J Forensic Sci, July 2006, Vol. 51, No. 4
doi:10.1111/j.1556-4029.2006.00177.x
Available online at: www.blackwell-synergy.com

Brenda A. Williams,[1] *B.A., M.Sc. and Tracy L. Rogers,*[2] *Ph.D.*

Evaluating the Accuracy and Precision of Cranial Morphological Traits for Sex Determination

ABSTRACT: Sex determination is a key analysis that forensic anthropologists perform in order to construct a biological profile of human remains. The techniques used in forensic investigations must meet the *Mohan* or *Daubert* criteria, for admissibility in a court of law. In this study, the precision and accuracy of 21 morphological characteristics of the skull were tested on a modern sample of 50 adult crania of European White ancestry. The following craniofacial features are identified as high-quality traits, defined by intraobserver error $\leq 10\%$ and accuracy $\geq 80\%$: mastoid size, supraorbital ridge size, general size and architecture, rugosity of the zygomatic extension, size and shape of the nasal aperture, and gonial angle. Ninety-six percent accuracy and 92% precision were achieved using 20 traits in combination. Fisher's exact probability tests revealed no significant differences ($p = 0.05$) in the levels of precision or accuracy between age categories. Sex-related bias in accuracy was found for the following cranial features: ramus symphysis ($p = 0.009$), zygomatic extension ($p = 0.0016$), and occipital markings ($p = 0.0013$). These traits demonstrated a greater tendency to be scored male than female.

KEYWORDS: forensic science, sex determination, morphology, skull, precision, accuracy, expert witness testimony

Forensic anthropologists actively engaged in casework may be called upon to provide expert witness testimony for cases involving unknown human remains. In Canada and the United States, legislation dictates that techniques used by expert witnesses must meet the *Mohan* (1) and *Daubert* (2) criteria for admissibility (3,4). Sex determination of unknown human remains is only one of several types of expert analyses that a forensic anthropologist may be called upon to present at trial.

In a review of several widely used osteology and forensic anthropology texts, Rogers (5) demonstrates that there is variability in the number of morphological traits of the skull recommended for sex determination and in the degree of description used to explain each trait. Most texts do not provide a comprehensive and descriptive suite of traits, but rather favor a discrete number of characteristics for analysis (5). Rogers was unable to discern the criteria by which researchers made recommendations because of the lack of data regarding the accuracy and precision of the techniques and/or characteristics recommended. Rogers evaluated 17 commonly used traits of the skull to establish their individual and collective value as sex determinants. The traits were evaluated and ranked using a historic cemetery collection of known sex (5).

This analysis tests the accuracy and precision of the 17 traits evaluated by Rogers (5) on a modern collection. The goals are: (1) to identify high-quality morphological traits for sex determination of the skull with $\geq 80\%$ accuracy, $\leq 10\%$ intraobserver error, and no age- or sex-related biases; and (2) to compare the results obtained from the two subsets (historic vs. modern) of a broader population (individuals of European/White ancestry) in order to identify potential confounding factors that affect the reliability and subsequent outcome of the traits being evaluated. Less than or equal to 10% intraobserver error was selected as a standard measure of the diagnostic reliability of individual traits for this analysis on the basis of previously reported standards for intraobserver concordance (6–9). Greater than or equal to 80% accuracy was chosen as the minimum standard as it is the minimum accuracy that can be attained using morphometrics of the skull for sex determination. The reliability of metric traits reported for sex determination of the skull lies between 80% and 90% for individuals of European ancestry (10). Information attained from morphological techniques that do not meet the minimum accuracy achieved by metric techniques complicate, rather than contribute to the interpretation of the data.

This study has implications for establishing the credibility of forensic anthropologists providing expert testimony. If the accuracy and error rates of techniques used by experts are unknown, the witness' testimony will not be admissible (2–4). A forensic anthropologist must be able to demonstrate that the findings are appropriate, that the techniques used for sex determination are founded on sound principles and methodology, producing an assessment that is statistically greater than chance (2–4).

