
[All ETDs from UAB](#)

[UAB Theses & Dissertations](#)

2023

Accuracy of Ancestry Estimation in Forensic Anthropology: An Examination of Select Nonmetric Methods

Mackenzie Miller
University Of Alabama At Birmingham

Follow this and additional works at: <https://digitalcommons.library.uab.edu/etd-collection>



Part of the [Arts and Humanities Commons](#)

Recommended Citation

Miller, Mackenzie, "Accuracy of Ancestry Estimation in Forensic Anthropology: An Examination of Select Nonmetric Methods" (2023). *All ETDs from UAB*. 79.
<https://digitalcommons.library.uab.edu/etd-collection/79>

This content has been accepted for inclusion by an authorized administrator of the UAB Digital Commons, and is provided as a free open access item. All inquiries regarding this item or the UAB Digital Commons should be directed to the [UAB Libraries Office of Scholarly Communication](#).

ACCURACY OF ANCESTRY ESTIMATION IN FORENSIC ANTHROPOLOGY: AN
EXAMINATION OF SELECT NONMETRIC METHODS

by

MACKENZIE MILLER

STEPHEN MERRITT, COMMITTEE CHAIR
LAUREN DOWNS
CHRIS KYLE

A THESIS

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Master of Arts

BIRMINGHAM, ALABAMA

2023

Copyright by
Mackenzie Miller
2023

ACCURACY OF ANCESTRY ESTIMATION IN FORENSIC ANTHROPOLOGY: AN
EXAMINATION OF SELECT NONMETRIC METHODS

MACKENZIE MILLER

ANTHROPOLOGY OF PEACE AND HUMAN RIGHTS

ABSTRACT

A goal of forensic anthropology is to aid in identifying unidentified human remains, often by creating a biological profile. An important piece of this profile is an ancestry estimate, and methodological accuracy is taken seriously. Traditional typological nonmetric methods of ancestry are heavily critiqued for being subjective and lacking scientific rigor in analysis. In addition, these methods are noted for conflating morphological traits with specific ancestral categories, ignoring the reality of continuous human variation. To combat this, scholars have proposed morphoscopic nonmetric methods as an alternative because they retain the accessibility popularized by typological methods but add a layer of scientific validity through a statistical analysis framework for determining ancestry. This study tests the hypothesis that morphoscopic methods are superior by analyzing the accuracy of two typological and two morphoscopic methods as reported in forensic case entries in the Forensic Anthropology Database for Assessing Methods Accuracy (FADAMA) ($n=241$). Data analysis looked at the four methods individually and the two types of methods collectively. Congruent with the literature, morphoscopic methods outperformed typological methods in methodological accuracy for both categories of analysis. Additionally, morphoscopic methods outperformed their expected research-based accuracy rate, while typological methods underperformed their expected accuracy rate. The results of this study speak to the superiority of objective nonmetric methods of

ancestry estimation that are rooted in statistical analysis of human variation. More broadly, this study lends itself to the discussion of ancestry estimation and its relationship with the concept of race. Some scholars postulate that ancestry estimation enforces the idea of biological race and invokes the racist past history of the discipline. However, the results of this study do not support the validity of biological race. Unlike typological methods, morphoscopic methods recognize that all morphological traits are present in all ancestral categories at different frequencies. Frequency distributions of traits correspond to social race categories used in the United States due to the lack of gene flow between “racial” groups historically, creating a concordance between social race and skeletal morphology. As a result, ancestry estimation remains a valuable part of American forensic anthropology.

Keywords: forensic anthropology, ancestry, ancestry estimation, nonmetric methods, race

DEDICATION

To Mom and Dad – endless thanks for your constant love, support, and willingness to watch *Bones* over and over with me. I wouldn't have gotten here without either of you.

ACKNOWLEDGMENTS

I would like to thank Dr. Stephen Merritt, Dr. Lauren Downs, and Dr. Chris Kyle for their support during this entire process. Thank you for your advice and your help brainstorming through all the twists and turns that it took to get here. This thesis would not have become the work that it is without all of you.

I am also extremely thankful for my amazing fellow cohort members. Audrey, Leslie, and Nick, thank you for both your support academically and your friendships personally. These last two years have been incredibly memorable and special because of you three.

Lastly, special thanks to Dr. Lisa Gezon and our seminar group. I appreciate you all coming with open minds and endless patience and letting me workshop (many) versions of this project with you.

TABLE OF CONTENTS

	<i>Page</i>
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGMENTS	vi
LIST OF TABLES	ix
CHAPTER	
1 INTRODUCTION	1
2 LITERATURE REVIEW	4
Development of Forensic Anthropology	4
History of Ancestry Estimation	7
Types of Ancestry Estimation Methods	9
Typological Methods	13
Morphoscopic Methods	16
Methods Analyzed in This Study	17
Generalized Morphology	18
Rhine (1990)	18
Hefner (2009)	19
Hefner and Ousley (2014)	20
3 METHODS	22
Overview	22
The Forensic Anthropology Database for Assessing Methods Accuracy	23
Accessing and Downloading FADAMA	24
Research Criteria	24
Coding the Data	25
Data Analysis	29
4 RESULTS	32

Descriptive Analysis.....	32
Inferential Analysis	35
Chi-square Tests	35
Wilcoxon Rank Sum Test.....	36
Two-sample <i>t</i> -test	37
5 DISCUSSION.....	38
6 CONCLUSION	48
LIST OF REFERENCES	50

LIST OF TABLES

<i>Tables</i>	<i>Page</i>
1	Examples of Case Entries with a Level 1 Accuracy Rate27
2	Examples of Case Entries with a Level 2 Accuracy Rate28
3	Examples of Case Entry with a Level 3 Accuracy Rate.....28
4	Examples of Case Entry with a Level 4 Accuracy Rate.....28
5	Examples of Case Entry with a Level 5 Accuracy Rate.....28
6	Frequency Distribution for Generalized Morphology Method33
7	Frequency Distribution for Hefner (2009) Method33
8	Frequency Distribution for Hefner and Ousley (2014) Method.....33
9	Frequency Distribution for Rhine (1990) Method.....34
10	Frequency Distribution for Typological Methods34
11	Frequency Distribution for Morphoscopic Methods35

CHAPTER ONE

INTRODUCTION

This research aims to analyze the accuracy rates of four nonmetric ancestry estimation methods used in American forensic anthropology by examining forensic case entries found in the Forensic Anthropology Database for Assessing Methods Accuracy (FADAMA). Nonmetric methods and metric methods of ancestry estimation are tools that forensic anthropologists use to assign a cranium or a set of human skeletal remains to a specific ancestral group. Most commonly, the decision is between three main ancestral groups: White, Black, and Asian. This study focuses on nonmetric methods because of their popularity amongst practicing forensic anthropologists due to their ease of use. Unlike metric methods of ancestry estimation, nonmetric methods require little formal equipment (such as various sets of calipers) and can be assessed visually with ease. As will be discussed within the literature review, two categories of nonmetric methods have arisen: typological and morphoscopic. While typological methods are the original nonmetric methods, they are known for being less reliable and having low accuracy rates and high levels of subjectivity during analysis. Morphoscopic methods arose as the field became discontent with the shortcomings of typological approaches. Presented as alternatives, morphoscopic methods still feature the same accessibility and ease of

analysis as typological methods but with an added layer of scientific rigor through the statistical analysis framework used to interpret results.

The four methods analyzed in this study are generalized morphology, Hefner (2009), Hefner and Ousley (2014), and Rhine (1990). These methods were selected because they are well-represented in the forensic anthropological literature and within FADAMA. For this study, generalized morphology and Rhine (1990) represent the typological methods sample, and Hefner (2009) and Hefner and Ousley (2014) are the morphoscopic methods sample. The working hypothesis is that morphoscopic methods will outperform typological methods, as indicated by a higher percentage of accurate ancestry estimations. The literature supports this hypothesis, as will be presented in detail below. In summary, morphoscopic methods are expected to outperform because of the added layer of statistical analysis and because of the focus on trait frequency distribution and degree of expression rather than on trait presence or absence. Conversely, typological methods tend to focus on the dichotomy of trait presence/absence. This leads to a conflation of certain traits with specific ancestral categories, to the point where traits are “expected” to only appear in individuals of a singular ancestral category.

The paper is organized as follows: a literature review, a discussion of the study’s methodology, a presentation of results, and a concluding discussion section. The literature review offers a brief history of the field of forensic anthropology and the creation of ancestry estimation methods before turning specifically to nonmetric methods. Sections on both typological and morphoscopic methods will outline their creation, basic protocol, and the field’s opinions of them. Finally, the literature review concludes with a discussion

of the four methods examined in this study. The methodology and results sections outline the specific protocol used in this study and the results found using it. Lastly, the discussion section outlines how the results of this study fit into the broader debate in forensic anthropology surrounding the use of ancestry estimation and its relation to the concept of distinct human races. Given the overarching anthropological field's view that distinct human races do not exist, many scholars wonder what an ancestry estimation measures. Some believe the estimation makes determinations about biological race. However, an ancestry estimation speaks to an individual's social race—the racial category that likely described them in life (Sauer, 1992). Ancestry estimation is possible in American forensic anthropology because of the concordance between social race categories and distinct skeletal morphologies of the American population. The forensic anthropological literature and the results of this study support this concordance, both of which report a high level of accuracy when describing social race via ancestry estimations.

CHAPTER TWO

LITERATURE REVIEW

This literature review will discuss the development of forensic anthropology as a recognized discipline and the creation of its methodology, especially as it relates to methods of ancestry estimation, to understand why the debate surrounding ancestry estimation and race is so salient to the field. From there, the topic will turn to the two main types of ancestry estimation methods: metric and nonmetric. Next, the section on nonmetric methods will focus on the creation of typological methods and subsequent morphoscopic methods. Finally, there will be a review of the four methods analyzed in this study.

