2025

An analysis of FORDISC 3.1 accuracy: estimating population affinity of Asian individuals in America using CT scans

https://hdl.handle.net/2144/51899

BOSTON UNIVERSITY

ARAM V. CHOBANIAN & EDWARD AVEDISIAN SCHOOL OF MEDICINE

Thesis

AN ANALYSIS OF FORDISC 3.1 ACCURACY: ESTIMATING POPULATION AFFINITY OF ASIAN INDIVIDUALS IN AMERICA USING CT SCANS

by

GIULIA ROMA DUNN

B.A., The College of New Jersey, 2021

Submitted in partial fulfillment of the
requirements for the degree of
Master of Science
2025

Approved by

First Reader ____

Sean D. Tallman, Ph.D.
Associate Professor of Anatomy and Neurobiology

Second Reader ____

Tara Moore, Ph.D.
Professor of Anatomy and Neurobiology

DEDICATION

I would like to dedicate this work to my brother, my father, and my late mother.

ACKNOWLEDGMENTS

I would like to thank Dr. Sean Tallman, Dr. James Pokines, Dr. Tara Moore, and department administrator Patty Jones for their assistance in the completion of this thesis and master's degree. A warm thanks to my classmates in my cohort, without whom I would not have been able to withstand the long nights of writing, reading, and editing.

AN ANALYSIS OF FORDISC 3.1 ACCURACY: ESTIMATING POPULATION AFFINITY FOR ASIAN INDIVIDUALS IN AMERICA USING CT SCANS

GIULIA ROMA DUNN

ABSTRACT

This study is an assessment of the popular population affinity estimation software FORDISC 3.1 (Ousley & Jantz 2005a). This software has been popularized in forensic anthropology and is a useful method of population affinity estimation. Other morphoscopic and morphometric approaches are used by forensic anthropologists, however these methods have been found to present difficulties in classifying individuals of groups outside of the tripartite system: African, European, Asian/Native American/Latin American (Plemons & Hefner 2016). This study uses computed tomography scans of 148 individuals of five different demographic groups from the New Mexico Decedent Image Database (NMDID) to assess the accuracy and functionality of FORDISC 3.1 in analyzing a modern United States population. It was suspected that individuals belonging to the Asian American population affinity group would be more likely to be misclassified than any other group, and more specifically, Asian American individuals who were assigned female at birth (AFAB) would be less accurately classified than all other demographic groups. The results of this study indicated that FORDISC misclassified Asian American AFABs and Native American AFABs at a higher rate than any other group. There was also a distinct overclassification of individuals in the sample as Latin American AMABs. These three population affinity groups have been shown through research to be the most consistently inaccurately

classified groups (Dudzik & Jantz 2018; Hughes et al. 2018). Asian populations and Latin American populations span large areas, and the lack of research into the genetic diversity within these groups likely is the cause of much of the misclassifications. To address the clear misunderstanding of population affinity in Asian American, Latin American, and Native American populations, more research needs to be done to differentiate the genetic variation and cranial morphology variation within and among these groups (González-José et al. 2008; Scott et al. 2023; Powell & Neves 1999; Willerslev & Meltzer 2021. The field of forensic and biological anthropology has long debated the applicability of population affinity estimates. However, until new non-population specific methods of biological profile estimation become mainstream, more work needs to be done on making the current population affinity estimation methods more applicable to diverse populations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA	African American
AAF	African American Female
AAM	African American Male
AFAB	Assigned female at birth
AIF	Asian Indian Female
AIM	Asian Indian Male
AMAB	Assigned male at birth
AS	Asian American
BF	Black Female
BIPOC	Black, Indigenous, and other people of color
BM	Black Male
CHM	Chinese Male
CT	Computed tomography
DCP	Data Collection Procedures
DFA	Discriminant function analysis
DICOM	Digital Imaging and Communications in Medicine
EA	European American
EAF	European American Female
EAM	European American Male
FDB	Forensic Anthropology Data Bank

GTM	Guatemalan Male
HF	Hispanic Female
HM	Hispanic Male
JF	Japanese Female
JM	Japanese Male
KM	Korean Male
LA	Latin American
N/A	Not Available
NA	Native American
NAF	Native American Female
NAM	Native American Male
NMDID	New Mexico Decedent Image Database
NOK	Next of Kin
WF	White Female
WF (HISP)	White Female with Hispanic Ethnicity
WM	White Male
WM (HISP)	White Male with Hispanic Ethnicity
VM	Vietnamese Males

INTRODUCTION

In forensic cases, the estimation of ancestry – now generally referred to as population affinity – is a potentially useful tool for anthropologists to winnow a group of missing persons. To estimate this portion of the biological profile, forensic anthropologists must have a thorough understanding of human variation, and due to the complex presentation of skeletal features pertaining to ancestry, it is one of the more difficult aspects of the biological profile to complete. There are two primary approaches to the estimation of population affinity by forensic anthropologists: morphoscopic and morphometric. Morphoscopic, or nonmetric, analysis of population affinity is based on the assessment and scoring of structural traits on the cranium, mandible, and dentition. Alternately, forensic anthropologists use osteometrics as a morphometric method in conjunction with statistical analyses, such as discriminant function analysis (DFA), to estimate population affinity.

Software, such as *FORDISC*, allows forensic anthropologists to apply customizable DFAs to craniometric data from unknown individuals. *FORDISC* allows the estimation of an individual's stature and sex, in addition to population affinity, based on standardized skeletal measurements. The *FORDISC* software is comprised of several reference groups; thirteen of these groups are from the Forensic Anthropology Data Bank (FDB) (Ousley & Jantz 2005a). The data within the FDB come from voluntarily submitted forensic case reports from across the United States. This databank is comprised of primarily African Americans (FDB title: "American Blacks", females=137, males=224), and European-Americans (FDB title: "American Whites", females=454, males=737). Other groups are also included in the FDB with fewer numbers such as Native Americans (FDB title: “American Indians”, females=32, males=69), Latin-American (FDB title: “Hispanic”, females=74, males=281), Chinese (males=80), Guatemalan (males=83), Vietnamese (males=51), and Japanese (females=58, males=84). In several of these smaller sized reference groups there are entire cranial measurements excluded from the databank. These sample sizes demonstrate the lack of variation available in the FDB and subsequently in *FORDISC*.

This study analyzed digital scans of 148 skulls of a modern US population and estimated population affinity using FORDISC 3.1. These estimates were then compared to the expected classifications as provided by Next of Kin and the database Investigator analyses. More research is being conducted to assess the functionality of modern population affinity methods in classifying Asian populations (Atkinson & Tallman 2019; Dudzik & Jantz 2016; Gonzalez-Jose et al. 2008; Hughes et al. 2021; Tallman & Winburn 2015). The intention of this study is to add to this research and address the applicability of the FORDISC 3.1 software in its reliability for classifying modern US populations and to posit how anthropologists can move forward to create more inclusive and comprehensive models of population affinity software. This is a concern now, more than ever, due to the increasing globalization and admixture of populations around the world (L’Abbe et al. 2013; Monte and Shin 2022; Manthey et al. 2018; Relethford 1994; Ross et al. 2002). FORDISC is a useful tool for anthropologists, but to ensure its continued applicability it must regularly be updated to be representative of the modern population seen around the world today.

POPULATION AFFINITY IN FORENSIC ANTHROPOLOGY

The field of forensic anthropology rests between the natural and the social sciences. However, with the recent social movements across the globe, it has become more important to infuse anthropological frameworks into forensic techniques. Population affinity estimation has been long associated with evolutionary trends and migration patterns of early humans (Spradley 2016). The estimation methods for population affinity in forensic anthropology are problematic due to the lack of perspectives appreciative of variation within ancestral groups. Additionally, the estimation of population affinity, as noted by DiGangi and Bethard (2021), is heavily reminiscent of the typological and hierarchical approach to the classification of humans. There has long been debate in the field of biological anthropology over the importance and relevance of ancestry and its relationship to race. There has even been the proposed separation of two schools of thought of biological anthropologists, the “splitters” and the “lumpers” (Lieberman & Reynolds 1978). The “splitters” believe in the biological separation of races, while the “lumpers” do not recognize this distinction (Adams & Pilloud 2021). It is clear that population affinity, ancestry, and race have been entangled throughout the history of biological anthropology, and since the 1970s there has been an effort to clarify how to standardize the concept of race in forensic anthropology. Forensic anthropologists, whether unwittingly or knowingly, perpetuate the biological race concept by applying current methodologies of population affinity estimation which disproportionately affects Black, Indigenous and other people of color (BIPOC) missing individuals (DiGangi & Bethard 2021; Tallman et al. 2021). The traditional methods used in population affinity estimation classify individuals primarily into three big groups—African, Asian, and European—which does not address diversity *within* these larger groups (Adams & Pilloud 2021; Atkinson & Tallman 2020). There is the assumption that all individuals under these broad categories will present phenotypic variation in similar ways. Moreover, there is little understanding for the causes behind the human variation used in population affinity estimation, and an understanding of these reasons may allow for a deeper appreciation for the variation seen between geographically diverse groups from the same greater category (Relethford 1994).

Throughout the history of anthropology in the United States, efforts have been primarily focused on assessing the relationship between cranial measurements and population affinity for American black and American white individuals (Komar & Grivas 2008; Winburn et al. 2022). Little research has been conducted on the osteometric variation of the crania in individuals of Asian descent in the U.S. (Atkinson & Tallman 2020; Go et al. 2019; Tallman 2016; Tallman 2019). As of the newest update to *FORDISC* (3.1), the only groups representative of modern individuals of Asian groups are Chinese, Japanese, and Vietnamese, and within each of these reference groups only sparse numbers of AFABs are included (Ousley & Jantz 2005a). Furthermore, the individuals included in these reference groups are from continental Asia, rather than Asian individuals in the United States, which calls into question their applicability on individuals in the United States (Ousley & Jantz 2005a). The small reference samples in the FDB are hypothesized to create inconsistencies in the estimated population affinity identifications of the unknown individuals being analyzed. Many methods used by forensic anthropologists are population specific, meaning that they can accurately be applied to particular population affinity groups (Flouri et al. 2022). Should an individual be categorized into the wrong population affinity group, it would likely impact the methodologies used for estimating the other aspects of the biological profile.

TERMINOLOGY OF POPULATION AFFINITY

Terminology in forensic anthropology is a widely debated topic as forensic anthropologists are positioned in between accommodating law enforcement and including as many population groups as possible while trying to holistically understand human variation (Marten et al. 2022). The field of forensic anthropology evolved from the work of anatomists and did not see much evolution or standardization until the 1970s (Bartelink et al. 2020). Many methods upon which forensic anthropologists base their population affinity estimates use the tripartite classification terminology which classifies individuals as European, African, or Asian (Adams & Pilloud 2021). This classification system does not consider the possibility of admixture, or variances between groups within a geographic region (e.g., East Asia vs South Asia; Atkinson & Tallman 2020). The *FORDISC* software (Ousley & Jantz 2005a) uses terminology that is explicitly being challenged today, such as “American Indian” when referring Native American individuals, and “Hispanic” in reference to Latin American or Latine individuals. The issues with terminology have been present in the exam administered by the American Board of Forensic Anthropologists up until 2003, with the usage of the term “Negroid” in reference to African population affinity classifications (Bartelink et al. 2020). The lack of standardization in terminology usage between forensic anthropologists further complicates the application of population affinity estimation as a portion of the biological profile. The terminology used in biological anthropology throughout the decades since its inception have provided ammunition to individuals hoping to use race as a divider (Pilloud et al. 2021). The typological verbiage that anthropologists have used to describe different groups of geographically diverse peoples are often taken out of context and used to support ideologies of enslavement, White supremacy, and bigotry (Adams & Pilloud 2021; Pilloud et al. 2021).

MISCLASSIFICATIONS OF POPULATION AFFINITY

Recent studies have demonstrated the inaccuracies in the classifications of *FORDISC*, specifically with individuals of Latin American population affinity (Dudzik & Jantz 2016; Hughes et al. 2018). Moreover, due to the narrative that Native American populations descended from ancient Asian groups, there is still a conflation of Asian craniometric morphology with Native American and Latin American (classified 'Hispanic' in *FORDISC*) craniometric morphology (Atkinson & Tallman 2020; Tallman 2020). However, studies on secular change in facial morphology (Jantz & Jantz 2000) as well as mandibular morphology (Kilroy et al. 2020), demonstrate that even in a short time span of approximately 150 years, there are changes in overall cranial morphology, as reflected in metric and nonmetric data used in population affinity estimations. This means that, operating under the assumption that conflation between Native Americans and Asians is due to the peopling of America, disregards the changes that both populations have experienced in the past several thousand years (Arnaiz-Villena et al. 2010).

Asian populations within the U.S. have been increasing for decades, with the most recent census indicating an almost 25% increase in the number of individuals in the U.S. that identify as Asian or Asian in combination with another race (Monte & Shin 2022). As of the 2020 census, 6.2% of the U.S. population identifies as Asian (Monte & Shin 2022). The steady increase in individuals that identify as Asian residing in the United States indicates a need for more inclusive methods of estimating population affinity, and a more anthropological approach to understanding the reasons for variation between population affinity groups categorized together as Asian. The purpose of this study is to assess the accuracy of *FORDISC 3.1* in classifying Asian American individuals in comparison to the accuracy with which the software classifies individuals of African American, European American, Latin American, and Native American population affiliation.

Several studies show that the lack of variability in the methods used for estimating population affinity can thoroughly skew results (Dunn et al. 2020). Within *FORDISC 3.*1, each reference group in the FDB is made up of less than 50% females. Of these thirteen reference groups, three have no females included and only one of the three Asian reference groups (Chinese, Vietnamese, and Japanese) has females included. The misclassification of Asian individuals is expected to be even higher in Asian assigned female at birth (AFAB) individuals as the *FORDISC* software lacks comparative female groups. Many nonmetric methods for estimation of the four facets of the biological profile were created on African American and European American assigned male at birth (AMAB) individuals and therefore their application to any other population affinity or sex group is impractical (Spradley 2016). However, many methods of estimation for sex and age still are conducted using reference casts, or models referenced when assessing remains to ensure accurate classification, for white males and females. These methods of the biological profile cannot accurately be applied to other demographic groups without consideration for the skeletal variation present (Plens et al. 2021). In this sense, FORDISC provides an advantage as it does represent more than just the racial binary of "black" versus "white". The underrepresentation of multiple population affinity groups in biological profile estimation methods results in more inaccurate classifications and potential misidentification of remains in forensic contexts. These are mistakes that cannot and should not be made when doing work to bring justice to a victim. The use of forensic anthropology in reducing a possible victim pool is valuable as long as the tools being used can accurately represent the community being served.

COMPUTED TOMOGRAPHY

A relatively new tool to the field of forensic anthropology is computed tomography (CT). This radiographic process is often used in forensic cases as a way to assess aspects of the decedent that are not visible externally (Baglivo et al. 2013). CT scans are especially useful as it is a non-invasive technique that allow for the preservation of the skeletal remains (Christensen et al. 2018). CT scans have been in use since the 1970s and since then their function in forensic settings has evolved to include virtual autopsies and virtual assessment of the biological profile (Abegg et al. 2021; Bollinger et al. 2007; Christensen et al. 2018; Lo et al. 2023). The scans themselves are two-dimensional slices that can then be converted into three-dimensional renderings (Carew

& Errickson 2019). While CT scans can be limited by their vulnerability to metal, the detail provided can be particularly useful in morphoscopic and morphometric analyses (Carew & Errickson 2019). In particular, the New Mexico Decedent Imaging Database (NMDID) is comprised of thousands of CT scans of individuals that have gone through the Office of the Chief Medical Examiner in Pima County New Mexico (Edgar et al. 2020; Vidoli 2022). Databases such as this are becoming more popular as they allow researchers to analyze several skeletal remains without needing the physical skeletal elements and allowing for the preservation of data which can be made available worldwide (Abegg et al. 2021; Lo et al. 2023). Skeletal repositories exist across the world; however, gaining access to these collections can be difficult and is not readily accessible to many researchers. The creation of online collections of skeletal scans allows for more research to be done (Kewal et al. 2016; Simmons-Ehrhardt 2021). The scans from the NMDID provide a sample of the modern American demographic that has not been thoroughly analyzed in the context of the biological profile and the methods used to achieve the bio profile.

PURPOSE

The morphometric analysis conducted in this study acts to further reinforce the novice concept that if population affinity is to remain a part of the biological profile it must be adjusted. The political and socio-cultural changes in the past decade have made it clear that differences between cultural groups that are often associated with population affinity groups must be considered when classifying remains. The tripartite framework that has for so long been accepted as an appropriate way to classify individuals is proving to be less accurate with the admixture of ancestry in the newest generations. The methods and software used to assess the biological profile, particularly population affinity, must be revalidated with modern populations and accordingly overhauled. It is hypothesized that individuals of Asian population affinity will be classified with less accuracy than the four other major demographic groups within the United States, and further females will be misclassified at a higher rate than males.

The following study is organized to provide more in-depth background information, explain the materials and methods used in this study and the results, and offer a discussion about how this study fits in the context of contemporary and future research. Contextual information about the current methods used in the field of biological anthropology for estimating population affinity, as well as their benefits and shortcomings are explained. A brief explanation of the past interpretations of race, ancestry, and population affinity in the field of biological anthropology is delivered to elaborate on the complexities of estimating population affinity. The peopling of the Americas is discussed to highlight the importance of understanding human variation and secular change when estimating population affinity. To better understand the software being analyzed in this study, the benefits, and flaws of the FORDISC 3.1 software are also discussed, particularly by addressing various studies on the misclassification of different population affinity groups using this FORDISC. The materials and methods, and results function here to indicate how the study collected data in a reliable and efficient manner to provide the most accurate interpretation of the functionality of FORDISC. This is followed by a discussion on how this study can impact future research, and how it can be framed in contemporary biological anthropology. This study was created with the intent of highlighting the need for more inclusivity in modern population affinity methods.