Regardless of the nature of the testimony, reports submitted by scientific experts become part of the court record. Any aspect of analysis performed by the expert may be called into question and subject to challenge during the qualification process or on cross-examination. If, for example, the evidence from other expert witness testimony (DNA) relating to sex determination conflicts with that found by the forensic anthropologist, it presents the opportunity for the credibility of the forensic anthropologist to be challenged. Accordingly, the forensic anthropologist may be called upon to explain why s(he) obtained inaccurate results relating to his/her finding of sex determination. Erroneous results in one aspect of the analyses may then be used to discredit other analyses conducted by the forensic anthropologist. It will be necessary for the forensic anthropologist to defend his/her finding using the theoretical principles, the accuracy, and the precision of the techniques used to make the assessment. If the techniques utilized by the forensic anthropologist are not based on reliable principles and methodology, and the accuracy and error rates are not known, the credibility of the expert and professional reputation could be damaged.

[1]Department of Anthropology, University of Toronto, Toronto, ON, Canada.
[2]Department of Anthropology and the Study of Religion, University of Toronto, Mississauga, ON, Canada.

Received 13 May 2004; and in revised form 10 Oct. 2004; 4 March 2006; accepted 26 March 2006; published 21 June 2006.

Materials and Methods

The study sample was selected from the William M. Bass Donated Skeletal Collection, curated in the Department of Anthropology at the University of Tennessee, Knoxville. The collection is comprised of approximately 400 complete and partial skeletons of Euro-American, Afro-American, and Hispanic ancestry. Individuals in the collection were born from the turn of the 20th century to the present. Fifty complete adult crania ($n = 50$), representing individuals of European ancestry of known age and sex, are the basis of this analysis. An equal proportion of females ($n = 25$) and males ($n = 25$) were selected. The female age distribution is 32–94 years, with a mean age of 62.6 years (Table 1). Males have an age distribution of 25–85 years, with a mean age of 51.9 years (Table 1). Known age was not available for two individuals ($n = 2$). Age estimation based on evaluation of the ectocranial sutures and the dentition indicates that the specimens were adult and therefore do not require exclusion from the study sample. A *post hoc* power analysis ($\alpha = 0.05$, $\beta = 0.2$) indicates that the study sample size is sufficient to support the results (accuracy) obtained in this analysis.

Rogers' (5) suite of 17 morphological characteristics of the skull were used to determine the sex of the specimens in this analysis with minor modifications. The features of the mandible were separated into four traits in order to evaluate each individual component and its contribution as a sex indicator. Although Rogers (5) did not include orbital margins in her analyses, the authors have included them in the current study because of common usage in determining sex (11–17).

An assistant (T. R.) selected the sample randomly to ensure that the analyses were conducted blind and to minimize potential bias of the evaluations conducted by the observer (B. W.). To facilitate a comparison of the results from the current study with those obtained by Rogers (5), the author established the following parameters *a priori* for the study sample: (1) that all specimens were of European White ancestry, (2) that a similar age and sex distribution of specimens were represented in the sample, and (3) the individuals had lived within the past 200 years. The assistant ensured that all potentially distinguishing marks (e.g., catalogue numbers) on the crania were hidden from the observer (B. W.). Specimens were presented and examined in no particular order. No preliminary analysis was undertaken by the author to determine the total range of variation within the sample before assessment, because such comparisons are not possible in single forensic cases.

During the first round of analysis, the crania were examined and scored individually over a period of 2 days. Following the initial round of observation, an intraobserver test was performed in which the entire sample ($n = 50$) was re-examined over two additional days. The order in which the crania were examined by B. W. for the intraobserver test differed from the initial round of observation to prevent possible memorization of scoring.

TABLE 1—*Distribution of study sample by age and sex (n = 50).*

Age Category (Years)	Sex	*n*	Age Category (Years)	Sex	*n*
20–29	M	3	20–29	F	0
30–39	M	4	30–39	F	2
40–49	M	4	40–49	F	2
50–59	M	6	50–59	F	5
60–69	M	3	60–69	F	8
70–79	M	2	70–79	F	3
80–89	M	2	80–89	F	3
90–99	M	0	90–99	F	1
Unknown	M	1	Unknown	F	1

Crania were observed and scored (Male, Female, and Indeterminate) based on the presence/absence of features. Traits were assessed using the data sheet provided (5) (Fig. 1). Features were observed in the order that they appear on the data sheet. Features that were not available for assessment due to ante—or perimortem trauma, or taphonomic damage were scored as N/A. Each criterion was given equal weighting and sex was assigned according to the category (Male, Female, or Indeterminate) in which the majority of features fell.