Development of Forensic Anthropology

Forensic anthropology arose as a sub-discipline of biological anthropology to apply the knowledge and methodology of the field to medico-legal issues concerning the recovery and analysis of human remains (Ubelaker, 2018). While the overall methodological toolkit is a compilation of skills from several fields, including biological anthropology and archaeology, the discipline's unique emphasis on techniques of human skeletal analysis makes it indispensable to the medico-legal system. The typical goal is to discover the person's identity when modern human remains are found. The identification

process is two-stepped, involving the creation of a biological profile by the forensic anthropologist and an attempt by the medico-legal system to match that profile for a positive identification (Sauer, 1992). The biological profile consists of four criteria that can be derived from the human remains using forensic anthropological methods: biological sex, age at time of death, stature, and ancestry. Law enforcement professionals can use this profile to generate a list of missing persons. More specific identifying markers, such as dental records and X-rays, can be compared to make a positive identification.

Analysis of human remains dates back to the eighteenth century and the publication of several European works outlining how to measure cadaver body proportions. Although these volumes were published initially to aid and educate artists in drawing accurate body proportions, interest in the work spread and led to dedicated scientific research in body proportions and stature calculation. This work continued into the nineteenth century and expanded beyond just stature as the field of physical anthropology began to take shape in Europe (Ubelaker, 2006). Prior to the development of forensic anthropology as a distinct discipline in North America, forensic analysis of remains was the job of anatomists and medical specialists interested in applied work, such as Jeffries Wyman, an anatomy professor at Harvard and the first curator of the Peabody Museum of American Archaeology and Ethnology. Wyman became a consultant on a high-profile case involving the murder of someone in the Harvard community by a faculty member in 1866 (Ubelaker, 2006). While Wyman among the first in the United States to provide forensic assistance to law enforcement, Harvard anatomy professor

Thomas Dwight is often credited as “the Father of American Forensic Anthropology” (Stewart, 1979). After an 1878 prize-winning essay on the medico-legal identification of the human skeleton, Dwight published several articles on issues of sex estimation, age at time of death, and stature estimation (Dwight 1881, 1890a, 1890b, 1894a, 1894b).

Other important figures in American forensic anthropology include George A. Dorsey, the first anthropologically trained professional to become involved in forensic cases, and Aleš Hrdlička, the first curator of the physical anthropology division at the Smithsonian Institution in Washington, DC. While best known for founding the American Association of Physical Anthropology and its journal, Hrdlička also started the tradition of forensic consultation of the Smithsonian by the FBI that continues into the present day (Ubelaker, 2006). Key publications by Hrdlička and Wilton Krogman in 1939 marked the beginning of the professionalization of American forensic anthropology. The need to identify recovered human remains from World War II and other military conflicts further spurred the discipline’s professionalization. In 1972, the American Academy of Forensic Sciences (AAFS) formed a physical anthropology section, which led to the first formal gatherings of forensic anthropologists to report on research and casework. A few years later, in 1977, the American Board of Forensic Anthropology (ABFA) was formed to develop standards of expertise in the field (Ubelaker, 2006).

As the field began to grow and standardize, interest in the discipline also increased with a growing number of forensic anthropology students. Training in the field involves learning methods for determining each of the four criteria in the biological profile. Numerous methods for each criterion exist, and several volumes exist that

compile them into single texts (e.g., Buikstra & Ubelaker, 1994; Krogman & İşcan, 1986). The choice of which method to employ depends on several factors, and some methods depend on knowing other aspects of the biological profile. For example, some methods of estimating stature are contingent upon the individual's biological sex, and some methods for sex determination are contingent upon knowledge of the individual's estimated ancestry. In other cases, some methods are population-specific and have the highest accuracy rate when applied only to individuals belonging to that population (Spradley, 2016). While this is admittedly complex, it is standard practice for the forensic anthropologist to be aware of the history and limitations of selected methods to select the most relevant and accurate methods on a case-by-case basis (Cunha & Ubelaker, 2019). Correct methodological selection is crucial when estimating ancestry, as many scholars claim that ancestry estimation is one of the most critical parts of the biological profile. Furthermore, it is fundamental to creating the rest of the biological profile, as some methods of sex and stature estimation depend on an ancestry estimate for successful application (Cuhna & Ubelaker, 2019; Klales & Kenyhercz, 2014; Thomas, Parks, & Richard, 2017).

History of Ancestry Estimation

The term “ancestry” describes an individual's geographic region or ancestral origin (Dunn et al., 2020). Historically, ancestry estimations were known as “race estimations” within forensic anthropology. However, the terminology shifted in the late twentieth century to reflect the sentiment of the broader field of anthropology that human

racess do not exist. Nevertheless, it is essential to note that methods for ancestry estimation came about as part of race-driven science in the nineteenth century. Physical anthropologists at the time were interested in the size and shape variation of different human crania. They used these differences to justify the idea of different human “race” categories. In North America, Samuel Morton began to define homologs and identifiable points on the cranium, known as cranial landmarks, that could be studied and measured to note differences in cranial size and shape. The burgeoning field of craniometrics used this data to erroneously correlate cranial morphology and mental abilities to rank various human groups in a race-based hierarchy that benefited white individuals (Dunn et al., 2020). Despite this false correlation between cranial size and mental ability, metric analyses of human crania continued well into the twentieth century and were the grounds for investigating human variation (Albanese & Saunders, 2006).

Post-race-based science, the field of biological anthropology took on craniometric research as a way of investigating biological distance (biodistance), which uses craniometric data as a measure of biological relatedness to provide an understanding of population structure. Gradually, craniometric analysis expanded to a forensic setting. It is used based on the theory that the craniofacial complex has moderate-to-high heritability and is useful to infer ancestral group membership (Dunn et al., 2020). Research to support this theory became more mainstream with the development of skeletal collection in the United States. The two largest collections, the Terry Collection and the Hamann-Todd Collection, were assembled with an emphasis on collecting individuals’ demographic information—including race—to assess human skeletal variation. Giles and

Elliot (1962) published a linear discriminant function method for determining race using the Terry Collection, which became the original standard classification statistic. However, it is worth noting that subsequent verification studies of their method found it to have a low classification accuracy rate, ranging from 52% to 76.4%, well below today's standards for accuracy (Albanese & Saunders, 2006). Nevertheless, this opened the door to metric analyses of the human cranium for race/ancestry estimation.

Nonmetric methods of ancestry estimation that involve assessing cranial features' size and shape also have their own race-based origins and can be traced back to Harvard-based E.A. Hooton in the early twentieth century. Early in his career, Hooton looked for metric and nonmetric traits that could potentially define races. Going further, he began to speculate that certain behaviors, along with biological traits, could be ascribed to different human races, particularly concerning criminality. Much of this culminated in his 1939 work, *The American Criminal: An Anthropological Study*. Hooton maintained that there are discernible morphological differences between criminals and civilians and noted frequency correlations between race and criminal acts (Hefner, Ousley, & Dirkmaat, 2012).

Types of Ancestry Estimation Methods

Methods of ancestry estimation exist in two types: metric and morphological (nonmetric) methods. Metric methods are more traditional and are considered more objective since cranial measurements are well-defined based on cranial landmarks (Cuhna & Ubelaker, 2019). As noted above, these methods rely on the theory of

moderate-to-high heritability of the craniofacial complex. Therefore, ancestry can be estimated from an examination of cranial measurements. These measurements are known as interlandmark distances (ILDs). *Data Collection Procedures for Forensic Skeletal Material 2.0* (Langley et al., 2016) and *Standards for Data Collection from Human Skeletal Remains* (Buikstra & Ubelaker, 1994) outline the most common ILDs used. The twenty-four most common measurements can capture the overall craniofacial complex to estimate ancestry (Dunn et al., 2020). Statistical analysis of the ILDs translates these measurements into an ancestry estimation. Also, as noted above, Giles and Elliot (1962) pioneered using linear discriminant function analysis (LDFA) for ancestry estimation, and it became the standard classification statistic. However, their poor accuracy rates left the field looking for alternatives. In 1999, Ross et al. introduced geometric morphometrics (GM) as a method of forensic anthropological identification, which incorporates ILD data in two- and three-dimensional space for analysis (Ross, McKeown, & Konigsberg, 1999). While LDFA is still the most commonly used classification statistic, GM methods are gaining popularity. In addition, methods using artificial intelligence and machine learning are also on the rise. These methods routinely outperform traditional classification statistics and are less susceptible to outliers and missing data than the older methods (Dunn et al., 2020). However, regardless of old versus newer methods, all metric methods are considered superior to morphological methods due to the objectivity of cranial landmarks and subsequent statistical analysis.

While metric methods are concerned with analyzing standardized cranial landmarks, morphological methods developed as physical anthropologists became

interested in identifying cranial skeletal anomalies (Hefner, 2009). Early research focused on these traits as evidence of individual variation. However, not until the discovery of DNA and the genomic revolution did nonmetric traits begin to be quantified rather than just described. This new quantification process allowed nonmetric traits to serve as proxies for genetic data that could measure relatedness and locality and identify family groups (Dunn et al., 2020). Two categories of nonmetric traits that could be analyzed exist: discrete/nonmetric and morphoscopic/macromorphoscopic/anthroposcopic. Discrete traits are the “typical” nonmetric traits, defined as “dichotomous, discontinuous, epigenetic traits; non-pathological variations of skeletal tissues that can be better classified as present or absent (or as a point on a morphological gradient) rather than quantified by a measurement” (Buikstra & Ubelaker, 1994). Five significant categories of discrete traits are observable on the cranium: (i) extra-sutural bones; (ii) proliferative ossifications; (iii) ossification failure; (iv) suture variation; and (v) foramina variation (Hefner et al., 2012).

