

Performance Evaluation as a Diagnostic Test for Traditional Methods for Forensic Identification of Sex

Evaluación del Rendimiento como una Prueba Diagnóstica de los Métodos Clásicos para la Determinación Forense del Sexo

****Iván Claudio Suazo Galdames; **Daniela Alejandra Zavando Matamala & **Ricardo Luiz Smith**

SUAZO, G. I. C; ZAVANDO, M. D. A & SMITH, R. L. Performance evaluation as a diagnostic test for traditional methods for forensic identification of sex. *Int. J. Morphol.*, 27(2):381-386, 2009.

SUMMARY: This study assesses the quality, as a diagnostic test, of the main indicators of morphological sexual dimorphism through direct anthropometry, biostatistics tools and clinical epidemiology. This study used 284 skulls of adult Brazilians, of which 187 were male and 97 female. A study of the cross-evaluation of the diagnostic test was performed; it was a qualitative approach based on visual examination of 16 traditional indicators of morphological sexual dimorphism, where each indicator determined the level of accuracy, sensitivity, predictive values, likelihood ratios, and odds ratio. All indicators studied had high levels of accuracy (84.75–72.89%). The best indicators were found in traits whose formation is related to the insertion and action of major muscle groups. In 14 of the 16 indicators, intraobserver error was <10%. The best indicators of morphological sexual dimorphism were mastoid process, zygomatic bone, mandible, and roughness of the occipital bone. The authors concluded that morphological dimorphism indicators present an adequate performance as diagnostic tests, however, the values of accuracy and sensitivity must be matched with more robust indicators that are independent of the distribution of the sample, and integrate diagnostic errors such as the likelihood ratios, odds ratios, and positive predictive values.

KEY WORDS: Forensic anthropology; Sex identification; Sexual dimorphism; Diagnosis test.

INTRODUCTION

The diagnosis of sex in skeletal remnants with varying degrees of conservation is necessary both in forensic and anthropology practice. There have been numerous methods for carrying out sex determination; they can be divided into morphological and molecular methods (Rosing *et al.*, 2007). Generally, in skeletal remnants in good condition, morphological indicators of sexual dimorphism allow a correct diagnosis in more than 95% of cases (Baughan & Demirjian, 1978; Franklin *et al.*, 2005; Kimmerle *et al.*, 2008). Performance in terms of reliability and error of quantitative methods in sex determination has been widely documented for most of the skeletal elements (Bidmos & Asala, 2003; Introna *et al.*, 1993; 1998). In 1955, Krogman introduced a set of 14 indicators that allowed the morphologic diagnosis of sex from a human skull (Krogman, 1955). Subsequently, Krogman & Iscan (1986) suggested the use of four additional characteristics: size and shape of the nasal opening, size of the nasal bone, zygomatic arch length, and shape of the chin.

As skeletal characteristics vary in different populations, specific analysis is required (Steyn & Iscan, 1998). This situation became evident after a study on an indigenous South African skull. The study could not determine the presence of sexual dimorphism from the traditional indicators described by Krogman along with other distinctive features between males and females, some of which were not described for the Caucasian population (Franklin *et al.*). Some authors have described the influence of the size and intrasexual variations in certain populations (Kimmerle *et al.*), determining errors in the diagnosis of sex made by traditional methods on the basis of the classical parameters described by Krogman.

Against this background, the purpose of this study is to assess the quality of the main morphological indicators of sexual dimorphism in dry human skulls by direct anthropometry and using the tools of epidemiology and clinical biostatistics for a diagnostic test. The skulls were

* Department of Morphology, Universidad de Talca, Chile.

** Department of Morphology and Genetic, Universidade Federal de São Paulo, Brazil.

taken from the collection in the Universidade Federal de São Paulo (UNIFESP), Brazil Skulls Museum.

MATERIAL AND METHOD

A cross-study evaluation of the diagnostic test was designed, using a set of tools used in epidemiology and biostatistics. The study had the approval of the Ethics Committee of UNIFESP/ Hospital de São Paulo.

