H5 Inter- and Intra-Observer Reliability Using the Walker Non-Metric Sexing Technique

Cheyenne J. Lewis, BS, 3108 Pine Avenue, Erie, PA 16504; Heather M. Garvin, PhD, Mercyhurst University, Dept of Applied Forensic Sciences, 501 E 38th Street, Erie, PA 16546; and Alex Taylor, BA, 922 E 31st Street, Erie, PA 16504*

After attending this presentation, attendees will better understand observer error rates using non-metric cranial traits and the effects of observer error when estimating the sex of unknown individuals.

This presentation will impact the forensic science community by providing knowledge on the reliability of using non-metric cranial sex traits and the practical consequences of observer error when estimating the sex of unknown skeletal remains in a forensic context.

While the validity of traditional non-metric techniques of sex estimation has been contested due to their subjective nature, some studies have found that certain non-metric traits do perform well when examined by multiple observers.[1,2] A study by Walker evaluated the use of five non-metric traits for sex estimation and provided descriptions and illustrations to demonstrate specific levels of expression for each trait.[3,4] The inter- and intra-observer error rates were calculated and the assigned scores between observers were found to fall within one score of the modal value 96% of the time and 99.5% for the intra-observer test.[3] These high levels of agreement suggest that the method can be reliably used. As the method is commonly used to analyze forensic cases, it is important to validate the reliability of the method. This study was designed to test the accuracy of Walker's reported inter- and intra-observer rates for each cranial trait. In addition, the practicality of this method was examined through the consequences of disagreement between and within observers in an effort to identify traits that need better clarification and description.

The Walker method was used to score five cranial traits in African American individuals from the Hamann-Todd Collection (total n=59). Three different observers scored the crania, two of whom scored the crania twice with at least a week between scoring rounds. Statistical analyses were performed to calculate the inter- and intra-observer error rates. The scores were also applied to the equation provided by Walker with the highest reported accuracy rates. The estimated sexes obtained were then compared across observers and to the known documented sex to evaluate the forensic consequences of observer error.

Pairwise comparisons completed using Cohen's weighted kappa revealed that there was moderate to substantial agreement (0.43-0.80) between and within observers for the mastoid process, orbital margin, and glabella, fair to substantial agreement (0.25-0.64) for the nuchal crest, and only slight to fair agreement (0.05-0.38) for the mental eminence. The percentage of agreement tests between observers indicated that the mastoid process, orbital margin, and glabella were the most consistently scored traits, while the mental eminence performed the worst with agreement ranging from 52.5% to 67.8%. Similarly, the intra-class correlation coefficient tests demonstrated that deviations between observers' scores for each round were greatest in the mental eminence with low agreement scores (0.07 and 0.20). For all tests, glabella performed best with high levels of agreement while mastoid and orbital margin exhibited moderate levels of agreement. The correct classification rates for each observer ranged from 67.8% to 79.7%, which are lower than Walker's published rates.[3] Overall, the results indicate that accuracy rates of this method are affected by observer error of certain traits which could have a vital impact in the forensic application of this method.

Results show that the mental eminence is the least reliable cranial trait for sex determination both between and within observers. The mental eminence had the lowest agreement scores for all inter- and intra-observer error tests followed by the nuchal crest scores; however, the equation used from the original study by Walker suggested to have the highest accuracy rates gives the mental eminence approximately equal weight as the glabella and mastoid process scores.[3] While the mental eminence did not prove to be a reliable trait, glabella and the mastoid process had high levels of agreement which likely enhances the efficacy of the method. The low reliability of the mental eminence could either be the result of low levels of sexual dimorphism or it may reflect the need for enhanced visual and descriptive tools to capture the sex difference in shape and more objectively score this feature. Until then, more reliance should be placed on methods utilizing the glabella, orbital, and mastoid regions as these traits proved to be reliably scored cranial traits both between and within observers.

References:

1. Walrath DE, Turner P, Bruzek J. Reliability test of the visual assessment of cranial traits for sex determination. Am J Phys Anthropol 2004:125:132-7.
2. Williams BA, Rogers T. Evaluating the accuracy and precision of cranial morphological traits for sex determination. J Forensic Sci 2006:51:729-35.

 ** Presenting Author*

3. Walker PL. Sexing skulls using discriminant function analysis of visually assessed traits. Am J Phys Anthropol 2008:136:39-50.
4. Buikstra JE, Ubelaker DH. Standards for data collection from human skeletal remains. Arkansas Archeological Survey Research Series No. 44, 1994.

Observer Error, Sex Estimation, Non-Metric

 ** Presenting Author*