PREVIOUS RESEARCH

BIOLOGICAL ANTHROPOLOGY AND ANCESTRY

Biological anthropology stems from the natural sciences as far back as the mid-19th century and was founded on interests in human origins, human anatomy, and race (Little & Kennedy 2010). Much of the inception of biological anthropology occurred in Philadelphia, Pennsylvania. At the time, Philadelphia was a major hub of trade, knowledge, and political revolution at the time (Mann 2009). In the early years of the field of biological anthropology, particularly in the 18th century, human variation was interpreted through race. This was a particularly salient subject due to the ongoing discussion regarding access to civil rights in the founding of the United States (Little & Kennedy 2010; Mann 2009). The early approaches to human variation followed a typological approach in which the classifications of ancestries were equivalent to the classification of races (DiGangi & Bethard 2021). The typological assessment is what drove race and ancestry estimation as these approaches were not measuring for biological distances of genes, but rather the geographic location from which an individual originated from (Edgar & Pilloud 2021). One of the first attempts to create quantifiable evidence of typological differences between ancestral groups and races was craniometry. Craniometry was a method of applying a quantitative analysis to classify remains (Gould 1981). Measurements of cranial capacity were obtained, and often these measurements were interpreted as estimates of intelligence (Gould 1981). Samuel G. Morton was one of the original supporters of the typological approach to race and ancestry assessment. In measuring the cranial capacity of crania of various ancestral groups, he created an improper interpretation of intelligence which supported his hierarchy of races based on intellect (Armstrong-Fumero 2014; Lewis et al. 2011). His vast collection of human crania and his degrees in medicine earned him a positive reputation throughout the 19th century, as he used his work to rank racial groups which has impacted the field of biological anthropology ever since (Gould 1981). The typological approaches to population affinity were the primary means for estimating an individual's ancestry, and it is argued that some of these methodologies were significantly biased to promote racist ideologies (Armstrong-Fumero 2014). This was often accompanied by the hierarchical structuring of these groups with Europeans consistently placed at the top and other "races" such as African or Native American were considered inferior (DiGangi & Bethard 2021). This hierarchical structuring of races fed into the misinformation of the 18th century. The hyper-fixation on the human physical form and its relation to race resulted in decades of research dedicated to racial and hierarchical classifications, which were then presented to the public as fact (DiGangi & Bethard 2021). The presentation of scientific information to the public is helpful in informing communities about certain states; however, in this context the presentation of scientific information was done in such a way as to perpetuate harmful race theory that has since remained rooted in culture and society. This configuration of races laid the foundation for decades of research based on typological assessment of humans using both qualitative and quantitative methods.

As biological anthropology evolved as a field into the 20th century, the typological approach to race and ancestry remained a prominent school of thought. In the early 1900s, the field was referred to as 'physical anthropology' and this was sustained by some of the more prominent researchers at the time when founding the *American Journal of Physical Anthropology* (Little & Kennedy 2010). Diving further into the 20th century, physical anthropology saw the rise of eugenics and the concept of a racial hierarchy. The concept of eugenics was the encouragement of procreation between individuals with desirable traits, and discouragement of procreation between individuals with undesirable characteristics (Rutherford 2020). Eugenics became a commonly discussed concept in the political arena throughout the early 20th century, and it took center stage during World War I and II. There was a wave of support for eugenics by political leaders in the United Kingdom and the United States in the 1910s and in many cases undesirable traits were equated with certain racial groups or disabilities, this resulted in mass sterilization of certain marginalized groups (Rutherford 2020). Eugenics provided a justification for political groups to use science to begin mass removals of certain racial groups which acted as a perpetuation of the scientific racism already evident at this time. These views on race were often supported by the research using the typological framework of population classification in physical anthropology (Little & Kennedy 2010). After World War I, American anthropologists were urged by the government and other institutional frameworks to focus time and energy into race science (Caspari 2009). However, not all members of the field at the time supported the racial hierarchy that was created by founders of physical anthropology (Smay & Armelagos 2000). Franz Boas was a critical adversary to race science for many decades and taught many of his students with this same philosophy (Caspari 2009). As with most changes in conceptual understandings, there were individuals promoting either school of thought and this influenced the future of the field of biological anthropology as a whole.

Although there were major proponents of dismantling race theory, throughout the mid-20th century racialized research remained. This was a time when race studies were the center of what was then still physical anthropology. While research supporting racialized science by anthropologists such as Carleton Coon was still ever present, a multitude of contemporary anthropologists in the 1950s began taking steps to undo the damage of research conducted by Morton, Coon, and others (Caspari 2003). For example, UNESCO had put forth a statement in 1950 to support the idea that race is a social construct (Rutherford 2020). The emphasis on separating race from biology is one that is still present in the works of many authors today, as the perpetual damage from the early works in biological anthropology still weigh heavy on society today. The works of individuals such as Boas, laid the foundation for physical anthropology to move toward a more scientifically sound and less typological approach to population affinity estimation; however, there is still work to be done to dismantle the tripartite and five-race systems that have built much of the discipline.

Biological anthropology continues to wrestle with its complicated origins. The typological approach to race and ancestry based on the idea that physiological characteristics automatically can be grouped together in racial classifications is no longer salient (Berg & Ta'ala 2014). However, still today the morphological methods utilized in assessing remains, is primarily an experience-based method that categorizes individuals based on phenotypic skeletal traits (L'Abbe et al. 2011). Efforts have been made by some of the leading organizations of biological anthropologists, such as the American Association of Biological Anthropologists (AABA), to encourage anthropologists to consider how their research interacts with race and racialized ideas. In fact, the AABA went as far as to put forth a statement on race concepts and the role that biological anthropology has played in perpetuating racial categories (Fuentes et al. 2019). The statement addressed the usage of race in forensic anthropology and more specifically what is classified as race and what is not, emphasizing that race is not aligned with the patterns of human variation among populations (Fuentes et al. 2019). There is continued promotion by anthropologists today to disassociate race from ancestry in anthropology, but further research into human variation and its presentation throughout populations around the world are necessary.

CURRENT METHODS OF ESTIMATING POPULATION AFFINITY

The estimation of population affinity is limited by the current methods which were created on reference groups that encompassed little human variation. The word "race" was replaced in anthropology by the term "ancestry"; however, the change was simply linguistic, as there has been little change in theoretical or methodological approaches (Ross & Williams 2021; Tallman et al. 2021). Efforts have been made to clarify the differences between "race" and "ancestry" in the field of biological anthropology by delineating the biogeographical nature of "ancestry" and the socially constructed foundation of "race" (Tallman et al. 2021). However, this clarification has yet to be dispersed throughout all methodology of anthropology. There has been minimal reconstruction of the methods or approaches to population affinity estimation after the linguistic switch from “race” to “ancestry”. Two popular types of methodology exist in estimating population affinity, morphological (or morphoscopic), and morphometric. Morphological methods are a subjective assessment of the general shape, or morphology, of skeletal elements and involves the scoring of particular traits on a scale. These scores often then are plugged into a statistical equation to produce an estimate. Morphometric methods measure the distances between certain landmarks on skeletal elements and these measurements are then statistically analyzed to produce an estimate. The usage of morphoscopic and morphometric methods in estimating population affinity is defended by the understanding that individuals clustered in similar geographic areas will experience similar genetic and environmental histories and therefore exhibit similar physical traits (Parsons 2021). In both cases, statistical analyses are used to create less subjective, and more objective population affinity estimates.

Morphological methods are often used in estimates of population affinity today. An example of a popular morphological method for population affinity estimation is Hefner’s (2009) cranial morphological scoring which attempts to classify individuals into “African, American Indian, Asian, or European” groups based on certain cranial traits. This method involves the scoring of eleven cranial traits including anterior nasal spine, inferior nasal morphology, interorbital breadth, the malar tubercle, nasal aperture width, nasal bone contour, nasal overgrowth, postbregmatic depression, supranasal suture, shape of the transpalatal suture, and the shape of the zygomaxillary suture (Hefner 2009). These traits were selected because of their expression of variation between population affinities.

Hefner’s (2009) study, while beneficial in that it created a more objective approach for scoring morphological traits of the cranium, does have its limitations. As previously mentioned, a large part of the functionality of a method is its reference groups. This is because the variation in the reference groups can be an indicator of the applicability of the method to different groups. The samples selected in Hefner’s (2009) study are not spatiotemporally accurate representations of contemporary United States populations. The Native American and Asian samples in particular cannot be assumed to be representative of modern US Native Americans and Asian American individuals. The Native American sample is comprised of prehistoric and protohistoric individuals with the age of remains spanning back to 1000-200 B.P., while the Asian sample is comprised of individuals selected from Japan and China (Hefner 2009). Neither of these samples encompass the secular change that has likely occurred in both populations throughout time and in different countries (Jantz & Meadows Jantz 2000; Weisensee & Jantz 2011). Population affinity methods such as this one are still in use today and do offer anthropologists an objective way of analyzing remains. Another population affinity method often used in biological and forensic anthropology due to its more objective structure is the OSSA method (Hefner & Ousley 2014). This method uses six of the previously mentioned cranial traits for analysis, these include anterior nasal spine, inferior nasal aperture, interorbital breadth, nasal aperture width, nasal bone structure, and post-bregmatic depression (Hefner & Ousley 2014). Once these traits are scored, they are transferred to a binary code and those numbers summed. The OSSA method has a limited reference group selection and only allows for individuals assessed using this method to be classified as “black”, “hispanic”, or “white” (Hefner & Ousley 2014). This method is not recommended by its authors for usage in every case due to its simplistic structure, and its inability to accurately classify individuals of mixed population affinity (Hefner & Ousley 2014). Other morphological methods include post-cranial methods. These, similarly, to morphoscopic methods, are also often developed with small reference groups. A key example of such a study would be Duray et al. (1999) in which the morphology of the cervical spinous processes are assessed for estimation of population affinity. This study scores the presence of bifurcation of the cervical spinous process, and only allows for individuals to be categorized into two groups: “black” or “white” (Duray et al 1999). These morphological methods allow for subjective analyses of traits to be analyzed in a statistical manner; however, these methods often are limited in their applicability to a variety of population affinities.

Morphometric methods were created with the intent of developing more objective methods to assess population affinity. Measurements are taken using different types of instrumentation to obtain quantifiable data on the structure of skeletal elements (Langley et al. 2016; Parsons 2021). A popular morphometric method includes the usage of the FORDISC 3.1 software, which allows the user to input measurements between different skeletal landmarks to obtain a population affinity estimate. The measurements submitted by the researcher are analyzed using a discriminant function analysis and compared to thirteen different reference groups (Ousley & Jantz 2005a). This software is helpful in that it is an accessible way for anthropologists to use a morphometric approach to population affinity, however, it is limited in its applicability to different population affinities.

The morphoscopic and morphometric methods most utilized today are often applied in conjunction with one another (Parsons 2021). Due to the limited reference groups and possible population affinity results for these methods, individuals often are classified differently by different methods. The lack of diverse reference groups used in these methodologies do not accurately represent the ever changing and varied population of the modern United States. The constant flow of individuals from around the globe indicates constant gene flow and suggests the need to adapt the methods for estimating population affinity to take these changes into account.

CRANIAL AND GENETIC VARIATION IN NORTH AMERICA

The methodologies of biological anthropology ancestry estimations have historically focused on classification of individuals into three large groups: African, European, or Asian. The Asian classification has long included Asian ancestry, as well as Hispanic, and Native American (Plemons & Hefner 2016). While efforts have been made by anthropologists, to direct population affinity away from this classification system, the interpretation of shared geographical ancestry amongst these groups has led to the assumption that their morphologic cranial traits will also share similarities. However, through research on the secular change visible in other populations, it can be inferred that similar change to cranial structure and morphology has occurred in Asian, Native American, and Hispanic groups (Jantz & Meadows Jantz 2000; Weisensee & Jantz 2011).

To be able to begin differentiating between these three groups, research has been done on the genetic flow observed during the peopling of the Americas from continental Asia.

For centuries anthropologists have worked to understand human variation, particularly in the Americas. Historically, to better understand this variation throughout time, anthropologists have studied the origins of Native American individuals as they have long been understood to be the direct link between continental Asia and North and South America. It is a supported theory that the Americas were populated by migration from Asia close to 15,000 years ago, though recent research has suggested even more ancient peopling. The most accepted understanding of the peopling of America is that the region was first inhabited by Asian populations following passage over the Bering Strait, which occurred over three separate flows of migration (González-José et al. 2008; Scott et al. 2023). The migration patterns associated with the peopling of the Americas is often utilized as an explanation for the craniofacial variation seen within and between groups in the Americas (Powell & Neves 1999). The research found that craniofacial variation seen within groups in the Americas does not directly align with genetic variation. In fact, early analyses of genetic variation in Native American populations indicated a lack of genetic diversity (Powell & Neves 1999; Willerslev & Meltzer 2021). This lack of genetic diversity, but presence of morphological diversity was concluded to be a result of the homogenous and isolated migration of individuals from one common area (Powell & Neves 1999). However, more recent research of ancient genomics in North America suggests that the lack of genetic diversity developed from a significant loss of genetic information of certain Native American lineages due to the demographic shift of North

America with the arrival of European settlers (Willerslev and Meltzer 2021). More research is necessary to better understand the genomic variation in North America as it is lacking in information (Willerslev & Meltzer 2021). A more thorough understanding of the genetic changes between Asian, Hispanic, and Native American groups will allow for a more comprehensive approach to estimating population affinity in these groups.

To address the need for more holistic understandings of genetic diversity and its relationship to skeletal morphology, researchers have used new technology in genetic testing to ascertain that environmental influences are not the sole proponent of change in cranial anatomy (Harvati & Weaver 2006). Cranial morphology and morphometric analysis of cranial landmarks are regularly used to assess population affinity and structure throughout history; therefore it is imperative to understand how human variation presents in cranial anatomy and its causes (Roseman 2004). The understanding of human variation is incredibly important in all subfields of biological anthropology, but it is particularly essential when considering the structure and characteristics of the cranium. The usage of cranial based methods for differentiation of population affinity becomes complicated particularly because of the research suggesting that cranial vault morphology is more reflective of carried genetics of population affinity groups, while craniofacial morphology presents with the residual effects of climactic change on morphology (González-José et al. 2008). Craniofacial structures such as nasal aperture have been found to be heavily reliant on climatic change or other environmental factors such as diet (Harvati & Weaver 2006). Parts of the neurocranium are also understood to be influenced by environmental and climatic factors, however, the basicranium diverges from this norm as its morphology relies heavily on genetics and is therefore a strong indicator of genetic variation within or between groups (Harvati & Weaver 2006). Most of the among-region variation witnessed in Homo sapiens crania is likely the result of neutral evolutionary forces, this acts as supplementary information to the already accepted concept that the history of a population is highly impactful on cranial characteristics (Roseman 2004). In understanding the regional variation of cranial morphology, researchers have been able to use morphoscopic and morphometric analyses to help understand migration patterns to further clarify patterns of variation. Craniometric data helped develop the four-wave theory of migration from Asia to America which has allowed for a more comprehensive understanding of potential migration patterns and timelines of the peopling of America, thus allowing resarchers to address the long-standing assumptions of genetic and morphologic similarities between Asian, Hispanic, and Native American populations (Sardi et al. 2004). Research using this understanding of cranial variation based in genetic and morphologic change has allowed for anthropologists to better address the change that has occurred over time.

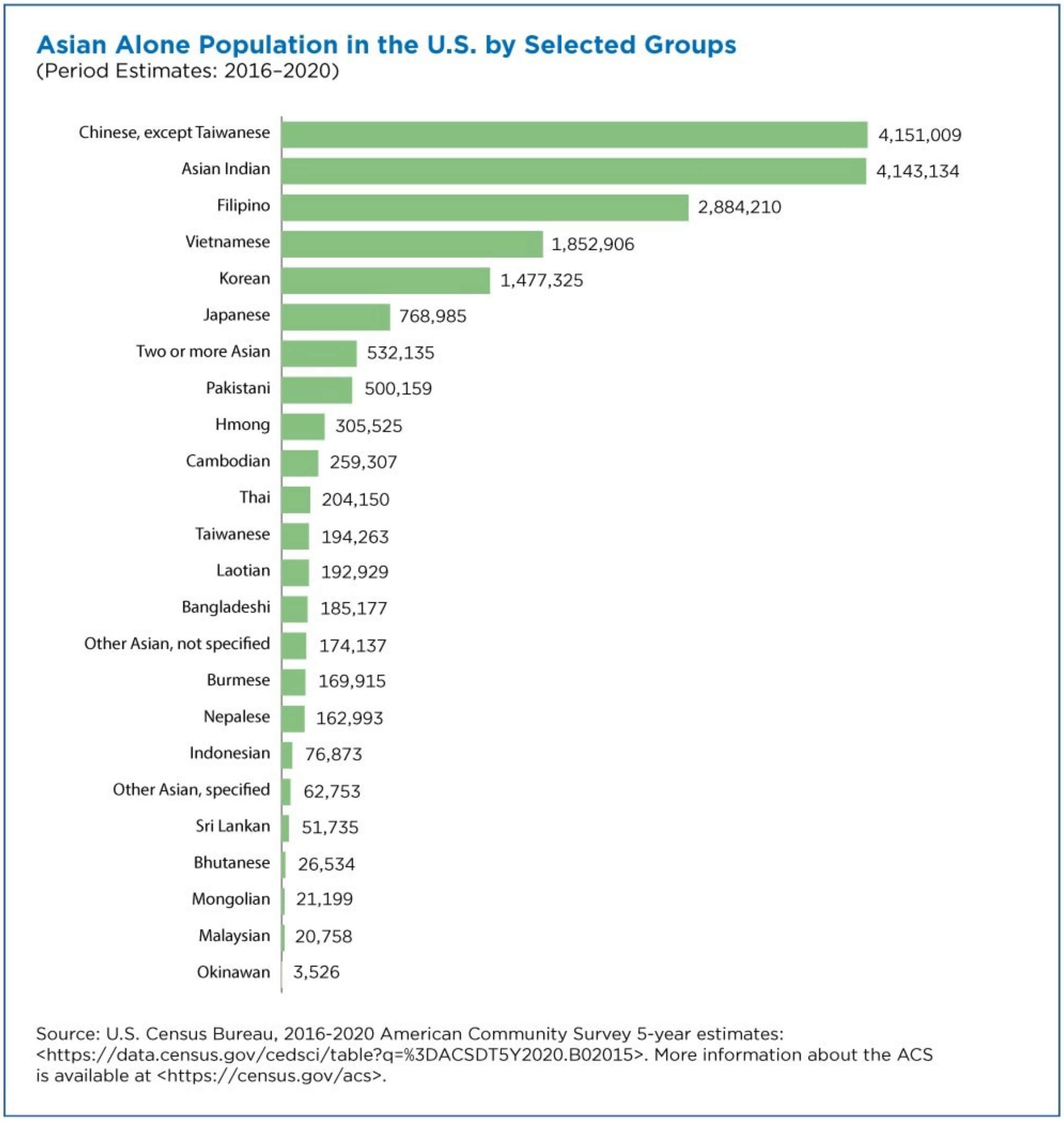
Several studies have shown that secular change in populations can occur in as little as 100 to 200 years; it can only be expected that a similar if not more aggressive level of change can be seen within Asian, Hispanic, and Native American populations over the past 15,000 years (Jantz & Meadows Jantz 2000; Weisensee & Jantz 2011). Native American groups are understood to have a common ancestral group descended from Asian Pleistocene populations (González-José et al. 2008), however, it cannot be assumed that these ancestral genetic ties supersede secular change that has resulted from climatic and demographic changes. Research on modern Asian populations and Asian-derived groups has become much more prevalent as the demographic structure of North America and the world continues to change (Atkinson & Tallman 2020; Dudzik & Jantz 2016; Hughes et al. 2018; Tallman & Winburn 2015). The extensive admixed history of Latin American populations creates further complexities in estimating their population affinity due to the conflation between population affinity groups genetically and morphologically (Ross et al. 2014 p. 155-164). The phenotypic characteristics of modern northeastern Asian populations are noticeably different from the characteristics seen in modern Native American and Hispanic populations (Sardi et al. 2004). To better understand variation within these three populations, more research needs to be conducted to address the secular changes within these groups that have led to further within and among group variation.