Early work on discrete traits attempted to classify certain traits as characteristics of specific groups, resulting in such cranial features as “Inca bones” and “os japonicum.” Several studies have sought to determine whether a suite of discrete traits could separate world population groups. Ossenberg (1976) used twenty-four cranial traits to compare “Native American Indians,” “Eskimos,” “African Americans,” and “African Blacks.” The calculated distance statistics were higher between the groups than those within them. This conclusion was tested further by Wijnsman and Neves (1986), who examined trait frequencies between “Brazilian Blacks,” “Whites,” and “mulattos.” Their results found

significant deviations in the pattern of discrete trait distances from a linear model of genetic differences, leading them to conclude that many discrete traits have low estimates of heritable variation. More recently, Hanihara and colleagues (2003) completed a study of the frequencies of twenty traits in several thousand individuals across global populations. They found variation in trait frequency to be in part due to geographical factors (rather than environmental). They also found distances in trait patterns to be similar to those in results from studies using craniometric data. The conclusion is that there is some evidence for discrete trait clusters in regional global populations. However, this cluster variation is a minimal part of overall worldwide variation, with the most variation being found within local group populations (Hanihara, Ishisa, & Dodo 2003).

The second trait category is macromorphoscopic/morphoscopic/anthroposcopic traits. Ousley and Hefner (2005) noted a difference between the discrete traits used by biological anthropologists for biological distance studies and those used by forensic anthropologists to estimate ancestry. They termed the forensic anthropological traits “macromorphoscopic traits,” defined as “quasi-continuous variables of the cranium that can be reflected as soft-tissue differences in the living” (Hefner et al., 2012, p. 295). Hefner (2009) later shortened the term to “morphoscopic traits” while retaining the exact definition. The trait category subdivides into five classes: (i) assessing bone shape, (ii) bone feature morphology; (iii) suture shape; (iv) feature presence or absence; and (v) feature degree of expression. Many traits noted by Hefner (2009) are a nod back to Hooton, who developed a standardized list of morphological traits known as “the Harvard list” that was useful for classifying human variation (Hefner et al., 2012). The study of

cranial morphological traits was not new, but the collection of traits studied and the definitions of their expressions were determined at the institutional level. Hooton sought to standardize the trait list studied and added extensive descriptions and figures that explained how to identify and score each trait (Dunn et al., 2020). The main goal of Hooton's list was to limit observer subjectivity of trait expression and decrease the likelihood of inter-observer error. While Hooton did not create the Harvard list with the intention of using the traits in a forensic context to determine ancestry, the field adopted the traits. Subsequently, the trait list method of ancestry estimation was developed and is still used by forensic anthropologists and taught to students of the discipline (e.g., Rhine, 1990). Even with standardization efforts to combat subjectivity, nonmetric methods are still considered less precise than metric methods. However, nonmetric methods are still popular today due to the lack of specialized equipment for analysis and the quick visual assessment speed (Albanese & Saunders, 2006). There are two methodological types of nonmetric ancestry estimation: typological and morphoscopic. The next section will discuss typological methods, while the following section will focus on morphoscopic methods.

Typological Methods

As stated previously, the trait list method of ancestry estimation remains one of the most popular nonmetric methods applied today and is the most commonly used typological method. Credit is attributed to Rhine (1990) for beginning the practice. For his method, he developed a list of forty-five different nonmetric traits that correspond

with three ancestry categories (“White,” “Black,” and “Amerindian”). Many of the traits Rhine (1990) identifies are scored dichotomously as present or absent and were used to generate “expected” trait lists for each ancestry category. However, many of the traits listed are expressed in life in a quasi-continuous manner, meaning that a discrete presence/absence differentiation strongly skews results as it does not capture degrees of expression. Nevertheless, lists of expected trait values can be found in many forensic anthropology textbooks (e.g., Burns, 2012; Byers, 2002), and several other typological methods have been developed (Bass, 2005; Buikstra & Ubelaker, 1994). For instance, many textbooks list a postbregmatic depression (PBD) as an expected trait for American Blacks, meaning it should be present in nearly all American Blacks. However, in Rhine’s (1990) study, only 33% of his American Black sample displays a PBD. Of the forty-five traits reported by Rhine (1990), nineteen did not match the expected race value but were still included (Hefner et al., 2012). To somewhat combat this discrepancy, Rhine calls the estimation of ancestry “as much art as science” and emphasizes the need for experience to estimate a skull’s ancestry accurately (1990, p. 18).

The proliferation of the typological methods, along with Rhine’s (1990) claims, has caused a shift in how to understand ancestry using nonmetric methods. Traits have become automatic race determiners: PBD indicates Black; Carabelli’s cusp means White; shovel-shaped incisors are Asian. Ambiguous or discordant trait values lead to assumptions of admixture or individual variation. However, it is wrong to consider traits in such a simple manner. Hefner and Ousley (2006) found that the percentage of individuals who display all of the expected trait values based on race ranged from only

17% to 58% of the population. This discrepancy results in an odd paradox where forensic anthropologists can claim high accuracy rates of ancestry estimation using these traits while the trait frequencies are much lower than assumed. In other words, more forensic anthropologists cite “expected” traits as determiners of an ancestry estimation than exist individuals who display those “expected” traits. This conundrum begs the question of how forensic anthropologists can accurately predict ancestry using a typological approach. Hefner and Ousley (2006) suggest that forensic anthropologists assess ancestry using the cranial gestalt and, after a known positive identification, choose traits *post hoc* that support this evaluation. Thus, traits appear as objective and valid indicators of ancestry, while ancestry estimation using a trait list approach is inherently subjective (Hefner, 2009). Berg and Tersigni-Tarrant (2014) performed a study to assess the accuracy of gestalt ancestry estimations. They found that the participant forensic anthropologists made decisions within seconds of viewing a skull. When given more time, research participants went back to re-examine traits and were able to change their original answers. Often, this resulted in participants changing correct answers to incorrect ones, lowering the overall accuracy of ancestry estimations from 75% (gestalt) to 60% (*post hoc*). Berg and Tersigni-Tarrant (2014) also found that the accuracy of ancestry estimations using the cranial gestalt and the anthropologist’s experience level are linked, with participants with the most experience scoring the highest. This finding echoes the claim made by Rhine (1990), which the discipline still believes, that nonmetric estimations of ancestry are more art than science and only field experience can hone

accuracy. This belief makes the method seemingly intuitive, untestable, unempirical, and unscientific (Hefner, 2009).

Morphoscopic Methods

The use of typological methods of ancestry estimation becomes specifically more problematic in light of the Daubert ruling that requires expert witnesses in court to use methods that are empirically supported, have estimated error rates, have method standardization, and are validated through peer review (Christensen, 2004). Hefner and colleagues (2012) note that rather than selecting expected trait expressions to fit an individual into an ancestral category, focusing on individual traits and the variable expression of those traits within a given population is more relevant. By studying the frequency of expression of specific traits, they can be combined into suites of significant traits and analyzed within a statistical framework to generate scientifically valid estimates of ancestry. Much of this work was undertaken by Hefner (2003, 2007, 2009). Following the example of Hooton, Hefner developed a set of character states for eleven traits, including written descriptions and visual illustrations of each character state. He then scored a collection of crania using this trait list and completed statistical analysis to show the frequencies of each trait in four ancestral groups (“African,” “American-Indian,” “Asian,” and “European”) that highlight intergroup variation. Results showed great promise for ancestry estimations based on morphoscopic traits in a statistical framework, with accuracy rates similar to those obtained using metric data (Hefner, 2009, p. 993). This work confirmed that a combination of traits is better than individual ones and that

some suites of traits work better than others at predicting ancestry. It is important to note that all character states for each trait were present in each ancestral group, only at different frequencies. Unlike with a typological method, this does not indicate admixture or individual idiosyncrasy. Instead, this highlights the reality of morphoscopic traits and human variation that exists with more nuance than can be captured by a present/absence dichotomy.

After establishing trait frequencies, it becomes possible to use various statistical methods to estimate ancestry from an unknown skull. The first step is to select character states for each trait present on the skull. The second is to select an appropriate statistic (e.g., *k*-nearest neighbor, discriminant function analysis, etc.) and appropriate reference groups (e.g., American Black, American White, etc.). Similar to metric methods, computer software is available to assist with analysis. Following analysis, the probability of ancestral group membership for the unknown skull and the error rates can be reported, along with the assigned ancestral group (Hefner et al., 2012). The results of this are ancestry estimation methods that are based on cranial morphology rather than cranial landmarks but are also statistically significant, differentiating them from typological methods.

Methods Analyzed in This Study

This study looks at the accuracy rates of methods from the two main categories of nonmetric methods: typological and morphoscopic. What follows are brief descriptions of each method.

General Morphology

The general morphology method is a typological method in which the forensic anthropologist assesses the cranial gestalt. There is no standardized list of traits to examine, but this method follows in the line of Rhine's (1990) work that correlates the most common "expected" traits to specific ancestral categories. It can also include the evaluation of other common "expected" traits listed in various typological methods of ancestry estimation (i.e., Bass, 2005; Buikstra & Ubelaker, 1994; Burns, 2012). After assessing the traits presented on the cranium, the number of traits associated with each ancestral category is tallied. The ancestry estimate is the category that has the most traits recorded. In cases where two or more categories have the same number of traits present, the forensic anthropologist typically either chooses the "minority" ancestral category as the ancestry estimate (i.e., any non-White category), notes the presence of ancestral admixture in the estimate (recording both categories), or leaves the estimate as unknown/indeterminant.

Rhine (1990)

Rhine's (1990) trait list method of ancestry estimation is discussed in detail above for its role in developing typological methods of ancestry estimation. In his study, Rhine (1990) examined how forty-five morphological traits presented across a sample of White, Black, and Native American (serving as a proxy for Asian) crania. Frequency distributions for each trait were calculated and used to generate "expected" trait lists for each ancestral category. While Rhine (1990) notes the potential for variability in trait

expression, he attributed it to individual idiosyncrasy and advises using expected trait lists for ancestry estimation. Attributing an ancestry estimation to a cranium using this method follows the same procedure used for the generalized morphology method, which tallies traits corresponding to each ancestral category, and the category with the most is considered the ancestry estimate. However, again, this method strongly conflates “expected” traits with only specific ancestral categories. For example, this means that the presence of a postbregmatic depression strongly indicates Black ancestry, and the presence of shovel-shaped incisors strongly indicate Asian ancestry using this method.

