5-2022

Reconsidering scales and the binary in forensic anthropology: a critical analysis of morphoscopic data utilized in sex estimation standards.

Bailey N. Watson
University of Louisville

Reconsidering Scales and the Binary in Forensic Anthropology: A Critical Analysis of Morphoscopic Data Utilized in Sex Estimation Standards

By

Bailey N. Watson

Submitted in partial fulfillment of the requirements

for Graduation *summa cum laude*

University of Louisville

May 2022

ACKNOWLEDGEMENTS

I would first like to express my gratitude for my mentor, Dr. Kathryn Marklein, whom I owe for my interest in anthropology in the first place after taking her introductory forensic anthropology class on a whim my freshman year. Your passion for the work you do is infectious and inspiring, and your regular words and actions of support for me and my academic ventures have been invaluable to me throughout the entirety of my undergraduate career.

I would also like to thank my thesis committee member and professor, Dr. Christopher Tillquist, for perpetually pushing me to do better and for always challenging me in ways that strengthened my critical thinking skills.

Finally, many thanks to my partner, Matthew, for providing me with much needed emotional support, for finding me new authors/research to read, and for always enthusiastically entertaining my lengthy ramblings about the human skeleton.

TABLE OF CONTENTS

ABSTRACT

This research explores sex estimation standards used in forensic anthropology in an effort to further the conversation about forensic anthropology's binary-focused language and methods. Discussions regarding sex estimation methodology are important in light of gender variance in the general population. Presently, there is minimal published research on the identification of gender non-conforming individuals in forensic anthropology.

Two researchers individually assigned scores to features associated with sexual dimorphism in the os coxae according to existing methods for 253 individuals, equally represented by self-reported males and females. These data were statistically analyzed for correlation and overlap between features.

Results mainly point to a high degree of variation among individuals, especially biological males who tended to be represented across most of the possible score values for a given feature. Females showed markedly less variation, likely due to evolutionary constraints on os coxa morphology associated with the ability to give birth.

This variation establishes a need for revised methods for sex estimation, to account for a spectrum of gender variance, especially as it may relate to marginalized non-binary, transgender, and intersex individuals.

1. INTRODUCTION & RESEARCH HYPOTHESIS

This project aims to explore commonly used methods in forensic anthropology for sex estimation with acknowledgement to the broad range of human variability as well as the social importance of gender experience and gender variance. Despite the accuracy and efficacy sex estimation methods have provided forensic anthropologists, well-established and reliable methods in sex estimation for female and male individuals should be scrutinized, not only in light of their intra-sexual variation (Marini et al, 1999) but with regard to the binary reality they uphold and perpetuate. While this research plan does not directly address the issue of inadequate training andmethods development related to identification of non-binary and trans individuals, it is hoped that the results from this project contribute to efforts in forensic anthropology to reconsider, or at least scrutinize, our ideas and practices (language and methods) relating to gender and sex identification and representation.

While many anthropologists support new, inclusive approaches to sex estimation or even the addition of gender estimation on the basis of recovery scene context and skeletal identifiers, research and training are not universal and generally occur on case-by-case bases (Tallman et al., 2021; Kincer, 2021). In tandem with explorations of possible avenues toward identifying trans people and non-binary individuals in the forensic record is also the need to reconsider binary-focused language and methods within forensic anthropology that potentially conceptualize sexual differences as reducible, oppositional, and static (Jones, 2014).

Similar to the pilot study (Watson & Marklein, 2021) in which metric variables from a subsample of the Forensic Anthropology Data Bank were statistically analysis using t-tests and which showed significant differences between cisgender female and male groups (see Tables S2 & S3), statistical analysis of morphoscopic variables in this study is expected to display similar results. There is, however, also expected to be significant overlap in these distributions with the morphoscopic variables as was shown with the metric variables in the pilot study (see Tables S2 & S3).

Additionally, it is anticipated that there will be greater overlap between older individuals, pushing the majority of these individuals closer to middle ranges as opposed to “very feminine” or “very masculine”, as compared with younger individuals. A reduction in the accuracy of features scored in standard sex estimation methods has been noted in older individuals (Lovell, 1989), which could relate to natural degeneration or hormonal changes associated with aging (e.g. menopause). Furthermore, a masculinization effect, in both biological males and females, has been noted in the morphology of the greater sciatic notch of older individuals (Walker, 2005). While these trends do not have a definite cause, it is possible they are related to the reduction of hormones associated with ageing. While such hormone loss has not been studied in relation to morphological characteristics of the skeleton associated with sex estimation, it has been documented to negatively affect the mass of both bone and skeletal muscle (see Horsten et al, 2012). If a difference associated with ages at which hormone level change occurs, it could have implications for future studies regarding the effects of Hormone Replacement Therapy (HRT) on morphological characteristics relevant to sex estimation.

2. BACKGROUND

2.1 Sex Estimation in Forensic Anthropology

A primary goal of forensic anthropology is the positive identification of an individual based on skeletal and contextual analysis (Ubelaker et al, 2018). Skeletal analysis typically includes an estimation of an individual’s biological age, sex, ancestry, and stature; depending on the case, individual variation or trauma analysis can also lend clues to an individual's positive identification.

Current sex estimation methods in forensic anthropology include morphoscopic (nonmetric) and morphometric analyses. The former considers morphological traits of the os coxa and cranial elements, though the os coxa is preferred due to greater accuracy, and traits are generally scaled from “1” (very feminine) to “5” (very masculine) (Buikstra and Ubelaker, 1996; Phenice, 1969). Observable traits are considered together when providing a sex estimation for an individual (Christensen et al, 2014). Morphometric analyses which rely on postcranial and cranial measurements may be analyzed multivariately (e.g., discriminant function analysis) or accordingly to population-specific means and ranges to estimate sex (Dabbs et al., 2010; Moore et al., 2016; Spradley and Jantz, 2011; Tise et al., 2012). Three-dimensional imaging technologies in recent years especially have contributed to the development and validation of sex estimation standards (Djorojevic et al., 2014; Gonzalez et al., 2009).

While the efficacy of these methods within forensic anthropology cannot be emphasized enough, these methods nonetheless are predicated on the existence of a retained sexual dimorphism in Homo sapiens and a subsequent sex binary, a scale from “more female” to “more male” (Christensen et al., 2014). An individual placed in the ambiguous range between the two is considered of “undetermined” sex. Although it is no intention of forensic anthropologists to exclude individuals of a population through these numerically scaled identifications (Tallman et al., 2021), these methods leave little to no room for accurate identification of gender non-conforming individuals and perpetuate a language and practice of sex/gender binary preference in anthropology. Additionally, they lend little room for discussions of biological variation *within* self-ascribed cis individuals.

2.2 Considering Gender Non-conforming Individuals

As violence rates against transgender and gender non-conforming individuals continually increase—2021 being the most violent year on record since the Human Rights Campaign began monitoring these populations in 2013—there is a pressing need for the forensic anthropology field to find new ways to provide representation and justice to individuals who do not fall with the preestablished sex and/or gender binary (Human Rights Campaign, 2022). In addition to high interpersonal violence rates, transgender, gender non-conforming, and intersex individuals, for whom typical sex estimations may be complicated, make up a significant portion of the population. A 2016 study identified at least 0.6% of the U.S. population (roughly 2 million Americans by today's population count) as transgender or gender non-conforming while intersex conditions account for somewhere around 1.7% of live births depending on the criteria for what constitutes an intersex condition (Flores et al., 2016; Intersex Human Rights Australia, 2019).