An intraobserver test ($n = 50$) was conducted to determine the precision of each trait and the overall sex determination. Precision was evaluated in terms of the percentage of cases in which the two separate rounds of assessment conflicted (5). Rank was assigned to individual traits based on the lowest percentage of intraobserver error. All traits with $>10\%$ intraobserver error were eliminated from further analyses (Table 2). Owing to time constraints, an interobserver test was not possible.

The accuracy of each characteristic was determined by comparing the blind assessments of a trait with the known sex. The percentage accuracy of the individual traits is presented in Table 3. Minimum accuracy (bold values) is reported for individual traits regardless of the round of examination. The most conservative accuracy values are reported because they provide the researcher with the potential minimum degree of success of obtaining a correct sex assignment utilizing individual characteristics.

The results of the accuracy and precision analyses were used to determine the overall value of individual traits. For each trait, the sum of the lowest percentage of intraobserver error (rank) was added to the highest percentage of accuracy (rank) to obtain a combined score from which traits were ranked and value was assigned (5).

Traits that met the critical values set for precision and accuracy were then tested for age- and sex-related bias by means of Fisher's exact probability test (two tailed) with a significance level of $\alpha = 0.05$. The overall accuracy for the final list of traits was determined by re-evaluating the ability of the traits to identify the sex of the sample correctly.

Results

The test of precision, in which the entire sample ($n = 50$) was re-examined, revealed four disagreements between the first and second assessments, resulting in an 8% intraobserver error for combined sexes. The precision of individual traits ranges from 2% to 20% (Table 2). Six of the 21 morphological traits of the cranium tested in this study did not meet the critical value of $\leq 10\%$ acceptable error (6,7) and were, therefore, excluded from further analyses. Each of the following traits scored extremely high levels of intraobserver error: orbit shape and position (12%), size of the occipital condyles (12.5%), forehead (16%), malar size and rugosity (18%), mandibular ramus breadth (18%), and parietal eminences (20%), far exceeding the acceptable level of precision set for this study. The remaining 14 morphological features scored at or below the critical value, producing an overall level of intraobserver error of 7.4%. These traits form the basis of all subsequent analyses and are listed in Table 3.

Once precision was established, it was necessary to determine the accuracy with which the remaining 14 traits could correctly predict the sex of the specimens. Table 3 provides the accuracy for each of the 14 morphological features of the skull for males, females, and combined sexes for each round of examination. Bold values denote the most conservative accuracy value obtained

ADULT VISUAL SEX DETERMINATION (CRANIUM)
RECORDING FORM

UNIVERSITY OF TORONTO MISSISSAUGA
FORENSIC ANTHROPOLOGY LABORATORY Date: ____________

Catalogue: ____________ Collection: ____________

Location: ____________

Observer(s): ____________ Time: _____/_____

FEATURE	MALE ♂		FEMALE ♀	
Size & architecture:	big/rugged	_____	small/smooth	_____
Forehead:	low, slopes post.	_____	rounded, full, vertical	_____
Frontal eminences:	small	_____	large	_____
Supraorbital ridges:	medium to lge	_____	sml to medium	_____
Orbits:	squared, low	_____	rounded, high	_____
	rnded margins	_____	sharp margins	_____
Nasal aperture:	high, thin sharp margins	_____	lower, wider rounded margins	_____
Nasals:	large	_____	small	_____
Malars:	posterior-lateral	_____	anterior-lateral	_____
Zygomatic:	extends	_____	does not	_____
Parietal eminences:	small	_____	large	_____
Mastoid:	medium-lge	_____	sml-medium	_____
Occipital:	well-muscled	_____	not marked	_____
Occipital condyles:	large	_____	small	_____
Palate:	lge, u-shape	_____	sml, parabolic	_____
Tooth size:	lgr, M1 5cusp	_____	smaller	_____
Mandible:	lge, high symp	_____	sml, low symph.	_____
	brd asc. ramus	_____	small asc. ramus	_____
	gon. angle <125	_____	gonial angle >125	_____
	gon. agl flares	_____	angle does not flare	_____
Chin:	square 2-point	_____	rounded 1-point	_____

