AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 136:39–50 (2008)

Sexing Skulls Using Discriminant Function Analysis of Visually Assessed Traits

Phillip L. Walker*

Department of Anthropology, University of California, Santa Barbara, California 93106

KEY WORDS discriminant analysis; sexual dimorphism; archaeology; forensics

ABSTRACT The accuracy of sex determinations based on visual assessments of the mental eminence, orbital margin, glabellar area, nuchal area, and mastoid process was tested on a series of 304 skulls of known age and sex from people of European American, African American, and English ancestry as well as on an ancient Native American sample of 156 individuals whose sex could be reliably determined based on pelvic morphology. Ordinal scores of these sexually dimorphic traits were used to compute sex determination discriminant functions. Linear, *k*th-nearest-neighbor, logistic, and quadratic discriminant analysis models were evaluated based on their capacity to minimize both misclassifications and sex biases in classification errors. Logistic regression discriminant analysis produced the best results: a logistic model containing all five cranial trait scores correctly classified 88% of the modern skulls with a negligible sex bias of 0.1%. Adding age at death, birth year, and population affinity to the model did not appreciably improve its performance. For the ancient Native American sample, the best logistic regression model assigned the correct pelvic sex to 78% of the individuals with a sex bias of only 0.2%. Similar cranial trait frequency distributions were found in same-sex comparisons of the modern African American, European American, and English samples. The sexual dimorphism of these modern people contrasts markedly with that of the ancient Native Americans. Because of such population differences, discriminant functions like those presented in this paper should be used with caution on populations other than those for which they were developed. Am J Phys Anthropol 136:39–50, 2008.

Physical anthropologists traditionally base their skeletal sex assessments principally on subjective visual assessments of sexually dimorphic features of the skull and pelvis. With the advent of powerful personal computers, these subjective judgments, to some extent, have been replaced by the use of multivariate discriminant techniques that employ measurements of sexually dimorphic features as the basis for sex assignments. Despite these methodological advances, the sex determinations of most investigators continue to rely heavily on visual assessments of sexually dimorphic traits.

There are several reasons for the failure of statistically based osteometric techniques to replace more subjective visual assessments. Not only is collecting osteometric data time consuming, but also it often requires expensive, specialized anthropometric equipment. More important than these practical considerations are the difficulties osteologists face in devising measurements that adequately capture subtle, visually apparent, sexually dimorphic shape variations; sexually dimorphic features that are easy to see are often very difficult to measure (e.g., Kanazawa, 1979; Graw et al., 1999). The poor preservation and small sample sizes that are typical of archaeological skeletal collections are another impediment to implementing many of the more powerful statistical techniques that require normally distributed ratio level data. These statistical procedures work best when many complete specimens of known sex can be used to develop and test the accuracy of discriminant functions (Cunha and van Vark, 1991; Novotny et al., 1993; Van Vark and Schaafsma, 1992). Often, collections large enough for such purposes are simply not available.

The accuracy of sex determinations based on visual inspection depends largely on the osteologist's familiarity with the population being studied. Although male skulls are generally larger and have heavier muscle attachments than those of females, substantial population differences exist. Even within a restricted geographical region and historical period, patterns of sexual dimorphism sometimes vary significantly (Cunha and van Vark, 1991; Kemkes and Gobel, 2006). Traits that are sexually dimorphic in one population may be much less so in another (e.g., Steyn and Iscan, 1998). Local populations also show sexually dimorphic changes in cranial size and shape over time spans as short as a few decades (Smith et al., 1986; Jantz and Meadows Jantz, 2000; Buretic-Tomljanovic et al., 2006; Buretic-Tomljanovic et al., 2007; Jonke et al., 2007).

This temporal and spatial variation within and between populations makes it necessary to reevaluate the diagnostic value of sexually dimorphic traits each time a new population is studied. In bioarchaeological research, this is problematic because known-sex skeletal material is typically unavailable. Without such documented collections, the diagnostic value of a feature must be established by analyzing the range of variation found in individuals whose sex can be determined with confidence based on highly sexually dimorphic features such as those of the pelvis. This information can then be

*Correspondence to: Phillip L. Walker, Department of Anthropology, University of California, Santa Barbara, California 93106, USA. E-mail: pwalker@anth.ucsb.edu

Received 15 July 2007; accepted 6 November 2007

DOI 10.1002/ajpa.20776
Published online 6 March 2008 in Wiley InterScience (www.interscience.wiley.com).

used as a standard for determining the sexes of individuals lacking pelvic material (Konigsberg and Hens, 1998; Walker et al., 1996; Walker, 2005).

Most researchers are not very systematic in the way they go through the subjective procedure described earlier of assessing the population-specific patterns of sexual dimorphism. Usually, knowledge of the range of variation in a population is slowly accumulated through years of personal experience. This informal learning process has some undesirable consequences. Differences in training and experience with specific populations produce systematic sex determination biases. Someone unfamiliar with the reduced mastoid dimorphism in some California Indian groups, for instance, would be likely to misclassify many females as males based on this trait. Consistent sex determination errors of this kind can distort male–female differences in reconstructed mortality patterns, health status, gender roles, and so on. Such interobserver inconsistencies can also easily produce statistically significant sex-ratio differences that might mistakenly be interpreted as biologically significant.

Another problem with the unsystematic visual assessment of cranial sex differences is the stultifying effect it has had on our understanding of population differences in sexual dimorphism. A cursory examination of crania from various parts of the world is sufficient to demonstrate that patterns of sexual dimorphism vary considerably through time and space. Unfortunately, with the exception of a few multivariate statistical studies of cranial dimensions (Van Vark et al., 1989), the osteological literature is essentially mute on these global patterns of sexual dimorphism and their probable causes.

Finding a solution to these problems is a matter of considerable urgency in view of the trend toward the reburial of archaeological collections in the United States, Australia, and elsewhere (Walker, 2000, 2004). This threat to the continued curation of museum collections is what motivated the research presented in this article. As a member of the National Science Foundation sponsored committee assembled to develop recording standards for skeletal material subject to repatriation (Buikstra and Ubelaker, 1994), the author took the assignment of creating a set of recommended methods for documenting adult sexually dimorphic skeletal features. One of the objectives of this project was to develop a procedure for systematically recording features not adequately captured by standard craniometric observations. Another was to devise a simple system that produced comparable results when used by different osteologists.

This article reports the results of tests designed to assess the interobserver and intraobserver error inherent in the widely used *Standards for Data Collection From Human Skeletal Remains* (Buikstra and Ubelaker, 1994) cranial trait scoring system describe earlier. A statistical method is also described for using these scores to develop population-specific, sex-determination discriminant functions whose performance is comparable with that of those commonly derived from ratio-level craniometric observations. The method is then tested on a modern known-sex skeletal sample and also on an ancient Native American sample for which the sex of individuals could be determined with some certainly based on sexually dimorphic pelvic features. Finally, population differences in cranial sexual dimorphism that limit the general applicability of such equations are discussed.