Though these individuals account for a sizeable percentage of the population, the forensic anthropology field has not made any comprehensive, standardized efforts towards identifying non-binary individuals. The impulse to contextualize individuals within a binary system is understandable as sex estimation methods based on sex/gender binaries are relatively clear, replicable, and yield high statistical accuracy (Krishan et al. 2016; Spradley and Jantz, 2011). However, as with ancestry estimation/biological affinity, another component of the traditional biological profile utilized for its general precision (Spradley and Jantz, 2016; Thomas et al., 2017), forensic anthropologists have an obligation to challenge and decolonize standards practiced in the field that perpetuate classificatory non-realities (Bethard and DiGangi, 2020; see Stull et al., 2021 for response; DiGangi and Bethard, 2021).

2.3 Review of current research

Current forensic anthropological research on topics related to the issues addressed in this study are quite limited and more narrowly focused than the aims of this research. One previous study, for instance, identifies characteristic marks of Facial Feminization Surgery (FFS) by using relevant surgical tools on pig bone. This research was completed in an effort to uncover a potential avenue for identifying transgender women who have undergone such surgeries (Buchanan, 2014). While this research is promising for identifying individuals who have received FFS surgeries, there are still significant limitations to such methodology. Although FFS may be a gender-affirming procedure for transfeminine individuals, it is often either inaccessible to these individuals, especially to younger individuals who may not have the financial capability or social support, or undesired altogether and is therefore not a viable method for standardized identification of these individuals in forensic anthropology. According to a report by the U.S. National Center for Transgender Equality, only around 7% of transwomen have received one or more FFS surgeries while another 43% express a future desire for an FFS surgery (James et al, 2016).

A similar research study evaluates the ability of FORDISC, a program commonly used in forensic anthropology for the purposes of sex and ancestry estimation, to adapt to morphological changes in transwomen after having undergone FFS (Schall et al, 2020). While FORDISC can analyze multiple cranial measurements at once, authors identified areas of the face commonly altered by FFS surgeries such as forehead contouring, rhinoplasty, or jaw contouring to determine whether altering these facial features would affect FORDISC sex and ancestry classifications. Unfortunately, this study did not yield the results anticipated by researchers, as the metric data for the observed transwomen remained largely unchanged by the surgeries, resulting in transwomen still being classified as males. This is likely due to the fact that features associated with biological males are overall larger than those associated with biological females, and FORDISC considers the magnitude of the measurement rather than the measurement's relative size relationship to other features. (Schall et al, 2020).

Most promising is a longitudinal study recently conducted in the Netherlands which addresses the effects of HRT on bone geometry. This research produced revolutionary results showing that, if started in early puberty, HRT has the ability to alter os coxae morphology to align with the experienced gender (Van der Loos et al., 2021). This implies that bone morphology can be affected by hormone exposure. It is likely that additional longitudinal studies will be key to understanding changes associated with the os coxa in individuals who do not identify with their assigned-at-birth gender as they provide substantial information regarding long-term changes. One approach to tackling the binary structural system involves a straightforward, scrutinizing examination of binary-classified skeletal data. For example, in a summer pilot study,Watson and Marklein (2021) tested the extent of difference (variation and overlap) in skeletal metrics (Table S1) between self-ascribed female and male individuals from a subsample of the Forensic Anthropology Data Bank and found significantly different measurements between sexeswith demonstrable overlap in male and female ranges (Tables S2-S3). While significant differences in all postcranial and cranial measurements (i.e., male measurements significantly larger than females) were expected, the degree of overlap between samples, ranging from 58% to100% postcranially and 84% to 100% cranially, demonstrates how supposed male and female extremes are not entirely distinctive, and that with so much overlap, we may be mis-categorizing people who are classified as one extreme or the other. Ultimately, focusing on theseaverage distributions upholds binary language and dilutes the importance of the variation we see within these two groups (cis-female and cis-male).

3. MATERIALS AND METHODS

3.1 The Bass Collection

The human skeletal sample included in this study is comprised of individuals within the William M. Bass Donated Skeletal Collection, housed in the Forensic Anthropology Center at the University of Tennessee, Knoxville. This collection is extensive with over 1,800 individuals from 1892 to the present and includes self-reported metadata on chronological age and self-identified sex. The Bass Collection contains a significantly higher percentage of older individuals as opposed to younger individuals as is expected of skeletal collections. The individuals in the collection are also overwhelmingly represented by White males (Campanacho et al, 2021).

The sample from the Bass Collection included as much age variation as was feasible within the limits of the collection to allow for the most equal representation of different age demographics possible. Age has been shown to contribute to changes in pelvic morphology (Walker, 2005), so comparable representations of individuals in chronological age-at-death groups (decadal) was prioritized. Equal representation of ancestry reflective of living population demographics, however, was ultimately unattainable; as such, the research sample included 98% self-identified White individuals, i.e., individuals who claim an ethnically American European ancestry. Individuals were only excluded for significant developmental or traumatic anomalies which inhibited accurate scoring of both the right and left os coxae during the recording process. Of 255 individuals, only 2 were removed for these reasons. The final data set included an equal distribution of self-reported females (n=127) and males (n=126) and represented a normal distribution of ages ranging from 21-99 years with the average age of 60 years.

3.2 Scoring procedure

Sex estimation methods included scoring of the greater sciatic notch (scores ranging from 1-5) and the preauricular sulcus (scores ranging from 0-4) as proposed by Buikstra and Ubelaker (1996) (Table 1). Keeping in mind that a score of ‘0’ on the scale for preauricular sulcus represents the absence of this feature, scores of ‘0’ are reported in this research as a score of ‘5’ to maintain a logical linear progression from 1-5 in which a 1 is the most prominent presentation of the preauricular sulcus, a 4 is the least prominent presentation of the preauricular sulcus, and a 5 represents the absence altogether of this feature.

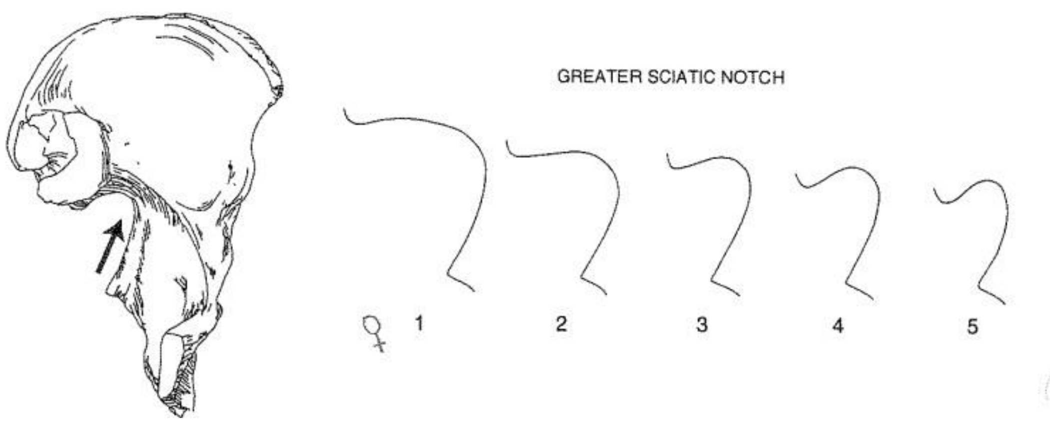

Figure 1. images representing scores 1-5 for the greater sciatic notch (Ubelaker & Buikstra, 1996)

