A Revised Method of Sexing the Human Innominate Using Phenice's Nonmetric Traits and Statistical Methods

Alexandra R. Klales,[1]* Stephen D. Ousley,[2] and Jennifer M. Vollner[3]

[1]*Department of Anthropology, University of Manitoba, Manitoba, MB R3T 5V5, Canada*
[2]*Department of Applied Forensic Sciences, Mercyhurst University, Erie, PA 16546*
[3]*Department of Anthropology, Michigan State University, East Lansing, MI 48824*

KEY WORDS phenice traits; pelvis; sex estimation; nonmetrics

ABSTRACT The traits of the pubis described by Phenice (Am J Phys Anthropol 30 (1969) 297–302) have been used extensively by physical anthropologist for sex estimation. This study investigates all three of Phenice's characteristics in an approach similar to Walker's (Am J Phys Anthropol 136 (2008) 39–50) study using observations from the cranium and mandible. The ventral arc, the subpubic contour, and the medial aspect of the ischio-pubic ramus were scored on a five-point ordinal scale from a sample of 310 adult, left innominates of known ancestry and sex from the Hamann-Todd Human Osteological Collection and the W.M. Bass Donated Skeletal Collection. Four observers with varying levels of experience blindly scored each trait using new descriptions and illustrations adapted from those originally created by Phenice. The scores were then analyzed with ordinal logistic regression. Using all three traits for sex classification, the mean correct classification rate was 94.5% cross-validated for experienced observers. Intra- and interobserver error in trait scoring was low for all three traits and agreement levels ranged from moderate to substantial. Tests of the method on an independent validation sample provided a classification accuracy of 86.2%. This revision of the Phenice (Am J Phys Anthropol 30 (1969) 297–302) technique is a reliable and valid method of sex estimation from the human innominate that meets the *Daubert* criteria for court admissibility. Am J Phys Anthropol 149:104–114, 2012. ©2012 Wiley Periodicals, Inc.

Accurate sex estimation is critical for forensic anthropologists establishing the biological profile of unidentified individuals and also for bioarcheologists recreating the demographics of past populations. Historically, the bones most frequently used for sex estimation have been the cranium and the bones of the pelvis. The innominate and the pubic bone in particular, has generally been accepted as the single best indicator of sex (Letterman, 1941; Phenice, 1969; Stewart, 1979; Krogman and Işcan, 1986; MacLaughlin and Bruce, 1986; Walker, 2005) due to differences between males and females related to childbirth and locomotion. Because of these aforementioned differences, the innominate has been examined previously through multiple metric and nonmetric studies to develop methods of sex estimation for use by physical anthropologists.

Numerous metric approaches have been developed for sex estimation in the innominate (Day and Pitcher-Wilmott, 1975; DiBennardo and Taylor, 1983; Patriquin et al., 2005; Bytheway and Ross, 2010), but nonmetric methods continue to be more frequently used for sex estimation due to ease of use and the inherent problems with defining measurement landmarks on the innominate (Walker, 2008; Klales, 2011). Interobserver errors using measurements of the innominate have been notoriously high because of poorly defined or difficult-to-locate measurement landmarks, though recent research by the authors has shown the utility of using discriminant function analyses of interlandmark distances for sex estimation (Klales et al., 2009; Vollner, 2009). Furthermore, many metric methods of sexing have proven to be highly specific to the population from which the sample originated and may not distinguish between the sexes in other populations (MacLaughlin and Bruce, 1986). Finally, apprehension for the use of statistical methods may be further limiting the widespread use of metric methods. On the contrary, nonmetric methods continue to be taught and utilized because of their practicality: they can be performed quickly and do not require specialized equipment. Nonmetric methods typically can be used even with fragmented remains and are the suggested method of sex estimation for the pelvis in many popular introductory osteology textbooks (for example Stewart, 1979; Krogman and Işcan, 1986; Bass, 2005; Byers, 2005). However, nonmetric methods are not without their limitations. Shortcomings of nonmetric methods in general, as summarized by Bruzek (2002: 158) include a "(1) high degree of observer subjectivity, (2) a lack of consistency in evaluation of traits, and (3) a strong dependence on the results of previous experiences of the observer." To these shortcomings, one could add that most nonmetric methods generally avoid the use of statistical methods for classification, which often optimize correct classification rates.

This research was presented in part at the 78th Annual Meeting of the American Association of Physical Anthropologists in Chicago (April 2009).

*Correspondence to: Alexandra R. Klales, 1515 E. Woodbank Way, West Chester, PA 19380, United States. E-mail: alexandra.klales@gmail.com

Received 9 October 2011; accepted 4 May 2012

DOI 10.1002/ajpa.22102
Published online 19 June 2012 in Wiley Online Library (wileyonlinelibrary.com).