COMMENTS: ____________

ESTIMATED SEX: ____________

after Rogers (5)

FIG. 1—*Adult visual sex determination (cranium).*

regardless of round. All further analyses report the minimum accuracy because it provides the most conservative assessment of accuracy that can be expected from individual traits. The minimum accuracy with which each trait could predict sex ranges from 60% to 100% for males, from 24% to 96% for females and from 58% to 92% for combined sexes based on the results from both rounds of analysis. By utilizing all 14 traits in combination, 92.0% overall accuracy and 92.6% precision were achieved for combined sexes.

TABLE 2—*Precision of individual traits of the skull (combined sexes).*

Trait	n	Error #	Error %	Rank
Occipital markings	50	1	2.0	1
Chin form	50	2	4.0	2
Size of supraorbital ridge	50	3	6.0	3
Zygomatic extension	50	3	6.0	3
Mandible—gonial angle	50	3	6.0	3
Mandible—symphysis height	41	3	7.3	3
Frontal eminences	50	4	8.0	4
Size of mastoid	50	4	8.0	4
Mandible—gonial eversion	50	4	8.0	4
Size of nasals	46	4	8.7	5
Size and architecture	50	5	10.0	6
Orbits—margins	50	5	10.0	6
Nasal aperture	50	5	10.0	6
Palate	50	5	10.0	6
Orbits—shape, position	50	6	12.0	7
Size of occipital condyles	48	6	12.5	8
Forehead	50	8	16.0	9
Size and rugosity of malars	50	9	18.0	10
Mandible—ramus breadth	50	9	18.0	10
Parietal eminences	50	10	20.0	11

Rank determined by lowest percentage of intraobserver error.

A marginal increase in average accuracy was observed for each sex when comparing the separate rounds of examination: 1.1% (males) and 1.3% (females). Collectively, the traits were not as accurate for females as males (Table 3). No difference in the proportion of low-scoring to high-scoring traits (7:7) was revealed for males between the separate rounds of examination. This pattern was not exhibited in the female sample ($n = 25$). A comparison of accuracy for individual traits reveals a reversal in the proportion of low-scoring to high-scoring traits from 5:9 to 9:5 in the second round, indicating a decrease in accuracy for females in the second round of evaluation. One trait (mastoid) did not change accuracy value between trials for combined sexes. An overall accuracy of 96% was achieved by using the entire trait list for combined sexes.

Fisher's exact probability tests were conducted for both males and females to evaluate whether age and/or sex were confounding factors affecting the reliability of the results achieved by the observer. Age categories were established following Rogers (5). Rogers (5) used three age categories: <25, 25–44 years, and 45+ years on the basis of male craniofacial growth and development. Owing to the age distribution of the study sample, only two age categories were applicable for the current study: 25–44 years ($n = 10$) and 45+ years ($n = 38$). The results indicate no age-related bias with respect to the levels of accuracy or precision for the age categories. No sex-related bias for the levels of intraobserver error between males and females was observed. Fisher's exact probability test results reveal sex-related bias regarding the levels of accuracy for each of the following traits: ramus height ($p = 0.009$), zygomatic extension ($p = 0.0016$), and occipital markings ($p = 0.0013$). These features are more likely to be scored male than female.

Discussion

The overall value of each trait was calculated by the sum of the precision rank (based on the lowest to the highest percentage of intraobserver error) and accuracy rank (based on the highest to the lowest percentage) to produce a combined scored. Features were ranked again on the basis of the combined score, the lowest combined score earning the highest overall value (Table 4). It should be noted that a low combined score does not necessarily directly reflect combined high accuracy and precision. The very nature of a combined score permits one value to be of greater than, equal to, or lesser than the other. As a result, traits may score high precision but report low accuracy. This indicates that the trait can be scored consistently but fails to perform well with respect to obtaining a correct sex assignment. Conversely, a trait may score high accuracy but is difficult to score consistently between trials. The most reliable traits then are those that obtain a low combined score, by demonstrating high accuracy and precision.