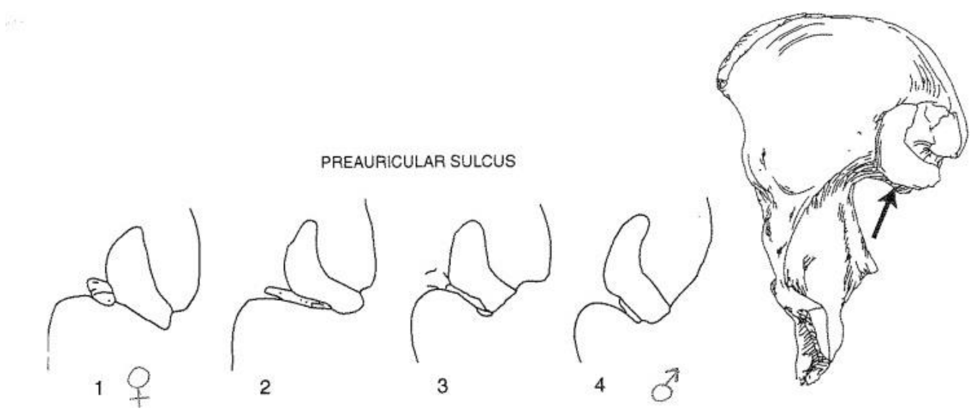

Figure 2. images representing scores 1-4 for the preauricular sulcus (Ubelaker & Buikstra, 1996)

Greater sciatic notch	See fig 1 (no textual descriptions)
Preauricular sulcus	See fig 2 1= wide, typically exceeding 0.5cm, and deep 2= wide (usually greater than 0.5 cm) but shallow 3= well defined but narrow, less than 0.5cm deep 4= narrow (< 0.5cm), shallow, and smooth-walled 0/5= absence of sulcus

Table 1. Greater sciatic notch and preauricular sulcus [Buikstra & Ubelaker, 1996]

Additionally, scoring of the ventral arc, ischiopubic ramus, and subpubic concavity was assigned a score from 1-5 based on the scoring criteria proposed by Klales and colleagues (2012) (Table 2) and a score from 1-3 following a modified scoring derived from presence/ absence method (1 or 2) proposed by Phenice (1969) (Table 3). Rather than giving each individual a 1 or 2 per a true Phenice scoring procedure, individuals were assigned a 1 or 3 in the same fashion, with a score of 2 representing an intermediate score falling somewhere between the "very feminine" (a score of '1' on this modified scale) and "very masculine" (a score of '3' on this modified scale) presentations. Though Phenice (1969) and Klales et al. (2012) scales assign scores to the same features, they are distinct in that the former operates on a very binary understanding of sex while the latter allows for a greater degree of human variability. Given the difference in theoretical approach between these scales, both were considered to assess for difference in correlation. All features were scored according to methodologies outlined in the original publications, taking into account any visual imagery (see figures 1-6) as well as textual descriptions (see tables 1-3) provided by these authors to illustrate score assignment.

If a feature for a given individual was believed to fall somewhere between two of the possible scores (e.g. exhibiting characteristics of both a 2 and a 3), range values were documented (e.g. 2-3) but the score which pushed the individual closer to one end of the scale (in this example, a score of '2') was utilized for final analyses. This decision to "push" individuals toward a binary classification is sometimes practiced in forensic anthropology, with anthropologists estimating individuals as more "female" ranges or "male ranges" depending on which sex feature scoring seems more closely aligned with (Gowland & Knüsel, 2012). Although left os coxae were preferred for visual scoring, per standard forensic practice (Klales et al., 2012), some scores were based on right os coxae when accurate scoring of a specific feature was unclear or inhibited due to traumatic injury or damage to the left os coxa.

Figure 3. Images representing scores 1-5 for ventral arc (Klales et al, 2012)

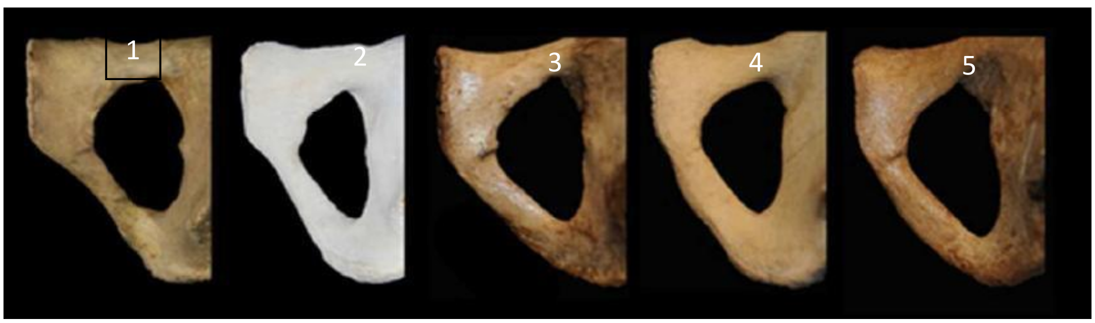

Figure 4. Images representing scores 1-5 for subpubic concavity (Klales et al, 2012)

Figure 5. images representing scores 1-5 for medial aspect of the ischiopubic ramus (Klales et al, 2012)

Subpubic concavity(1-5)	See fig 4 1= Well-developed concavity present inferior to symphyseal face and along length of inferior ramus 2= Slight concavity present inferior to face extended partially down inferior ramus. 3= No concavity present, bone is nearly straight (may be a very slight indentation just below the symphyseal face). 4= Small convexity, especially pronounced along inferior pubic ramus. 5= Large convexity, especially pronounced along inferior pubic ramus
Medial aspect of ischiopubic ramus (1-5)	See fig 5 1= Ascending ramus is narrow dorso-ventrally with a sharp ridge of bone present below the symphyseal face 2= Ascending ramus is narrow dorso-ventrally with a plateau/rounded ridge of bone present below the symphyseal face 3= Ascending ramus is narrow dorso-ventrally with no ridge present. 4= Ascending ramus is medium width dorso-ventrally w/ no ridge present5= Ascending ramus very broad dorso-ventrally w/ no ridge present
Ventral arc (1-5)	See fig 3 1= Arc present at approximately or at least a 40° angle in relation to symphyseal face with a large triangular portion of bone inferiorly placed to arc. 2= Arc present at approximately a 25–40° angle in relation to symphyseal face with a small triangular portion of bone inferiorly placed to arc. 3= Arc present at a slight angle (less than 25°) to the symphyseal face with a slight, nontriangular portion of bone inferiorly placed to arc. 4= Arc present approximately parallel to the symphyseal face with hardly any additional bone present inferior to arc. 5= No arc present (therefore, no additional bone present inferior to the arc)

Table 2. Subpubic concavity (1-5), Medial aspect of ischiopubic ramus (1-5), and Ventral arc (1-5) [Klales et al,2012]

Subpubic concavity (1-3)	Presence= score of “1”/ female (fig 6: A) Absence= score of “3”/ male (fig 6: B)
Medial aspect of ischiopubic ramus (1-3)	Presence= score of “1”/ female (fig 6: C) Absence= score of “3”/ male (fig 6: D)
Ventral arc (1-3)	Presence= score of “1”/ female (fig 6: E) Absence= score of “3”/ male (fig 6: F)

Table 3 (above). Subpubic concavity (1-3), Medial aspect of ischiopubic ramus (1-3), Ventral arc (1-3) [Phenice, 1969]

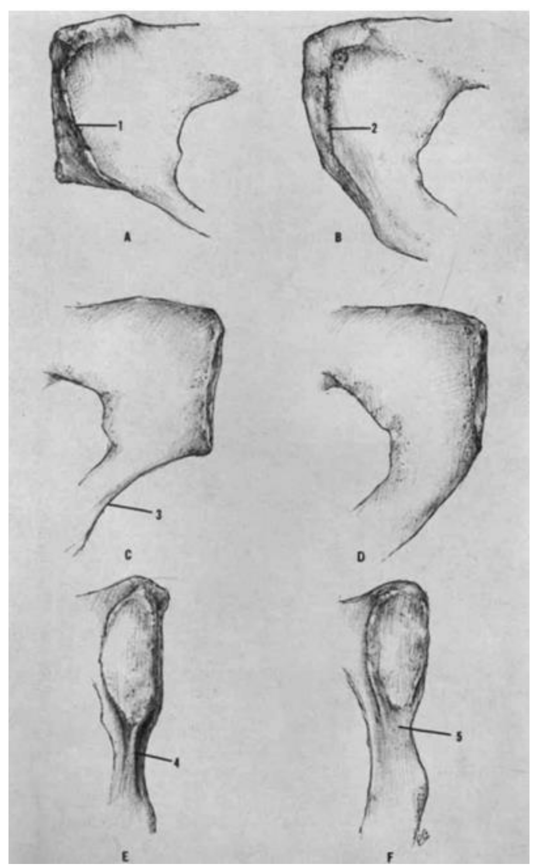

Figure 6 (right). images representing '1' (presence) and'3' (absence) for ventral arc (A&B), subpubic concavity (C&D), and medial aspect of ischiopubic ramus (E&F) (Phenice, 1969)

3.3 Interrater Reliability

Individuals were scored by both researchers (BNW and KEM). To establish reliability between these scores, both researchers individually scored all features using the methods described above for the same 20 individuals. Reliability was evaluated by means of Cohen's Kappa index (Landis & Koch, 1977) to quantify this agreement. Considering the scores assigned are ordinal, weighted kappa values were also highlighted to demonstrate closeness in the instance of a disagreement.

