J Forensic Sci, Mar. 2003, Vol. 48, No. 2
Paper ID JFS2001378_482
Available online at: www.astm.org

John Albanese,[1] *M.Sc.*

A Metric Method for Sex Determination Using the Hipbone and the Femur*†

ABSTRACT: Since the earliest descriptions of the pubis length measurement, it has been recognized that the location of the key landmark in the acetabulum has to be estimated. Using samples from the Terry Collection ($n = 324$) and the Coimbra Collection ($n = 232$), the purpose of this research is to, first, test the reproducibility of a new alternative to the traditional measurement of the pubis, and second, to use the best measurement of the pubis along with other measurements of the hipbone and femur to develop a metric method that can be used with confidence to determine the sex of individuals of various geographic origins and time periods. In this study, it was found that, first, the alternative pubis measurement, known as the superior pubis ramus length (SPRL), can be collected more reliably with less mean intra-observer error (0.57%) than the more commonly used manner of measuring the pubis (2.7%). Second, a logistic regression sex determination method using the SPRL, along with other measurements of the hipbone and femur, has an allocation accuracy of 90% to 98.5% (depending on the model used and the manner of testing) across independent samples. Third, traditional racial categorization was irrelevant to the accuracy of the method. Fourth, measurement error greater than 2% in the measurement of the pubis can be the difference between a correct and an incorrect allocation of sex, particularly in borderline cases.

KEYWORDS: forensic science, forensic anthropology, human identification, sex determination, skeletal, pelvis

Determination of sex of an unknown individual is one of the critical questions addressed when human skeletal remains are found in both forensic investigations and in studies of past populations. Beginning with the earliest investigations into the development of sex determination methods, the pelvis in general and the pubis in particular have been recognized as the best sources of information for determining the sex of an unknown individual (1). However, since the earliest descriptions of the pubis length measurement, it has been recognized that the location of the key landmark in the acetabulum has to be estimated. While the description of the pubis length measurement is well defined, relying on an estimated landmark position will result in relatively high measurement error. Using samples from the Terry Collection and the Coimbra Collection, the purpose of this research is to, first, test the reproducibility of an alternative to the traditional measurement of the pubis; second, to present an alternative approach to constructing a reference sample for developing sex determination methods; and third, to use the best measurement of the pubis along with other measurements of the hipbone and femur to develop a metric method that can be used to determine the sex of individuals of various geographic origins and time periods.

[1] Department of Anthropology, McMaster University, Hamilton, Ontario, L8S 4L9, Canada.

* Funding was provided in part by a Doctoral Fellowship from the Social Science and Humanities Research Council of Canada (grant # 752-99-1267); School of Graduate Studies, McMaster University; and an Ontario Graduate Scholarship.

* Some equipment was provided by the Department of Anthropology, McMaster University.

† Based on Research Presented at the 70th Annual Meeting of the American Association of Physical Anthropologists in Kansas City, Missouri, March 18–31, 2001; and the 29th Annual Meeting of the Canadian Association for Physical Anthropology in Winnipeg, Manitoba, October 25–27, 2001.

Received 14 Nov. 2001; and in revised form 5 Oct. 2002; accepted 21 Oct. 2002; published 3 Feb. 2003.

Measurement of the Pubis and Metric Approaches for Determining Sex

The length of the pubis is usually defined as the distance between the superior margin of the pubic symphysis and the acetabular margin of the pubis. Although the acetabular margin of the pubis is clearly visible in subadults, in most individuals, the pubis, ischium and ilium are completely fused by the age when the pubis is actually useful for sex determination (1,2). Fusion begins around the time of puberty and the three elements are completely fused by the mid to late teens (3). Thus, while the description of the pubis length measurement may be clear, difficulty in locating the acetabular margin of the pubis in adults often results in problems with the reproducibility of the measurement.

The difficulty in locating the acetabular landmark has been recognized for as long as the pubis has been measured. In one of the earliest descriptions of the pubis length measurement, Schultz (4) used a diagram of what he referred to as an "infantile" hipbone to highlight the landmark that is difficult or impossible to locate in adults. Many descriptions of the pubis length that have been published in North America in the last 50 years quote or paraphrase Washburn's (5) description and suggestions for estimating the point where the pubis, ischium and ilium meet in adults (for example see 6, 7). These "hints" include looking for irregularities and notches in the acetabulum and on the inside surface of the pubis, and holding the acetabulum up to the light to look for differences in bone thickness. In Europe, the same difficulties in locating the point are described in various sources (2). The confusion, which still persists in North America and Europe, surrounding the position of the acetabular landmark was summarized by Olivier over 35 years ago:

> It would appear that the [point where the ischium, ilium and pubis meet] in fully ossified bone is variously located by different authors. For Schultz it is the point A in the figure

> (cotyloid point) but this is a variable location, situated where the inner border of the hip bone[s] meet each other. In monkeys it is easy to find. In [humans] it is often a notch of the articular margin at this level as well as a more internal roughness of the bone. Several authors have defined this point more exactly: for Genovés it is that point of the inner articular margin which is nearest to the anterosuperior iliac spine. For Gaillard it is the intersection of the long axis of the pubis and the ischium (2 pp. 248–9).

Despite the problems with locating the acetabular landmark, this measurement is the only pubis measurement described in two recent measurement reference volumes (see 6,8).

Building on the work of Thieme (9) and Washurn (5), an alternative to the pubis length was described by Schulter-Ellis and colleagues (10,11) and used to develop a sex determination method that had great promise, but does not seem to be used widely. Their measurement of the pubis (PS-A) is described as the length between the superior margin of the pubic symphysis to the nearest rim of the acetabulum. They note that the main advantage of the PS-A is that the problems associated with estimating the position of the landmark where the pubis, ischium and ilium meet is avoided, but they do not provide any evidence to support this statement. They do not show that their PS-A, with a floating landmark on the acetabulum rim, is reproducible with a reasonably low margin of error nor do they demonstrate that it is more reproducible than the traditional pubis length.

The sex determination method developed by Schulter-Ellis and colleagues may not be used widely because it is needlessly complex and was not adequately tested. First, they use ratios as variables in discriminant function equations. Second, they use their *a priori* knowledge of documented sex in order to decide when to include femur measurements in cases where the pelvic discriminant functions were inconclusive or incorrect. Third, the method was tested on the same sample that was used to develop the method. Lastly, they used 100 "Whites" and 100 "Blacks" to develop two different methods for sex determination without presenting any evidence that suggests that race-specific methods are necessary. Although the allocation accuracy of their method was as high as 96%, the method of testing and other complexities suggest that their method may not perform well in actual cases. These problems may contribute to the lack of widespread use of this method, particularly when there are other good options for sex determination when the pelvis is complete.

Bruzek (12) independently tested the method developed by Schulter-Ellis and colleagues using identified skeletons from the Paris and Coimbra Collections and found that the Schulter-Ellis method had an allocation accuracy of over 90%: Paris males, 91.5%; Paris females, 100%; Coimbra males, 93.5%; and Coimbra females 95.1%. The major shortcoming was that there was an 8.5% difference in allocation accuracy between Paris Collection males and females, a problem that is common with other metric sex determination methods (13).

