## A65 Sensitivity Analysis of Craniometric Measurements and Modeling Techniques to Assess Impact of Measurement Error on FORDISC® Results

*Michala K. Stock, MA*, University of Florida, 2033 Mowry Road, Rm G-17, Gainesville, FL 32610; Katie M. Rubin, MS, University of Florida, CAPHIL, PO Box 103615, Gainesville, FL 32610; and Alexander McGhee, MS, University of Florida, Dept of Mechanical and Aerospace Engineering, Gainesville, FL 32611*

After attending this presentation, attendees will better understand how the magnitude of variation in different craniometric variables affects the results of FORDISC® analyses.

This presentation will impact the forensic science community by challenging the current rule-of-thumb tolerance for measurement error in craniometrics and highlighting the use of mathematical modeling techniques in assessing the sensitivity of metric analyses.

This research uses iterative manipulation of craniometric measurements to assess the impact of measurement error on FORDISC® classification of sex and ancestry.[1]

Metric analysis of skeletal material is an important tool for forensic anthropologists assessing the biological profile of unknown decedents. As measurements constitute the fundamental data for these analyses, recent efforts have focused on standardizing definitions and procedures to produce the highest degree of precision and accuracy when taking measurements. Part of this undertaking has involved investigating which measurements are more problematic for repeatability, as assessed through both interobserver and intraobserver measurement error studies; however, the impact of the magnitude of measurement error on analytical results has not been well captured. Instead, the widely used rule of thumb for skeletal metrics is an arbitrary ±2mm tolerance level for all measurements, even though the size of standardized measurements from an individual can vary by an order of magnitude within both the skull and the postcrania. The discriminant function program FORDISC® 3.1 is routinely used to assess sex and ancestry from the cranium; it remains unclear how measurement error affects the results of these analyses.

Thirteen standard craniometric measurements were taken from 30 skeletal individuals of varying demographics curated at the C.A. Pound Human Identification Laboratory. To establish a baseline FORDISC® classification for each individual, all 13 original measurements for a given individual were entered into the program, and a discriminant function analysis was run without transformations against all groups in the Forensic Anthropology Data Bank. For this project, whether the baseline classification reflected the decedent's true demographic group was unimportant, as the goal was not to assess the accuracy of the program itself. For each individual, each original measurement was then varied in 1mm increments — up to 5mm greater than and less than the original measured value — and the FORDISC® analysis re-run in each iteration. Only one measurement was altered at a time, resulting in 131 FORDISC® runs per individual. The classification, Mahalanobis distance, posterior probability, and type F typicality results of all runs were recorded and compared. Deviations in classification from the baseline classification were recorded as "disparate classifications." To complement this analysis, Latin Hypercube Sampling (LHS), which varies all measurement values orthogonally in an N-dimensional hyperspace within the range permitted by FORDISC® and generates a new FORDISC® analysis for each iteration, was performed on one individual. The LHS method is widely used to test the general sensitivity of measurement variables to variation due to the few number of sample points needed to produce high-quality results.

Average disparate-classification rates by measurement when each measurement is varied by ±2mm (in accordance with the rule of thumb) ranges from 0%–33%, with cranial base length (basion-nasion length) creating the most deviation and the vault chord lengths creating the least deviation in classification. Additionally, the variance of the posterior probabilities and typicalities about the values generated using the original measurements is asymmetrical, but without a general, clearly defined magnitudinal or directional bias. The LHS sampling analysis revealed that midfacial measurements, especially when varied simultaneously, are the most sensitive to variation, even when sensitivity comparisons are scaled by size. Overall, the findings of these analyses suggest that accepting measurement error of 2mm is inappropriate for some craniometric measurements, particularly the relatively smaller breadths and heights measured in the midface. Conversely, the magnitude of within-demographic group variation for other measurements, particularly those in the vault, subsume even large amounts of measurement error, indicating that these are more tolerant measurements. Moving forward, the use of modeling techniques, such as those used above, will allow for the creation of empirically tested error ranges unique to each craniometric measurement, which satisfies *Daubert* criteria. FORDISC® is a powerful tool, and the more that can be understood concerning how different factors influence the outcomes of its analyses, the more successful and reliable forensic practitioners will be.

**Reference(s):**

1. Jantz R.L., Ousley S.D. *FORDISC® 3: Computerized discriminant functions, version 3.1*. Knoxville (TN): Univ. of Tennessee, 2005.

**Metric Analysis, Error Analysis, Latin Hypercube Sampling**



*Presenting Author