3.4 Descriptive Statistics/ Overlap

Mean, standard deviation, and range were calculated for all feature scores according to self-reported sex. Histograms and scatterplots were generated to provide a visual supplement for discussion of overlap.

3.5 Correlations

Polychoric correlation was employed to assess correlation between scored features. Polychoric correlation tests for correlation between two ordinal variables with the assumption that the variables have similar distributions. Correlations were calculated between preauricular sulcus, greater sciatic notch, and the three features of Phenice and Klales et al methods. 1-5 scores and 1-3 scores of these latter features were tested independently with greater sciatic notch and preauricular sulcus.

4. RESULTS

4.1 Interrater Reliability

Intra-rater agreement was generally high (see Table 4) across scored features (substantial or almost perfect agreement) with the exception of the medial aspect of the ischiopubic ramus following Klales scoring (1-5) which showed only moderate agreement.

	KAPPA	WEIGHTED KAPPA	SE	95% CON. INT
SPC5	0.560	0.772*	0.131	0.303 to 818
MA5	0.186	0.481	0.142	-0.092 to 0.465
VA5	0.588	0.797*	0.123	0.346 to 0.829
SPC3	0.829**	0.852**	0.109	0.615 to 1.000
MA3	0.620*	0.693*	0.139	0.348 to 0.892
VA3	0.844**	0.889**	0.102	0.643 to 1.000
GSN	0.628*	0.822**	0.120	0.393 to 0.864
PAS	0.934**	0.971**	0.063	0.810 to 1.000

Table 4. Results of Cohen's Kappa: SPC5 (subpubic concavity, 1-5 scale), MA5 (medial aspect of ischiopubicramus, 1-5 scale), VA5 (ventral arc, 1-5 scale), SPC3 (subpubic concavity, 1-3 scale), MA3 (medial aspect ofischiopubic ramus, 1-3 scale), VA3 (ventral arc, 1-3 scale), GSN (greater sciatic notch), PAS (preauricular sulcus);
*substantial agreement; ** almost perfect agreement

4.2 Descriptive Statistics/ Overlap

Self-reported females, on average, consistently exhibit lower scores than self-reported males. Overlap in score ranges between females and males occurs in all feature traits (Table 5). However, histograms illustrate a clearer picture of this overlap in regard to number of individuals, as relatively few individualsof the total sample fall within this overlap. The preauricular sulcus (Figure 11) represents the greatest degree of overlap of all scored features.

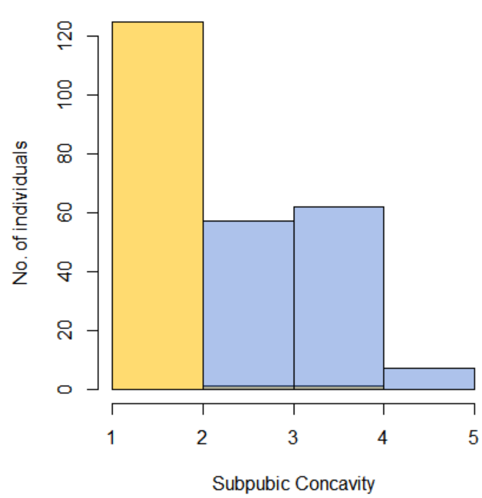

Figure 7. Histogram showing distribution and overlap of scores for subpubic concavity; yellow=females/ blue=males

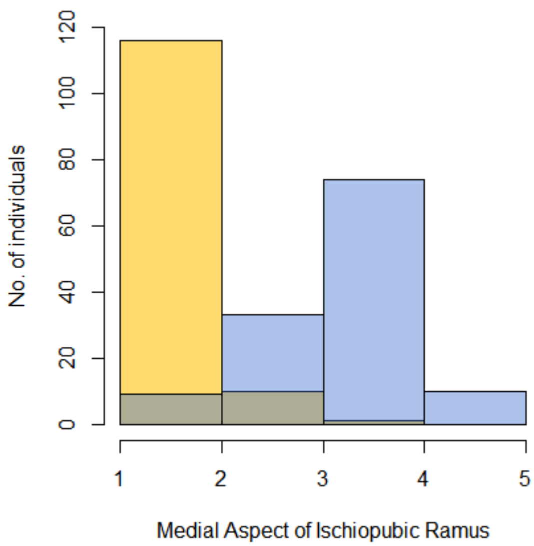

Figure 8. Histogram showing distribution and overlap of scores for medial aspect of ischiopubic ramus; yellow=females/ blue=males

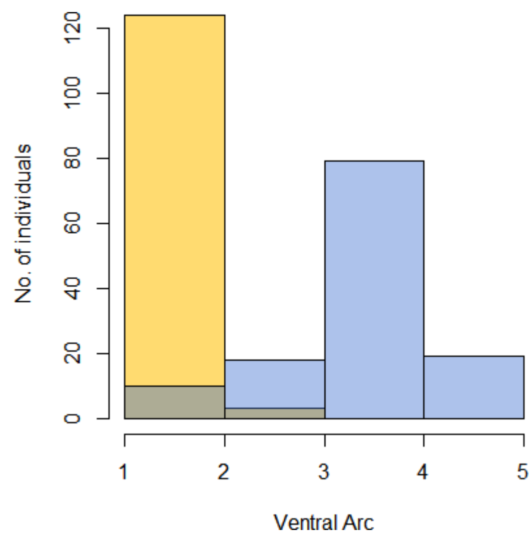

Figure 9. Histogram showing distribution and overlap of scores for ventral arc; yellow=females/ blue=males

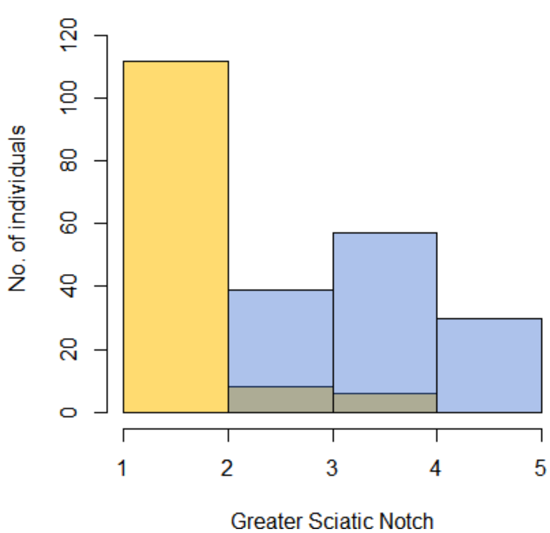

Figure 10. Histogram showing distribution andoverlap of scores for greater sciatic notch; yellow=females/ blue=males