Materials and Methods

Representativeness, Sources of Variation and Sample Selection

The composition of the reference sample used to develop sex determination methods can have an impact on the applicability of the methods that are developed (13–17). It has been assumed for decades that the Terry Collection and other similar collections are not representative of populations in the U.S. (18) and may no longer be useful for the development of forensic identification methods (19). However, with only a few exceptions (for example, 19,20), this lack of representativeness has never been investigated in detail and the impact of the biases of the collection on skeletal variation—rather than the applicability of methods—has never been assessed (21). However, if representativeness is considered in shades of gray rather than as black and white, then the level of representativeness of any collection will vary depending on the research question. Furthermore, representativeness can be maximized and biases can be minimized or even exploited with alternative approaches to sampling in order to address specific research questions (22). With careful sampling, it is possible to construct a large reference sample from the Terry Collection or a combination of identified collections that captures a wide range of modern human variation and that can be used to develop highly accurate sex determination methods that can be applied to 21st century populations. The key is considering the demographic data (age, year of birth, etc.) and the historical details such as socioeconomic, political and legal issues associated with the construction of the collections. In the past, this information has been used to dismiss the collection as antiquated and not representative. In this study the same information is used to minimize bias and maximize representativeness of human variation when constructing a reference sample. The underlying assumption in sample selection for this study is that if a greater amount of human variation is included in the sample used to develop a sex determination method, then that method can be applied with confidence in a wide range of cases. The samples used in this study are from two very different skeletal collections: the Robert J. Terry Anatomical Collection (Smithsonian Institution, Washington, D.C.) and the Coimbra Identified Skeletal Collection (Museum of Anthropology, University of Coimbra, Coimbra, Portugal). Some details regarding the Coimbra Collection[2] can be found in publications by Rocha (23) and Cunha (24). In summary, careful planning went into the assembly of the Coimbra Collection and supporting documentation but did not result in a random sampling of either the cemetery from which the skeletons were derived nor the greater population of the District of Coimbra (22,25). Despite these biases, the collection has been used to develop and test forensic identification methods and in palaeopathological investigations (24, see also 12,26,27)

Some details regarding the size and basic demographic profile of the Terry Collection have been described in numerous sources (for example, 28,29, but see also 22). However, both the demographic profile and the historical context of the collection have had an impact on the variation that is present in the collection and must be considered in some detail when constructing a reference sample (21). The Terry Collection is the result of the joint efforts of Drs. Robert J. Terry and Mildred Trotter over six decades at the medical school at Washington University in St. Louis, Missouri. The collection was derived from anatomy school cadavers, which were mostly unclaimed bodies from various hospitals and institutions in Missouri. Only a relatively small number who died after 1955, about 10% of the entire collection, were people who bequeathed their bodies for medical use. This 10% of the collection has been described as middle class American (20,30), however, Ericksen (20) found that the differences between the bequeathed and nonbequeathed individuals are not conclusive using data from the proximal femur. She hypothesizes that the practice of bequeath-

[2] There are two series of identified skulls and one series of identified complete skeletons at the University of Coimbra (23,24). In this paper, Coimbra Collection refers to the 505 identified skeletons (Colecção de Esqueletos Identificados).

ment can be associated with higher socio-economic status at the time of death even though the individual may have lived under different conditions during their growth and development period. Conversely, the other 90% of the collection consists of individuals who were possibly of very low socio-economic status *only at the time of death.* About 55% of the 1618 individuals who died before 1955, died during the Great Depression (1929–1939) and may not necessarily have lived in poverty during their growth period. As Terry noted several decades ago when describing the cadavers before dissection and maceration:

> . . .these bodies commonly bear the marks of undernourishment and in many cases of the wasting effects of chronic ailment that brought death. Whereas *these conditions scarcely affect at all the longitudinal measurements* they render some of the transverse and circumferential measurements of questionable value (31 pp 435; emphasis added).

A similar impact on the skeleton should be expected. For a number of social and historical reasons, the Terry Collection, like other skeletal collections in North America, *had* a very unbalanced sex ratio. Unlike other skeletal collections and largely through Trotter's efforts, this imbalance was, in part, corrected in the Terry Collection (32). Trotter was instrumental in drafting a major change in the Missouri laws on bequeathment of human remains. After this change in the mid-1950s, Trotter focused on including the skeletons of younger "White" females in the collection (32).

The trend in physical and forensic anthropological research in the last 60 years in the U.S. has been to randomly sample major collections and to develop race-specific sex determination methods in order to control for differences in sexual dimorphism between populations (for example, 10,11,18,33–37). In these examples, a major source or the exclusive source of data is the Terry Collection. Much of the variation attributed to racial differences in sexual dimorphism in various studies that have sampled the Terry Collection may be attributable to Trotter's efforts to correct for the lack of young "White" females. In any random sample of the Terry Collection, there are different proportions of males and females of different "races" from various years of birth and age cohorts. A random sampling of the collection for the development of race-specific sex determination methods will result in a poor sampling of sexual dimorphism and the pseudo-significance of "race" because it is so closely correlated to age at death, year of birth and other variables associated with how the collection was constructed. In a random sample of "Whites," there are a disproportionately high number of younger adult females born early in the 20th century compared with a male sample composed of older individuals born in the middle decades of the 19th century. In a randomly selected sample of "Blacks," the impact of Trotter's approach to adjusting the collection has less of an impact because there are many more "Black" females than "White" females in the collection that were born in the middle to the end of the 19th century. Because of Trotter's approach to collecting, a comparison of "Black" and "White" females, is a comparison of age and year of birth differences between younger females born in the 20th century who were described as "White" to older females born late in the 19th century who were described as "Negro" when they were included in the collection. Using logistic regression (see below for details), it is possible to assess "race" with an allocation accuracy of 69.5% using only age at death and year of birth (no skeletal data) in a sample that includes all the females in the Terry Collection over 18 years of age for whom age and year of birth data are available (n = 671).

Rather than control for "race" in a sample from the Terry Collection, the underlying sources of variation (age at death and year of birth) that are highly correlated with "race" due to the collection process need to be considered when selecting samples. Giles (38, pp. 102) minimized the importance of the effects of age on sex determination but he found that cranial "discriminant functions tend to mis-classify younger males and older females." He went on to state that this pattern of misclassification may be "present but undetected" in morphological sex determination methods. Walker (39) and Meindl and colleagues (14) confirmed Gile's hypothesis and have shown that age at death can be a critical factor in the level of sexual dimorphism in the cranium. Walker found that younger males have a morphology that is referred to as a typically female pattern and older females have a morphology that is referred to as a typically male pattern. After reviewing several different cranial sex assessment methods, Meindl and colleagues found that "greater age produces an increasingly male morphology" (14 pp. 81). Anderson (40) has suggested that a similar but inverted pattern may occur with the ventral arc. Although his sample size is small and not distributed across a wide range of ages, Anderson reports that the ventral arc is less visible in females under 20 years of age and becomes more visible in males older than 70 years of age. When testing the Phenice method (41), which includes the ventral arc, Lovell (42) found that accuracy decreased with age at death because of age-related irregularities in the pubic bone. Sutherland and Suchey (43) reproduce Anderson's results but not Lovell's conclusions regarding age-related changes in the ventral arc. Still other research suggests that age may also be a factor in the absolute length of the pubis, where growth of the pubis continues into the third decade of life in females (44). However, others have attributed this age related difference in pubis length as a misinterpretation of differential mortality (45). Lastly, the effects on sexual dimorphism of well documented secular change in the late 19th and throughout the 20th centuries are not well defined or fully understood. Because of the manner in which both the Terry and Coimbra Collections were collected, a random approach to sample selection will result in a poor and uneven sampling of variation associated with age at death and year of birth. A sex determination method developed from such a sample would be of limited use.

