



## A2 An Evaluation of the Inter-Observer Reliability of Aging Methods From the Pubic Symphysis

Tyler E. Dunn, MS\*, Creighton University School of Medicine, Omaha, NE 68178; Sarah C. Kindschuh, PhD, Defense POW/MIA Accounting Agency Laboratory, Offutt Air Force Base, NE 68113; Brittany S. Walter, PhD, Defense POW/MIA Accounting Agency Laboratory, Offutt Air Force Base, NE 68113

**Learning Overview:** The goal of this presentation is to assess the inter-observer reliability of pubic symphysis scores for age estimation, to evaluate a potential area of error in pubic symphyseal aging, and to quantify the comparative inter-observer variation between a component and phase-based methods and among analysts with different levels of experience.<sup>1</sup>

**Impact on the Forensic Science Community:** This study impacts the forensic science community by reviewing the replicability of morphological assessments used for age estimation from the pubic symphysis and by validating the inter-observer reliability of these methods for use in forensic casework.

Any method applied to forensic casework must be proven valid and repeatable based on the *Daubert* standard; as such, an assessment of inter-observer reliability is necessary to ensure that these methods have replicability acceptable for use in a forensic context. Pubic symphyseal morphology has a long history of forensic applications and has the ability to provide relatively narrow age estimate intervals.<sup>2,3</sup> McKern and Stewart devised a component-based approach for age estimation from the pubic symphysis in which a summed composite score is calculated based on three independently evaluated morphological areas of the symphyseal face: (1) the Dorsal demiface (D); (2) the Ventral demiface (V); and (3) the symphyseal Rim (R).<sup>2</sup> The Brooks and Suchey method of age estimation from the pubic symphysis is a phase-based system, which relies on the overall assessment of the symphyseal face into one of six developmental phases.<sup>3</sup> The purpose of this study is to evaluate inter-observer reliability between these methods and to assess inter-observer reliability among analysts with differing levels of anthropological experience for each method.

The study sample consisted of 167 os coxae from United States service members from the USS *Oklahoma* loss at Pearl Harbor on December 7, 1941, identified at the Defense POW/MIA Accounting Agency (DPAA) Laboratory. All individuals were male, and damaged or pathologically altered pubic bones were excluded from analysis. Eight analysts independently assessed each os coxa in the blind. The analysts were divided into three cohorts: inexperienced ( $n=5$ ), experienced ( $n=3$ ), and a combined cohort of inexperienced and experienced ( $n=8$ ). The inexperienced cohort was comprised of individuals with some level of postgraduate anthropological training and no casework experience; the experienced cohort was comprised of DPAA forensic anthropologists with advanced degrees in anthropology and, at a minimum, one year of casework experience. Each os coxa was scored following the McKern/Stewart and Brooks/Suchey methods, and each analyst had access to the original publication and plastic casts for both methods. Inter-observer agreement was determined using an Intraclass Correlation (ICC) analysis to assess inter-observer reliability (mixed model and 95% confidence) and evaluated using the Cicchetti scale. ICCs were calculated for each cohort for the McKern/Stewart method component and composite scores and for the Brooks/Suchey method phases.

The lowest ICCs were consistently observed in the inexperienced cohort, while the experienced cohort demonstrated the highest ICCs. The inexperienced cohort had ICCs of 0.70 and 0.76, and the experienced cohort had ICCs of 0.86 and 0.88 for the McKern/Stewart and Brooks/Suchey methods, respectively. When observations were divided into the independent components of the McKern/Stewart method, the inexperienced group had ICCs of D:0.60, V:0.63, R:0.64, and the experienced group had scores of D:0.79, V:0.87, and R:0.76. The McKern/Stewart method ICCs were lower than the Brooks/Suchey method ICCs for all analyst cohorts.

All scores demonstrate ICCs that indicate good to excellent agreement, an indication that both approaches are broadly consistent across observers and that observations when using these methods are replicable. The differences between the experienced and inexperienced groups indicate that analysts are more consistent when they have more osteological experience and regularly apply the method. The consistently higher ICCs for the Brooks/Suchey method may be a result of detailed descriptions of the changes in symphyseal morphology provided for this method, compared to relatively sparse descriptions provided for the McKern/Stewart method. Within the experienced group, the symphyseal rim component of the McKern/Stewart system had the lowest ICC relative to the other two components, which is possibly a result of poor documentation and guidance for scoring this component in the original publication. Overall, the inter-observer agreement for developmental scores of the pubic symphysis are good to excellent, and these methods are recommended for forensic applications, barring large deviations between the reference sample and individual being assessed.

### Reference(s):

1. Shirley, N.R., and P.A. Ramirez Montes. "Age Estimation in Forensic Anthropology: Quantification of Observer Error in Phase Versus Component-Based Methods." *Journal of Forensic Sciences*, vol. 60, no. 1, 2014, pp. 107–111.
2. McKern T, Stewart T. *Skeletal age changes in young American males analysed from the standpoint of age identification*. Natick, MA: Quartermaster Research and Development Center, Environmental Protection Research Division, 1957.
3. Brooks, S., and J.M. Suchey. "Skeletal age determination based on the os pubis: A comparison of the Acsadi-Nemeskeri and SucheyBrooks methods." *Human Evolution*, vol. 5, no. 3, 1990, pp. 227-238.

### Age Estimation, Inter-Observer Error, Pubic Symphysis