MATERIALS AND METHODS

The list of cranial traits (nuchal crest, mastoid process, glabella/supraorbital area, supraorbital margin, and mental eminence) included in the *Standards for Data Collection from Human Skeletal Remains* (Buikstra and Ubelaker, 1994) scoring system was selected based on a review of the literature and the author's personal experience. Although they show some population differences, these are all highly sexually dimorphic traits traditionally used by physical anthropologists. Illustrations for use in scoring these traits had previously been published by Broca (1875) and Acsádi and Nemeskéri (1970). Although the Acsádi and Nemeskéri diagrams provided a good starting point, they required considerable modification. The Acsádi and Nemeskéri coding system (-2 = hyperfeminine, -1 = feminine, 0 = androgynous, 1+ = masculine, 2+ = hypermasculine) was developed specifically for sexing people of European ancestry and does not encompass the full range of human variation. People from other geographical areas often diverge significantly from this European sexual dimorphism pattern. In addition, the Acsádi and Nemeskéri diagrams are so schematic that students sometimes have difficulty relating them to the cranial differences they illustrate. Because of these problems, the author developed a new scoring system for the *Standards for Data Collection from Human Skeletal Remains* that encompasses the extremes observed in a worldwide sample of skulls in museum collections. Based on this survey, a set of representative specimens was selected and a preliminary set of drawings was prepared to illustrate each trait. For the orbital margin, glabellar area, and nuchal areas, drawings were made from sagittal sections of plaster casts. The final diagrams (Fig. 1) were produced by modifying these preliminary illustrations digitally to obtain a geometrically even increase in each feature through a scale of 1–5. This substitution of a 1–5 scale for the Acsádi and Nemeskéri -2 to $+2$ scale is methodologically important because it generalizes the system and removes the implicit assumption that the morphological condition assigned a zero value represents the optimal cut-point for separating males from females. The verbal explanations that accompany the diagrams (Table 1) were developed in response to the questions of volunteers who tested the system.

Collections studied

The scoring system described earlier, which was published in 1994 (Buikstra and Ubelaker, 1994), was tested by the author through scoring 304 modern adult American and British skeletons with documented ages and sexes in the collections of the Cleveland Museum of Natural History (Hamann-Todd Collection), the Smithsonian Institution (Terry Collection), and Saint Bride's Church in London, England (Table 2). The skeletons in the Hamann-Todd and Terry Collections are from the bodies of Americans of recent African and European ancestry used for medical school dissection (Cobb, 1932; Hunt and Albanese, 2005). Most of these people were born during the last half of the 19th century (Table 3). The Saint Bride's sample consists of people buried in the church's crypt whose sex and age at death have been determined based on coffin plate inscriptions and other documentary evidence (Scheuer and Bowman, 1995; Walker, 1995).

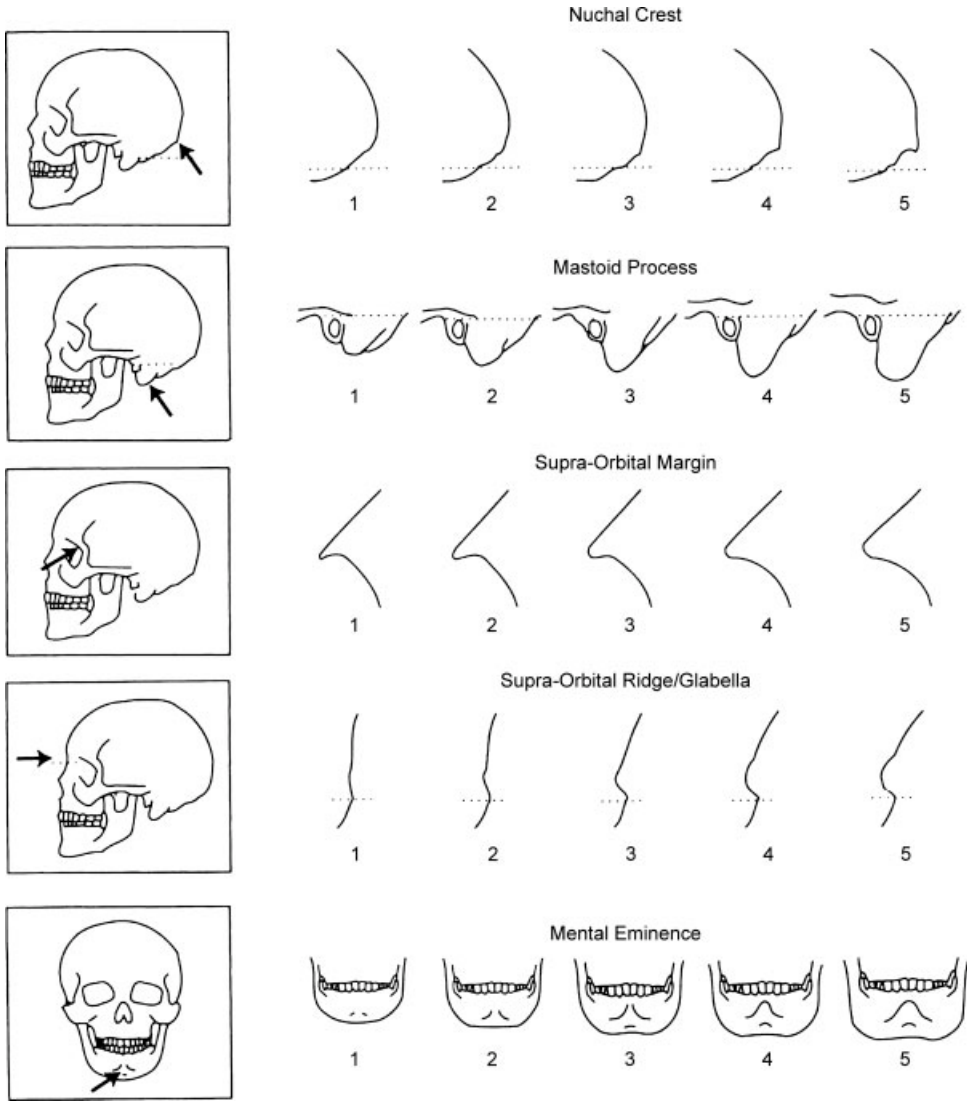

Fig. 1. Standard for scoring cranial traits (from Buikstra and Ubelaker, 1994). The numbers below diagrams are the scores to be assigned to specimens whose morphology most closely resembles the condition illustrated.

Most of these people were born during the last half of the eighteenth century.

Specimens from these collections were selected using a random, stratified sampling strategy with the goal of obtaining approximately equal numbers of adult males and females from each decade of life (Table 2). Damaged crania with unobservable traits were excluded from the sample.

To determine the applicability of these techniques to archaeological collections, cranial trait scores were recorded by the author and two colleagues (Corina Kellner and Angie Carrithers) on crania from archaeological sites in the Santa Barbara Channel area of southern California (SBA-104, SBA-46, SBA-46.1, SBA-52, SBA-60, SBA-71, SCRI-100, SCRI-3, SCRI-83, SRI-2.1, SRI-2.2, SRI-3, SRI-41, SRI-9.1). These are all individuals for which a reliable sex determination could be made based on os coxae morphology using standard sex determination procedures (Buikstra and Ubelaker, 1994; Walker, 2005). These collections are curated by the Santa Barbara Museum of Natural History and the Department of Anthropology, University of California, Santa Barbara. The sites span the period from about 7,000 B.P. until the time of European contact.

Intraobserver and interobserver error tests

A series of tests were performed to assess the magnitude of the intraobserver and interobserver error inherent in the scoring system. Ten skulls were selected to represent the range of variation in each trait. To assess intraobserver error, the author scored each trait on these test specimens five times over a period of about a month. The consistency with which the scoring system could be used by different observers was tested with 20 volunteers. Six of the volunteers were professional physical anthropologists with years of osteological experience. The rest were mostly undergraduate students with little or no previous osteological training. Instruction concerning scoring procedures was limited to providing the volunteers with the text of Table 1, and answering any questions they had after reading it.

TABLE 1. Scoring procedures for sexually dimorphic cranial traits (from: Buikstra and Ubelaker, 1994)

Skull scoring procedures

Hold the skull at arms length a few inches from the diagram. Orient it so that its features can be directly compared with those illustrated. Move the skull from diagram to diagram until the closest match is obtained. Score each trait independently, ignoring the other features.

