Bioarchaeology of the Near East, 14:11–26 (2020)

Estimating the sex of Ancient Egyptian skeletal remains: Methods from Tell el-Amarna

Gretchen R. Dabbs
Department of Anthropology, Southern Illinois University,
1000 Faner Dr., Carbondale, IL 62901, USA
email: gdabbs@siu.edu

Abstract: *The objective of this paper is to provide univariate and multivariate metric sex estimation techniques developed and tested specifically on New Kingdom Egyptian skeletal remains, which the literature lacks. Three samples from Tell el-Amarna were used. The South Tombs Cemetery development sample (STC_{DS} n=155; n_f=99, n_m=56) was used to establish sectioning points for univariate metric standards and multivariate equations using discriminant function analysis (stepwise 0.05 to enter, 0.10 to exit). The sectioning points and equations were tested on the cross-validated development sample and on a random hold-out sample from the South Tombs Cemetery (STC_{TS} n=59; n_f=34; n_m=25) and the totality of adult individuals with metric data from the North Tombs Cemetery (NTC_{TS} n=70; n_f=57; n_m=13). Univariate sectioning points identify sex in concordance with sex estimates based on pelvic and cranial morphology in 63.2–89.4% (cross-validated STC_{DS}) of cases. Test samples showed similar levels of concordance (STC_{TS} 52.5–95.2%; NTC_{TS} 63.8–100.0%). Fisher's exact tests show no statistically significant difference between the concordance rates for the three samples (all $p>0.002$, the alpha value with Bonferroni correction). Multivariate equations utilizing either multiple measurements of the same element or measurement of multiple elements produced sex estimates in concordance with those based on pelvic and cranial morphology in 81.3–92.6% (cross-validated STC_{DS}) of cases. Test samples show similar levels of concordance (STC_{TS} 80.6–96.3%; NTC_{TS} 78.3–100.0%; $p>0.05$ for all seven equations). These metric sex estimation techniques are of particular use when the pelvic and cranial morphology is ambiguous, when the skeletal material is incomplete, when the skeletal sample is comingled, and when the skeletal sample is curated by element, not individual.*

Key words: New Kingdom; discriminant function; Amarna Period

Introduction

Sex estimation from skeletal remains is an important basic assessment in bioarchaeology that has potential to impact further data collection, such as age or stature estimates. Further, distribution of individuals based on sex may have implications in

Received 7 January 2020; accepted 19 June 2020; published online 15 August 2020 on www.anthropology.uw.edu.pl

interpreting cemetery spatial organization (c.f., Stevens 2017) and can be further extrapolated into larger site assessments. The most reliable techniques for assessing sex in skeletal remains are based on the pubic region (Phenice 1969), a fragile and easily damaged skeletal element, although more comprehensive morphological techniques exist (Buikstra & Ubelaker 1994). Further, sex can also be estimated based on metric analysis of postcranial remains (Egyptian examples include Raxter 2007; Dabbs 2010; Marlow & Kozieradzka-Ogunmakin 2016) and on cranial morphology (Buikstra & Ubelaker 1994 after Acsádi & Nemeskeri 1970), in that order (Spradley & Jantz 2011).

Systematic variation in human proportions is a well-established fact (Ruff 1994; Holliday and Ruff 1997), with postcranial metric variables being directly affected by these varying proportions (Işcan et al. 1998; King et al. 1998; Gonzalez-Reimers et al. 2000; Mall et al. 2000). Thus, while using metric variables to assess sex can be more accurate than observing cranial morphology and can provide high rates of accurate classification (Spradley & Jantz 2011), this is only true when the technique being used was developed on populations similar in geographic and temporal composition as that under study.

Marlow (2016) and Raxter (2007) have demonstrated the need for ancient Egyptian specific metric standards for estimating sex by using ancient Egyptian remains to test methods developed on modern skeletal samples. Raxter (2007) examined univariate sex estimation techniques based on femur maximum vertical head diameter, humerus maximum vertical head diameter, and the circumference of the tibia at the nutrient foramen. In all cases, techniques developed using modern skeletal samples performed poorly when applied to ancient Egyptian samples (femur 54%, Stewart 1979; humerus 51%, Stewart 1979; tibia 56%, Symes & Jantz 1983). Marlow (2016) tested several multivariate methods that utilize the cranium, axis, femur, tibia, humerus, radius, first metacarpal, and first metatarsal. Overall, her results show poor classification rates for most (9/12 methods) of the techniques tested, with classification rates for several being as low as 30–40%. Males were particularly poorly classified (Marlow 2016), suggesting ancient Egyptians were more gracile than the modern American individuals used to develop the original techniques (Masali 1972).

Additionally, the taphonomic and curatorial condition of ancient Egyptian skeletal samples necessitates metric methods. Skeletal samples are often incomplete, fragmentary, or otherwise taphonomically damaged. Further, curation of ancient Egyptian human skeletal materials, particularly those excavated in the early 20th century and before, is often by skeletal element, not by individual, resulting in loss of contextual data and separation of postcranial elements from the most sexually dimorphic and diagnostic skeletal remains (c.f., Hrdlička Collection of Deir el-Medina human remains; Tomsová & Schierová 2016).