In this study an effort was made to include the full range of adult ages and a wide range of years of birth represented in each collection. Any variation that may possibly be related to age at death or secular change is sampled and included in the analysis even if year of birth and age at death are not included as predictor variables. Figure 1 is a scatter plot of age at death by year of birth of the samples from both collections. There are adults that are older and younger than 40 years of age with years of birth before and after 1900. A portion of the sample falls into each of the four quadrants of the scatter plot in Fig. 1. All years of birth and ages are well represented. The diagonal linear appearance of the plot is the result of the limits of the Terry and Coimbra Collections in years of birth. The sample from the Terry collection is 324 individuals and from the Coimbra Collection it is 232 individuals. Years of birth range from approximately 1832 to 1913 for the Coimbra Collection sample and 1850 to 1930 in the Terry Collection sample. Ages range from 19 to 79 in the samples from both collections. The upper age limit was arbitrarily set to avoid missing data related to extreme joint problems and misleading data that resulted from age-related loss of robusticity. Many of the individuals in the Terry Collection died before the 1935 Social Security Act and would have continued to work in manual labor employment well beyond 65 years of age. The skeletal impact of this behavior adds to the complexity of

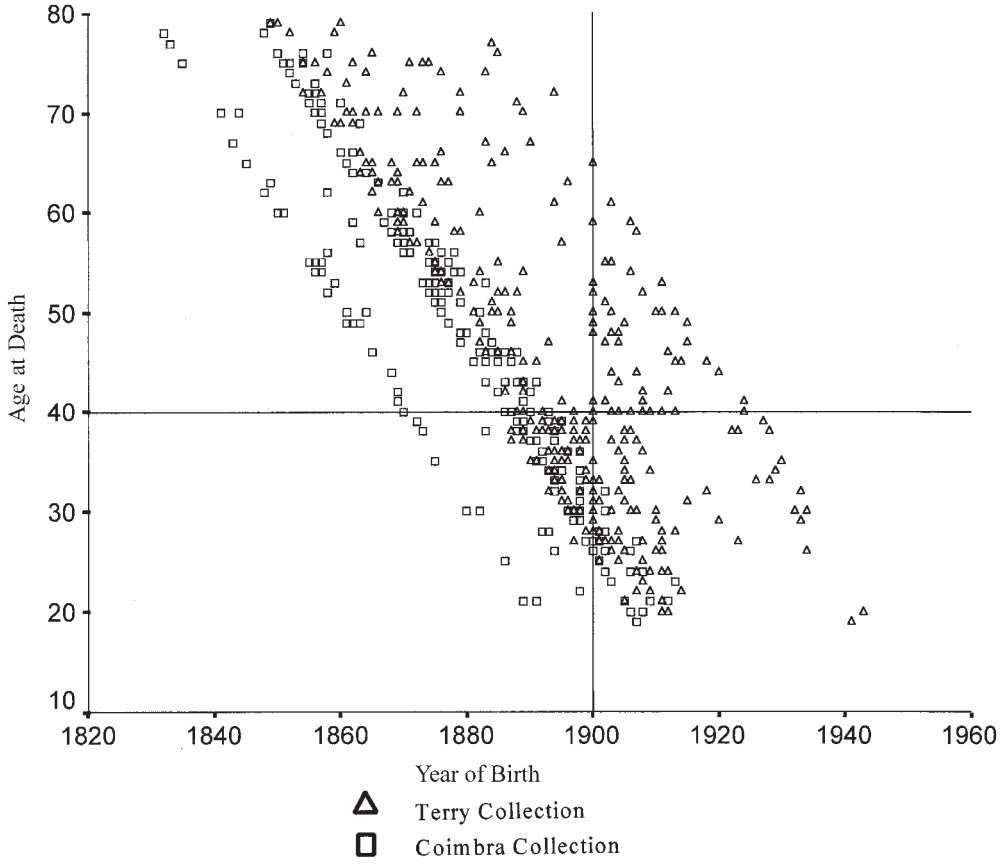

FIG. 1—*Age at death by year of birth for the entire sample from the Terry and Coimbra Collections (n = 556).*

interpreting skeletal variation in the Terry Collection since it is in contrast to the wasting from malnutrition and diseases described by Dr. Terry. The lower age limit is dependent on chronological and biological criteria. First, documented age at death is 18 years or greater. Second, all epiphyses, with the exception of the sternal end of the clavicle had to be at least partially fused. In some cases epiphyseal lines were visible but epiphyses were never separate from diaphyses on any long bones.

Measurements Collected

Several standard femur and hipbone measurements were collected for the entire sample including hipbone height, iliac breadth, pubis length, ischium length, maximum femur length, maximum femur head diameter, anterior-posterior diameter of the femur at mid-shaft, transverse diameter of the femur at mid-shaft, and epicondylar breadth of the femur (2,6,7). These measurements are defined in Table 1. Two new measurements were also collected. The new measurement of the pubis is referred to as the superior pubis ramus length (SPRL). The second measurement is an alternative for measuring the ischium and is referred to as the Acetabular-Ischium Length (AIL). All measurements were collected by the author to the nearest millimeter.

The SPRL is measured using sliding calipers from the superior margin of the pubic symphysis to the superior-anterior apex of the lunate surface in the acetabulum. Unlike the measurement of the pubis described by Schulter-Ellis and colleagues, the SPRL described here has a fixed, easily recognized landmark in the acetabulum. Occasionally, this landmark may be affected when there are extreme arthritic changes to the acetabulum. In most cases, it is easy to measure around any arthritic lipping on the rim of the acetabulum. In the few extreme cases, using spreading calipers may be advantageous. The AIL is measured from the same landmark on the acetabulum to the most inferior point on the ischium and not perpendicular to the SPRL. The maximum length should be measured. Both measurements are illustrated in Fig. 2.

Intra-observer Error and Measurement of the Pubis

In a randomly selected sub-sample of just over 10% of the entire sample ($n = 65$), all measurements were re-collected to test their reproducibility and the level of intra-observer error. The sample was divided about equally between males and females and approximately proportionally by collection: 13 males and 13 females from the Coimbra Collection, and 19 females and 20 males from the Terry Collection. To calculate the percent intra-observer error, the absolute difference between the two measurements was divided by the first measurement and then multiplied by 100 for each individual in the intra-observer error sample. The mean error was then calculated. Only the pubis and ischium data are presented here. For the traditional pubis length measurement, the mean intra-observer error is 2.7%. For the new SPRL measurement, the intra-observer error is 0.57%. Figure 3 is a scatter plot of the measurement error on a case-by-case basis for each of the 65 individuals. The errors for the traditional pubis measurement by the SPRL error are graphed. For the SPRL, more than half of the sample has an error of zero,

TABLE 1—*Definitions of traditional measurements*.*

Measurement	Definition
Hipbone height	Distance from the most inferior point on the ischial tuberosity to the most superior point on the iliac crest (2,6). Measured using an osteometric board. Also known as innominate height.
Iliac breadth	Distance from the anterior superior iliac spine to the posterior superior iliac spine (2,6). Measured using an osteometric board.[†]
Pubis length[‡§]	Distance from the superior margin of the pubic symphysis to the point in the acetabulum where the ischium, ilium, and pubis meet (2,6,7). Measured using sliding calipers.
Ischium length[§]	Distance from the point in the acetabulum where the ischium, ilium, and pubis meet to the most inferior point on the ischial tuberosity perpendicular to the pubis length (6,7). Measured using sliding calipers.
Maximum femur length	Distance from the femur head to the most inferior point on the medial condyle (2,6,7). Measured using an osteometric board.
Maximum diameter of femur head	Maximum diameter of the head of the femur at the border of the articular surface (6). Measured using sliding calipers.
Epicondylar breadth of femur	Distance between the most projecting points on the medial and lateral condyles (2,6). Measured using sliding calipers.

*Anterior-posterior diameter of the femur and transverse diameter of the femur at mid-shaft are not statistically significant predictor variables and were not used for other illustrative purposes, and have been left out.

[†]Some sources recommend that spreading calipers be used, however, using an osteometric board was found to be much quicker and easier while providing identical results.

[‡]Various definitions of the acetabular landmark for the pubis length are discussed in the text.

[§]Definitions of the SRPL and AIL are in the text and illustrated in Fig. 2.

and in almost 95% of the cases, the error is less than 2% or 1 mm. For the traditional pubis length measurement, less than 30% of the cases are below 2%, while the majority of cases have an error greater than 2%.