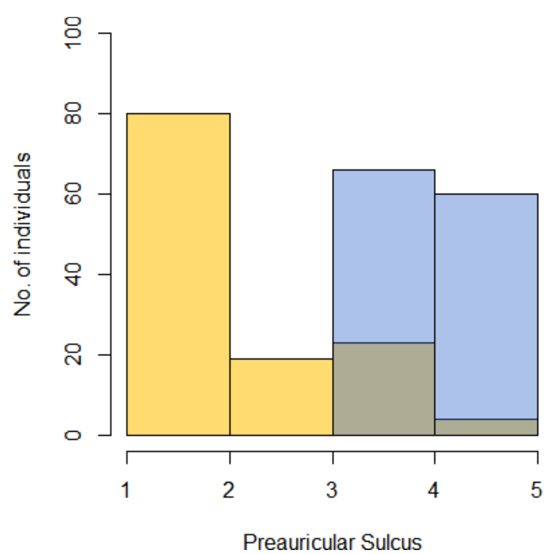

Figure 11. Histogram showing distribution andoverlap of scores for preauricular sulcus; yellow=females/ blue=males

		Females			Males			
		mean	SD	range	mean	SD	range	Overlap range
Klales et al (2012) 1-5	Subpubic Concavity	1.189	0.467	1-4	3.452	0.816	2-5	2-4
	Medial aspect- ischiopubic ramus	1.614	0.667	1-4	3.667	0.748	1-5	1-4
	Ventral arc	1.268	0.495	1-3	3.841	0.794	1-5	1-3
Modified Phenice (1969) 1-3	Subpubic Concavity	1.031	0.216	1-3	2.448	0.546	1-3	1-3
	Medial aspect- ischiopubic ramus	1.150	0.358	1-2	2.524	0.547	1-3	1-2
	Ventral arc	1.063	0.275	1-3	2.603	0.538	1-3	1-3
Standards (1996) *1-5*	Greater Sciatic Notch	1.603	0.811	1-4	3.873	0.839	2-5	2-4
	Preauricular Sulcus	2.230	1.227	1-5	4.444	0.559	3-5	3-5

Table 5. reported values for mean, SD (standard deviation), and range

4.3 Correlations

Results from polychoric correlations are reported in Tables 6-11. Matrices which consider individuals holistically (Tables 6 & 9) as opposed to divided on the basis of self-ascribed sex (Tables 7, 8, 10 & 11) show notably higher correlations overall. Among all individuals, correlations between subpubic concavity and other variables (MA, 0.8056; VA, 0.811; and GSN, 0.790) are markedly higher than other correlation coefficients across all matrices.

Correlations between the preauricular sulcus and other features are highly variable throughout the matrices. In Table 6, the correlation between preauricular sulcus and the medial aspect of the ischiopubic ramus represents the lowest correlation coefficient (0.7170) in this matrix. Meanwhile, in Table 11, the correlation between preauricular sulcus and greater sciatic notch represents the second highest correlation coefficient (0.3074) in this matrix.

Correlations between the medial aspect of the ischiopubic ramus also show a comparable variability. The medial aspect shows negative correlations (GSN, -0.1114 and PAS, -0.1802) in Table 8 yet contributes to the highest positive correlation (0.3553) in the Table 7 matrix.

	MA	SPC	VA	GSN	PAS
MA	1				
SPC	0.8056	1			
VA	0.7992	0.8108	1		
GSN	0.6863	0.7895	0.7530	1	
PAS	0.6214	0.7132	0.7508	0.7170	1

Table 6. Correlation matrix (all individuals) comparing Klales scores (1-5) and Ubelaker and Buikstra's features for individual traits: MA (medial aspect of ischiopubic ramus), SPC (subpubic concavity), VA (ventral arc), GSN(greater sciatic notch), PAS (preauricular sulcus)

	MA	SPC	VA	GSN	PAS
MA	1				
SPC	0.3553	1			
VA	0.2540	0.2046	1		
GSN	0.1172	0.2938	0.1109	1	
PAS	0.0622	-0.0332	0.1663	0.0834	1

Table 7. Correlation matrix (males only) comparing Klales scores (1-5) and Ubelaker and Buikstra's features for individual traits: MA (medial aspect of ischiopubic ramus), SPC (subpubic concavity), VA (ventral arc), GSN(greater sciatic notch), PAS (preauricular sulcus)

	MA	SPC	VA	GSN	PAS
MA	1				
SPC	0.1515	1			
VA	0.1798	0.3870	1		
GSN	-0.1114	0.2386	0.1201	1	
PAS	-0.1802	0.2085	0.1402	0.3074	1

Table 8. Correlation matrix (females only) comparing Klales scores (1-5) and Ubelaker and Buikstra's features for individual traits: MA (medial aspect of ischiopubic ramus), SPC (subpubic concavity), VA (ventral arc), GSN(greater sciatic notch), PAS (preauricular sulcus)

	IRR	SPC	VA	GSN	PAS
MA	1				
SPC	0.9006	1			
VA	0.8725	0.8905	1		
GSN	0.7580	0.8531	0.7927	1	
PAS	0.7428	0.7846	0.7836	0.7170	1

Table 9. Correlation matrix (all individuals) comparing Phenice scores (1-3) and Ubelaker and Buikstra's features for individual traits: MA (medial aspect of ischiopubic ramus), SPC (subpubic concavity), VA (ventralarc), GSN (greater sciatic notch), PAS (preauricular sulcus)

	IRR	SPC	VA	GSN	PAS
MA	1				
SPC	0.4774	1			
VA	0.2416	0.2144	1		
GSN	0.2083	0.3097	0.1240	1	
PAS	0.2084	0.0176	0.0444	0.0834	1

Table 10. Correlation matrix (males only) comparing Phenice scores (1-3) and Ubelaker and Buikstra's features for individual traits: MA (medial aspect of ischiopubic ramus), SPC (subpubic concavity), VA (ventralarc), GSN (greater sciatic notch), PAS (preauricular sulcus)

	IRR	SPC	VA	GSN	PAS
MA	1				
SPC	0.2272	1			
VA	0.0798	-0.8473	1		
GSN	-0.1737	0.7466	-0.1263	1	
PAS	-0.0759	0.1855	0.0381	0.3074	1

Table 11. Correlation matrix (females only) comparing Phenice scores (1-3) and Ubelaker and Buikstra's features for individual traits: MA (medial aspect of ischiopubic ramus), SPC (subpubic concavity), VA (ventralarc), GSN (greater sciatic notch), PAS (preauricular sulcus)

5. DISCUSSION

5.1 Interrater reliability

Overall, relatively high reliability was expected. In no instance during scoring did either researcher assign a score more than one score higher or lower than the other (i.e., if one researcher gave a '2', the other researcher always assigned a '1' or '3' in a case of disagreement). However, reliability for one feature, the medial aspect of the ischiopubic ramus was notably lower across both scoring systems employed. A probable reason for this trend rests in the forensic anthropological practice of pushing individuals towards binaries (Gowland & Knüsel, 2012). Range values were noticeably commonly documented in the scoring of this feature and forcing these ranges towards the more binary value ultimately led to more varied score assignment between researchers. The impulse to assign an individual a range value because they exhibit characteristics of two different scores begs the need for a revision of these scales. A viable future solution may be a mere expansion of the scales to include a greater range of possible scores and therefore allow for a more thorough representation of the vast range of human variability. Expanded scales could possibly result in lower rates of reliability among researchers due to increased numerical options for scoring, so any possible revision of these scales would necessitate substantial testing.

5.2 Overlap

Ranges in which overlap is present comprise a major portion of the possible range values; for instance, for the medial aspect of the ischiopubic ramus (1-5 scale), both males and females are represented between a score of ‘1’ and ‘4’ (Table 5). However, the histogram for the medial aspect of the ischiopubic ramus (Figure 8) show that the number of individuals who fall within this overlap range, is a relatively marginal portion of total individuals.