In cases where there is disparity between the accuracy and precision ranks used to calculate the combined score, weight should be given to precision over accuracy. Precision is a measure of the observer's ability to reproduce his/her results and is both a reflection on the researcher's capability as well as the features being examined (18). Whether the researcher or the criteria are at fault, an observer may be able to assess correctly the sex of a particular skeletal sample, but when required to repeat those results, has a greater than one in 10 chance of making an incorrect assignment. This has particular significance in forensic applications where sample size is comprised of a single individual and observations are limited owing to time constraints (5).

The results reported in this study indicate that it is possible to achieve high accuracy and precision using a discrete number of morphological features of the skull to determine the sex of unknown skeletal remains. Because the aim of this study is to

TABLE 3—*Accuracy results for individual traits with $\leq$ 10% intraobserver error.*

Trait	n	M+ F (%)		n	Males (%)		n	Females (%)	
		1	2		1	2		1	2
Size and architecture	50	92.0	**88.0**	25	**80.0**	88.0	25	92.0	**88.0**
Frontal eminences	50	66.0	**64.0**	25	88.0	**76.0**	25	**24.0**	52.0
Size of supraorbital ridge	50	**86.0**	90.0	25	**84.0**	92.0	25	**88.0**	**88.0**
Orbits—margins	50	**76.0**	78.0	25	76.0	**72.0**	25	76.0	84.0
Nasal aperture	50	**84.0**	86.0	25	**80.0**	92.0	25	88.0	**80.0**
Size of nasals	46	72.0	**68.0**	23	**71.0**	79.2	23	74.0	**65.0**
Zygomatic extension	50	84.0	**82.0**	25	**100.0**	**100.0**	25	68.0	**64.0**
Size of mastoid	50	**92.0**	**92.0**	25	92.9	**88.0**	25	**92.0**	96.0
Occipital markings	50	**58.0**	60.0	25	**80.0**	88.0	25	36.0	**32.0**
Size and shape of palate	50	**74.0**	**74.0**	25	**76.0**	92.0	25	72.0	**56.0**
Mandible—symphysis height	41	68.0	**58.0**	21	**91.0**	95.0	20	**42.0**	53.0
Mandible—gonial angle	50	**80.0**	86.0	25	92.0	**84.0**	25	**68.0**	88.0
Mandible—gonial eversion	50	**58.0**	72.0	25	64.0	**60.0**	25	52.0	**44.0**
Chin form	50	78.0	**72.0**	25	88.0	**72.0**	25	**72.0**	**72.0**

Bold denotes minimum accuracy achieved regardless of round.

TABLE 4—*Comparison of ranking results of accuracy, precision, and combined rank with the results obtained by Rogers (5).*

Trait	Williams and Rogers					Rogers (5)				
	n	*A*	*P*	Total (*A*+*P*)	Combined Rank	*n*	*A*	*P*	Total (*A*+*P*)	Combined Rank
Size of supraorbital ridge	50	2	**3**	5	1	49	4	2	6	1
Size of mastoid	50	1	4	5	1	49	9	1	10	3
Size and architecture	50	3	6	9	2	49	12	1	13	5
Zygomatic extension	50	6	3	9	2	49	2	4	6	1
Nasal aperture—size, shape	50	4	6	10	3	49	1	5	6	1
Mandible—symphysis height	38	9	4	13	4	49	8	3	11	4
Chin form	50	11	2	13	4	49	5	3	8	2
Orbits—margins	50	8	6	14	5	49	—	—	—	—
Occipital markings	50	13	1	14	5	49	6	2	8	2
Mandible—ramus breadth	50	4	10	14	5	49	8	3	11	4
Mandible—gonial angle	50	11	3	14	5	49	8	3	11	4
Size of nasals	46	10	5	15	6	49	7	4	11	4
Size and rugosity of malars	50	5	10	15	6	49	3	3	6	1
Forehead	50	7	9	16	7	49	10	3	13	5
Frontal eminences	50	12	4	16	7	49	14	3	17	7
Orbits—shape, position	50	9	7	16	7	49	11	6	17	7
Palate—shape, size	50	10	6	16	7	49	13	3	16	6
Size of occipital condyles	47	10	8	18	8	49	16	1	17	7
Mandible—gonial eversion	50	14	4	18	8	49	8	3	11	4
Parietal eminences	50	10	11	21	9	49	15	2	17	7

identify high-quality traits, features that do not meet the critical values for acceptable levels of precision ($\leq$ 10%) and accuracy ($\geq$ 80%) were excluded from further analyses. Of the 21 original characteristics that were selected for analysis, 15 craniofacial features do not meet the critical values necessary to ensure the reliability of these traits.