Hefner (2009)

Hefner (2009) developed an ancestry estimation method that involves analyzing eleven morphoscopic traits to determine their frequency distributions across four sample ancestral groups: “African,” “American Indian,” “Asian,” and “European.” The eleven morphoscopic traits are the anterior nasal spine (ANS), inferior nasal aperture (INS), inter-orbital breadth (IOB), malar tubercle (MT), nasal aperture width (NAW), nasal bone contour (NBC), nasal overgrowth (NO), postbregmatic depression (PBD), supra-nasal suture (SPS), transverse palatine suture (TPS), and zygomaticomaxillary suture (ZS). This method scores all traits along a continuum of either three or five states of character expression, except for NO and PBD, which are scored dichotomously as present or absent. Population-level reference data can be found online in the Macromorphoscopic Databank (MaMD) created by Hefner (2018). Using nonparametric statistical analysis is

recommended to compare the un-assigned cranium with the reference sample to make an ancestry estimate.

Hefner and Ousley (2014)

Hefner and Ousley (2014) built upon Hefner's (2009) morphoscopic method to develop the optimized summed scored attributes (OSSA) method of ancestry estimation. The OSSA method differentiates between American Blacks and Whites only, which is a noted flaw of the method, since there is no way to confirm that an unknown cranium will absolutely fit into either of those ancestral categories (Albanese & Saunders, 2006). Rather than use eleven traits, the OSSA method focuses on six of Hefner's (2009) traits: ANS, INA, INS, NAW, NBC, and PBD. The scoring system for the traits (excluding PBD) corresponds to the same character expressions as used in Hefner (2009). However, the scores are dichotomized as a 0 or 1 to maximize the between-group differences between the two populations. Score dichotomization is done heuristically by analyzing reference sample data, such as that found in MaMD. Generally, scores more common in Blacks are compressed to a 0, and scores more common in Whites are compressed to a 1. After compressing all six trait scores into a 0 or 1, they are summed to achieve the cranium's OSSA score. Hefner and Ousley (2014) determined the sectioning point of a score of 3 or below to indicate a Black cranium and a score of 4 or above to indicate a White cranium. The authors note several shortcomings with the OSSA method, including that the method does not take multivariate relationships among the traits into account and that all six traits must be present for the analysis to be successful (Hefner & Ousley,

2014, pp. 887–8). Importantly, the choice of sectioning point also plays a role in the final classification. While the method’s original sectioning point is 3, a study by Kenyhercz and colleagues found that the method performed more accurately when the sectioning point shifts to 4, where 4 and below corresponded to a Black cranium and a score of 5 or above corresponded to a White cranium (Kenyhercz, Klales, Rainwater, & Fredette, 2016). Again, this change is not reflected in the original study’s methodology. Forensic anthropologists who use the OSSA method distinguish whether it is the original Hefner and Ousley (2014) method or the Kenyhercz et al. (2016) method with the updated sectioning point.

CHAPTER THREE

METHODS

Overview

The purpose of this study is to review the accuracy rates of select nonmetric methods of ancestry estimation with the working hypothesis that morphoscopic methods have higher levels of accuracy than typological methods. Two methods from each category were selected for analysis to test this hypothesis. The generalized morphology method and the Rhine (1990) method are the typological sample, and the Hefner (2009) method and the Hefner and Ousley (2014) method are the morphoscopic sample. All data in the study comes from identified forensic cases entries in the Forensic Anthropology Database for Assessing Methods Accuracy (FADAMA; <https://www-app.igb.illinois.edu/sofadb/faq.php>). In some ways, an accuracy rate assessment derived from known forensic data is more valuable than research-based accuracy validation because skeletal collections do not always represent the demographics of a modern population represented in forensic casework (Thomas et al., 2017). Methods are not universally applicable, and a method derived using a specific skeletal collection may not perform as well as expected on an individual outside of that population. As a result, a single method can have two accuracy rates, depending on the context in which it is studied and applied, i.e., on a historic skeletal collection versus on a modern population represented in forensic

casework. Beginning with the study conducted by Thomas and colleagues (2017) on ancestry methods accuracy using Federal Bureau of Investigation (FBI) solved case files, subsequent studies have also used comprehensive forensic case databases to assess accuracy rates of the whole biological profile (Hughes, Juarez, & Yim, 2021) and population affinity determinations (Winburn & Algee-Hewitt, 2021).

The Forensic Anthropology Database for Assessing Methods Accuracy

Hosted by members of the Society of Forensic Anthropologists (SOFA), FADAMA (Hughes & Juarez, 2018) is an online repository for data from identified skeletal material in a forensic setting. The development of the database was motivated by the need to address inadequate measures of assessing forensic anthropological methodology when applied to casework (Juarez et al., 2020). Access to the database is free but requires potential users to submit proof of membership in a nationally or internationally recognized forensic organization or be sponsored by an organization member. Admitted users can access to the two FADAMA functions: uploading forensic case data to be added to the database and accessing database files for research purposes. To date, over eighty-five methods of sex, ancestry, age at time of death, and stature estimation are represented in FADAMA. Users can download the entirety of the database (presently over six hundred case files) as a .csv file or filter for desired criteria before downloading. Potential search filters include identification information, forensic anthropology report information, and forensic anthropological methods used. FADAMA does not store any identifying information about the practicing forensic anthropologists

or the decedents in any case. Uploaded cases have case numbers, but these are FADAMA-specific numbers that do not correspond to external case reports.

Accessing and Downloading FADAMA

In January 2023, an application was submitted to the University of Alabama, Birmingham's (UAB) Internal Review Board (IRB) outlining the project. It was reviewed and received an exempt classification, deemed "not human subject research." After this exemption was received, an application to FADAMA was submitted with Dr. Elizabeth Gardner, associate professor of forensic chemistry in the UAB Department of Criminal Justice, serving as a sponsor with American Academy of Forensic Sciences (AAFS) membership. As a result, FADAMA granted permission to access in February 2023.

Initially, the research design was to filter the FADAMA database for cases that fit the research criteria before downloading the .csv file. However, technical problems arose when trying to download the search results, and they could not be resolved after contacting FADAMA technical support. As a result, the entire database of 631 cases was downloaded on 28 February 2023 and manually filtered to match research criteria.

Research Criteria

As this is a study on method accuracy in ancestry estimation, the first criterion was to select cases that include a forensic anthropologist ancestry estimation. Case entries with a blank ancestry estimation were removed; however, cases that contain an "unknown" or "indeterminant" estimation were included, as these results still speak to

method accuracy. This filter reduced the number of eligible cases from 631 to 528. Following this, case entries that did not list an identified ancestry were removed, reducing the number of eligible cases to 498. An additional 23 cases were removed for listing the decedent as a foreign national; all the methods involved in this study were developed on U.S.-based skeletal collections and are designed to be used on an American population. Lastly, case entries that did not use one of the methods studied in this research (i.e., generalized morphology; Hefner, 2009; Hefner & Ousely, 2014; Rhine, 1990) were removed for a final total sample size of 241. Hefner (2009) noted no sexual dimorphism in the expression of morphoscopic traits, so individuals of both sexes were pooled together for analysis. Most of the case entries noted the use of multiple methods, so a random number generator was used to assign those cases to a specific method sample. This most notably affected the sample size for Hefner and Ousley (2014), the method most commonly listed in conjunction with another method. The final method sample sizes are as follows: Generalized Morphology ($n=59$); Hefner 2009 ($n=50$); Hefner and Ousley 2014 ($n=19$); and Rhine 1990 ($n=113$).

Coding the Data

While each case entry had many data points associated with it, the ones of interest for this study included (1) the forensic anthropologist report of ancestry; (2) identified race/ethnicity; (3) race/ethnicity notes; (4) the ancestry method used; and (5) the method's estimated outcome. In addition, supplement information from each case entry's "Case Notes" section was read and, if applicable, factored into the analysis. Each aspect

of the case entry is an open-response form, so responses varied for the five relevant data points. However, consistent with themes in the broader forensic anthropological literature (Maier, Craig, & Adams 2020), various terms were synonymously grouped—“Asian” and “Mongoloid”; “White,” “European,” and “Caucasian”; and “Black,” “African,” and “African-American.” Additionally, identifiers of Hispanic and Native American ancestry historically fall under the broader category of Asian ancestry traits. While some scholars in the field are attempting to develop a methodology that allows for a distinct Hispanic ancestral classification (e.g., Hurst, 2012), this has not become standardized across the field. As a result, an Asian ancestry estimation under Data Point (5) coupled with a Hispanic/Native American estimation in Data Points (1), (2), or (3) was not discounted unless the information provided in the “Case Notes” section showed a discrepancy of the current classification norm.

A holistic examination of each of the five data points and supplemental case notes was conducted for each case entry to assign it an estimation accuracy level ranging from 1 (entirely incorrect) to 5 (fully correct). This scoring system involved first looking at the correlation between information provided for the forensic anthropology ancestry estimation, the identified race/ethnicity, and the selected method’s estimated outcome. If the information provided for these three data points did not correspond, the case entry received an accuracy level 1 rating. When the information provided for identified race/ethnicity notes did not correspond with the information provided for the forensic report, the case entry was also scored as a level 1 (see Table 1). Case entries were scored at a level 2 accuracy rate when the information provided in the forensic report included the

mention of potential ancestral admixture and where the proposed admixture corresponded to the identified race over the report’s principal ancestry estimation. Case entries were also scored at level 2 when the “Case Notes” section mentioned the addition of *post hoc* information that affected the forensic report’s ancestry estimation (see Table 2). Case entries received a level 3 accuracy rate when the forensic report included an admixture estimate, but the identified race/ethnicity corresponded to the report’s broader ancestry estimate (see Table 3). Case entries were scored at a level 4 accuracy rate when the forensic report included quantifiers about the certainty of the ancestry estimate that also corresponded with the identified race/ethnicity and did not contain mention of potential admixture (see Table 4). Case entries were scored at a level 5 accuracy rate when the information provided in the forensic report corresponded to identified race/ethnicity, race/ethnicity notes (when applicable), and the method’s estimated outcome (when applicable) (see Table 5).

