A12 Applying Posterior Probability Thresholds to Traditional Cranial Trait Sex Estimation Methods

Patricia Avent, University of Illinois Urbana Champaign, Urbana, IL 61901; Cris E. Hughes, PhD, University of Illinois, Department of Anthropology, Urbana, IL 61801; Heather M. Garvin, PhD, Des Moines University, Des Moines, IA 50312-4198*

Learning Overview: After attending this presentation, attendees will understand how Posterior Probability-Informed Threshold (PPIT) cutoffs can be used to improve the accuracy and reporting methods of sex estimations based on traditional cranial non-metric traits.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by presenting PPITs as a replacement to the dichotomous distinction (male or female estimate) based on a single sectioning point.

Currently, sex estimation methods using the traditional five cranial non-metric traits produce a final sex estimation (male or female) using the resulting score; for example, the discriminant function score, and a single sectioning point.[1,2] When sex estimations are gleaned from discriminant scoring alone, statistical confidence in the assigned sex is not typically assessed. The ultimate sex assignment is the same regardless of whether the individual falls close to the sectioning point or at the extremes of one of the sexes. Although some practitioners may informally report their confidence in the assessment (e.g., probable male), these confidence statements are subjective, not standardized, and not necessarily based on statistical results. The present study examines sex estimation outcomes using Garvin et al.'s non-metric cranial traits data.[1] The purpose of the present study is threefold: (1) to evaluate how PPITs impact accuracy, (2) assess the balance between sample inclusion and accuracy for the proposed PPIT approach, and (3) make recommendations for the use and interpretation of specific thresholds for this sex estimation method.

To assess how PPITs affect accuracy, posterior probabilities associated with the discriminant function analyses for Garvin et al.'s 2014 sample (n=491) were used. Five PPIT cutoffs were developed (≥0.95, ≥0.85, ≥0.75, ≥0.65, and ≥0.50), such that a threshold of "≥0.85" means that only individuals whose posterior probability was greater than or equal to 0.85 should be considered for sex estimation using this method. Thus, individuals with posterior probabilities less than 0.85 would result in an "indeterminate." Unsurprisingly, accuracy rates increased as the PPIT cutoff increased. Furthermore, accuracy rates were sample-dependent, with the United States White sample generally producing greater accuracy rates (0.89–0.99). Accuracy rates were also sex-dependent, with females generally displaying greater accuracies than males. Importantly, increasing the PPIT mitigated both population and sex differences in accuracies.

The study next examined how sample inclusion was impacted by the use of thresholds, given that high accuracy rates will not be beneficial in practice if the method categorizes the majority of cases as indeterminate. Sample inclusion frequency was calculated for each threshold range by dividing the number of individuals that had a posterior probability that met the threshold criteria by the total number of individuals in the available sample (those that did and did not meet the threshold criteria). Sample inclusion frequencies consistently decrease as the threshold increases. While the ≥0.95 PPIT had the greatest accuracy (0.97), it also had the lowest sample inclusion (0.49), as expected. Sample inclusion and accuracy rates were most closely balanced in the ≥0.65 and ≥0.75 PPITs.

Finally, this study examined whether the difference in accuracy rates among the thresholds differed from one another statistically to help inform a standardized approach to method interpretation of sex. A Chi-squared test and subsequent correspondence analysis indicated that significant differences in accuracies exist among the thresholds, with PPITs ≥0.95 and ≥0.85) clustering together, and away from, the remaining thresholds. PPIT ≥0.50 was not the poorest performing accuracy and clustered with the ≥ 0.75 accuracy rates.

The following approach to PPIT applications to cranial trait scores is recommended: cases with PPs < 0.75 should be considered indeterminate, cases with PPs of 0.75–0.84 should be considered as "possible" estimated sex (e.g., "possible male"), while cases with PPs ≥ 0.85 should be considered as "probable" estimated sex (e.g., probable male). Utilizing PPITs can not only increase method accuracy and provide a means of reporting estimate confidence, but can also provide standardized interpretations of sex estimation results among forensic anthropologists.

Reference(s):

1. Garvin, Heather M., Sabrina B. Sholts, and Laurel A. Mosca. Sexual dimorphism in human cranial trait scores: effects of population, age, and body size. *American journal of physical anthropology* 154, no. 2 (2014): 259-269.
2. Walker, Phillip L. Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* 136, no. 1 (2008): 39-50.

Anthropology, Sex Estimation, Accuracy

*Presenting Author