A nonprobability sample of 284 adult skulls of Brazilians, in good condition, were selected from the collection of UNIFESP. In the event of partial destruction of a segment, it was decided to use the skull, except for analysis and calculations that include the indicators destroyed in the segment; skulls with abnormal growth, asymmetries, and obvious dimorphism were excluded. The sample was made up of 187 male skulls and 97 female skulls. The sample size is consistent with those described in the literature (Krogman; Dominok, 1959; Giles & Elliot, 1963; Olivier & Tissier, 1977; Rösing, 1977; Rösing *et al.*, 2005; Rogers, 2005).

Sixteen indicators of classical morphological sexual dimorphism, as described by Krogman and Krogman & Iscan were analyzed: Size and architecture of the skull, forehead shape, frontal eminences, superciliary arches, orbital shape, piriform aperture, nasal bone, zygomatic bone, zygomatic arch, parietal eminences, mastoid process, occipital bone, occipital condyles, shape of the palate, general appearance of the mandible, and chin shape.

The diagnosis for each indicator and the skull was in a separate tab. Neither was the investigator in sight, nor the diagnoses for the skull made on the basis of the record or other indicators of sex according to the UNIFESP archive's collection of skulls.

The diagnostic tests based on the observation of the morphological indicators of sexual dimorphism were evaluated in terms of accuracy, intraobserver error sensitivity, positive predictive value, likelihood ratios, and odds ratio.

RESULTS

All indicators studied had high levels of accuracy (84.75-92-89%). The best indicators in terms of overall accuracy, intraobserver error, and odds ratio values were found in features whose formation was related to the insertion and action of major muscle groups. The lowest accuracy of all the

morphological indicators for the diagnostic tests in this study was obtained from those based on the observation of the shape. In 14 of the 16 indicators, intraobserver error was <10%. Table I shows the summary of the values of accuracy, intraobserver error, odds ratio, sensitivity, predictive values, and likelihood ratios for the 16 classical morphological indicators studied as diagnostic tests to determine sex.

The best indicators of morphological sexual dimorphism based on overall accuracy, combined ranking and reliability, and odds ratio were the mastoid process, the zygomatic bone, mandible, and ridges of the occipital bone. Indicators with lower overall accuracy were shape of the palate, chin, and orbit. Epidemiological statistical tools were used to help ensure a more robust analysis, especially when considering the overall sensitivity of indicators and values of probability coefficients. In Table II, a combined ranking obtained from the scores of accuracy, intraobserver error, and odds ratio tests are given.

DISCUSSION

The determination of sex is a common diagnostic procedure in forensic anthropology. When the diagnosis is based on the observation of cranial morphology, each parameter becomes a test. The examiner and the forensic or morphological physical anthropologist analyzes the indicators by applying a single group or all of them, to obtain the best diagnostic test for sexual dimorphism. These conditions allow the use of tools of epidemiology and biostatistics to better describe the applicable indicators in terms of accuracy and prediction.

In this study, 16 of the 17 classic morphological indicators of sexual dimorphism were analyzed. Excluded from this analysis was the assessment of teeth because of the wide variety of characteristics of teeth in the skulls of the sample.

All the indicators studied had high accuracy (84.75–92.89%); these levels are similar to those reported by Rogers and Williams & Rogers (2006). Most of the best indicators in terms of overall accuracy, intraobserver error, and odds ratio values were found in features whose formation was related to insertion and action of major muscle groups, such as size and general architecture of the skull, mastoid process, zygomatic bone, ridges of the occipital bone, and general appearance of the mandible. By contrast, there is evidence that the decrease in muscle action or severe malnutrition predisposing atrophy of bone tissue can decrease the accuracy of these methods (Suazo *et al.*, 2008b).

Table I. Summary of the values of accuracy, intraobserver error, odds ratios, sensitivity, predictive values, and likelihood ratios for the 16 classical morphological indicators of diagnostic tests to determine sex.