The field of anthropology has worked hard to separate itself from its complicated foundations, however, many methods used today still perpetuate inaccurate theories about distribution of populations. The standards used by many anthropologists do aim to recognize the change forward that the field is making both in technological advances as well as in understanding a need to address variation (Buikstra & Ubelaker 1994). The standardization of data and the methods of collection are of importance to allow for anthropologists to compare data and to create a more comprehensive understanding of human variation and how anthropologists address population affinity and other aspects of the biological profile. Ancestry estimation should be grounded in the idea that there is structured human variation and there are reasons for this variation. Some reasons may be geographic in that certain environments selected for or against certain cranial traits, but the current approaches to population affinity in forensic anthropology are concerned with typology and categorization more than having a complete understanding of the processes that cause these differences (Tallman et al. 2021). As the field of biological anthropology moves forward, it is clear that the goal is to improve the methods and technology that are currently in use today to be more inclusive and accurate in estimating aspects of the biological profile, particularly population affinity.

COMPOSITION OF MODERN UNITED STATES POPULATION

The oversimplification of population affinity is often tied to the classifications of race. In many cases the two are used interchangeably even though studies have suggested that the connection between inherited phenotypic skin traits and cranial traits may not be as connected as once expected (Relethford 2002). Modern populations of humans repeatedly show higher levels of variation within regional groups than between different regional groups, this low level of among group variation is a strong challenge to classifications based on “race” (Relethford 1994). The United States has become a widely diverse country that can no longer be divided into only three or four major groups based on race. Recent studies show that, due to the increase in global migration, there is an increase in gene flow. Therefore, amendments must be made to the methods of population affinity estimation (Go et al. 2019). As previously addressed, the methods currently in use are often limited by the reference groups used. However, current understandings of human variation in forensic anthropology are informed by the data collected on reference groups, most notably American White, American Black, and Native American groups from a specific period in time (Komar & Grivas 2008). With these limited groups available, this leaves large gaps in understanding the scope of cranial variation in Asian population affinity groups.

Two of the four morphoscopic and morphometric methods recently discussed contain reference groups that are representative of Asian populations. Of the two studies with Asian reference groups, neither have Asian American references. This is notable due to the consistent increase in Asian American identifying individuals in the United States. According to recent US Census data, 73.2% of individuals who identified themselves as "Asian alone" were citizens of the United States by birth or naturalization (Monte & Shin 2022) (Figure 1). With such a large percentage of the Asian demographic in the United States identifying as Asian American, it is imperative that more research be done to better understand the variation within Asian population affinity groups. The mass migrations to the United States from all over the globe need to be considered when using the methods popular in forensic anthropology today.



Asian Alone Population in the U.S. by Selected Groups
(Period Estimates: 2016–2020)

Source: U.S. Census Bureau, 2016-2020 American Community Survey 5-year estimates: <https://data.census.gov/cedsci/table?q=%3DACSDT5Y2020.B02015>. More information about the ACS is available at <https://census.gov/acs>.

Figure 1. Distribution of individuals that identified as “Asian alone” on the United States Census.

FORDISC 3.1

What is FORDISC 3.1?

FORDISC 3.1 is a software used by forensic anthropologists nationwide to assist in the estimation of population affinity for unidentified remains. The software was created in 1986 with a grant from the National Institute of Justice, and uses various skeletal measurements obtain an estimate of population affinity, sex, and stature based on statistical analyses (Ousley & Jantz 2005a). One of the most common uses of FORDISC is to estimate population affinity using craniometric data. The reference groups used for population affinity estimation were obtained from the Forensic Data Bank (FDB) which is comprised of individuals born after 1900 (Ousley & Jantz 2005a). The FDB was created with the primary intent of comparing remains from the 20th century to remains from the Terry and Hamann-Todd collections (Ousley & Jantz 2005a). The collection of data within the FDB continues to grow slowly as forensic anthropologists around the U.S. submit information from the cases they complete. As the FDB continues to expand the information is input into FORDISC to increase the sizes of the reference groups. Forensic anthropologists use FORDISC to assist in identification of remains as the software uses discriminant function analyses (DFAs) to compare the obtained measurements to other groups to find the most probable population affiliation. However, most of the individuals in the FDB are American whites, Hispanic, and American black individuals with 1191, 355, and 361 individuals respectively. This is in comparison to the significantly smaller sample sizes seen for American Indians (Native American) (n=91), Chinese (n=80), Guatemalan (n=83), Japanese (n=142), and Vietnamese (n=51) individuals. While reference groups for other Asian ancestral groups are available in FORDISC, they are derived from historical collections, and due to secular change, may not accurately reflect the Asian populations in America today.

FORDISC also contains reference groups from a worldwide collection curated by W.W. Howells. This collection known as the Howells' dataset is comprised of cranial measurements collected between 1965 to 1980 (Howells 1996). The collection is comprised of approximately 2504 individuals (AFAB=1156, AMAB=1348) from 28 detailed population groups (Howells 1996). Both Howells' (1973, 1989) studies were used to delineate the different groups included in FORDISC (Ousley & Jantz 2005a). The dataset is comprised of up to 82 measurements obtained from each individual in the collection and are primarily in agreement with previous standards of craniometric measurement standards (Howells 1996; Pang et al. 2022). The Howells dataset was intended to consist of 50 males and 50 females from each population, and there are inclusions of test samples as well as unidentified population samples as well in the dataset (Pang et al. 2022). Currently, the Howells dataset is one of the largest craniometric databases available on the internet, accessible through the University of Tennessee Knoxville website (https://web.utk.edu/~auerbach/HOWL.htm). The Howells dataset provides a worldwide collection of cranial measurements; however, this data is not representative of modern human populations and should be used cautiously—if at all—for forensic casework.

Measurements used in FORDISC 3.1

There are up to thirty-seven measurements taken on human skull for population affinity estimation in FORDISC, as highlighted by Langley et al. (2016). A series of inconsistencies between the measurements in the original Data Collection Procedure (DCP) (Moore-Jansen et al. 1994) and the FDB led to a reevaluation of certain measurements and therefore an updated version of the DCP was necessary. The measurements defined by the DCP 2.0 are meant to assist forensic anthropologists in their use of FORDISC. Langley et al. (2016) provide explicit definitions for landmarks and measurements, and diagrams are also provided to indicate where specific landmarks are located, and where linear measurements should be collected. Cranial, mandibular, and postcranial measurements are outlined in the DCP 2.0 which allows the reader to estimate not only population affinity and sex through FORDISC, but also stature (Langley et al. 2016). However, the reference group data for stature for 19th century individuals are from the Terry collection, and the reference groups for black and white males is from World War II soldiers which may not be accurately representative of modern US populations (Ousley & Jantz 2005a). The measurements taken with guidance from the DCP 2.0 should be applied to remains in the FDB or that will be analyzed against the FDB reference groups. Measurements taken in accordance with the Howells methods, should be compared to the Howells reference group in FORDISC. The Howells group consists of several population affinity groups from across the world, each group is given a three-letter abbreviation to represent the name of the group (Ousley & Jantz 2005a). For the purposes of this study, the DCP 2.0 landmark definitions and measurements were used to collect data from each individual and were compared to the reference groups from the FDB as these groups represent the most modern groups in FORDISC.

Statistics of FORDISC 3.1

Forensic anthropologists in the U.S. rely on the FORDISC software to assist in the estimation of population affinity for unidentified remains, and therefore require statistically significant methods of estimation (Parsons 2021). The user interface allows analysts to create a case folder which is made up of three pages (i.e., FDB, Howells, or postcranial measurements/reference comparisons) (Ousley & Jantz 2005a). The software allows users to analyze data using unique DFAs based on the available measurements. The statistical procedure of DFAs allow for the classification of unknown individuals into specific groups based on known measurements of the reference groups. The measurements of the unknown individual are then converted into unique discriminant scores which are then compared to the mean DFA score of each reference group. After entering the measurements collected from the unidentified individual, users can conduct an analysis with all reference groups included on the first attempt. It is recommended that the analysis on one individual be run several times, each time removing the reference group with the lowest likelihood of similarity. Different probabilities – posterior and typicality – are presented alongside the classification results. Posterior probabilities are the chance that an unknown individual is a member of each group and sum to 1, while typicality probabilities, particularly F-Typicality's, are the likelihood that an unknown individual is a member of a specific group based on the F distribution (Ousley & Jantz 2005a). The most common DFA used is the linear discriminant function analysis, which are derived from a linear combination of the original measurements. These distances are also called Mahalanobis Distances which are important as the closest mean score of reference to a reference group for an unknown individual means a smaller Mahalanobis distance (Ousley & Jantz 2005a). The statistics provided by FORDISC allow the user to make an educated estimate as to the probability that an unknown individual is a part of a certain group. If a probability is presented on the results page in red, the probability is too low and therefore should be ignored as the classification cannot be considered reliable (Ousley & Jantz 2005a). FORDISC also checks for potential measurement errors when measurements of an unknown case are submitted. The software compares the measurements to the standard univariate range of values from the various populations, and if a measurement is outside of this range, it is recommended that it be removed from the analysis (Ousley & Jantz 2005a). Once measurements are processed, the results page will indicate that a measurement is outside of the standard range with "+" or "-" signs, the more of each symbol next to a measurement, the more likely it is that the measurement is incorrect (Ousley & Jantz 2005a).

The statistical framework upon which FORDISC is built allows for an objective and qualitative estimation of population affinity. This has become increasingly important in forensic contexts due to the growing pressures established by the *Daubert* rulings (Parsons 2021). The posterior probabilities provided by FORDISC in population affinity estimation allow the researcher to report an estimate with the statistical backing to present the information confidently. However, as with any method, there are limitations and therefore these results while potentially statistically significant have to be interpreted with consideration of the shortcomings of the software.

Misclassifications using FORDISC 3.1

The lack of population and sex diversity in the FDB as well as the small sample sizes can cause issues with population affinity identification in which individuals of one ancestral makeup are identified as another group (Dudzik & Jantz 2016; Hughes et al. 2018; Manthey et al. 2018). Assessments of the reliability of the FORDISC software have been conducted on various populations. It's applicability in the United States is of particular concern as it is so readily used in medicolegal contexts. There has been a significant change to the demographic structure of the United States with a noticeable increase in the populations of Latin American individuals immigrating into the country (Dudzik & Jantz 2016; Hughes et al. 2018; Monte & Shin 2022). This change suggests that analyses need to be done to better understand how the FORDISC software classifies these individuals as to ensure that misclassifications are at a low. Recent research suggests that when craniometric measurements of Hispanic individuals are analyzed with Native American and White populations included, the Hispanic individuals are consistently classified poorly and were likely to classify as Asian (Dudzik & Jantz 2016). This research also noted how the broad classification of Hispanic to refer to any Spanish-speaking Latin American country consequentially includes Native American, European, and African ancestry due to the high levels of admixture of these groups in the Latin Americas (Dudzik & Jantz 2016, Hughes et al. 2018). It is suggested that any population estimates, or probability patterns provided by FORDISC should be considered with population history and context in mind, particularly when assessing the probability estimates of Latin American and Asian individuals due to the vast amounts of heterogeneity in these populations (Hughes et al. 2018). The overarching theme when assessing the classifications of Latin American individuals using FORDISC is that the misclassifications of these individuals into Asian reference groups is directly tied to their association with Native American populations through the peopling of America. Conducting analyses of population affinity using this framework does not account for the secular change that is likely to have occurred over the past several centuries.

FORDISC has also been applied to various populations from around the world (L'Abbe et al. 2013; Manthey et al. 2018). South African and Italian populations are some of the groups that have been used to assess the reliability of FORDISC 3.1 in estimating non-United States populations (L'Abbe et al. 2013; Manthey et a. 2018). The applicability of FORIDSC software to South African populations of the 20th century was tested due to the struggle many South African anthropologists face in trying to make an estimate of population affinity. A likely cause of this difficulty is the diversity that is seen in South Africa without proper population-specific methods to address this diversity (L'Abbe et al. 2013). A South African sample was compared to the the FDB, the Howell's database, and a South African database; the Howell's database was used to compare two African groups within the database to the South African sample (L'abbe et al. 2013). The population affinity of the South African population was estimated correctly more often than sex, when compared to the North American sample 73% of South Africans (both black and white) classified correctly, and when compared to the South

African database 71% of South Africans (both black and white) classified correctly (L'Abbe et al. 2013). The misclassifications between the South African sample and the reference groups were most often seen in sex, as black AFABs were more likely to misclassify as black AMABSs but not as white AFABs; the same misclassification pattern was noted for the other sex and population combinations (L'Abbe et al. 2013). The higher rates of correct classifications for population affinity were notable due to the longstanding differences between North American populations and South African populations; however, when assessing the comprehensive statistics, the South African sample classified at a better rate when compared to the South African database than when compared to the FORDISC databases (L'Abbe et al. 2013). A possible explanation for the closeness in classification is that due to the similar social behaviors despite the historical, and regional differences. An example of similar social behavior would be the response to the legal realization of separating racial groups which limited gene flow throughout the past several decades leading to some similarities in population structure (L'Abbe et al. 2013). When using FORDISC to assess the crania of non-US populations, it is important to remember that its performance will likely be poor due to the limited reference groups available in FORDISC, and it is not by fault of the statistical model. When assessing European (Italian) crania using the FORDISC software, it was found that the morphological changes in crania were significant, and it supported claims of secular change throughout time (Manthey et al. 2018). The secular change visible in many of these populations suggests that newer and more population specific reference groups would improve the accuracy and reliability of FORDISC on a grander global scale.

NEW MEXICO DECEDENT IMAGE DATABASE AND COMPUTED TOMOGRAPHY

The sample for this study was collected from the New Mexico Decedent Image Database (NMDID). This database is comprised of CT scans from roughly 15,000 deceased individuals routed through the New Mexico Office of Medical Investigator between 2010 and 2017 (Berry & Edgar 2017; Vidoli 2022). The NMDID allows researchers to take advantage of a much more demographically diverse assemblage, as most of the collections in the United States are primarily adult males (Komar & Grivas 2008). Demographics for each individual were collected through next of kin (NOK) reporting. A series of 60 questions were developed to create demographic profiles for each individual which make up the thirty variables available for researchers to query (Vidoli 2022). The CT scans from the NMDID are regularly used by researchers for a variety of research topics, for access to the database a request must be submitted by the researcher and approved by a board of a minimum of two individuals at the NMDID (Vidoli 2022). The CT scans are available for download once a research proposal has been approved. The NMDID represents very modern forensic cases and therefore is a useful tool for forensic anthropologists aiming to understand the modern variation in U.S. populations. These CT scans are readily available to researchers and provide a broader sample size of modern individuals than is available in most other skeletal collections.

CT scans were originally created in the 1970s as a method of 3D images that are more comprehensive than the 2D images from other radiographic scans (Ali Hasan et al.

2016). In recent years, computed tomography scanners have become more readily available and therefore more commonly used in medicolegal contexts (Richard et al. 2014). The usage of computed tomography allows for the permanent archival of 3D models of skeletal material, this allows for a longer-term preservation (Franklin & Marks 2021). This long-term preservation allows for osteological analysis to be conducted during the period after an investigation as the skeletal material is stored digitally (Franklin & Marks 2021). This method of scanning remains for analysis to preserve the integrity of the artifacts was originally used in bioarchaeology and paleontology as fossils and centuries-old skeletal material is fragile (Giovanetti et al. 2022).

There has been an increase in the use of medical imaging technology in forensic contexts in the past several decades (Baglivo et al. 2013; Giovanetti et al. 2022). Radiographic imaging in forensic contexts allows researchers to analyze structures not visible when the deceased individual is fully fleshed and allows for the remains to be kept fleshed. Radiographs also allow for research to take place on a certain sample from anywhere in the world, thereby reducing the geographic constraints for research (Stull et al. 2014; Vidoli 2022). Between 2000 and 2011 the forensic sciences saw an exponential increase in the amount of research using radiographic imaging, with CT scans being the most frequently employed for forensic contexts (Baglivo et al. 2013). Studies have been conducted to assess the accuracy of craniometric measurements on CT scans versus the original skull, the studies showed that one measurement in consistently showed significant differences and that was Orbital Height (OBH) (Richard et al. 2014). The differences noted between the technology assessment of these measurements and those taken on the physical skeletal material was attributed to the way in which the different thicknesses of CT slices impact the 3D imaging of the remains (Richard et al. 2014). However, CT scans have been proven to be generally reliable methods for collecting measurements (Stull et al. 2014; Richard et al. 2014) and, due to the changes in technology, are now regularly collected on individuals brought into forensic laboratories around the world (Baglivo et al. 2013). CT scans have proven especially useful in research analyzing the estimation methods for the biological profile. Three-dimensional models of CT scans are rendered by researchers which, once cleaned, allow for an optimal surface for collecting morphoscopic and morphometric data (Kelley & Tallman 2022). The use of CT scans allows for less damage to remains as well as larger and more diverse sample sizes which will assist in studies such as this one which aim to develop better understandings of variation in modern human populations.

CONTEMPORARY CRITIQUES OF POPULATION AFFINITY ESTIMATION

Forensic anthropology has long held discourse around the functionality of the facets of the biological profile. The biological profile is comprised of estimations of sex, stature, population affinity, and age. With the change of the social climate in the United States, anthropologists have begun to take a closer look at the benefits and limitations of population affinity estimation (Armstrong-Fumero 2014; Cunha & Ubelaker 2020; DiGangi & Bethard 2021; Flouri et al. 2022; Ousley et al. 2008; Tallman et al. 2020). Some researchers are in support of population affinity inclusion in the biological profile, while others argue that its inclusion perpetuates racialized ideologies in science. Those in favor of population affinity estimation often highlight the repeated success of various methods in estimating the population affinity of different samples of individuals (Ousley et al. 2008). These successes are attributed by many to the geographic patterning that is seen throughout the globe (Ousley et al. 2008). Other supporters of population affinity estimation note its importance in the functionality of methods for estimating other facets of the biological profile (Cunha & Ubelaker 2020). Many methods for estimating, sex, stature, and age, become exponentially more accurate when the individual's population affinity is known, and therefore this is a strong argument for the inclusion of population affinity in the biological profile (Cunha & Ubelaker 2020). However, these arguments in support of population affinity inclusion do not consider the potential damages that may arise from its inclusion.

Other researchers assess the inclusion of population affinity from a more social lens. The inclusion of population affinity in the biological profile, particularly for forensic anthropologists, can result in an unintentional bias in which marginalized individuals are disproportionately negatively affected (Marten et al. 2023). Because of the highly racialized nature of the field of biological anthropology, the classification of individuals into categories based on typological differences and racial hierarchies can perpetuate structural violence against marginalized communities (DiGangi & Bethard 2021; Tallman et al. 2020). The focus on the methodology and applied practices of identification leaves little room for forensic anthropologists to address the holistic impact of estimating an individual's population affinity.