Despite the demonstrable need for ancient Egyptian specific methods for estimating sex, few have been published and those that have been published are generally limited in scope to a small number of variables on limited skeletal elements. Raxter (2007) provided sectioning points for the femoral head diameter, humeral head diameter, and tibia circumference at the nutrient foramen, indicating each metric correctly classified individuals in 89% of cases. Raxter (2007) also indicates that a multivariate function produced using the three aforementioned variables along with estimated stature, estimated body mass, maximum length of the tibia, maximum humeral length, and maximum femoral length correctly classified 93% of ancient Egyptian individuals in a temporally and geographically diverse sample. However, the exact function is not clearly provided in the publication (Raxter 2007). Dabbs (2010) provided one univariate and four multivariate functions using scapula metrics that correctly classified 84.6–88.0% of the cross-validated development sample of non-elite Amarna Period Egyptians. Recently, Marlow (2016) tested Dabbs' technique on a temporally and geographically distinct sample of ancient Egyptians, reporting high levels of overall concordance (86.4–100%), depending on the function used, with the three best functions performing at 95.8–100% concordance. Marlow and Kozieradzka-Ogunmakin (2016) provided multivariate functions for sex estimation developed on a geographically and temporally diverse ancient Egyptian sample. The functions were then tested on a separate sample from Saqqara consisting of remains dating from the Old Kingdom and the Ptolemaic period. Two craniometric functions performed well on the developmental sample (86.4–91.8% concordance), but the overall concordance rate for the test sample was somewhat lower (79.3–81.0%), with very poor performance in the temporally later component of the test sample. They also provided a multivariate function using the femoral head diameter and the proximal epiphyseal breadth of the tibia, which performed well in both the developmental (93%) and test samples (89%), but the female classification rates were very poor, averaging 69.7% (Marlow & Kozieradzka-Ogunmakin 2016).

This paper addresses multiple issues in sex estimation via postcranial metric analysis for ancient Egyptian remains. First, it provides univariate sectioning points for 33 variables collected from the postcranial long bones, extending beyond what most previous authors have examined. While these univariate techniques have high rates of classification concordance, multivariate techniques are often better classifiers and each long bone has been analyzed to provide a multivariate classification function that provides higher classification rates than any of the univariate sectioning points for that individual element. Further, a multi-element multivariate function with high classification rates is also provided. Finally, the sectioning points and single element and multi-element functions were tested against two separate test samples to demonstrate utility beyond the developmental sample.

The site

Tell el-Amarna, or simply Amarna, is the archaeological site representing the ancient city of Akhetaten, Akhenaten's capital city, built to venerate the Aten (**Figure 1**). Akhetaten was occupied for a short period of about 15–20 years during the reign of Akhenaten (c. 1353 BCE), built after he shifted the focus of the state level religion from the traditional pantheon of gods to focus on the Aten (Kemp 2012). The city was large, with 30,000–50,000 residents estimated and shows the hallmarks of a fully functioning urban area, with temples, housing areas, administrative offices, and craft production sites. After Akhenaten's death, the city was abandoned in short order, with the court returning to Memphis. Horemheb, who was a high official in Akhenaten's court and later king in his own right, eventually ordered the city dismantled (Kemp 2012). The abandonment and dismantling of the city likely did much to preserve it for 3,300 years.

Figure 1. Amarna situated within Egypt, image credit K. Underwood.

Amarna has been under excavations for over a century, with multiple expeditions focusing on various aspects of the city during that time. The cemeteries of the non-elites, however, were only recently identified via Geographic Information Systems (GIS) investigation (Fenwick 2003) and the four non-elite cemeteries identified to date have been the focus of a large-scale excavation and analysis project since 2005 (Stevens 2017) (**Figure 2**).

