A14 Long-Term Observer Error, Observer Experience, and the Value of Trait Standardization in Macromorphoscopic Trait Analysis

Kelly R. Kamnikar, MA, Michigan State University, 355 Baker Hall, 655 Auditorium Road, East Lansing, MI 48824; and Joseph T. Hefner, PhD, Michigan State University, Dept of Anthropology, 355 Baker Hall, East Lansing, MI 48824*

After attending this presentation, attendees will understand the implications of observer experience and trait standardization on intra-observer error in the analysis of macromorphoscopic trait scores commonly used in the forensic estimation of ancestry. Attendees will also gain insight into the impact of this type of error in the estimation of ancestry.

This presentation will impact the forensic science community by addressing the role of experience, standardization, and validation in long-term observer error studies and the impact of these factors on the estimation of ancestry using macromorphoscopic trait scores.

As one part of a much larger investigation into the macromorphoscopic traits used in the estimation of ancestry from skeletal remains, a 14-year (2002 to 2016) intra-observer error study was conducted. Motivated by the development of a large macromorphoscopic database, which could potentially utilize data collected in 2002, quantification of the impact on observations caused by observer error, technological improvements in macromorphoscopic trait collection, and observer experience was necessary. To maximize comparisons between the two samples, only ten macromorphoscopic traits were assessed: (1) anterior nasal spine; (2) inferior nasal aperture; (3) interorbital breadth; (4) malar tubercle; (5) nasal aperture width; (6) nasal bone contour; (7) nasal overgrowth; (8) postbregmatic depression; (9) posterior zygomatic tubercle; and, (10) zygomaticomaxillary suture.

Paired data were collected from 185 American Black (n=127) and White (n=58) individuals from the Robert J. Terry Collection. The 2002 sample was collected on paper forms using then-standard texts, references, and line drawings. The 2016 sample was collected using Macromorphoscopic Traits v.1.61 (MMS), a data collection program designed specifically for macromorphoscopic trait analysis. Following data collection, the 2002 and 2016 samples were combined into a single data table for analysis. A traditional (unweighted) Cohen's Kappa is used to quantify disagreement between two assessments, but does not control for the degree of disagreement. Therefore, when the ratings are ordered ($1 < 2 < 3$), as in macromorphoscopic trait expressions, a quadratic weighted Cohen's Kappa is better suited to assess intra-observer error.

There was good agreement between the two observation periods for seven of the ten traits. The frequency of agreement ranged from 74.05% (postbregmatic depression) to 28.11% (zygomaticomaxillary suture). The three underperforming traits are: (1) malar tubercle (confidence interval κ_w = 0.029 – 0.046; levels = 4); (2) zygomaticomaxillary suture (CI κ_w = 0.031 – 0.093; levels = 3); and, (3) posterior zygomatic tubercle (CI κ_w = 0.029 – 0.046; levels = 4). Discrepancy between observations can be attributed to the following explanations. First, these three traits are not commonly used in forensic ancestry estimations and are therefore less familiar. Second, difficulty distinguishing between the various character state manifestations due to poorly worded definitions or inadequate capture of trait variants may occur. Finally, the experience level of the observer is potentially the most interesting factor influencing low observer agreement. Intermediate values for all ten traits demonstrated a lower proportion of agreement between observation periods; the 2016 sample demonstrated a higher proportion of intermediate

Copyright 2017 by the AAFS. Unless stated otherwise, noncommercial *photocopying* of editorial published in this periodical is permitted by AAFS. Permission to reprint, publish, or otherwise reproduce such material in any form other than photocopying must be obtained by AAFS.

scores, possibly indicating extreme trait values are less likely as the experience of the observer increases. As a final point, classification accuracies from a canonical analysis of the principal coordinates using all ten traits and a reduced model using the seven traits with good observer agreement were calculated and assessed for statistically significant differences between the 2002 and 2016 samples. Despite the noted error, all classification accuracies were promising and similarly distributed. The 2002 sample correctly classified 92.2% (seven variables) and 93.5% (ten variables) of the total sample. Similarly, the 2016 sample correctly classified 91.1% (seven variables) and 88.2% (ten variables)/

The results of this research suggest a moderate level of observer error should be expected as an analyst becomes more familiar with a methodology. As new technologies supplant older approaches, there is potential to reduce observer differences. Although observer error was significant for three of the ten traits documented in this study, their influence on classification accuracies was not demonstrably significant.

This project was supported by an award from the National Institute of Justice, Office of Justice Programs, United States Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the authors and do not necessarily reflect the views of the Department of Justice.

Ancestry, Macromorphoscopic Trait Analys, Inter-Observer Error

Copyright 2017 by the AAFS. Unless stated otherwise, noncommercial *photocopying* of editorial published in this periodical is permitted by AAFS. Permission to reprint, publish, or otherwise reproduce such material in any form other than photocopying must be obtained by AAFS.