The following are descriptions of the procedures to use in scoring cranial traits. These are followed by descriptions of minimal (score = 1) and maximal (score = 5) expressions of the traits.

Nuchal crest

View the lateral profile of the occipital and compare it with the diagrams. Feel the surface of the occipital with your hand and note any rugosities on its surface. The important feature to consider in scoring this trait is the development of bone on the external surface of the occipital associated with the attachment of the nuchal muscles. Ignore the contour of the underlying bone (e.g., the presence or absence of an occipital bun in scoring this trait.

a. Minimal expression (score = 1)

The external surface of the occipital is smooth with no bony projections visible from when the lateral profile of the occipital is viewed.

b. Maximal expression (score = 5)

A massive nuchal crest that projects considerable distance from the bone and forms a well defined ledge or hook of bone.

Mastoid process

Score this feature by comparing its size with that of surrounding structures such as the external auditory meatus and zygomatic process of the temporal bone. Mastoid processes vary considerably in their proportions. The most important variable to consider in scoring this trait is the volume of the mastoid not its length.

a. Minimal expression (score = 1)

A very small mastoid process that projects only a small distance below the inferior margins of the external auditory meatus and the digastric groove.

b. Maximal expression (score = 5)

A massive mastoid process with lengths and widths several times that of the external auditory meatus

Orbital margin

Hold your finger against the margin of the orbit in the area lateral to the supra-orbital foramen. Look at each of the diagrams to determine which diagrams it feels like it matches most closely.

a. Minimal expression (score = 1)

Extremely sharp, border feels like the edge of a dull knife

b. Maximal expression (score = 5)

A thick rounded margin with a curvature that approximates that of a pencil

Glabella-supra-orbital ridge

View the cranium from its lateral side and compare the profile of the glabella/supra-orbital area with the profiles in the diagrams.

a. Minimal expression (score = 1)

The contour of the frontal is smooth with little or no projection in the glabellar area.

b. Maximal expression (score = 5)

The glabella and/or supra-orbital ridge are massive and from a rounded loaf shaped projection

Mental eminence

Hold the mandible between your thumbs and your index fingers with your thumbs on either side of the mental eminence. Move your thumbs medially so that they delimit the lateral borders of the mental eminence.

a. Minimal expression (score = 1)

Area of the mental eminence is smooth. There is little or no projection of the mental eminence above the surrounding bone.

b. Maximal expression (score = 5)

A massive mental eminence that occupies most of the anterior portion of the mandible

The score most frequently assigned to each trait of each specimen was used to evaluate interobserver agreement. This modal score was subtracted from the scores each volunteer assigned to the specimen to determine the average deviation from the modal value.

Univariate statistical analysis

The value of each trait when used alone as a sex discriminator was evaluated by calculating the proportion of males and females in the test sample assigned each score (Tables 4 and 5). When proportionately more males were assigned a specific score, individuals with that score were classified as males; when proportionately more females were assigned a specific score, individuals with that score were classified as females. This approach was used to determine cut-points between scores that can be used to assign sexes to crania of unknown sex from comparable populations. It was also used to calculate the trait-by-trait probabilities that individuals assigned specific scores were either male or female (Table 5). This was done using the following equations:

$$\text{pm} = \%\ \text{males}/(\%\ \text{females} + \%\ \text{males})$$

$$\text{pf} = \%\ \text{females}/(\%\ \text{females} + \%\ \text{males})$$

where pm and pf are the probabilities that a person having that score is either male or female and % male and % female are the proportions of males or females in the test sample assigned the score. Using these data, predicted sexes were assigned based on the assumption that individuals with male probabilities of >0.5 were likely to be males and individuals with female probabilities of >0.5 were likely to be females. For each trait, the optimal cut-point for discriminating between males and females was determined through inspection of the probabilities in Table 5. For most of the traits in the modern population, the cut point is between the scores of 2 and 3: individuals

TABLE 2. Number of individuals in the modern skeletal samples studied by age, sex, and population affinities

	Age group						
Population/sex	17–29	30–39	40–49	50–59	60–69	70–79	Total
African Americans							
Male	13	11	9	12	8	7	60
Female	11	11	9	4	8	10	53
Total	24	22	18	16	16	17	113
European Americans							
Male	10	11	8	15	11	6	61
Female	10	6	9	11	8	8	52
Total	20	17	17	26	19	14	113
English							
Male	8	4	5	7	13	6	43
Female	1	4	3	11	8	8	35
Total	9	8	8	18	21	14	78
All groups							
Male	31	26	22	34	32	19	164
Female	22	21	21	26	24	26	140
Total	53	47	43	60	56	45	304

TABLE 3. Statistics on birth year and age at death for the skeletal collections studied

	Hamann-Todd	Terry	St Bride's	Total
Males				
N	43	78	43	164
Mean Age at death	40.2	53.2	52.5	49.6
Mean birth year	1882	1879	1767	1850
Mean death year	1922	1932	1819	1900
Females				
N	41	64	35	140
Mean Age at death	42.9	55.5	59.4	52.8
Mean birth year	1879	1884	1765	1853
Mean death year	1922	1939	1825	1906
Both sexes				
N	84	142	78	304
Mean age at death	41.5	54.2	55.6	51.1
Mean birth year	1880	1881	1766	1851
Mean death year	1922	1935	1822	1902

with scores less than 3 are more likely to be females and individuals with scores of 3 or greater are more likely to be males. The only exceptions to this are the distributions of male and female orbital margin scores in the modern sample and the mental eminence scores in the ancient Native American sample. For the modern sample, only individuals with orbital scores of 1 have a probability of being female that is greater than 0.5; for the ancient Native American sample, only individuals with mental eminence scores greater than 3 have a probability of being male that is greater than 0.5.

Overall, average scores for both males and females in the Native American crania were shifted markedly toward higher, more robust, values. This is especially true for the orbital margin, which tends to be much more rounded in Native Americans than in people of European or African ancestry and in the mental eminence, which tends to be much more pronounced in the Native American females than it is in the females of European or African ancestry.

Multivariate sex determination models

The performance of a variety of discriminant function techniques (linear, *k*th-nearest-neighbor, logistic, and quadratic) was tested using the known sex, or pelvic sex, in the case of the Native American sample, as the dependent variable. The scores for various combinations of cranial traits likely to be encountered in archaeological and forensic contexts were used as independent variables. The goal of these statistical tests was to describe differences between individuals of known sex and then use these differences as the basis for assigning sexes to individuals of unknown sex. Each of these statistical techniques makes specific assumptions about the distributions of the independent variables: typical requirements are ratio level data, with multivariate normality, and variance/covariance matrices of the variables that are homogeneous across groups (McLachlan, 2004). Since the independent variables in the current context are cranial trait scores, they clearly fail the ratio level data criterion and they also may not have multivariate normality.

Several factors mitigate these statistical problems. First, as described earlier, this cranial trait scoring system was intentionally devised to obtain an evenly distributed increase between intervals (1–2, 2–3, 3–4, and 4–5). Although crude, these scores are thus likely to approximate the distributional characteristics of ratio level data. When each set of ordinal scores was recoded as a series of four binary dummy variables to work around these distributional uncertainties, the performance of the more complicated models was virtually identical to that of the simpler models in which the ordinal scores were entered as a single variable. Perhaps this is because, as Jaccard and Wan (1996: p 4) point out, "for many statistical tests, rather severe departures (from intervalness) do not seem to affect Type I and Type II errors dramatically." Also encouraging, is the fact that these cranial data convincingly pass the graphic standard test for multivariate normality devised by Thompson (1990). In any event, the goal of performing discriminant analysis using these cranial trait data is to discriminate between males and females; the test of the efficacy of a specific discriminant procedure in this context is not how well the data fit the assumptions of the technique, but how well the procedure solves the classification problem at hand.

