex target sample. However, given that skeletons were only included in the reference sample if they were assessed to be unambiguously male or female, and that an inter-observer error test of morphological sex estimation revealed no systematic bias in the first author's ability to assign sex to unknown individuals using the methodology described (Marlow 2016), the impact of this limitation should be minimal.

The population-specific methods were developed in the SPSS statistical software package (version 20.0) using discriminant function analysis with a stepwise approach. This procedure is a multivariate statistical approach that can be used to predict a binary categorical variable, such as sex, from a group of independent variables, such as skeletal measurements (Albanese 2003).

Test sample

The test sample consists of the skeletal remains of adult individuals recovered from the late Old Kingdom (Fifth–Sixth Dynasties; c. 2494–2181 BC) and Ptolemaic Period (c. 332–30 BC) cemeteries at Saqqara, located immediately west of the Step Pyramid complex of Pharaoh Djoser (c. 2667–2648 BC). Presently located around 40km south-west of the capital city of Cairo, the Saqqara necropolis was one of several extensive cemetery sites (including neighbouring Giza) that served the population of the ancient Memphite region as burial grounds from the Early Dynastic (c. 3000–2686 BC) to the Byzantine Period (AD 395–641). To date, the Saqqara-West site, which represents only a fraction of the entire Saqqara necropolis, has yielded almost 700 burials uncovered during excavation works conducted annually since 1996 by a team from the Polish Centre of Mediterranean Archaeology, University of Warsaw (Myśliwiec et al. 2004; Myśliwiec 2008; Myśliwiec et al. 2010; Myśliwiec 2013).

The sample consists of 119 adults excavated and examined between 2006 and 2012 by the second author (IK-O). The sex of nine individuals was positively identified as male by the presence of mummified external genitalia and/or well-preserved soft tissue of the face exhibiting facial hair. These individuals were classified as the 'known sex' subsample. An additional 110 individuals were confidently sexed using non-metric identification of sex-specific characteristics of the bony pelvis and skull, as described previously. Of these 110 individuals, 72 were male and 38 were female. Age estimations were primarily based on pubic symphyseal and auricular surface changes (Brooks & Suchey 1990; Lovejoy et al. 1985), and further supplemented by the degree of dental attrition and cranial suture closure (Brothwell 1965; Miles 1962; Smith 1984). Skeletal elements exhibiting pathological changes were not included. The majority of individuals (n=91) date to the Ptolemaic Period; 28 individuals date to the late Old Kingdom.

Metric data were collected as part of the skeletal analysis performed on site at Saqqara by IK-O. This study utilises five dimensions of the cranial and post-cranial skeleton:

- glabello-occipital length of cranium (GO) – the most anterior point on the frontal bone in the midline to the most distal point on the occiput in the midline, measured in the midsagittal plane,
- maximum breadth of cranium (MW) – the greatest breadth of the cranium, perpendicular to the median sagittal plane (avoiding the supramastoid crest),
- maximum bi-zygomatic diameter of cranium (DB) – the maximum width between the lateral surfaces of the zygomatic arches measured perpendicular to the median sagittal plane,
- femoral head diameter (FHD) – the maximum diameter of the femoral head, wherever it occurs,
- proximal epiphyseal breadth of tibia (PEB) – the maximum distance between the condyles (usually slightly below the articular surfaces).

These dimensions were selected for two reasons. First, they were included in the discriminant functions developed using data from the study reference sample and are therefore required to test the functions. Second, they were common to the two pieces of research conducted independently by IK-O and the first author (EJM).

Each of the population-specific discriminant functions created by EJM was blind-tested on all individuals in the test sample for whom the required measurements were available. A comparison of sex estimates using morphological versus metric methods was only made after all scores had been calculated and recorded. The metric estimate of sex was deemed to be “correct” if it was in agreement with the morphological estimate. Accuracy rates in percent were calculated for males and females separately by dividing the total number of correct sex estimates for the equation by the number of individuals to whom the function could be applied, and multiplying the result by 100. In this context, ‘accuracy’ actually refers to consistency with the morphological sex estimate, given that the skeletal reference sample is of estimated sex. A weighted mean accuracy (consistency) rate for males and females combined (‘total accuracy’) was obtained by adding the counts of correct sex estimation across the two sexes and dividing by the total number of cases across the sexes, then multiplying the result by 100.

