11-16-2018

Who Needs Data? I've Got Experience!

Dawnie Wolfe Steadman
University of Tennessee, Knoxville, osteo@utk.edu

Invited Commentary: Who Needs Data? I've Got Experience!

Dawnie Wolfe Steadman[1]*

[1]Department of Anthropology, University of Tennessee, Knoxville, Tennessee, USA.

*Correspondence to: Dawnie Wolfe Steadman, Department of Anthropology, University of Tennessee, 1621 Cumberland Avenue, Strong Hall Room #502A, Knoxville, TN 37996 USA. E-mail: osteo@utk.edu.

Short Title: Who Needs Data? I've Got Experience!

KEY WORDS: FORENSIC ANTHROPOLOGY, QUANTITATIVE METHODS, MEDICO-LEGAL CASEWORK, HUMAN IDENTIFICATION, SKELETAL ANALYSIS.

In 2009, the National Academy of Sciences unleashed a blistering report on the current state of forensic science oversight and practice in the United States. The report skewered the current scientific integrity and standards of a number of forensic subdisciplines, including fingerprints, blood spatter, bite marks and ballistics, and offered specific recommendations to improve forensic science. Anthropology was not specifically mentioned in the report but the same criticisms leveled at the other disciplines, including poor training in quantitative methods, little attention to standards, and the abject lack of errors associated with methods, were obviously relevant to forensic anthropology. To the credit of the discipline, anthropologists have produced scores of articles that focus on quantitative analyses (e.g., Algee-Hewitt 2016; Algee-Hewitt 2017; Hefner et al. 2014; Kooi and Fairfield 2013; Megyesi et al. 2005; Slice and Algee-Hewitt 2015; Stoyanova et al. 2015; Stoyanova et al. 2017), validation studies (e.g., Jooste et al. 2016; Kenyhercz et al. 2017a; Kenyhercz et al. 2017b; Kim 2016; Milner and Boldsen 2012; Savall et al. 2016; Suckling et al. 2016), and cognitive bias in our methods (Nakhaeizadeh et al. 2014a; Nakhaeizadeh et al. 2014b; Nakhaeizadeh et al. 2018). The rise of "computational anthropology," the use of advanced computing to produce models, simulations and predictive modeling to address complex problems in anthropology, offers unlimited opportunities for forensic anthropology. The technological explosion of quantitative computing and "big data" analysis tools has formed an important space in anthropology, including demography, ethnography and past and present migration studies, yet forensic anthropology has remained largely unchanged by these advancements. Despite the warnings of the NAS and recent computational efforts in the field, those who are most likely to practice forensic anthropology continue to rely on traditional techniques and personal experience rather than quantitative approaches. Here I explore some of the historical and structural reasons behind the slow acceptance of new methods that are based on large (and/or simulated) datasets and computational approaches that are highlighted in these special issues and offer some guidance for moving forward.

Forensic anthropology is the application of the principles of skeletal biology to medicolegal questions, such as the identification of unknown remains and interpretation of trauma on the bones. Forensic anthropologists are called into the most challenging forensic contexts in that they work with human remains that are decomposed, fragmentary, burnt, cremated, incomplete and otherwise visibly unidentifiable. Unlike other forensic fields that focus on a single set of techniques and tasks (e.g., fingerprints, DNA, ballistics), forensic anthropologists can be asked to complete four disparate tasks in a single case – search and recovery of human remains, personal identification, trauma analysis, and estimation of the postmortem interval. Other specialized tasks may include facial reproduction techniques to garner leads from the public concerning the identity of unknown remains and isotope analysis of human remains to assess geographic origins and migration history that may also assist with identification. These distinct responsibilities preclude the ability to develop a single set of best practices and leads to a crowded playing field of methods. For example, the task of personal identification begins with assessing the biological profile - estimations of sex, ancestry, age and stature from the skeleton – each of which may consist of multiple methods. While stature estimates rely on regression formulae derived from various sized reference samples, most of the other methods are historically based on qualitative observations of skeletal morphological traits.