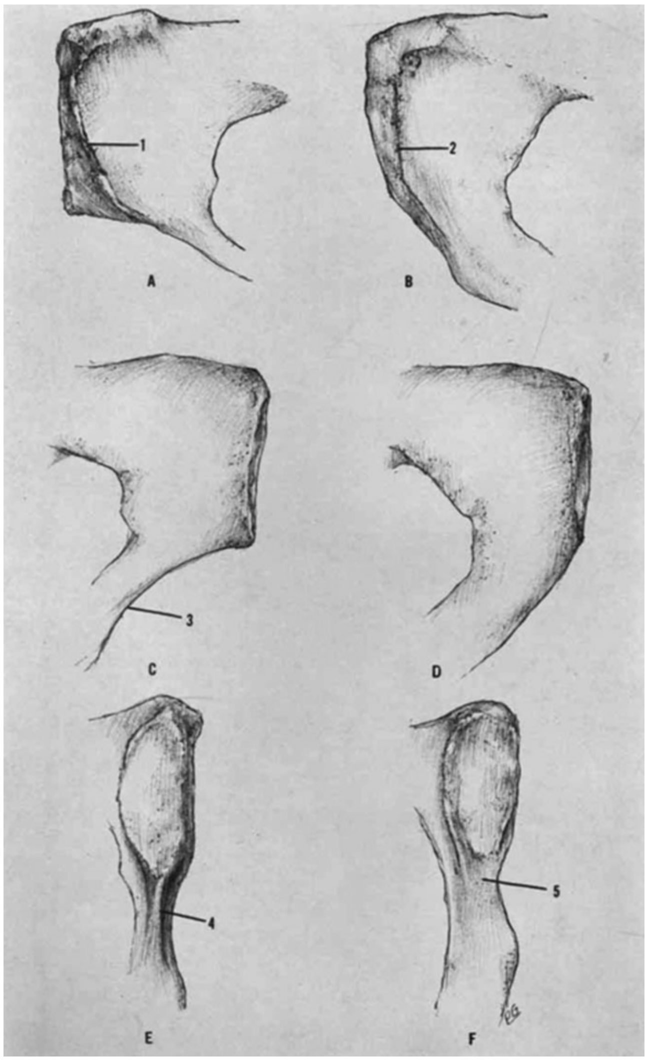

Fig. 1. Phenice's (1969) nonmetric traits of the pubis for the female (left) and male (right) form. From top to bottom: ventral arc (1), the subpubic concavity (3), and the medial aspect of the ischio-pubic ramus (4) (From Phenice, Am J Phys Anthropol, 1969, 30, 297–302, © Wiley, reprinted with permission).

The three most often cited nonmetric traits for sex estimation of the innominate were first described by Phenice (1969): the ventral arc (VA), the subpubic concavity/contour (SPC), and the medial aspect of the ischio-pubic ramus (MA) (Fig. 1). According to Phenice (1969), females frequently have (1) an elevated ridge of bone on the ventral surface of the pubis known as the ventral arc, (2) a lateral curvature of the subpubic concavity, and (3) an elevated ridge of bone on the medial aspect of the ischio-pubic ramus, while males typically do not exhibit these traits. Presence of the traits is therefore considered indicative of a female, while absence of the traits indicate a male. Phenice reported sex classification accuracy of over 95% using a simple majority rule when all three traits did not indicate the same sex. Many validation studies have since confirmed the utility of Phenice's traits both individually and in combination; however, most of these studies failed to produce accuracy rates as high as Phenice's (Table 1). The inconsistencies in classification accuracy may be the result of the shortcomings of nonmetric methods in general, as summarized by Bruzek (2002) (above).

TABLE 1. Classification accuracies in validation studies of the Phenice Method (1969)

Validation study	Classification accuracy (%)
Kelley (1978)	90
Schon (1979)	96
Rosenberg (1986)	99
Sutherland and Suchey (1987)	96
Lovell (1989)	83
MacLaughlin and Bruce (1990)	59–83
McBride et al. (2001)	89
Ubelaker and Volk (2002)	88

The observed differences between males and females, as described by Phenice (1969) are often easily seen; however, previous studies utilizing this technique fail to describe the full range of variation seen in each trait's expression or why these differences occur. To rectify the aforementioned shortcomings, a comprehensive description of each trait and the morphological expression and developmental differences between males and females is presented below. Additional grades of expression for Phenice's (1969) original traits were also employed in order to encompass a wider range of variation and to allow for greater differentiation between the sexes.

TRAIT DESCRIPTIONS

Ventral arc

The ventral arc is the female condition consisting of "a slightly elevated ridge of bone which extends from the pubic crest and arcs inferiorly across the ventral surface of the lateral most extension of the subpubic concavity where it blends with the medial border of the ischio-pubic ramus" (Phenice, 1969: 298). While a true ventral arc is only present in female pubic bones, a ridge of bone along the ventral side of the pubis can also be found in males (Phenice, 1969; Anderson, 1990; Sutherland and Suchey, 1991); however, the difference in orientation and the degree of the angle of this bony ridge allows distinction between the sexes when present. The morphological differences in arc manifestation are a result of variation in the attachment site for the gracilis and adductor (brevis and magnus) muscles (Anderson, 1990; originally noted by Todd, 1921). These muscle attachments sites in males were found to be more medially and inferiorly placed than in females and result in a ventral ridge of bone that is parallel to the face of the pubic symphysis in males (Anderson, 1990). The overall morphology of the region is also helpful for differentiating between the sexes although previously ignored by Phenice. Stewart (1979) noted a "triangle" area of bone found in females inferior and medial to the ventral arc that Kerley (1977) suggested is the result of the adolescent growth spurt in the female pubis.

Subpubic concavity

Phenice (1969: 300) described the subpubic concavity as "a lateral recurve which occurs in the ischio-pubic ramus of the female a short distance below the lower margin of the pubic symphysis... [which] is absent in the male pelvis." The concavity of the female ischio-pubic ramus results in a greater subpubic angle where the two innominates articulate, and a generally more gracile form when compared with males. Coleman (1969) noted that the developmental differences that occur during puberty explain the differences between males and females in the subpubic angle. On the basis of longitudinal radiographic data from the Fels study (Roche, 1992), Coleman (1969) found two factors account for the shape differences in the ischio-pubic ramus: (1) the inferior margin of the ischial tuberosity grows laterally in females, while growing inferiorly in males and (2) differential directional growth at the middle of the ischio-pubic ramus occurs. LaVelle (1995) also compared growth in this area and found significant regional differences between males and females during the adolescent growth phase that explain morphological differences. Females tended to exhibit an increase in growth during this period in the region of the pubis and ischium which results in a longer pubis, a larger pelvic outlet, and a more obtuse subpubic angle (LaVelle, 1995). These differences in growth are related to the sexual dimorphism between males and females associated with parturition.

Medial aspect of the ischio-pubic ramus

Phenice (1969) wrote that the medial aspect of the ischio-pubic ramus is the most ambiguous and least understood of these three traits for sex estimation. The male condition is described as "a broad surface which is found on the ischio-pubic ramus immediately below the symphyseal surface," while the female condition consists of a sharp "ridge which is found here... in contrast to the broad surface in the male" (Phenice, 1969: 300). In addition to the ridge, the female form is narrower than the male form. Lengthening of the pubic bone during the adolescent growth phase in females may account for the narrower surface found in females; however, little research has currently be been done as to why these sex differences occur.

