Assessing the Precision of Cranial and Mandibular Morphoscopic Traits from 3D Surface Scans

A Thesis Presented for the

Master of Arts

Degree

The University of Tennessee, Knoxville

Sarah Thomas Schwing

May 2021

To my best friend and to my biggest cheerleader,
my mom and dad.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Lee Meadows Jantz. This project has not been without its hiccups and I sincerely appreciate your guidance, patience, support, and encouragement throughout this process. I am so grateful to have had the opportunity to work with you and to have been one of your students. Thank you so much.

I would also like to thank my committee members, Dr. Dawnie Wolfe Steadman and Dr. Ben Auerbach, for their guidance, suggestions, and feedback throughout this project. I truly appreciate your availability and flexibility throughout this process. Thank you, also, to the Forensic Anthropology Center and all senior staff for the opportunity to work with the William M. Bass Donated Skeletal Collection and for providing access to the necessary technology to make this project a reality.

Thank you, Sun for all of your help with data analyses, and for all of your time and patience over these past two semesters.

Many, many thanks to my friends and colleagues Hayden Mckee-Zech and Sara Fatula for your constructive feedback, problem-solving assistance, hand-holding, and immense kindness during this past year especially, but over the past three years, as well. Thank you for continuing to answer my texts and phone calls, you guys are shining stars!

Finally, sincere thanks to my parents and all of my family, but especially to my mom. You are my role model and my very best friend, and I could not have done this without you.

Abstract

Virtual anthropological (VA) methods have been successfully used to capture metric data in the form of standard caliper measurements as well as volumetric data from various human skeletal elements. Virtual anthropological investigations of morphoscopic traits have increased over the past two decades, however, greater attention has been paid to investigations of metric data and to the use of CT technologies. Few studies have focused on morphoscopic data and fewer have employed 3D surface scans in data collection. Morphoscopic VA studies largely pertain to age estimation using the os coxa; fewer pertain to sex estimation and, to the author's knowledge, no VA study has investigated ancestry traits to date. As little research exists on the precision of morphoscopic data obtained from 3D surface scans, or virtual cranial skeletal elements in general, error rates for these data are not known, making these data, and studies using them, inadmissible in a court of law.

In an effort to address these voids, this study was devoted to understanding the utility of 3D surface scans for morphoscopic data collection, specifically sex and ancestry related morphoscopic data. This study found slightly higher agreement amongst Walker (2008) cranial traits for sex estimation, versus OSSA (Hefner and Ousley 2014) traits for ancestry estimation, and a minimal impact of experience level on score agreement. No statistically significant differences in Weighted Kappa values were found for cranial or mandibular traits during inter-format comparisons between graduate and professional participants. Correct classification accuracy rates were also ranged from 76-88% using trait evaluations from 3D surface scans.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

Forensic anthropologists seek a robust understanding of the human skeletal system, one that includes the dimensions of human skeletal growth and development, human skeletal variation, trauma, pathology, and taphonomy. Forensic anthropologists use this knowledge within medicolegal and human disaster contexts to aid law enforcement, medical examiners, and survivors of decedents in the location, recovery, identification, and repatriation of deceased individuals and by providing insight into the circumstances surrounding a decedent's death.

Forensic anthropological analyses typically focus on the construction of a biological sketch - termed a biological profile - of the deceased, which includes, at minimum, estimations of sex, age, ancestry, and stature. The biological profile is informed by the anthropologist's holistic understanding of human osteology, for example their ability to visually distinguish developmental processes from degenerative processes in the skeleton, and to perform similar morphoscopic evaluations on skeletal remains. Parameters of the biological profile, particularly those of sex and ancestry, can be informed by both metric and morphoscopic (nonmetric) methods, and are typically constructed using physical skeletal remains. Although specific methodology is left to the discretion of the anthropologist (Hefner and Ousley 2014), both quantitative and qualitative assessments are frequently performed to supplement one another and to provide a more robust analysis (Klales 2013; Lewis and Garvin 2016).

Over the course of the past century, numerous metric and morphoscopic methods have been developed for use in the construction of the biological profile. These methods make use of different traits, software programs, and foundational texts and are typically used in concert, as their foci on size and shape variation are complimentary in anthropological analyses (Lewis and Garvin, 2016). Most metric and morphoscopic methods, however, were developed for use on physical bone and were not intended for use on virtual skeletal elements. However, in recent years, increasing numbers of researchers have turned to virtual platforms to evaluate the precision of widely used metric and morphoscopic methods on three-dimensional (3D) virtual models and their utility as alternatives to dry bone (Bertoglio et al. 2020). This growing body of research is known as virtual anthropology (VA) and encompasses multiple methods of assessing human skeletal remains virtually or digitally. Forensic anthropologists have become increasingly interested in VA techniques for the unique insights that they afford which dry bone often cannot.

From radiographs to photogrammetry, CT scans, and 3D surface scans, technology is playing an ever-increasing role in forensic case documentation and analyses of the biological profile (Bertoglio et al. 2020; Decker et al. 2011; Deduoit et al. 2007; Garvin and Stock 2016; Villa et al., 2013). Digitally formatted human skeletal elements have been regarded with particular interest within forensic anthropology over the past two decades (Cooke and Terhune 2015; Freiss 2012; Petaros et al. 2017; Weber 2013) and are being increasingly endorsed as a non-damaging and permanent means of cataloging skeletal collections for the purposes of preserving fragile remains, data conservation, and ease of data sharing (Algee-Hewitt and Wheat 2016; Decker et al. 2011; Freiss 2012; Grabherr et al. 2008; Rebedizo 2016; Weber 2013). Such virtual methods have even been used to conduct virtual autopsies of human remains, augmenting physical autopsies and enabling forensic anthropologists, pathologists, and radiologists to document and preserve the initial condition of remains, to assess age and sex, and to discern any individuating characteristics that may lead to a positive identification (Decker et al. 2011: p.1112; Deduoit et al. 2007, 2014).

Given VA's origins in paleoanthropology, virtual anthropological research is used more extensively for metric methods than for morphoscopic methods (Bertoglio et al. 2020). Metric investigations have featured both cranial (Abdel Fatah et al. 2014; Algee-Hewitt and Wheat 2016; Friend and Stock 2016; Lewis and Garvin 2016; Stull et al. 2014) and post-cranial elements (Ali et al. 2018; Decker et al. 2011; Kim et al. 2019; Stull et al. 2014), as opposed to morphoscopic methods, which have concentrated on few skeletal elements, namely the os coxa (Colman et al. 2019; Hisham et al. 2019; Villa et al. 2013; Telmon et al. 2005). Further, both metric and morphoscopic bodies of literature are largely populated by studies employing CT technologies as opposed to other virtual technologies, such as photogrammetry or 3D surface scanning. Virtual anthropology is becoming more appreciated amongst forensic anthropologists however, alternative technologies must also be explored as access to medical imaging equipment (e.g. CT scanners) is limited for most practicing forensic anthropologists. Three-dimensional surface scans have enjoyed success in metric studies (Algee-Hewitt and Wheat 2016; Friend and Stock 2016; Lewis and Garvin 2016), are more accessible (given their more economical price tags), and provide detailed resolution, in some cases greater than that offered by CT scans (Errickson et al. 2014) and, thus, should be investigated for their conveyance of morphoscopic data.

In an investigation of two morphological methods used for sex estimation from the os coxa, Colman et al. (2019) found results to be interchangeable between dry bone and virtual bone for the widely used Klales (2012) method, however, reported lower reliability for the second method investigated. Given such results, additional methods and biological parameters should be investigated for their precision and interpretability from virtual human skeletal elements, as Colman et al. suggest that virtual skeletons are a promising source for modern comparative human skeletal collections (2019: p.1853).