Table 1 — Examples of Case Entries with a Level 1 Accuracy Rate

FA Report: Ancestry	Identified Race/Ethnicity	Race/Ethnicity Notes (optional)	Ancestry Method	Estimated Outcome (optional)
[Asian or Asian American]	[White]	[Non Hispanic]	Hefner & Ousley (2014)	Black
[Hispanic]	[African-American/		Rhine (1990)	Hispanic
[Mestizo]	[White]	[Non Hispanic]	Rhine (1990)	Mongoloid (Hispanic)

Table 2 — Examples of Case Entries with a Level 2 Accuracy Rate

FA Report: Ancestry	Identified Race/Ethnicity	Ancestry Method	Estimated Outcome (optional)	Case Notes
[Black, possible Hispanic admixture]	[Hispanic]	Rhine (1990)	Black	
[Black]	[African-American/Black]	Rhine (1990)	Black	...ancestry estimation done again after ID was provided, and since last name is common of certain African

Table 3 — Example of Case Entry with a Level 3 Accuracy Rate

FA Report: Ancestry	Identified Race/Ethnicity	Ancestry Method
[Black with possible American Indian]	[African-American/Black]	Generalized Morphology

Table 4 —Example of Case Entry with a Level 4 Accuracy Rate

FA Report: Ancestry	Identified Race/Ethnicity	Ancestry Method	Estimated Outcome
[Possibly White]	[White]	Hefner & Ousley (201)	White

Table 5 — Example of Case Entry with a level 5 accuracy rate

FA Report: Ancestry	Identified Race/Ethnicity	Ancestry Method	Estimated Outcome
[Asian (to include Hispanic and Native American)]	[Native American]	Hefner (2009)	Asian

Data analysis

Data analysis began once all case entries were assigned an accuracy rate level. Data analysis was conducted using descriptive and inferential statistics. All descriptive statistics were calculated by hand, and all inferential statistical tests were run using an online calculator. Frequency distribution tables illustrate the results for each of the four methods (generalized morphology; Hefner, 2009; Hefner & Ousley, 2014; Rhine, 1990) and the two categories of methodology (typological and morphoscopic) to show the distribution of case entries for each method and method type among the five accuracy level ratings. In addition, measures of central tendency, variance, standard deviation, and standard error were calculated for each of the four methods and the two methodological categories. For inferential statistics, generalized morphology case entries and Rhine (1990) case entries were aggregated ($n=172$), and Hefner (2009) case entries and Hefner and Ousley (2014) case entries were aggregated ($n=69$) to assess differences between the categories of typological and morphoscopic methods. Several different inferential statistical tests were performed, as outlined below.

A chi-square test was performed to determine if the application-based accuracy rates of both categories of methods statistically differed from their reported research-based accuracy rates. Komar and Grivas (2008) note that reference skeletal collections used to developed methods do not always align with the demographics of modern populations found in forensic cases. As a result, methodological accuracy rates can vary between research and applied forensic casework. Thomas and colleagues (2017) echo this sentiment; however, their study found that ancestry estimation methods have comparable

accuracy rates for both research-based studies and applied forensic casework, with some applied-based methods having a slightly higher accuracy rate than research-based methods. To determine these results are supported by this study, applied-based accuracy rates for both categories of methods (typological and morphoscopic) were created by finding the ratio of case entries with levels 2-5 accuracy rates to overall case entries. While the distinction of these levels is useful when comparing the four individual methods, they are all considered accurate ancestry estimations and were examined together to reflect an overall accuracy rating for the two categories of methods. Research-based accuracy rates were determined based on what is available in the literature. Due to high levels of subjectivity, there are few published research-based accuracy rates for typological methods (see Albanese & Saunders, 2006; Hefner et al., 2012). Hughes and colleagues (2011) completed a study on assessing trait list subjectivity on accuracy rates and found an overall 90% rate of accuracy classification, which will serve as a proxy rate here since neither general morphology nor Rhine (1990) has a published accuracy rate. For the morphoscopic methods, a weighted average of Hefner's (2009) and Hefner and Ousley's (2014) published accuracy rates was determined to be 89.23% and was used for comparison. A chi-square test for each methodological category was then completed to compare the observed frequency of correct ancestry estimations to what is expected, using the research-based accuracy rating.

Additionally, a Wilcoxon Rank Sum Test (Mann-Whitney U Test) was performed to determine whether differences in the accuracy rates of typological and morphoscopic methods are statistically significant. One-tailed and two-tailed tests were performed. In

all tests, typological methods were Sample 1, and morphoscopic methods were Sample 2. Both one-tailed and two-tailed tests were run with the null hypothesis of (H_o : Sample 1 and Sample 2 have the same distribution) and a significance level of ($\alpha=0.05$). For one-tailed tests, the alternative hypothesis was (H_a : Sample 1 has a distribution with smaller values than Sample 2). For two-tailed tests, the alternative hypothesis was (H_a : Sample 1 has a distribution with different values than Sample 2). One-tailed and two-tailed tests were run with the entire sample sizes of typological and morphoscopic case entries ($n=172$ and $n=69$, respectively) and then again with a randomized sample from each category ($n=60$).

Two-sample t -tests were also performed using the same protocol as outlined above. These tests were performed because although the Wilcoxon Rank Sum Test is recommended for non-normally distributed data, parametric tests such as the two-sample t -test can also be used on non-normal data with large sample sizes like the ones in this study. De Winter and Dodou (2010) found that both tests have similar error rates and statistical power when examining five-point ordinal data and recommend performing both tests as data verification method. Both tests were run in this study since the data collected matched the criteria for both the Wilcoxon Rank Sum Test and the two-sample t -test.

CHAPTER FOUR

RESULTS

Descriptive Analysis

Tables 6-10 show the frequency distributions for each of the four methods. All of the methods have a similar frequency distribution, with the majority of case entries classified at an accuracy level 5 rank. Data for all methods have the same median and mode of Accuracy Level 5; however, the mean varies for each method's data set. Case entries that use Hefner (2009) have the highest mean level ranking at 4.62/5 and the lowest standard deviation (*sd* 1.10). Second highest are case entries that use generalized morphology with a mean ranking of 4.31/5 (*sd* 1.38), followed by those using Rhine (1990) with 4.22/5 (*sd* 1.45) and Hefner and Ousley (2014) with 4.21/5 (*sd* 1.51). It is noted that the Hefner (2009) method was utilized with the highest level of accuracy in determining ancestry estimations when using descriptive analysis.

Table 6 — Frequency Distribution for Generalized Morphology Method

Variable	Count of Case Entries	Cumulative Count	Percent	Cumulative Percent
Accuracy Level 1	6	6	10.169	10.169
Accuracy Level 2	3	9	5.085	15.254
Accuracy Level 3	4	13	6.78	23.034
Accuracy Level 4	0	13	0	23.034
Accuracy Level 5	46	59	77.966	100

Mean: 4.31, Median: 5, Mode: 5; Variance: 1.91; Standard Deviation: 1.38; Standard Error: 0.18

Table 7 — Frequency Distribution for Hefner (2009) Method

Variable	Count of Case Entries	Cumulative Count	Percent	Cumulative Percent
Accuracy Level 1	4	4	8	8
Accuracy Level 2	0	4	0	8
Accuracy Level 3	0	4	0	8
Accuracy Level 4	3	7	6	14
Accuracy Level 5	43	50	86	100

Mean: 4.62, Median: 5, Mode: 5; Variance: 1.22; Standard Deviation: 1.10; Standard Error: 0.16

Table 8 — Frequency Distribution for Hefner and Ousley (2014) Method

Variable	Count of Case Entries	Cumulative Count	Percent	Cumulative Percent
Accuracy Level 1	3	3	15.789	15.789
Accuracy Level 2	0	3	0	15.789
Accuracy Level 3	1	4	5.263	21.052
Accuracy Level 4	1	5	5.263	26.315
Accuracy Level 5	14	19	73.684	100

Mean: 4.21, Median: 5, Mode: 5; Variance: 2.29; Standard Deviation: 1.51; Standard Error: 0.35

Table 9 — Frequency Distribution for Rhine (1990)

Variable	Count of Case Entries	Cumulative Count	Percent	Cumulative Percent
Accuracy Level 1	15	15	13.274	13.274
Accuracy Level 2	3	18	2.655	15.929
Accuracy Level 3	9	27	7.965	23.894
Accuracy Level 4	1	28	0.885	24.779
Accuracy Level 5	85	113	75.221	100

Mean: 4.22, Median: 5, Mode: 5; Variance: 2.10; Standard Deviation: 1.45; Standard Error: 0.14

Frequency distribution tables were also created for each of the two methodological categories (typological and morphoscopic) by aggregating Generalized Morphology and Rhine (1990) and Hefner (2009) and Hefner and Ousley (2014), respectively. Tables 10 and 11 present the results. Again, both data sets had the same median and mode. However, the dataset of case entries that uses morphoscopic methods has a higher mean of a 4.51/5 accuracy level than the typological methods dataset, which has a mean of 4.25/5.