RESULTS

Interobserver and intraobserver error

The deviations of the scores assigned by each observer from the modal scores show that for all traits, 50% or

TABLE 4. Distribution of cranial trait scores for the populations studied

		Score (frequency)					Score (%)				
Trait	Sex	1	2	3	4	5	1	2	3	4	5
Nuchal area	Males										
African American		9	14	20	16	1	15.0	23.3	33.3	26.7	1.7
European American		7	12	22	17	3	11.5	19.7	36.1	27.9	4.9
English		3	17	14	7	2	7.0	39.5	32.6	16.3	4.7
Native American		1	7	28	18	8	1.6	11.3	45.2	29.0	12.9
Total		20	50	84	58	14	8.8	22.1	37.2	25.7	6.2
	Females										
African American		21	21	5	4	2	39.6	39.6	9.4	7.5	3.8
European American		19	23	9	1	0	36.5	44.2	17.3	1.9	0.0
English		18	13	3	1	0	51.4	37.1	8.6	2.9	0.0
Native American		3	24	30	12	4	4.1	32.9	41.1	16.4	5.5
Total		61	81	47	18	6	28.6	38.0	22.1	8.5	2.8
Orbital margin	Males										
African American		24	32	3	1	0	40.0	53.3	5.0	1.7	0.0
European American		12	33	15	1	0	19.7	54.1	24.6	1.6	0.0
English		14	25	4	0	0	32.6	58.1	9.3	0.0	0.0
Native American		1	11	37	18	2	1.4	15.9	53.6	26.1	2.9
Total		51	101	59	20	2	21.9	43.3	25.3	8.6	0.9
	Females										
African American		36	16	1	0	0	67.9	30.2	1.9	0.0	0.0
European American		34	17	1	0	0	65.4	32.7	1.9	0.0	0.0
English		25	10	0	0	0	71.4	28.6	0.0	0.0	0.0
Native American		9	23	34	12	0	11.5	29.5	43.6	15.4	0.0
Total		104	66	36	12	0	47.7	30.3	16.5	5.5	0.0
Glabellar area	Males										
African American		2	11	28	16	3	3.3	18.3	46.7	26.7	5.0
European American		2	2	20	22	15	3.3	3.3	32.8	36.1	24.6
English		9	23	10	1	0	20.9	53.5	23.3	2.3	0.0
Native American		2	16	26	18	4	3.0	24.2	39.4	27.3	6.1
Total		15	52	84	57	22	6.5	22.6	36.5	24.8	9.6
	Females										
African American		25	20	7	1	0	47.2	37.7	13.2	1.9	0.0
European American		21	13	16	2	0	40.4	25.0	30.8	3.8	0.0
English		22	12	1	0	0	62.9	34.3	2.9	0.0	0.0
Native American		18	37	12	9	1	23.4	48.1	15.6	11.7	1.3
Total		86	82	36	12	1	39.6	37.8	16.6	5.5	0.5
Mastoid process	Males										
African American		1	9	29	15	6	1.7	15.0	48.3	25.0	10.0
European American		3	10	30	17	1	4.9	16.4	49.2	27.9	1.6
English		12	26	4	1	0	27.9	60.5	9.3	2.3	0.0
Native American		0	19	28	11	3	0.0	31.1	45.9	18.0	4.9
Total		16	64	91	44	10	7.1	28.4	40.4	19.6	4.4
	Females										
African American		19	17	15	2	0	35.8	32.1	28.3	3.8	0.0
European American		14	27	10	1	0	26.9	51.9	19.2	1.9	0.0
English		10	23	2	0	0	28.6	65.7	5.7	0.0	0.0
Native American		17	41	15	3	0	22.4	53.9	19.7	3.9	0.0
Total		60	108	42	6	0	27.8	50.0	19.4	2.8	0.0
Mental eminence	Males										
African American		0	11	26	16	7	0.0	18.3	43.3	26.7	11.7
European American		0	16	37	7	1	0.0	26.2	60.7	11.5	1.6
English		0	14	27	1	1	0.0	32.6	62.8	2.3	2.3
Native American		0	11	25	21	3	0.0	18.3	41.7	35.0	5.0
Total		0	52	115	45	12	0.0	23.2	51.3	20.1	5.4
	Females										
African American		10	24	19	0	0	18.9	45.3	35.8	0.0	0.0
European American		10	32	9	1	0	19.2	61.5	17.3	1.9	0.0
English		10	24	1	0	0	28.6	68.6	2.9	0.0	0.0
Native American		3	20	39	12	0	4.1	27.0	52.7	16.2	0.0
Total		33	100	68	13	0	15.4	46.7	31.8	6.1	0.0

more of the observers agreed with each other and independently assigned the modal score to the trait. The mean values of these deviations in the interobserver error trials (Tables 6 and 7) show that there is a negative bias for all traits except the development of the glabellar area. In other words, when a value other than the modal score was assigned, this score was usually lower than the modal value.

TABLE 5. The percent of males and females in the modern American/English (known sex) and ancient Native American (sex inferred from pelvic morphology) samples assigned a specific cranial trait scores and the probability of someone with each score being male

	Score				
Trait/group	1	2	3	4	5
Nuchal area					
American/English					
Male %	11.6	26.2	34.1	24.4	3.7
Female %	41.4	40.7	12.1	4.3	1.4
Male probability[a]	0.22	0.39	0.74	0.85	0.72
Native American					
Male %	1.6	11.3	45.2	29.0	12.9
Female %	4.1	32.9	41.1	16.4	5.5
Male probability[a]	0.28	0.26	0.52	0.64	0.70
Orbital margin					
American/English					
Male %	30.5	54.9	13.4	1.2	0.0
Female %	67.9	30.7	1.4	0.0	0.0
Male probability[a]	0.31	0.64	0.90	1.00	1.00
Native American					
Male %	1.4	15.9	53.6	26.1	2.9
Female %	11.5	29.5	43.6	15.4	0.0
Male probability[a]	0.11	0.35	0.55	0.63	1.00
Glabellar area					
American/English					
Male %	7.9	22.0	35.4	23.8	11.0
Female %	48.6	32.1	17.1	2.1	0.0
Male probability[a]	0.14	0.41	0.67	0.92	1.00
Native American					
Male %	3.0	24.2	39.4	27.3	6.1
Female %	23.4	48.1	15.6	11.7	1.3
Male probability[a]	0.11	0.33	0.72	0.70	0.82
Mastoid process					
American/English					
Male %	9.8	27.4	38.4	20.1	4.3
Female %	30.7	47.9	19.3	2.1	0.0
Male Probability[a]	0.24	0.36	0.67	0.9	1.0
Native American					
Male %	0.0	31.1	45.9	18.0	4.9
Female %	22.4	53.9	19.7	3.9	0.0
Male probability[a]	0.00	0.37	0.70	0.82	1.00
Mental eminence					
American/English					
Male %	0.0	25.0	54.9	14.6	5.5
Female %	21.4	57.1	20.7	0.7	0.0
Male probability[a]	0.00	0.30	0.73	0.95	1.00
Native American					
Male %	0	18.3	41.7	35.0	5.0
Female %	4.1	27	52.7	16.2	0
Male probability[a]	0.00	0.40	0.44	0.68	1.00

[a] Male probabilities were calculated as follows: pm = % males/(% females + % males) where pm = the probability of a person with a specific score being male, % females = the proportion of females assigned that score and, % males = the proportion of females assigned that score.