Comfort levels with methods developed in the 1980s and 1990s based exclusively on macroscopic morphological age-, sex- or population-based variation (called here “traditional methods”) seem to preclude the inclusion of important new methods that use large sample sizes (including simulated data and machine learning), quantitative data such as 3D imaging techniques, and sophisticated statistical analyses. For instance, transition analysis is a maximum likelihood approach developed for bioarchaeological samples using three skeletal indicators and choice of priors, providing an output of a maximum likelihood estimate and 95% confidence interval for each indicator as well as for the combined evidence (Milner and Boldsen 2012), while Slice and Algee-Hewitt (2015) and Stoyanova et al. (2015; 2017) apply digital methods to quantify the specific patterns on the pubic symphysis to predict age with smaller errors than traditional methods. For ancestry, Fordisc is a computer software that can compare measurements of an unknown skull to those of reference groups of varying sample sizes and provide a number of statistical measures of variable performance and group membership (Ousley and Jantz 2005), while Hefner and Ousley (2014) provide a simple computer tool to analyze ordinal data derived from morphological cranial traits to provide a probability of white or black ancestry. Despite the availability of these and other quantitative methods and tools for the biological profile, forensic anthropologists are reluctant to actually employ them in forensic practice.

While some forensic anthropologists do incorporate validated quantitative methods and approaches into their casework, it is clear that this practice is not widespread. For example, Parsons (2017) examined the forensic anthropologist work product (reports) from three medical examiners offices around the country. She analyzed the accuracy of forensic anthropology conclusions when the identity of the decedent ultimately became known. While she found that the accuracy of age, sex, ancestry and stature was overall good (with stature estimation being the poorest performer of the biological profile), the methods employed were traditional (e.g. Suchey-Brooks pubic symphysis [Brooks and Suchey, 1990], Rhine trait list for ancestry [Rhine 1990]) and, except for Fordisc use for stature and ancestry estimation, none of the methods were computational. In fact, all of the methods employed by these offices were introduced before 2002, as the original Fordisc program was introduced in 1993 (Parsons 2017:195).

Garvin and Passalacqua (2012) conducted a survey among anthropologists to assess the most common aging methods employed in casework and how forensic anthropologists constructed the final age estimates when multiple methods were employed. As found by Parsons (2017), the Suchey-Brooks (Brooks and Suchey,1990) method for the pubic symphysis was the most favored technique, followed by additional traditional macroscopic techniques. Moreover, when given a series of possible answers to the question of how the respondent reconciles the results of multiple methods to create a final age estimate, the most chosen answer (41.9%) was that, “I determine an age range on the basis of my experience, the results, and an overall ‘gestalt’ of the remains” (2012:431). As a free comment, one respondent summed up the general sense of where we are in the field: “I typically use whatever information is available from the utilized method(s)…in order to report my final age range, which is ultimately based on my experiences and confidence with the techniques I used” (Garvin and Passalacqua 2012:431). It is clear that anthropologists are maintaining their hold on methods for which they are most comfortable and that experience and gestalt remains the primary means of decision-making.

So what might explain the disconnect between major advances in data collection, analysis and computation available in forensic anthropology and the methods that practicing forensic anthropologists use? Where is the translation between theory and practice lost? It is likely that historical, structural and pedagogical factors each play a role.

The historical model of physical anthropology is what I call the “see and select” approach, in which an observer examines morphological traits to assess whether they fit into one category or another (e.g., presence of a ventral arc on a pubic bone is a female trait while the absence of that trait is classified as male; round orbits are typical of Asian groups while black individuals have square orbits) and then tallying the traits to arrive at a conclusion. Such observational techniques are easy to teach to young college students and are operationalized by graduate students in casework as guided by their mentors. The evidence of this legacy of "see and select" is prevalent in case reports today. In one transcript, an anthropologist expert argued that ancestry was determined with "great confidence" given that favored morphological traits that the anthropologist "always uses" were consistent with one racial group (no quantitative methods were applied). Based on comparing antemortem-postmortem radiographs, the expert further concluded that, "based on my experience, this is the same individual beyond a reasonable doubt," despite not providing any statistical justification about how common the two features compared occur in any particular reference sample and hence the probability of a match (Steadman et al. 2006).