Further, VA investigations of metric assessments have found high accuracy rates and precision as compared to traditional methods using dry bone, with similar reported error rates (Algee-Hewitt and Wheat 2016; Ali et al. 2016; Friend and Stock 2016; Stull et al. 2014; Telmon et al. 2005). Metric methods, however, have been met with greater investigation and rigor than have morphoscopic methods (Bertoglio et al. 2020), thus, precision and classification accuracy rates for the latter are unknown. Morphoscopic studies have also explored few biological parameters, namely age and less frequently sex, and have focused on investigations of the os coxa and few additional skeletal elements (Stull et al. 2014). Further, to the author's knowledge, there are currently no VA studies that investigate morphoscopic traits or methods commonly used for ancestry estimation, save for very few dental traits (i.e. crenelation, reduced hypocone) (Bertoglio et al. 2020). Additionally, very few studies evaluate paired dry bone and virtual samples (CT or otherwise), restricting conclusions regarding the precision and utility of virtual elements that can be drawn from these studies.

Given these voids within the literature, this investigation explores the precision of morphoscopic data obtained from paired dry bone and 3D surface scan samples, specifically morphoscopic traits that inform the parameters of sex and ancestry: Walker's (2008) traits for sex estimation, OSSA (Hefner and Ousley 2014) traits for ancestry estimation, and seven mandibular morphoscopic traits for sex and ancestry estimation.

Statement of Problem

Virtual anthropological investigations of morphoscopic traits, specifically those pertaining to sex and ancestry, remain scant within the literature. The past decade has seen increased efforts at age and sex estimation, however, these VA studies remain focused on CT technologies and are few for 3D surface scanning technologies or investigations of cranial elements. Given the limitations in this research, it is currently unknown whether morphoscopic traits for sex and ancestry estimation can be reliably interpreted from 3D surface scans and, thus, whether 3D surface scans can be utilized as alternatives to physical dry bones.

This study will contribute to virtual anthropology by investigating two widely used morphoscopic methods informing sex and ancestry estimates: Walker's (2008) cranial traits for sex estimation and OSSA (Hefner and Ousley 2014) cranial traits for ancestry estimation. Findings of high precision and low error rates would support the evaluation of these morphoscopic traits virtually and allow for the use of 3D surface scans as alternatives or supplements to physical bone.

Three hypotheses were outlined for investigation in this research. The first hypothesis stated that agreement for cranial traits would exceed that for mandibular traits, regardless of experience. This was hypothesized given the frequency with which cranial traits are referenced throughout introductory forensic anthropology textbooks and courses, likely increasing participants' knowledge, familiarity, and comfort with the included cranial traits, and resulting in greater agreement for these traits. The second hypothesis stated that agreement will increase with experience level for cranial traits, though not necessarily for mandibular traits. Following the first hypothesis, the agreement between elements would likely differ due to lesser familiarity with the mandible, however, would also differ between experience levels, as individuals with professional experience have had more practice at evaluating skeletal remains and identifying the character states of these methods than have graduate students. The third hypothesis stated that the Walker (2008) and OSSA (Hefner and Ousley 2014) traits would display similar levels of agreement between scans and dry bone. This was hypothesized as the level of familiarity and experience between the Walker (2008) and OSSA (Hefner and Ousley 2014) methods would likely be similar and given that most traits within each method are evaluated visually. As such, it was hypothesized that the two methods would be equally challenging to assess.

Summary

Despite increased incorporation of technological methods into forensic anthropology, research in some areas is still lacking, particularly regarding the use of virtual human skeletal remains in research and casework. Research regarding the compatibility of morphoscopic data between virtual human skeletal remains and dry bone is limited, and research specific to 3D surface scans even more so. As such, this study investigated the agreement of cranial and mandibular morphoscopic traits for sex and ancestry estimation between paired dry bone and 3D surface scans.

CHAPTER 2: BACKGROUND AND LITERATURE REVIEW

Background

Virtual anthropology emerged in the 1990s and was largely intended for use with hominoid fossil remains in paleoanthropology (Weber and Bookstein 2011; Weber 2013). Weber (2013) describes VA as a multidisciplinary approach to understanding the functional and comparative morphology of humans, their relatives, and their ancestors by offering novel insights into the internal and external structures of objects by producing virtual models in two, three, and four (i.e. space-time) dimensions and integrating the disparate fields of anthropology, mathematics, statistics, medicine, computer science, and engineering. Virtual anthropology maintained an early focus on metric methods and was largely characterized by fossil reconstruction and geometric morphometric (GMM) techniques (Weber 2013). However, in the decades since its inception, VA has become recognized as a tool for further understanding (macro)morphoscopic osteology by physical anthropologists, specifically forensic anthropologists, and now encompasses a wide realm of techniques used for osteological analyses including metric, morphoscopic, and morphometric techniques (Decker et al. 2011).

In recent decades, technology has been increasingly incorporated into forensic anthropology for such uses as forensic scene documentation and 3D printed evidence for use in the courtroom (Errickson et al. 2014), development of rapid-use triage tools for mass disasters (Ali et al. 2018), development of novel methods for establishing the biological profile (Slice and Algee-Hewitt 2015), and even to aid autopsy procedures by conducting virtual autopsies (“virtopsies”) prior to physical autopsy (Decker et al. 2011; Deduoit et al. 2007). In one virtual autopsy reported by Deduoit et al. (2007), practitioners assessed the utility of metric and morphoscopic methods on MSCT images of human remains. The Suchey-Brooks (1988) pubic symphyseal age estimation method and the İşcan (1985) fourth sternal rib-end age estimation method were used to establish a probable age range for a set of unidentified human remains. Analyses of MSCT images revealed conclusions similar to those obtained from subsequent analyses of dry bone, suggesting that virtual assessments of the biological profile are possible (Dedouit et al. 2007). Researchers regard such virtual methods as "innovative discovery methods" and suggest that they offer novel insights and perspectives into the human skeleton and are even producing better classification rates over some traditional methods (Abdel Fatah et al. 2014: p.591; Decker et al. 2011: p.1107).

Recently, scholarship within VA has increasingly sought the compatibility of metric data collected from dry bone with those from virtual bones (Algee-Hewitt and Wheat 2016; Colman et al. 2019; Friend and Stock 2016). Although the utility of virtual methods for metric data collection appears promising, given several accounts of precision within the literature (Algee-Hewitt and Wheat 2016; Ali et al. 2018; Friend and Stock 2016; Stull et al. 2014), the same cannot be said for the collection of morphoscopic data from virtual models. Morphoscopic traits, particularly those relevant to sex and ancestry estimation, have been the focus of far less scholarship within VA. Most anthropomorphic investigations have focused on age estimation from the os coxa and, to the author's knowledge, there have not been any similar investigations focused on ancestry traits. Despite a progressive shift towards a virtual forensic anthropology, as described by Colman and colleagues (2019) it is, as yet, unknown whether morphoscopic methods designed for use on dry bone can be reliably applied to virtual remains. As such, the applicability of traditional morphoscopic methods to virtual bone will be the central question of this investigation.

This chapter will describe a wide range of research that has been performed within VA, limitations of and gaps within the current literature, and the potential for 3D surface scans to contribute to VA.