Intra-observer and inter-observer error

The ability to replicate measurements reliably and precisely is an essential component of osteometric-based studies. In the present study, precision and reliability were evaluated using tests of intra-observer error, to establish the level of error associated with repeated measurements taken by the same observer, and inter-observer error, to establish the level of error associated with repeated measurements taken by different observers.

A random number generator was used to select a subsample of skeletons or isolated crania from each of the three collections used to create the study reference sample. All skeletal dimensions were then remeasured on a separate day after all original measurements had been collected. In total, measurements were retaken for 22 skeletons.

Precision was assessed using three separate measures:

1. percent intra-observer error,
2. paired samples t-test
3. technical error of measurement (TEM).

To allow comparison of TEM between different dimensions, percent or relative TEM was additionally calculated using the following equation: $(\text{TEM} \div \text{mean}) \times 100 = \%\text{TEM}$ (Ulijaszek & Kerr 1999). In this form, the %TEM represents an estimate of error magnitude relative to the size of the measurement; smaller percentages represent more precise measurements.

The reliability of the measurement of each dimension listed was assessed using an equation that calculates the reliability coefficient, R, a value that ranges from 0 (not reliable) to 1 (complete reliability). This coefficient reveals what proportion of the between-subject variance in a measured population is free from measurement error. In the case of a measurement with an *R*-value of 0.9, 90% of the variance could be attributed to factors other than measurement error (Ulijaszek & Lourie 1994). Although there are no recommended values for R, values greater than 0.95 are generally taken to indicate small errors and good quality control (Goto & Mascie-Taylor 2007; Ulijaszek & Kerr 1999). The reliability of the measuring instruments used in the study was further assessed by examining the Pearson correlation coefficient, *r*, which places a numerical value on the degree of agreement between test and retest measurements taken on the same individual. As a general rule, an *r* of 0.80 or higher indicates reliability (Spatz 2001).

A test of inter-observer error was conducted to determine the level of error associated with repeated measurements made by the two study authors (EJM and IK-O) for the five skeletal dimensions listed previously. The inter-observer error test sample consisted of two crania, two femora, and two tibiae, and forms part of the unprovenanced but very well preserved skeletal material held within the Tissue Bank at the KNH Centre for Biomedical Egyptology, University of Manchester, UK. Each observer used the same recording form, the same equipment (for example, the same type and brand of sliding or cranial calipers), and worked from the same list of skeletal landmark and/or dimension definitions. Each observer collected their measurements independently and without reference to the other. Similar to the test of intra-observer error, inter-observer precision and reliability were evaluated by calculating mean percent absolute error, TEM, %TEM, and *R* for each of the five skeletal dimensions used in this study.

Results

Intra-observer and inter-observer error

A total of 22 randomly selected individuals were included in the intra-observer error test sample. Of these, 16 individuals were represented by a complete skeleton; six were represented by a cranium only. Overall, the results of the intra-observer error test, given in **Table 1**, revealed that the measurements used in this study can be precisely and reliably measured.

Across all five dimensions, the average mean percent error is 0.4, with no individual measurement exceeding 1%. Similarly, no dimension exceeded 1% for %TEM. The results of the paired samples t-test demonstrated that for femoral head diameter (FHD) the retaken measurements were statistically significantly higher than the original measurements, using a significance level of 0.05. Examination of the Pearson correlation coefficient for each dimension revealed that the method of measuring the skeletal dimensions included in the study was reliable ($r \geq 0.80$) for all five dimensions. Similarly, the coefficient of reliability, *R*, scores for all five dimensions was greater than 0.99, suggesting that less than 1% of the variance associated with the means of the original and retaken measurements was the result of measurement error.

The results of the inter-observer error test are given in **Table 2**. Overall, the results appear to suggest that the five skeletal dimensions used in the study can be reliably measured by two different observers; however, a larger sample size would be required to confirm this finding.

Across all five dimensions, the average mean percent absolute error is 0.90, with no individual measurement exceeding 3%. As with the test of intra-observer error, the highest level of error is associated with measurement of the femur. The skeletal dimension exhibiting the lowest mean percent absolute error, the lowest TEM and the lowest %TEM was obtained for the maximum bi-zygomatic diameter of the cranium (DB). All five dimensions exceeded the critical value of reliability, *R*, of 0.95, which is indicative of small errors and good quality control. Overall, these results suggest that the measuring techniques of the two study authors (EJM and IK-O) are

Table 1. Results of the intra-observer error test.