Indicator	Accuracy %	Error %	Odds Ratios	Sensitivity		Predictive values		Likelihood ratios			
				Male	Female	Male	Female	Male		Female	
				%	%	%	%	+	-	+	-
Size and architecture of the skull	80.60	3.3	16.07	82.22	77.65	87.57	69.52	3.68	0.22	4.36	0.27
Forehead shape	80.14	10.0	18.67	77.65	84.94	90.84	66.38	5.15	0.26	3.80	0.19
Frontal eminences	75.47	10.0	10.24	73.98	78.26	86.48	61.53	3.40	0.31	3.13	0.29
Superciliary arches	78.75	6.6	11.28	79.21	77.89	87.03	66.66	3.58	0.26	3.74	0.27
Orbital shape	75.50	13.3	9.88	74.86	76.84	85.89	61.86	3.26	0.32	3.06	0.30
Piriform aperture	83.30	5.0	8.61	74.40	74.70	84.50	61.20	2.95	0.34	2.96	0.33
Nasal bones	79.38	11.6	13.52	81.14	75.86	87.11	66.66	3.36	0.24	4.03	0.29
Zygomatic bone	83.51	3.3	25.30	83.70	83.10	90.30	73.14	4.98	0.19	5.12	0.20
Zygomatic arch	75.90	5.0	9.53	76.83	74.19	85.00	67.72	2.97	0.31	3.20	0.33
Parietal eminences	80.66	6.6	4.76	70.45	66.66	80.00	54.30	2.10	0.44	2.25	0.47
Mastoid process	84.75	1.6	30.51	92.09	80.00	84.80	84.40	3.01	0.098	10.1	0.33
Ridges of the occipital bone	80.90	3.3	15.35	86.50	70.52	84.60	73.60	2.90	0.19	5.26	0.33
Occipital condyles	79.30	8.3	14.10	80.00	77.70	87.26	67.30	3.57	0.25	3.90	0.27
Shape of the palate	72.89	8.3	5.82	80.89	57.89	78.26	61.79	1.90	0.32	3.03	0.52
Mandible	81.40	3.3	17.88	82.60	79.01	88.66	69.56	3.95	0.22	4.54	0.25
Chin shape	75.41	8.3	7.99	79.30	67.50	83.00	62.06	2.44	0.3	3.27	0.40

Table II. Combined ranking derived from the score of accuracy, intraobserver error and odds ratios of diagnostic tests based on the observation of the 16 classical morphological sexual dimorphism indicators.

Indicator	Score A Accuracy	Score B Error	Score C Odds Ratios	Score A+B+C	Combined Ranking
Mastoid process	1	1	1	3	1
Zygomatic bone	2	2	2	6	2
Mandible	4	2	4	10	3
Ridges of the occipital bone	5	2	6	13	4
Size and architecture of the skull	7	2	5	14	5
Forehead shape	8	6	3	17	6
Piriform aperture	3	3	13	19	7
Occipital condyles	10	5	7	22	8
Nasal bones	9	7	8	24	9
Superciliary arches	11	4	9	24	10
Parietal eminences	6	4	16	26	11
Zygomatic arch	12	3	12	27	12
Frontal eminences	14	6	10	30	13
Orbital shape	13	8	11	32	14
Chin shape	15	5	14	34	15
Shape of the palate	16	5	15	36	16

Rösing indicates that the evaluation of gender based on qualitative indicators should be contextualized to the population. Park & Nowosielski-Slepowron (1983) described socioenvironmental influences (e.g., nutrition, food, weather, diseases) that affect the development and the appearance of the bones.

The lowest accuracy of all the morphological indicators for the diagnostic tests in this study was based on the observation of the shape of the palate, which is again consistent with the findings reported by Williams & Rogers. In relation to this indicator, information on the state of teeth and the presence of the maxillary canines seem to be

important. Suazo *et al.* (2008a) examined the accuracy and reliability of the diagnostic tests based on the shape of the palate of subjects divided into groups with different levels of edentulism. Suazo *et al.* (2008a) concluded that the loss of teeth alters the diagnostic value of the observation of the shape of the palate, which mainly affects the identification of women's skulls.