Morphoscopic Traits

Forensic anthropologists use a number of analytical methods to construct the biological profile. Analytical methods can be categorized as metric or nonmetric (morphoscopic) in nature and are both commonly used, though the latter are frequently regarded as more subjective and less reliable given the choice and variability in trait selection left to the analyst (Hefner and Ousley 2014; Petaros et al. 2017). In recent decades, however, measures have been taken to increase the precision, reliability, and reproducibility of methods (Decker et al. 2011), both metric and morphoscopic in nature, by standardizing data collection methods (Klales 2013), incorporating more robust statistical procedures (Hefner and Ousley 2014), increasing sample sizes (Warren et al. 2008), and discerning error rates for widely used methods to make them admissible in courts of law (Williams and Rogers 2006).

Metric versus Morphoscopic

Metric methods are those that necessitate the use of specialized measuring devices (e.g. spreading calipers, mandibulometer, osteometric board) to evaluate continuous, quantitative traits. Continuous skeletal traits are standardized measurements and include long bone lengths, shaft circumferences, orifice diameters, and points of maximum and minimum curvature, among others (Langley et al. 2016b). Standardized measurements can be used to calculate discriminant functions by hand to estimate sex, following schemes such as Spradley and Jantz, 2011, or can be entered into the program FORDISC to obtain estimates of sex, ancestry, and stature (Ousley and Jantz 1998). FORDISC is a computer software program that uses discriminate function analysis and linear regression to classify unidentified human remains using known samples from the Forensic Databank (Klales 2020; Ousley and Jantz 1998; Walrath et al. 2004). Given the use of rigorously tested and standardized measurements (Hefner and Linde 2018; Hefner 2003), and the calibrated measuring devices of metric methods, they are typically highlighted as less subjective and more reliable (Langley et al. 2018).

Contrastingly, morphoscopic methods evaluate quasicontinuous or discrete traits of the cranial and post-cranial skeleton. Morphoscopic traits are evaluated visually following line drawings, illustrated guides or atlases, and with an understanding of the extensive human skeletal variation amongst traits. Traits of a given element may be evaluated relative to the size or shape of a neighboring feature or bony structure (e.g. mastoid process relative to EAM), though are typically evaluated without the use of additional objects or measuring aids. Researchers contend that morphoscopic methods are inherently more subjective, from how individual traits are evaluated and scored to the traits chosen to assess a given biological parameter (e.g. sex, ancestry) (Byrnes et al. 2017; Hefner and Linde 2018; Hefner and Ousley 2014; Langley et al. 2018; Rogers 2005). Over the past two decades, however, steps have been taken to mitigate subjectivity in morphoscopic methods, particularly through standardization of trait and character state definitions, the development of line drawings (Walker 2008) and atlases illustrating character states - and variation within and between states - and, importantly, by urging implementation of these definitions and pictorial models in research, method development, and forensic casework (Hefner and Linde 2018).

As analysts may select divergent traits, trait lists, or morphoscopic methods that utilize different skeletal elements or skeletal regions entirely (e.g. cranial elements, versus post cranial) (Hefner and Ousley 2014), morphoscopic methods have been re-considered and approached with heightened systematic and statistical diligence in recent decades (Hefner and Linde 2018; Pink et al. 2016). Despite greater subjectivity, morphoscopic traits are continuously and routinely used by forensic anthropologists as they can be applied to fragmentary or poorly preserved remains, quickly, without use of additional tools, and in field contexts, if necessary (Buikstra and Ubelaker 1994; Klales 2020; Lewis and Garvin 2016; Petaros et al. 2017; Pink et al. 2016; Walker 2008; Walrath et al. 2004; Warren et al. 2008). Morphoscopic methods are also favored for their ability to better capture subtle variation in shape that is often difficult or impossible to capture via metric methods (Decker et al. 2011; Kimmerle et al. 2008; Lewis and Garvin 2016; Walker 2008; Warren et al. 2008). Further, several morphoscopic investigations and validation studies have also reported high classification accuracies, supporting continued application of these methods within forensic contexts (Hefner and Ousley 2014; Kenyhercz et al. 2017; Pink et al. 2016; Walker 2008).

Morphoscopic versus Nonmetric

Morphoscopic traits are frequently referred to as nonmetric traits (Buikstra and Ubelaker 1994; Hefner and Linde 2018). Morphoscopic and nonmetric traits are closely related, sharing similar characteristics of a quasicontinuous or discrete, regularly variable, and non-pathological nature (Hefner and Linde 2018). Morphoscopic and nonmetric traits have different application histories and are used to different ends by biological and forensic anthropologists (i.e. for ancestry and sex estimation, respectively) (Pink et al. 2016). Nonmetric traits are frequently used in analyses of biological distance (biodistance) to discern biological relatedness within or between populations in biological anthropological and bioarchaeological contexts (Buikstra and Ubelaker 1994; Pink et al. 2016). Nonmetric traits are often dichotomous in nature and are used in such analyses as it is known that their trait states are commonly inherited (Buikstra and Ubelaker 1994; Pink et al. 2016). Buikstra and Ubelaker divide nonmetric traits into four classes: 1) extrasutural bone, 2) abnormal ossifications, 3) ossification failure, and 4) variation in foramen number and location (1994: p.85; Hefner and Linde 2018: p.2; Pink et al. 2016: p.92).

Morphoscopic traits are most commonly recognized and utilized in ancestry estimation (Pink et al. 2016). Morphoscopic traits are typically used to estimate the population (ancestral) affinity of a single individual of unknown identity within a forensic context. Hefner and Linde divide morphoscopic traits into a series of five classes or trait types: 1) bone shape, 2) bony feature, 3) suture shape, 4) absence or presence of feature, and 5) feature prominence or protrusion (2018: p.2).

Nonmetric and morphoscopic traits have long informed the biological profile, and are favored for their expeditious, simple, and effective nature, as mentioned previously. Amongst these methods are the Walker (2008) method for sex estimation (Lewis and Garvin 2016), and Hefner and Ousley's (2014) OSSA method, two of the most commonly employed methods by forensic anthropologists. Similarly, the morphoscopic mandibular traits evaluated are among those most frequently recognized as useful for sex and ancestry estimation. Brief summaries of these methods are presented below, and trait state diagrams are located in Chapter 3: Materials and Methods. (The author recognizes recent and ongoing conversations regarding morphoscopic traits (c.f. Bethard and Digangi 2020; Digangi and Bethard 2021; Stull et al. 2020), however discussion of these conversations is beyond the scope of the current research project.)

Precision versus Accuracy

This investigation seeks to define the precision of morphoscopic data from 3D surface scans in comparison to those from dry bone. As these data are subjective interpretations of dry bone and virtual models, the goal is not to obtain the correct answer (accuracy), as there is no correct answer, but to obtain high agreement between scores assigned to divergent formats (e.g. dry and virtual). Precision is achieved when similar results are obtained through repeated assessments, indicating little variation and good reliability of the method, whereas accuracy speaks to the ability to obtain a true value (Stull et al. 2014). High precision of VA methods in this investigation will be indicated by high agreement between dry and virtual formats.

Why Surface Scans?

Freiss (2012) notes that digitizing techniques and CT imaging have largely dominated the emergent realm of virtual anthropology. However, in recent years, surface scanning has joined these techniques as one of the most widely recognized methods of 3D image construction (Algee-Hewitt and Wheat 2016; Freiss 2012), gaining much of its recognition for its ability to provide accurate and precise metric data without the need for exceedingly expensive or medical-grade equipment. Despite this trend, however, many suggest that surface scanning techniques and their applications remain in their infancy in forensic, and other, anthropological subdisciplines (Freiss 2012; Garvin and Stock 2016). This notion seems particularly true in regard to virtual assessment of sex and ancestry. Most literature using 3D surface scans pertains to methods of age estimation using the os coxa and metric data collection from the cranium. Additionally, investigations of intra- and interobserver error between 3D virtual skeletal elements and their dry bone counterparts are few, suggesting this aspect of VA has not yet been thoroughly explored.