The preauricular sulcus showed the greatest number of individuals in the overlap range (Figure 11). This is unsurprising as the presentation of the preauricular sulcus is quite a variable feature. During scoring, this feature received the most variable scores of any scored feature, as females were not uncommonly assigned scores of 3-5, which represent more ‘masculine’ presentation of this feature. Furthermore, prior research suggests that preauricular sulcus scoring alone is not a strong enough indicator alone to provide substantiation for a female sex estimate, though the absence of the sulcus entirely (a score of ‘5) has been shown to suggest a male sex estimate (Karsten, 2017). The patterns shown by the distribution of preauricular sulcus scores in this study is consistent with Karsten's findings and others (Dee, 1981), as seen in figure 11 and evident in the range values for the preauricular sulcus (Table 5). Reported females are represented in every possible score assigned to this feature, while males are more evidently confined to the upper parts of the range (associated with minimal to no expression of sulcus). Additionally, a score of '5' representing the absence of this sulcus displays minimal overlap with reported females and is comprised primarily of reported males (Figure 11).

5.3 Correlations

Higher correlations in matrices which consider individuals holistically (Tables 6 & 9) are higher than matrices which consider female and male individuals separately (Tables 7, 8, 10, & 11) due to an expected effect regarding the homogeneity of the samples considered. When biological males and females are considered together, the full range of scores is represented and therefore the sample shows greater variability and heterogeneity. On the other hand, female only and male only groups are less diverse on their own (i.e., increased homogeneity), which results in lower correlations overall. The relatively high positive correlations noted in the entirety of matrices considering males and females holistically supports the current practice of considering these features together rather than individually.

Greater sciatic notch and subpubic concavity seem to be highly correlated, as compared to all correlation coefficients across the matrices which was expected. Both of these features are angular and therefore collectively contribute to overall breadth of the os coxae. Biological females have notably wider os coxa than biological male counterparts to allow for the ability to give birth (González et al, 2016).

The variability of correlations involving the preauricular sulcus seen throughout the matrices is ultimately unsurprising. While the correlations presented here do suggest the uncertainty expressed by Karsten (2017) regarding the assignment of sex based on the preauricular sulcus, given the correlation variability across matrices, it does not substantiate the tendency for this feature to aid in the sex estimation of biological males. Correlations for the preauricular sulcus in male only matrices (Tables 7 & 10) are generally on the lower side compared to other scores in these matrices. This, however, could be explained by the higher variability in males as males tend to encompass a broader range of scores than females across most features (see Table 5).

The variability of correlations involving the medial aspect of the ischiopubic ramus likely, in part, relates to the fact that this feature serves as a muscle attachment site. Specifically, the ischiopubic ramus is an attachment site for muscles involved in the adduction of the lower limbs (Wobser et al, 2021). Muscle attachment sites may not be the most reliable indicators of sex, on their own, since muscles may be affected by life history events and are not restrained by biological sex. Accordingly, Phenice (1969) presents the ischiopubic ramus as the least reliable feature within his scoring system, clarifying that this feature should not be used alone to determine biological sex. Furthermore, lower interrater reliability for the medial aspect of the ischiopubic ramus may have also, in part, contributed to the variability of correlation results.

5.4 Age

Age is a possible confounding variable to sex estimation in this research as changes in hormone production associated with aging may affect os coxa morphology. Walker (2005) argued that “age related increase in sciatic notch sexual dimorphism” associates with a shift towards more a masculine morphology with age (388). This research suggested that individuals who died before age 50 have more feminine morphology than older individuals, an age effect which is more pronounced in males. Therefore, young males are more likely to be incorrectly estimated as female while older females are more likely to be incorrectly estimated as male (Walker, 2005). According to histogram distributions of ages and sciatic notch scores for biological males and females (Figures 12 & 13), this age effect seems to be almost entirely absent. This may relate to the relatively small sample size within decadal groups considered here. However, it is interesting to note that all scores of '4', a number within the more masculine range, do not appear in the females of this sample until age 46 at the earliest, a nod to Walker's (2005) proposition of a masculinization effect with age. Overall, distributions of other features when compared by age suggested a similar lack of age effect. Future research should address possible age effects on morphology change more directly, however, to understand this issue more clearly and help to prevent future incorrect sex estimations.

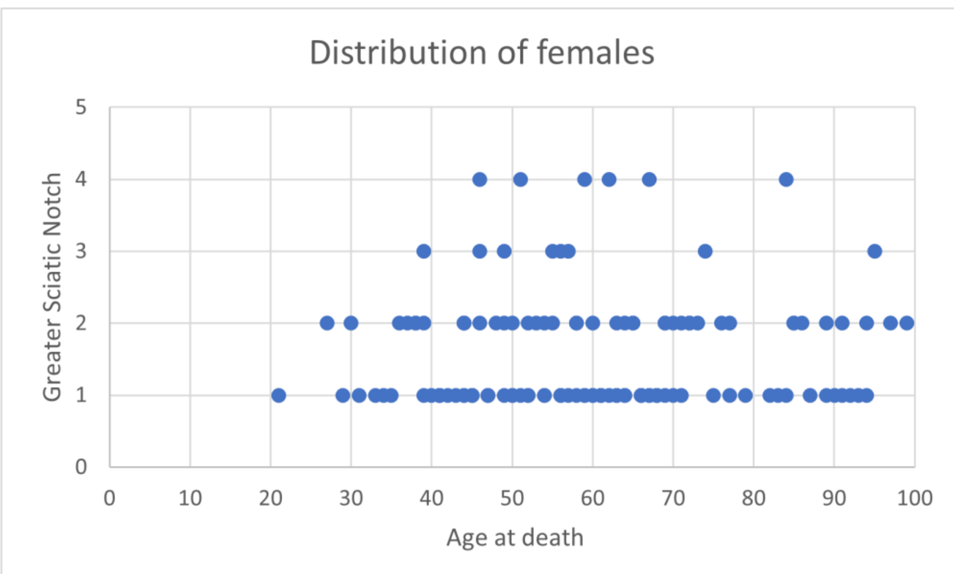

Figure 12. A scatterplot showing the distribution of scores assigned to females for one feature, the greater sciatic notch as it relates to age-at-death.

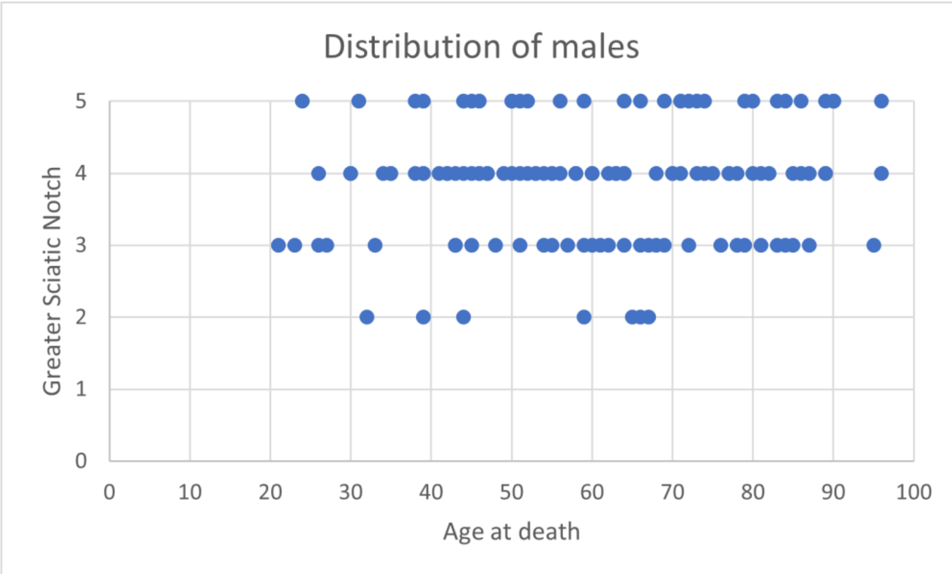

Figure 13. A scatterplot showing the distribution of scores assigned to males for one feature, the greatersciatic notch as it relates to age-at-death.