A review of all the cases revealed that there is no pattern to the errors for either measurement. Using the Spearman correlation statistic, correlations were weak and not significant between the measurement errors and the collection (Terry or Coimbra), the date of data collection, or sex of the individual. Data collection was consistent throughout the data collection period and there is no relationship between errors in the traditional pubis length and the SPRL measurements. The larger intra-observer error in the traditional pubis length measurement can be attributed to the difficulty in locating the acetabular landmark.

The intra-observer error of the AIL and the traditional ischium length (not shown graphically) follows a different pattern than the comparison of the SPRL and the traditional pubis length measurement. Both the AIL and the traditional ischium length measurements have a low, virtually identical intra-observer error at 0.98% and 0.94%, respectively[3].

Statistical Approach

Several multivariate statistical approaches can be used to predict a binary dependent variable, such as sex, from a group of independent variables. Discriminant function analysis is the most commonly used approach in skeletal sex determination methods; however, other more statistically robust methods are available. One underused yet very powerful approach is logistic regression.

[3] Extensive tests of inter-observer measurement error of all of the pubis and ischium measurements will be presented in a future publication.

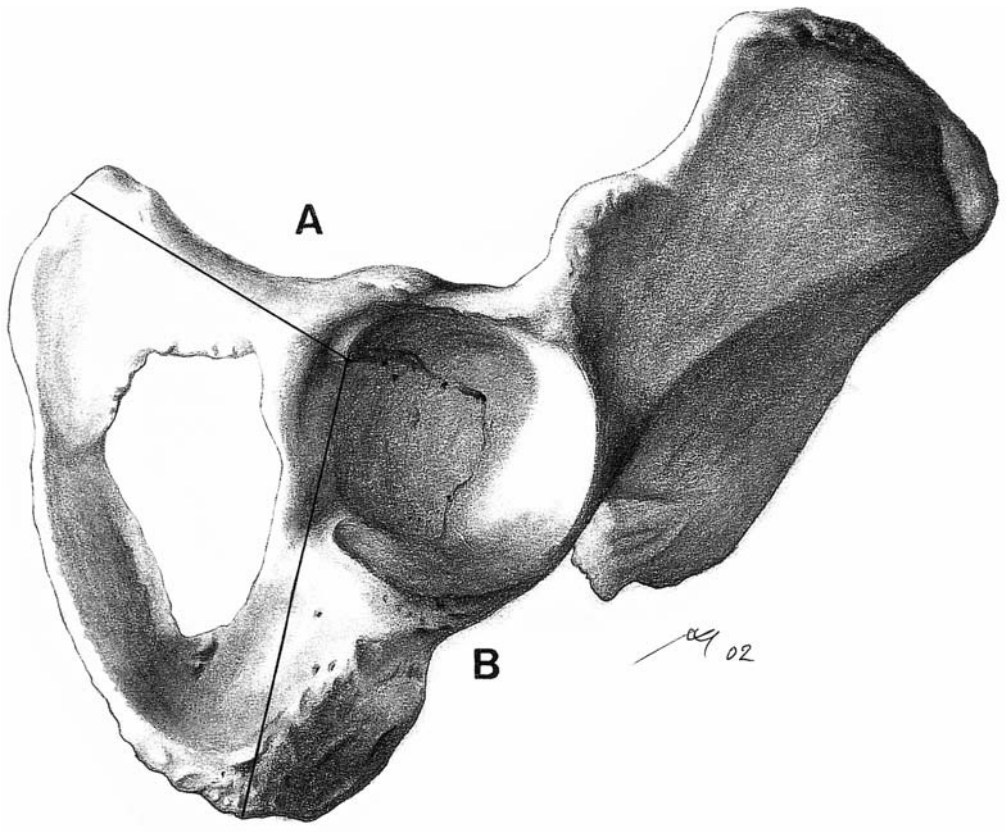

FIG. 2—*The superior pubis ramus length (A) and the acetabular ischium length (B).*

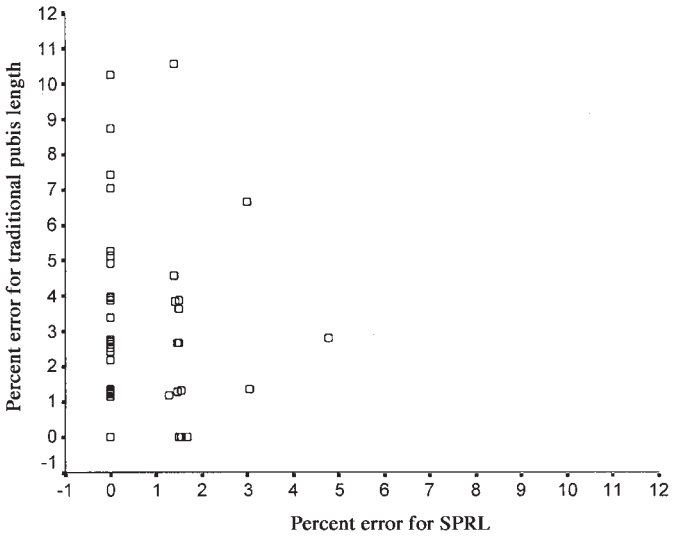

FIG. 3—*Plot of percent measurement error of the traditional pubis length* with percent measurement error for the superior pubis ramus length (SPRL)† for each case (n = 65).*

*Mean intra-observer measurement error for the traditional pubis length is 2.7%.

†Mean intra-observer measurement error for the SPRL is 0.57%.

Norusis (46) suggests several reasons why logistic regression is a better choice than discriminant function analysis when predicting a binary dependent variable (see also 13 for a discussion of logistic regression and sex determination). First, with discriminant function analysis, there is the assumption of a normal distribution of the independent variables. Although skeletal metric data are usually normally distributed when samples are large, logistic regression does not require a normal distribution to optimize prediction accuracy and categorical variables can be used as independent variables along with metric data. Second, logistic regression analysis does not require equal variance-covariance matrices in the two groups (female and male), a condition that is necessary for discriminant function analysis. Norusis adds that "even when the assumptions required for discriminant analysis are satisfied, logistic regression still performs well" (46, pp.119). Aside from the underlying statistical assumptions, a major benefit of a logistic regression model over a discriminant function model when allocating individuals—as opposed to discriminating between groups—is that the probability of the event is calculated. Separate posterior probability and/or typicality probability statistics must be considered for an analogous approach when using discriminant functions. With a few exceptions (36,47,48), this approach is missing from most discriminant function sex determination methods.

The logistic model is an S-shaped function of the form

$$P = \frac{1}{(1 + e^{-Z})}$$

where P is the probability of the event (male or not male in this case) and Z is a linear combination of the independent variables such as,

$$Z = \beta_o + \beta_1 X_1 + \beta_2 X_2$$

Calculated probabilities are always between 0 and 1. If P is greater than 0.5, then the individual is considered male. If P is less than 0.5, the individual is considered female. For example, if P = 0.84, the individual is male and there is an 84% probability that the individual is male. Alternatively, there is only a 16% probability (1-P) that the individual is female. Despite the clear benefits of using the logistic regression model, with only a few exceptions (13, 49–51) it has been underutilized in forensic anthropological research (13).

There are four main methods used to assess the fit of a logistic model to the data. First, any later version of SPSS (version 9.0 for Windows was used for this study) will automatically calculate the allocation accuracy of the model applied to the sample used to develop the model. Second, a histogram of probabilities of the sample used to develop the method can be generated to assess the range of probability scores. Ideally, the histogram should have few probabilities in the mid range and two large spikes near 0 and 1 for females and males, respectively. Third, a goodness of fit statistic known as the −2 log likelihood (−2LL) is calculated. The lower the −2LL statistic, the better the fit of the model to the data. Fourth, the allocation accuracy can be calculated for a hold-out sample not used to develop the method.