For most traits, the Kruskal-Wallis test indicates no significant difference between observers in the distribution of the scores they assigned to each cranial trait. In other words, no observers systematically assigned scores that were either higher or lower than those that other observers assigned to the same series of test specimens. The exception to this was the mastoid process (χ^2 = 42.7, d.f. = 21, P = 0.003). Evidently, the instructions and diagrams for scoring the mastoid process are interpreted differently by different observers.

TABLE 6. Results of intraobserver-error tests in which the author independently scored 10 test skulls five times over a period of about a month

		Frequency distribution of deviations of assigned scores from the modal value				Average deviations from the modal value of assigned scores	
Trait		−1	0	1	2	Mean	Mean (absolute value)
Glabella	N	1	35	14		0.26	0.58
	%	2.0	70.0	28.0	0.0		
Orbit	N	1	21	28		0.54	0.47
	%	2.0	42.0	56.0	0.0		
Mental	N	14	34	2		−0.24	0.50
	%	28.0	68.0	4.0	0.0		
Mastoid	N	8	28	14		0.12	0.46
	%	16.0	56.0	28.0	0.0		
Nuchal	N		41	8	1	0.52	0.47
	%	0.0	82.0	16.0	2.0		

The values in the table represent the frequencies of deviations from the modal values assigned to each trait of each test skull.

These tests indicate that for most traits there is considerable interobserver scoring agreement. Even students with no previous osteological experience were able consistently to assign scores that agreed with those of very experienced osteologists. Overall, 96% of the scores assigned fell within one score of the modal value assigned by all observers.

In the intraobserver tests, 99.5% of the repeated observations fell within this same range. These results are consistent with similar tests of interobserver error conducted on the same traits scored using a modified version of the current system (Walrath et al., 2004) as well as a simplified male–female coding scheme (Williams and Rogers, 2006).

Univariate statistical analysis

Depending on the trait selected, the univariate approach described earlier results in a correct sex assignment for between about 69% and 83% of the people in the modern sample (Table 8). Overall, the scores for the mastoid process and the glabellar area were the best sex discriminators and the nuchal area and orbital margin the worst. These figures are misleading, however, because the traits varied significantly in their sex bias (i.e., the difference in the proportion of males and females misclassified). For glabellar area, mastoid, and orbital margin, the percent of misclassifications was lower for males; for the nuchal crest and mental eminence, the error rate was lower for females (Table 8).

This univariate approach produced higher classification errors for the Native Americans. The nuchal area is the least sexually dimorphic feature in this population and its scores correctly predicted the sexes of only 57% of the individuals in the sample. The glabellar area and the mastoid process were the most sexually dimorphic features among the Native Americans; the scores for these traits produced correct sex assignments for about 70% of the crania in the sample.

TABLE 7. Results of interobserver-error tests in which 20 people independently scored 10 test skulls.

		Frequency distribution of deviations of assigned scores from the modal value						Average deviations from the modal value of assigned scores	
Trait		−3	−2	−1	0	1	2	Mean	Mean (absolute value)
Glabella	*N*	0	1	30	145	43	1	0.06	0.40
	%	0.0	0.5	13.6	65.9	19.5	0.5		
Orbit	*N*		3	50	135	32	0	−0.11	0.53
	%	0.0	1.4	22.7	61.4	14.5	0.0		
Mental	*N*	0	16	61	120	22	1	−0.31	0.57
	%	0.0	7.3	27.7	54.5	10.0	0.5		
Mastoid	*N*	0	13	55	108	44	0	−0.17	0.35
	%	0.0	5.9	25.0	49.1	20.0	0.0		
Nuchal	*N*	1	4	32	162	21	0	−0.10	0.39
	%	0.5	1.8	14.5	73.6	9.5	0.0		

The values in the table represent the frequencies of deviations from the modal values the testers assigned to each trait of each test skull.

TABLE 8. Test results of different discriminant analysis techniques and sex determination models containing various combinations of cranial trait scores as independent variables using the modern American/English skulls from people of known sex

Model	Statistical method	Dummy variables[a]	% Correctly classified[b]			% Sex bias[c]
			Total	Males	Females	
Sex = glabella	Univariate	–	82.6	84.1	80.7	3.4
Sex = mastoid	Univariate	–	78.6	78.7	78.6	0.1
Sex = mental	Univariate	–	76.6	75.0	78.6	−3.6
Sex = nuchal	Univariate	–	71.4	62.2	82.1	−19.9
Sex = orbit	Univariate	–	68.8	69.5	67.9	1.7
Sex = glabella + mastoid	*k*-nearest neighbor	No	83.3	75.8	92.2	−16.4
Sex = glabella + mastoid	*k*-nearest neighbor	Yes	83.3	75.8	92.2	−16.4
Sex = glabella + mastoid	Linear	No	84.0	84.8	83.0	1.9
Sex = glabella + mastoid	Linear	Yes	84.0	84.8	83.0	1.9
Sex = glabella + mastoid	Logistic	No	84.0	84.8	83.0	1.9
Sex = glabella + mastoid	Logistic	Yes	84.0	84.2	83.7	0.6
Sex = glabella + mastoid	Quadratic	No	84.0	84.8	83.0	1.9
Sex = glabella + mastoid + mental	*k*-nearest neighbor	No	88.9	90.3	87.2	3.1
Sex = glabella + mastoid + mental	*k*-nearest neighbor	Yes	86.9	86.7	87.2	−0.6
Sex = glabella + mastoid + mental	Linear	No	86.9	87.3	86.5	0.7
Sex = glabella + mastoid + mental	Linear	Yes	87.9	89.1	86.5	2.6
Sex = glabella + mastoid + mental	Logistic	No	87.3	87.9	86.5	1.4
Sex = glabella + mastoid + mental	Logistic	Yes	87.3	87.9	86.5	1.4
Sex = glabella + mastoid + mental	Quadratic	No	87.9	89.1	86.5	2.6
Sex = orbit + glabella + nuchal + mastoid	*k*-nearest neighbor	Yes	88.2	82.3	95.0	−12.7
Sex = orbit + glabella + nuchal + mastoid	Linear	Yes	88.5	90.2	86.4	3.8
Sex = orbit + glabella + nuchal + mastoid	Logistic	Yes	88.8	90.9	86.4	4.4
Sex = orbit + glabella + nuchal + mastoid + mental	*k*-nearest neighbor	no	89.8	87.8	92.1	−4.3
Sex = orbit + glabella + nuchal + mastoid + mental	Linear	No	87.5	87.2	87.9	−0.7
Sex = orbit + glabella + nuchal + mastoid + mental	Logistic	No	87.8	87.8	87.9	−0.1
Sex = orbit + glabella + nuchal + mastoid + mental	Quadratic	No	90.1	92.7	87.1	5.5

[a] For the multivariate models, the percent of individuals classified correctly was determined using the "leave one out" method.
[b] % Sex bias = (% males correctly classified − % female correctly classified).
[c] For the models using dummy variables, each cranial trait score was converted into a series of four binary variables that were each entered into the model as independent variables.

Multivariate sex determination models

The results of the multivariate analysis of the modern sample show that, in general, when more cranial traits are included in the discriminant models, better results are obtained (Table 8): the average percent of individuals from the modern samples classified correctly using the "leave one out" procedure was 89% for the models that included all five variables, 88% for those with three, and 84% with those with two independent variables. The quadratic discriminant analysis model containing scores for all five of the cranial traits performed best with 90% of the individuals classified correctly (Table 8).