Furthermore, three-dimensional surface scans are relatively easy to produce, with smaller learning curves than alternative 3D imaging devices (e.g. CT scanners), and are generally quicker to generate (Freiss 2012). The NextEngine 3D Ultra HD laser scanner is far cheaper than many other 3D surface scanners, starting at around $3000 (NextEngine Inc., Malibu, CA), making them a cost-effective option for forensic anthropology labs and eliminating the cost of lab fees, regular maintenance, or equipment-upkeep fees that may be required of medical technology.

Surface scanning techniques, similar to CT scanning techniques, are “non-contact methods” which contribute to efforts of conservation and preservation of skeletal remains (Freiss 2012: p.8). CT scanners operate via emission of X-rays, a type of ionizing radiation, directed at the object of interest (Freiss 2012). Although CT scanners are commonly used on human patients, these types of radiation have the potential to be harmful with prolonged exposure. The harmful potential of CT radiation to skeletal remains is not well understood at this point, however, limited research suggests the adverse effects of DNA fragmentation within osseous tissues and permanent discoloration of dental remains can occur (Freiss 2012). Surface scanners, on the other hand, operate with non-ionizing radiation, using lasers or white light to capture an object’s surface morphology (Freiss 2012). As such, surface scanning is considered both a non-destructive and non-invasive measure (Ramsthaler et al. 2010), a particularly important notion when DNA extraction is also to be performed on a set of skeletal remains for identification purposes.

Further advantages of 3D surface scans are the lack for need of a certified radiologist to interpret the images (Dedouit et al. 2014) and the ability of users to manipulate the object of interest during image evaluation, offering views from multiple perspectives, similar to analyses of dry bone (Freiss 2012). The NextEngine 3D surface scanner can also preserve the color (texture) of an object allowing the observation of some contextual data (e.g. sun bleaching, soil staining, metal staining, etc.), although this may be at the expense of a quick acquisition time (Freiss 2012). The NextEngine laser scanner also comes accompanied by ScanStudio (NextEngine Inc., Malibu, CA), a 3D model processing and editing software, eliminating the need for users to invest in software or pay licensing fees. A number of studies that have employed NextEngine generated surface scans in the collection of metric data have also found high precision and accuracy (Algee-Hewitt and Wheat 2016; Friend and Stock 2016; Sholts et al. 2011), a notion that highlights the potential utility of NextEngine 3D surface scans for morphoscopic data collection. Recent studies have also implemented 3D surface scans in the development of a novel quantitative age estimation method utilizing the morphology of the pubic symphysis, achieving high accuracy rates comparable to, or better than, those using physical bone (Slice and Algee-Hewitt 2015).

Literature Review

In recent scholarship, digitally formatted human skeletal elements (e.g. CT scans, 3D surface scans) have been used for quantitative analyses, typically through the employment of geometric morphometric (GMM) techniques (Garvin and Stock 2016). These GMM techniques have typically taken one of two forms, (1) extracting traditional anthropometric data from digital skeletal remains or (2) have used shape analyses to examine patterns in morphological variation (Freiss 2012; Garvin and Stock 2016), most frequently in novel investigations of sex and ancestry. Techniques of the latter form have also been used to obtain anthropometric data previously unavailable to researchers through the simple caliper measurements that characterize traditional data collection methods (e.g. volume, surface area). Most of these studies, however, have employed CT scans in their investigations. Few studies have used 3D surface scans. Investigations employing 3D surface scans have reported low error rates (≤ ~2mm) and high precision between this and other methods (e.g. measurements from dry bone and Microscribe digitizer) (Freiss 2012; Friend and Stock 2016; Algee-Hewitt and Wheat 2016).

Anthropometric Investigations in VA

Recently, Friend and Stock (2016) demonstrated precision in standard craniometric measurements taken from NextEngine 3D surface scans. In this investigation, two observers compared measurements taken on dry bone to corresponding measurements taken on virtual crania using the GMM package Geomorph in R. Their results revealed interobserver errors of 1.39mm for dry bone measurements and 2.14mm for virtual measurements, and intraobserver error between dry bone and virtual measurements of 2.07mm for observer 1 and 1.58mm for observer 2 (Friend and Stock, 2016). These error ranges are very close to the ± 2mm error range traditionally accepted in analyses of dry bone (Langley et al. 2018) and support the use of 3D surface scans in metric assessments. These results are also very similar to those reported in studies employing CT scans, suggestive that 3D surface scans may be just as useful in virtual assessment of the biological profile.

Abdel Fatah et al. (2014) also investigated improved sex estimation techniques through the identification of endo- and ecto-cranial sexual dimorphism using 3D CT scans. Through observation of 43 measurements, differences between dry bone and computer-automated measurements were small, with errors of < 2mm, a finding in good accordance with similar research and acceptable under current standards for measurement errors of ≤ 2mm (Abdel Fatah et al. 2014; Langley et al. 2018).

Similarly, Sholts et al. (2011) demonstrated minimal differences in measurements obtained from virtual cranial models generated by multiple observers. In a novel investigation of cranial volume and surface area, NextEngine 3D surface scans of six crania were used to investigate the relationship between model quality and data quality in relation to differential scan generation. Comparisons of cranial volume and surface area across six unique models, three generated by each of two observers, revealed small mean volume and area differences. Models independently generated by observers 1 and 2 revealed averaged errors of approximately 1.6% for cranial volume and 2.6% for surface area (Sholts et al. 2011). These results suggest good replicability between virtual assessments and smaller error between images generated under different settings than reported for studies using CT imaging.

Sholts and Wärmländer (2012) also use 3D surface scans in a novel GMM method for ancestry estimation using the zygomaticomaxillary (ZMS) suture. Using 120 craniofacial scans, linear discriminant functions were developed through Fourier transformation of suture outlines and were used to predict ancestry of European and Native American individuals with an average accuracy rate of 83%. Sholts and Wärmländer report that this method achieves higher classification accuracies for these ancestral populations than does the traditional "typological" approach of cranial morphoscopic assessments (2012: p. 234.e5). Furthermore, evidence of sexual dimorphism in ZMS outline was also found, though not discussed, through this digital morphometric approach, supporting the notion that digital morphometric techniques, and virtual models in general, offer novel insights unavailable through simple linear measurements or macroscopic evaluations.

Further, Nuzzolese et al. (2019) used panoramic radiographs of the mandible in a GMM analysis to investigate patterns of sexual dimorphism between males and females.

Using a series of 25 landmarks, sexually dimorphic variation was found to be significant in the mandible, presenting correct sex classification rates of 94%. Such promising results of GMM analyses of the mandible highlight the presence of sexually dimorphic shape variation within the mandible and the potential for use of this bone in morphoscopic investigations as well.

Few post-cranial skeletal elements have been investigated in metric-focused VA studies. Similar to findings of cranial studies, however, post-cranial investigations have reported good accuracy and low error rates. In a recent study, Ali et al. (2018) assessed the accuracy of measurements of virtual scapulae and their ability to predict sex following Dabbs and Moore-Jansen's (2010) two-variable discriminant function and a novel logistic regression discriminant function developed by the authors. The Dabbs and Moore-Jansen (2010) method developed for use on dry bone was tested on virtual scapulae from postmortem CT scans. A sample of 290 scapulae were measured for maximum height and maximum length, the two measurements required of both the Dabbs and Moore-Jansen (2010) discriminant function and novel logistic discriminant function introduced. The Dabbs and Moore-Jansen method presented a correct sex classification rate of 94.1% and Ali et al.'s (2018) equation revealed a slightly higher correct sex classification accuracy of 94.5%. Ali et al.'s (2018) virtual validation of the Dabbs and Moore-Jansen method and findings of their novel method support notions that metric data obtained from virtual skeletal elements are precise and can be reliably used in place of from dry bone. Such findings also suggest that metric methods developed for use on dry bone can be reliably applied to virtual skeletal elements.