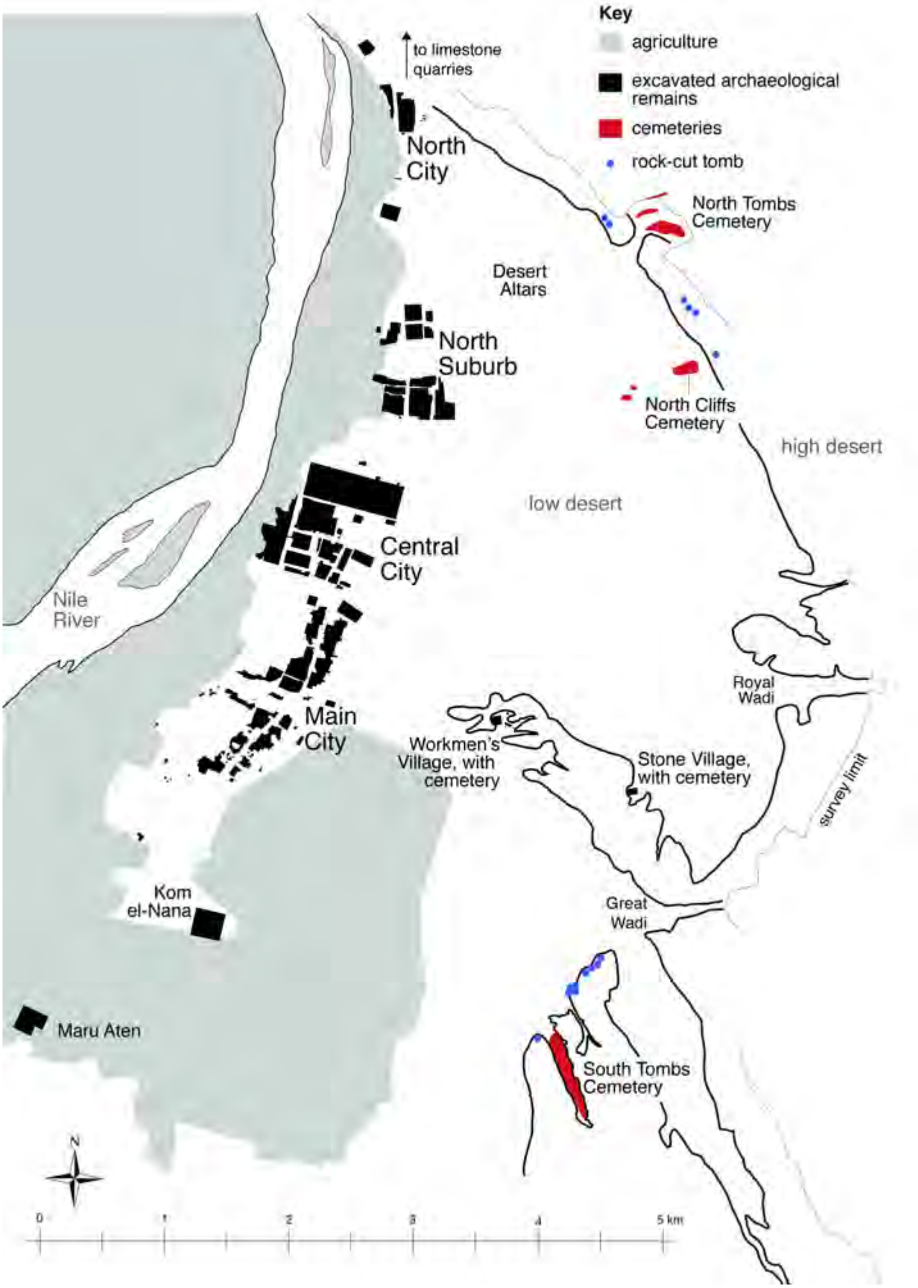

Figure 2. Amarna, as known today, showing the locations of the North and South Tombs Cemeteries in relationship to the excavated city, image credit Amarna Project.

The South Tombs Cemetery is the largest of the non-elite cemeteries, situated in a *wadi* between two low cliff rises that contain the southern grouping of elite tombs (The South Tombs). This cemetery was excavated from 2005–2013 and skeletal analysis is nearing completion (n=429). The North Tombs Cemetery is situated in a *wadi* separating Tombs 1/2 from the remaining tombs in the grouping known as the North Tombs. This cemetery was partially excavated in 2015 and 2017 and skeletal analysis is ongoing (n=252). In both cemeteries, individuals were buried in simple pit graves, generally wrapped in layers of textile and a plant-fiber mat (Stevens 2017). Both cemeteries have experienced multiple looting episodes, with the first estimated to be during the early period immediately after, or even during, the Amarna Period itself (Kemp et al. 2013).

Materials and methods

This study utilized three distinct samples for development and testing of the metric sex estimation methods presented below. The South Tombs Cemetery Development Sample (STC_{DS}) (n=155; n_f=99, n_m=56) consists of individuals from the STC with firm sex estimations based on pelvic remains. Using the protocol outlined in Buikstra and Ubelaker (1994), these individuals have been identified as either female or male, without any "probable" or "possible" qualifiers. No individuals with ambiguous sex estimates were included in the development sample. Generally speaking, the individuals in the STC_{DS} are well preserved and mostly complete.

The STC Test Sample (STC_{TS}) (n=59; n_f=34, n_m=25) consists of a random hold-out sample of individuals with sex estimation based on pelvic remains (Excel random number generator was used to create this subsample), individuals who have sex estimates of "probable female" or "probable male" based on pelvic remains (Buikstra & Ubelaker 1994 summary scores 2 and 4), or individuals identified as either female or male using cranial morphology or previously published scapular metric standards (Dabbs 2010), or some combination thereof. Individuals in the STC_{TS} are, generally speaking, less complete and more fragmentary than those in the STC_{DS}.

The North Tombs Cemetery Test Sample (NTC_{TS}) (n=70; n_f=57, n_m=13) includes all adult individuals from the North Tombs Cemetery excavations at Amarna with a sex estimate (either female/male or "probable female"/"probable male") and measurements of fully mature postcranial remains. These sex estimates may be based on pelvic or cranial morphology, or previously published scapular metric standards (Dabbs 2010) not included in this study, or a combination of these characters. The general level of preservation of these individuals is good to excellent and most are substantially complete. The NTC_{TS} has a four times more females than males, which is consistent with the overall demographic makeup of that cemetery (Dabbs 2019)

Table 1. Development and two test samples used in this study.