The inclusion of population affinity in the biological profile is a charged topic in the field of forensic and biological anthropology. Its inclusion can perpetuate harmful biases in medicolegal contexts, but it is often viewed as essential for the appropriate usage of other biological profile estimation methods. While there is no clear answer on if population affinity should be included in the biological profile, it is clear that should it be included, extensive rehabilitation of its methods, terminology, and applicability needs to be addressed. The biological profile and population affinity can be helpful tools for biological anthropologists in providing an estimate of an unidentified individual's identity during life, however, a holistic approach is necessary to ensure that no biases are being perpetuated.

MATERIALS AND METHODS

This study analyzed craniometric data collected from three-dimensional digitized models of CT scans of five general ancestral groups. The intention of this study was to assess the accuracy of the FORDISC software, and to determine how it classifies a sample of modern individuals from the United States (Vidoli 2022). Historically, FORDISC has been analyzed to assess in accurately classifying population affinity and sex of different groups (Manthey et al. 2018; Dudzik & Jantz 2016; Guyomarc'h and Bruzek 2011; Hughes et al. 2019). Most research conducted to address the concern of population affinity accuracy has been done on populations of Hispanic ethnicity or Latin American descent (Dudzik & Jantz 2016; Hughes et al. 2019). It is hypothesized that Asian American individuals will be classified with less accuracy, and that Asian American AFABs specifically will show the lowest levels of accurate classifications. The intent of this study is to shed light on the functionality of a widely used software in modern contexts.

Study Sample

A randomized sample of 70 adult individuals were collected for five different demographic groups using the identification by NMDID "Investigator Report". The five groups were Asian American, African American, European American, Native American, or Latin American. These groups were selected from the NMDID to represent the most common classifications of population affinity seen in the United States (Monte & Shin 2022). As indicated by Kelley and Tallman (2022), these assigned titles and population groupings are not all encompassing of the variation seen in population affinity groups.

The Asian population affinity group within the NMDID was small and therefore it was not possible to further narrow down these groups past the five most common demographic groups in the United States. The sample of individuals that represents the Asian-American population affinity group are coded in the NMDID as “races”. Individuals identified as Korean (n=12), Filipino (n=4), Chinese (n=7), Asian Indian (n=18), Japanese (n=8), Vietnamese (n=14), and Other Asian (n=37) with a specification of “Not Hispanic or Latino” in the ethnicity query were included in the Asian-American population affinity group. The Latin American sample are comprised of individuals whose “race” is identified as Hispanic (n=486), Other (n=3), and Unknown (n=1), all with the specification of Hispanic or Latinx ethnicity (Table 1). Native American (NMDID title: Native American), European American (NMDID title: White), and African American (NMDID title: Black/African American) are all also classified as “races” and all groups applied the specification filter of “Not Hispanic or Latino” ethnicity (Table 1). Birth country for all groups was not specified due to the small sample sizes that come from restricting searches to only individuals born inside the United States; however, the majority of individuals in the NMDID with a listed birth location were born in the U.S. (Vidoli 2022).

An approximately equal number of assigned fenale at birth (AFAB) and assigned male at birth (AMAB) individuals in each demographic group were selected, with 35 AFABs and 35 AMABs selected for each group. The selection criteria of ethnicity, population affinity, and sex were used to access a list of potential individuals which were then chosen using a random number generator through the Google search engine based on the number of individuals available in each group. For example, if a group had 300 AFAB individuals to select from the random number generator was set to 300 and processed, once a number was selected by the generator the investigator counted down the list and added the individual at that place in the list. This process was repeated for 35 individuals. This selection process was not possible for the Asian American demographic group due to the small number of AFAB individuals and therefore all AFABs in the Asian American demographic group were selected.

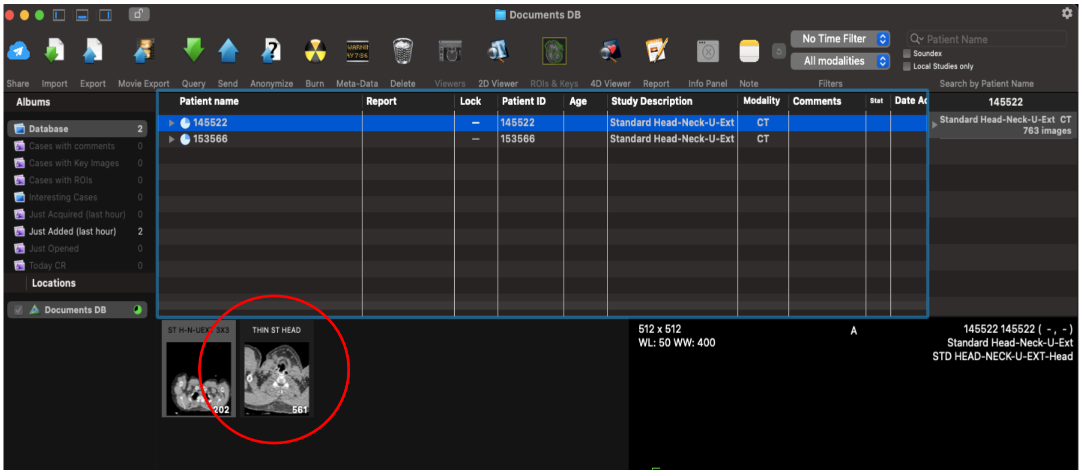
The original number of scans collected was 350, however, due to various constraints including technology limitations and imaging quality, the final number of individuals included in the sample was 148. The final samples for each group excluded scans of individuals under 18 years of age at death and scans missing crania (Appendix Table A.1). The number of individuals in each demographic group were obtained in Excel (Table 1). Of the 148 individuals analyzed, 16% were identified by NMDID investigators as Asian (Other Asian, Asian Indian, Vietnamese, Japanese, Chinese, and Korean), 26% were identified as European American (White), 24% were classified as African American (Black or African American), 20% were identified as Native American, and the remaining 15% were classified as Latin American (Hispanic). The results of the FORDISC analyses (Table 1) estimated 25% of individuals as White (European American), 24% as Black (African American), 18% as Hispanic (Latin American), 13% as Guatemalan (Latin American), 9% as Native American, 5% as Japanese (Asian American), 5% as Vietnamese (Asian American), and 0.7% as Chinese (Asian American).

Table 1. Final number of individuals (n) and corresponding percentages of the sample collected from the NMDID.

Demographic Group		**%**	**n**	**TOTAL**
African American	AMAB	11%	17	35
	AFAB	12%	18	
Asian American	AMAB	9%	14	23
	AFAB	6%	9	
European American	AMAB	14%	20	38
	AFAB	12%	18	
Latin American	AMAB	7%	10	22
	AFAB	8%	12	
Native American	AMAB	9%	14	30
	AFAB	11%	16	
TOTAL		100%	148	148

Analysis of Scans: OsiriX

Each scan retained its identification number provided by the NMDID so that measurements and data could be associated with the scan for statistical analysis. Due to the large number of scans being analyzed and the complexity of the ID numbers, there was little concern for investigator bias as it would be difficult to accurately remember the population affinity associated with each six-digit ID number. CT scans were downloaded from the NMDID into the DICOM viewer, Osiri-X for further analysis. Once each scan was uploaded into the DICOM viewer, the “Thin-Bone Head” scan setting was selected for surface rendering as it was confined to the cranium and upper thoracic sections limiting the amount of excess imaging needed to be removed. This slice of the CT scan was often selected as it offered the clearest image, in certain cases different slices were used due to clarity or visibility of the cranium (Figure 2). While previous studies have noted the functionality of volume renders collected from DICOM viewers (Kelley & Tallman 2022), the surface rendering tool was used to create a file that could be transferred to a 3D rendering software (Figure 3). To ensure the best quality of imaging, the settings were adjusted based on each individual scan to allow for visualization of sutures and cranial landmarks. In most cases this involved adjusting the pixels value and the resolution of the images. A consistent set back presented itself as there was frequent interference from metal. For several scans, the 3D surface rendered image was impacted by the metal table upon which the decedents were scanned, as well as any metal items on the decedents themselves such as jewelry, dental hardware, or medical interventions like EKG leads. The metal interfered with the external structure of many of the skulls resulting in a wavy, washboard effect on the cranial vault. In certain circumstances when there was metal in or around the dentition of the maxilla and mandible, spiky starburst projections were visible on the renderings in those areas. The unexpected excess materials impacted the resolution and clarity of certain images, if the images were too severely impacted, they were removed from the study entirely.



Patient name	Report	Lock	Patient ID	Age	Study Description	Modality	Comments	Stat	Date Ac
145522		–	145522		Standard Head-Neck-U-Ext	CT			
153566		–	153566		Standard Head-Neck-U-Ext	CT			

145522

Standard Head-Neck-U-Ext CT
763 images

512 x 512
WL: 50 WW: 400

145522 145522 (- , -)
Standard Head-Neck-U-Ext
STD HEAD-NECK-U-EXT-Head

Figure 2. Home page of OsiriX Dicom viewer. The “Thin ST Head” (red circle) was selected to gather the most accurate image.

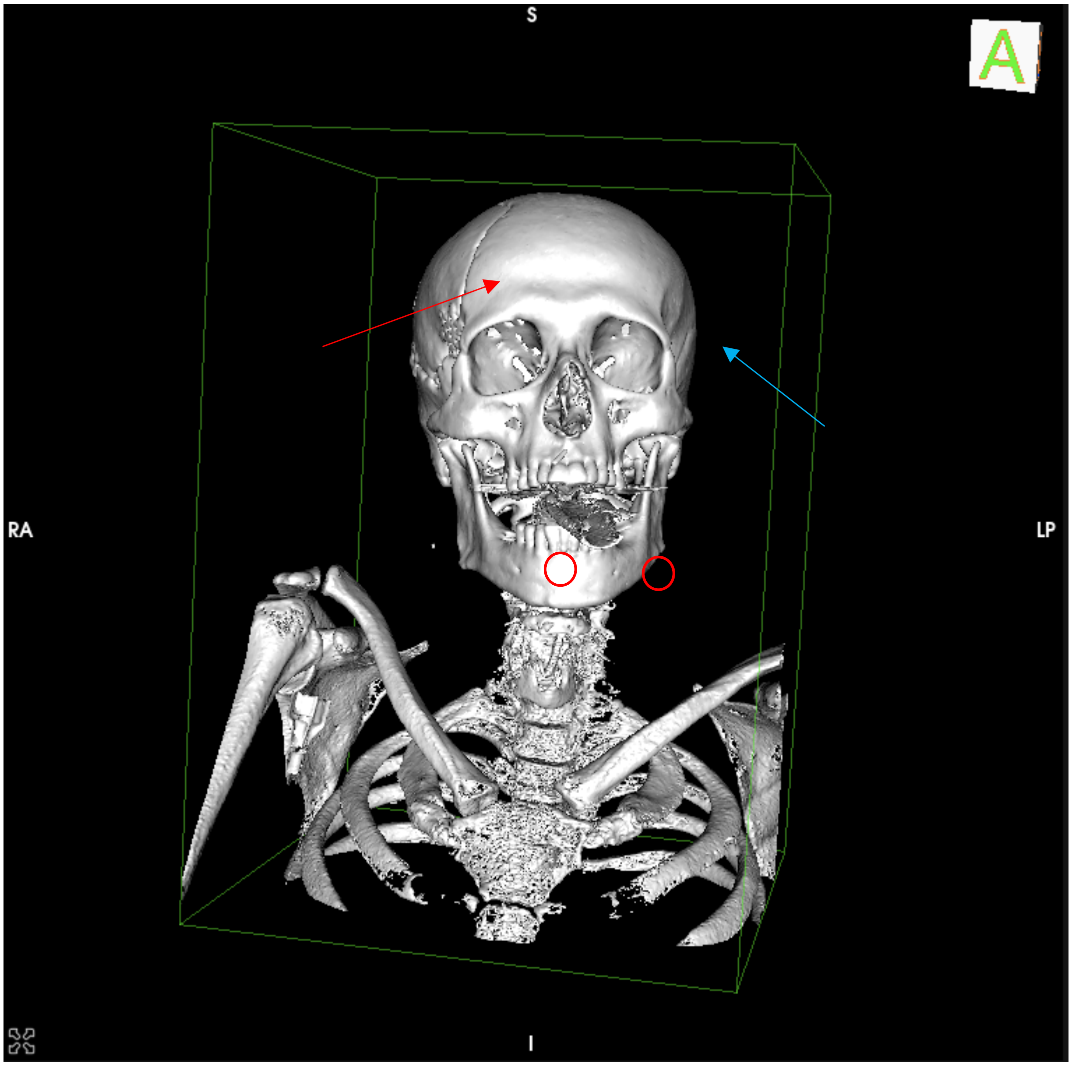


Figure 3. A 3D Surface rendering of an individual in the sample. Fractures (red arrow), sutures (blue arrow), and foramina (red circle) are visible allowing for the most accurate measurements.

Analysis of Scans: MeshMixer

Once a surface rendering was created in the DICOM viewer, 3D digital models of the scans were created using MeshMixer. The MeshMixer software allows for surface or volume renderings to be turned into 3D models for optimization of the image, analysis, and other functions as well. After uploading the scans into the software, the "Select" feature was used to isolate the skull and remove any excess or unwanted skeletal and medical content (Figure 3). Clusters of red dots were often visible on 3D renderings in MeshMixer, and were indicative of excess material that was likely made visible because of interference from metal. As much material was removed as possible to provide a clear view of the craniometric landmarks necessary for measuring, this included other skeletal elements, the table upon which they were scanned, medical intervention, jewelry, etc. (Figure 4). Each individual required different amounts of cleaning and therefore different selection tools were used for each individual; the two options were the “brush” tool and the “lasso” tool. The lasso was much more effective at removing large areas of material, and the brush was helpful for more specific areas that required more precision (Figure 5). After individuals were processed through Osiri-X and MeshMixer, approximately 43% of the sample remained. Of the initial 350 individuals selected for analysis, 202 were excluded due to either complications with the scans, or problems with the integrity of the skull such as fractures or surgical intervention. The final number of scans for analysis was 148 and was comprised of 75 AMABs and 73 AFABs which allowed for an approximately even distribution of AFABs and AMABs for analysis (Table 2, Appendix Table A.1).

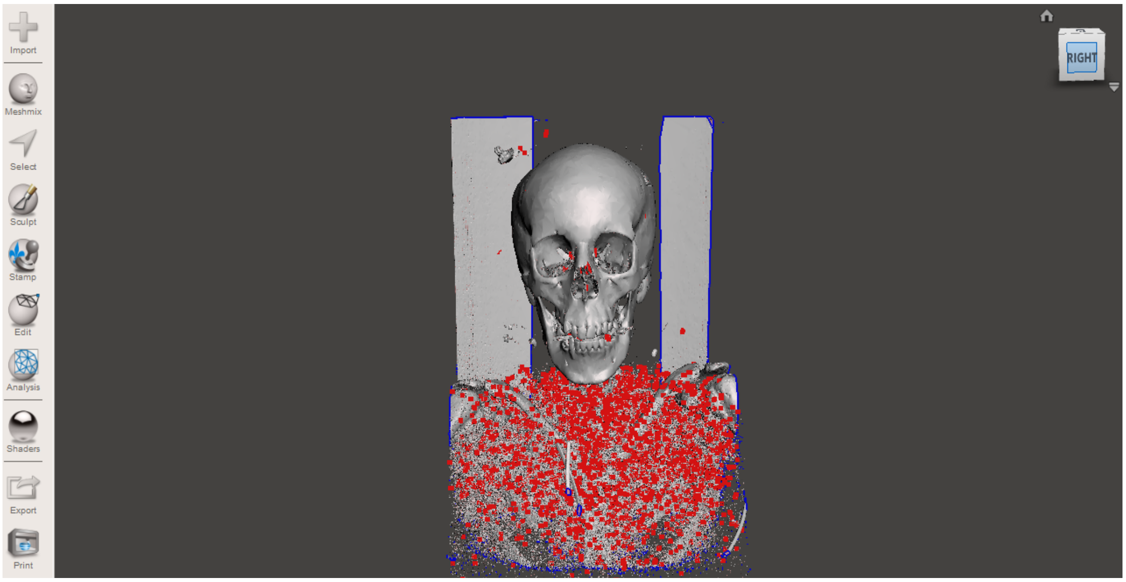


Figure 4. 3D Surface Rendering of individual from the sample in the MeshMixer software. Red dots indicate clusters of material that were later removed.

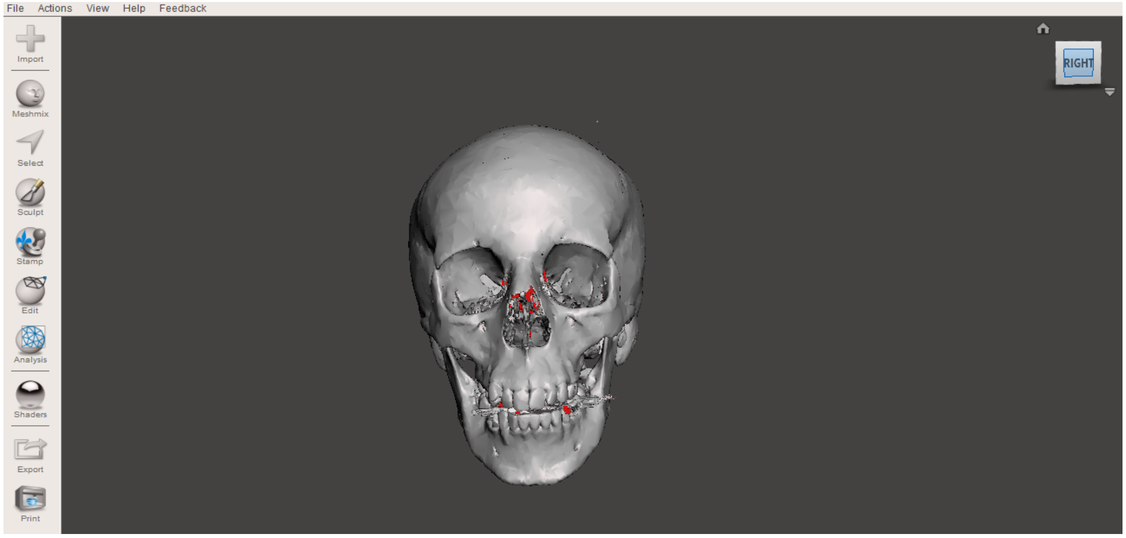


Figure 5. 3D Model of skull of individual from the sample with excess skeletal material removed. “Select” feature (red circle) was used to remove excess material, and the “Analysis” (black arrow) feature was used to measure between craniometric landmarks.

Table 2. The number of individuals (n) and corresponding percentages in the final analysis using the classifications by the NMDID. These population affinity groups were then combined to create the five larger demographic groups.