RATIONALE

Phenice (1969:298) acknowledged that every individual will not represent the perfect male or female form in each or all of the traits and suggested that "more detailed research of these criteria could result in refinements of the technique which would allow even greater accuracy." Therefore, simply scoring the extremes of a trait fails to encompass the range of variation found in the pubis. As a result, many subsequent studies have included an intermediate or ambiguous stage between the female and male expression in order to accommodate individuals lacking the classic male or female forms of the trait (cf. Kelley, 1978; MacLaughlin and Bruce, 1990). The Phenice traits can be improved by classifying the range of variation and also sex classification can be improved because currently, the technique weighs all traits equally and fails to objectively assign greater weight to more informative traits.

The most significant weaknesses of the Phenice (1969) technique, and traditional nonmetric techniques in general, are that (1) observations are forced into only a few character states that do not often accommodate the range of variation; (2) an estimated sex is based on simple majority rules that weigh each observation equally, even though some traits show greater sex differences than others; (3) and the estimated sex is not associated with a posterior probability to quantify uncertainty. In light of the *Daubert* decision (*Daubert vs. Merrell Dow Pharmaceuticals, Inc.,*1993), scientific conclusions must follow four guidelines: (1) conclusions must be based on established theories using reliable techniques and methods, (2) the tests or techniques must be peer reviewed, (3) the techniques must have estimated error rates in order to assess validity, and (4) the underlying scientific methods must be generally accepted by the peer community, maintaining the Frye standard (Melnick, 2005). In many ways the *Daubert* decision simply reminds forensic scientists that they must "do good science." More recently, the U.S. National Academy of Sciences (NAS) (2009) evaluated the current state of the forensic sciences and reinforced the *Daubert* decision with specific recommendations for improvements. Reliability, or repeatability, is a measure of interobserver agreement in measurements or scoring of traits, while validity reflects the accuracy of the answers provided by the method. The Phenice (1969) technique of sex estimation fails to meet several of the *Daubert* requirements, for example, the reliability of the scoring procedure and estimated error rates, calling the utility of this technique for sex estimation into question, especially in light of the NAS (2009) report. The technique does not provide a posterior probability of sex classification, so certainty in sex estimations cannot be quantified. In two-group classifications, the posterior probability is the probability that a classified individual is a member of a particular group as opposed to another group and is calculated for both reference groups. The sum of posterior probabilities is 100%; for instance, an individual with certain observations may have a posterior probability of 90% for being a male and 10% for being a female. Naturally, anthropologists should use methods that have quantifiably higher observer agreement and more frequently produce correct answers.

Two recent studies (Walker, 2008; Hefner, 2009) addressed some of the shortcomings of previous analyses of nonmetric traits. First, the new studies specified up to five ordinal grades of trait expression used in traditional nonmetric methods in order to standardize the scoring of those traits. Second, the authors analyzed those traits using statistical classification methods in order to best weight traits in combination, to estimate error rates, and to provide posterior probabilities. Walker (2008) examined and analyzed traditional cranial and mandibular traits for sex estimation in several different groups. Various statistical classification methods were applied to the ordinal data to assess sex and showed high classification accuracies (88% to 90%) when all traits were used. Interobserver differences were reported to be small, so the method would appear to be both reliable and valid. Hefner (2009) undertook a similar approach with ordinal scoring of cranial traits for ancestry estimation in African, European, Native American, and Asian samples. Hefner's study reported high classification accuracies (84–93%) depending on trait combination and statistical classification method used and also had relatively low observer differences, further showing the utility of this method in anthropological applications.

Although Phenice (1969) and Kelley (1978) maintained that the traits of the innominate were overwhelmingly expressed at the extremes for each sex, scoring the Phenice traits on an ordinal scale and analyzing them using statistical methods is a logical approach. The purpose of this study is to improve upon Phenice's (1969) technique for sex estimation through ordinal scoring of traits, analyzing the scores using statistical classification, and comparing scores to quantify intra- and interobserver agreement, thereby providing a method of sex estimation with estimates of reliability and validity.

MATERIALS AND METHODS

Samples

The material used for this study is derived from two skeletal collections: the Hamann-Todd Human Osteological Collection (HTH) housed at the Cleveland Museum of Natural History and the W.M. Bass Donated Skeletal Collection housed at the University of Tennessee, Knoxville (UTK). A stratified random sample of individuals was selected from each collection so both sexes and ancestral groups were represented as evenly as possible. The HTH sample consists of 170 innominates from American black and white males and females born for the most part in the 19th century (Table 2). The UTK sample consists of 140 male and female innominates from black, white, Hispanic, Asian, Mexican and Japanese populations all born during the 20th century (Table 2). All individuals utilized in this research are of documented age, sex, and ancestry. Left innominates used in this study were from individuals that were developmentally mature with all epiphyses fused, with an intact pubic bone, and with no apparent pathological conditions affecting the bone.

TABLE 2. Sample composition

Sample	Population[a]	*n*
HTH	BF	41
(n = 170)	WF	42
	BM	44
	WM	43
UTK	BF	7
(n = 140)	WF	46
	JF	1
	BM	30
	WM	46
	HM	8
	MM	1
	AM	1

[a] Populations are labeled according to descriptions in the sample's database.
BF, black female; WF, white female; JF, Japanese female; BM, black male; WM, white male; HM, Hispanic male; MM, Mexican male; AM, Asian male.

Nonmetric traits

This study focuses on the three traits originally described by Phenice (1969), but an attempt was made to better understand the range of expression for each trait. For this study, character states were described and illustrated in order to expand upon the presence/absence used by Phenice (1969) in a fashion similar to other recent studies focusing on nonmetric traits (Walker, 2008; Hefner, 2009). Five grades of expression were chosen instead of three, as commonly used in nonmetric studies, to encompass more of the range of variation. Illustrations were created for each character state by ARK to facilitate standardization in trait scoring (Fig. 2). Each trait was