In a more comprehensive study, Stull et al. (2014) investigate the accuracy of both cranial and post-cranial measurements obtained from volumetric CT scans. Measurements obtained from the cranium and mandible as well as the clavicle, scapula, humerus, radius, ulna, femur, tibia, and fibula were compared in this study. All skeletal measurements were taken within three contexts, CT scans of elements *in situ*, CT scans of dry elements, and physical measurements of dry elements. Mean percent differences across the three data sources ranged from 0.9 to 3.7%, with cranial measurements demonstrating smaller percent differences than post-cranial elements, and most inter-format comparisons presenting differences of < 2mm from dry bone measurements. Dry bone versus dry CT and dry bone versus *in situ* CT comparisons presented similar results, with the former performing slightly better, and both outperform the dry bone versus *in situ* CT comparison. Additionally, in the same investigation, measurements from CT scans of seven dry crania were compared to paired dry bone measurements by each of three observers. Mean percentage Technical Error of Measurement (%TEM) for intraobserver comparisons were 0.56 to 1.06% (0.46-0.71mm) and mean %TEM for interobserver comparisons were 2.26% (2.6mm). These results demonstrate high accuracy and error rates similar to those reported in studies using 3D surface scans (Friend and Stock 2016), as well as good replicability between formats and observers. This study also addresses the paucity of research on post-cranial measurements from virtual sources, as limited elements have been investigated in VA studies of inter-format comparisons.

Decker et al. (2011) also proposed a novel metric approach to pelvic sex estimation using two traditional nonmetric traits and one common medical index, not typically used in anthropometric evaluations. Using *in situ* CT scans of 100 pelves, a series of 35 landmarks were placed and 20 measurements extracted. An interobserver test of 10 of these measurements revealed minimal differences with an average overall error of 2.22%. Four of the original 20 measures - width of sciatic notch, innominate height, sub-pubic angle, and transverse pelvic outlet - were found to have the greatest impact on sex estimation and high correlations between right and left sides and were, thus, used to develop a 4-variable logistic regression for sex estimation from the pelvis. An accuracy rate of 100% was found following leave-one-out cross validation using this logistic regression. Sex estimation subsequently determined by FORDISC 3.0 revealed an average sex classification accuracy of 86%. As this method appears to improve upon those relying on simple linear caliper measurements, further consideration of medical indices is recommended within anthropometric contexts. Additionally, the low error rates reported in this study are similar to those reported in similar investigations and contribute to the body of literature supporting the use of virtual skeletons for metric data collection.

Anthropomorphic Investigations in VA

In contrast with investigations of metric data, virtual skeletal elements have not been extensively evaluated for the collection of morphoscopic data. This body of research is small and largely populated by studies employing CT images in investigations of age estimation. These studies also frequently focus on the os coxa, primarily the pubic symphysis; fewer studies have investigated the auricular surface of the os coxa, the cranium, or the mandible. Additionally, few investigations of morphoscopic traits have assessed inter-format precision - comparing dry bone data with those obtained virtually - leaving precision, reliability, and replicability of these data, and data collection methods, unknown.

In one of few such studies, Villa et al. (2013) address this paucity in the literature with a multi-method investigation of the precision of morphoscopic data from dry bone, CT scans, and 3D surface scans. Following the Suchey-Brooks (1990) pubic symphysis and Buckberry-

Chamberlain (2002) auricular surface age estimation methods, three observers compared age estimations of 12 pubic symphyses and eight auricular surfaces across the three formats. Results of their investigation revealed greatest agreement between dry bone and 3D surface scans, though presented substantial agreement only for the auricular surface (Kappa value = 0.66) (Villa et al. 2013). Comparisons of dry bone and 3D surface scans for the pubic symphysis, as well as all comparisons of CT scans with dry bone, presented only fair agreement (Kappa values = 0.26-0.36) (Villa et al., 2013). Comparisons were also performed between the two virtual formats (CT scans and 3D surface scans), revealing moderate agreement for both the pubic symphysis and the auricular surface (Kappa value = 0.45 and 0.59, respectively). These results support the findings of agreement between the respective virtual methods and dry bone. These findings are supported by Freiss (2012) in a brief review of 3D surface scanning technology, who describes similar heightened compatibility of dry bone with 3D surface scans, attributing greater compatibility to the user's ability to rotate and manipulate 3D surface scans in (morpho)space, offering users observations at multiple angles, very similar to in-person analyses of dry bone.

In an earlier investigation, Telmon et al. (2005) similarly investigated the utility of CT scans of the pubic symphyseal surface in estimating Suchey-Brooks defined phases, contrasting results with traditional methods using dry bone. This study appears to be one of the first to compare physical methods to virtual methods and demonstrates high precision with age estimates based on phase assignment, with Kappa values ≥ 0.82 exhibiting almost perfect agreement between formats (Telmon et al. 2005). Furthermore, this study remains one of few published studies comparing divergent formats and, as such, has served as much the basis for more recent and similar studies (Villa et al. 2013).

Hisham et al. (2019) also applied the Suchey-Brooks pubic symphysis age estimation method to a large virtual Malaysian sample. This study sought to determine the applicability of this method to Malaysian skeletal remains, as standardized biological profile methods are few for this population (Hisham et al. 2019). Results of their study were promising with an average sex classification accuracy of 97.18% and substantial to almost perfect agreement for both intra- and interobserver comparisons of phase assignment (Kappa values = 0.832 and 0.763, respectively). Of great import is that this study sought to validate use of the Suchey-Brooks method on a Malaysian population, not to determine the precision of data obtained virtually. This study does, however, demonstrate the applicability of traditional dry-bone methods to virtual elements, as well as great replicability of virtual data collected from MDCT scans. As these data were not compared to dry bone data, however, conclusions regarding consistency or precision between CT images and dry bone, or use of virtually developed standards on dry bone, cannot be made from this study.

Furthermore, and similar to Hisham et al. (2019), Savall et al. (2016) tested the applicability of the Suchey-Brooks age estimation method on a modern French population. MSCT scans representative of each stage were established and used to estimate the age of 680 male MSCT pubic symphyses. Repeat assessments on 80 specimens and evaluation of the subsample by a second observer, allowed for examination of both intra- and interobserver reliability. Results of both reliability tests were high, presenting Kappa values of 0.88 and 0.73, respectively. Inaccuracy and bias, however, were large for individuals over the age of 55, and both erred by more than 20 years for individuals over the age of 65, suggesting this method may not be the most reliable when applied to a modern French population, and that a new method may need to be devised (Savall et al. 2016). Substantial agreement was found between observers, however, suggesting good replicability of virtual evaluations and similar to findings of previous VA investigations. Conclusions regarding the use of the Suchey-Brooks method on modern French individuals, however, may be spurious and result from misinterpretation of features in a virtual format, as little is understood regarding the use of dry bone methods on virtual bone. As such, further investigations should consider the compatibility between virtual and dry formats when investigating this method.