Population Affinity	**Count**		**Percentage**
African American	AMAB	17	11.5%

	AFAB	18	12.2%
Asian Indian	AMAB	4	2.7%
	AFAB	2	1.4%
Chinese	AMAB	1	0.7%
	AFAB	0	0%
Japanese	AMAB	1	0.7%
	AFAB	2	1.4%
Korean	AMAB	1	0.7%
	AFAB	0	0%
Latin American	AMAB	10	6.8%
	AFAB	12	8.1%
Native American	AMAB	14	9.4%
	AFAB	16	10.8%
Other Asian	AMAB	4	2.7%
	AFAB	5	3.4%
Vietnamese	AMAB	3	2.0%
	AFAB	0	0%
White	AMAB	20	13.5%
	AFAB	18	12.2%
Total	AMAB	75	51%
	AFAB	73	49%
	Both	148	100%

The MeshMixer software allowed measurements to be taken according to the DCP 2.0 (Langley et al. 2016) standards using the "Units/Dimension" feature under the "Analysis" tab in MeshMixer (Figures 4 and 5). This provided measurements in millimeters and allowed for precision in marking locations on the cranium. A total of 28 cranial osteometric measurements were taken on each cranium and mandible (Table 3, Figures 6 and 7). While Langley et al. (2016) outline 37 cranial measurements for analysis, not all are included in the FORDISC software (Ousley & Jantz 2005a) and therefore not all 37 were measured. Certain measurements were excluded for certain individuals, based on the visibility of cranial landmarks and sutures necessary to take accurate measurements. Measurements such as Basion-Bregma Height (BBH), Basion-Nasion Length (BNL), Basion-Prosthion Length (BPL), Foramen Ovale Length (FOL), Foramen Ovale Breadth (FOB), and Maxillo-Alveolar Length (MAL) were difficult to obtain due to the positioning of the remains when they were scanned; the cranial landmark basion and opisthion could often be blocked by the first and second cervical vertebrae, while alveolon was blocked by the hyoid or excess material that could not be removed. Mandibular measurements such as Mandibular Body Breadth, Maximum Ramus Height, Mandibular Length, and Mandibular Angle, were not recorded because the standard measurement collection process for these measurements requires a mandibulometer, which was not reliably replicable with the MeshMixer Units/Dimensions feature. Edentulous individuals did not have any mandibular or maxillary measurements collected as the morphology of the dental arcade is altered when teeth are lost antemortem (Dechow et al. 2010). Maxillo-Alveolar Breadth (MAB) was often impacted by metal dental hardware or absent teeth which impacted the visibility of ectomolare (ecm). Biasterionic Breadth (ASB), Bimaxillary Breadth (ZMB), and Zygoorbitale Breadth (ZOB) were not taken on a large majority of the sample as the craniometric landmarks asterion (ast), zygomaxillare anterior (zma), and zygoorbitale (zo) were consistently too difficult to identify on the scans. Frontal Chord (FRC), Parietal Chord (PAC), and Occipital Chord (OCC) were difficult to measure due to lack of suture visibility at bregma, and lambda. Measurements were collected in a spread sheet with the correlated scan number to maintain consistency. Any measurements that were not observable were marked with “N/A” and were not included in the FORDISC analysis.

Table 3. List of measurements taken on each individual in the sample as outlined by Langley et al. (2016; p. 65-72) with both the acronym and full title of each measurement provided.

Measurements	
GOL	g-op; Maximum Cranial Length
XCB	eu-eu; Maximum Cranial Breadth
ZYB	zy-zy; Bizygomatic Breadth
BBH	ba-b; Basion-Bregma Height
BNL	ba-n; Cranial Base Length
BPL	ba-pr; Basion-Prosthion Length
MAB	ecm-ecm; Maxillo-Alveolar Breadth
AUB	ra-ra; Biauricular Breadth
NPH	n-pr; Nasion-Prosthion Height
WFB	ft-ft; Minimum Frontal Breadth
FMT-FMT	fmt-fmt; Upper Facial Breadth
NLH	Nasal Height
NLB	Nasal Breadth
OBB	d-ec; Orbital Breadth
OBH	Orbital Height
EKB	ec-ec; Biorbital Breadth
DKB	d-d; Interorbital Breadth
FRC	n-b; Frontal Chord
PAC	b-l; Parietal Chord
OCC	l-o; Occipital Chord
FOL	Foramen Magnum Length
FOB	Foramen Magnum Breadth
MDH	Mastoid Height
ID-GN	Chin Height
Mandibular Body Ht	Height of Mandibular Body
GO-GO	go-go; Bigonial Breadth
CDL-CDL	cdl-cdl; Bicondylar Breadth
Minimum Ramus Brdth	Minimum Ramus Breadth

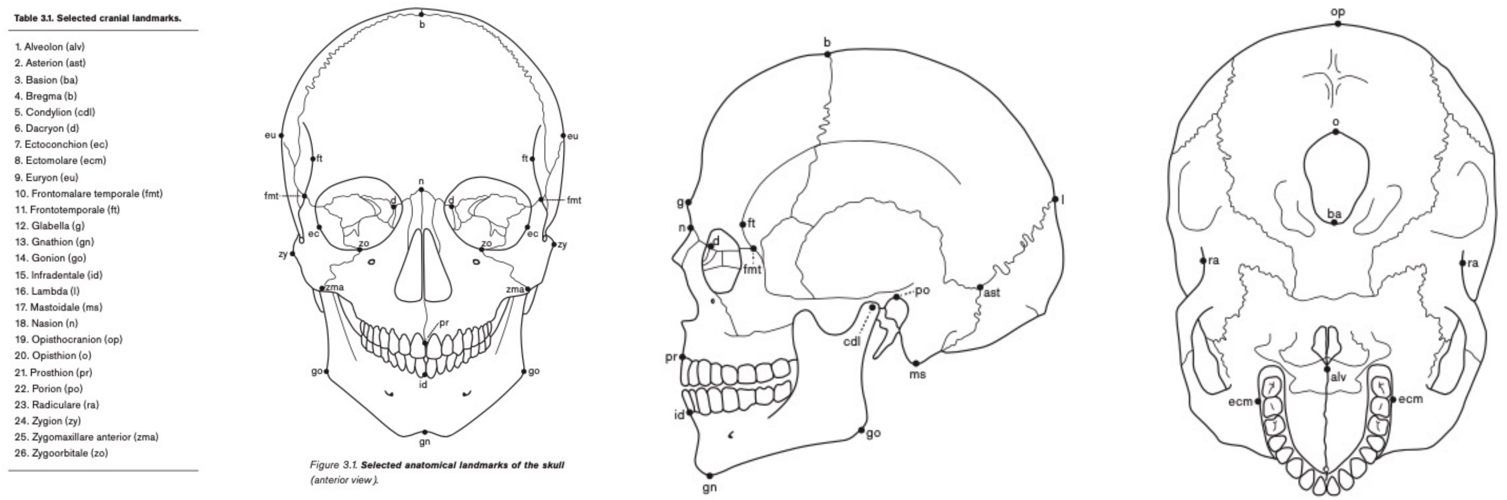


Table 3.1. Selected cranial landmarks.

1. Alveolon (alv)
2. Asterion (ast)
3. Basion (ba)
4. Bregma (b)
5. Condylion (cdl)
6. Dacryon (d)
7. Ectoconchion (ec)
8. Ectomolare (ecm)
9. Euryon (eu)
10. Frontomalare temporale (fmt)
11. Frontotemporale (ft)
12. Glabella (g)
13. Gnathion (gn)
14. Gonion (go)
15. Infradentale (id)
16. Lambda (l)
17. Mastoidale (ms)
18. Nasion (n)
19. Opisthocranion (op)
20. Opisthion (o)
21. Prosthion (pr)
22. Porion (po)
23. Radiculare (ra)
24. Zygion (zy)
25. Zygomaxillare anterior (zma)
26. Zygoorbitale (zo)

Figure 3.1. ***Selected anatomical landmarks of the skull*** *(anterior view).*

Figure 6. Mapping of craniometric landmarks as outlined by the DCP 2.0 (Langley et al. 2016, p.

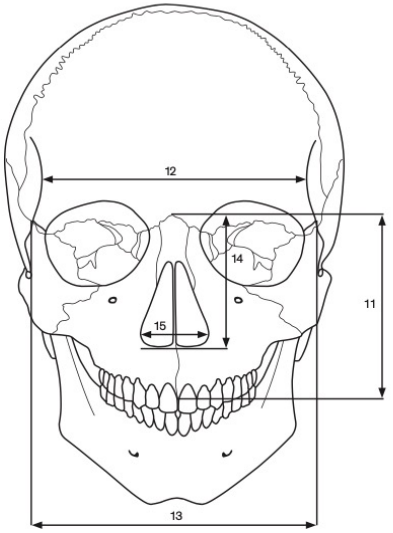


Figure 3.6. **Selected cranial measurements** *(anterior view).*

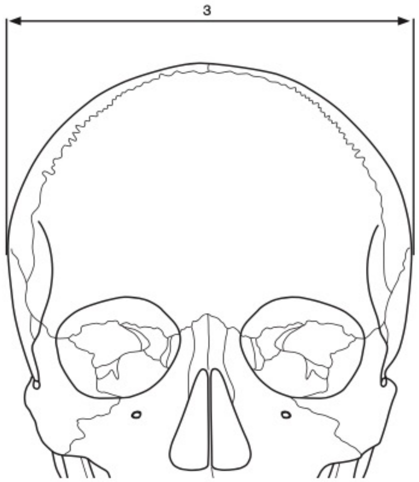


Figure 3.4. **Maximum Cranial Breadth**

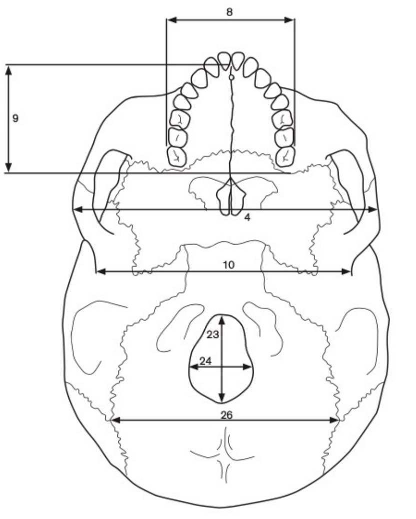


Figure 3.5. **Measurements taken with the skull resting on bregma with the base facing up.**

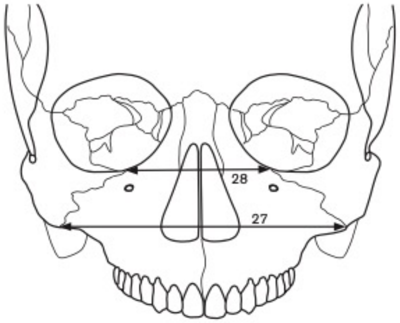


Figure 3.10. **Bimaxillary breadth and zygoorbitale breadth.**

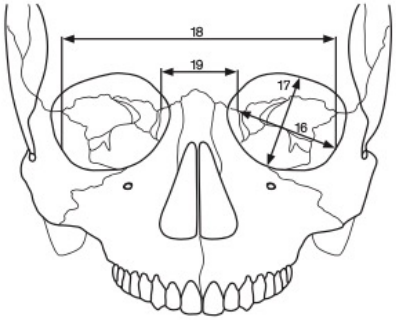


Figure 3.7. **Selected measurements of the orbital region.**

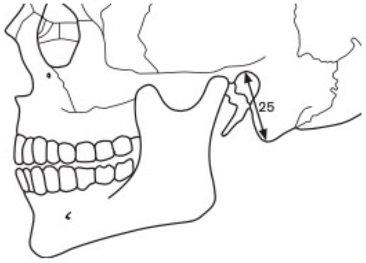


Figure 3.8. **Mastoid height.**

Figure 7. Craniometric measurement guidelines as described and depicted by the DCP 2.0 (Langley et al. 2016, p. 66-70).

FORDISC 3.1 Analysis

Measurements were entered into the *FORDISC* 3.1 software and processed with all thirteen reference groups included in the first analysis. Each individual was processed through *FORDIS*C more than once; each round in the software involved removing any reference groups with a posterior probability below 0.005 and then the measurements were re-run with the new number of reference groups. Measurements were processed until all the remaining reference groups produced a posterior probability of 0.005 or greater. The posterior probability and F-Typicality were recorded in a separate spreadsheet for each FORDISC reference group. Once a classification was provided by *FORDISC* 3.1, the scan number was entered into the NMDID search engine to collect the data on the investigator and Next of Kin classifications. In cases where a Next of Kin identification was not available, "N/A" was entered into the spreadsheet, this was not a criterion for selection because in adding this designation it severely limited the number of potential individuals for the sample. *FORDISC* 3.1 classifications were shorthanded using most of the *FORDISC* 3.1 nominations to allow for clarity: AIF (American Indian AFAB), AIM (American Indian AMAB), BF (Black AFAB), BM (Black AMAB), CHM (Chinese AMAB), GTM (Guatemalan AMAB), HF (Hispanic AFAB), HM (Hispanic AMAB), JF (Japanese AFAB), JM (Japanese AMAB), VM (Vietnamese AMABs), WF (White AFAB), WM (White AMAB). Similar abbreviations were developed for the NMDID Next of Kin and Investigator classifications: AAF (African American AFAB), AAM (African American AMAB), AIF (Asian Indian AFAB), AIM (Asian Indian AMAB), HF (Hispanic AFAB (NOK)), HM (Hispanic AMAB (NOK)), NAF (Native

American AFAB), NAM (Native American AMAB), OAF (Other Asian AFAB), OAM (Other Asian AMAB), KM (Korean AMAB), JF (Japanese AFAB), JM (Japanese AMAB), VM (Vietnamese AMABs) WF (White AFAB), WM (White AMAB), WF (HISP) (White AFABs with Hispanic Ethnicity), WM (HISP) (White AMABs with Hispanic Ethnicity). The observations were then further separated so that sex and population affinity could be analyzed as individual variables.

Statistics Conducted on FORDISC 3.1 Estimates

Descriptive statistical analyses such as frequency distributions, Chi Square Tests, and Standardized Residual tests were conducted using Microsoft Excel. To distinguish between the Next of Kin classifications and those by the NMDID, the ones by the NMDID are called the “NMDID Investigator” classifications. Once sex and population affinity were separated for FORDISC, and NMDID Investigator, the individual columns were highlighted, and frequency distributions were run using the “Analyze Data” function on Excel. This was repeated for each variable (ex. FORDISC Population Affinity, FORDISC Sex, etc.) until the counts for each group were obtained. To run the appropriate statistical analyses, the classifications obtained from FORDISC, and the NMDID Investigator reports were adjusted to represent the five major demographic groups of the United States. The African American (AA) demographic group was comprised of Black, and African American classifications; the Asian American (AS) group was comprised of Japanese, Chinese, Korean, Vietnamese, Asian Indian, and Other Asian classifications; the European American (EA) group was made up of White classification; the Latin American (LA) group was made of Hispanic, White with

Hispanic Ethnicity, and Guatemalan classifications; and the Native American (NA) demographic group was comprised of Native American and American Indian classifications.

Separate Chi Square tests were conducted for population affinity and sex between FORDISC and NMDID Investigator Report. These tests were conducted using Excel by using the CHISQ.TEST function and referring to the cells that contained the appropriate data. Degrees of freedom were determined by subtracting one from the total number of categories, which in the case of FORDISC vs Investigator Report the degrees of freedom was 4 for population affinity and 1 for sex. The Chi-Square tests were conducted with an alpha value of 0.05, and therefore any p-values that were calculated to be below 0.05 were considered statistically significant. Once it was determined that the results were statistically significant, individual Standardized Residual tests were run on each population affinity and sex category. This test is conducted to measure outliers as well as to measure the strength of differences between the expected and observed values. Standardized residuals greater than 2 or less than -2 are considered statistically significant. The statistically significant standardized residuals indicate that those categories are the cause of the majority of the differences observed between the observed and expected data.

Intraobserver Error Analysis

Measurements were taken twice on random 10% of the study sample to explore intraobserver error; however, previous research has demonstrated that measurements from CT scans are generally consistent with measurements on skeletal material (Stull et al. 2014; Baglivo et al. 2013; Richard et al. 2013). Intraclass correlation coefficients (ICC) were conducted for every measurement collected to test the reliability of the investigator between the first and second rounds of measurements. The intraobserver error sample consisted of 14 individuals, and each individual had two columns of data per measurement (GOL 1st, GOL 2nd, XCB 1st, XCB 2nd, etc.). This data was imported into SPSS and measurements were changed from "string" variables to "numeric" and the variable measure was changed from "nominal" to "scale", this allowed the software to view all values as numeric so that the ICC could be conducted. Any non-numeric values such as "N/A" for non-observable measurements became dots in the system that were bypassed when running the tests. To run the ICC the investigator selected "Analyze" from the SPSS menu, then selected "scale" and "reliability analysis". The intraclass correlation coefficient statistics were selected under the "statistics" tab in the new pop-up screen and a two-way mixed model and consistency type were used. The intraobserver error rates were taken to ensure reliability of the researcher in taking the same measurements on a smaller portion of the sample.

RESULTS

The purpose of this study was to assess the reliability of FORDISC 3.1 in assessing modern United States populations. Individuals of five separate demographic groups were analyzed using the measurement collection standards outlined by Langley et al. (2016). The intention of the author in conducting this research was to assess how accurately FORDISC is able to classify modern US populations, particularly Asian American individuals, due to the minimal research on Asian American population affinity estimation which has left a gap in knowledge for anthropologists handling modern populations. The goal of this research is to investigate whether modern Asian American individuals are accurately classified by FORDISC, and to address potential conflation between different demographic groups. The FORDISC software is limited in its function due to the few reference groups available for comparison, this study addresses the need for more population specific population affinity estimation methods as well as updates to the current systems in use.

Cross-Tabulations: Accuracy for Each Demographic Group

Cross-tabulations were used to assess the number of individuals in each demographic group that were correctly identified by FORDISC. Of the 148 individuals, 36% had their population affinity accurately estimated by FORDISC, and 64% did not. Sex was classified correctly for 80% of the 148 individuals, and incorrectly for 20% of the individuals (Table 4). The two groups with the lowest number of correct classifications were Asian Americans and Native Americans. Of the 23 individuals classified as Asian American, only three were classified correctly by FORDISC, and of the 30 individuals classified as Native American, FORDISC only classified four correctly (Table 5). Within these two demographic groups AFABs were never classified correctly by FORDISC. AMABs were classified correctly for only 20% of the Asian American demographic group, and for 29% of the Native American group. When assessing the misclassifications of Native American individuals, it was noted that the most common classification for individuals of this demographic group was Latin American (Hispanic). Asian American individuals were more likely to misclassify as Latin American (35%) and European American (30%) than any other group (Table 6). Asian American individuals were more likely to classify as Guatemalan males than any other Latin American classification. Other non-Asian population affinity groups that Asian American individuals were classified incorrectly into included African American (13%) and Native American (9%) (Table 6). The groups with the highest number of accurate classifications were the African American demographic group, with 71% of AMABs accurately classified, and 67% of AFAB cases accurately classified.