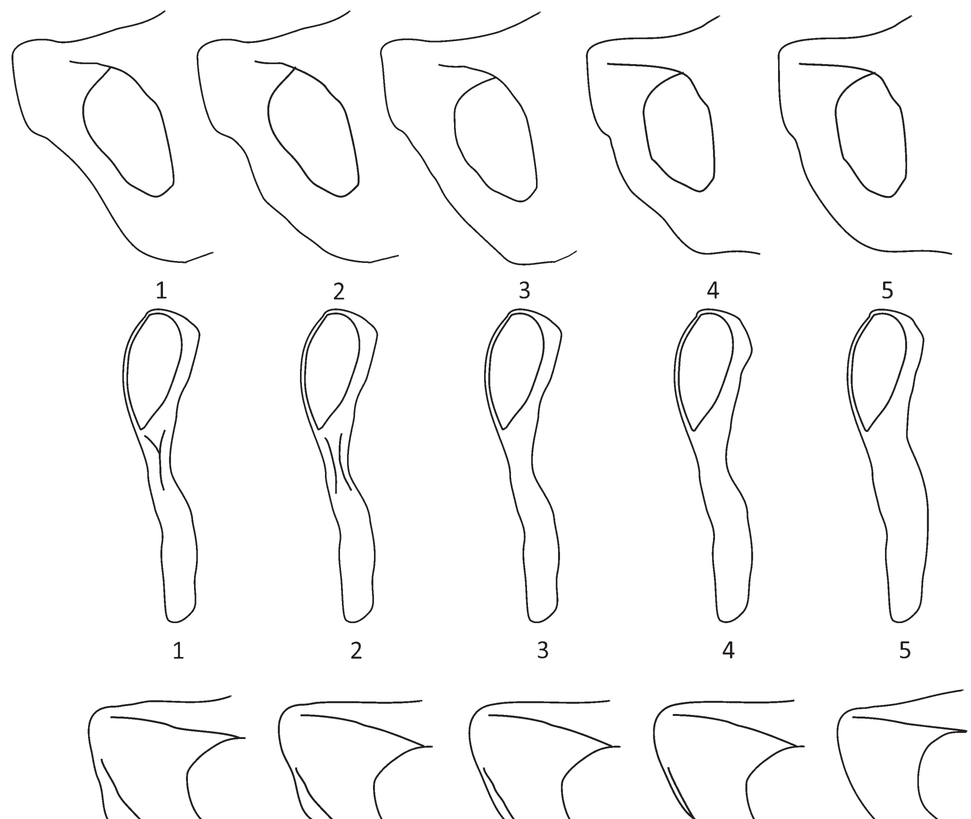

Fig. 2. Character states and ordinal scores for the SPC (top), the MA (middle), and the VA (bottom).

Fig. 3. Proper orientation of the innominate to score the VA. Ventral surface of pubic bone is showing with the superior pubic ramus aligned horizontally. Left innominate, medial side on left. (Photo credited to Laura Diefenbach and courtesy of Dr. Dennis Dirkmaat of Mercyhurst University). Color photographs can be found online at http://nonmetricpelvissexing.weebly.com. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Fig. 4. Proper orientation of the innominate to score the SPC. Dorsal surface of pubic bone is showing with the superior pubic ramus aligned horizontally. Right innominate, medial side on left. (Photo credited to Laura Diefenbach and courtesy of Dr. Dennis Dirkmaat of Mercyhurst University). Color photographs can be found online at http://nonmetricpelvissexing.weebly.com. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

scored on an ordinal scale from one to five without any assumption of "maleness" (masculinity) or "femaleness" (femininity). All innominates from the HTH collection were blindly scored using the new graphics and descriptions by a total of four observers, while the UTK validation sample was scored by ARK.

For this study, slight modification to Phenice's (1969) descriptions of all three traits was necessary. The ventral arc was originally described as solely the presence or absence of a linear muscle attachment ridge on the ventral surface. This trait was expanded to include the angle or degree of orientation of the bony ridge in relation to the symphyseal face, as well as, to take into consideration the overall morphology of the region inferior and medial to the arc. Phenice (1969) described the subpubic concavity as the lateral curvature of the medial edge of the ascending ramus (in dorsal view) situated inferior to the symphyseal face. We use the term subpubic "contour" in order to describe the trait neutrally because the subpubic area ranges in morphology from a subpubic concavity, seen more often in females, to a subpubic convexity seen more often in males. Subsequent researchers have included the entire length and shape of the inferior pubic ramus in assessment of curvature. For the purpose of this study, the area below the symphyseal face, as

Fig. 5. Proper orientation of the innominate to score the MA. Medial surface of the pubic bone is showing with the symphyseal face aligned vertically. Left innominate, posterior side on left. (Photo credited to Laura Diefenbach and courtesy of Dr. Dennis Dirkmaat of Mercyhurst University). Color photographs can be found online at http://nonmetricpelvissexing.weebly.com. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

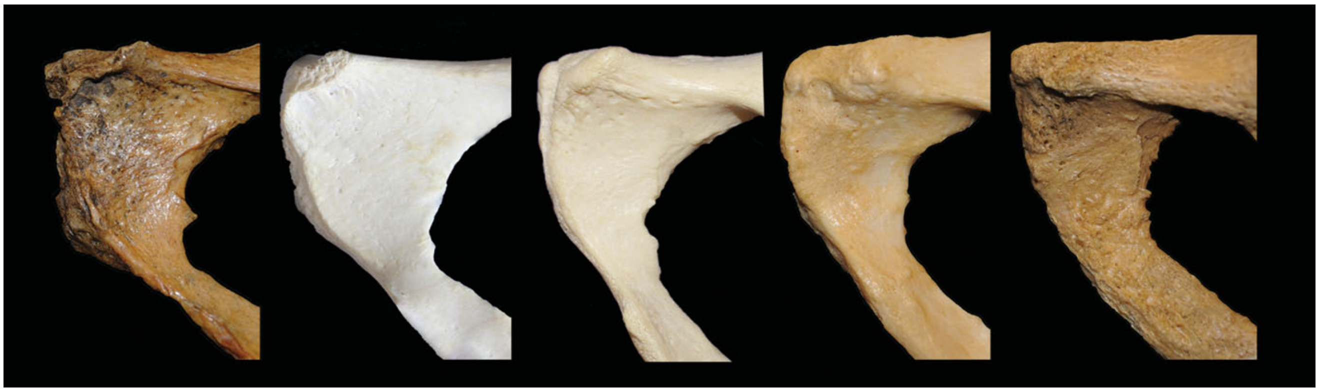

Fig. 6. Character states for the VA (from left to right: Score 1 to Score 5). Ventral surface of pubic bone is showing with the superior pubic ramus aligned horizontally. Left innominate, medial side on left. Color photographs can be found online at http://nonmetricpelvissexing.weebly.com. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

well as the entire medial edge is considered in dorsal view. The inclusion of the inferior portion of the pubic bone takes into account the subpubic concavity and the subpubic angle. Finally, the medial aspect of the ischio-pubic ramus has been expanded to include the dorso-ventral width of the ascending ramus, as well as the presence or absence of a ridge or plateau of bone.