Dim.[1]	N	Precision			Reliability		Paired t-test	
		Mean % error	TEM, mm	%TEM	*R*	*r*	T-value	P-value
GO	20	0.36	0.61	0.34	0.993	0.992	-0.459	0.652
MW	18	0.13	0.17	0.12	0.999	0.999	-0.487	0.633
DB	13	0.17	0.25	0.20	0.998	0.999	-1.594	0.137
FHD	14	0.62	0.28	0.65	0.994	0.999	-2.639	0.020
PEB	9	0.88	0.58	0.85	0.988	0.998	0.483	0.642

[1] Dimensions: GO – glabello-occipital length of cranium; MW – maximum width of cranium; DB – maximum bizygomatic diameter of cranium; FHD – femoral head diameter; PEB – proximal epiphyseal breadth of tibia.

Table 2. Results of the inter-observer error test between the two study authors.

Dim.[1]	N	Precision			Reliability
		Mean % error	TEM, mm	%TEM	*R*
GO	2	0.50	0.95	0.52	0.988
MW	2	0.64	0.63	0.47	0.965
DB	2	0.08	0.07	0.06	>0.999
FHD	2	2.50	0.56	1.20	0.996
PEB	2	0.79	0.43	0.55	0.997

[1] Dimensions: GO – glabello-occipital length of cranium; MW – maximum width of cranium; DB – maximum bizygomatic diameter of cranium; FHD – femoral head diameter; PEB – proximal epiphyseal breadth of tibia.

similar enough to allow conclusions to be drawn from tests based on data collected by different observers.

Population-specific discriminant functions

The population-specific discriminant functions created are given in **Table 3**. The table additionally provides the cross-validated accuracy rates associated with each function, both in total, and in males and females separately. The first function was created using cranial dimensions of the complete study sample, the second function using measurements from the femur and tibia of the complete study sample, and the third function using cranial measurements from the Late Period Giza subsample. This approach was taken because it enabled analysis of the effect of the composition of the reference sample on accuracy rates.

Table 3. Population-specific discriminant functions.

	Function 1	**Function 2**	**Function 3**
GO[1]	0.108		0.160
MW[2]			-0.074
DB[3]	0.146		0.166
FHD[4]		0.247	
PEB[5]		0.178	
Constant	-38.163	-22.718	-39.949
Sectioning point[6]	-0.005	-0.004	-0.0105
Male accuracy, %[7]	86.2	94.1	93.8
Female accuracy, %[7]	86.7	91.4	89.4
Overall accuracy, %[7]	86.4	93.0	91.8

[1] GO – glabello-occipital length of cranium.
[2] MW – maximum width of cranium. [3] DB – maximum bizygomatic diameter of cranium. [4] FHD – femoral head diameter.
[5] PEB – proximal epiphyseal breadth of tibia.
[6] Discriminant scores greater than the sectioning point indicate a male individual; scores less than the sectioning point indicate a female individual. [7] Cross-validated.

Accuracy of population-specific discriminant functions

Table 4 shows the calculated accuracy rates associated with the application of each function to the skeletal test sample from Saqqara-West.

Of the three discriminant functions tested, only Function 1, requiring measurement of GO and DB, was able to produce acceptable accuracy levels in both males and females. However, when the results are broken down by time period, the result for Ptolemaic Period females was a little below the 80% cut-off point at which metric

Table 4. Accuracy rates, broken down by sex and time period, associated with the application of discriminant functions to the sample of skeletons from the Saqqara-West necropolis.

Population	**Correct sex estimates**				
	Total	**Male, n/N**	**Male, %**	**Female, n/N**	**Female, %**
Function 1					
Old Kingdom	88.2	8/10	80.0	7/7	100.0
Ptolemaic Period	78.0	20/25	80.0	12/16	75.0
Total	81.0	28/35	80.0	19/23	82.6
Function 2					
Old Kingdom	94.7	12/12	100.0	6/7	85.7
Ptolemaic Period	87.5	46/46	100.0	17/26	65.4
Total	89.0	58/58	100.0	23/33	69.7
Function 3					
Old Kingdom	82.4	7/10	70.0	7/7	100.0
Ptolemaic Period	78.0	20/25	80.0	12/16	75.0
Total	79.3	27/35	77.1	19/23	82.6

sex estimation methods are generally considered useful (Rogers 1999). Considering the known-sex subsample only, of the five individuals to whom Functions 1 and 3 could be applied, one individual was misclassified, whereas all seven individuals to whom Function 2 could be applied were correctly sexed.