This confusion between personal experience and scientific rigor is not uncommon in forensic anthropology. An expert forensic anthropologist is more likely to employ methods with which they are most comfortable, which means those with which they have the most experience. Not only is this human nature but fits the historical impetus of forensic anthropology that was based on observations of anatomical variability. Forensic anthropologists are end users of this variation – not caring much for quantitative methods that sort that variation systematically or theories as to why such variation exists. And, in fact, we are quite good at applying these traditional techniques and getting the "right" answers. In Parsons' study (2017), 100% of the cases for which sex was estimated using morphological traits of the pelvis and skull (not employing the regression methods of Walker [2008]) were correct. Similarly, age was correct for

89 % of the cases examined among the three medical examiner offices. So if, as a discipline, we are pretty good at using our current methods, no matter how “traditional” they may seem, why employ methods that require an advanced understanding of statistics?

The most simple answer is provided in this issue - that data is power. Computational anthropology provides the means to mine existing data with machine learning methods but also produce simulated data that can resolve the statistical issue of small reference sample sizes. These techniques provide the power to sort and classify massive volumes of data, create predictive models, calculate realistic errors in a statistically responsible manner, and develop probability values for methods that can be presented in court. Deep machine learning methods will identify biological patterns in morphological data that can be exploited for classification purposes and promote discovery of new areas of inquiry based on fine-grained biological patterning that can only be revealed by computational analyses. Deep learning techniques benefit not only biological sorting for biological profile and facial approximation but the multidimensional data derived from remote sensing techniques (Wescott and colleagues, this issue).

Another factor that may impact the adoption of quantitative methods by forensic anthropologists is the structural limitation of the courts. The forensic anthropologist is an independent expert who provides a service (in the form of scientific opinion) to the medicolegal community and ultimately the courts. Forensic anthropologists provide testimony in criminal and civil trials in state and federal courts and must comply with the rules of evidence admissibility for each court. Minimally, this means the methods or techniques must be generally accepted by the discipline (the *Frye Rule*, 1923). More rigorous rules mandate the methods must be reliable and relevant as determined by the judge, who is guided by determining whether the method has been peer reviewed and tested, has known or potential errors and is generally accepted in the discipline (*Daubert v. Merrill Dow Pharm. 1993*). In legal contexts, precedence is critical so new techniques that had not been previously admitted into court may be legally challenged. For instance, counsel may raise a pre-trial motion (Daubert hearing) to challenge a specific method or conclusion (evidence) whereby the onus is on the expert to explain the method to the judge and demonstrate the scientific basis of the technique and how it satisfies the relevant rules of evidence. One of the distinct advantages of the newer quantitative methods in forensic anthropology, including those in this volume, is that they are based on much larger sample sizes and the associated errors are better statistically defined than those of traditional methods (what I call "Daubert-friendly" methods). However, forensic anthropologists must understand the underlying theory and statistics in order to explain the method to the judge (and potentially later the jury). Many forensic anthropologists lack the statistical training necessary to effectively communicate a technique to the courts, which leads to the third limitation of forensic anthropologists to accept computational methods – training.

One of the most important steps we can take to ensure computational methods are used by forensic practitioners, and hence introduced into court, is to modify our pedagogy to include computational theory and methods in both undergraduate and graduate programs. In fact, cross-training in forensic anthropology and computational anthropology would be ideal. Few practicing forensic anthropologists are also broadly trained in computational anthropology (though some are represented in these two issues), while a quantitative anthropologist may or may not ever take a forensic anthropology course, not to mention become a forensic expert. While it is unnecessary for one expert to become fully proficient in the other field, it is imperative that they share some of the same foundational training such that they can speak a common language and understand the fundamental needs of each field. Moreover, anthropology students could benefit from the theoretical mathematical in other fields, such as nonlinear dynamics in physics that have proven useful for taphonomic work (Dautartas 2018) and could be applicable for aging. Introducing computer modeling and simulations in lab exercises early in forensic training, as is commonly done in paleoanthropology and archaeology courses, would also be helpful.