5.5 Variation

The data considered here holistically points to a need to acknowledge variation among humans, particularly in regard to sex variance. Overlap ranges show that, while minimal, self-reported males and females are not uncommonly assigned scores outside the supposed "feminine" (1-2) and "masculine" (4-5) ranges that exist in sex estimation standards, displaying a range of feature expression among both the male and female sex. The overlap ranges are especially informative with regard to the Phenice (1969) 1-3 scale and Klales et al (2012) 1-5 scale. The former had an overlap range of 1-3 between males and females for two of the three variables scored. If individuals from both sexes are being assigned scores representing every possible score value, then this is surely not a productive way of estimating sex in the skeleton. Phenice scoring shows high correlations between feature scores (Table 9), however this probably has more to do with the forcing of individuals into a category where the options are female, male, or ambiguous.

Variation is greater in males, especially in histogram distributions. In figures 8 and 9 which show distributions for medial aspect of the ischiopubic ramus and ventral arc, for instance, males are represented by every possible feature score value. This constraint of female feature traits is most likely related to the birthing capabilities of biological females; evolutionary processes that maintain female ability to birth children keep them, for the most part, in the lower part of these ranges because certain characteristics like a wider breadth of the os coxa are a requirement for this biological process (González et al, 2016). Therefore, biological males show a wider range of feature scores because they lack this pressing biological need for particularly shaped os coxae (González et al, 2016).

Life history may also play a role in skeletal variation as secular changes in the human skeleton likely associated with lifestyle have been noted in the past two hundred years (Klales, 2016) though the particular mechanisms that drive these changes are currently uncertain. A more physically active lifestyle, for instance, may relate to bone morphology changes (see Infantino et al, 2021); however, whether this is related to the factor of body size as associated with muscle building or weight or alternatively with changes in hormone production or bone density isn't clear without deeper investigation.

The extent of human variation illustrates the critical idea that using these standards which operate under the assumption of a clear, defined sex binary should be scrutinized on a case by case situation. Perhaps a more useful way of thinking about these scales is considering them a scale from more gracile features "1" to more robust features "5" rather than thinking of them as a binary of feminine to masculine. This may allow for individuals to be understood on a spectrum of different levels of feature expression rather than boxed into biological categories that may or may not be informational.

6. CONCLUSION

All things considered, sex estimation standards are frequently accurate in assessing biological sex. The current research study upholds these standards, considering high correlations between all features when individuals are considered holistically rather than divided by sex (Tables 6 & 9). However, as biological sex becomes secondary to personal identity over time, we must question if these methods are providing anthropologists useful information in their application. As the main goal of forensic anthropology is positive identification of individuals, forensic *anthropologists* must apply a biocultural approach that transcends biological sex. This responsibility involves identifying non-binary people as non-binary, transgender people as transgender, and cisgender people as cisgender. Estimating a transwoman's skeleton as biologically male, for instance, does no justice to the individual and can inhibit the positive identification of that individual. If forensic anthropologists want to accurately assume information about life history in death, then they must work towards accurately deciphering this critical social information, whether that be through revised methods and/or context of what an individual is found with. The underlying goals of this exploration into sex estimation methods demonstrates a need for more nuanced biocultural methods and aims to fuel future research on this topic.

REFERENCES

Christensen, A. M., Passalacqua, N. V., & Bartelink, E. J. (2014). *Forensic anthropology: Current methods and practice*. Academic Press.

Christensen, Angi M., and Nicholas V. Passalacqua. (2018). *A Laboratory Manual for Forensic Anthropology*. Elsevier, AP, Academic Press.

Bethard, J. D., & DiGangi, E. A. (2020). Letter to the editor—moving beyond a lost cause: Forensic anthropology and ancestry estimates in the United States. *Journal of Forensic Sciences*, *65*(5), 1791–1792. https://doi.org/10.1111/1556-4029.14513

Buchanan, S. (2014). *Bone modification in male to female transgender surgeries: Considerations for the forensic anthropologist* (thesis). LSU Digital Commons.

Buikstra, J. E., & Ubelaker, D. H. (1996). Standards for data collection from human skeletal remains. *Human Biology*, *68*(1), 155.

Campanacho, V., Alves Cardoso, F., & Ubelaker, D. H. (2021). Documented skeletal collections and their importance in forensic anthropology in the United States. *Forensic Sciences*, *1*(3),228–239. https://doi.org/10.3390/forensicsci1030021

Dabbs, G. R., & Moore-Jansen, P. H. (2010). A method for estimating sex using metric analysis of the scapula. *Journal of Forensic Sciences*, *55*(1), 149–152. https://doi.org/10.1111/j.1556-4029.2009.01232.x

Dee, P. M. (1981). The preauricular sulcus. *Radiology*, *140*(2), 354–354. https://doi.org/10.1148/radiology.140.2.7255707

DiGangi, E. A., & Bethard, J. D. (2021). Uncloaking a lost cause: Decolonizing ancestry estimation in the United States. *American Journal of Physical Anthropology*. https://doi.org/10.1002/ajpa.24212

Djorojevic, M., Roldán, C., García-Parra, P., Alemán, I., & Botella, M. (2014). Morphometricsex estimation from 3D computed tomography OS coxae model and its validation in skeletal remains. *International Journal of Legal Medicine*, *128*(5), 879–888. https://doi.org/10.1007/s00414-014-1033-x

Flores, A. R., Herman, J. L., Gates, G. G., & Brown, T. N. T. (2016). *How many adults identify* *as transgender in the United States?* Williams Institute. Retrieved from https://williamsinstitute.law.ucla.edu/publications/trans-adults-united-states/

González, N. C., Pérez, J. R., Chamero, B., Cambra-Moo, O., & Martín, A. G. (2016). Geometric morphometrics reveals restrictions on the shape of the female os coxae. *Journal of Anatomy*, *230*(1), 66–74. https://doi.org/10.1111/joa.12528

Gonzalez, P. N., Bernal, V., & Perez, S. I. (2009). Geometric morphometric approach to sex estimation of human pelvis. *Forensic Science International*, *189*(1-3), 68–74. https://doi.org/10.1016/j.forsciint.2009.04.012

Gowland, R., & Knüsel Christopher. (2012). *Social Archaeology of funerary remains*. Oxbow Books.

Horstman, A. M., Dillon, E. L., Urban, R. J., & Sheffield-Moore, M. (2012). The role of androgens and estrogens on healthy aging and longevity. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *67*(11), 1140–1152. https://doi.org/10.1093/gerona/gls068

Human Rights Committee. (2022). *Fatal violence against the transgender and gender non-conforming community in 2022*. Human Rights Campaign. Retrieved March 1, 2022, from https://www.hrc.org/resources/fatal-violence-against-the-transgender-and-gender-non-conforming-community-in-2022

Infantino, N. A., McCormack, W. P., & Almstedt, H. C. (2021). Bone mineral density and hip structure changes over one-year in collegiate distance runners and non-athlete controls. *Bone Reports*, *14*, 101056. https://doi.org/10.1016/j.bonr.2021.101056

Intersex Human Rights Australia. (2019). *Intersex population figures*. Retrieved from https://ihra.org.au/16601/intersex-numbers/.

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. (2016). *The Report of the 2015 U.S. Transgender Survey*. Washington, DC: National Center for Transgender Equality.