Discussion

This study is unique because it is the first to both create and test metric sex estimation methods that are specific to the ancient Egyptians. The findings of the test demonstrate that Functions 1 and 2 are accurate enough to be used to estimate the sex of skeletal remains from Old Kingdom (but not Ptolemaic Period) contexts at Saqqara; Function 1 could additionally be considered in individuals from either time period.

None of the functions tested performed particularly well in the Ptolemaic Period subsample. This indicates that the skeletal proportions of the individuals included in this subsample differed to those of the individuals in the reference sample. There are several reasons why this might be the case. The first is that the reference and target populations experienced different conditions during the crucial periods of growth and development. Given the composition of the reference (predominantly Predynastic Period and Old Kingdom) and target samples (Ptolemaic Period), these two groups could be considered to differ with respect to subsistence strategy, socioeconomic status, and access to resources, as well as in other ways such as genetic makeup and child-rearing practices. Previous research has demonstrated that growth in body size during infancy and early childhood is relatively similar in populations growing under optimal environmental conditions. This suggests that any deviations in growth during this period are a reflection of environmental and socioeconomic differences rather than genetic differences (Graitcer & Gentry 1981; Habicht et al. 1974). Differences in the growth of children from low socioeconomic backgrounds compared to those from high socioeconomic backgrounds reported in the literature include reduced stature and body weight, delayed skeletal maturation, and prolongation of the growth period (Dreizen et al. 1967; Graitcer & Gentry 1981; Habicht et al. 1974; Martorell et al. 1979). Other researchers have documented changes in stature in response to changing subsistence strategies (Cardoso & Gomes 2009; Larsen 1982; Mummert et al. 2011; Taylor 2010; Tobias 1962), notably the switch to agriculture. According to several authors, this switch results in an initial decrease in food quality and quantity and a concomitant decrease in stature and health status (Cohen & Armelagos 1984; Larsen 1995; Nickens 1976; Starling & Stock 2007; Stini 1971; Stock et al. 2011:347–367; Zakrzewski 2003).

Previous research has documented that the switch from a hunter–gatherer to an agricultural subsistence economy in ancient Egypt developed hand-in-hand with the formation of the state and the rise of hierarchical society (Hassan 1988; Marcus 2008; Wenke 1989, 1991). One theory linking the adoption of agriculture with the development of social stratification is that the elite class developed from groups of individuals with the desire to control agricultural surplus (Bard 1992; Castillos 2007; Frood 2010). Ultimately, this would lead to a society divided in terms of the quality and quantity of resources at the disposal of members of different social groups. For example, evidence from CT scans of mummies suggests that individuals of high socioeconomic status had access to 'luxury' foods that were high in saturated fats (Allam et al. 2011; David et al. 2010; Thompson et al. 2013). Thus, individuals growing in an environment of optimal, or at least plentiful, nutritional resources might be expected to exhibit different adult body size or proportions to individuals who may have suffered protein-calorie malnutrition at some point during childhood given the known relationship between stature, social status and access to resources (Komlos 1990; Steckel 1995).