Again, these statistics on the total number of specimens classified correctly disregard the important issue of bias in the sex determination errors (Meindl et al., 1985; Williams and Rogers, 2006). In bioarchaeological research, sex determination techniques producing similar male and female sex determination errors are highly desirable. If the misclassification errors are heavily biased toward one of the sexes, this can lead to spurious conclusions regarding sex differences in mortality pat-

TABLE 9. Logistic discriminant analysis equations for predicting sex from various combinations of cranial trait scores for the American/English (pooled African American, European American, and English collections) and ancient Santa Barbara Channel area Native American samples

Population-specific discriminate functions[a]	% Classified correctly		% Sex bias[b]
	Males	Females	
American/English			
Y = glabella × −1.375 + mastoid × −1.185 + mental × −1.151 + 9.128	88.4	86.4	2.0
Y = glabella × −1.568 + mastoid × −1.459 + 7.434	85.4	82.9	2.5
Y = glabella × −1.525 + mental × −1.485 + 7.372	86.6	82.1	4.5
Y = mental × −1.629 + mastoid × −1.415 + 7.382	79.9	83.6	−3.7
Y = orbit × −1.007 + mental × −1.850 + 6.018	78.1	77.9	0.2
Y = nuchal × −0.7 + mastoid × −1.559 + 5.329	76.8	82.9	−6.1
Native American			
Y = orbit × −0.499 + mental × −0.606 + 3.414	78.1	77.9	0.2
Y = mental × −0.576 + mastoid × −1.136 + 4.765	74.1	72.7	1.4
Y = glabella × −0.797 + mastoid × −1.085 + 5.025	69.5	82.9	−13.4

[a] These models were selected so that the equations only include statistical significance coefficients; significant effects are suggested when confidence intervals do not contain 0. The value of y is a discriminant function score in which the cut point between males and females is 0. Skulls with values of > 0 are more likely to be females and skulls with scores of < 0 are more likely to be males.
[b] % Sex bias = (% males correctly classified − % female correctly classified).

terns, gender-related burial patterns, and so on. In comparison to the other techniques, quadratic discriminant analysis and especially kth nearest neighbor analysis performed poorly in terms of this sex bias criterion. They produced discriminant functions whose classification error rates were much greater for one sex than the other (Table 8). Logistic and linear discriminant analysis models, in contrast, performed very well in this regard with average sex biases in classification error rates of between 1 and 2%.

Based on inspection of these results, logistic regression appears to be the technique best suited to these data owing to its ability to make sex determinations that have both low misclassification and low sex bias rates. Logistic regression has several additional desirable characteristics in the present context: 1) it is designed to accommodate dependent variables that only have two values (in this case male and female) and produces predicted values that can be interpreted as probabilities of group membership, 2) it does not assume a linear relationship between the dependent and independent variables, 3) it does not require multivariate normality of independent variables, 4) it does not assume homoscedasticity, and 5) in general, it has less stringent requirements than linear discriminant analysis (Press and Wilson, 1978; Hosmer and Lemeshow, 1989).

Owing to these considerations, logistic regression was used to derive discriminant functions for use in the identification of skulls from populations with patterns of cranial sexual dimorphism comparable with that of the test samples (Table 9). These discriminant functions can be used by simply substituting the appropriate cranial trait scores into the equations and then using 0 as the cut point that optimally distinguishes between males and females. Skulls with calculated discriminant function scores of less than 0 are more likely to be from males and skulls with scores of >0 are more likely to be from females. The probability of being female (pf) associated with this value can be calculated using the following equation:

$$\text{pf} = 1/(1 + e^{-y})$$

and the probability of being male (pm) is thus:

$$\text{pm} = 1 - \text{pf}$$

where y is the value produced by the discriminant function equations in Table 9, and e^{-y} is the negative exponential function of y using the base of the natural logarithm (2.71828).

For example, for a skull with the following scores: glabella = 4, mastoid = 3, mental = 4, the calculations would be as follows using the appropriate equation from Table 9:

$$y = (\text{glabella} \times -1.375) + (\text{mastoid} \times -1.185) + (\text{mental} \times -1.150) + 9.128$$

$$y = (4 \times -1.375) + (3 \times -1.185) + (4 \times -1.151) + 9.128$$

$$y = -4.527$$

Because of the way in which sex was coded in the discriminant analysis (0 = male, 1 = female), this value of less than 0 means that this skull is more likely to be from a male than from a female. The probability of this sex assignment can be calculated as follows:

$$\text{pf} = 1/(1 + e^{-(-4.527)})$$

$$\text{pf} = 1/(1 + 92.513)$$

$$\text{pf} = 0.011$$

The probability of this skull being from a female is thus 0.011 and the probability of it being from a male is 0.989 (i.e., 1–0.011).

In developing these discriminant functions for the modern samples, models were tested that included the person's year of birth and age at death as independent variables. The coefficients associated with birth year and age show a modest level of statistical significance ($P = \sim 0.03$). These age-related effects undoubtedly derive from the secular trends (Jantz and Meadows Jantz, 2000) and age-related changes in cranial sexual dimorphism (Walker, 1995) that have been documented

TABLE 10. Sex-specific comparisons of population differences in cranial trait scores

		African American vs. English			African American vs. European American			European American vs. English			American/English vs. Native American		
Trait	Sex	Avg. diff.[a]	χ^2	*P*	*Avg. diff.*[a]	χ^2	*P*	Avg. diff.[a]	χ^2	*P*	Avg. diff.[a]	χ^2	*P*
Glabella	Females	0.3	4.18	0.24	−0.32	8.35	0.08	0.62	12.45	***0.006***	−0.22	38.90	***<0.0001***
	Males	0.05	2.33	0.68	−0.64	16.5	***0.002***	0.68	20.93	***0.0001***	0.43	14.49	***0.013***
Mastoid	Females	0.23	12.47	***0.006***	0.04	4.35	0.23	0.19	4.16	0.25	0.11	19.14	***0.001***
	Males	0.41	8.97	0.06	0.22	4.76	0.31	0.19	8.66	0.07	0.49	26.76	***<0.0001***
Mental	Females	0.43	13.06	***0.001***	0.15	5.71	0.13	0.28	5.43	0.14	−0.41	72.08	***<0.0001***
	Males	0.57	15.74	***0.001***	0.43	10.86	***0.013***	0.14	3.18	0.37	0.21	35.16	***<0.0001***
Nuchal	Females	0.33	2.85	0.58	0.12	5.12	0.28	0.22	2.58	0.46	−0.59	73.60	***<0.0001***
	Males	0.05	5.55	0.24	−0.18	1.52	0.82	0.23	5.66	0.23	−0.19	36.16	***<0.0001***
Orbit	Females	0.05	0.72	0.7	−0.03	0.08	0.96	0.08	0.9	0.64	−1.05	119.54	***<0.0001***
	Males	−0.08	1.88	0.6	−0.4	12.01	***0.007***	0.31	5.68	0.13	−1.23	108.18	***<0.0001***

Population differences that are statistically different at the < 0.05 level are in bold italics.
[a] Avg. diff. = Differences between the average cranial trait scores of the groups compared (i.e., African-American female mean score – English female mean score, etc.). χ^2 = the value comparing the frequency distributions of the traits in each population. *P* = the probability associated with the χ^2 value. Values in bold italics are statistically significant at the < 0.05 level.

through previous studies of these same collections. The inclusion of these additional dependent variables, however, produced little (<2%) improvement in the accuracy of the discriminant functions and, for practical reasons (the precise year of birth and age at death will almost never be known for a skull of unknown sex), they are not included in the discriminant functions presented in Table 9.