Sample	n	Description
South Tombs Cemetery Developmental Sample (STC_{DS})	155	Only individuals from the STC with female/male estimates based on pelvic remains
South Tombs Cemetery Test Sample (STC_{TS})	59	Individuals from the STC with female/male estimates based on pelvic remains (random hold-out sample); individuals with "probable" female/male estimates based on cranial and/or pelvic morphology or scapular metrics
North Tombs Cemetery Test Sample (NTC_{TS})	70	Individuals from the NTC with either female/male or "probable female/male" estimates based on pelvic or cranial morphology and/or postcranial metrics on fully fused elements

and is not due to selection criteria of this study. **Table 1** briefly outlines the different samples used in this study.

The complete long bones available for each individual were recorded to the nearest millimeter in 33 dimensions listed in **Table 2**. Most of these measurements have been previously described in Buikstra and Ubelaker (1994), and if so, the descriptions are not repeated here, but the measurements are identified by corresponding number in **Table 2** for ease of reference. Measurements not included in Buikstra and Ubelaker (1994) are described below with appropriate citation. Where no citation is provided, the measurement was developed specifically for the Amarna Cemeteries Project.

- **Humerus**
  - **Distal Articular Breadth** – distance between the medial and lateral margins of the distal articular surface at its most extreme distal aspect (after Byrd & Adams 2003:#41A);
  - **Deltoid Tuberosity Circumference** – maximum circumference of the deltoid tuberosity (measured with a fabric tape);
- **Femur**
  - **Distal Articular Breadth** – distance from the most lateral part of the articular surface to the most medial part of the articular surface at the most distal point (after Auerbach & Ruff 2004, 2006);
  - **Head Circumference** – circumference around the head of the femur at its maximum (measured with a fabric tape);
- **Tibia**
  - **Maximum Length** – maximum length of the tibia from the proximal tip of the intercondylar eminence to the distal tip of the medial malleolus (after Auerbach & Ruff 2004). Mark the midpoint with a pencil;

- **Length Without Eminences** – distance from the superior articular surface of the lateral condyle to the tip of the medial malleolus (Moore-Jansen et al. 1994). Measured using hole in the osteometric board;
- **Anteroposterior Diameter at 50%** – distance between the anterior and posterior surfaces at the midpoint of the maximum length (previously marked) (after Auerbach & Ruff 2004, 2006);
- **Mediolateral Diameter at 50%** – distance between the medial and lateral surfaces at the midpoint of the maximum length (previously marked) (after Auerbach & Ruff 2004, 2006);
- **Physiological Length** – distance between the articular surface of the lateral condyle and the distal articular surface (not including the malleolus) (measured with anthropometer).

All postcranial elements included in this study exhibit full fusion of the epiphyses for the element, which is not to say the individual had necessarily finished growing. For example, if the humeral epiphyses are fused on an individual, but the femoral epiphyses are not yet, the humerus would be included in this study, but not the femur. This is particularly important in the NTC_{TS}, which consists of mainly young adult individuals, reflecting the overall demographics of the North Tombs Cemetery (Dabbs 2019). During data collection, both the left and right antimeres were recorded, but for this study, where both were present, only the left was used. If the left antimere was not present, or was damaged, incomplete, or otherwise compromised in some way (usually healed trauma altering length or breadth), the right was used in its stead.

Data were tested for normality using Q-Q plots before further statistical analysis. All statistical analyses were performed using SPSS v. 26 (IBM Corp.; Armonk, NY). Independent samples t-tests were used to establish statistical significance between the means for males and females in the STC_{DS}. Sexual dimorphism scores are calculated as the difference in mean value divided by the sum of standard deviations for each observation (Dabbs 2010). The univariate sectioning point was calculated as the simple average of the mean values for males and females (after Spradley & Jantz 2011) in the STC_{DS}. For the sectioning points, individuals with values below the sectioning point indicate female and values above indicate male. If an individual metric was exactly equal to the sectioning point, the individual was identified as indeterminate for this analysis and was included as an incorrect assessment in the calculation of percentage correct listed in **Tables 3–4**. Classification rates reported were calculated by sex, recorded as number correct divided by total observed for individual sex. The overall concordance rates reported is the total number of individuals correctly classified divided by the total number of individuals observed.

It should be acknowledged that given the actual sex of these individuals is unknown, all indications of “correct” sex estimation based on any metric variable(s) reported in this paper are actually reports of concordance with the sex estimate based on pelvic or cranial morphology and/or metric analysis of the scapula (Dabbs 2010), which was not included in this study. However, given the degree of conservatism used in estimating sex in the original data collection and for overall simplicity in reporting,

Table 2. Number of individuals observed, mean, standard deviation (SD), and p-values for all measurements included in this analysis for the STC_{DS}; parenthetical numbers correspond with measurement descriptions in Buikstra and Ubelaker (1994); all other measurements described in Methods section.