A small body of research also focuses on the utility of virtual skeletal models for the estimation of sex; this literature, however, is also primarily restricted to investigations of the os coxa. Specifically, Johnstone-Belford et al. (2018) examine the applicability of Phenice's (1969) sex estimation method to MDCT images and the replicability of these data. Although data from MDCT images were not directly compared to data from their dry bone counterparts, as was done in previous studies (Telmon et al. 2005; Villa et al. 2013), agreement of repeated virtual assessments was found to be almost perfect (Kappa value = 0.91-0.98) and correct sex classification rates averaged 92.41% for pooled males and females (Johnstone-Belford et al. 2018). These results support the notion that CT images can be evaluated with good replicability and that the Phenice sex estimation method can correctly predict sex on MSCT scans almost as accurately as when applied to dry bone (1969: p.298).

Further, and similar to Telmon et al. (2005) and Villa et al. (2013), Colman et al. (2019) investigated the precision of morphoscopic traits of the os coxa, examining two sets of traits on CT scans: Klales (2012) traits for sex estimation of the pubis and a set of five morphoscopic traits recognized by the Workshop of European Anthropologists (WEA traits), which includes the preauricular sulcus, greater sciatic notch, pubic angle, arc compose, and ischial body. CT scans of 27 individuals were scored for all pelvic traits using two different virtual platforms, MeshLab and an unspecified in-house software program, alongside the paired sample of dry bones (Colman et al. 2019). Dry-virtual score agreement using both software programs was moderate to almost perfect for both Klales (2012) and WEA traits (Kappa values = 0.58-0.84 and 0.42-0.87, respectively). Use of the in-house software revealed greater dry-virtual agreement for Klales (2012) traits than for WEA traits (Kappa values = 0.74-0.84 and 0.42-0.85, respectively). Agreement of Klales (2012) traits using MeshLab was slightly lower than that for in-house software, with Kappa values of 0.58-0.76. Results for WEA traits were similar between viewing softwares however, did exhibit a broader range of values when using MeshLab (Kappa values of 0.42-0.85 and 0.33-0.87, respectively). Further, inter-observer results for repeat dry and virtual assessments of Klales (2012) traits presented agreement of 0.61-0.77 for dry comparisons, 0.62-0.72 for in-house software comparisons, and 0.62-0.80 for MeshLab comparisons. Inter-observer results for WEA repeat assessments presented agreement of 0.24-0.88 for dry comparisons, 0.33-0.83 for in-house software comparisons, and 0.05-0.83 for MeshLab comparisons. Results of this study demonstrate promise for virtual sex estimation, particularly using Klales (2012) traits, as agreement was typically found to be moderate, substantial or almost perfect. The findings of this study, however, also suggest that viewing environment should be carefully considered in standardization of virtual methods, as it may have an impact on data interpretation.

Ramsthaler et al. (2010) also present one of few investigations of cranial morphoscopic traits. In this study, four observers investigated a suite of 17 morphoscopic cranial traits used for the estimation of sex on a sample of 50 cranial CT (CCT) scans. Of the 17 traits observed, arcus superciliaris, glabella, processus mastoideus, and margo supraorbitalis - traits closely resembling Walker's (2008) cranial morphoscopic traits for sex estimation - performed best, with individual accuracy rates ranging from 69 to 85% (Ramsthaler et al. 2010). Processus mastoideus and margo supraorbitalis both achieved 69% accuracy, glabella 81% accuracy, and arcus superciliaris 85% accuracy. Walker's (2008) univariate classification accuracies for mastoid process, supra-orbital margin, and glabella were 78.6, 68.8, and 82.6%, respectively. As can be seen, Ramsthalter et al.'s (2010) accuracy rates are comparable to Walker's (2008) however, Ramsthaler et al. (2010) divide two of Walker's (2008) traits into two constituent traits - recognizing supraorbital margin as glabella and arcus superciliaris, separately, and nuchal crest as protuberantia occipitalis and planum nuchale. These four constituent traits achieve accuracy rates of 60 to 85%, with both supra-orbital margin traits outperforming both nuchal crest traits. Additionally, an average sex classification accuracy of 96% was found, demonstrating great promise for the utility of CCT scans in cranial morphoscopic data collection, as well as for the implementation of virtual reference collections as proxies for physical collections when the latter are unavailable.

Finally, Bertoglio et al. (2020) explored the compatibility of cranial morphoscopic data observed on dry bone with data from 3D CT reconstructions. Twenty-nine morphological features of the cranium were investigated for compatible identification and bone quality on CT reconstructions. Twenty of the 29 traits investigated presented almost perfect agreement between formats, with Kappa values ranging from 0.61 to 1.00, however, only six of the 20 traits were correctly scored by the two observers. These six traits corresponded to various facial foramina and medium to large sized morphoscopic cranial features (i.e. inter-sutural bones). Although this study reports good agreement for only one third of the traits investigated, the primary goal of this study was to evaluate the use of CT reconstructions as proxies to physical bones, rather than to validate particular methods for use on virtual remains. This study revealed important findings regarding the weaknesses of CT reconstructions, namely consistent bone loss in regions characterized by thin compact bone, specifically in the maxillary region, which could lead to spurious findings in investigations using virtual remains.

Limitations of Previous Research

Researchers have increasingly engaged in investigations of virtual remains for the construction of the biological profile, however, divergent methods and technologies have not been addressed to the same extent within the literature (Decker et al. 2011; Stull et al. 2014). As has been previously noted, the majority of these studies have pertained to metric investigations and have most frequently implemented CT scans. Morphoscopic investigations in VA have examined few traits or methods and have largely targeted the parameters of age and sex, very few investigations have explored morphoscopic traits informative of ancestry.

Morphoscopic studies in VA have reported virtual assessments of the Suchey-Brooks (1988) pubic symphysis age estimation method (Hisham et al. 2019; Savall et al. 2016; Telmon et al. 2005; Villa et al. 2013), the Phenice (1969) pubis sex estimation method (Johnstone-Belford et al., 2018), the Buckberry-Chamberlain (2002) auricular surface age estimation method (Villa et al. 2013), the İşcan (1985) fourth sternal rib-end age estimation method (Deduoit et al. 2007), the Klales (2012) pubis sex estimation method (Colman et al. 2019), as well as several traits that fall outside of these methods. The Suchey-Brooks (1988) pubic symphysis age estimation method has received the greatest attention of these methods, as indicated in the previous section. Several studies investigating this method, however, have been virtual investigations of the applicability of this method to various populations lacking population specific standards - in other words, validation studies - rather than investigations of the precision between data from dry bone and virtual bone. As such, conclusions regarding the compatibility of data from dry bone and virtual remains and the resulting compilation of these data for research, or the utility of virtual skeletal collections by those who lack access to physical collections, cannot be drawn. As correct classification accuracies were high for virtual data in the above virtual validation studies, it would seem that there is agreement between formats, however, research directly comparing the two formats must also be performed.

Moreover, investigations of Walker's (2008) cranial morphoscopic traits for sex estimation and OSSA (Hefner and Ousley, 2014) cranial morphoscopic traits for ancestry estimation are seemingly absent from the VA literature. This is surprising given the frequency with which Walker (2008) traits, in particular, are used in forensic investigations and the extensive discussion of these traits throughout the literature as useful and typically only second to those of the pelvis (Langley and Dudzik 2016; Stull et al. 2013).

As no known virtual anthropological research has investigated these methods for sex or ancestry estimation, or the aforementioned morphoscopic traits of the mandible, save for chin shape, this research will contribute to these voids within the literature.

Research Objectives and Goals

The primary objectives of this research were to 1) investigate the precision of methods developed for use on dry bone on virtual bone, specifically morphoscopic methods for sex and ancestry estimation, and 2) to understand the utility of 3D surface scans as alternatives to dry bone for morphoscopic data collection.