In addition to training, co-authorship on publications can go a long way towards increasing the possibility that forensic anthropologists will be better consumers of computational anthropological methods. All too often we stay in our own lanes, reading papers only generated by like minds in a limited journal repertoire. Co-authorship is imperative to promote the relevance of the presented mathematical models to non-math specialists and publish in journals accessible to both disciplines. Adding a hypothesis-driven approach is particularly useful to clearly communicate the goals of the model or simulation or to explain that simulated data is not equivalent to "fake data." By partnering with math-averse forensic anthropologists we can reduce the intimidation factor, incorporate these important quantitative methods into the mainstream of forensic practice, and thereby introduce them into the courts.

There is no doubt that forensic anthropologists need computational anthropologists. While our work is on single cases, forensic anthropologists are sitting atop mounds of research data waiting to be mined but few may have the training to do so. Forensic anthropologists are very good at leveraging technology developed in other fields for forensic applications but often fall short in using the data derived from that technology to full effect. Massive data sets may require computational platforms just to clean and organize the data while the proper analyses necessitate advanced statistics or deep learning techniques to categorize the data. For instance, Wescott et al. (this volume) provide an excellent overview of the current methods in remote sensing to detect clandestine graves and decomposed bodies – methods that create massive data sets that require significant computational work to reduce the data to usable algorithms. Given these large data sets, it would be interesting to see if there can be enough known images generated to create predictive models to help statistically prioritize target areas.

At the same time, those trained in large-scale quantitative and "big data" analyses need forensic anthropologists to ensure that the methods employed and interpretations based on intricate quantitative patterns are well-founded in biological theory. It is also up to the forensic anthropologists to provide the pertinent questions that drive computational questions. Some of the most problematic issues in forensic anthropology lie beyond the biological profile, which has received the most attention by computational anthropologists to date. In reality, forensic anthropologists are rarely called to testify on the biological profile and personal identification because DNA is often used to confirm the identification before trial. Instead, forensic anthropologists testify on taphonomy and trauma and, in some cases, how a body was located and recovered. Our current techniques in postmortem interval estimation and trauma interpretation cry out for better predictive modeling and pattern matching probabilities, respectively, and there is a tremendous amount of data that has already been collected in these areas. I encourage forensic and computational anthropologists to work together on these and other relevant problems to ensure that scientific rigor becomes preferenced over personal experience in forensic anthropology.

Received 21 May 2018; revision accepted for publication 30 May 2018.

Literature Cited

Algee-Hewitt, B. F. B. 2016. Population inference from contemporary American craniometrics. *Am. J. Phys. Anthropol.* 160:604–624.

Algee-Hewitt, B. F. B. 2017. Geographic substructure in craniometric estimates of admixture for contemporary American populations. *Am. J. Phys. Anthropol.* 164:260–280.

Brooks, S., and J. M. Suchey. 1990. Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Hum. Evol.* 5:227–238.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Dautartas, A. 2018. A Comparison of human and domestic pig decomposition using multivariate methods. PhD dissertation, University of Tennessee.

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Garvin, H. M., and N. V. Passalacqua. 2012. Current practices by forensic anthropologists in adult skeletal age estimation. *J. Forensic Sci.* 57:427–433.

Hefner, J. T., M. K. Spradley, and B. Anderson. 2014. Ancestry assessment using random forest modeling. *J. Forensic Sci.* 59:583–589.

Hefner, J. T., and S. D. Ousley. 2014. Statistical classification methods for estimating ancestry using morphoscopic traits. *J. Forensic Sci.* 59:883–890.