Mortuary evidence suggests that the Ptolemaic Period individuals buried at Saqqara-West belonged to the non-elite social class. These individuals were predominantly deposited in shallow pit graves dug in the desert sand that contained minimal or no burial goods (Myśliwiec 2008, 2013; Myśliwiec et al. 2010). By comparison, the Old Kingdom burials, which were located beneath the sand and colluvial deposits into which the former graves were dug, consisted of rock-hewn shaft tombs equipped with subterranean burial chambers and grave goods. The individuals buried in these tombs at Saqqara-West were therefore considered to belong to the upper or governing classes (Myśliwiec et al. 2004, 2010; Myśliwiec 2013). Similarly, the Old Kingdom Giza skeletal remains, which form a large proportion of the reference sample used in this study, were excavated from the Western Mastaba Field near the Great Pyramid of Khufu, which was thought to hold the tombs of the governing classes and high officials (Der Manuelian 2009:23; Reisner 1942). The presence of a class division at Giza is additionally supported by the findings of more recent excavations at the site. For example, in the end of season report for the 1997 Koch-Ludwig Giza Plateau Mapping Project, Lehner (1997) notes that a large number of young male cattle bones were found in what could be a Fourth Dynasty 'rubbish dump'. Cattle are generally viewed as a costly source of meat; therefore, Lehner (1997) suggests that his team "...could [have been] digging the discarded remains of an expensive way of life." Despite this, the remains of fish, a cheap and common source of meat protein, were additionally found in abundance. This may suggest that a 'working class' also lived in close proximity to the 'elites' and that the two social groups were not completely segregated in this early period (Lehner 1997). Although the evidence appears to support the theory that the reference and Ptolemaic Period samples were not comparable in terms of social status and access to resources, and therefore may not have experienced similar conditions during growth and development, this is only one interpretation.

A second explanation for the differences in skeletal size and proportions of the reference sample and Ptolemaic Period subsample is that they exhibited different levels of genetic diversity. A number of studies examining cranial and dental traits for time-successive series of ancient Egyptian skeletal remains suggest that the basic population demonstrated a high degree of uniformity throughout the entire Dynastic Period (Berry & Berry 1972; Irish 2006), but underwent a significant change in the Ptolemaic Period (Berry et al. 1967). During this time Memphis was a flourishing centre for trade and commerce, and experienced a significant amount of immigration and infiltration by other ethnic groups (Angel 1972; Thompson 1988:82). According to Keita (1992) this migration is likely to have had a "major genetic impact" on the Egyptian population that probably occurred immediately prior to and during the Ptolemaic Period. This is supported by the findings of Schillaci and colleagues (2009), who found that the highest level of group diversity, as assessed using non-metric craniodental data and measures of biological distance, occurred during the Ptolemaic and Roman Periods.

Discriminant Function 3, which was created using data from the Late Period Giza ('Gizeh E') cranial sample, was unable to produce an acceptable accuracy rate in both males and females when applied to the test sample, even when broken down by time period. Several of the points discussed previously may also be relevant here in explaining this finding. In addition, a number of authors have suggested that the Gizeh E series is an atypical Egyptian sample, the crania possessing morphological features that are distinct from other Predynastic and early Dynastic Period samples from the Egyptian Nile Valley (Howells 1973; Keita 1990; Nikita et al. 2012; Zakrzewski 2004). Other authors have suggested that Near Eastern and European crania may be present in the 'E' series, which contains crania from the final periods of dynastic Egypt, periods that were characterised by increased foreign rule and immigration (Brace et al. 1993; Keita 1990). Alternatively, it is possible that the entire series represents a non-indigenous, immigrant population, the individuals of which were buried together in a separate and discrete cemetery; such practices are attested in other parts of the Memphite necropolis (Thompson 1988:88). As such, it may be fair to suggest that the development of metric sex estimation methods based on the metric measurements of this series for use in ancient Egyptian samples is not appropriate, unless use of the resulting functions is restricted to other specimens from the same series.

A significant advantage of the test sample used in the present study is that a small proportion was individuals whose sex was known thanks to the preservation of soft tissue and facial hair. It is expected that ongoing excavations at Saqqara-West may reveal additional individuals of known sex. When combined with CT scanning or other imaging of named mummies, it might be possible that a known-sex reference sample of sufficient size will become available to researchers exploring questions and issues related to sex estimation in ancient Egyptian skeletal remains. Ultimately, this would be the only really useful way of creating metric sex estimation equations, as it would negate some of the problems associated with the use of an estimated sex reference sample.

Conclusion

This is the first study to both create and test the accuracy of new sex estimation methods using reference and test samples from different ancient Egyptian subpopulations. The results of the test demonstrated that Functions 1 and 2 may be used by other researchers to estimate sex in ancient Egyptian skeletal remains, provided the target sample is similar in composition (both in terms of historical period and geographic location) to the reference and test samples used herein.

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 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.

