



A6 Reliability and Validity of the Walker and Klales, et al. Methods

Mackenzie Walls*, Forensic Anthropology Program, Washburn University, 1700 SW College Avenue, Topeka, KS 66621; Alexandra R. Klales, PhD, Washburn University, Forensic Anthropology Program, Soc & Anthro Dept, 1700 SW College Avenue, HLRC#218, Topeka, KS 66621; Kate M. Lesciotto, JD, MS, Pennsylvania State University, Dept of Anthropology, 409 Carpenter Bldg, University Park, PA 16802; Timothy P. Gocha, PhD, University of Nevada, Las Vegas, Dept of Anthropology, 4505 S Maryland Parkway, Las Vegas, NV 89154; and Heather M. Garvin, PhD, Des Moines University, Dept of Anatomy, 3200 Grand Avenue, Des Moines, IA 50312-4198

After attending this presentation, attendees will better understand the reliability and validity of two commonly used morphoscopic sex estimation methods.

This presentation will impact the forensic science community by reporting observer consistency in scoring commonly used skull and pelvis traits and investigating the role that experience level plays in obtaining accurate sex estimations.

To have forensic utility, methods must be both valid and reliable. Validity refers to how well a method measures what it is supposed to (i.e., accuracy) and reliability refers to the ability to produce consistent results. External validity and reliability of methods must be assessed prior to their incorporation into standard operating procedures, and, to comply with *Daubert* standards, forensic methods must be tested and potential error rates published. As a result, there has been a push to translate traditional qualitative methods into quantitative methods, in which morphoscopic traits are assigned ordinal scores; however, these scoring methods do not eliminate method subjectivity completely, and thus it is important to assess validity and reliability in these trait scoring methods.

Two popular sex estimation methods, Walker and Klales, et al., provide an ordinal scoring method for traits of the skull and pelvis, respectively.^{1,2} Walker utilizes the nuchal crest, glabella, supraorbital margin, mental eminence, and mastoid process. Klales et al. utilizes the ventral arc, subpubic contour, and medial aspect of the ischio-pubic ramus.^{1,2} The validity and reliability of these methods have been tested only a handful of times since their incorporation into forensic casework, and the results have varied. Recognizing the need for updated standards in the field, trait data have been collected from more than 2,500 individuals as a part of a National Institute of Justice (NIJ) grant to: (1) assess the reliability and validity of these methods; (2) evaluate the impacts of population differences, secular change, and asymmetry on sex estimation; and, (3) create a free morphological database for sex estimation using these traits/methods: MorphoPASSE. This presentation addresses the first of these goals: trait scoring validity and reliability.

Three observers with varying levels of experience (expert/grant Principal Investigator (PI), experienced, and inexperienced) collected ordinal score data using the Walker and Klales et al. traits. This interobserver sample consisted of 222 individuals from the Hamann-Todd (HTH) and Bass skeletal collections. The expert observer scored the individuals twice, with a year between scoring events for intraobserver analyses. Additional trait data were contributed by four other researchers, with various levels of experience, thereby facilitating additional tests of observer error. Their data were collected from the HTH ($n=174$), Bass ($n=57$), and the Operation Identification and donated collections at Texas State University ($n=57$) and included individuals also scored by the grant PI. Interobserver error was assessed using the Intraclass Correlation Coefficient (ICC), while intraobserver error was assessed with quadratic weighted Kappa (wK). Sex classification accuracy was evaluated using the logistic regression equations provided in the original publications.

Intraobserver agreement was nearly perfect for the pelvis and substantial for the skull. The ICC indicated excellent levels of agreement between the three observers that scored the entire sample of 222 individuals, with the exception of the orbital margin (good agreement). The expert and experienced observer achieved higher agreement than the inexperienced observer, indicating that experience does play a role to some degree. When the additional contributed trait data were included in analyses, excellent agreement was obtained between the three expert observers for all traits except the mental eminence. Classification accuracy was high for all three experience levels for the pelvis (96.6% expert, 93.6% experienced, and 78.2% inexperienced), but was generally lower, with a high sex bias, for the skull regardless of experience (73.5% expert, 61.4% experienced, and 70.7% inexperienced). In conclusion, the reliability results indicate general consistency in trait scoring among observers, while the validity results suggest that experience plays a larger role in the accurate application of the methods than has been previously reported; observers with a great deal of experience can expect much higher sex classification accuracy than observers with less experience.

This research was funded by an NIJ Grant. Opinions, points of view expressed in this research, and products discussed represent a consensus of the authors and do not necessarily represent the official position, policies, or endorsement of the United States Department of Justice or the NIJ.

Reference(s):

1. Walker P.L. Sexing skulls using discriminant function analysis of visually assessed traits. *Am J Phys Anthropol.* 2008;136:39-50.
2. Klales A.R., Ousley S.D., Vollner J.M. A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *Am J Phys Anthropol.* 2012;149:104-114.

Walker Method, Klales et al. Method, Sex Estimation