This study had low intraobserver error and most indicators were within an acceptable level of error, in line with similar studies ($\leq 10\%$) (Williams & Rogers). Only the nasal bones and shape of the orbit showed higher levels of error (11.6% and 13.3%, respectively). According to Ofodile

(1994) and Ofodile & Bokhari (1995), the high variance in the size of the nasal bones is related to racial issues. This might explain the lower reliability of this indicator in a sample having a high racial mix, such as the skulls used in this study.

The reliability of diagnostic tests analyzed is directly related to the experience of the observer. Rösing *et al.* (2007) reported that investigators with little experience in the forensic diagnosis of sex in a population-specific analysis produce higher levels of error and less accurate results; the examiner needs to provide essential utility value in the practice of any diagnostic test. A test with high accuracy but a low reproducibility has little practical value and should be avoided.

The highest levels of reliability in this study were found in indicators that showed increases, peaks, and visible and palpable ridges (size and general architecture of the skull, mastoid process, zygomatic bone, ridges of the occipital bone, and appearance of the mandible), whereas indicators based on the form (shape of the palate, the chin, forehead shape, frontal eminences, nasal bones, shape of the orbit) had the lowest reliability. In this regard, Rösing *et al.* (2007) indicate that in assessing qualitative aspects of the skull, techniques that complement the observation, such as palpation or weight and thickness, favor the diagnosis and reduce the intra- and interobserver error.

For an indicator to be useful in discriminating between male and female skulls, it is necessary to have a consistent relationship between accuracy and reliability. Williams & Rogers recommend the use of a combined ranking built from the scores of accuracy and reliability; their results are similar to those reported in this study.

Epidemiological statistics give an important indicator of overall performance of a diagnostic test, the odds ratio. In this study, to the combined ranking of accuracy and reliability the value of odds ratio was added as an indicator of overall performance of the test, as it incorporates the error or false positive and negative diagnoses to assess the usefulness of the test. Comparing the results of this study and the indicators in the ranking that include combined odds ratio as reported in similar studies (Rogers), the zygomatic bone and mastoid process indicators were found to be comparable, but the piriform aperture indicator shows differences. Although there was a high level of accuracy (score 3) and a low intraobserver error (5%), the high frequency of false positive and negative diagnoses (Odds ratio 8.61, score 13) showed that the result of ranking combined under a deal (Rank 7). In the Williams & Rogers study, piriform aperture has a good place in the combined ranking (Rank 3), despite the high intraobserver

reported error (10%). This indicates that the incorporation of the odds ratio values for the evaluation of diagnostic tests provides valuable information when integrating accuracy and reliability.

Sensitivity is a value that not only allows the researcher to make an initial approach to the classification – a screening – but it is also useful to the low specific diagnosis. If an appraiser knows that there is a group of skulls belonging to males, a highly sensitive indicator, such as the mastoid process can be used to select those possibly male skulls; more specific evidence can be subsequently applied to make the diagnosis. It is said that this sensitivity is influenced by the distribution of the sample or in epidemiological terms, the prevalence of the condition being studied. This has a double significance for the diagnosis of sex using human skulls: (i) as already mentioned, the sensitivity reported in the literature has a population-specific behavior; constitutional, environmental, and nutrition can influence the performance indicators of morphological sexual dimorphism, and (ii) even comparable populations in the composition of the sample has an influence on the sensitivity of the indicators.

Epidemiological tools allow solving the problem of the influence of the prevalence and composition of the sample in the sensitivity parameter; however, the use of positive likelihood ratios (LR +), also called reasons of verisimilitude, are independent of the composition of the sample. In this study, the LR + values were moderate to low, 5.15-1.9 in males, but in females, the mastoid process indicator showed a high LR + (10.1). Other indicators present similar behavior for males with values of between LR + 5.26-2.25. The high value of LR + of mastoid process allows classification of a skull with high accuracy as having female characteristics, in this case, the indicator had a score of 3 in the ranking of sensitivity to females. However, its diagnostic value is greater; this should be considered when weighing the morphological indicators of sexual dimorphism.

Steadman *et al.* (2006) published an extensive study on the statistical basis for identification in forensic anthropology, which states that the epidemiological tools, especially the values of variability coefficients, provide sufficient statistical basis to support the diagnosis made by experts.