Measurement	Female			Male			p-value
	n	Mean	SD	n	Mean	SD	
Clavicle							
Length (35)	69	135.6	7.8	35	144.2	9.5	<0.001
Anteroposterior Diameter (36)	67	10.1	1.0	36	12.1	1.9	<0.001
Superoinferior Diameter (37)	67	9.1	1.0	36	10.7	1.3	<0.001
Humerus							
Maximum Length (40)	83	291.5	16.9	43	307.4	13.5	<0.001
Distal Articular Breadth	74	37.6	3.8	36	42.1	2.9	<0.001
Epicondylar Breadth (41)	86	54.6	3.3	44	61.6	2.8	<0.001
Circumference at Deltoid Tuberosity	79	60.0	4.4	43	64.2	6.2	<0.001
Max Diameter at 50% (43)	82	19.7	1.5	43	20.5	1.5	0.008
Min Diameter at 50% (44)	82	14.8	1.3	43	15.7	1.3	<0.001
Vertical Head Diameter (42)	83	38.3	2.2	45	43.6	2.52	<0.001
Radius							
Maximum Length (45)	84	225.4	13.2	45	246.3	11.0	<0.001
Anteroposterior Diameter (46)	84	10.2	0.9	44	11.7	1.1	<0.001
Mediolateral Diameter (47)	84	12.8	1.1	45	14.0	1.4	<0.001
Ulna							
Maximum Length (48)	80	246.1	13.5	47	267.5	12.3	<0.001
Physiological Length (51)	81	216.4	13.2	46	234.9	11.1	<0.001
Anteroposterior Diameter (49)	78	12.0	1.6	46	14.0	1.7	<0.001
Mediolateral Diameter (50)	78	12.2	1.7	46	13.3	1.3	0.001
Femur							
Maximum Length (60)	82	412.7	22.1	46	438.9	19.3	<0.001
Bicondylar Length (61)	81	409.1	21.8	46	436.6	19.2	<0.001
Distal Articular Breadth	75	62.5	5.2	42	70.6	3.9	<0.001
Epicondylar Breadth (62)	78	70.5	3.5	46	77.9	3.2	<0.001
Subtrochanteric AP Diameter (64)	89	23.1	2.0	51	25.2	2.3	<0.001
Subtrochanteric ML Diameter (65)	89	27.4	2.4	51	29.6	2.4	<0.001
Maximum Head Diameter (63)	84	39.6	2.3	44	44.7	2.8	<0.001
Head Circumference	83	125.5	7.3	40	143.0	7.8	<0.001
Anteroposterior Diameter (66)	76	25.1	2.2	42	28.1	2.3	<0.001
Mediolateral Diameter (67)	76	23.6	1.6	42	25.6	2.0	<0.001
Tibia							
Maximum Length	83	348.7	19.1	48	373.4	20.6	<0.001
Length w/o Eminences	74	345.1	19.7	38	368.9	17.7	<0.001
Anteroposterior Diameter	78	26.1	2.1	42	29.6	2.1	<0.001
Mediolateral Diameter	78	18.5	1.7	41	20.6	2.5	<0.001
Physiological Length	77	330.9	19.5	40	351.2	16.3	<0.001
Fibula							
Maximum Length (75)	79	338.6	19.1	45	359.0	16.7	<0.001

for the purposes of this paper, the designation of "correct" assessment should be read as "in concordance with previous sex estimates based on other skeletal morphological or metric features".

Discriminant function analysis was performed using stepwise procedure (0.05 to enter; 0.10 to exit) and leave-one-out cross-validation for all variables included in this analysis to identify potential equations that have higher classification rates than the simple sectioning points reported below. These analyses were performed for all 33 variables as a unit and then individually by element. The multivariate equation includes the clavicle, humerus, ulna, femur, and tibia, and would be useful when the skeleton is largely complete, but does not include pelvic or cranial remains, or those remains are too fragmentary or otherwise taphonomically damaged to be useful for sex estimation. Discriminant function analyses were also performed on each individual element to identify equations that may be better classifiers than sectioning points when a single element is present and complete. The sectioning point for the multivariate equations is 0 for all.

Results

The first step was to demonstrate variation between males and females within the STC_{DS}. Independent samples t-tests show that for all variables included in this analysis, the male mean is larger than the female mean and the difference is statistically significant ($p \leq 0.008$ for all variables). **Table 2** summarizes these data and provides the p-value for each comparison. It is appropriate to note here that due to preservation and completeness, few individuals have every metric variable recorded. In each table, the n presented is the number of recorded individuals for that metric, not the number of individuals in the total sample.

Table 3 summarizes data on the sexual dimorphism score and the sectioning point between males and females for univariate techniques. **Table 3** also presents data on the percentage of individuals correctly classified to sex based on univariate analysis of the individual metric under consideration using the sectioning point. Correct classification for the univariate tests ranged from 63.2–89.4% on the cross-validated STC_{DS}.

Table 4 reports the results of overall classification for the STC_{TS} and the NTC_{TS} using sectioning points under the same classification protocol as described above, along with the number of individuals included in the assessment for each sample. Both samples exhibit classification rates consistent with the STC_{DS}, with the STC_{TS} ranging from 52.5–95.2% and the NTC_{TS} ranging from 63.8–100.0%. The classification rates for each sample (STC_{DS}, STC_{TS}, NTC_{TS}) were compared using Fisher's Exact Test and those results are reported in **Table 4** as well. Due to the large number of comparisons made, a Bonferroni correction (α=0.05/33=0.002) was applied in order to take the most conservative approach avoiding Type I (false positive) errors.