Fourteen features of the cranium produced levels of error that were within the acceptable range established for precision in this study (Table 2). The overall precision of these traits was 7.4%, well below the critical level of 10% error. An examination of intraobserver error by trait identified four borderline traits. General size and architecture, orbital margins, and the size and shape of the palate and nasal aperture each yielded the maximum level ($\leq$ 10%) of acceptable error, achieving 90% precision. The importance of these four traits individually is of debate; however, in combination, the impact of these criteria could significantly influence the outcome of sex assignment in the case of fragmentary remains with limited criteria available for observation. Traits that exceed 10% intraobserver error should be considered unreliable (6–8), and for this reason it is recommended that they be excluded from analysis.

Of the four borderline criteria, Rogers (5) did not assess orbital margins. In Rogers' (5) evaluation of individual morphological features of the skull, only general size and architecture exhibited a high level of precision (Table 4). Nasal aperture did not perform well in Rogers' (5) study, achieving the second lowest precision rank score of all the traits evaluated. Palate size and shape scored in the top three criteria for precision. The results obtained by Rogers (5) suggest that overall size and architecture, and palate size and shape are useful criteria for sex determination and that difficulty in assessment of the traits in the current research may be attributed to the experience of the observer.

The advantage of utilizing sample populations of known age and sex to evaluate techniques and/or methodology is the ability of the researcher to determine its accuracy and thus provide commentary on its applicability and utility. Accuracy is a measure of the number of correct assessments, and is typically reported in terms of percentages (18). Accuracy is one of several essential criteria considered by forensic experts when selecting a technique or methodology to use in analyses (19). There are several potential measures of accuracy that may be reported in an analysis of this type: highest, average, or minimum accuracy. The minimum measure of accuracy of individual traits or in traits in combination is the most valuable measure for expert analyses. Minimum accuracy values provide a more realistic measure of the ability of traits to determine sex in cases of unknown sex. Reporting the highest accuracy results achieved for individual features simply reports the results achieved in the best-case scenario.

The overall accuracy of the 14 traits considered in this analysis is 92.0%. This value exceeds the acceptable level of $\geq$ 80% accuracy established for this analysis. An examination of the individual traits for accuracy identified six traits that did not perform well in this study: orbital margins (76%), palate size and shape (74%), chin form (72%), size of nasals (68%), frontal eminences (64%), and gonial eversion of the mandible (58%). Four of these traits report accuracy values that fall significantly below the critical level of acceptable error and in some cases, report results that are only slightly better than chance. The rank score results for accuracy reported by Rogers (5) confirm the poor results obtained for frontal eminences and palate size and shape. In contrast to the current study, chin form scored in the top five for accuracy (rank scores) for Rogers (5).

Because of slight modifications to the original suite of traits examined by Rogers (5) for this study, a direct comparison of the accuracy and precision results (%) of the individual mandibular features for each is not possible. The overall value of individual traits, determined by the sum of both the accuracy and precision rank scores, however, enables the observer to provide a general commentary of the performance of the features of the mandible as a whole. In Rogers' (5) evaluation, the mandible scored high precision but achieved moderate accuracy. In this study, three of the four mandibular criteria evaluated, excluding ramus breadth, scored high precision and low accuracy. Gonial angle was the only single mandibular feature of the four evaluated to perform well for accuracy and precision in this analysis, scoring 80.0% and 86.0%, respectively. Maat et al. (20) likewise report poor performance of the morphology of the mandible as a sex indicator, particularly when evaluated as an isolated element. These results confirm the poor reliability of individual features of the mandible in this study for sex assessment of unknown skeletal remains.

Owing to the difficulty in visually assessing the degree of sexual dimorphism of dentition, and the consequent potential for increased error when making a sex assessment on a single, isolated mandible (20), it is recommended that both metric and morphological assessments be undertaken where possible. These recommendations should also be applied to cases that wholly consist of fragmentary or incomplete mandibles that possess limited criteria for analysis.