Another interesting aspect is the relationship between an indicator of sensitivity and positive predictive value. Most studies, when analyzing the performance of a morphological indicator of sexual dimorphism, expressed their results in terms of number of hits or true positives when applying the diagnostic test and is expressed in percentage terms; this

value corresponds to the sensitivity of the indicator. In this study, the sensitivity for diagnosis of male indicator mastoid process was very high (92.09%), indicating that it is a good indicator for identifying male skulls, the problem is that this value does not reflect the number of incorrect diagnoses, that is, the number of female skulls diagnosed as male skulls. Noting that the mastoid process to correct this defect supplements the analysis of the sensitivity of the indicators in calculating the positive predictive value, this indicator reflects the false positives and allows a balance of more accurate performance indicator for the diagnostic test based on the mastoid process. The positive predictive value was 84.8%, indicating that some female skulls were misdiagnosed by observing the indicator; this is reflected in the decreased sensitivity to females (80%).

Therefore, to make a proper assessment of the performance of a morphological indicator of sexual dimorphism, the values of sensitivity, which often refers to the literature, should be supplemented by statistics that diminish the impact of the composition of the sample, such as likelihood ratios and diagnostic data, incorporating the false positive and positive predictive values.

Finally, the use of statistical tools applied to the epidemiological diagnosis of sex is not sufficiently explored, in contrast to the extensive development in clinical epidemiology and evidence-based medicine. This opens new perspectives for the systematic evaluation of the morphological sexual dimorphism indicators.

SUAZO, G. I. C.; ZAVANDO, M. D. A & SMITH, R. L. Evaluación del rendimiento como una prueba de diagnóstica de los métodos clásicos para la determinación forense del sexo. *Int. J. Morphol.*, 27(2):381-386, 2009.

RESUMEN: En el estudio fue evaluada la calidad, como una prueba diagnóstica de los principales indicadores morfológicos de dimorfismo sexual a través de la antropometría, herramientas bioestadísticas y epidemiología clínica. Fueron utilizados 284 cráneos de individuos adultos brasileños, de los cuales 187 eran varones y 97 mujeres. Se realizó un estudio de corte transversal de evaluación de prueba diagnóstica, mediante un enfoque cualitativo en base al examen visual, de 16 indicadores tradicionales morfológicos de dimorfismo sexual, determinándose para cada uno de ellos los niveles de exactitud y de sensibilidad, valores predictivos, coeficientes de probabilidad y Odds ratios. Todos los indicadores estudiados presentaron altos niveles de exactitud (84,75-72,89%). Los mejores indicadores se encontraron en los rasgos cuya formación está relacionada con la inserción y la acción de los grandes grupos musculares. En 14 de los 16 indicadores, el error intraobservador fue <10%. Los mejores indicadores morfológicos de dimorfismo sexual fueron: proceso mastoideo, hueso cigomático, mandíbula y rugosidades del hueso occipital. Los autores concluyeron que los indicadores morfológicos de dimorfismo presentan un adecuado rendimiento como pruebas de diagnóstico; sin embargo, los valores de exactitud y sensibilidad deben ser complementados con indicadores más sólidos que sean independientes de la distribución de la muestra, e integren a los errores diagnósticos, tales como los coeficientes de probabilidad, odds ratios y valores predictivos positivos.

PALABRAS CLAVE: Antropología forense; Determinación del sexo; Dimorfismo sexual; Prueba diagnóstica.