Table 10 — Frequency Distribution for Typological Methods

Variable	Count of Case Entries	Cumulative Count	Percent	Cumulative Percent
Accuracy Level 1	21	21	12.209	12.209
Accuracy Level 2	6	27	3.488	15.697
Accuracy Level 3	13	40	7.558	23.255
Accuracy Level 4	1	41	0.581	23.836
Accuracy Level 5	131	172	76.163	100

Mean: 4.25, Median: 5, Mode: 5; Variance: 2.02; Standard Deviation: 1.42; Standard Error: 0.11

Table 11 — Frequency Distribution of Morphoscopic Methods

Variable	Count of Case Entries	Cumulative Count	Percent	Cumulative Percent
Accuracy Level 1	7	7	10.145	10.145
Accuracy Level 2	0	7	0	10.145
Accuracy Level 3	1	8	1.449	11.594
Accuracy Level 4	4	12	5.797	17.391
Accuracy Level 5	57	69	82.609	100

Mean: 4.51, Median: 5, Mode: 5; Variance: 1.52; Standard Deviation: 1.23; Standard Error: 0.15

Inferential analysis

Chi-square tests

As noted above, several inferential statistical tests were run to determine the significance of the data. The first was a chi-square test to determine if the applied casework accuracy rates of the ancestry estimation methods, as reported in FADAMA, are statistically distinct from the research-based accuracy rates reported in the literature. Again, aggregate samples for typological and morphoscopic methods were created by combining Generalized Morphology and Rhine (1990) for the typological method dataset ($n=172$) and combining Hefner (2009) and Hefner and Ousley (2014) to form an overall morphoscopic method dataset ($n=69$). By analyzing the ratio of case entries with an accuracy level rating between 2-5 to the overall size of each of the dataset samples, applied accuracy rates were determined for each of the two methodological categories. Typological methods had an applied accuracy rate of 87.79%, and morphoscopic methods

had an applied accuracy rate of 89.96%. As noted in the methods section above, the research-based accuracy rates for the two categories are 90% for typological methods and 89.23% for morphoscopic methods. The research-based rates helped calculate the “expected” number of case entries with accurate and inaccurate ancestry estimates for each methodological category. While applied typological methods underperformed their research-based counterparts, a chi-square test determined this difference to be not statistically significant, with a two-tailed p -value of 0.3341. And while the applied morphoscopic methods slightly outperformed research-based accuracy rates, this difference is not statistically significant, with a chi-square test producing a p -value of 0.8674.

Wilcoxon Rank Sum Test

Results from the Wilcoxon Rank Sum Tests show no statistically significant differences between the accuracy rate of typological methods and morphoscopic methods of ancestry estimation. A two-tailed test examining the total sum of each methodological category’s dataset yielded a p -value of 0.4065. A one-tailed test resulted in a p -value of 0.2033, indicating that the claim that morphoscopic methods have a greater accuracy rate than typological methods cannot be supported by the current data. The same two tests were repeated using randomly generated samples of each dataset ($n=65$), and the same conclusions were supported ($p=0.2501$ and $p=0.1251$, respectively).

Two-sample t-test

Parametric two-sample *t*-tests were completed to support the findings of the Wilcoxon Rank Sum Tests. Again, two-tailed and one-tailed tests compared aggregate datasets for typological and morphoscopic methods. The initial two-tailed test returned a *p*-value of 0.0241, and the initial one-tailed test yielded a *p*-value of 0.012, both below the significance level ($\alpha=0.05$). The random sample ($n=65$) two-tailed test yielded a *p*-value of 0.0286. The same random sample datasets were used for a one-tailed test ($p=0.0143$). Given that these test results did not align with the results from the Wilcoxon Rank Sum Tests, the same two tests were repeated with a different random sample ($n=65$) from each methodological category to determine if there was a difference in power. In the two-tailed test, $p=0.2503$; in the one-tailed test, $p=0.1251$. A third round of two- and one-tailed *t*-tests were conducted using another random sample of the same size to determine which set of findings was anomalous. The result of this two-tailed test was $p=0.1899$ and $p=0.049$ for the one-tailed test. Based on the results of all three rounds of random sample testing, it is concluded that the hypothesis that differences in the accuracy rates of typological and morphoscopic methods of ancestry estimation are statistically significant cannot be supported. While the initial tests using the entire datasets ($n=172$ and $n=69$, respectively) returned *p*-values of statistical significance, several randomized sample tests did not. Most likely, these initial *p*-values ($p=0.0241$ and $p=0.012$) are due to the size differences in the datasets. When the sample sizes are equal, these statistically significant results disappear.

CHAPTER FIVE

DISCUSSION

Results from the descriptive analysis show that the Hefner (2009) method of ancestry estimation has the highest level of accuracy, followed by generalized morphology, Rhine (1990), and Hefner and Ousley (2014). Surprisingly, the two morphoscopic methods had the highest and lowest level of accuracy out of the four methods. It is worth noting, however, that two of the case entries for Hefner and Ousley (2014) have “Hispanic” listed as either the result of the application of the method or as the forensic anthropologist report’s estimated ancestry, while the Hefner and Ousley (2014) method is designed only for American Blacks and Whites. These discrepancies are likely the result of errors on behalf of the practicing forensic anthropologist rather than a fault of the methodology. When comparing the two methodological types, the morphoscopic methods had a higher mean accuracy rating than the typological methods. In addition, the morphoscopic methods had a lower percentage of case entries with an accuracy level 1 rating than did typological methods, indicating fewer case entries where the forensic anthropologist’s report of ancestry was wholly incorrect using those methods. This finding supports the body of forensic literature and the hypothesis of this study, stating that morphoscopic methods outperform typological methods of ancestry estimation in terms of accuracy rates.

Results of the chi-square tests show that differences in applied casework accuracy rates and research-based accuracy rates for both typological and morphoscopic methods are not statistically significant. However, it is worth noting that morphoscopic methods have a higher applied accuracy rate than is reported in the literature. In contrast, typological methods have a lower applied accuracy rate than is noted in the literature. This discrepancy speaks directly to points made by Hefner and Ousley (2006) and Hefner (2009) that frequencies for “expected” traits analyzed in typological approaches are much lower in applied work than they are perceived to be in the literature surrounding typological methodologies. Choosing traits *post hoc* that correspond to a specific typologically determined ancestry estimation is not supported with a higher accuracy rate and leaves much to be desired in terms of methodological objectivity.

Results from the Wilcoxon Rank Sum Tests show no statistical differences between the accuracy rates of typological and morphoscopic ancestry estimation methods. Further, the hypothesis that morphoscopic methods outperform typological methods could not be confirmed statistically. Results from the two-sample *t*-tests showed the same, with one caveat. The tests that compared the whole samples for typological methods ($n=172$) and morphoscopic methods ($n=69$) did show statistically significant differences in overall accuracy rates, and the hypothesis that morphoscopic methods outperform typological methods could be supported. However, the results were not statistically significant in re-runs of the same tests using random samples ($n=65$). Since these test results echoed those found using the Wilcoxon Rank Sum Test, these results are

assumed to be correct. The anomalous results from the first *t*-tests using the full sample sizes were likely due to the disparity in the sample sizes.

While the results of the inferential statistical tests show no statistically significant differences between the accuracy rates of typological and morphoscopic methods of ancestry estimation, the descriptive statistics show morphoscopic methods outperform typological methods in terms of accuracy rates. These results support the arguments made by several scholars (Hefner, 2009; Hefner et al., 2012; Hefner & Ousley, 2014; Klales & Kenyhercz, 2014) that morphoscopic methods should become the standard for nonmetric methods of ancestry estimation moving forward. This preference for morphoscopic methods is due to their higher accuracy rate, their levels of standardization and objectivity, and their scientific credibility. As mandated by the Daubert standard, morphoscopic ancestry estimation methods require empirical support, estimated error rates, standardization, and peer-reviewed validation (Hefner & Ousley, 2014). The added layer of statistical analysis that separates typological and morphoscopic methods also helps reduce levels of observer subjectivity and the conflation of certain “expected” traits as being definitive identifiers of a specific ancestral group, as, for example, Rhine’s (1990) assumption that the presence of a postbregmatic depression indicates an individual of Black ancestry. Instead, the premise of morphoscopic methods is that no trait is found exclusively in only one population. Rather, the frequency of different character expressions of a suite of traits can be utilized to make assessments of ancestry that are scientific and statistically valid (Hefner et al., 2012).

The results of this study fall firmly within the debate within forensic anthropology surrounding ancestry estimation and its relationship to race. The emphasis on the scientific and statistical validity of morphoscopic methods begs the question of what these methods say about ancestry. Do these methods support the argument that race is biologically determinable? Moreover, by extension, does this put the work of forensic anthropology at odds with the broader field of anthropology as a discipline? The American Anthropological Association (AAA) released its Statement on Race in 1998, which classifies the existence of distinct human races as a cultural worldview and the existence of racial classifications as a myth (AAA, 1998). Despite having no biological basis, race as a social construct has a significant bearing on humans' lives and is impossible to dismiss with the notion that "race is not real." These theoretical breakthroughs in anthropology and other disciplines have resulted in the overall race concept dividing into three distinct conceptions of race: social, bureaucratic, and biological. Both social and bureaucratic race are socially constructed concepts whose definitions and classifications vary depending on the specific cultural context. In contrast, the biological concept of race is theoretically based on human phenotypic and genotypic variation that does not exist (Albanese & Saunders, 2006). Ancestry estimation in forensic anthropology centers around providing data on an individual's social or bureaucratic race based on morphological variation while rejecting the concept of biological race. This distance from biological race increased with the field's change in the 1990s from determining "race" to determining "ancestry." Further, the categories of ancestral groups are subject to change depending on the cultural context. For example,

the choices of “White, Black, or Asian” have expanded to “White, Black, Asian, Native American, or Hispanic” to reflect changes in how the American population perceives the folk concept of race/ethnicity.

Nevertheless, some scholars argue that this is insufficient to sever the assumed relationship between ancestry estimation and race and question the utility of ancestry estimation in the field moving forward. DiGangi and Bethard (2021) adopt an extreme position and maintain that not only does the practice of ancestry estimation enforce the idea of biological race but also that the field’s continued use of ancestry estimation keeps forensic anthropology a race-based and racist discipline. They assume this position by reflecting on the history of the field and the development of its methods as part of race-based science to prove the existence of a hierarchy of human biological races. Even though the field has denounced this history and uses these methods for the purpose of identification rather than categorization, this contributes to a “color-blind” use of the methods that equates to a form of “racism-lite” which supports narratives of white supremacy, even while practitioners believe they are being anti-racist (2021, p. 424). Inherent racist undercurrents are especially true for nonmetric methods of ancestry estimation, as DiGangi and Bethard (2021) claim that morphological traits have never been studied with respect to geographic variation or heritability, making them unsuitable variables for assessing ancestry. However, they fail to note the distinctions in how morphoscopic and typological methods analyze morphological traits. While typological uses of morphological traits have no tangible connection to geographic variation or heritability, the analysis of suites of morphological traits in a statistical framework of

frequency distribution across various ancestral groups does speak to geographic variation, especially when these distributions are noted to be significant (Hefner, 2009; Stull et al., 2020). This connection to geographic variation is, in part, why the Hefner (2009) method of ancestry estimation performed the highest out of all of the methods analyzed in this study. In creating of this method, Hefner (2009) showed that morphoscopic traits could be associated with geographic variation when the traits are analyzed collectively. A verification study by Klales and Kenyhercz (2012) supports this conclusion.