Four specific research questions were established for investigation in this study: 1) Do scores assigned to scans agree with those assigned to dry bone, 2) Is agreement different between the Walker (2008) and OSSA (Hefner and Ousley 2014) methods, 3) Do cranial traits exhibit greater agreement than mandibular traits, and 4) Do individuals with more experience exhibit greater agreement in scores than those with less?

Further aims of this research were to understand the utility of a technology that is accessible to forensic anthropologists working outside of medical examiner's offices, or those with limited access to medical imaging technology, and to contribute to the ongoing discussions regarding the use of virtual skeletal collections as proxies for comparative or research collections and in forensic settings to facilitate casework at a distance or to preserve evidence.

CHAPTER 3: MATERIALS AND METHODS

Sample

Three-dimensional surface scans were generated for 92 crania and mandibles of donors from the William M. Bass Donated Skeletal Collection housed at the University of Tennessee, Knoxville (UTK). Crania constituted 42 of these specimens and mandibles the remaining 50. Crania and mandibles were not from the same individuals. Both males (n = 58) and females (n = 34) were included in the total sample, as were individuals of Black (n = 21), White (n = 69), Asian (n = 1), and Hispanic (n = 1) ancestral affinity. The cranial sample was comprised of Black (n = 21) and White (n = 21) individuals, as the OSSA method (Hefner and Ousley, 2014) currently only differentiates between these two groups, and the mandibular sample was more diverse, as can be seen in Table 3.1, below. The mean age amongst the cranial sample was 50.5, with a range of 23-88 years, and 43 amongst the mandibular sample, with a range of 18-81 years.

Exclusion criteria for the crania were not confined to any one region, as the morphoscopic traits under investigation span the facial cranium, the neurocranium, and the cranial vault. Crania presenting obvious trauma or pathologies were excluded, as were juveniles and those autopsied or damaged in such a way so as to obscure any morphological traits under investigation. Mandibles were included on the basis of no visible trauma or pathology to the ascending ramus, corpus, or coronoid process, minimal loss of dentition, non-extensive alveolar resorption, and no medical interventions present (e.g. reconstructive metal mesh, plates, or screws). Dental implants were not an exclusion criteria provided the surrounding alveolar bone did not appear pathological.

Table 3.1: Sample Demographics.

	Ancestry	N
Cranial Sample (Sex and Ancestry)	Black Americans	21
	White Americans	21
Mandibular Sample (Sex and Ancestry)	White Americans	48
	Hispanic Americans	1
	Asian Americans	1

Morphoscopic Methods Investigated

Walker Method

In accordance with the prevailing re-evaluation of metric and nonmetric methodologies of the early 21st century, Walker (2008) expanded on Buikstra and Ubelaker's (1994) widely used cranial and mandibular morphoscopic traits for sex estimation. Buisktra and Ubelaker's (1994) method, adapted from an earlier method developed by Ascadi and Nemeskeri (1970), includes the nucal crest, mastoid process, supra-orbital margin, supra-orbital ridge (glabella), and mental eminence and was presented in the bioarchaeological guide *Standards for Data Collection from Human Skeletal Remains*. Buikstra and Ubelaker's (1994) method evaluates the same traits as Ascadi and Nemeskeri (1970) but modifies the scoring method, re-assigning ordinal scores of 1 to 5 to character states. This novel numbering system functioned to remove the bias that 0 was the optimal sectioning point between males and females following the previous -2 to +2 scale (Walker 2008).

To mitigate the subjectivity of this method, Walker clarified character state descriptions and developed a series of corresponding univariate and multivariate discriminant function models based on the ordinal scores assigned to these five cranial and mandibular traits. Linear, kth-nearest-neighbor, logistic, and quadratic discriminant functions were performed on combinations of two to all five morphoscopic traits. Logistic regression analyses were found to perform the best, providing high classification accuracies ranging from 84 to 87.8% and low sex biases of approximately 1 to 2%. Walker's (2008) proposed logistic discriminant equations include six formulae appropriate for use with American and English populations and three formulae for use with Native American populations. Further, Walker provides equations that analysts can use to determine male and female posterior probabilities using the discriminant function results obtained (2008: p.47). Validation studies and similar investigations of these traits have also enjoyed high accuracy, except for the mental eminence, supporting continued use of this method and its constituent traits (Lewis and Garvin 2016; Williams and Rogers 2006).

OSSA Method

To address the lack of statistical rigor surrounding morphoscopic traits used in ancestry estimation, Hefner and Ousley developed the Optimized Summed Scored Attributes (OSSA) method (2014; Tallman and Go 2018). The OSSA method is a "heuristic" method based on six morphoscopic traits of the craniofacial skeleton and skeletal vault and their frequency distributions amongst Black and White American populations (Hefner and Ousley 2014: p.884). In this method, anterior nasal spine (ANS), inferior nasal aperture (INA), inter orbital breadth (IOB), nasal aperture width (NAW), nasal bone contour (NBC), and post bregmatic depression (PBD) are scored ordinally on a scale from 1 to 5, with scores closer to 1 representing more gracile morphologies and those closer to 5 more robust morphologies, or discretely as present/absent on a scale from 0 to 1 (Hefner 2009; Tallman and Go 2018). Following previous findings of character state frequency distributions, the scores for each trait are dichotomized into scores of 0 or 1 and are based on the character states most commonly associated with Black and White American populations. Dichotomization of all trait scores into 0 or 1 helps to achieve maximum difference between these two ancestral populations (Hefner and Ousley 2014; Tallman and Go 2018). The dichotomized scores of all six traits are then summed totaling a score ranging from 0 to 6. This final OSSA score is used against a sectioning point of 3, whereby individuals with scores ranging from 0 to 3 are classified as

American Blacks and those with scores from 4 to 6 as American Whites (Hefner and Ousley 2014).

In their investigation of 11 novel ancestry classification methods, Hefner and Ousley (2014) reported that the OSSA method achieved classification accuracies roughly as good as, or better than, the 10 additional methods investigated, including several machine learning methods (e.g. artificial neural networks, support vector machines, random forest models, and decision trees). Further, validation studies of the OSSA method corroborate the effectiveness of the method, although Kenyhercz et al. (2017) report increased classification accuracies following a slightly higher sectioning point of 4 in comparison to the proposed 3. Tallman and Go (2018) further validated the effectiveness of the OSSA model in an investigation that adapted the ancestry classification method into a tool for estimating sex amongst Japanese, Thai, and Filipino populations. Correct sex classification accuracies of 81-82% were found using the modified OSSA method (Tallman and Go 2018), providing further support for the OSSA method's effectiveness.

Hefner and Linde (2018) write that nearing the end of the 20th century, and continuing to present day, forensic anthropology has seen a re-evaluation of methods, spanning all parameters of the biological profile, and lending to a need for increased statistical rigor, known error rates, and reproducibility. These two methods both incorporate widely used and recommended morphoscopic traits for estimations of sex (Klales 2020) and ancestry (Hefner 2009) and, thus, were chosen for investigation in this research. These methods are also curiously absent from the VA literature and their importance warrants investigations of their utility in the proposed virtual method.