Acknowledgements

We would like to extend our 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. We additionally wish to thank Prof. K. Myśliwiec and colleagues of the Saqqara-West necropolis excavation project, and the Polish Centre of Mediterranean Archaeology, University of Warsaw, for coordinating our archaeological expeditions and supporting us financially. Further thanks are extended to the reviewers and Editor of *Bioarchaeology of the Near East* for their invaluable comments on the manuscript. The data collection at Saqqara-West was conducted as part of the research project ("Dry Moat" in Saqqara as a unique time capsule, a source for the history of Memphite necropolis reflecting changes of natural environment) funded by the National Science Centre, Poland (research grant No. UMO-2013/11/B/HS3/04472). The raw data used in this paper are available from the authors upon request.

References

Albanese J. (2003), *A metric method for sex determination using the hipbone and femur*, Journal of Forensic Science 48:263–273.

Allam A.H., Thompson R.C., Wann L.S., Miyamoto M.I., el-Din A., el Maksoud G., Al-Tohamy Soliman M., Badr I., el-Rahman Amer H., Sutherland M.L., Sutherland J.D., Thomas G.S. (2011), *Atherosclerosis in ancient Egyptian mummies*, Journal of the American College of Cardiology Imaging 4:315–327.

Angel J.L. (1972), *Biological relations of Egyptian and Eastern Mediterranean populations during Pre-dynastic and Dynastic times*, Journal of Human Evolution 1: 307–313.

Bard K.A. (1992), *Toward an interpretation of the role of ideology in the evolution of complex society in Egypt*, Journal of Anthropological Archaeology 11:1–24.

Barrier I.L.O., L'Abbé E.N. (2008), *Sex determination from the radius and ulna in a modern South African sample*, Forensic Science International 179:85.e1–85.e7.

Berry A.C., Berry R.J. (1972), *Origins and relationships of the ancient Egyptians. Based on a study of non-metrical variations in the skull*, Journal of Human Evolution 1:199–208.

Berry A.C., Berry R.J., Ucko P.J. (1967), *Genetical change in ancient Egypt*, Man 2:551–568.

Brace C.L., Tracer D.P., Yaroch A., Robb J., Brandt K., Nelson A.R. (1993), *Clines and clusters versus "race": a test in ancient Egypt and the case of a death on the Nile*, Yearbook of Physical Anthropology 36:1–31.

Brooks S., Suchey J.M. (1990), *Skeletal age determination based on the* os pubis*: a comparison of the Ascadi-Nemeskeri and Suchey-Brooks methods*, Human Evolution 5:227–238.

Brothwell D.R. (1965), *Digging up bones: The excavation, treatment, and study of human skeletal remains*, London: The British Museum (Natural History).

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

Cardoso H.F.V., Gomes J.E.A. (2009), *Trends in adult stature of peoples who inhabited the modern Portuguese territory from the Mesolithic to the late 20th century*, International Journal of Osteoarchaeology 19:711–725.

Castillos J.J. (2007), *The beginning of class stratification in early Egypt*, Göttinger Miszellen 215:9–25.

Cohen M.N., Armelagos G.J. (eds.) (1984), *Paleopathology at the origins of agriculture*, Orlando, FL: Academic Press Inc.

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.

David A.R., Kershaw A., Heagerty A. (2010), *Atherosclerosis and diet in ancient Egypt*, Lancet 375:718–719.

Der Manuelian P. (2009), *Mastabas of Nucleus Cemetery G2100. Part I: Major mastabas of G2100–2200*, Boston: Museum of Fine Arts.

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.

Dreizen S., Spirakis C.N., Stone R.E. (1967), *A comparison of skeletal growth and maturation in undernourished and well-nourished girls before and after menarche*, The Journal of Pediatrics 70:256–263.

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

Frood E. (2010), *Social structure and daily life: Pharaonic* [in]: "A companion to Ancient Egypt", A.B. Lloyd (ed.), Oxford: Wiley-Blackwell Oxford.

Goto R., Mascie-Taylor C.G.N. (2007), *Precision of measurement as a component of human variation*, Journal of Physical Anthropology 26:253–256.

Graitcer P.L., Gentry E.M. (1981), *Measuring children: one reference for all*, Lancet 2:297–299.

Habicht J.P., Martorell R., Yarbrough C., Malina R.M., Klein R.E. (1974), *Height* *and weight standards for preschool children. How relevant are ethnic differences in growth potential?*, Lancet 1:611–614.

Hassan F.A. (1988), *The Predynastic of Egypt*, Journal of World Prehistory 2:135–185.