The rates of correct classification for the three samples was not statistically significantly different for any of the univariate sectioning points provided ($p \geq 0.007$ for all univariate sectioning points).

Table 3. Sexual dimorphism values, sectioning point, and percent correctly classified using individual variables on STC_{DS}; sexual dimorphism calculated as $(x_f - x_m)/(s_f + s_m)$, where x = mean and s = standard deviation; values below sectioning point indicate female, values above indicate male; overal % correct calculated as (total correct/total observed).

Measurement	Sexual Dimorphism	Sectioning Point	% Female Correct	% Male Correct	Overall % Correct
Clavicle					
Maximum Length	0.50	139.9	73.9	65.7	71.2
Anteroposterior Diameter	0.71	11.1	85.1	72.2	80.6
Superoinferior Diameter	0.73	9.9	79.1	80.6	79.6
Humerus					
Maximum Length	0.52	299.4	69.9	74.4	71.4
Distal Articular Breadth	0.68	39.9	82.4	75.0	80.0
Epicondylar Breadth	1.13	58.1	88.4	88.6	88.5
Circumference at Deltoid Tuberosity	0.39	62.1	74.7	55.8	68.0
Max Diameter at 50%	0.27	20.1	69.5	51.2	63.2
Min Diameter at 50%	0.36	15.2	69.5	51.2	63.2
Vertical Head Diameter	1.12	41.0	86.7	84.4	85.9
Radius					
Maximum Length	0.87	235.9	77.4	84.4	79.8
Anteroposterior Diameter	0.72	10.9	75.0	75.0	75.0
Mediolateral Diameter	0.49	13.4	71.4	71.1	71.3
Ulna					
Maximum Length	0.83	256.8	80.0	83.0	81.1
Physiological Length	0.77	225.6	76.5	76.1	76.4
Anteroposterior Diameter	0.61	13.0	75.6	63.0	71.0
Mediolateral Diameter	0.34	12.7	62.8	69.6	65.3
Femur					
Maximum Length	0.63	425.8	72.0	73.9	72.7
Bicondylar Length	0.67	422.9	74.1	73.9	74.0
Distal Articular Breadth	0.89	66.6	77.3	85.7	80.3
Epicondylar Breadth	1.09	74.2	87.2	87.0	87.1
Subtrochanteric AP Diameter	0.48	24.2	74.2	64.7	70.7
Subtrochanteric ML Diameter	0.47	28.5	65.2	62.7	64.3
Maximum Head Diameter	1.00	42.1	92.9	79.5	88.3
Head Circumference	1.16	134.2	91.6	85.0	89.4
Anteroposterior Diameter	0.67	26.6	77.6	73.8	76.3
Mediolateral Diameter	0.53	24.6	71.1	73.8	72.0
Tibia					
Maximum Length	0.62	361.0	74.7	72.9	74.0
Length w/o Eminences	0.64	357.0	71.6	73.7	72.3
Anteroposterior Diameter	0.83	27.8	76.9	78.6	77.5
Mediolateral Diameter	0.48	19.6	76.9	63.4	72.3
Physiological Length	0.57	341.0	72.7	70.0	71.8
Fibula					
Maximum Length	0.57	348.8	73.4	71.1	72.6

The multivariate equations presented in **Table 5** all have higher cross-validated classification rates than any of the individual variables using sectioning points for that element on the STC_{DS} (81.3–92.6%). The equations performed equally well on the STC_{TS} (80.6–96.3%) and the NTC_{TS} (78.3–100.0%). Classification rates using the

Table 4. Overall percent correctly classified using sectioning points reported in **Table 3** on the STC_{TS} and the NTC_{TS}; χ^2 and p-values reported compare classification rates of STC_{DS} (see **Table 3**) to the STC_{TS} and the NTC_{TS}; alpha=0.002 with Bonferroni correction.