Individuals in the sample were classified as European American AMABs by FORDISC more often than any other group with 14% of cases in the sample classified as European American AMABS, followed by Guatemalan AMABs, and then African American AFABs both making up 13% of the classifications. The group with the fewest FORDISC classifications was Chinese Males which made up 0.7% of the FORDISC classifications. Certain individuals were identified by NMDID investigators as Korean, Other Asian, or Asian Indian, three categories which are not accounted for in FORDISC reference groups and therefore did not have a direct match in FORDISC. This limitation of the reference groups available in FORDISC most likely had an impact on the percentage of Asian American individuals in the sample that were accurately classified (Table 7). The results of this study suggest that individuals of Asian American population affinity, particularly AFAB individuals are consistently misclassified by FORDISC. The cross-tabulations of accuracy suggest that FORDISC is a stronger at estimating sex than population affinity in a modern US sample.

Table 4. Accuracy of classifications of FORDISC estimates when compared to the NMDID Investigator reported classifications for the entire sample.

	Population Affinity		**Sex**	
Classified correctly by FORDISC	36%	53	80%	118
Classified incorrectly by FORDISC	64%	95	20%	30
Total	100%	148	100%	148

Table 5. Accuracy of classifications of FORDISC estimates compared to NMDID Investigator classifications for each demographic group.

Population affinity	**Classified correctly**		**Classified incorrectly**		**Total**
African American	69%	24	31%	11	100%
Asian American	13%	3	87%	20	100%
European American	42%	16	58%	22	100%
Latin American	41%	9	59%	13	100%
Native American	13%	4	87%	26	100%

Table 6. Accuracy of FORDISC population affinity estimates compared to NMDID Investigator classifications by sex for each demographic group.

Demographic group		**Classified correctly**		**Classified incorrectly**	
African American	AMAB	71%	12	29%	5
	AFAB	67%	12	33%	6
Asian American	AMAB	20%	3	80%	12
	AFAB	0%	0	100%	8
European American	AMAB	40%	8	60%	12
	AFAB	44%	8	56%	10
Latin American	AMAB	36%	3	64%	7
	AFAB	25%	3	75%	9

Native American	AMAB	29%	4	71%	10
	AFAB	0%	0	100%	16
TOTAL			53		95

Table 7. Breakdown of number of individuals (n) and corresponding percentages of how FORDISC classified the 23 Asian American individuals in the sample. Most Asian American individuals were classified by FORDISC as Latin American or European American.

Demographic group	%	n
African American	13%	3
Asian American	13%	3
European American	30%	7
Latin American	35%	8
Native American	9%	2
Total	100%	23

Frequency distributions were used to compare how FORDISC functioned in estimating sex and population affinity of a sample to the expected classifications by the NMDID reports, it was not used to assess accuracy. Of the 148 individuals in the sample, FORDISC classified an equal number of the sample as African American individuals as is present. FORDISC classified only 9% of the sample to be Native American, when in reality it comprised 20% of the sample, and the software classified 31% of the sample to be Latin American even though this demographic group only made up 15% of the actual sample. These frequency distributions show the severe under-classification of individuals into the Native American demographic group, and over classification of individuals into the Latin American demographic group.

Of the 148 individuals in the sample, 51% were reported as AMAB by the NMDID Investigator report, and 49% were reported as AFAB. FORDISC classifications for sex showed 66% of individuals were estimated as AMABs and 34% individuals were estimated as AFABs (Tables 8 and 9). There was a significant overclassification by FORDISC of individuals in the sample as AMABs compared to the actual composition of the sample. While frequency distributions do not indicate accuracy, when assessed using a Chi-Square analysis, they can indicate how differently FORDISC classifies the sample in comparison to the NMDID Investigator classifications and NoK identifications.

Table 8. A comparison of the frequency distributions of the number of individuals (n) and corresponding percentages of each demographic group for NMDID Investigator classification, and FORDISC classification of the same sample.

Population affinity	NMDID Investigator classification		FORDISC classification	
	n	%	n	%
African American	35	24%	36	24%
Asian American	23	16%	16	11%
European American	38	26%	37	25%
Latin American	22	15%	46	31%
Native American	30	20%	13	9%
Total	148	100%	148	100%

Table 9. Frequency distribution of the number of individuals (n) and percentages of each sex for NMDID Investigator classification, FORDISC classification, and NoK classification in the sample.

Sex	NMDID Investigator classification		FORDISC classification	
	n	%	n	%
AMAB	75	51%	97	66%
AFAB	73	49%	51	34%
Total	148	100%	148	100%

Chi-Square Analyses: FORDISC vs Investigator Report

The results of the Chi-Square tests that were conducted on population affinity between FORDISC and NMDID Investigator reported classifications provided a *p*-value of 1.12031E-07 and therefore are considered statistically significant at a 95% confidence interval (Table 10). This indicates that there is a significant difference in how FORDISC classified the individuals in the sample compared to the expected classification of the individuals based on the NMDID investigator report. Standardized residual tests (Table 7) were conducted to determine which of the demographic groups exhibited most of the differences between the observed (FORDISC) and expected (Investigator report) classifications. Standardized residual values greater than 2 or less than -2 are considered statistically significant. Native American and Latin American population affinity groups resulted in standardized residual values of -3.1 and 5.1 respectively (Table 10). Because these values are greater than 2 or less than -2, they are considered statistically significant and are interpreted as the cause of a majority of the differences between the NMDID investigator report and the FORDISC classification. European American (-0.2), and African American (0.2), both had standardized residuals within the range of not statistically significant, and therefore these classifications are likely not responsible for the differences between FORDISC and NMDID Investigator classifications. The Asian American population affinity group had a standardized residual value of -1.5, and although this is not statistically significant, it is quite close to the accepted threshold of -2. Therefore, with a larger sample this value may in fact be statistically significant and is something that should be considered in future research.

Chi-Square and standardized residual tests were also conducted on sex. The differences between the NMDID investigator reports of sex and those observed using FORIDSC were statistically significant at a 95% confidence interval with a p-value of 0.0002979 (Table 11). Once statistical significance was determined, standardized residual equations were used to determine which of the two groups were responsible for the majority of the differences calculated by the Chi-square test. In the case of the individual sex characteristic, both categories of male (2.5) and female (-2.6) almost equally contributed to the differences.

To better understand if there were particular population affinity and sex combinations that were presenting with higher differences than others, the paired sex and population affinity estimates were analyzed using the same tests as before. Male and female classifications of European American, African American, Native American, Asian American, and Latin American were analyzed. The Chi-square test resulted in a *p*-value of 1.62493E-14 which is statistically significant at a 95% confidence interval (Table 12). Once it was determined that the differences between the observed and expected values for these categories were statistically significant, standardized residual equations were conducted to discover which groups displayed the most diversity between FORDISC estimates and the NMDID Investigator reported identification. The groups with statistically significant standardized residual values were Native American females (-3.5), Asian females (-2), and Latin American males (8.2) (Table 9). These scores suggest that Latin American males make up most of the differences seen between FORDISC estimates and NMDID Investigator reported scores.

Table 10. Chi-square test and standardized residuals with percentage and counts (n) of each demographic group used. Standardized residuals less than -2 or greater than 2 are considered statistically significant.

Demographic Group	NMDID Investigator		FORDISC		Chi-Square	Standardized Residuals
	%	n	%	n		
African American	24%	35	24%	36	0.03	0.2
Asian American	16%	23	11%	16	2.1	-1.5

European American	26%	38	25%	37	0.03	-0.2
Latin American	15%	22	31%	46	26.2	5.1
Native American	20%	30	9%	13	9.6	-3.1
Total	100%	148	100%	148		
				Degrees of freedom		4
				p-value		1.12031E-07

Table 11. Chi-square test and standardized residuals with percentage and counts (n) of each sex. Standardized residuals less than -2 or greater than 2 are considered statistically significant.

Sex	NMDID Investigator		FORDISC		Chi-Square	Standardized Residuals
	%	n	%	n		
AMAB	51%	75	66%	97	6.5	2.5
AFAB	49%	73	34%	51	6.6	-2.6
Total	100%	148	100%	148		
				Degrees of freedom		1
				p-value		0.000297913

Table 12. Chi-square test and standardized residuals of each sex within each demographic group with percentages and counts (n). Standardized residuals less than -2 or greater than 2 are considered statistically significant.

Demographic group		NMDID Investigator		FORDISC		Chi-Square	Standardized Residuals
		%	n	%	n		
African American	AMAB	11%	17	11%	17	0	0
	AFAB	12%	18	13%	19	0.06	0.3
Asian American	AMAB	9%	14	9%	13	0.07	-0.3
	AFAB	6%	9	2%	3	4	-2.0
European American	AMAB	14%	20	14%	20	0	0
	AFAB	12%	18	11%	17	0.06	-0.2
Latin American	AMAB	7%	10	24%	36	67.6	8.2
	AFAB	8%	12	7%	10	0.3	-0.6
Native American	AMAB	9%	14	7%	11	0.6	-0.8
	AFAB	11%	16	1%	2	12.3	-3.5
Total		100%	148	100%	148		
					Degrees of freedom		9
					p-value		1.62493E-14

Next of Kin vs FORDISC classification

Frequency distributions were conducted on the individuals with NoK classifications. A total of 35 individuals had NoK classifications, and of these 80% were NoK identified as Hispanic, 17% as White, and 3% as Native American (Table 10). Of the 35 individuals, 49% were identified by NoK as AMABs, and 51% as AFABs (Table 10). FORDISC classified 29% of these 35 individuals as Hispanic, 34% as White, 11% as African American, 9% as Native American, 9% as Guatemalan, 5% as Japanese, and 3% as Vietnamese (Table 13). Of the 35 individuals, FORDISC classified 60% as AMABs, and 40% as AFABs (Table 14).

To assess if there was a significant difference between the NoK identification and the FORDISC classifications, Chi-square tests were conducted on sex and population affinity separately. Due to the small sample size of individuals with NoK idenfiticaion, a different grouping method was selected for population affinity. Individuals were grouped into "Latin American", "European American", or "Not Latin American or European American" classifications. The Chi-Square test examining the differences between NoK identification, and FORDISC population affinity estimates was statistically significant at a 95% confidence interval with a *p*-value of 8.2264E-36 (Table 10). The majority of this variation was exhibited in the "Not Latin American (LA) or European American (EA)" classification as only one of the 35 individuals in the sample had a NoK identification as a population affinity other than Latin American or European American; however, FORDISC classified thirteen individuals of the sample as "Not Latin American or

European American" (Table 13). This indicates a large overclassification of individuals into the wrong population affinity group within this smaller sample. The single individual who had an NoK identification as a group other than LA or EA was a Native American male, which FORDISC classified as Latin American. Of the thirteen individuals not classified into LA or EA groups by FORDISC, the highest classification was African American (31%), followed by Native American (23%) and Guatemalan, (23%), then Japanese (15%), and Vietnamese (8%) (Table 12). There was an over classification of individuals as African American, and while 23% of the "Not Latin or European American" group were identified by FORDISC as Native American, the software failed to accurately classify the one Native American individual in the sample. The Chi-square analysis conducted on the sex values between NoK and FORDISC were not statistically significant at a *p*-value of 0.1761 (Table 14 and 15). Therefore, no further tests were conducted as this p-value indicates that there is no significant difference in how FORDISC classifies sex compared to the Next of Kin identification of sex. The Next of Kin sample was significantly smaller than the overall sample used for this study and therefore the statistics conducted on this sample may not be reflective of the true functionality of FORDISC in this context.

Table 13. Percentages and number of individuals (n) in each demographic group for NoK identification and FORDISC classifications.

Demographic group	**Next of Kin (NoK) identification**		**FORDISC classification**		**Chi-Square**	**Standardized residual**
	%	**n**	**%**	**n**		
Latin American	80%	28	29%	10	11.6	-3.4
European American	17%	6	34%	12	6	2.4

Not Latin American or European American	3%	1	37%	13	144	12
Total	100%	35	100%	35		
					Degrees of freedom	2
					p-value	8.2264E-36

Table 14. Percentages and number of individuals (n) in each sex for NoK identification and FORDISC estimates. Standardized residual values were not collected due to the *p*-value being greater than the 0.05 alpha value for significance.

Sex	**Next of Kin (NoK) classification**		**FORDISC classification**		**Chi-Square**
	Percentage	**Count**	**Percentage**	**Count**	
AMAB	49%	17	60%	21	0.9
AFAB	51%	18	40%	14	0.9
Total	100%	35	100%	35	
			Degrees of freedom		3
			p-value		0.17611983

Table 15. Breakdown of the number of individuals and corresponding percentages of the thirteen individuals classified by FORDISC as “Not Latin American or European American”.

FORDISC classification within Not LA or EA	**Number of individuals**	**Percentage**
African American	4	31%
Guatemalan	3	23%
Japanese	2	15%
Native American	3	23%
Vietnamese	1	8%
Total	13	100%

To ensure that all measurements that were collected were reliable, intraobserver error rates were collected using an Intraclass Correlation Coefficient (ICC) test. The results showed that the study investigator was heavily reliable in collecting most of the measurements with all the ICC single measure values being above 0.5, indicating a high level of reliability (Table 16). The one measurement that showed the least amount of intraobserver reliability was Orbital Breadth with an ICC value of 0.585 which is still within the range of moderate reliability. The overall high values for the ICC analyses suggest that the investigator was reliable in taking the measurements used for subsequent analyses.

Table 16. Intraclass correlation coefficient results for each measurement to test reliability of the investigator in taking measurements.

Measurement	**Single measures**
GOL	.997
XCB	.970
ZYB	.910
BBH	.992
BNL	.974
BPL	.986
MAB	.971
AUB	.905
NPH	.989
WFB	.922
FMT	.968
NLH	.946
NLB	.951
OBB	.585
OBH	.823
EKB	.650
DKB	.702
FRC	.771
PAC	.836
OCC	.666
FOL	.844
FOB	.813
MDH	.887
ID-GN	.972
MandBodHt	.959
go-go	.988
cdl-cdl	.868
MinRamBrdth	.978

Conclusion: Results

In assessing the frequency distributions, cross-tabulations, chi-square tests, and standardized residuals it is evident that the Asian American demographic group is consistently misclassified. The groups with the fewest accurate classifications, with corresponding FORDISC reference groups, were Native American AFABs, Japanese AMABs, Japanese AFABs, Vietnamese AMABs, and Chinese AMABs with zero matches. Asian American individuals were one of the least accurately classified groups, along with Native Americans and in both demographic groups, AFAB individuals were misclassified at a higher rate than AMABs. The Chi-Square results indicated that there was significant difference between the observed NMDID identifications and the estimated FORDISC classifications. The Chi-Square tests simply assess how FORDISC classified the same sample compared to the NMDID investigators or NMDID Next of Kin, it does not assess the accuracy of FORDISC estimates in comparison to the classifications of the NMDID. However, the Chi-Square test does indicate that most of the differences between the FORDISC classifications and NMDID identifications comes from Native American, and Latin American demographic groups. This is important because it suggests that FORDISC over or underclassifies into each of these groups and is not ideal for estimating a modern US population.

DISCUSSION

The goal of this study was to assess how accurate FORDISC 3.1 would be in classifying individuals of a modern US population. Of particular interest to the researcher, was the applicability of FORDISC in the estimation of Asian American individuals, due to the minimal representation of this group in the software's reference groups. It was hypothesized that Asian American individuals would be accurately classified fewer times than any other group, and Asian American AFABs would be even more likely to be misclassified than any other group, both of these claims were supported by the cross-tabulation analysis in which it was shown that Asian American and Native American individuals were classified less accurately than all other groups, and more specifically AFABs in both demographic groups were never accurately classified by the FORDISC software. There were clear differences in how FORDISC classified the sample compared to the expected classifications, suggesting that the software is not entirely accurate when estimating the population affinities of a sample of modern US individuals of unknown population affinities. The morphometric methods of analysis used in this study are valuable when assessing population affinity, however, proper reference groups are needed to develop accurate estimates.

Craniometric analysis has become a very useful tool for forensic anthropologists to understand the complex origins of various population affinity groups as they are ever changing, particularly in the United States (Hughes et al. 2016). The creators of FORDISC 3.1 note that it is a useful tool in assessing craniometric data of remains that are of a known population affinity; however, it is not often used on samples of known population (Ousley & Jantz 2005a). For remains of unknown origin, this software is often had difficulty in classifying individuals accurately due to the limited reference groups to which one can compare the unknown individual (Dudzik & Jantz 2016; Hughes et al. 2019; L'Abbe et al. 2013; Manthey et al. 2018). The results of the Chi-Square tests between FORDISC and the Investigator Reported classifications of both sex and population affinity are indicative of there being statistically significant differences between the classifications created by FORDISC and those of the NMDID investigator and NoK classifications. The majority of the differences seen between the observed FORDISC estimates and the expected NMDID Investigator classifications were caused by an overclassification of individuals by FORDISC into the Latin American AMAB group, and an under-classification of individuals into the Asian American and Native American AFAB groups. This suggests that FORDISC will classify individuals as Latin American AMABs much more often than any other group, which can cause complications in the context of forensic investigations as it can spark implicit biases in investigators or impact the direction of an investigation (Appel et al. 2023). This is of concern due to the usage of FORDISC by anthropologists in forensic contexts, within which most remains are of unidentified and of unknown origin.

There are thousands of unidentified remains discovered each year within the United States, and a software such as FORDISC would be helpful for anthropologists to create estimates for potential identification (Parsons 2021). However, if the software cannot be reliably applied to unidentified remains, it begs the question of it is ethical to use a software that is not representative of the possible groups that an individual within the United States can belong to. While the FORDISC help file (Ousley and Jantz 2005a) does state that the software should be used on known populations, it is a common practice for this software to be used by anthropologists in various subfields of biological anthropology to estimate the potential population affinity of unknown individuals (Hughes et al. 2018, 2021). The continued research on the accuracy of FORDISC in estimating population affinity of known and unknown remains assists in the overall development and evolution of the software so that it can best serve those who use it.

In assessing the cross-tabulation results, it was clear that certain demographic groups were misclassified at higher rates than others. African American AFABs and AMABs were the only demographic group with more correct classifications by FORDISC than incorrect classifications which could likely a result of the large reference groups available in the FORDISC software and its origins as being representative of African American populations in the United States (Ousley & Jantz 2005a). It is interesting to note that while European American AMABs and AFABs were the second demographic group with the highest correctly classified individuals, there were still more individuals incorrectly classified. This is interesting as it was expected that due to the large numbers of European American individuals in the FORDISC reference groups for both AMABs and AFABs this demographic group would likely have more accurately classified individuals than inaccurately classified individuals. These numbers indicating that African American and European American individuals are classified correctly more often than any other group align with recent research into the accuracy of population affinity estimation methods (Hughes et al. 2021). Other groups were not classified with such strong rates of accuracy, and these also followed expected patterns.