Population differences

For the modern samples, models were also tested that included independent variables coding for various population groupings (American vs. English, African vs. European ancestry, and so on). The coefficients associated with these population variables were, in most cases, not statistically significant and in none of the models did the addition of a population variable substantially add to the accuracy of the discriminant function.

Nevertheless, population differences do exist in the distributions of these traits in the modern samples that are of considerable physical anthropological interest. Inspection of mean cranial trait scores (Table 10) shows that for most same-sex comparisons, the African Americans and European Americans are more robust (i.e., they tend to have higher average scores) than the English. For several traits, these differences in robusticity are statistically significant: African American males and females have significantly greater mental eminence development than their same-sex English counterparts, African American females have larger mastoid processes than their English counterparts do and European American males and females have much more prominent glabellar areas than their English counterparts. Comparisons of African American and European American cranial trait distributions show a mixed pattern with significant differences only present between the males (Table 10). The African American males have significantly larger mental eminences than the European American males. On the other hand, European American males have more pronounced glabellar areas and more rounded orbital margins than African Americans males.

The trait distributions of these modern people contrast markedly with that of the Native Americans (Table 10, Fig. 2). In general, the Native American are more robust (i.e., have higher cranial trait scores) and show less sexual dimorphism in their cranial traits. The greater

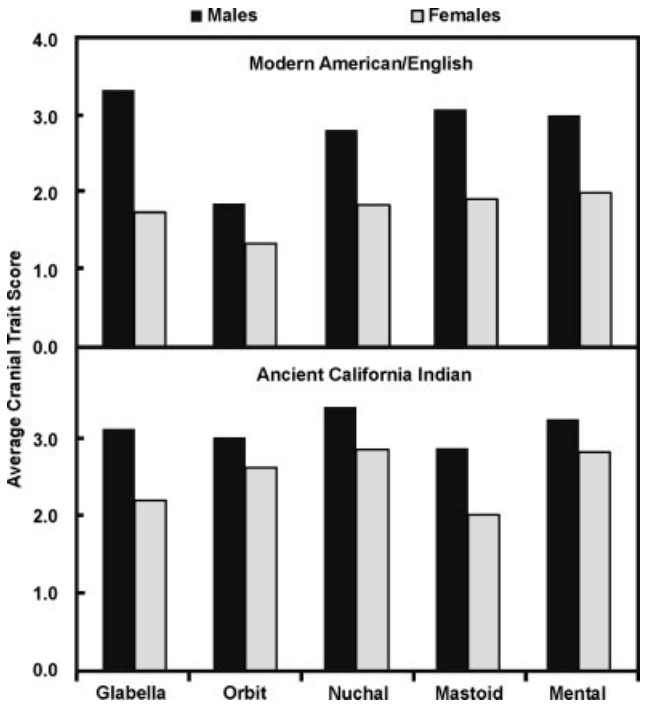

Fig. 2. Average cranial trait scores of males and females for the modern (American/English) and ancient Native American samples. For the Native American sample, sex was inferred based on pelvic morphology.

Native American robusticity is especially apparent for the orbital margin and the nuchal area. The only exceptions to this are the glabellar and mastoid areas, for which the modern males have average robusticity scores that are marginally higher than those of the Native American males. The reduced sexual dimorphism of the Native Americans in comparison with the modern samples, may to some extent be an artifact of pelvic sex determination errors, which are likely to be on the order of 5–10% (Phenice, 1969; Ubelaker and Volk, 2002; Walker, 2005). However, the difference in sexual dimorphism is large enough that it seems likely to involve some significant underlying physiological difference between these populations.

DISCUSSION

The results of this study are promising. The analysis of both the modern known-sex collections and the archaeological collections with sexes inferred from pelvic remains show that the cranial trait scoring system developed for the *Standards for Data Collection From Human Skeletal Remains* (Buikstra and Ubelaker, 1994) can be used effectively by observers with minimal osteological training. The performance of these scores as independent variables in sex determination equations is excellent. For the logistic discriminant analysis models, the proportion of modern crania whose sex was correctly assigned ranged between 84% and 88% depending on the number of traits used in the analysis. This accuracy level is similar to, and in some cases (e.g., Cunha and van Vark, 1991) considerably better than that of conventional discriminant functions based upon cranial measurements (Giles and Elliot, 1963; Giles, 1964; Steyn and Iscan, 1998; Franklin et al., 2005).

For the modern samples, adding information on age at death and population (African American, European American, and English) did not appreciably increase the performance of the discriminant functions developed for the recent cranial samples. Although it is an empirical issue needing further exploration, this suggests that these equations may perform well when used to sex skulls from many different modern human populations with European or African ancestral affinities. In a modest test along these lines, the author scored four cranial traits on 11 modern Greek crania from people of known age and sex at the University of Athens and got promising results: with one misclassification, this small sample produced a correct sex determination rate of 91%.

The practical implications of the subjective aspects of the scoring system described in this paper are difficult to judge. The intraobserver error study conducted as part of the current research and an earlier study using a modified version of the current scoring system (Walrath et al., 2004) both show that the level of intraobserver error inherent in this system is quite low. Tests of interobserver error, in contrast, show that they are larger and in some cases statistically significant (Walrath et al., 2004; Williams and Rogers, 2006). As a result, population differences in the distributions of the scores collected by different observers should be interpreted with caution in the absence of evidence that size of the intraobserver error is less than the magnitude of the differences identified.

Trait distribution comparisons of the three modern populations studied (African Americans, European Americans, and English) show interesting similarities and differences that would have been obscured if traditional racial categories had been used. Americans of African and European ancestry were found to be much more similar to each other than either group is to the Europeans in the English sample (Table 10). In general, American males and females showed more robust cranial traits than their English counterparts. Although the causes of these English–American differences are unknown, they may well be related to differences in environmental influences such as nutritional deficiencies or dietary consistency that have been shown to significantly influence craniofacial growth (Héctor, 1981; Kiliaridis et al., 1985; Miller and German, 1999). This interpretation is consistent with a morphological study of the Terry Collection crania (Jantz and Meadows Jantz, 2000) that concluded that changes in nutrition and health were the likely causes of a secular trend identified in cranial morphology, and a previous study that has linked population difference in the pelvic morphology of these same English and American samples to living conditions conducive to vitamin D deficiency in nineteenth century England (Walker, 2005).

CONCLUSIONS

This study shows that the accuracy of sex determination discriminant functions incorporating scores of the traits traditionally visually assessed by physical anthropologists when making their sex determinations is comparable to that of discriminant functions derived from craniometric observations. The advantages of using cranial trait scores instead of cranial measurements for sex determinations include the ease of data collection, their suitability for use on fragmentary archaeological and forensic materials, and the capacity of scores to encapsulate morphological information that is difficult to quantify using standard anthropometric techniques. The main disadvantage of cranial trait scores as sex determination criteria rests in their greater subjectivity in comparison to observations made with calipers or other anthropometric devices.

The marked difference between these modern people and the ancient Native American sample provides an important note of caution regarding the indiscriminate use of sex determination equations such as those presented in this paper. Population differences such as these mean that discriminant functions that perform well on one population may produce high error rates and sex determination biases when applied to another population. In many respects, the spatial and temporal distribution of such population differences in human cranial morphology is uncharted territory. Exploring this important dimension of human phenotypic variation is likely to be a highly productive area for future investigations.