DiGangi and Bethard (2021) also express skepticism that the process of ancestry estimation works since accuracy rates such as those examined in this study and those reported in Thomas et al. (2017) are determined using solved/identified cases. DiGangi and Bethard (2021) argue that it is impossible to truly know how ancestry estimation methods perform because the effect of an incorrect estimate on identification efforts is unknown. Albanese and Saunders (2006) also question whether it is possible to know the actual accuracy rates of ancestry estimation methods for much of the same reason. If practicing forensic anthropologists are unsure of their ancestry estimation, they argue that it is better not to provide one at all because the high risk of false information outweighs the value that a potentially correct estimate could have on a positive identification. However, the fears of both Albanese and Saunders (2006) and DiGangi and Bethard (2021) are unfounded. Hughes and colleagues (2021) analyzed cases from FADAMA to assess the overall accuracy of information presented in the forensic anthropologist report. They found that cases with inaccurate estimates in the forensic report occur among identified cases and that these inaccurate estimates are not necessarily a prohibiting factor

to a positive identification. Additionally, ancestry estimates had the second highest accuracy rate of all four biological profile components, lagging behind only biological sex estimation. Stull et al. (2020) turn the onus of proof back on critics by saying there needs to be data that supports claimed negative impacts of ancestry on investigations, as that is noticeably missing from the current literature.

DiGangi and Bethard (2021) also stress the problem of ancestry estimations attempting to assign an etic identity to an individual when identity is inherently emic. They argue that the determinants of social race identities are various sociocultural factors with no analogs in the skeletal material. Sauer (1992:110) argues against this claim directly, stating that the ancestry estimation made by a forensic anthropologist is indeed emic as it uses categories that are likely to have been used to describe the individual in life. This distinction is vital to understanding forensic anthropologists' motivations when making ancestry estimations. The process of ancestry estimation has little to do with the existence of biological races. Instead, it involves translating information about biological traits into a social race classification (Sauer, 1992). The fact that forensic anthropology is an applied field cannot be overstated. Compared to broader biological anthropology, forensic anthropology is concerned with answering practical questions about an individual that can assist in identification. Because of this, the field is concerned with no other definition of race than the social definition, as that will be most salient for identification.

Indeed, the only reason ancestry estimation is practical in United States forensic anthropology is the strong association between skeletal morphology and social race in the

American population (Ousley et al., 2018). In large part, this has to do with the country's history and institutionalized racism that kept social race groups segregated for an extended period of time. These factors have notably kept groups morphologically distinct due to a lack of gene flow between them. Sauer (1992) noticed this distinction and subsequent concordance between social race and skeletal morphology, which Ousley and colleagues (2009) verified. Even with increased levels of gene flow within recent years, groups defined by social race categories (e.g., Whites and Blacks) have retained enough unique skeletal morphology that ancestry estimations can provide salient information about individuals to lead to possible identifications, at least up until now.

However, it is worth noting that this phenomenon does not address the existence of individuals of mixed races whose skeletal morphology blurs the lines between two ancestral groups. As mentioned in the literature review, ancestry estimation methods do not accurately account for mixed race individuals. In the broader forensic anthropological literature, there is no consensus on how to estimate ancestry for mixed race individuals. Much of this has to do with the insidious assumption perpetuated by typological methods that the presence of certain traits automatically assumes a specific ancestral group. With this outlook, there is no room for an unknown cranium to display several traits from two or more ancestral categories without calling into question the validity of the methodology. For their part, morphoscopic methods do slightly more to address the reality of mixed race individuals. This is because analysis focuses on suites of traits rather than on singular traits and because results are presented as probability percentages of belonging to different ancestral groups. These two factors allow forensic anthropologists

more space to consider the possibility of a mixed ancestry estimation. However, there is no recognized protocol for offering a mixed ancestry estimate so the decision to do so is left up to the practicing forensic anthropologist. There could become a time in the future where the population of mixed race individuals is large enough to warrant further investigation into determining specific categorical criteria, like what Hurst (2012) did with the Hispanic population. More likely than not, if this does become a reality, it would be most beneficial to examine a suite of traits and their degrees of expression within the population. Work could potentially be done to include a “mixed” ancestry reference sample in MaMD to be used in morphoscopic analysis. However, genetic variability in the population might make trait expressions too inconsistent to warrant this type of analysis. If this becomes the case, it is time to re-examine ancestry estimation methodology or develop another form that would work to identify mixed race individuals.

In line with changing methodology, Kenyhercz and colleagues (2017) note in their verification study of Hefner and Ousley’s (2014) method that the sectioning point between Blacks and Whites needs to shift. This supported by the data from this study showing that case entries that used the Hefner and Ousley (2014) method had the lowest accuracy rate. A possible explanation for this shift is secular change in the distribution of trait expressions. Jantz and Meadows Jantz (2000) note that secular change is possible and that the cranium has changed in size and shape for both U.S. Black and White populations from the nineteenth century to the 1970s. It should be noted that these changes are not enough to significantly diminish the morphological differences

frequently found between the two groups. However, it should also be appreciated that secular change does happen and that there could become a point when ancestry estimation is no longer useful to distinguish between social races. At this time, the field should remove it from the methodological toolkit. This possibility highlights the need for constant verification and validation of methodological accuracies, like this study.

CHAPTER SIX

CONCLUSION

This research examined four nonmetric ancestry estimation methods used in United States forensic anthropology by assessing their applied accuracy rates. Two methods were selected to represent each of two main categories of nonmetric methods: typological and morphoscopic. A critical reading of the literature led to the hypothesis that morphoscopic methods will outperform typological methods in accuracy. This hypothesis was tested by analyzing case entries from in the Forensic Anthropology Database for Assessing Methods Accuracy. While inferential analysis showed no statistical significance between the accuracy rates of both categories of methods, descriptive analysis supported the hypothesis.

The results of this study support the literature highlighting the superiority of morphoscopic methods over typological methods of ancestry estimation for being more accurate, objective, and scientifically valid. In addition, the results of this study lend themselves to the debate surrounding ancestry estimation and its relation to race that is currently taking place in the field of forensic anthropology. While the practice of ancestry estimation does not in any way corroborate the existence of a biological basis for race, morphoscopic methods of ancestry estimation do show a significant correlation between suites of morphological trait expressions and their presentation in various ancestral

groups. This means is that there is a concordance between skeletal morphology and social race categories in the United States due to its history of institutionalized racism.

Morphoscopic methods, rather than typological methods, highlight this concordance and can be used to assign unknown individuals to relevant ancestral categories to aid in identification. However, due to the reality of secular change, it is possible that this correlation will not exist in the future. As a result, accuracy studies such as this are essential to ensuring that the methods used by the field of forensic anthropology are still relevant and able to help assist the process of forensic identification in a meaningful way.

REFERENCES

- AAA statement on race. (1998). *American Anthropologist*, 100(3), 712-713. <https://doi.org/10.1525/aa.1998.100.3.712>
- Albanese, J., & Saunders, S. R. (2006). Is it possible to escape racial typology in forensic identification? *Forensic Anthropology and Medicine*, 281–316. https://doi.org/10.1007/978-1-59745-099-7_12
- Bass, W. M. (2005). *Human Osteology: A Laboratory and Field Manual*. Missouri Archaeological Society.
- Berg, G. E., & Tersigni-Tarrant, M. A. (2014). Sex and ancestry determination: assessing the “gestalt”. In *Proceedings of the 66th Annual Meeting of the American Academy of Forensic Sciences* (pp. 17-22).
- Buikstra, J. E., & Ubelaker, D. H. (1994). *Standards for Data Collection from Human Skeletal Remains*. Arkansas Archeological Survey.
- Burns, K. R. (2012). *Forensic Anthropology Training Manual* (3rd ed.). Routledge.
- Byers, S. N. (2002). *Introduction to Forensic Anthropology*. Allen and Bacon.
- Christensen, A. M. (2004). The impact of Daubert : Implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). *Journal of Forensic Sciences*, 49(3), 1–4. <https://doi.org/10.1520/jfs2003185>
- Cunha, E., & Ubelaker, D. H. (2019). Evaluation of ancestry from human skeletal remains: A concise review. *Forensic Sciences Research*, 5(2), 89–97. <https://doi.org/10.1080/20961790.2019.1697060>
- de Winter, J., & Dodou, D. (2010). Five-Point Likert Items: t Test Versus Mann–Whitney–Wilcoxon. *Practical Assessment, Research & Evaluation*, 15(11), 1–12.