Asian American, Native American, and Latin American groups all misclassified at the highest rates out of all demographic groups. The misclassification of these groups follows the expected trends which have been seen in previous research (Dudzik & Jantz 2018; Hughes et al. 2021). All of these groups are groups are considered minorities and are frequently underrepresented in biological anthropology methodology (Scott et al. 2023). Throughout the development of anthropological approaches to population affinity estimation, these three population affinity groups (Asian, Latin, and Native American) have been lumped together as one ancestral group due to their connection as Asian or Asian-derived groups (Plemons & Hefner 2016). What is of note is the way in which most individuals that were classified as Latin American, more specifically, Guatemalan AMAB by FORDISC, were not identified as "Hispanic males" by the Investigator or NoK, but rather the majority of individuals classified to this group by FORDISC were Asian American. This could indicate high variability within the Guatemalan male reference group and low variability in the Asian reference groups in the FDB which is leading to conflation of other groups with Guatemalan males. Regardless of the cause, these groups are at risk of misclassification in forensic contexts.

The Forensic Data Bank consists of individuals intended to be representative of the United States population. Many individuals in the FDB were born after 1930 and within the United States; however, this does not apply the Native American or Asian populations (Ousley & Jantz 2005a). The Native American population consists of only 91 individuals who were born during the 19th century and primarily are members of Southwestern tribes; the Asian sample is comprised of primarily males with only 58 Asian females across the three reference groups, and all individuals are from outside of the United States. The low sample size of these populations and their lack of representation of modern individuals, likely is the cause of the under classification of Native American and Asian individuals in this study. Biological anthropology has long operated on the assumption that there is little heterogeneity within Native American groups, based on the inaccurate postulation that a common language is indicative of homogeneous origin (Ross 2002). However, studies on microevolution within the Americas have worked to emphasize the heterogeneity of these groups and the multi-wave migration that led to these differences; future work needs to be done to further emphasize the heterogeneity between the Native American populations and modern Asian populations (Powell & Neves 1999; Ross 2002). Many studies have shown the changes that can occur within population groups over as short of a span as 100 to 200 years and therefore the continued interpretation of Asian and Native American groups to be homogenous is incorrect (Jantz & Meadows Jantz 2000; Jellinghaus et al. 2018; Weisensee & Jantz 2011). When assessing population affinity and morphology of Native American populations, the discussion spans a much longer time frame. Estimating affinity with the assumption that there has been no morphological change to populations does not result in an accurate analysis of the population. This same approach is necessary when assessing the overclassification of individuals by FORDISC as Latin American. The “Latin American” classification encompasses a large group of ethnicities, races, and geographical regions, as does the classification of “Asian”. The United States census does not identify Hispanic, Latino, or Spanish as racial categories, but rather ethnic groups (Berg & Ta’ala **2014**). And like race, ethnic group classifications change and fluctuate based on culture, region, and time period (Parsons 2021). There is also the consideration of phenotypic appearance of the individual in comparison to their skeletal morphology. There was a clear overclassification of individuals into the Latin American reference group, a large majority of which were Native American or Asian American. There is a trend in population affinity research that shows that these three groups often are conflated which likely has to do with the continued misunderstanding of the variation between these three broader populations (Dudzik & Jantz 2018; Hughes et al. 2019) The overclassification of individuals as Latin American and under classifications of individuals in the sample as Asian American or Native American suggest that FORDISC has difficulty correctly assigning these individuals to their appropriate groups.

Within the NMDID sample analyzed, several were classified as Hispanic by NoK, but were classified as White by the investigator. This suggests that the social/peer-perceived classification that the individual identified as might not have always directly correlated with their phenotypic presentation that would classify them into an ethnic or social group. Forensic anthropologists are one of the many participants in the medicolegal death investigation system that perpetuate the errors in identification of ethnicity or race (Appel et al. 2023). It is also interesting to note that the two groups with the highest levels of misclassification have a history of admixture which were at one point one of the largest population groups in Mexico (Ross et al. 2014 p.155-164). Even into the 21st century, the group of individuals of mixed Spanish and Native American ancestry was the largest within Mexico, and the next largest being indigenous individuals (Ross et al. 2014 p.155-164). Mexico has a high level of admixture of Native American populations, while other Latin American population groups such as Caribbean groups have a much larger genetic overlap with African populations (Dudzik & Jantz 2016). Native American individuals within the NMDID were frequently misclassified as Latin American by FORDISC. The ethnic group of "Hispanic" often is a large group that encompasses admixture of various groups to varying degrees (Dudzik & Jantz 2016). Using this title is an oversimplification of the heterogeneous structure of the population. The misclassification of Latin American individuals or individuals of Hispanic ethnicity is becoming an increasing issue as the population of immigrants from Latin America is increasing annually (Hughes et al. 2019; Ross et al. 2014 p. 155-164). It becomes even more complicated when considering the ineffectiveness of current anthropological methods for estimating admixed population groups due to the consistent gene flow within and between groups (Dudzik & Jantz 2016). The complicated genetic history of indigenous populations in the Americas results in complexities in estimating population affinity.

APPLICABILITY OF FORDISC: MISCLASSIFICATIONS

The applicability of FORDISC in forensic contexts is of growing concern. The misclassification of remains disproportionately affects marginalized demographic groups. The overclassification of individuals as males is potentially a result of the lack of female reference groups, and a majority of this misclassification is likely the result of the overclassification of individuals as Guatemalan AMAB. The minimal female reference groups is of important consideration when discussing the accuracy rates of the Asian American AFAB sample as all but one of the Asian reference groups in FORDISC are only representative of males and are missing female counterparts. It is also interesting to note that Asian American AFABs and Asian American AMABs were the two most likely demographic groups to be classified as Guatemalan AMABs. The lack of AFAB Asian American reference groups for individuals in the sample to classify into resulted in an over saturation of the Latin American AMAB classification. Another group that presented in a similar pattern was Native American AFABs, in which none were estimated correctly by FORDISC. The FORDISC software only has 32 Native American AFABs in the Native American reference group, and as previously stated, the population is not representative of modern populations. The observed rates of misclassifications call into question the applicability of these estimates in forensic contexts, particularly in communities with large Native American, Asian American, or Latin American populations.

Native American, Asian American, and Latin American population affinity groups show substantial conflation. There are many reasons why this presents issues in law enforcement. There is already a serious mistrust of law enforcement by Native American communities due to the generations of neglect and abuse at the hands of the government and law enforcement (Redner-Vera & Galeste 2015). The misclassification of remains likely adds to the distrust that is present within these communities. Other difficulties present themselves with the high misclassification of Latin American populations, particularly due to the current large influx of migrating populations from Latin America into the United States (Monte & Shin 2022). It is important to have methodology that can accurately estimate the current demographic groups observed in the United States especially when it is concerning migrants crossing the border as identification of these individuals is difficult even without misclassification of their remains.

The misclassifications of Asian American, Native American, and Latin American individuals is a strong indicator of the need for more population specific methods in estimating population affinity and other facets of the biological profile (Ramsthaler et al. 2007). Most formulae and methods for estimation of the bio profile are applicable to European American, African American, and Native American individuals which disproportionately impacts the classifications and applicability of methods to other population affinity groups. The introduction of discriminant function analysis through FORDISC is a helpful step toward more accurate classification methods. However, as seen by the data, there is still work to be done on the accuracy of sex classifications using FORDISC, not necessarily due inaccurate calculations, but rather due to a lack of diversity within the reference groups to which the investigator is meant to compare their measurements.

CONCLUSIONS

The field of forensic anthropology has long dealt with the conceptual difficulties in estimating population affinity. As the field has developed, software such as FORDISC have been created to try and ease the burden upon anthropologists in their attempts to estimate population affinity (Ousley & Jantz 2005a). However, it is clear that the complexities of race, ethnicity, ancestry, and population affinity have created barriers, that even a computational software such as FORDISC cannot supersede (Hughes et al. 2018). This study has found that while FORDISC can be a beneficial software in estimating population affinity for certain demographic groups in the modern United States, it is not always the best tool for estimating sex or population affinity on these more modern groups. Forensic anthropology has struggled to differentiate between race, ethnicity, ancestry, and population affinity (DiGangi & Bethard 2021; Flouri et al. 2022; Ross & Williams 2021; Tallman et al. 2021). In conducting this study, larger classification groups were created so that statistical analyses could be conducted. When creating these larger groups, it was important to not fall into the heavily relied on tripartite system. However, this became a complicated task when assessing demographic groups such as Asian American, or Latin American due to the geographic span of these population affinity groups.

The current methodologies in forensic anthropology used to estimate the facets of the biological profile usually require population specific equations to be used. If an individual is estimated to be of one population affinity, but truly belongs to another, this can impact the subsequent population specific methods used to estimate sex, age, or stature (Bright et al. 2018; Muñoz et al. 2018). The misclassification of a set of remains also presents other issues, particularly for investigators. Providing law enforcement with the incorrect population affinity estimate can result in the investigation excluding certain demographics of individuals that should be included and severely hindering the rest of identification (Appel et al. 2023). The results of this study suggest that this may occur more frequently for Latin American and Native American individuals. These groups are not interchangeable in language, culture, or region and therefore the conflation of Asian or Native American individuals as Latin American puts them at a disadvantage of identification post-mortem. In many cases populations that represent minority communities in the United States are disproportionately affected by the lack of diversity in population specific methodology for the biological profile (Tallman et al. 2020). The marginalization of these communities even in death is something that the field of anthropology is actively trying to combat through research into more inclusive methods and language of population affinity estimation (Appel et al. 2023; Atkinson & Tallman 2019; Bartelink et al. 2020; DiGangi & Bethard 2021; Hughes et al. 2018; Hughes et al. 2021).

Limitations

This study would benefit from an increase in sample size particularly of Asian American individuals. To more accurately assess the proportion of individuals estimated correctly using the FORDISC software, it would be best who have a large sample of solely Asian American individuals with equal numbers of AFABs and AMABs to assess the potential impact of sexual dimorphism on the classification trends using FORDISC. A comparison group that is repeatedly shown to classify well using FORDISC, such as an African American sample, would be helpful to provide a comparison of how drastic of a difference there is between the classification accuracies. To mitigate the potential for non-American born individuals, an even more modern sample of Asian American individuals, with proof of birth in the United States would also greatly benefit a future study as it would be a comprehensive look at modern Asian American individuals. Larger samples are much more easily accessible in digital databases, especially for modern populations and therefore the continued use of CT images for this study would be appropriate.

To address the need for a truly modern population to obtain a sample from, it would be helpful to use remains from an evolving database of CT scans within the United States that continues to develop to represent the demographics of the current US population. With the innovations to technology, it is expected that more reliable radiographic reports and CT imaging will become available. The utilization of this updated technology would improve this study as there were certain craniometric measurements that were unobtainable due to the quality of the scans. If this study were to be repeated, doing so on a physical skeletal collection would likely ensure fuller collection of measurements for each individual, however it likely would reduce the variability of the population from which the researcher can draw their sample, as well as limit the modernity of the population. While an ideal sample would have an increased number of Asian American individuals, it is important to note that many skeletal collections open for researchers are not accurately representative of the modern United States population and this would need to be considered.

Final Thoughts

This study was conducted in an effort to shed light on the potential disparities and inequalities that are present in some of the software and methodologies used by forensic anthropologists in the United States. The FORDISC 3.1 software is a useful tool for forensic anthropologists, but its usefulness is challenged by its ability to maintain modern populations in its database. The implementation of a more universal uploading system for forensic casework, or the included data from various geographic regions would improve the functionality of the software and its applicability to numerous populations. Further studies into population affinity and its practical usage in forensic contexts are necessary as the United States faces a consistently fluctuating population and an even faster changing social climate. The estimation of population affinity is complicated in its history and should be practiced with caution as to not allow for biases. To best serve the community anthropologists need to better understand human variation, particularly relating to the manifestation of variation in cranial features. A more holistic approach to understanding variation within and among groups is a beneficial way to move forward to ensure the protection of marginalized communities and the retention of accuracy and reliability in anthropological methods.

APPENDIX

Table A.1. Age, sex, and population affinity for each individual in the final sample of 148 individuals.

Scan number	NMDID PA	NMDID sex	NoK PA	NoK sex	FORDISC PA	FORDISC sex	Age
100122	W	M	H	M	LA	M	25
100144	NA	M	N/A	N/A	NA	F	47
104670	AA	F	N/N	N/A	AA	F	30
103609	NA	F	N/A	N/A	W	M	68
101607	LA	F	H	F	LA	F	39
100487	OA	F	N/A	N/A	W	F	28
103894	NA	F	N/A	N/A	W	F	36
102114	J	F	N/A	N/A	LA	F	102
101839	AA	F	N/A	N/A	AA	F	88
100625	W	M	H	M	NA	M	29
100585	W	M	N/A	N/A	GT	M	49
101933	W	M	N/A	N/A	W	M	46
108025	W	M	N/A	N/A	W	M	46
193855	AA	F	N/A	N/A	AA	F	49
197385	AA	F	N/A	N/A	AA	F	36
199960	NA	M	N/A	N/A	AA	M	38
199818	LA	F	H	F	AA	F	42
198394	AA	M	N/A	N/A	GT	M	57
198324	W	F	N/A	N/A	W	F	64
196394	AA	F	N/A	N/A	W	F	50
195754	W	F	W	F	NA	M	50
195094	AA	M	N/A	N/A	W	M	56
194935	AA	M	N/A	N/A	AA	M	53
100308	LA	F	H	F	J	F	60
103122	NA	F	N/A	N/A	W	F	49
108962	OA	M	N/A	N/A	W	M	21
109336	NA	M	N/A	N/A	W	M	39
109799	LA	M	H	M	W	M	39
110687	NA	M	N/A	N/A	LA	M	28
109005	LA	F	H	F	W	F	42
108966	AA	F	N/A	N/A	LA	F	28
186316	LA	F	H	F	W	F	45
186377	W	F	W	F	W	F	53
187788	AA	M	N/A	N/A	NA	M	20
188138	AI	M	N/A	N/A	GT	M	58

187544	NA	F	N/A	N/A	V	M	56
190979	W	F	N/A	N/A	AA	M	19
189158	AA	M	N/A	N/A	AA	F	59
186416	W	F	N/A	N/A	W	M	41
109139	NA	M	N/A	N/A	AA	M	23
110033	CH	M	N/A	N/A	V	M	21
113436	W	M	W	M	J	F	53
185685	NA	M	N/A	N/A	CH	M	29
187419	LA	F	H	F	AA	F	77
114991	AI	F	N/A	N/A	LA	F	25
115283	LA	M	N/A	N/A	GT	M	20
115869	NA	M	NA	M	LA	M	59
115724	V	M	N/A	N/A	GT	M	55
117773	NA	F	N/A	N/A	LA	M	31
115850	AA	M	N/A	N/A	LA	M	64
194639	AA	F	N/A	N/A	NA	F	48
115690	W	F	N/A	N/A	AA	F	38
118131	AA	F	N/A	N/A	AA	F	61
118582	V	M	N/A	N/A	GT	M	25
119886	NA	F	N/A	N/A	W	F	93
120087	W	M	W	M	W	M	64
121970	W	M	H	M	V	M	51
122000	AA	M	N/A	N/A	AA	M	39
122739	AA	F	N/A	N/A	AA	M	33
123347	NA	F	N/A	N/A	GT	M	30
126901	W	M	N/A	N/A	J	M	48
126929	NA	F	N/A	N/A	LA	M	18
127201	AA	M	N/A	N/A	AA	M	56
113895	OA	F	N/A	N/A	GT	M	23
120981	W	F	N/A	N/A	J	M	76
127928	NA	F	N/A	N/A	LA	M	34
127989	W	F	H	F	AA	F	25
128264	NA	F	N/A	N/A	GT	M	53
128391	AA	M	N/A	N/A	AA	M	38
128425	LA	F	H	F	LA	F	43
128608	AA	M	N/A	N/A	AA	M	44
129420	NA	M	N/A	N/A	LA	M	35
129648	OA	F	N/A	N/A	W	F	26
130189	W	M	N/A	N/A	W	M	39
130290	LA	M	H	M	NA	M	25
131133	W	M	H	M	LA	M	30
131537	OA	M	N/A	N/A	W	M	67

131714	AA	F	N/A	N/A	AA	M	39
131893	W	F	N/A	N/A	AA	M	37
131933	AA	F	N/A	N/A	AA	F	64
134487	W	F	H	F	GT	M	49
134518	NA	F	N/A	N/A	V	M	43
134686	LA	F	N/A	N/A	AA	F	48
134954	V	M	N/A	N/A	J	M	78
134998	K	M	N/A	N/A	NA	M	70
135190	NA	M	N/A	N/A	GT	M	77
135233	AI	M	N/A	N/A	NA	M	66
135543	AI	M	N/A	N/A	AA	M	25
135673	NA	M	N/A	N/A	NA	M	25
136718	AA	F	N/A	N/A	AA	F	27
137356	OA	F	N/A	N/A	GT	M	63
138121	NA	M	N/A	N/A	GT	M	36
139022	LA	M	N/A	N/A	W	F	22
139296	LA	M	N/A	N/A	NA	M	60
139325	LA	M	H	M	LA	M	27
139871	LA	F	H	F	GT	M	49
139954	W	M	N/A	N/A	J	M	22
140836	OA	M	N/A	N/A	W	M	29
140892	AA	M	N/A	N/A	AA	M	55
141103	NA	M	N/A	N/A	NA	M	62
141232	OA	F	N/A	N/A	AA	M	19
142882	LA	M	H	M	LA	M	30
142925	W	M	N/A	N/A	NA	M	47
144387	AA	F	N/A	N/A	GT	M	51
144783	W	F	N/A	N/A	W	F	53
144834	AA	F	N/A	N/A	SAA	F	31
144911	AA	M	N/A	N/A	AA	M	61
147301	W	F	N/A	N/A	LA	M	55
147386	AI	M	N/A	N/A	AA	F	46
148298	LA	F	H	F	W	F	29
149162	AA	M	N/A	N/A	AA	M	52
151758	NA	F	N/A	N/A	LA	F	53
152085	W	F	H	F	LA	F	86
153566	NA	F	N/A	N/A	V	M	21
153614	W	F	H	F	AA	F	62
153969	NA	M	N/A	N/A	LA	M	60
154683	NA	F	N/A	N/A	LA	F	51
154730	W	M	H	M	LA	M	21
158033	AA	F	N/A	N/A	AA	F	33

158267	J	F	N/A	N/A	GT	M	64
158596	AA	F	N/A	N/A	LA	F	34
158699	LA	M	H	M	LA	M	60
159787	LA	M	H	M	W	M	70
159879	AA	F	N/A	N/A	GT	M	64
159977	AA	M	N/A	N/A	LA	M	25
160978	W	M	W	M	W	M	53
161393	LA	F	H	F	GT	M	53
162676	W	M	H	M	W	M	64
162830	NA	F	N/A	N/A	V	M	41
163166	W	M	N/A	N/A	GT	M	53
163857	AA	M	N/A	N/A	AA	M	41
164429	AA	F	N/A	N/A	AA	F	53
165304	AA	M	N/A	N/A	AA	M	61
166504	AA	F	N/A	N/A	AA	F	38
166727	NA	M	N/A	N/A	LA	M	34
169100	W	F	N/A	N/A	W	F	67
169152	J	M	N/A	N/A	W	M	55
169608	OA	M	N/A	N/A	J	M	21
170105	W	F	N/A	N/A	J	F	75
170356	W	F	W	F	W	M	86
170627	AI	F	N/A	N/A	W	F	51
171208	W	F	H	F	W	F	36
172750	NA	F	N/A	N/A	V	M	37
173775	LA	M	H	M	W	M	49
176577	W	F	N/A	N/A	W	F	67
176735	LA	F	N/A	N/A	LA	F	74
176827	W	M	N/A	N/A	W	M	55
177151	NA	M	N/A	N/A	W	M	44

REFERENCES

Abegg, C., Balbo, I., Dominguez, A., Grabherr, S., Campana, L., & Moghaddam, N. (2021). Virtual anthropology: A preliminary test of macroscopic observation versus 3D surface scans and computed tomography (CT) scans. *Forensic Sciences Research*, *6*(1), 34–41. https://doi.org/10.1080/20961790.2020.1817270

Adams, D. M., & Pilloud, M. A. (2021). Perceptions of Race and Ancestry in Teaching, Research, and Public Engagement in Biological Anthropology. *Human Biology*, *93*(1), 1–25.