Howells W.W. (1973), *Cranial variation in man. A study by multivariate analysis of patterns of differences among recent human populations*, Papers of the Peabody Museum of Archeology and Ethnology 67, Cambridge, MA: Peabody Museum.

Irish J.D. (2006), *Who were the ancient Egyptians? Dental affinities among Neolithic through Postdynastic peoples*, American Journal of Physical Anthropology 129:529–543.

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

Keita S.O.Y. (1990), *Studies of ancient crania from Northern Africa*, American Journal of Physical Anthropology 83:35–48.

Keita S.O.Y. (1992), *Further studies of crania from ancient northern Africa: an analysis of crania from first Dynasty Egyptian tombs, using multiple discriminant functions*, American Journal of Physical Anthropology 87:245–254.

Komlos J. (1990), *Height and social status in eighteenth-century Germany*, The Journal of Interdisciplinary History 20:607–621.

Larsen C.S. (1982), *The anthropology of St. Catherines island 3. Prehistoric human biological adaptation*, Anthropological Papers of the American Museum of Natural History 57:157–276.

Larsen C.S. (1995), *Biological changes in human populations with agriculture*, Annual Review of Anthropology 24:185–213.

Lehner M. (1997), *End of field season full report. Pyramids: the inside story*, unpublished report available online, accessed April 2016.

Lovejoy C.O., Meindl R.S., Pryzbeck T.R., Mensforth R.P. (1985), *Chronological metamorphosis of the auricular surface of the ilium: a new method for the determination of adult skeletal age at death*, American Journal of Physical Anthropology 68:15–28.

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.

Marcus J. (2008), *The archaeological evidence for social evolution*, Annual Review of Anthropology 37:251–266.

Marlow E.J. (2016), *Metric sex estimation of ancient Egyptian skeletal remains, Part I: The applicability of "modern" methods*, Bioarchaeology of the Near East 10:1-25.

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.

Martorell R., Yarbrough C., Klein R.E., Lechtig A. (1979), *Maturation, body size, and skeletal maturation: interrelationships and implications for catch-up growth*, Human Biology 51:371–389.

Migliano A.B., Vinicius L., Lahr M.M. (2007), *Life history trade-offs explain the evolution of human pygmies*, Proceedings of the National Academy of Sciences USA 104:20216–20219.

Miles A.E.W. (1962), *Assessment of the ages of a population of Anglo-Saxons from their dentitions*, Proceedings of the Royal Society of Medicine 55:881–886.

Milner G.R., Wood J.W., Boldsen J.L. (2000), *Paleodemography* [in:] "Biological anthropology of the human skeleton", M.A. Katzenberg, S.R. Saunders (eds.), New York: Wiley-Liss Inc., pp. 467–497.

Mummert A., Esche E., Robinson J., Armelagos G.J. (2011), *Stature and robusticity during the agricultural transition: evidence from the bioarchaeological record*, Economics and Human Biology 9:284–301.

Murphy A.M. (2005), *The femoral head: sex assessment of prehistoric New Zealand Polynesian skeletal remains*, Forensic Science International 154:210–213.

Myśliwiec K. (ed.) (2008), *Saqqara III. The Upper Necropolis*, Warsaw: Centre of the Mediterranean Archaeology of the Polish Academy of Sciences.

Myśliwiec K. (ed.) (2013), *Saqqara V. Old Kingdom structures between the Step Pyramid Complex and the Dry Moat, Part I and II*, Warsaw: Centre of the Mediterranean Archaeology of the Polish Academy of Sciences.

Myśliwiec K., Kuraszkiewicz K., Czerwik D., Rzeuska T., Kaczmarek M., Kowalska A., Radomska M., Godziejewski Z. (2004), *Saqqara I. Tomb of Merefnebef*, Warsaw: Centre of the Mediterranean Archaeology of the Polish Academy of Sciences.

Myśliwiec K., Kuraszkiewicz K., Kowalska A., Radomska M., Rzeuska T.I., Kaczmarek M., Kozieradzka I., Godziejewski Z., Ikram S., Zatorska A. (2010), *Saqqara IV. The Funerary Complex of Nyankhnefertem*, Warsaw: Centre of the Mediterranean Archaeology of the Polish Academy of Sciences.

Nickens P.R. (1976), *Stature reduction as an adaptive response to food production in Mesoamerica*, Journal of Archaeological Science 3:31–41.