To utilize the illustrations correctly, the scorer must orient the innominate properly. For assessment of the ventral arc, the ventral surface of the pubic bone should be perpendicular to the viewer and the superior pubic ramus should be aligned horizontally or straight (Fig. 3). For this trait, the area lateral to the symphyseal face should be assessed. For the subpubic contour, the dorsal surface of the pubic bone should be held perpendicular to the scorer (Fig. 4). The region immediately below the symphyseal face and the entire inferior pubic ramus should be considered. To assess the medial aspect of the ischio-pubic ramus, the symphyseal face should be held perpendicular to the viewer (Fig. 5). The superior and inferior borders of the symphyseal face should be aligned vertically (see Fig. 5). To assess this trait the region below the inferior edge of the symphyseal face to the approximate mid-region of the ramus should be considered.

Photographs of real bone examples for each trait score are presented in Figures 6–8. In addition, color photographs can be found online at the site entitled “Nonmetric Pelvis Sexing” (http://nonmetricpelvissexing.weebly.com). Written descriptions of the scores are as given in the following text.

Ventral arc

1. Arc present at approximately or at least a 40° angle in relation to the 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 the 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).

Subpubic contour

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 the ischio-pubic ramus

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 with no ridge present.
5. Ascending ramus is very broad dorso-ventrally with no ridge present.

STATISTICAL METHODS

Frequencies for each ordinal score for all three traits were calculated for the HTH and UTK samples and for the different groups in each sample. The Freeman-Halton test, an extension of Fisher’s exact test for more than two columns in a contingency table (SAS Institute, 2002), was used to test for significant differences in trait frequencies between the groups from the two samples utilized (HTH and UTK). In addition, polychoric correlations (necessary when comparing ordinal scales) of the traits in each sample were calculated.

Classification

The scores for each trait were analyzed using ordinal logistic regression (LR). Logistic regression involves the

Fig. 7. Character states for the SPC (from left to right: Score 1 to Score 5). Dorsal surface of pubic bone is showing with the superior pubic ramus aligned horizontally. Right innominate, medial side on left. Color photographs can be found online at http://nonmetricpelvissexing.weebly.com. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Fig. 8. Character states for the MA (from left to right: Score 1 to Score 5). Medial surface of the pubic bone is showing with the symphyseal face aligned vertically. Left innominate, posterior side on left. Color photographs can be found online at http://nonmetricpelvissexing.weebly.com. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

direct calculation of posterior probabilities for classification and is suited to classification into two groups using discrete or ordinal data because it does not require normally distributed data or the same level of variation in groups (Press and Wilson, 1978). For these data, logistic models were fit to trait scores as ordinal data. R (R Development Core Team, 2011), SAS version 9.0 (SAS Institute, 2002), and SPSS 17.0 (SPSS Inc., 2005) were used for all statistical analyses.

Classification accuracy must be estimated and cannot merely be the accuracy when applied to the sample, because it is biased optimistically: each individual is classified based on a classification function that it influenced. In logistic regression, different software packages produce different values for estimated classification rates because they use different algorithms. SAS estimates classification accuracy using a less biased procedure than some other software packages (Peng and So, 2002), but in this paper a more conservative but laborious procedure was followed to estimate classification accuracies: leave-one-out-cross-validation (LOOCV; Lachenbruch and Mickey, 1986). In LOOCV, each individual in a reference group is removed from its reference group, the function is calculated and applied to the removed individual, and its classification is recorded. That individual is then added back into its group, and the procedure is applied to each individual in the reference groups. When all individuals have been classified in this manner, the percentage of correctly classified individuals is the estimated unbiased classification accuracy. Many packages offer LOOCV for linear discriminant functions, but in this case was performed using logistic regression in SPSS and manual tabulation of classifications. Classification accuracies for males and females were estimated separately using LOOCV and are a measure of validity. Accuracy estimates for each sex were estimated and are important because a pronounced sex bias usually indicates different patterns of variation in the scores for each sex that cannot be overcome by the classification method. External validity, the application of a method to a completely different sample, is another important consideration. One of the authors (ARK) scored 140 innominates from the UTK material as an independent validation sample. The calibration sample (HTH) was used to calculate the classification function, which was then applied to the validation sample (UTK), and the correct classification percentage of the UTK sample is an estimate of external validity. The UTK validation sample represents a modern sample of individuals born, in many cases, 100 years later than the HTH individuals.

Intraobserver and interobserver differences

The HTH sample was originally scored by two of the authors (ARK and JMV) and also by two additional individuals with limited osteological experience. One of the scorers had no previous coursework in osteology and had no prior familiarity with the Phenice (1969) technique for sex estimation. The other scorer was a student who had previously taken one introductory course in fragmentary osteology and was aware of the Phenice (1969) technique. However, she had limited experience in using the Phenice technique to assess sex of the human pelvis. Therefore, there are three experience levels represented: two second year graduate students with experience in the technique (Observers A and B), a senior undergraduate student with knowledge of the technique, but without practical application of the technique (Observer C), and one incoming first year graduate student with no prior experience (Observer D). Several months after the first observations, JMV blindly scored the entire HTH sample again for tests of intraobserver agreement. Reli-

TABLE 3. Classification accuracies (%) using logistic regression for experienced observers (A and B) and inexperienced observers (C and D)

Scorer	Males	Females	Combined	Sex Bias
A[a]	99	92	95.5	7
B[b]	97	90	93.5	7
C[c]	87	84	85.5	3
D[d]	77	77	77	0

[a] Sample (males n = 88, females n = 83).
[b] Sample (males n = 77, females n = 86).
[c] Sample (males n = 90, females n = 90).
[d] Sample (males n = 91, females n = 85).

ability was assessed using Cohen's (1968) Weighted Kappa (K) for intraobserver agreement. There has been debate on what constitutes agreement for K, so for this study significance was based on parameters outlined in Landis and Koch (1977): K = 0.0 shows no agreement, K = 0.01 to 0.20 is slight agreement, K= 0.21 to 0.40 is fair agreement, K= 0.41 to 0.60 is moderate agreement, K=0.61 to 0.80 is substantial agreement, and K=0.81 to 1.0 is almost perfect to perfect agreement. Interobserver error was tested using the intraclass correlation coefficient (ICC) which measures the proportion of the variance that is attributable to object measurements (Shrout and Fleiss, 1979). A two-way random, average model was used, with consistency type 95% tolerance interval. Assessing reliability is important because unreliable methods cannot be valid and appropriate methods for data types must be used (Foster and Huber, 1999; Klales and Ousley, 2010).