The results reported in this study indicate that it is possible to achieve high accuracy and precision using a discrete number of morphological features of the skull to determine the sex of unknown skeletal remains. Because the aim of this study is to identify high-quality traits, only features that meet the established critical values for acceptable levels of precision ($\leq$ 10%) and accuracy ($\geq$ 80%) for this study are recommended. Of the 21 original characteristics that were selected for analysis, 15 features do not meet the critical values necessary to ensure the reliability of these traits.

Six features achieved high precision and accuracy as defined by this study (Table 5). Of these six criteria, three are commonly recommended for sex determination of the skull (supraorbital ridges, mastoid, and overall size and architecture (11–17)). Two traits, zygomatic extension and nasal aperture, are not well recognized in the literature.

Five of the six criteria identified as high-quality traits in this analysis (mastoid, supraorbital ridge, size and architecture, zygomatic extension, and nasal aperture) performed similarly in Rogers' (5) examination. The results of the current study are strengthened by those reported by Rogers (5), and establish the effectiveness of these five traits as skeletal indicators of sex for crania of historic and modern populations of European descent. The differences in ranking of these traits between the two studies are minimal, with rank scores differing from one to three positions, but all scoring within the top five criteria. For traits that scored lower than one rank position compared with Rogers (5) in this study, the difference may be attributed to the decision of the observer not to assess the range of variation between the sexes before examination. The size of the supraorbital ridge is the only trait to receive the same combined rank score in each analysis.

Although an interobserver test was not conducted, by maintaining the same parameters as Rogers (5), it is possible to compare the results and provide commentary on the overall utility of the morphological features of the crania examined. The comparison is valid, as both study samples: (1) are of known age and sex; (2) they exhibit only minor differences with respect to time and geography (individuals were born within the last 200 years); and (3) are subsets of a single, broader population (European/White ancestry). The success of the six traits in this analysis indicates that further testing on Euro-American White populations is not necessary. Efforts should now be directed toward evaluating the accuracy and precision of these traits on new populations.

In conclusion, the results achieved by both Rogers (5) and the current study indicate that it is possible to achieve high accuracy and precision using a discrete number of morphological features of the skull for sex determination. In this study, the size of the mastoid, supraorbital ridge, general size and architecture, rugosity of the zygomatic extension, size and shape of the nasal aperture, and gonial angle each performs well, producing results that are $\geq$ 80% accuracy and $\geq$ 10% intraobserver error. It is necessary, however, to acknowledge the limitations in the ability of this analysis to detect differences in the levels of precision and accuracy for both age and sex owing to the small sizes used. Overall accuracy based on the six characteristics is 94%, 2% less than reported for the entire suite of traits. On the basis of these results, forensic anthropologists should consider the above-named traits of primary value when using cranial morphology to determine the sex of unknown skeletal remains. Several features performed poorly and should not be considered as single reliable sex indicators for cases in which sex is unknown: tooth size; parietal eminences; size of the occipital condyles; orbit shape and position; frontal eminences; palate size and shape; and slope of the forehead.

Further research is necessary to establish the levels of accuracy and precision of cranial characteristics as sex indicators for populations of different biogeographical origins (e.g., European White, American Black, and Hispanic). Intra- and interobserver tests should be performed to better demonstrate the reliability of and/or identify potential problems with the characteristics being tested on the particular population in question. Sample sizes need to be large enough to ensure that researchers may make meaningful interpretations based on the results of the statistical applications used.

The increasing reality of forensic anthropologists being called upon to provide expert witness testimony demonstrates the need for rigorous testing of the techniques being used and the reporting of results of their analyses.

At present, the paucity of reported data regarding levels of precision and accuracy makes it difficult to ascertain whether or not the techniques and methodology commonly used in expert analyses by forensic anthropologists meet the guidelines as required by *Daubert*, *Mohan*, and the *Federal Rules of Evidence*. The current study confirms the accuracy and reliability of the six traits (mastoid, supraorbital ridge, size and architecture, zygomatic extension, nasal aperture, and gonial angle) for sex determination and hopes to encourage researchers to undertake similar investigations in order to preclude potential challenges of admissibility of evidence as experienced in other areas of forensic specialization.