Norusis (46) notes that all the same problems with variable selection algorithms found in regression and discriminant functions can also be found in logistic regression. The two major approaches to automated variable selection involve the Wald statistic and the Likelihood Ratio (LR) test. Either can be used in a forward or backward stepwise procedure. Both work equally well provided that sample sizes are large (46). In this study, experiments with both statistics in the forward and backward stepwise procedures resulted in exactly the same model. For consistency, the Forward LR was selected whenever automated variable selection was used.

A two-step approach was used for variable selection. First, various possible scenarios were considered such as complete recovery of skeletal remains (hipbone and femur), dismemberment, and postmortem damage to the pubis, ischium, ilium and/or femur. Second, within each of these scenarios, the Forward LR option was used to select the best predictors from the available variables. Because of missing data for some individuals, sample size varies for each model.

The sample of over 550 individuals from both the Terry and Coimbra Collections was divided into two sub-samples. Models were developed using a sub-sample of 422 individuals (75%) from both collections, which is referred to as the ***model sample***. The model was tested on a hold-out sample of 134 individuals (25%) referred to as the ***test sample*** which were not used to develop the models. The test sample was selected randomly from the overall sample using the random sample selection feature in SPSS. The composition by sex and collection of the model and test samples are presented in Table 2.

TABLE 2—*Model sample* and test sample by collection and sex.*

	Model Sample			Test Sample			
	Te	Co	Total	Te	Co	Total	Total
F	129	92	221	41	26	67	288
M	109	92	201	45	22	67	268
T	238	184	422	86	48	134	556

*The sample of 422 is the pool from which models were developed. Sample size varies from 401–418 for various models because of missing data. See Table 3 for specific sample sizes.

Abbreviations: F = female, M = male, T = total, Te = Terry Collection, Co = Coimbra Collection.

Results

The best-fit model, which included hipbone height, iliac breadth, SPRL, maximum femur head diameter, and epicondylar breadth of the femur, correctly allocated 98% of the model sample. Allocation accuracy for males and females was identical at 98%. When the method was tested on the hold-out test sample, the results were slightly better: 98.5% of both the males and the females were allocated correctly. Stated another way, out of 134 test cases, only one male and one female were allocated incorrectly. Rather than graph the probabilities of the model sample to test the fit of the model, probabilities for each individual in the test sample are presented in Fig. 4. Note the near-perfect fit of the model. With the exception of the one male and one female who were allocated incorrectly, all the males have high probabilities (≥0.80) and all the females have low probabilities (≤0.2). In 93% of the test cases, the scores indicated that there is 90% or greater probability of a correct allocation. Thus, there are both consistently high probability scores for each individual allocation and a high overall allocation accuracy. A summary of the various models and allocation accuracies are presented in Table 3. Coefficients for each model are presented in Table 4.

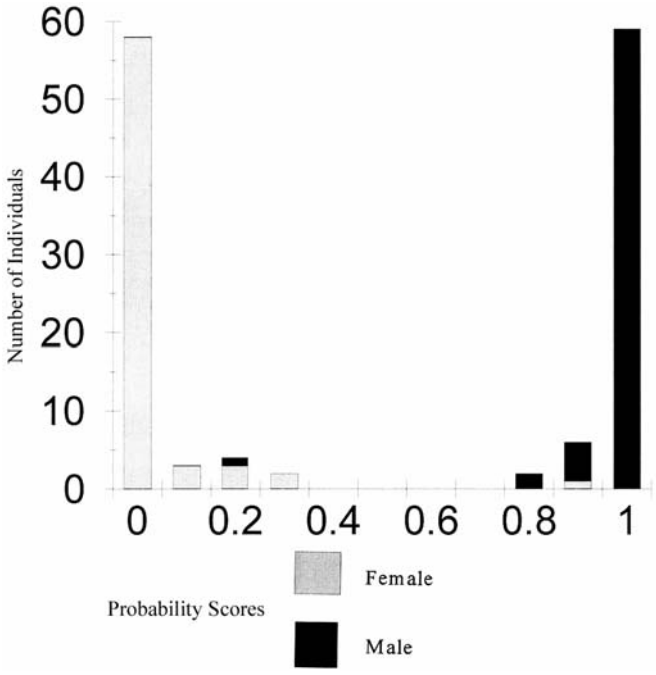

FIG. 4—*Calculated probabilities (P) for the hold-out test sample.*

In various situations, only fragmentary remains are available for analysis. For example, if only the hipbone is available for analysis, then Model 4 would be used to assess sex. With this model, allocation accuracy for the model sample is 96% for each sex. When applied to the test sample, allocation accuracy is 96%: 94% for females and 98.5% for males. The pubis can be susceptible to postmortem damage, and therefore, two models that do not require the SPRL (Models 20 and 26) have also been developed and tested. Even with a test sample that includes very small individuals from the Coimbra Collection and larger individuals from the Terry Collection, allocation accuracy is 97% for females and 95.5% for males (96% overall) for Model 20, and 95.5% for females and 91% for males (93% overall) for Model 26. Model 26 does not require either the SPRL or the AIL, and therefore, can be further tested using data from the Forensic Anthropology Data Bank (FDB). The FDB sample ($n = 213$) used for testing was selected to include only data from positively identified individuals 19 years of age and older and no data from the Terry Collection. Allocation accuracy for Model 26 when applied to the FDB data was 94% overall (89%

TABLE 3—*Assessment of the fit of logistic regression models.**

			Tested on Model Sample					Tested on Hold-Out Sample ($n = 67$ males and 67 females)				
			Female Correct‡		Male Correct‡			Female Correct		Male Correct		
Model†	n	−2LL	n	%	n	%	Total	n	%	n	%	Total
1	401	41	203	98.1	190	97.9	98.0	66	98.5	66	98.5	98.5
2	401	85	198	95.7	184	94.9	95.3	65	97.0	65	97.0	97.0
3	414	75	211	97.7	188	95.0	96.4	64	95.5	66	98.5	97.0
4	404	91	201	96.2	187	95.9	96.0	63	94.0	66	98.5	96.3
5	418	119	208	95.0	186	93.5	94.3	65	97.0	66	98.5	97.8
8	402	46	203	97.6	188	96.9	97.3	65	97.0	66	98.5	97.8
10	402	56	201	97.1	190	97.4	97.3	64	95.5	65	97.0	96.3
11	415	86	208	96.3	186	93.5	95.0	63	94.0	65	97.0	95.5
12	416	81	210	96.8	190	95.5	96.2	62	92.5	65	97.0	94.8
13	402	107	199	96.1	185	94.9	95.5	62	92.5	64	95.5	94.0
14	403	95	195	93.8	183	93.9	93.8	62	92.5	62	92.5	92.5
15	416	98	207	95.4	190	95.5	95.4	62	92.5	64	95.5	94.0
17§	419	111	207	95.0	187	93.0	94.0	60	89.5	63	94.0	91.8
19§	419	155	204	93.6	187	93.0	93.3	63	94.0	61	91.0	92.5
20	401	117	195	94.2	182	93.8	94.0	65	97.0	64	95.5	96.3
22	418	134	209	95.4	183	92.0	93.8	63	94.0	62	92.5	93.3
26	402	122	194	93.7	180	92.3	93.0	64	95.5	61	91.0	93.3

*Only models with an overall allocation accuracy of 90% or higher, and a difference in allocation accuracy between males and females of less than 5% in both the model sample and the test sample are presented. Hence, there are gaps in the model numbers.

†See Table 4 for the coefficient for each model.

‡Sample size varies with each model.

§Enter method instead of Forward: LR was used to select independent predictor variable.