Jones, G. (2014). *Not a yes or no question: critical perspectives on sex and gender in forensic anthropology*. Electronic Theses and Dissertations. 5246. Department of Sociology, Anthropology, and Criminology. University of Windson

Karsten, J. K. (2017). A test of the preauricular sulcus as an indicator of sex. *American Journalof Physical Anthropology*, *165*(3), 604–608. https://doi.org/10.1002/ajpa.23372

Kincer, C. D. (2020). *Centering transgender personhoods in forensic anthropology and expanding sex estimation in casework and research* (thesis). OpenBU, Boston.

Klales, A. R. (2016). Secular change in morphological pelvic traits used for sex estimation. *Journal of Forensic Sciences*, *61*(2), 295–301. https://doi.org/10.1111/1556-4029.13008

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*(1), 104–114. https://doi.org/10.1002/ajpa.22102

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, *261*. https://doi.org/10.1016/j.forsciint.2016.02.007

Landis, J.R.; Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics. 33 (1): 159-174. https://doi.org/10.2307%2F2529310

Lovell, N. C. (1989). Test of Phenice's technique for determining sex from the OS pubis. *American Journal of Physical Anthropology*, *79*(1), 117–120. https://doi.org/10.1002/ajpa.1330790112

Marini, E., Racugno, W., & Borgognini Tarli, S. M. (1999). Univariate estimates of sexual dimorphism: The effects of intrasexual variability. *American Journal of Physical Anthropology*, *109*(4), 501–508. https://doi.org/10.1002/(sici)1096-8644(199908)109:4<501::aid-ajpa6>3.0.co;2-7

Moore, M. K., DiGangi, E. A., Niño Ruíz, F. P., Hidalgo Davila, O. J., & Sanabria Medina, C. (2016). Metric sex estimation from the postcranial skeleton for the Colombian population.*Forensic Science International*, *262*. https://doi.org/10.1016/j.forsciint.2016.02.018

Phenice, T. W. (1969). A newly developed visual method of sexing the OS pubis. *American Journal of Physical Anthropology*, *30*(2), 297–301. https://doi.org/10.1002/ajpa.1330300214

Schall, J. L., Rogers, T. L., & Deschamps-Braly, J. C. (2020). Breaking the binary: The identification of Trans-women in forensic anthropology. *Forensic Science International*, *309*, 110220. https://doi.org/10.1016/j.forsciint.2020.110220

Spradley, K. M., & Jantz, R. L. (2016). Ancestry estimation in forensic anthropology: Geometric morphometric versus standard and nonstandard Interlandmark distances. *Journal of Forensic Sciences*, *61*(4), 892–897. https://doi.org/10.1111/1556-4029.13081

Spradley, M. K., & Jantz, R. L. (2011). Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences*, *56*(2), 289–296. https://doi.org/10.1111/j.1556-4029.2010.01635.x

Stull, K. E., Bartelink, E. J., Klales, A. R., Berg, G. E., Kenyhercz, M. W., L'Abbé, E. N., Go, M. C., McCormick, K., & Mariscal, C. (2020). Commentary on: Bethard JD, Digangi EA.letter to the editor—moving beyond a lost cause: Forensic anthropology and ancestry estimates in the United States. J forensic sci. 2020;65(5):1791–2. doi: 10.1111/1556-4029.14513. *Journal of Forensic Sciences*, *66*(1), 417–420. https://doi.org/10.1111/1556-4029.14616

Tallman, S. D., Kincer, C. D., & Plemons, E. D. (2021). Centering transgender individuals in forensic anthropology and expanding binary sex estimation in casework and research. *Forensic Anthropology*, 1–20. https://doi.org/10.5744/fa.2020.0030

Thomas, R. M., Parks, C. L., & Richard, A. H. (2017). Accuracy rates of ancestry estimation by forensic anthropologists using identified forensic cases. *Journal of Forensic Sciences*, *62*(4), 971–974. https://doi.org/10.1111/1556-4029.13361

Tise, M. L., Spradley, M. K., & Anderson, B. E. (2012). Postcranial sex estimation of individuals considered Hispanic. *Journal of Forensic Sciences*, *58*. https://doi.org/10.1111/1556-4029.12006

Ubelaker, D. H., Shamlou, A., & Kunkle, A. (2018). Contributions of forensic anthropology to Positive Scientific Identification: A critical review. *Forensic Sciences Research*, *4*(1), 45–50. https://doi.org/10.1080/20961790.2018.1523704

Van der Loos, M. A., Hellinga, I., Vlot, M. C., Klink, D. T., den Heijer, M., & Wiepjes, C. M. (2021). Development of Hip Bone Geometry During Gender-Affirming Hormone Therapy in Transgender Adolescents Resembles That of the Experienced Gender When Pubertal Suspension Is Started in Early Puberty. *Journal of Bone and Mineral Science*, *36*(5), 931–941. https://doi.org/10.1002/jbmr.4262

Walker, P. L. (2005). Greater Sciatic Notch Morphology: Sex, age, and population differences. *American Journal of Physical Anthropology*, *127*(4), 385–391. https://doi.org/10.1002/ajpa.10422

Watson, B., & Marklein, K.E. (2021). Transcending the binary in forensic anthropology: a critical analysis of metric data utilized in sex estimation standards. Poster presented at 2021 Summer Research Opportunity Program Poster Presentation, Louisville, KY.

Wobser, A. M., Adkins, Z., & Wobser, R. W. (2021). *Anatomy, abdomen and pelvis, bones (ilium, ischium, and pubis)*. National Center for Biotechnology Information. Retrieved March 23, 2022, from https://pubmed.ncbi.nlm.nih.gov/30137809/

SUPPLEMENTARY MATERIALS

Postcranial measurements	**Cranial measurements**
CLAVXLN (L) Maximum clavicular length (left)	**WFB** (Minimum frontal breadth)
HUMHDD (L) Maximal humeral head diameter (left)	**ZYB** (Bizygomatic breadth)
FEMXLN (L) Maximum femoral length (left)	**NLH** (Nasal height)
FEMHDD (L) Maximal femoral head diameter (left)	**GOL** (Maximum cranial length)
OSCOXHT (L) Os coxa height (left)	
ILIABR (L) Iliac breadth (left)	

Table S1. Measurements assessed univariately in pilot study

Postcranial measurements		**Mean (in mm)**	**Range (in mm)**	**% overlap**	**P-value**
CLAVXLN (L)	**F**	140.70	121-165	97.94%	<0.001
	M	158.04	126-183	100%	
HUMHDD (L)	**F**	42.31	35-49	100%	<0.001
	M	48.86	22-60	58.48%	
FEMXLN (L)	**F**	437.75	352-514	94.49%	<0.001
	M	475.21	403-566	94.70%	
FEMHDD (L)	**F**	42.32	35-50	90.71%	<0.001
	M	48.35	40-60	80%	
OSCOXHT (L)	**F**	201.51	75-235	99.31%	<0.001
	M	222.54	159-265	88.76%	
ILIABR (L)	**F**	155.62	96-179	99.77%	<0.001
	M	161.28	127-194	96.55%	

Table S2. Mean and range data for biological male and female postcranial measurements analyzed in study. P-values reflect results from Student's t-tests. Percent overlap captures the percent of individuals within the female/male sample that fall within the range of the other female/male group.

Cranial measurements		**Mean (in mm)**	**Range (in mm)**	**% overlap**	**P-value**
WFB	**F**	93.61	81-122	99.74%	<0.001
	M	96.68	80-113	99.84%	
ZYB	**F**	121.63	108-135	98.83%	<0.001
	M	130.36	112-149	83.65%	
NLH	**F**	49.02	34-104	99.07%	<0.001
	M	52.67	43-64	100%	
GOL	**F**	177.79	155-196	98.39%	<0.001
	M	187.89	164-211	86.26%	

Table **S3.** Mean and range data for biological male and female cranial measurements analyzed in study. P-values reflect results from Student's t-tests. Percent overlap captures the percent of individuals within the female/male sample that fall within the range of the opposite female/male group.