Measurement	STC_{TS}		NTC_{TS}			
	n	**Overall % Correct**	**n**	**Overall % Correct**	χ^2	**p-value**
Clavicle						
Maximum Length	31	90.3	8	75.0	4.75	0.093
Anteroposterior Diameter	31	80.6	8	100.0	1.90	0.387
Superoinferior Diameter	31	64.5	8	100.0	5.58	0.062
Humerus						
Maximum Length	39	87.2	47	63.8	6.08	0.048
Distal Articular Breadth	36	88.9	37	89.2	2.62	0.270
Epicondylar Breadth	42	85.7	45	88.9	0.27	0.875
Circumference at Deltoid Tuberosity	40	52.5	49	83.7	10.02	0.007
Max Diameter at 50%	38	60.5	46	76.1	2.99	0.224
Min Diameter at 50%	38	84.2	46	80.4	8.87	0.012
Vertical Head Diameter	45	84.4	48	81.3	0.59	0.744
Radius						
Maximum Length	29	69.0	37	75.7	1.69	0.431
Anteroposterior Diameter	29	89.7	36	88.9	5.42	0.067
Mediolateral Diameter	29	62.1	36	80.6	2.72	0.256
Ulna						
Maximum Length	31	74.2	33	75.8	0.98	0.612
Physiological Length	31	77.4	33	72.7	0.24	0.888
Anteroposterior Diameter	31	64.5	33	81.8	2.49	0.288
Mediolateral Diameter	31	80.6	33	69.7	2.73	0.256
Femur						
Maximum Length	21	76.2	32	68.8	0.37	0.832
Bicondylar Length	22	77.3	33	75.8	0.13	0.938
Distal Articular Breadth	20	80.0	20	95.0	2.58	0.276
Epicondylar Breadth	21	95.2	34	85.3	1.33	0.514
Subtrochanteric AP Diameter	29	69.0	37	83.8	2.78	0.249
Subtrochanteric ML Diameter	30	66.7	28	89.3	6.76	0.034
Maximum Head Diameter	27	88.9	34	85.3	0.26	0.879
Head Circumference	27	92.6	36	83.3	1.51	0.470
Anteroposterior Diameter	21	61.9	37	83.8	3.53	0.171
Mediolateral Diameter	21	66.7	37	91.9	7.04	0.030
Tibia						
Maximum Length	43	81.4	48	70.8	1.43	0.489
Length w/o Eminences	36	83.3	48	70.8	2.04	0.360
Anteroposterior Diameter	35	85.7	46	78.3	1.13	0.568
Mediolateral Diameter	35	71.4	47	74.5	0.11	0.945
Physiological Length	35	82.9	49	73.5	1.73	0.420
Fibula						
Maximum Length	39	79.5	32	71.9	0.81	0.667

equations presented in **Table 5** on the STC_{TS} and the NTC_{TS} are presented in **Table 6**, along with the results of Fisher's Exact Tests to compare the overall classification rates for the cross-validated STC_{DS}, STC_{TS}, and NTC_{TS} ($p \geq 0.260$ for all equations).

Discussion and conclusion

The metric analysis of postcranial skeletal remains has been demonstrated previously to be superior in classifying individuals by sex when compared to the overall assessment of cranial morphology (Spradley & Jantz 2011). Applying metric formulae developed for sex estimation in one population to a different population has been demonstrated to have less discriminatory power than when applied to the original population (c.f., Işcan et al. 1998; King et al. 1998). Therefore, in areas where the bioarchaeological analysis of human remains is likely to include poorly preserved, in-

Table 5. Multi-element and single element multivariate equations for sex estimation and cross-validated accuracy rate from STC_{DS}.

Element(s)	Equation	n	% correct
Multivariate	$(-0.3869*Clav_{Max})+(-0.4032*Hum_{Max})+(-2.7407*Hum_{MaxDiam})+(3.0205*Hum_{HeadDiam})+(0.8175*Ulna_{Max})+(-1.4840*Fem_{STMLdiam})+(2.0994*Tibia_{APdiam})-124.7818$	65	92.3
Clavicle	$(0.9941*Clav_{APDiam})+(1.3206*Clav_{SIDiam})-24.8158$	103	87.4
Humerus	$(0.5027*Hum_{Epicond})+(-0.9671*Hum_{MaxDiam})+(0.9668*Hum_{HeadDiam})-50.1962$	122	92.6
Radius	$(0.1001*Rad_{Max})+(0.9406*Rad_{APDiam})-34.5389$	128	81.3
Ulna	$(0.1031*Ulna_{Max})+(0.7904*Ulna_{APDiam})+(0.6056*Ulna_{MLDiam})-44.9761$	124	84.7
Femur	$(0.2940*Fem_{HeadCirc})-40.2109$	123	89.6
Tibia	$(0.0420*Tibia_{WithoutE})+(0.6560*Tibia_{APDiam})-33.7880$	110	81.8

Table 6. Fisher's Exact Test comparing classification rates between cross-validated STC_{DS} accuracy (see **Table 4**) and STC_{TS} and NTC_{TS} samples, reporting number observed for each test group, overall percentage correct classification, Pearson's χ^2 value and p-value.

Equation	STC_{TS}		NTC_{TS}		Pearson's χ^2	p-value
	n	% Correct	n	% Correct		
Multivariate	8	87.5	2	100.0	2.116	0.549
Clavicle	31	80.6	8	100.0	2.318	0.509
Humerus	38	84.2	41	87.8	4.022	0.260
Radius	29	86.2	36	86.1	0.854	0.837
Ulna	31	87.1	33	84.8	0.126	0.989
Femur	27	96.3	36	83.3	2.585	0.460
Tibia	35	85.7	46	78.3	0.806	0.848

complete, and/or inappropriately curated skeletal individuals, the development of geographically and temporally specific metric tools for estimating sex is highly valuable to the overall analysis of the human condition.

This study provides novel geographically and temporally specific metrics for the assessment of sex in New Kingdom Egyptian skeletal remains from the 18th Dynasty capital city of Tell el-Amarna. These techniques have high classification rates (up to 89.4% using univariate methods and 92.6% using multivariate methods), are simple to use, use metric data collected as part of a standard osteological analysis, and can be applied even if the individual is represented by a single element. The techniques reported herein have similarly high correct classification rates when applied to the two test samples.