TABLE 4—*Coefficients for logistic models.**

Model†	Hip Bone Height	Iliac Breadth	SPRL	AIL	Max Di. Of Femur Head	Epicondylar Breadth	Constant
1	0.5950	−0.5192	−1.1104		1.1696	0.5893	−61.5345
2		−0.1600	−0.5951	0.2920	1.0365	0.3901	−30.5291
3	0.2572		−0.9852		0.7303	0.3177	−40.5313
4	0.4323	−0.2217	−0.7404	0.3412			−30.3590
5	0.3084		−0.8092	0.2657			−28.3111
8	0.4868	−0.4903	−1.0597	0.2901	1.6241		−45.2528
10	0.5267	−0.3785	−0.8156			0.7758	−52.8262
11	0.2896		−0.8794			0.5783	−42.6362
12	0.2694		−0.9850		1.0484		−32.7486
13		−0.0963	−0.5042	0.4154		0.6453	−32.9320
14		−0.1428	−0.6382	0.3447	1.3968		−22.6111
15			−0.6964	0.2171	0.8489	0.3285	−31.3620
17			−0.6285		0.9906	0.4058	−30.9086
19			−0.4776			0.8032	−28.6274
20	0.2007	−0.4445		0.1734	0.5697	0.3915	−41.9071
22	0.3943		−0.8007				−25.4936
26	0.2326	−0.4321			0.6235	0.4085	−40.2291

*See Table 3 for an assessment of the fit of each model.
†See text for an example of the application of a logistic regression model to determine sex.

for females and 96% for males). This approach to testing is equivalent to applying Model 26 in 213 forensic cases and suggests that older collections, such as the Terry Collection, can still be very useful for developing (and not just testing) forensic methods provided that the reference samples are carefully constructed.

The combination of high accuracy and small differences between male and female accuracy in a diverse test sample is essential for any useful sex determination method (13). Therefore, in order to maximize the reliability of the models in predicting sex in various situations, several criteria had to be met for models to be included in Tables 3 and 4. First, in both the model sample and the test sample, overall allocation accuracy had to be 90% or higher. Second, the difference in allocation accuracy between males and females had to be less than 5%. Both criteria are somewhat arbitrary yet more strict than other published guidelines (13) in order to establish confidence in allocation when the methods are applied in actual cases. A model developed in this study (Model 6) that is analogous to Washburn's (5) ischium-pubis index or Novotny's ischio-pubic index (52) does not meet either the 90% accuracy or the 5% sex difference criteria. Model 6 performed acceptably on the model sample: 91% of females and 90% of males were allocated correctly. However, when applied to the test sample, the allocation accuracy for Model 6 is only 88%: 85% and 91% for females and males, respectively. The pubis and the ischium are important sources of information when determining sex but relying only on measurements of these two bones may produce misleading results.

When using sex determination methods, it is as important to know when a method may fail as when it may provide useful information. An incorrect assessment of sex can be very misleading in both archaeological contexts and forensic investigations. A case by case review of all the individuals used in the model sample that are allocated incorrectly using Model 1 revealed a pattern ($n = 8$ or 2% of 401). Females with unusually large joints relative to the size of their pubis (SPRL) are allocated incorrectly. All the males with extremely small joints relative to their pubis (SPRL) are allocated incorrectly. In contrast to many other univariate and multivariate metric methods for determining sex, it is not simply the shorter, less robust males and the taller, more robust females that are allocated incorrectly (for example, see 53). Very short males from the Coimbra Collection and taller females from the Terry Collection were consistently classified correctly even though the Coimbra males are so much shorter than the Terry Collection females that there is no significant difference in femur length ($t = -1.531$, $p = 0.128$, $n = 64$ females, $n = 58$ males) between males from one collection and females from another when birth cohort is held constant (1875–1899). Furthermore, there are no other patterns in misallocation related to age (including 19–21 year old individuals who had not completed growth), sex, or "race." Just over half of the individuals in the sample from the Terry Collection used in this study were described as "Negro" at the time they were included in the collection. The high allocation for all models presented and no difference in allocation accuracy by "race" strongly suggest that race-specific sex determination methods are not necessary.

The following example illustrates how the method is used and its reliability. Model 1 is used to determine the sex of an individual from the test sample from the Coimbra Collection. The data for this individual are as follows: hipbone height is 204 mm, iliac breadth is 151 mm, SPRL is 66 mm, maximum femur head diameter is 42 mm, and epicondylar breadth of the femur is 75 mm. In this case, the femur head is more than one standard deviation smaller than the Coimbra male mean, and equivalent to the Terry female mean. The epicondylar breadth is more than one standard deviation smaller than the Coimbra male mean, less than 1 mm larger than the Terry female mean, and about one standard deviation larger than the Coimbra female mean. Furthermore, this individual has a ventral arc very similar to the pattern described by Sutherland and Suchey (43) in their Fig. 7.

Using Model 1,

$$P = \frac{1}{(1 + e^{-(-61.3545+0.5950(204)-0.5192(151)-1.1104(66)+1.1696(42)+0.5393(75))})}$$

$$P = 0.8147$$

Therefore, there is an 81.5% probability that the individual is male despite the small size and the presence of a ventral arc. The documented sex of the individual is male. There is no doubt that the documentary data are correct for this individual and there has been

TABLE 5—*Changes in calculated probability (P) with different intra-observer errors.**

Scenario†	SPRL (mm)	Calculated Probability (*P*)	Predicted Sex
Actual Data	66	0.8147	Male
SPRL + 0.6%	66.396	0.7390	Male
SPRL − 0.6%	65.604	0.8722	Male
SPRL + 2.7%	67.782	0.3780‡	Female
SPRL − 2.7%	64.218	0.9695	Male

*Model 1 is used to illustrate how *P* changes with various hypothetical errors for the SPRL.

†All independent variables are kept constant except for SPRL which is modified by the amount indicated.

‡An error equivalent to the mean error for the traditional pubis length is the difference between a correct and incorrect allocation.

no mixing of identity cards and skeletons. The cause of the death transcribed from hospital records is suicide by gunshot to the head. The cranium of this individual has a clear entrance wound in the palate, clear exit wound in the frontal, and damage to the eye orbits that is consistent with the peri-mortem effects of such a gunshot wound.

This case can also be used to illustrate the importance of having highly reproducible measurements of the pubis, particularly for borderline cases. Table 5 shows how the calculated probability changes depending on the level of intra-observer error. The magnitude and the direction of the error are both critical factors that must be considered. Row one shows the actual data for this individual. The SPRL is 66 mm and all the other measurements—which are not shown in the table—are used to calculate *P* and determine the sex of the individual. In row two, all the data are the same except that the measurement error of 0.6% is added to the SPRL to simulate a positive error in measurement. In row 3, again all things are equal except that a negative measurement error of 0.6% is subtracted from the SPRL to simulate a negative error in measurement. In row 4, the mean error found in the traditional pubis length is added to the SPRL to simulate a larger positive measurement error scenario. In row 5, the mean error found in the traditional pubis length is subtracted from the SPRL to simulate a larger negative measurement error. The critical scenario appears in row 4. In this scenario, there is a hypothetical positive measurement error in the SPRL that is equivalent to the mean error of the traditional pubis length. The individual that is definitely male is actually classified as female. A positive measurement error of greater than 2% is the difference between a correct and incorrect allocation. This problem is avoided with the SPRL because in almost 95% of cases, measurement error for the SPRL is less than 2% (see Fig. 3).