RESULTS

The effects of ancestry were not found to be significant in previous research (Vollner et al., 2009), but were tested to see if significant variation existed between different ancestral groups. Analysis of differences using the Freeman-Halton test showed no significant differences at $P < 0.05$ between the black and white groups within the HTH and UTK samples. Therefore, ancestry groups from each sample were pooled for all analyses.

Reliability

Tests of intraobserver error using Cohen's Weighted Kappa provided values indicating moderate to substantial levels of agreement based on parameters outlined by Landis and Koch (1977). The weighted K value for the ventral arc was 0.645 (substantial agreement), for the subpubic contour weighted K was 0.579 (moderate agreement), and for the medial aspect weighted K was 0.694 (substantial agreement). Interobserver error using the ICC for all four observers with multiple levels of experience for the entire sample rendered high values of agreement: 0.9 for the ventral arc, 0.8 for the subpubic concavity, and 0.8 for the medial aspect of the ischio-pubic ramus. As in Cohen's K values, no single trait showed significantly higher agreement than the others. Measures of both intra- and interobserver differences indicate high reliability of the method for scoring the features.

Validity

Classification accuracy rates are presented for the experienced observers (Observers A and B) and for the inexperienced observers (Observers C and D) in Table 3.

TABLE 4. Classification accuracies (%) using logistic regression by trait and trait combinations for observer A

Trait(s)	Males	Females	Combined	Sex bias
MA	79.3	72.3	75.8	7.0
SPC	82.8	90.4	86.6	7.6
VA	80.5	96.4	88.5	15.9
MA,VA	95.4	90.4	92.9	5.0
VA,SPC	95.4	91.6	93.5	4.8
MA,SPC	92.0	85.5	88.8	6.5

The ventral arc (88.5% correct classification) proved to be the best trait for separating males and females, similar to Phenice (1969), followed by the subpubic contour (86.6% correct classification) and then the medial aspect of the ischio-pubic ramus (75.8% correct classification) (Table 4). The lower classification rate using the medial aspect of the ischio-pubic ramus alone was echoed in somewhat lower classification accuracy when it was paired with either of the other two traits. In fact, the ventral arc and subpubic contour together produced the highest rate for all functions, though with some sex bias. However, classification accuracies were among the highest, and sex bias the lowest when all three traits were used in combination. The logistic regression equation to apply to all three trait scores is:

$$2.726(\mathrm{VA}) + 1.214(\mathrm{MA}) + 1.073(\mathrm{SPC}) - 16.312$$

The probability of being female based on the score can be calculated by the equation $p_f = 1/(1 + e^{\text{score}})$ and the probability for males, $p_m = 1 - p_f$ (Press and Wilson, 1978). An unknown individual is classified into the sex with the greater probability; an individual is classified as female if the score is less than zero.

The validity of the method using the HTH group as the calibration sample and the UTK group as the validation sample provided a combined sex classification accuracy of 86.2%. Females achieved 98.0% classification accuracy, while males achieved 74.4% classification accuracy because the UTK sample males had lower scores, especially for the ventral arc, which is weighted heavily in the equation. Also, in the HTH sample, females showed greater variation in their scores than males, whereas in the UTK sample, males showed greater variation in their scores than females.

Frequencies and correlations

Score frequencies were fairly similar for the pooled ancestry HTH and UTK samples in all traits (Table 5), but notable differences are due to a higher percentage of extreme scores (5) in HTH males for VA and SPC and a higher percentage of low scores (1) for all three traits in UTK females (Tables 6–8). Several of the differences were statistically significant at $P < 0.05$ using the Freeman-Halton test. Score frequencies for the ventral arc were nearly evenly distributed between the scores for both samples, with the slight exception of score (1) in the HTH sample. Frequencies of the subpubic contour scores were bimodal in both the samples, while scores for the medial aspect of the ischio-pubic ramus were both skewed to the left with scores (3) and (4) being the most frequently assigned. Using polychoric correlations, all three traits were significantly correlated to each other at $P < 0.01$ for both samples (Table 9). Because the three traits analyzed are significantly correlated, the

TABLE 5. Frequencies (%) for each trait score in the calibration and validation sample

		Trait score				
Sample	Trait	1	2	3	4	5
HTH	VA	8.8	25.9	22.4	21.2	21.8
	SPC	12.4	28.8	11.8	28.8	18.2
	MA	2.9	10.0	32.9	44.7	9.4
UTK	VA	23.6	17.9	18.6	22.9	17.1
	SPC	27.1	19.3	14.3	32.1	7.1
	MA	10.0	12.1	33.6	31.4	12.9

TABLE 6. Score frequencies for the ventral arc by population group

		Trait score				
Sample	Population	1	2	3	4	5
HTH	BF	0.10	0.51	0.32	0.07	0.00
	WF	0.26	0.52	0.19	0.02	0.00
	BM	0.00	0.00	0.20	0.27	0.52
	WM	0.00	0.00	0.19	0.49	0.33
UTK	BF	0.86	0.14	0.00	0.00	0.00
	WF	0.54	0.39	0.07	0.00	0.00
	BM	0.00	0.07	0.27	0.43	0.23
	WM	0.02	0.08	0.29	0.37	0.24

TABLE 7. Score frequencies for the subpubic contour by population group

		Trait score				
Sample	Population	1	2	3	4	5
HTH	BF	0.22	0.56	0.12	0.10	0.00
	WF	0.29	0.50	0.12	0.07	0.02
	BM	0.00	0.09	0.11	0.52	0.27
	WM	0.00	0.02	0.12	0.44	0.42
UTK	BF	0.71	0.29	0.00	0.00	0.00
	WF	0.65	0.33	0.00	0.02	0.00
	BM	0.00	0.17	0.17	0.57	0.10
	WM	0.04	0.11	0.28	0.41	0.15

simple majority rule proposed by Phenice (1969) does not utilize the variables most effectively.