TABLE 5—*Individual traits of the cranium that achieved $\geq$ 80 % accuracy and $\leq$ 10% intraobserver error (combined sexes).*

Trait	n	Minimum Accuracy (%)	Intraobserver Error (%)	Combined Value ($A+P$)
Mastoid	50	92.0	8.0	1
Supraorbital ridge	50	86.0	6.0	1
Size and architecture	50	88.0	10.0	2
Zygomatic extension*	50	82.0	6.0	2
Nasal aperture	50	84.0	10.0	3
Mandible–gonial angle	50	80.0	6.0	5

*Sex-related bias ($p = 0.0016$) tendency to score as male.

Acknowledgments

The authors would like to thank Dr. Lee Meadows-Jantz, at the University of Tennessee Knoxville, for providing assistance and access to the William M. Bass Donated Skeletal Collection; Dr. Michael Schillaci, at the University of Toronto Scarborough, for his support and feedback regarding this analysis; and Dr. Micheal Brand, for his assistance with the data collection.

References

1. Regina vJ-LJ. [2000] S.C.R. File No. 26830.
2. United States Supreme Court in Daubert v. Merrell Dow Pharmaceuticals Inc., 509 U.S. 579 [1993].

3. Rogers T, Allard TT. Expert testimony and positive identification of human remains through cranial suture patterns. J Forensic Sci 2004;49(2): 203–7.
4. Christensen AM. The impact of Daubert: implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). J Forensic Sci 2004;49(3):1–4.
5. Rogers T. Sex determination and age estimation: skeletal evidence from St. Thomas' Cemetery Belleville, Ontario[M.A. thesis]. Hamilton, ON: McMaster University, 1991
6. Novotny V, Iscan MY, Loth SR. Morphologic and osteometric assessment of age, sex, and race from the skull. In: Iscan MY, Helmer RP, editors. Forensic analysis of the skull. New York: Wiley-Liss, 1993:71–88.
7. Nichol CR, Turner CG. Intra- and interobserver concordance in classifying dental morphology. Am J Phys Anthropol 1986;69:299–315.
8. Molto JE. The assessment and meaning of intraobserver error in population studies based on discontinuous cranial traits. Am J Phys Anthropol 1979;5:333–44.
9. Rogers T, Saunders S. Accuracy of sex determination using morphological traits of the human pelvis. J Forensic Sci 1994;39(4):1047–56.
10. Giles E. Discriminant function sexing of the human skeleton. In: Stewart TD, editor. Personal identification in mass disaster. Washington, DC: Museum of Natural History, 1970:99–107.
11. Bass WM. Human osteology: a laboratory and field manual. 4th ed. Columbia: Missouri Archaeological Society, 1995.
12. Burns KR. Forensic anthropology training manual. New Jersey, NJ: Prentice-Hall, 1999.
13. Byers S. Introduction to forensic anthropology: a textbook. Boston: Allyn & Bacon, 2002.
14. Schwartz JH. Skeleton keys: an introduction to human skeletal morphology, development, and analysis. New York: Oxford University Press, 1995.
15. White TD. Human osteology. 2nd ed. San Diego: Academic Press, 2000.
16. Krogman WM, Iscan MY. The human skeleton in forensic medicine. Springfield: Charles C. Thomas, 1986.
17. Graw M, Czarnetski A, Haffner H-T. 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 1999;108:91–6.
18. Sokal RR, Rohlf FJ. Biometry. 3rd ed. New York: Freeman and Company, 1995.
19. St. Hoyme LE, Iscan MY. Determination of sex and race: accuracy and assumptions. In: Iscan MY, Kennedy KAR, editors. Reconstruction of life from the skeleton. New York: Wiley-Liss, 1989:53–93.
20. Maat GJR, Mastwijk RW, Van der Velde EA. On the reliability of non-metrical morphological sex determination of the skull compared with that of the pelvis in the low countries. Int J Osteoarchaeol 1997;7: 575–80.

Additional information and reprint requests:
Brenda Williams, M.Sc.
Forensic Science Program
South Building Room 3030
3359 Mississauga Road North
University of Toronto at Mississauga
Mississauga, ON
Canada L5L 1C6
E-mail: brenda.williams@utoronto.ca