- DiGangi, E. A., & Bethard, J. D. (2021). Uncloaking a lost cause: Decolonizing ancestry estimation in the United States. *American Journal of Physical Anthropology*. <https://doi.org/10.1002/ajpa.24212>
- Dunn, R. R., Spiros, M. C., Kamnikar, K. R., Plemons, A. M., & Hefner, J. T. (2020). Ancestry estimation in forensic anthropology: A Review. *WIREs Forensic Science*, 2(4). <https://doi.org/10.1002/wfs2.1369>
- Dwight, T. (1881). The sternum as an index of sex and age. *Journal of anatomy and physiology*, 15(Pt 3), 327.
- Dwight, T. (1890a). Sternum as an index of sex, height, and age. *Journal of anatomy and physiology*, 24(Pt 4), 527.
- Dwight, T. (1890b). The closure of the cranial sutures as a sign of age. *The Boston Medical and Surgical Journal*, 122(17), 389-392.
- Dwight, T. (1894a). Methods of estimating the height from parts of the skeleton. *Medical Record (1866-1922)*, 46(10), 293.
- Dwight, T. (1894b). The Range and Significance of Variation in the Human Skeleton: The Shattuck Lecture for 1894. *The Boston Medical and Surgical Journal*, 131(4), 73-76.
- Giles, E., & Elliot, O. (1962). Race identification from cranial measurements. *Journal of Forensic Sciences*, 7(2), 147-157.
- Hanihara, T., Ishida, H., & Dodo, Y. (2003). Characterization of Biological Diversity Through Analysis of Discrete Cranial Traits. *American Journal of Physical Anthropology*, 121(3), 241–251. <https://doi.org/10.1002/ajpa.10233>
- Hefner, J. T. (2003). *Assessing nonmetric cranial traits currently used in forensic determination of ancestry* (thesis). The University of Florida, Gainesville, FL.
- Hefner, J. T. (2007). *The statistical determination of ancestry using cranial nonmetric traits* (dissertation). The University of Florida, Gainesville, FL.
- Hefner, J. T., & Ousley, S. D. (2006). Morphoscopic traits and the statistical determination of ancestry II. In *Proceedings of the 58th Annual Meeting of the American Academy of Forensic Sciences* (pp. 20-25). American Academy of Forensic Sciences.

- Hefner, J. T. (2009). Cranial Nonmetric Variation and Estimating Ancestry. *Journal of Forensic Sciences*, 54(5), 985–995. <https://doi.org/10.1111/j.1556-4029.2009.01118.x>
- Hefner, J. T., & Ousley, S. D. (2014). Statistical Classification Methods for Estimating Ancestry Using Morphoscopic Traits. *Journal of Forensic Sciences*, 59(4), 883–890. <https://doi.org/10.1111/1556-4029.12421>
- Hefner, J. T., Ousley, S. D., & Dirkmaat, D. C. (2012). Morphoscopic Traits and the Assessment of Ancestry. In D. Dirkmaat (Ed.), *A Companion to Forensic Anthropology* (pp. 287–310). essay, Blackwell Publishing Ltd.
- Hefner, J. T. (2018). The macromorphoscopic databank. *American Journal of Physical Anthropology*, 166(4), 994–1004. <https://doi.org/10.1002/ajpa.23492>
- Hooton, E. A. (1939). *The American Criminal: an Anthropological Study*. Harvard University Press.
- Hughes, C. E., Juarez, C. A., Hughes, T. L., Galloway, A., Fowler, G., & Chacon, S. (2011). A Simulation for Exploring the Effects of the “Trait List” Method’s Subjectivity on Consistency and Accuracy of Ancestry Estimations*. *Journal of Forensic Sciences*, 56(5), 1094–1106. <https://doi.org/10.1111/j.1556-4029.2011.01875.x>
- Hughes, C.E., & Juarez, C. A. (2018). Learning from Our Casework: The Forensic Anthropology Database for Assessing Methods Accuracy (FADAMA). NIJ 2018 DU-BX-0213.
- Hughes, C. E., Juarez, C., & Yim, A. D. (2021). Forensic anthropology casework performance: Assessing accuracy and trends for biological profile estimates on a comprehensive sample of identified decedent cases. *Journal of Forensic Sciences*, 66(5), 1602–1616. <https://doi.org/10.1111/1556-4029.14782>
- Hurst, C. V. (2012). Morphoscopic Trait Expressions Used to Identify Southwest Hispanics*. *Journal of Forensic Sciences*, 57(4), 859–865. <https://doi.org/10.1111/j.1556-4029.2012.02080.x>
- Jantz, R. L., & Meadows Jantz, L. (2000). Secular change in Craniofacial Morphology. *American Journal of Human Biology*, 12(3), 327–338. [https://doi.org/10.1002/\(sici\)1520-6300\(200005/06\)12:3<327::aid-ajhb3>3.0.co;2-1](https://doi.org/10.1002/(sici)1520-6300(200005/06)12:3<327::aid-ajhb3>3.0.co;2-1)

- Juarez, C. A., Hughes, C. E., & Yim, A. D. (2020). Technical note: A report on the Forensic Anthropology Database for Assessing Methods Accuracy. *American Journal of Physical Anthropology*, 174(1), 149–150. <https://doi.org/10.1002/ajpa.24167>
- Kenyhercz, M. W., Klales, A. R., Rainwater, C. W., & Fredette, S. M. (2016). The optimized summed scored attributes method for the classification of U.S. blacks and whites: A validation study. *Journal of Forensic Sciences*, 62(1), 174–180. <https://doi.org/10.1111/1556-4029.13243>
- Klales, A. R., & Kenyhercz, M. W. (2014). Morphological Assessment of Ancestry Using Cranial Macromorphoscopies. *Journal of Forensic Sciences*, 60(1), 13–20. <https://doi.org/10.1111/1556-4029.12563>
- Komar, D. A., & Grivas, C. (2008). Manufactured populations: What do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection. *American Journal of Physical Anthropology*, 137(2), 224–233. <https://doi.org/10.1002/ajpa.20858>
- Krogman, W. M., & İşcan, M. Y. (1986). *The Human Skeleton in Forensic Medicine*. Charles C. Thomas.
- Langley, N. R., Meadows Jantz, L., Ousley, S. D., Jantz, R. L., & Milner, G. (2016). Data collection procedures for forensic skeletal material 2.0. University of Tennessee, Knoxville, TN.
- Maier, C., Craig, A., & Adams, D. M. (2020). Language use in ancestry research and estimation. *Journal of Forensic Sciences*, 66(1), 11–24. <https://doi.org/10.1111/1556-4029.14601>
- Ossenberg, N. S. (1976). Within and between race distances in population studies based on discrete traits of the human skull. *American Journal of Physical Anthropology*, 45(3), 701–715. <https://doi.org/10.1002/ajpa.1330450337>
- Ousley, S. D., & Hefner, J. T. (2005). The statistical determination of ancestry. In *Proceedings of the 57th annual meeting of the American Academy of Forensic Sciences* (pp. 21-26).
- Ousley, S., Jantz, R., & Freid, D. (2009). Understanding race and human variation: Why forensic anthropologists are good at identifying race. *American Journal of Physical Anthropology*, 139(1), 68–76. <https://doi.org/10.1002/ajpa.21006>

- Ousley, S., Jantz, R. L., & Hefner, J. T. (2018). From Blumenbach to Howells: The slow, painful emergence of theory through forensic race estimation. In C. C. Boyd Jr. & D. C. Boyd (Eds.), *Forensic anthropology: Theoretical Framework and Scientific Basis* (pp. 65–97). essay, John Wiley & Sons Ltd.
- Rhine, S. (1990). Non-metric Skull Racing. In G. W. Gill & S. Rhine (Eds.), *Skeletal attribution of Race* (pp. 9–20). essay, Maxwell Museum of Anthropology.
- Ross, A. H., McKeown, A. H., & Konigsberg, L. W. (1999). Allocation of crania to groups via the “new morphometry.” *Journal of Forensic Sciences*, *44*(3), 584–587. <https://doi.org/10.1520/jfs14513j>
- Sauer, N. J. (1992). Forensic Anthropology and the Concept of Race: If Races Don't Exist, Why Are Forensic Anthropologists So Good at Identifying Them? *Social Science & Medicine*, *34*(2), 107–111. [https://doi.org/10.1016/0277-9536\(92\)90086-6](https://doi.org/10.1016/0277-9536(92)90086-6)
- Spradley, M. K. (2016). Metric Methods for the Biological Profile in Forensic Anthropology: Sex, Ancestry, and Stature. *Academic Forensic Pathology*, *6*(3), 391–399. <https://doi.org/10.23907/2016.040>
- Stewart, T. D. (1979). *Essentials of Forensic Anthropology: Especially as Developed in the United States*. Charles C. Thomas.
- Stull, K. E., Bartelink, E. J., Klales, A. R., Berg, G. E., Kenyhercz, M. W., L'Abbé, E. N., Go, M. C., McCormick, K., & Mariscal, C. (2020). Commentary on: Bethard JD, Digangi EA. Letter to the Editor—Moving Beyond a Lost Cause: Forensic anthropology and ancestry estimates in the United States. *J forensic sci.* 2020;65(5):1791–2. doi: 10.1111/1556-4029.14513. *Journal of Forensic Sciences*, *66*(1), 417–420. <https://doi.org/10.1111/1556-4029.14616>
- Thomas, R. M., Parks, C. L., & Richard, A. H. (2017). Accuracy rates of ancestry estimation by forensic anthropologists using identified forensic cases. *Journal of Forensic Sciences*, *62*(4), 971–974. <https://doi.org/10.1111/1556-4029.13361>
- Ubelaker, D. H. (2006). Introduction to Forensic Anthropology. In A. Schmitt, E. Cunha, & J. Pinheiro (Eds.), *Forensic Anthropology and Medicine: Complementary Sciences From Recovery to Cause of Death* (pp. 3–12). essay, Humana Press Inc.
- Ubelaker, D. H. (2018). A history of forensic anthropology. *American Journal of Physical Anthropology*, *165*(4), 915–923. <https://doi.org/10.1002/ajpa.23306>

Wijsman, E. M., & Neves, W. A. (1986). The use of nonmetric variation in estimating human population admixture: A test case with Brazilian blacks, whites, and Mulattos. *American Journal of Physical Anthropology*, 70(3), 395–405. <https://doi.org/10.1002/ajpa.1330700313>

Winburn, A. P., & Algee-Hewitt, B. (2021). Evaluating population affinity estimates in forensic anthropology: Insights from the forensic anthropology database for assessing methods accuracy (fadama). *Journal of Forensic Sciences*, 1210–1219. <https://doi.org/10.1111/1556-4029.14731>