Discussion

Previous research has shown that the applicability of sex determination methods can be restricted by the reference sample used to develop the method (13,15–17). Regardless of the statistical approach, single measurement sex determination methods can be particularly susceptible to these problems because the methods are dependent on absolute size differences in means of males and females for any given measurement. However, there is evidence that some non-metric pelvic methods may also have similar constraints related to the limits imposed by the reference sample. The high allocation accuracy (96%) for the Phenice method (41) for a sample from the Terry Collection was not duplicated by MacLaughlin and Bruce (15) on three separate samples from England (82%), the Netherlands (68%), and Scotland (59%), by Lovell (42) on a sample of medical school cadavers from British Columbia (83%), nor by Rogers and Saunders (54) on a 19th century cemetery sample from Ontario (88%). Similarly, using only the ventral arc, Sutherland and Suchey (43) had an overall accuracy of 96% for a large sample of individuals autopsied in the County of Los Angeles, while Rogers and Saunders (54) had an accuracy of 87% on the cemetery sample from Ontario. These are differences in allocation accuracies that are comparable to those seen when some single measurement sex determination methods are applied across diverse samples (see 13). Ubelaker and Volk (55) suggest that this difference in accuracy when using the Phenice method may be due to the experience of the investigator; however, their methodology makes it difficult to separate the effects of experience from other issues. Ubelaker and Volk's results follow the pattern describe by Rogers and Saunders (54) and the results from this current study: accuracy when using only data from the pubis and ischium (Phenice method or SPRL and AIL) is between 85% and 90%, but accuracy for a combination of data from the pubis and other parts of the pelvis is over 95%. Lovell (42) did not find any significant differences in allocation accuracy based on experience. MacLaughlin and Bruce (15) did find that the difference in allocation between experienced and inexperienced was significant but the allocation accuracy for experienced investigators was still only 78.9%.

Despite some of the applicability problems reported for some metric methods, the benefits of metric methods are usually the ease of measurement as opposed to scoring presence, absence, or pronouncement of a trait, and the ease of statistical analysis particularly when large samples are involved (see also 38). In actuality, metric and non-metric approaches are often different ways of assessing the same variation. Two independent studies (40,56) have shown that differential growth in females at the symphyseal end of the pubis is responsible for the presence of a ventral arc in most females and its absence in most males. Measurement of the pubis or the scoring of the ventral arc should assess the same sexual dimorphism in the pubis, and it is recommended that the traditional pubis length and the ventral arc should not be considered two independent sources of information for assessing sex (56). The example above used to demonstrate the logistic model suggests that there are some exceptions. Discrepancies between methods may be due to a combination of factors including: 1) measurement error for metric approaches (a good example is the traditional pubis length in this current study); 2) lack of well defined standards in scoring non-metric traits (15); and 3) idiosyncracies of individuals in specific cases who have a combination of what are considered typical male or typical female traits (such as having a relatively short pubis and a ventral arc).

The very high allocation accuracies for the model sample, the test sample, and supplementary testing on forensic data from such different collections as the Terry Collection, the Coimbra Collection, and the FDB suggest that the metric methods presented here can be applied in a wide range of cases. The high accuracy of the methods, particularly Model 1, can be attributed to several factors. First, the very low intra-observer error associated with the SPRL measurement is critical in determining sex, particulary with borderline cases when other methods may fail.

Second, a wide range of human variation was sampled and included in the reference sample. This non-random approach to sampling: 1) considered the historical context of the collections in order to minimize bias and maximize representativeness, 2) considered other sources of variation such as year of birth and age at death even if these variables were not used as predictor variables,

3) included data from two very different collections, and 4) did not divide the reference sample into racial categories.

Third, the combination of long bone data *and* pelvic data contributes a great deal to the accuracy of the method. It is recommended in many sources (for example, 57) that all available data be used from the entire skeleton when attempting to determine sex. However, because of the emphasis placed on the pelvis when determining sex, it is implied that if the pelvis is complete, other data from the rest of the skeleton, with the exception of the cranium, does not have to be considered since the pelvis is the most sexually dimorphic part of the skeleton. Others bluntly state, "in many cases, particularly those involving the cranium and pelvis, qualitative morphological observations are sufficient for accurate sex attribution," (34, pp. 421). Qualitative morphological data from only the pubis would have resulted in an incorrect assessment of sex in the example from the Coimbra Collection presented in this paper. Data from long bones only tend to be considered if the pelvis is not recovered or is too damaged for assessment. Many published descriptions of sex determination methods that use bones other than the pelvis or cranium (for example, 34,35,58–60) begin by stating that there is a need for "other" methods for determining sex in cases where both the cranium and the pelvis are not recovered or are too damaged for analysis. The non-pelvic/non-cranial methods are never tested in conjunction with pelvic and/or cranial sex determination methods and it is never recommended that the method be used in conjunction with other data. Even Schulter-Ellis and colleagues who clearly describe the value of including the femur when determining sex, defer to the femur head measurement only when they suspected that their hipbone method was not correct. This current study shows that it makes a great deal of biological sense to draw information from both pelvic measurements and femur measurements when determining sex.

The femur in general, and the femur joints in particular, are highly sexually dimorphic, and can be used to estimate the overall size of a person. The size of the pubis *relative* to the hipbone and the femur is a highly effective way of maximizing accuracy when determining sex. Furthermore, by including the maximum femur head diameter and the SPRL, it is possible to assess the relative length of the symphyseal end of the pubis—where there is differential growth between sexes—with the acetabular end of the pubis, which is approximated by the maximum femur head diameter. Paradoxically, leaving out the acetabular portion of the pubis when using the SPRL measurement results in more information, provided that the size of the acetabulum is considered in some way. The benefits of including femur data whenever it is available is reflected in the increase in allocation accuracy by about 2% from Model 4 (hipbone only) to Model 1 (hipbone and femur) and by the total elimination of the difference in allocation accuracy between males and females. See Table 3 for details. Even when the pubis is damaged and the SPRL measurement cannot be collected, allocation accuracy is still very high for Model 22 and Model 26 because data from both the hipbone and the femur are used to determine sex.

Conclusion

This study demonstrates that allocation accuracy of better than 98% is possible for very different samples from Europe and North America from the 19th and 20th centuries using a metric sex determination method. This high accuracy for such diverse samples can be attributed to several factors. First, the superior pubis ramus length measurement, a new, highly reproducible alternative of the traditional pubis length measurement is used. Second, historical and demographic information—including year of birth, and age at death—was used to construct reference samples that included a substantial amount of normal human skeletal variation. Third, data from the pelvis and a long bone are considered together to determine sex. Fourth, race-specific methods were not developed.

Acknowledgments

I would like to thank the administration at the Museum of Anthropology at the University of Coimbra in Portugal for access to the Coimbra Collection and the curators of the FDB at the University of Tennessee, Knoxville for data. Thanks to Dr. Shelley Saunders for assistance throughout the research, Hugo Cardoso for the illustration in Fig. 2, and Sandra Albanese for her assistance with database management and editorial suggestions.