DISCUSSION

The Phenice (1969) technique has been tested on numerous population groups with varying degrees of success. This study has shown the utility of modifying Phenice's (1969) traits and using them within a statistical framework to optimize accurate sex estimation. The application of Walker's (2008) methodology to create a revised method of sex estimation for the human innominate shows promise for use by forensic anthropologists creating biological profiles and by bioarcheologists attempting to recreate population demographics. Classification accuracies for experienced observers matched and sometimes surpassed those found by Phenice and previous validation studies of his method. All three traits were highly correlated with one another, and the use of all three traits together produced higher classification accuracies than the use of one trait or two traits in combination.

The expansion of Phenice's original presence or absence of the male or female condition into five grades of expression, with no assumption of male or femaleness, was also valuable, because every grade of character expression was seen in our samples. This result contradicts with Phenice's opinion that intermediate forms do not frequently occur. Additionally, score frequency differences in males and females between the two samples contradict Kelley (1978), who found that that intermediate forms are predominantly a female condition and are seen less frequently in males. Intermediate expression was seen most in the UTK sample males.

TABLE 8. Score frequencies for the medial aspect of the ischio-pubic ramus by population group

		Trait score				
Sample	Population	1	2	3	4	5
HTH	BF	0.07	0.24	0.41	0.24	0.02
	WF	0.05	0.17	0.50	0.29	0.00
	BM	0.00	0.00	0.27	0.57	0.16
	WM	0.00	0.00	0.14	0.67	0.19
UTK	BF	0.14	0.29	0.43	0.14	0.00
	WF	0.24	0.24	0.46	0.07	0.00
	BM	0.00	0.00	0.20	0.43	0.37
	WM	0.04	0.07	0.20	0.54	0.15

TABLE 9. Polychoric trait correlations in each sample

Sample		MA	SPC	VA
HTH	MA	1		
	SPC	0.646	1	
	VA	0.512	0.746	1
UTK	MA	1		
	SPC	0.652	1	
	VA	0.604	0.805	1

The results of this study are especially important in light of the *Daubert* guidelines. The newly proposed revision of Phenice's (1969) method for nonmetric sex estimation provides posterior probabilities and has low estimated error rates, comparable and often lower than metric methods for sex estimation with the innominate. This method will likely prove useful for other populations as well, depending on the level of sexual dimorphism. While the UTK validation sample results showed lower classification accuracy in males, the method was tested recently on a modern South African sample and a 99.2% correct classification rate was reported (Kenyhercz, 2012). Variation in scores between males and females differed in the two population samples studied; however, classification accuracies remained high for both sexes in both groups, suggesting that the observed variation did not significantly impact use of this method. Also, using the proposed method is quick and simple. The authors have created a spreadsheet, available upon request, which accepts trait scores, estimates sex, and calculates posterior probabilities for classifications.

Finally, this study has also shown that observers with varying degrees of experience can effectively use this method for sex estimation. The use of illustrations with written descriptions assists both experienced and inexperienced observers in consistent scoring of the traits. Inexperienced observers achieved good classification accuracies for sex, while the experienced observers achieved the highest classification rates. Therefore, while this method can be utilized by those without experience, its utility for sex estimation increases if the scorer learns the Phenice (1969) characteristics prior to using this revised method for sex estimation.

CONCLUSIONS

Numerous studies have validated the use of the Phenice (1969) technique; however, this study was conducted to statistically test both its reliability and validity in view of the *Daubert* criteria for scientific standards. Nonmetric methods of sex estimation continue to be predominantly used for the pelvis; therefore, the standardization of trait expression and statistical methodologies employed in this study created a valid and reliable method for using Phenice's traits. In addition, all three traits can be scored using the illustrations and descriptions quickly and efficiently by people with varying degrees of experience and without the need for specialized equipment. This method of sex estimation shows promise for use in both forensic and bioarcheological applications by incorporating the popularly used Phenice (1969) technique into a systematic scoring system with a statistical framework designed to provide classification accuracies and known error rates.

ACKNOWLEDGMENTS

The authors thank Lyman Jellema, Cleveland Museum of Natural History, for access to the Hamann-Todd Collection and assistance with photography, and Dr. Lee Meadows-Jantz, University of Tennessee, for access to the William M. Bass Donated Collection. Thanks go to Rebekah Baranoff and Alexis Dzubak, for assisting in the data collection process and to Laura Diefenbach for photographing innominates. Innominate photos are courtesy of Dr. Dennis Dirkmaat of Mercyhurst University, Dr. Todd Fenton of Michigan State University, and Dr. Robert Hoppa of the University of Manitoba. Thanks go to Dr. Dennis Dirkmaat, Dr. Steve Symes, and Mr. Luis Cabo for their support. Finally, the authors would like to thank Michael Kenyhercz for editorial assistance and the reviewers for their comments and suggestions.

LITERATURE CITED

Anderson BE. 1990. Ventral arc of the os pubis: anatomical and developmental considerations. Am J Phys Anthropol 83:449–458.

Bass WM. 2005. Human osteology: a laboratory and field manual. Columbia, MO: Missouri Archaeological Society.

Bruzek J. 2002. A method for visual determination of sex, using the human hip bone. Am J Phys Anthropol 117:157–168.

Byers SN. 2005. Introduction to forensic anthropology, 3rd ed. Upper Saddle River, NJ: Pearson.

Bytheway JA, Ross AH. 2010. A geometric morphometric approach to sex determination of the human adult os coxa. J Forensic Sci 55:859–864.

Cohen J. 1968. Weighted kappa: nominal scale agreement with provisions for scaled disagreement or partial credit. Psychol Bull 70:213–220.

Coleman WH. 1969. Sex differences in the growth of the human bony pelvis. Am J Phys Anthropol 31:125–152.