Appel, N. S., Edgar, H. J. H., Berry, S. D., & Hunley, K. (2023). Error and bias in race and ethnicity descriptions in medical examiner records in New Mexico: Consequences for understanding mortality among Hispanic/Latinos. *Forensic Science International: Synergy*, *7*, 100338. https://doi.org/10.1016/j.fsisyn.2023.100338

Arnaiz-Villena, A., Parga-Lozano, C., Moreno, E., Areces, C., Rey, D., & Gomez-Prieto, P. (2010). The Origin of Amerindians and the Peopling of the Americas According to HLA Genes: Admixture with Asian and Pacific People. *Current Genomics*, *11*(2), 103–114. https://doi.org/10.2174/138920210790886862

Atkinson ML, Tallman SD. (2019). Nonmetric cranial trait variation and ancestry estimation in Asian and Asian-derived groups. Journal of Forensic Sciences 65:692–706.

Baglivo, M., Winklhofer, S., Hatch, G. M., Ampanozi, G., Thali, M.J., Ruder, T. D. (2013) The rise of forensic and post-mortem radiology—Analysis of the literature between the year 2000 and 2011. *Journal of Forensic Radiology and Imaging.* 1, 3-9. https://doi.org/10.1016/j.jofri.2012.10.003

Bartelink, E. J., Boyd, D. C., France, D. L., Pokines, J. T., & Prince-Zinni, D. (2020). The American Board of Forensic Anthropology Turns 40 Historical Perspectives and Current Trends in Certification for Forensic Anthropology. *Forensic Anthropology*, *3*(2), 109–118. https://doi.org/10.5744/fa.2020.1017

Berg, G.E., Ta'ala, S.C. (2014). 'Biological Affinity in Medicoloegal, Public, and Anthropological Contexts', in *Biological Affinity in Forensic Identification of Human Skeletal Remains Beyond Black and White*. 1st edn. Boca Raton: CRC Press, pp. 17-26.

Birkby, W. H. (1966). An evaluation of race and sex identification from cranial measurements. *American Journal of Physical Anthropology*, *24*(1), 21–27. https://doi.org/10.1002/ajpa.1330240103

Carew, R. M., & Errickson, D. (2020). An Overview of 3D Printing in Forensic Science: The Tangible Third-Dimension. *Journal of Forensic Sciences*, *65*(5), 1752–1760. https://doi.org/10.1111/1556-4029.14442

Caspari, R. (2003). From types to populations: A century of race, physical anthropology, and the American Anthropological Association. *American Anthropologist*, *105*(1), 65–76. https://doi.org/10.1525/aa.2003.105.1.65

Caspari, R. (2009). 1918: Three perspectives on race and human variation. *American Journal of Physical Anthropology*, *139*(1), 5–15. https://doi.org/10.1002/ajpa.20975

Christensen, A., Smith, M., Gleiber, D., Cunningham, D., & Wescott, D. (2018). The Use of X-ray Computed Tomography Technologies in Forensic Anthropology. *Forensic Anthropology*, *1*(2), 124–140. https://doi.org/10.5744/fa.2018.0013

Dechow, P. C., Wang, Q., & Peterson, J. (2010). Edentulation Alters Material Properties of Cortical Bone in the Human Craniofacial Skeleton: Functional Implications for Craniofacial Structure in Primate Evolution. *The Anatomical Record*, *293*(4), 618–629. https://doi.org/10.1002/ar.21124

DiGangi, E. A., & Bethard, J. D. (2021). Uncloaking a Lost Cause: Decolonizing ancestry estimation in the United States. *American Journal of Physical Anthropology*, *175*(2), 422–436. https://doi.org/10.1002/ajpa.24212

Dudzik B., Jantz RL. (2016). Misclassifications of Hispanics using Fordisc 3.1: Comparing cranial morphology in Asian and Hispanic populations. *Journal of Forensic Sciences* 61:1311–1318.

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), e1369. https://doi.org/10.1002/wfs2.1369

Duray, S. M., H. B. Morter, and F. J. Smith (1999) Morphological variation in cervical spinous processes: Potential applications in the forensic identification of race from the skeleton. *Journal of Forensic Sciences* 44:937-944.

Edgar, HJH; Daneshvari Berry, S; Moes, E; Adolphi, NL; Bridges, P; Nolte, KB (2020). New Mexico Decedent Image Database. Office of the Medical Investigator, University of New Mexico. doi.org/10.25827/5s8c-n515.

Edgar, H., Pilloud, M., (2021). A Reassessment of Assessing Race: "Ancestry" Estimation and Its Implications for Forensic Anthropology and Beyond. *Forensic Anthropology 4*(4), 67-72. http://doi.org/10.5744/fa.2021.0026

Flouri, D.E., Alifragki, A., Garcfa-Donas, J.G., Kranioti, E.F. (2022). Ancestry Estimatio: Advances and Limitations in Forensic Applications. *Research and Reports in Forensic Medical Science* 12:12-24.

Fuentes A., Ackermann RR., Athreya S., Bolnick D., Lasisi T., Lee S-H., McLean S-A., Nelson R. (2019) AAPA statement on race and racism. *American Journal of Physical Anthropology*, 169(3):400–402.

Go, M. C., Jones, A. R., Algee-Hewitt, B. F. B., Dudzik, B., & Hughes, C. E. (2019). Classification trends among contemporary Filipino crania using Fordisc 3.1. *Forensic Anthropology*, *2*(4), 293–303. https://doi.org/10.5744/fa.2019.1005

González-José, R., Bortolini, M. C., Santos, F. R., & Bonatto, S. L. (2008). The peopling of America: Craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *American Journal of Physical Anthropology*, *137*(2), 175–187. https://doi.org/10.1002/ajpa.20854

Gould, S.J. (1981) *The Mismeasure of Man*. New York: W.W. Norton.

Hasan, H., Khursheed Alam, M., Yusof, A., Matsuda, S., Shoumura, M., & Osuga, N. (2016). Accuracy of Three Dimensional CT Craniofacial Measurements Using Mimics and InVesalius Software Programs. *Journal of Hard Tissue Biology*, *25*(2), 219–224. https://doi.org/10.2485/jhtb.25.219

Harvati, K., & Weaver, T. D. (2006). Human cranial anatomy and the differential preservation of population history and climate signatures. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology*, *288A*(12), 1225–1233. https://doi.org/10.1002/ar.a.20395

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

Herrera, M. D., & Tallman, S. D. (2019). Craniometric variation and ancestry estimation in two contemporary Caribbean populations. *Forensic Science International*, *305*, 110013. https://doi.org/10.1016/j.forsciint.2019.110013

Herrera, B., Peart, D., Hernandez, N., Spradley, K., & Hubbe, M. (2017). Morphological variation among late holocene Mexicans: Implications for discussions about the human occupation of the Americas. *American Journal of Physical Anthropology*, *163*(1), 75–84. https://doi.org/10.1002/ajpa.23186

Hughes CE, Dudzik B, Algee-Hewitt BFB, Jones A, Anderson BE. (2018). Understanding (mis)classification trends of Latin Americans in Fordisc 3.1: Incorporating cranial morphology, microgeographic origin, and admixture proportions for interpretation. *Journal of Forensic Sciences* 64:353–366.

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

Howells, W. W. (1996). Howells' craniometric data on the internet. *American Journal of Physical Anthropology*, *101*(3), 441–442. https://doi.org/10.1002/ajpa.1331010302

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

Jellinghaus, K., Hoeland, K., Hachmann, C., Prescher, A., Bohnert, M., & Jantz, R. (2018). Cranial secular change from the nineteenth to the twentieth century in modern German individuals compared to modern Euro-American individuals. *International Journal of Legal Medicine*, *132*(5), 1477–1484. https://doi.org/10.1007/s00414-018-1809-5

Kelley, S. R., & Tallman, S. D. (2022). Population-Inclusive Assigned-Sex-at-Birth Estimation from Skull Computed Tomography Scans. *Forensic Science*, *2*, 321–348. https://doi.org/10.3390/ forensicsci2020024

Krishan, K., Chatterjee, P. M., Kanchan, T., Kaur, S., Baryah, N., & Singh, R. K. (2016). A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. *Forensic Science International*, *261*, 165.e1-165.e8. https://doi.org/10.1016/j.forsciint.2016.02.007

Kilroy GS, Tallman SD, DiGangi EA. (2020). Secular change in morphological cranial and mandibular trait frequencies in European Americans born 1824–1987. *American Journal of Physical Anthropology* 173:589–605.

Koehler, J. J. (2011). If the Shoe Fits They Might Acquit: The Value of Forensic Science Testimony. *Journal of Empirical Legal Studies*, *8*(s1), 21–48. https://doi.org/10.1111/j.1740-1461.2011.01225.x

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

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research | *Journal of Chiropractic Medicine (15)* 155-163 https://doi.org/10.1016/j.jcm.2016.02.012

Langley, N., Jantz, L., Jantz, R., Ousley, S., Maijanen, H., & McNulty, S. (2016). *DCP 2.0 Osteometric Data*. 1-116 https://doi.org/10.17632/6xwhzs2w38.1

Langley, N. R., Meadows Jantz, L., McNulty, S., Maijanen, H., Ousley, S. D., & Jantz, R. L. (2018). Error quantification of osteometric data in forensic anthropology. *Forensic Science International*, *287*, 183–189. https://doi.org/10.1016/j.forsciint.2018.04.004

Lee, M., & Gerdau-Radonic, K. (2020). Variation within physical and digital craniometrics. *Forensic Science International*, *306*, 110092. https://doi.org/10.1016/j.forsciint.2019.110092

Lieberman, L., & Reynolds, L. T. (1978). The Debate over Race Revisited: An Empirical Investigation. *Phylon (1960-)*, *39*(4), 333–343. https://doi.org/10.2307/274899

Lo, M., Mariconti, E., Nakhaeizadeh, S., Morgan, R.M. (2023). Preparing computed tomography images for machine learning in forensic and virtual anthropology. *Forensic Science International: Synergy*, *6,* 100319. https://doi.org/10.1016/j.fsisyn.2023.100319

Manthey, L., Jantz, R. L., Vitale, A., & Cattaneo, C. (2018). *Population specific data improves Fordisc®'s performance in Italians | Elsevier Enhanced Reader*. 263.e1-263.e7 https://doi.org/10.1016/j.forsciint.2018.09.023

Marten, M. G., Winburn, A. P., Burgen, B. R., Seymour, S. K., & Walkup, T. (2023). What makes a "good" forensic anthropologist? *American Anthropologist*, *125*(3), 582–596. https://doi.org/10.1111/aman.13874

Monte, L. M., & Shin, H. B. (2022). *20.6 Million People in the U.S. Identify as Asian, Native Hawaiian or Pacific Islander*. Census.Gov. Retrieved October 21, 2022, from https://www.census.gov/library/stories/2022/05/aanhpi-population-diverse-geographically-dispersed.html

Ousley S, Jantz R. (2005a). FORDISC Help File: Version 1.35. 1-221. Available from: http://math.mercyhurst.edu/~sousley/Fordisc/

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* 139:68–76.

Pang, J., Dong, Y., Turner, C., Li, C., & Liu, X. (2022). Analysis of data consistency identifies measurement abnormality in Howells' craniometric test data set. *American Journal of Biological Anthropology*, *179*(4), 687–692. https://doi.org/10.1002/ajpa.24631

Parsons, H.R., (2021). Ancestry Estimation in Practice: An Evaluation of Forensic Anthropology Reports in the United States. *Forensic Anthropology, 4*(4), 192+. http://dx.doi.org.proxy.library.upenn.edu/10.5744/fa.2020.0047

Pilloud, M., Skipper, C., Horsley, S., Craig, A., Latham, K., Clemmons, C., Zejdlik, K., Boehm, D., & Philbin, C. (2021). Terminology Used to Describe Human Variation in Forensic Anthropology. *Forensic Anthropology*. https://doi.org/10.5744/fa.2021.0001

Plemons, A., Hefner, J., (2016) Ancestry Estimation Using Macromorphoscopic Traits. *Academic Forensic Pathology 6*(3). 400-412. https://doi.org/10.23907/afpj.2016.041

Plens, C. R., De Souza, C. D., Albanese, J., Capp, T. T. L., & Saavedra de Paiva, L. A. (2021). Reflections on methods to estimate race and ancestry on reference osteological samples in the Brazilian context. *Ethics, Medicine and Public Health*, *18*, 100680. https://doi.org/10.1016/j.jemep.2021.100680

Powell, J. F., & Neves, W. A. (1999). Craniofacial morphology of the first Americans: Pattern and process in the peopling of the New World. *American Journal of Physical Anthropology*, *110*(S29), 153–188. https://doi.org/10.1002/(SICI)1096-8644(1999)110:29+<153::AID-AJPA6>3.0.CO;2-L

Ramsthaler, F., Kreutz, K., & Verhoff, M. A. (2007). Accuracy of metric sex analysis of skeletal remains using Fordisc® based on a recent skull collection. *International Journal of Legal Medicine*, *121*(6), 477–482. https://doi.org/10.1007/s00414-007-0199-x

Redner-Vera, E., & Galeste, M.-A. (2015). Attitudes and Marginalization: Examining American Indian Perceptions of Law Enforcement Among Adolescents. *Journal of Ethnicity in Criminal Justice*, *13*(4), 283. https://doi.org/10.1080/15377938.2014.984043

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., . . . Ruiz-Linares, A. (2012). Reconstructing native american population history. *Nature, 488*(7411), 370-4. doi:https://doi.org/10.1038/nature11258

Ross, A.H., Slice, D.E. and Ubelaker, D.H. (2014). 'Population Affinities of Hispanic Crania Implications for Forensic Identification', in *Biological Affinity in Forensic Identification of Human Skeletal Remains Beyond Black and White*. 1st edn. Boca Raton: CRC Press, pp. 155–164.

Ross AH, Williams SE. (2021). Ancestry studies in forensic anthropology: Back on the frontier of racism. Biology 10:602.

Richard AH, Parks CL, Monson KL. (2014). Accuracy of standard craniometric measurements using multiple data formats. *Forensic Science International, 242*, 177–185.

Rutherford, A. (2021). Race, eugenics, and the canceling of great scientists. *American Journal of Physical Anthropology*, *175*(2), 448–452. https://doi.org/10.1002/ajpa.24192

Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science*, *309*(5736), 892–896.

Sardi, M. L., Rozzi, F. R., González-José, R., & Pucciarelli, H. M. (2005). South Amerindian craniofacial morphology: Diversity and implications for Amerindian evolution. *American Journal of Physical Anthropology*, *128*(4), 747–756. https://doi.org/10.1002/ajpa.20235

Scott, G. R., Navega, D., Vlemincq-Mendieta, T., Dern, L. L., O'Rourke, D. H., Hlusko, L. J., & Hoffecker, J. F. (n.d.). Peopling of the Americas: A new approach to assessing dental morphological variation in Asian and Native American populations. *American Journal of Biological Anthropology*, *n/a*(n/a). https://doi.org/10.1002/ajpa.24878

Shamlou, A. A., & Tallman, S. D. (2022). Frontal Sinus Morphological and Dimensional Variation as Seen on Computed Tomography Scans. *Biology*, *11*(8), Article 8, 1-23. https://doi.org/10.3390/biology11081145

Simmons-Ehrhardt, T. (2021). Open osteology: Medical imaging databases as skeletal collections. *Forensic Imaging, 26,* 200462. https://doi.org/10.1016/j.fri.2021.200462

Spradley, K. M., & Jantz, R. L. (2016). Ancestry Estimation in Forensic Anthropology: Geometric Morphometric versus Standard and Nonstandard Interlandmark Distances. *Journal of Forensic Sciences*, *61*(4), 892–897. https://doi.org/10.1111/1556-4029.13081

Stull, K. E., Tise, M. L., Ali, Z., & Fowler, D. R. (2014). Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. *Forensic Science International. 238*, 133-140. https://doi.org/10.1016/j.forsciint.2014.03.005

Tallman, S. D., & Winburn, A. P. (2015). Forensic Applicability of Femur Subtrochanteric Shape to Ancestry Assessment in Thai and White American Males. *Journal of Forensic Sciences*, *60*(5), 1283–1289. https://doi.org/10.1111/1556-4029.12775

Tallman, S. D., Parr, N. M., & Winburn, A. P. (2020). Assumed Differences; Unquestioned Typologies: The Oversimplification of Race and Ancestry in Forensic Anthropology. *Forensic Anthropology*, 1-24. https://doi.org/10.5744/fa.2020.0046

Uldin, T. (2017). Virtual Anthropology—A Brief Review of the Literature and History of Computed Tomography Review. *Forensic Sciences Research (FSR)*, *2*(4), 165–173. https://doi.org/10.1080/20961790.2017.1369621

Vidoli, G. M. (2022). The New Mexico Decedent Image Database: Demographics, life, and the body (https://nmdid.unm.edu/). *American Journal of Biological Anthropology*, *179*(2), 331–332. https://doi.org/10.1002/ajpa.24584

Weisensee, K. E., & Jantz, R. L. (2011). Secular changes in craniofacial morphology of the portuguese using geometric morphometrics. *American Journal of Physical Anthropology*, *145*(4), 548–559. https://doi.org/10.1002/ajpa.21531

CURRICULUM VITAE

[redacted]

[redacted]