Nikita E., Mattingly D., Lahr M.M. (2012), *Three-dimensional cranial shape analyses and gene flow in North Africa during the Middle to Late Holocene*, Journal of Anthropological Archaeology 31:564–572.

Özer I., Katayama K. (2008), *Sex determination using the femur in an ancient Japanese population*, Collegium Antropologicum 32:67–72.

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.

Reisner G.A. (1942), *The history of the Giza Necropolis*, vol. 1, Cambridge, USA: Harvard University Press.

Ríos Frutos L. (2005), *Metric determination of sex from the humerus in a Guatemalan forensic sample*, Forensic Science International 147:153–157.

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(suppl. 2):73–76.

Ruff C. (2002), *Variation in human body size and shape*, Annual Review of Anthropology 31:211–232.

Schillaci M.A., Irish J.D., Wood C.C.E. (2009), *Further analysis of the population history of ancient Egyptians*, American Journal of Physical Anthropology 139:235–243.

Šlaus M., Tomičić Ž. (2005), *Discriminant function sexing of fragmentary and complete tibiae from Medieval Croatian sites*, Forensic Science International 147:147–152.

Smith B.H. (1984), *Patterns of molar wear in hunter-gatherers and agriculturalists*, American Journal of Physical Anthropology 63:39–56.

Spatz C. (2001), *Basic statistics: Tales of distributions*, Belmont, CA: Wadsworth/Thomson Learning.

Starling A.P., Stock J.T. (2007), *Dental indicators of health and stress in early Egyptian and Nubian agriculturalists: a difficult transition and gradual recovery*, American Journal of Physical Anthropology 134:520–528.

Steckel R.H. (1995), *Stature and the standard of living*, Journal of Economic Literature 33:1903–1940.

Steyn M., Işcan M.Y. (2008), *Metric sex determination from the pelvis in modern Greeks*, Forensic Science International 179:86.e1–e6.

Stini W.A. (1971), *Evolutionary implications of changing nutritional patterns in human populations*, American Anthropologist 73:1019–1030.

Stock J.T., O'Neill M.C., Ruff C.B., Zabecki M., Shackelford L., Rose J.C. (2011), *Body size, skeletal biomechanics, mobility and habitual activity from the Late Palaeolithic to the mid-Dynastic Nile Valley* [in:] "Human bioarchaeology of the transition to agriculture", R. Pinhasi, J.T. Stock (eds.), Oxford: Wiley-Blackwell, pp. 347–367.

Tabachnick B.G., Fidell L.S. (1996), *Using multivariate statistics*, New York: Harper Collins College Publishers.

Taylor M.S. (2010), *Adult stature and health among early foragers of the western Gulf coastal plains*, Plains Anthropologist 55:55–65.

Thompson D.J. (1988), *Memphis under the Ptolemies*, Princeton, New Jersey: Princeton University Press.

Thompson R.C., Allam A.H., Lombardi G.P., Wann L.S., Sutherland M.L., Sutherland J.D., Soliman M., Frohlich D., Mininberg D.T., Monge J.M., Vallodolid C.M., Cox S.L., el-Maksoud G., Badr I., Miyamoto M., Nur el-Din A.H., Narula J., Finch C.E., Thomas G.S. (2013), *Atherosclerosis across 4000 years of human history: the Horus study of four ancient populations*, Lancet 381:1211–1222.

Tobias P.V. (1962), *On the increasing stature of the Bushmen*, Anthropos 57:801–810.

Ulijaszek S.J., Kerr D.A. (1999), *Anthropometric measurement error and the assessment of nutritional status*, British Journal of Nutrition 82:165–177.

Ulijaszek S.J., Lourie J.A. (1994), *Intra- and inter-observer error in anthropometric measurement* [in:] “Anthropometry: The individual and the population”, S.J. Ulijaszek, C.G.N. Mascie-Taylor (eds.), Cambridge: University Press, pp. 30–55.

Wenke R.J. (1989), *Egypt: origins of complex societies*, Annual Review of Anthropology 18:129–155.

Wenke R.J. (1991), *The evolution of early Egyptian civilization: issues and evidence*, Journal of World Prehistory 5:279–329.

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.

Zakrzewski S.R. (2004), *Intra-population and temporal variation in ancient Egyptian crania*, American Journal of Physical Anthropology 123:215.