Daubert v. Merrell Dow Pharmaceuticals. 1993. US Supreme Court 509.U.S.579,113S.Ct.2786, 125L. Ed.2d 469.

Day MH, Pitcher-Wilmott RW. 1975. Sexual differentiation in the innominate bone studied by multivariate analysis. Ann Hum Biol 2:143–151.

DiBennardo R, Taylor JV. 1983. Multiple discriminant function analysis of sex and race in the postcranial skeleton. Am J Phys Anthropol 61:305–314.

Foster KR, Huber PW. 1999. Judging science: scientific knowledge and the federal courts. Cambridge, MA: MIT Press.

Hefner JT. 2009. Cranial nonmetric variation and estimating ancestry. J Forensic Sci 54:985–995.

Kelley MA. 1978. Phenice's visual technique for the os pubis: a critique. Am J Phys Anthropol 48:121–122.

Kerley ER. 1977. Forensic anthropology. In: Tedeschi CG, Eckert WG, Tedeschi LG, editors. Forensic medicine, V. II. Philadelphia: W.B. Saunders.

Klales AR. 2011. Current practices in sex estimation for the adult skeleton. Electronic Survey. Available at:http://kwiksurveys.com?u=SexEstimationSurvey. Results accessed 30 September 2011.

Klales AR, Ousley SD. 2010. The utility of Cohen's Kappa for testing observer error in discrete data and alternatives. [Abstract] Proc Am Acad Forensic Sci 16:340.

Klales AR, Vollner JM, Ousley SD. 2009. A new metric procedure for the estimation of sex and ancestry from the human innominate. [Abstract] Proc Am Acad Forensic Sci 15:311.

Kenyhercz MK. 2012. Sex estimation using pubic bone morphology in a modern South African sample: a test of the Klales et al. method. [Abstract] Am J Phys Anthropol S54:179–180.

Krogman WM, Işcan MY. 1986. The human skeleton in forensic medicine. Springfield, IL: Thomas.

Lachenbruch PA, Mickey MR. 1968. Estimation of error rates in discriminant analysis. Technometrics 10:1–11.

Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. Biometrics 33:159–174.

LaVelle M. 1995. Natural selection and developmental sexual variation in the human pelvis. Am J Phys Anthropol 98:59–72.

Letterman GS. 1941. The greater sciatic notch in American Whites and Negroes. Am J Phys Anthropol 28:99–116.

Lovell NC. 1989. Test of Phenice's technique for determining sex from the os pubis. Am J Phys Anthropol 79:117–120.

MacLaughlin SM, Bruce MF. 1986. Population variation in sexual dimorphism in the human innominate. Hum Evol 1: 221–231.

MacLaughlin SM, Bruce MF. 1990. The accuracy of sex identification in European skeletal remains using the Phenice characters. J Forensic Sci 35:1384–1392.

McBride DG, Dietz MJ, Vennemeyer MT, Meadors SA, Benfer RA, Furbee NL. 2001. Bootstrap methods for sex determination from the os coxae using the ID3 algorithm. J Forensic Sci 46:427–431.

Melnick RL. 2005. A Daubert motion: a legal strategy to exclude essential scientific evidence in toxic tort litigation. Am J Pub Health 95 (S1):30–35.

National Academy of Sciences. 2009. Strengthening forensic sciences in the United States: a path forward. Washington, DC: The National Academies Press.

Patriquin ML, Steyn M, Loth SR. 2005. Metric analysis of sex differences in South African black and white pelves. Forensic Sci Int 147:199–127.

Peng C-YJ, So T-SH. 2002. Logistic regression analysis and reporting: a primer. Understanding Stat 1:31–70.

Phenice TW. 1969. A newly developed visual method for sexing the os pubis. Am J Phys Anthropol 30:297–302.

Press SJ, Wilson S. 1978. Choosing between logistic regression and discriminant analysis. J Am Stat Assoc 73:699–705.

R Development Core Team. 2011. R: a language and environment for statistical computing Version 2.13.0. Vienna, Austria: R Foundation for Statistical Computing.

Roche AF. 1992. Growth, maturation, and body composition. The Fels longitudinal study 1929–1991. Cambridge: Cambridge University Press.

Rosenberg KR. 1986. The functional significance of Neandertal pubic morphology. Doctoral dissertation. Ann Arbor, MI: University of Michigan.

SAS Institute. 2002. SAS/STAT1 Version 9.0. Cary, NC: SAS Institute Inc.

Schon R. 1979. Tests on the reliability of the Phenice method and selected, pelvimetric dimensions for sexing human innominates: a preliminary report based on known sex skeletons from the J.B.C. Grant Collection. Unpublished manuscript. Toronto, ONT: University of Toronto.

Shrout PE, Fleiss JL. 1979. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428.

SPSS for Windows, Rel. 14.0. 2005. Chicago, IL: SPSS.

Stewart TD. 1979. Essentials of forensic anthropology. Springfield, IL: Charles C. Thomas.
Sutherland LD, Suchey JM. 1987. Use of the ventral arc in sex determination of the os pubis. [Abstract] Proc Am Acad Forensic Sci. Colorado Springs, CO: American Academy of Forensic Sciences.
Sutherland LD, Suchey JM. 1991. Use of the ventral arc in pubic sex determination. J Forensic Sci 36:501–511.
Tabachnick BG, Fidell LS. 2001. Using multivariate statistics, 4th ed. Boston: Allyn and Bacon.
Todd TW. 1921. Age changes in the pubic bone. III. The pubis of the white female. Am J Phys Anthropol 4:26–40.
Ubelaker DH, Volk CG. 2002. A test of the Phenice method for the estimation of sex. J Forensic Sci 47:19–24.
Vollner JM. 2009. Metric assessment of sex and ancestry in the human innominate. Master's Thesis. Erie, PA: Mercyhurst College.
Vollner JM, Klales AR, Ousley SD. 2009. Sexing of the human innominate using non-metric traits and statistical analysis. Am J Phys Anthropol 138(S48):263–264.
Walker PL. 2005. Greater sciatic notch morphology: sex, age, and population differences. Am J Phys Anthropol 127:385–391.
Walker PL. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. Am J Phys Anthropol 136:39–50.