References

1. Stewart TD. Essentials of forensic anthropology. Springfield: CC Thomas, 1979.
2. Olivier G. Practical anthropology. Springfield: CC Thomas, 1969.
3. Steele D, Bramblett CA. The anatomy and biology of the human skeleton. College Station: Texas A&M University Press, 1988.
4. Schultz AH. The skeleton of the trunk and limbs of higher primates. Am J Phys Anthropol 1930;2:303–438.
5. Washburn SL. Sex differences in the pubic bone. Am J Phys Anthropol 1948;6:199–207.
6. Moore-Jansen PM, Ousley SD, Jantz RL. Data collection procedures for forensic skeletal material. Third edition. Knoxville: University of Tennessee Department of Anatomy, 1994.
7. Bass WM. Human osteology: a laboratory and field manual. Third edition. Columbia: Missouri Archaeological Society, 1987.
8. Buikstra JE, Ubelaker DH. Standards for data collection from human skeletal remains. New York: Wiley-Liss, 1994.
9. Thieme FP. Sex in the Negro skeleton. J Forensic Med 1957;4:72–81.
10. Schulter-Ellis FP, Schmidt DJ, Hayek LA, Craig J. Determination of sex with a discriminant analysis of new pelvic bone measurements: Part I. J Forensic Sci 1983;28:169–80.
11. Schulter-Ellis FP, Hayek LA, Schmidt DJ. Determination of sex with a discriminant analysis of new pelvic bone measurements: Part II. J Forensic Sci 1985;30:178–85.
12. Bruzek J. Fiabilité des fonctions discriminantes dans la détermination sexuelle de l'os coxal critiques et propositions. Bull et Mém de la Société d'Anthropologie de Paris 1992;4:67–104.
13. Saunders SR, Hoppa RD. Sex allocation from long bone measurements using logistic regression. Can Soc Forens Sci J 1997;30:49–60.
14. Meindl RS, Lovejoy CO, Mensforth RP, Carlos LD. Accuracy and direction of error in the sexing of the skeleton: implications for paleodemography. Am J Phys Anthropol 1985;68:79–85.
15. MacLaughlin SM, Bruce MF. The accuracy of sex identification in European skeletal remains using the Phenice characters. J Forensic Sci 1990;35:1384–92.
16. Albanese J. A comparison of the Terry Collection and the Grant Collection using the head of the femur and head of the humerus: implications for determining sex. Can Soc Forens Sci J 1997;30:167.
17. Albanese J. Skeletal variability in recent North Americans: implication for the development of new sex determination methods. Paper presented at the 25th Annual Meeting of the Canadian Association for Physical Anthropology. London, Ontario, 1997.
18. Giles E. Sex determination by discriminant function analysis of the mandible. Am J Phys Anthropol 1964;22:129–36.
19. Ousley SD, Jantz RL. The forensic data bank: documenting skeletal trends in the United States. In: Reichs K, ed. Forensic anthropology. Second edition: Springfield: CC Thomas, 1998;441–58
20. Ericksen MF. How "representative" is the Terry Collection? Evidence from the proximal femur. Am J Phys Anthropol 1982;59:345–50.
21. Albanese J. The use of skeletal data for the study of secular change: methodological implications of combining data from different sources. Am J Phys Anthropol Suppl 2002;34:36.
22. Albanese J. Identified skeletal collections and the study of human variation [dissertation]. Hamilton, On: McMaster University, 2003.
23. Rocha MA. Les collections ostéologiques humaines identifiées du Musée Anthropologigue de l'Université de Coimbra. Anthropol Port 1995;13:7–38.
24. Cunha E. Testing identification records: evidence from the Coimbra identified skeletal collections (Nineteenth and Twentieth centuries). In:

Saunders SR, Herring A, editors. Grave reflections: portraying the past through cemetery studies. Toronto: Canadian Scholars Press, 1995; 179–98
25. Santos AL. A Skeletal picture of tuberculosis [dissertation]. Coimbra, Portugal: University of Coimbra, 2000.
26. Bruzek J. A method for visual determination of sex, using the human hipbone. Am J Phys Anthropol 2002;117:157–68.
27. Santos AL, Roberts CA. A picture of tuberculosis in young Portuguese people in the early 20th century: a multidisciplinary study of the skeletal evidence. Am J Phys Anthropol 2001;115:38–49.
28. Ubelaker DH. Methodological considerations in the forensic applications of human skeletal biology. In: Katzenberg MA, Saunders SR, editors. Biological anthropology and the human skeleton. New York: Wiley-Liss, 2000;41–67
29. Iscan MY. A comparison of techniques on the determination of race, sex and stature from the Terry and Hamann-Todd Collections. In: Gill GW, Rhine S, editors. Skeletal attribution of race. Albuquerque: Maxwell Museum of Anthropology, 1990;73–81
30. Angel L. Colonial to modern skeletal change in the U.S.A. Am J Phys Anthropol 1976;45:723–36.
31. Terry RJ. On measuring and photographing the cadaver. Am J Phys Anthropol 1940;26:433–47.
32. Trotter M. Robert James Terry, 1871–1966. Am J Phys Anthropol 1981;56:503–8.
33. Iscan MY, Miller-Shavitz P. Discriminant function sexing of the tibia. J Forensic Sci 1984;29:1087–93.
34. Holman DJ, Bennett KA. Determination of sex from arm bones. Am J Phys Anthropol 1991;84:421–6.
35. Marino EA. Sex estimation using the first cervical vertebra. Am J Phys Anthropol 1995;97:127–33.
36. Ousley SD, Jantz RL. FORDISC 2.0: personal computer forensic discriminant functions. Knoxville: The University of Tennessee, 1996.
37. Smith SL. Attribution of foot bones to sex and population groups. J Forensic Sci 1997;42:186–95.
38. Giles E. Discriminant function sexing of the human skeleton. In: Stewart TD, ed. Personal identification in mass disasters. Washington, DC: National Museum of Natural History, 1970;99–109
39. Walker PL. Problems of preservation and sexism in sexing: some lessons from historical collections for palaeodemographers. In: Saunders SR, Herring A, editors. Grave reflections: portraying the past through cemetery studies. Toronto: Canadian Scholars press, 1995;31–47.
40. Anderson BE. Ventral arc of the os pubis: anatomical and developmental considerations. Am J Phys Anthropol 1990;83:449–58.
41. Phenice TW. A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 1969;30:297–301.
42. Lovell NC. Test of Phenice's technique for determining sex from the os pubis. Am J Phys Anthropol 1989;79:117–20.
43. Sutherland LD, Suchey JM. Use of the ventral arc in sex determination. J Forensic Sci 1991;36:501–11.
44. Tague RG. Maternal mortality or prolonged growth: age at death and pelvic size in three prehistoric Amerindian populations. Am J Phys Anthropol 1994;95:27–40.
45. Fuller K. Adult females and pubic bone growth. Am J Phys Anthropol 1998;106:323–8.
46. Norusis MJ. SPSS advanced statistics student guide. Chicago: SPSS Inc, 1990.
47. Iscan MY, Yoshimo M, Kato S. Sex determination from the tibia: standards for contemporary Japan. J Forensic Sci 1994;39:785–92.
48. Kieser JA, Moggi-Cecchi J, Groeneveld HT. Sex allocation of skeletal material by analysis of the proximal tibia. Forensic Sci Int 1992;56: 29–36.
49. Dudar JC, Dupras TL. Sex estimation from the adult human rib: what to use? Discriminant function or logistic regression? Am J Phys Anthropol Suppl 1996;22:99–100.
50. Albanese J. A new metric method for sex determination using the pelvis and femur. Am J Phys Anthropol Suppl 2001;32:31.
51. Albanese J. Measurement error and metric methods for sex determination. Paper presented at the 29th Annual Meeting of the Canadian Association for Physical Anthropology. Winnipeg, Manitoba, 2001.
52. Ferembach D, Schwidetzky I, Stoulkal M. Recommendations for age and sex diagnoses of skeletons. J Hum Evol 1980;9:517–94.
53. Dibennerdo R, Taylor JV. Classification and misclassification in sexing the black femur by discriminant function analysis. Am J Phys Anthropol 1982;58:145–51.
54. Rogers T, Saunders S. Accuracy of sex determination using morphological traits of the human pelvis. J Forensic Sci 1994;39:1047–56.
55. Ubelaker DH, G VC. A test of the Phenice method for the estimation of sex. J Forensic Sci 2002;47:19–24.
56. Budinoff LC, Tague RG. Anatomical and developmental bases for the ventral arc of the human pubis. Am J Phys Anthropol 1990;82:73–9.
57. Krogman WM, Iscan MY. The human skeleton in forensic medicine. Second edition. Springfield: CC Thomas, 1986.
58. Black K. A new method for assessing the sex of fragmentary skeletal remains: femoral shaft circumference. Am J Phys Anthropol 1978;48: 227–32.
59. France DL. Sexual dimorphism in the human humerus [dissertation]. Boulder: University of Colorado, 1983.
60. Berrizbeitia EL. Sex determination with the head of the radius. J Forensic Sci 1989;34:1206–13.

Additional information and reprint requests:
John Albanese, Lecturer
Department of Anthropology
McMaster University
Hamilton, Ontario
L8S 4L9
Canada
E-mail: albaneg@mcmaster.ca