A56 Sex Estimation Using Discriminate Function Analysis of Non-Metric Cranial Traits: An Inter-Observer Error Study

Janna M. Andronowski, MSc, 250 S Stadium Hall, Knoxville, TN 37917; Kyle A. McCormick, MA, 824 W Hill Avenue, Apt 301, Knoxville, TN 37902-3015; and Michael W. Kenyhercz, MS, University of Alaska Fairbanks, 403 Salcha Street, 310 Eielson Bldg, Fairbanks, AK 99775*

The goal of this presentation is to explore inter-observer error rates of sexually dimorphic non-metric cranial traits using the Walker five-trait scoring system.

This presentation will impact the forensic science community by providing further treatment to inter-observer error rates for non-metric cranial traits. While Walker assessed inter- and intra-observer error rates for sexually dimorphic cranial traits, the sample size was small (n=10) and the confounding variables of observer interaction and observer experience were not partitioned out from this assessment.[1] The current study attempts to evaluate inter-observer error derived from the scoring method by controlling for experience level.

Skeletal sex estimation is traditionally based on the subjective morphological assessment of traits of the pelvis and skull. To standardize morphological assessment, Walker developed a five-trait scoring system to evaluate the expression of non-metric cranial sex traits. The traits include the supraorbital margin, the mastoid process, the supraorbital ridge/glabella, the nuchal crest, and mental eminence. Each trait is scored on a scale of 1 to 5, which shows the progression of the traits from gracility (1) to robust (5).

Minimal treatment has been given to inter-observer error rates for visually scored traits of the skull. As such, the objectives of this study are to: (1) evaluate inter-observer error derived solely from the original scoring method as proposed by Walker; and, (2) assess how inter-observer error affects sex estimation.

The current work evaluates a large sample of individuals of known age and sex (n=78; F=37, M=41) using the Walker five-trait scoring system. Data were obtained from the William M. Bass Donated Skeletal Collection housed at the University of Tennessee, Knoxville. Ages range from 24 to 88 years old. Each of the traits was scored in accordance with Walker.[1] The left side was evaluated in the event of paired traits. Trait scores were compared among three anthropologists with similar levels of experience to negate error attributed to varied levels of experience. As such, any variation in scores among observers can be ascribed to inherent limitations in the scoring method as opposed to differences in experience level.

Statistical software was used to perform Fleiss' Kappa, the Intra-Class Coefficient (ICC) and Discriminant Function (DF) analyses.[2] The ICC was applied to account for close observations between observers and was used with a two-way model that evaluated both consistency and absolute agreement. Trait score differences among observers were assessed and subsequently compared to those reported by Walker.[1]

Results indicate that, for most traits, there is consistent scoring agreement among observers. The majority of the traits showed substantial absolute agreement (>0.7) with the exception of the mental eminence, which showed a moderate agreement of 0.54. Further, the observations were substantial to almost perfectly consistent between anthropologists.

Agreement between observers was the same for each DF at approximately 72%. In cases of disagreement, sex bias was idiosyncratic to the observer. A score of 1 for female and -1 for male was assigned to each disagreeing observer's sex estimation. Observer Three had a sex bias score of -10, or a strong male bias. Observer Two had a sex bias score of 13, or a strong female bias. Observer One had a sex bias score of 2, or no appreciable sex bias. Though the DF identified a sex bias in two of the three observers, correct classifications of sex were not impacted by the observer bias as each scorer showed correct classifications higher than 77%.

The current study provides further treatment to inter-observer error rates for non-metric cranial traits. Though results indicate that there is some subjectivity associated with the scoring system, trait scores can be reliably assigned by observers with comparable levels of experience. Thus, when possible, sex estimation from the skull should be confirmed through blind peer review.

In contrast to inter-observer error results reported by Walker, the majority of traits scored revealed substantial absolute agreement among observers.[1] As such, by controlling for observer experience level, the inter-observer error inherent in the scoring system is quite low.

References:

1. Walker PL. Sexing skulls using discriminant function analysis of visually assessed traits. *Am J Phys Anthropol*. 2008;136(1):39-50.
2. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2011.

Inter-Observer Agreement, Cranial Non-Metrics, Sex Estimation