REFERENCES

- Baughan, B. & Demirjian, A. Sexual dimorphism in the growth of the cranium. *Am. J. Phys. Anthropol.*, 49:383-90, 1978.
- Bidmos, M. A. & Asala, S. A. Discriminant function sexing of the calcaneus of the South African whites. *J. Forensic Sci.*, 48:1213-8, 2003.
- Dominok, G. W. On age and sex determination from the morphology of the human cranial vault. *Zentralbl. Allg. Pathol.*, 100:54-64, 1959.
- Franklin, D.; Freedman, L. & Milne, N. Sexual dimorphism and discriminant function sexing in indigenous South African crania. *Homo*, 55:213-28, 2005.
- Giles, E. & Elliot, O. Sex determination by discriminant function analysis of crania. *Am. J. Phys. Anthropol.*, 21:53-68, 1963.
- Introna, F. Jr.; Cantatore, F.; Dragone, M. & Colonna, M. Sexual dimorphism of deciduous teeth in medico-legal identification. *Boll. Soc. Ital. Biol. Sper.*, 69:223-30, 1993.
- Introna, F. Jr.; Di Vella, G. & Campobasso, C. P. Sex determination by discriminant analysis of patella measurements. *Forensic Sci. Int.*, 95:39-45, 1998.
- Kimmerle, E. H.; Ross, A. & Slice, D. Sexual dimorphism in america: geometric morphometric analysis of the craniofacial region. *J. Forensic Sci.*, 53:54-7, 2008.
- Krogman, W. M. The human skeleton in forensic medicine. *Postgrad. Med.*, 17A:48-72, 1955.

- Krogman, W. E. & Iscan, M. Y. *The Human Skeleton in Forensic Medicine*. Springfield, Charles C. Thomas, 1986.
- Ofodile, F. A. Nasal bones and pyriform apertures in blacks. *Ann. Plast. Surg.*, 32:21-6, 1994.
- Ofodile, F. A. & Bokhari, F. The African-American nose: Part II. *Ann. Plast. Surg.*, 34:123-9, 1995.
- Olivier, G. & Tissier, H. Sexual dimorphism of correlations in craniometry. *Bull. Assoc. Anat.*, 61:259-67, 1977.
- Park, A. W. & Nowosielski-Slepowron, B. J. Postnatal skull growth of the rat relative to environmental variance. *Acta Morphol. Neerl. Scand.*, 21:141-63, 1983.
- Rogers, T. L. Determining the sex of human remains through cranial morphology. *J. Forensic Sci.*, 50:493-500, 2005.
- Rösing, F. W. Anthropometry of six tribal populations in Angola. *Z. Morphol. Anthropol.*, 68:107-22, 1977.
- Rösing, F. W.; Graw, M.; Marré, B.; Ritz-Timme, S.; Rothschild, M. A.; Rötzscher, K.; Schmeling, A.; Schröder, I. & Geserick, G. Recommendations for the forensic and age determination from the skeleton. *Anthropol. Anz.*, 63:221-32, 2005.
- Rösing, F. W.; Graw, M.; Marré, B.; Ritz-Timme, S.; Rothschild, M. A.; Rötzscher, K.; Schmeling, A.; Schröder, I. & Geserick, G. Recommendations for the forensic diagnosis of sex and age from skeletons. *Homo*, 58:75-89, 2007.
- Suazo, I.; Zavando, D. & Smith, R. Accuracy of palate shape as sex indicator in human skull with maxillary teeth loss. *Int. J. Morphol.*, 26(4):989-93, 2008a.
- Suazo, I.; Zavando, D. & Smith, R. Evaluating accuracy and precision in Morphologic Traits for sexual dimorphism in malnutrition human skull: A comparative study. *Int. J. Morphol.*, 26(4):876-83, 2008b.
- Steadman, D. W.; Adams, B. J. & Konigsberg, L. W. Statistical basis for positive identification in forensic anthropology. *Am. J. Phys. Anthropol.*, 131:15-26, 2006.
- Steyn, M. & Iscan, M. Y. Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Sci. Int.*, 98:9-16, 1998.
- Williams, B. A. & Rogers, T. Evaluating the accuracy and precision of cranial morphological traits for sex determination. *J. Forensic Sci.*, 51:729-35, 2006.

Correspondence to:
Prof. Dr. Iván Suazo Galdames
Department of Morphology
Universidad de Talca
Avenida Lircay s/n oficina N°104
CHILE.

Phone: 56-71-201576, Fax: 56-71-201576

E-mail: isuazo@utalca.cl

Received: 16-12-2008
Accepted: 22-01-2009