A123 Histological Age Estimation of the Femur Using Random Forest Regression.

Michael W. Kenyhercz, PhD, Defense POW/MIA Accounting Agency, Joint Base Pearl Harbor-Hickam, HI 96816; Christian Crowder, PhD, Southwestern Institute of Forensic Sciences, Dallas, TX 75207; Victoria M. Dominguez, PhD, Lehman College City University of New York, Bronx, NY 10468*

Learning Overview: After attending this presentation, attendees will be re-introduced to the histological age estimation method developed by Crowder and Dominguez in 2012, and be introduced to the use of random forest regression to generate age estimates.[1] Additionally, an application to perform these regressions will be presented.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by providing a more flexible statistical approach for histological age estimation than has been previously applied, resulting in a higher quality of forensic practice and reducing errors in histological analysis.

Estimating adult age is a challenging aspect of biological profile construction. Histological methods have been touted as a more accurate approach to adult age estimation because they are based on the continuous turnover of primary by secondary cortical bone and therefore follow a more predictable rate of change. However, research demonstrates that past methods succumb to the same issues affecting gross methods. In 2012, Crowder and Dominguez introduced a method that addressed issues such as subjective variable definitions, difficulty reproducing microscopic field sizes, spatial variation of histological structures, and the statistical significance of histological variables.[1] They employed stepwise linear regression to evaluate the following variables: osteon population density (intact and fragmentary), mean osteonal cross-sectional area, and mean anterior cortical width. Three separate regressions, one each for females, males, and a pooled model, were performed. The female model achieved the highest r^2 at 0.682 (SEE=9.64 years); males had the lowest r^2 at 0.510 (SEE=11.95 years); and the pooled model was in-between with an r^2 of 0.587 (SEE=11.19 years). While this method minimized observer error through improved variable definitions and accounted for a large amount of spatial variation for histological features, it failed to improve a common limitation of adult age estimation—the use of linear regression analysis.

The current study will re-evaluate the variables from the Crowder and Dominguez study using random forest regression.[1] Random forest regression is a flexible, non-linear method utilizing decision trees and aggregated estimates. Essentially, a forest of decision trees, of a user-defined size (typically between 100 and 1,000), is constructed in which random variables are tested at each decision node and individual estimates of age are generated; the number of estimates is equal to the number of decision trees.

The sample includes 236 (120 males, 116 females) femur cross-sections of known-age individuals from three histological collections.[2,3] Histomorphometric data were collected using a transmitted light microscope and a firewire camera. The topographic sampling method was modeled after Iwaniec and colleagues and Stout and Paine.[4,5] Using a Merz counting reticule at 200x magnification (field area=0.2304mm^2), the method evaluates alternating fields in ten columns from the periosteal to the endosteal cortex at the anterior femur midshaft. This sampling strategy is designed to capture the remodeling variability within the anterior cross-section. Osteon areas and cortical widths were calculated using imaging software.

The current study utilizes the Random Forest package in R to generate three models (female, male, pooled sex) using 800 decision trees. Each tree utilizes separate, bootstrapped training and testing sets to avoid overfitting and to reduce overall bias. Individual estimates are then aggregated by taking the mean of each estimate as the final predicted age. Each individual estimate is retained and can be directly modeled for a Prediction Interval (PI). Random forest regression models, unlike linear regressions, make no normality assumptions, and, as a consequence, PIs cannot be easily generated. Individual age estimates from the model are summarized using quantiles (2.5 and 97.5) to generate a 95% PI. The r^2 values for the resulting models each exceeded those obtained via linear regression reported by Crowder and Dominguez: female=0.796 (MD=3.85 years), male=0.692 (MD=4.56 years), and general model=0.719 (MD=4.17 years).[1] Similar to the results of Crowder and Dominguez, the most important variable in each model is fragmentary osteon population, followed by intact osteon population, cortical thickness, and, lastly, osteon area.

Random forest regression offers a flexible alternative to traditional methods without suffering from the assumptions of linear models. Random forest regression improves estimates and model fits by all available metrics and can be easily implemented with the newly developed application named keRley. This research will benefit the forensic community by generating more accurate age estimates without violating any statistical assumptions.

Reference(s):

1. Crowder C., Dominguez V.M. A new method for histological age estimation of the femur. *Proceedings of the American Academy of Forensic Sciences*, 64th Annual Scientific Meeting, Atlanta, GA. 2012.
2. Ericksen M.F. Histological estimation of age at death using the anterior cortex of the femur. *Am J Phys Anthropol* 1991;84:171-179.
3. Kerley E.R. The microscopic determination of age in human bone. *Am J Phys Anthropol* 1965; 23:149-164.
4. Iwaniec U.T., Crenshaw T.D., Schoeninger M.J., Stout S.D., Ericksen M.F. Methods for improving the efficiency of estimating total osteon density in the human anterior mid-diaphyseal femur. *Am J Phys Anthropol* 1998;107:13–24.
5. Stout S.D., Paine R.R. Brief communication: Histological age estimation using rib and clavicle. *Am J Phys Anthropol* 1992;87: 111-15.

Forensic Anthropology, Histomorphology, Random Forest Regression

*Presenting Author