Jooste, N., E. N. L'Abbé, S. Pretorius et al. 2016. Validation of transition analysis as a method of adult age estimation in a modern South African sample. *Forensic Sci. Int.* 266:580.e1–580.e7.

Kenyhercz, M. W., A. R. Klales, C. W. Rainwater et al. 2017a. The optimized summed scored attributes method for the classification of U.S. blacks and whites: A validation study. *J. Forensic Sci.* 62:174–180.

Kenyhercz, M. W., A. R. Klales, K. E. Stull et al. 2017b. Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales et al. method. *Forensic Sci. Int.* 277:259.e1–259.e8.

Kim, J. 2016. Understanding population-specific age estimation using multivariate cumulative probit regression for Asian skeletal samples. PhD dissertation, University of Tennessee.

Kooi R., and S. Fairgrieve. 2013. SEM and stereomicroscopic analysis of cut marks in fresh and burned bone. *J. Forensic Sci.* 58:452–458.

Megyesi, M. S., S. P. Nawrocki, and N. H. Haskell. 2005. Using accumulated degree-days to estimate the postmortem interval from decomposed human remains. *J. Forensic Sci.* 50:1–9.

Milner, G., and J. Boldsen. 2012. Transition analysis: A validation study with known-age modern American skeletons. *Am. J. Phys. Anthropol.* 148:98–110.

Nakhaeizadeh, S., I. E. Dror, and R. M. Morgan. 2014a. Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Sci. Justice* 54:208–214.

Nakhaeizadeh, S., I. Hanson, and N. Dozzi. 2014b. The power of contextual effects in forensic anthropology: a study of biasability in the visual interpretations of trauma analysis on skeletal remains. *J. Forensic Sci.* 59:1,177–1,183.

Nakhaeizadeh, S., R. M. Morgan, C. Rando et al. 2018. Cascading bias of initial exposure to information at the crime scene to the subsequent evaluation of skeletal remains. *J. Forensic Sci.* 63:403–411.

National Academy of Science. 2009. *Strengthening Forensic Science in the United States: A Path Forward*. Washington, D.C.: National Academy Press.

Ousley, S. D., and R. L. Jantz. 2005. FORDISC 3.0: Personal Computer Forensic Discriminant Functions. University of Tennessee.

Parsons, H. R. 2017. The accuracy of the biological profile in casework: An analysis of forensic anthropology reports in three medical examiners' offices. PhD dissertation, University of Tennessee.

Rhine, S. 1990. Nonmetric skull racing. In *Skeletal Attribution of Race: Methods for Forensic Anthropology*, G. Gill and S. Rhine, eds. Albuquerque: University of New Mexico, 9–20.

Savall, F., C. Rérolle, F. Hérin et al. 2016. Reliability of the Suchey-Brooks method for a French contemporary population. *Forensic Sci. Int.* 266:586.e1–586.e5.

Slice, D. E., and B. F. Algee-Hewitt. 2015. Modeling bone surface morphology: A fully quantitative method for age-at-death estimation using the pubic symphysis. *J. Forensic Sci.* 60:835–843.

Steadman, D. W., B. J. Adams, and L. W. Konigsberg. 2006. Statistical basis for positive identification in forensic anthropology. *Am. J. Phys. Anthropol.* 131:15–26.

Stoyanova, D., B. F. Algee-Hewitt, and D. Slice. 2015. An enhanced computational method for age-at-death estimation based on the pubic symphysis using 3D laser scans and thin plate splines. *Am. J. Phys. Anthropol.* 158:431–440.

Stoyanova, D. K., B. F. Algee-Hewitt, J. Kim et al. 2017. A computational framework for age-at-death estimation from the skeleton: Surface and outline analysis of 3D laser scans of the adult pubic symphysis. *J. Forensic Sci.* 62:1,434–1,444.

Suckling, J. K., M. K. Spradley, and K. Godde. 2016. A longitudinal study on human outdoor decomposition in Central Texas. *J. Forensic Sci.* 61:19–25.

Walker, P. L. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. *Am. J. Phys. Anthropol.* 136:39–50.