Performance of these metric evaluation techniques on the two test samples is particularly promising, although further testing will be required to support the suggestion that these equations and sectioning points are directly applicable throughout ancient Egypt's temporal and geographic extent. However, the author does suggest researchers would be better off using these equations over those developed on modern American skeletal samples, even without testing.

Acknowledgements

The excavations at the Amarna cemeteries are directed by Anna Stevens and are performed under permits from the Ministry of Antiquities of Egypt as part of the Amarna Project, which is directed by Barry Kemp. Thanks also go to the team of excavators who work diligently to excavate the skeletal remains included in this project. The South Tombs Cemetery project was funded by the British Academy, National Geographic, King Fahd Center for Middle East Studies (University of Arkansas), the McDonald Institute for Archaeological Research, Amarna Research Foundation, Michela Schiff-Giorgini Foundation, Seven Pillars of Wisdom Trust, Robert Kiln Trust, and public donations to the Amarna Trust. The North Tombs Cemeteries project was funded by the National Endowment for the Humanities. Any views, findings, conclusions, or recommendations expressed here are exclusively those of the author and do not necessarily represent those of the National Endowment for the Humanities. Additional thanks are extended to the reviewers of this manuscript who dedicated their time and attentions to making it a better piece of science. All errors, however, remain my own.

References

Acsádi J., Nemeskéri J. (1970), *History of human life span and mortality*, translated by K. Balas, Budapest: Akadémiai Kiadó.

Auerbach B.M., Ruff C.B. (2004), *Human body mass estimation: A comparison of "morphometric" and "mechanical" methods*, American Journal of Physical Anthropology 125:331–342

Auerbach B.M., Ruff C.B. (2006), *Limb bone bilateral asymmetry: Variability and commonality among modern humans*, Journal of Human Evolution 50:203–218.

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

Byrd, J.E., Adams B.J. (2003), *Osteometric sorting of commingled human remains*, Journal of Forensic Sciences 48:717–724.

Dabbs G.R. (2010), *Sex determination using the scapula in New Kingdom skeletons from Tell El-Amarna*, HOMO—Journal of Comparative Human Biology 61:412–420.

Dabbs G.R. (2019), *Preliminary results from the North Tombs Cemetery at Tell el-Amarna*, Bioarchaeology International 3(3):174–186.

Fenwick H. (2003), *Desert survey* [in:] "Tell el-Amarna, 2003", B.J. Kemp (ed.), Journal of Egyptian Archaeology 89:11–12.

Gonzalez-Reimers E., Velasco-Vazquez J., Arnay-de-la-Rosa M., Santolaria-Fernandez F. (2000), *Sex determination by discriminant function analysis of the right tibia in the Prehispanic population of the Canary Islands*, Forensic Science International 108:165–172.

Holliday T.W., Ruff C.B. (1997), *Ecogeographic patterning and stature prediction in fossil hominids: Comment on Feldesman and Fountain*, American Journal of Physical Anthropology 103:137–140.

Işcan M.Y., Loth S.R., King C.A., Shihai D, Yoshino M. (1998), *Sexual dimorphism in the humerus: A comparative analysis of Chinese, Japanese, and Thais*, Forensic Science International 98:17–29.

Kemp B.J. (2012), *The city of Akhenaten and Nefertiti: Amarna and its aspects of antiquity*, London: Thames & Hudson.

Kemp B.J., Stevens A., Dabbs G.R., Zabecki M., Rose J.C. (2013), *Life, death and beyond in Akhenaten's Egypt: Excavating the South Tombs Cemetery at Amarna*, Antiquity 87(335):64–78.

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 51:985–989.

Mall G., Graw M., Gehring K.D., Hubig M. (2000), *Determination of sex from femora*, Forensic Science International 113:221–231.

Marlow E.J. (2016), *Metric sex estimation of ancient Egyptian skeletal remains. Part I: Testing of published methods*, Bioarchaeology of the Near East 10:1–25.

Marlow E.J., Kozieradzka-Ogunmakin I. (2016), *Metric sex estimation of ancient Egyptian skeletal remains. Part II: Testing of new population-specific methods*, Bioarchaeology of the Near East 10:27–46.

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

Moore-Jansen P.H., Ousley S.D., Jantz R.L. (1994), *Data collection procedures for forensic skeletal material*, Knoxville: University of Tennessee, Department of Anthropology Report of Investigations No. 48.

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.

Ruff C.B. (1994), *Morphological adaptation to climate in modern and fossil hominids*, Yearbook of Physical Anthropology 37:65–107.

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

Stevens A. (2017), *Death and the city: The cemeteries of Amarna in their urban context*, Cambridge Archaeological Journal 28:103–126.

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

Symes S.A., Jantz R.L. (1983), *Discriminant function sexing of the tibia*, Paper presented at the 35th Annual Meeting of the American Academy of Forensic Sciences, Cincinnati.

Tomsová J., Shierová Z. (2016), *Skeletal material from Deir El-Medina in the Egyptological collection of the Hrdlička Museum of Man in Prague*, Annals of the Náprstek Museum 37:41–69.