ACKNOWLEDGMENTS

I thank the Leverhulme Trust and Canon John Oates for allowing me to study the St. Bride's Church skeletal collection. Louise Scheuer provided much help and advice during the work at St. Bride's Church. I am grateful for the assistance of Dave Hunt and Carol Butler at the Smithsonian Institution and Lyman Jellema and Bruce Latimer at the Cleveland Museum of Natural History. I also thank Sotiris Manolis and Constantinos Eliopoulos for their help during my visit to Athens and Alejandro Pérez–Pérez, Judy Suchey, Torstein Sjøvold, Gonzalo Troncho, Daniel Turbón, and my students for participating in the interobserver-error tests. I am indebted to Corina Kellner and Angie Carrithers helping with the scoring of the Santa Barbara Channel area skeletal collections, to Dirk Brandts for his work on Figure 1, and to Cynthia Brock for her editorial advice.

LITERATURE CITED

Acsádi G, Nemeskéri J. 1970. History of human life span and mortality. Budapest: Akadémiai Kiadó.

Broca P. 1875. Instructions craniologiques et craniométriques de la Société d'Anthropologie de Paris. Bull Soc Anthropol 16:534–536.

Buikstra JE, Ubelaker DH. 1994. Standards for data collection from human skeletal remains. Proceedings of a seminar at the Field Museum of Natural History, organized by Jonathan Haas. Fayetteville, AK: Arkansas Archeological Survey. 206p.

Buretic-Tomljanovic A, Giacometti J, Ostojic S, Kapovic M. 2007. Sex-specific differences of craniofacial traits in Croatia: the impact of environment in a small geographic area. Ann Hum Biol 34:296–314.

Buretic-Tomljanovic A, Ostojic S, Kapovic M. 2006. Secular change of craniofacial measures in Croatian younger adults. Am J Hum Biol 18:668–675.

Cobb W. 1932. Human archives. Cleveland, OH: Case Western Reserve University.

Cunha E, van Vark G. 1991. The construction of sex discriminant functions from a large collection of skulls of known sex. Int J Anthropol 6:53–66.

Franklin D, Freedman L, Milne N. 2005. Sexual dimorphism and discriminant function sexing in indigenous South African crania. Homo 55:213–228.

Giles E. 1964. Sex determination by discriminant function analysis of the mandible. Am J Phys Anthropol 22:129–135.

Giles E, Elliot O. 1963. Sex determination by discriminant function analysis of crania. Am J Phys Anthropol 21:53–68.

Graw M, Czarnetzki A, Haffner HT. 1999. The form of the supraorbital margin as a criterion in identification of sex from the skull: investigations based on modern human skulls. Am J Phys Anthropol 108:91–96.

Héctor MP. 1981. Growth of the functional components of the rat skull and its alteration by nutritional effects. A multivariate analysis. Am J Phys Anthropol 56:33–41.

Hosmer D, Lemeshow S. 1989. Applied logistic regression. New York, NY: Wiley.

Hunt DR, Albanese J. 2005. History and demographic composition of the Robert J. Terry anatomical collection. Am J Phys Anthropol 127:406–417.

Jaccard J, Wan CK. 1996. LISREL approaches to interaction effects in multiple regression. Thousand Oaks, CA: Sage Publications.

Jantz RL, Meadows Jantz L. 2000. Secular change in craniofacial morphology. Am J Hum Biol 12:327–338.

Jonke E, Prossinger H, Bookstein FL, Schaefer K, Bernhard M, Freudenthaler JW. 2007. Secular trends in the facial skull from the 19th century to the present, analyzed with geometric morphometrics. Am J Orthod Dentofacial Orthop 132:63–70.

Kanazawa E. 1979. A three-dimensional measurement of the human parietal curvature. Anat Anz 146:363–376.

Kemkes A, Gobel T. 2006. Metric assessment of the "mastoid triangle" for sex determination: a validation study. J Forensic Sci 51:985–989.

Kiliaridis S, Engstrom C, Thilander B. 1985. The relationship between masticatory function and craniofacial morphology. I. A cephalometric longitudinal analysis in the growing rat fed a soft diet. Eur J Orthod 7:273–283.

Konigsberg LW, Hens SM. 1998. Use of ordinal categorical variables in skeletal assessment of sex from the cranium. Am J Phys Anthropol 107:97–112.

McLachlan GJ. 2004. Discriminant analysis and statistical pattern analysis. New York: Wiley.

Meindl RS, Lovejoy CO, Mensforth RP, Don Carlos L. 1985. Accuracy and direction of error in the sexing of the skeleton: implications for paleodemography. Am J Phys Anthropol 68: 79–85.

Miller JP, German RZ. 1999. Protein malnutrition affects the growth trajectories of the craniofacial skeleton in rats. J Nutr 129:2061–2069.

Novotny V, Iscan MY, Loth SR. 1993. Morphologic and osteometric assessment of age, sex, and race from the skull. In: Iscan MY, Helmer RP, editors. Forensic analysis of the skull: craniofacial analysis, reconstruction, and identification. New York: Wiley-Liss. p 71–88.

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 30:297–301.

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

Scheuer L, Bowman J. 1995. Correlation of documentary and skeletal evidence in the St. Bride's crypt population. In: Saunders S, Herring A, editors. Grave reflections: portraying the past through cemetery studies. Toronto: Canadian Scholars Press, Inc. p 49–70.

Smith BH, Garn SM, Hunter WS. 1986. Secular trends in face size. Angle Orthod 56:196–204.

Steyn M, Iscan MY. 1998. Sexual dimorphism in the crania and mandibles of South African whites. Forensic Sci Int 98: 9–16.

Thompson B. 1990. MULTINOR: a fortran program that assists in evaluating multivariate normality. Educ Psychol Meas 50:845–848.

Ubelaker DH, Volk CG. 2002. A test of the Phenice method for the estimation of sex. J Forensic Sci 47:19–24.

Van Vark G, Schaafsma W. 1992. Advances in the quantitative analysis of skeletal morphology. In: Saunders S, Katzenberg M, editors. Skeletal biology of past peoples: research methods. New York: Wiley Liss. p 225–257.

Van Vark GN, van der Sman PG, Dijkema J, Buikstra JE. 1989. Some multivariate tests for differences in sexual dimorphism between human populations. Ann Hum Biol 16:301–310.

Walker PL. 1995. Problems of preservation and sexism in sexing: some lessons from historical collections for paleodemographers. In: Saunders S, Herring A, editors. Grave reflections: portraying the past through skeletal studies. Toronto: Canadian Scholars' Press. p 31–47.

Walker PL. 2000. Bioarchaeological ethics: a historical perspective on the value of human remains. In: Katzenberg MA, Saunders SR, editors. Biological anthropology of the human skeleton. New York: Wiley Liss. p 3–39.

Walker PL. 2004. Caring for the dead: finding a common ground in disputes over museum collections of human remains. In: Grupe G, Peters J, editors. Documenta archaeobiologiae 2: conservation policy and current research. Yearbook of the State Collection of Anthropology and Palaeoanatomy. Rahden/ Westf: Verlag M. Leidorf. p 13–27.

Walker PL. 2005. Greater sciatic notch morphology: sex, age, and population differences. Am J Phys Anthropol 127:385–391.

Walker PL, Drayer F, Siefkin S. 1996. Malibu human skeletal remains: a bioarchaeological analysis. Report California Department of Parks and Recreation, editor: Resource Management Division, California. Department of Parks and Recreation P.O. Box 942896 Sacramento, California 93106.

Walrath DE, Turner P, Bruzek J. 2004. Reliability test of the visual assessment of cranial traits for sex determination. Am J Phys Anthropol 125:132–137.

Williams BA, Rogers T. 2006. Evaluating the accuracy and precision of cranial morphological traits for sex determination. J Forensic Sci 51:729–735.