



### A63 Performance Assessment for Osteometric Pair-Matching

*John E. Byrd, PhD\*, 95-033 Hokuwa Street, #51, Mililani, HI 96853-5530; and Carrie B. LeGarde, MA, Defense POW/MIA Accounting Agency, 106 Peacekeeper Drive, Bldg 301, Offutt AFB, NE 68113*

After attending this presentation, attendees will understand how to measure the performance of a forensic test method and will understand what to expect from osteometric pair-matching.

This presentation will impact the forensic science community by providing attendees with error rates that can be used to understand results in the laboratory.

Byrd and Legarde provided osteometric methods for testing the association of paired bones when resolving a commingled assemblage.<sup>1</sup> The overall approach utilized a significance test with a null hypothesis which states that two bones are of a size to have originated from the same person. The reasoning is counterfactual in nature, as in: If one sampled left and right femora of individuals within a random sample of the relevant population and applied this test to each individual, the test would rarely reject the association. This approach does not provide a means to prove that two bones must have originated from the same person. To do that, one must employ a second line of counterfactual reasoning: If one sampled left and right femora of individuals within a random sample of the relevant population and applied the test to bones from different individuals, one would rarely accept the null hypothesis. The first line of reasoning is a straightforward significance test whose efficacy in future applications can be easily evaluated. The second line of reasoning is not straightforward — its veracity will be determined by the circumstances in which the method is applied. Specifically, the diversity of sizes among the commingled individuals will drive how rare it will be to accept the null hypothesis.

Byrd and Adams recommended using  $p < 0.10$  as a standard for rejecting an association of two bones.<sup>2</sup> This standard was intended to balance power against risk in applications. It is not intended to be a draconian cutoff, and analysts are encouraged to use the attained  $p$ -value to evaluate the association. As a rubric for evaluating performance, one can view a null hypothesis rejection ( $P < 0.10$ ) as a “positive” result and acceptance as a “negative.” The False Positive Rate (FPR) is the rate at which associations that actually come from the same individual are rejected. When applying the Byrd-Legarde paired whole bone models (six bones) to DPAA reference data, the FPR ranges from 0.05 to 0.20, with a mean FPR of 0.10. Acknowledging that with  $p < 0.10$  as a standard, one can expect to attain variation in the FPR, these results largely meet expectations. One bone, the ulna, has an FPR (0.20) that is significantly higher than expected (one-sided binomial test  $p = 0.00083$ ). Vickers et al. applied the Byrd-Legarde models to the Forensic Databank (FDB) and attained very similar results (mean FPR = 0.10) to include higher error with the ulna.<sup>3</sup> LeGarde applied a larger variety of models (12 models from three bones) to the Bass Collection and found the FPR to range from 0.07 to 0.13, with a mean FPR of 0.10.<sup>4</sup> She also applied the Byrd models to the collection at Chiba, Japan, and attained an FPR range 0.04 to 0.15, with a mean FPR of 0.09.

One cannot directly address the second counterfactual given above because it is difficult to project what variation in body sizes will hold in future forensic cases; however, one can use the size distribution of the reference sample as a proxy. A sample ( $N = 1,000$ ) of the right and left long bone measurements in the reference sample were randomly paired for comparison and any instances of the same individual removed from the analysis. This yielded a large number of random pairings of bones from different persons. These comparisons provided a basis for assessing false negatives along with other possible outcomes. With these random comparisons combined with



## Anthropology - 2017

the above metrics, the false discovery rate ( $Q_{\text{actual}}$ ), sensitivity, specificity, positive predictive values, negative predictive values, and efficiency were estimated. The results from the six whole bone models (DPAA reference data) are in simple terms:  $Q_{\text{actual}}$  0.01-0.03; sensitivity 0.67-0.84; specificity 0.80-0.95; positive predictive value 0.97-0.99; negative predictive value 0.17-0.52; and efficiency 0.69-0.86. The meaningfulness of these results will be determined by how representative the size distribution of the reference data is. Under strong assumptions about the size distributions of future cases, it is possible to project the future performance of these models.

### Reference(s):

1. Byrd J.E., LeGarde C. Osteometric Sorting. In: Adams B.J., Byrd J.E., editors. *Commingled Human Remains: Methods in Recovery, Analysis, and Identification*. Amsterdam: Academic Press, 2014:167-192.
2. Byrd J.E., Adams B.J. Osteometric sorting of commingled human remains. *J Forensic Sci.* 2003;48: 717-724.
3. Vickers S., Lubinski P.M., DeLeon L.H., Bowen J.T. Jr. Proposed method for predicting pair matching of skeletal elements allows too many false rejections. *J Forensic Sci.* 2015;60(1):102-106.
4. LeGarde C. Asymmetry of the humerus: The influence of handedness on the deltoid tuberosity and possible implications for osteometric sorting. (Thesis). Missoula (MT): The University of Montana, 2012.

### Performance, Osteometric, Statistics