Traits Investigated

Cranial Traits

Ten cranial traits and seven mandibular traits were evaluated for paired dry bone and virtual 3D models. The cranial traits considered in this investigation were Walker's (2008) cranial traits for sex estimation: supra-orbital margin, supra-orbital ridge (glabella), mastoid process, and nuchal crest, depicted in Figure 3.1, and OSSA (Hefner and Ousley 2014) cranial traits for ancestry estimation: anterior nasal spine, inferior nasal aperture, inter orbital breadth, nasal aperture width, nasal bone contour, and post-bregmatic depression, depicted in Figures 3.2-3.7. Walker's (2008) mandibular trait, mental eminence, was excluded from this study as approximately one-third of the cranially paired mandibles exhibited extensive alveolar resorption extending inferiorly into the mandibular corpus, and a large proportion of these pathological mandibles were from Black individuals. As OSSA only distinguishes between Black and White individuals, a Black sample was needed and, thus, prioritized over the inclusion of the mental eminence trait.

Mandibular traits

Among the seven mandibular traits included were those most commonly cited as useful for sex or ancestry estimation using the mandible (Bass 2005; Berg and Ta'ala 2015; Rhine 1993; Warren et al. 2008). These traits were: pinching of the ascending ramus, undulation of the inferior border, chin profile (anterior projection of the mental eminence), shape of the inferior body, gonial eversion, shape of mandibular notch, and height of the coronoid process, and are depicted in Figures 3.8-3.14. Character states for these traits ranged from one to three in number following a modified scoring procedure developed, for previous research, by the author (see Appendix A, Table A.1).

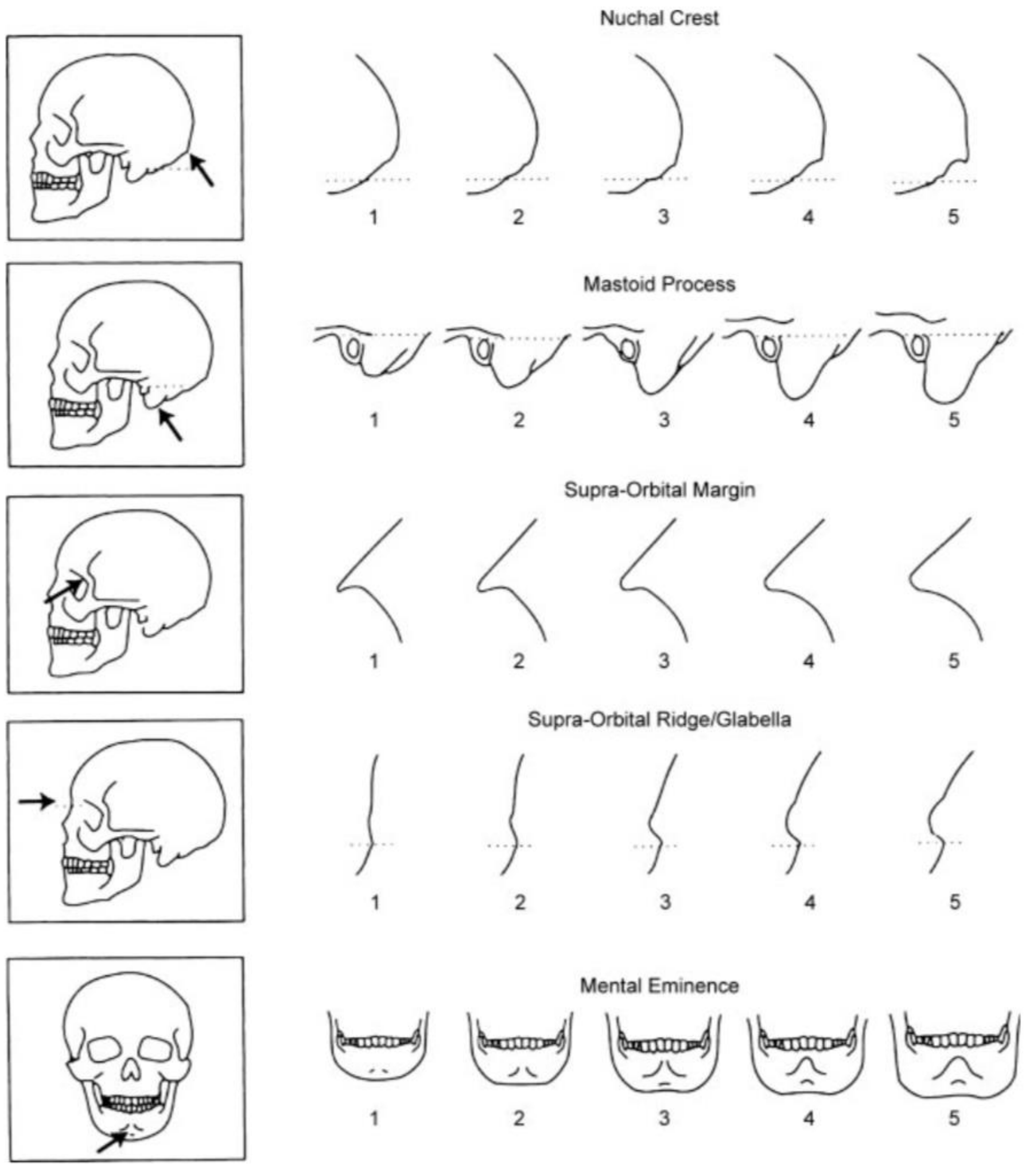

Figure 3.1: Walker (2008: p.41) cranial and mandibular traits for sex estimation.

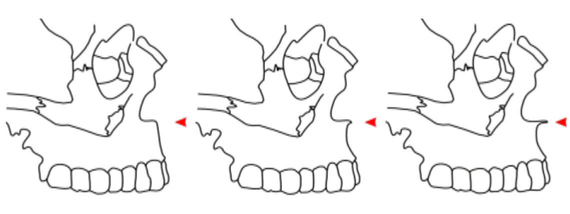

Figure 3.2: OSSA trait: Anterior Nasal Spine (image from Macromorphoscopic Databank).

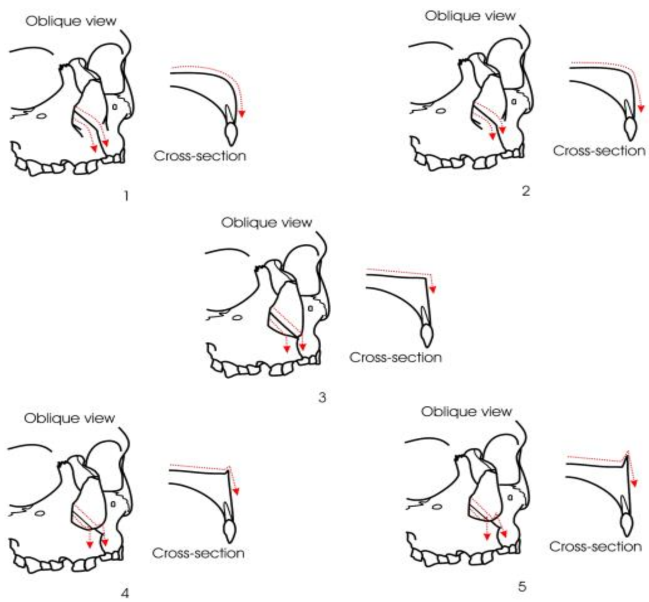

Figure 3.3: OSSA trait: Inferior Nasal Aperture (image from Macromorphoscopic Databank).

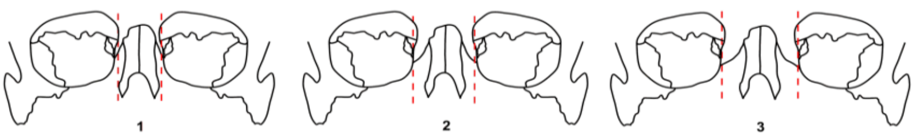

Figure 3.4: OSSA trait: Interorbital Breadth (image from Macromorphoscopic Databank).

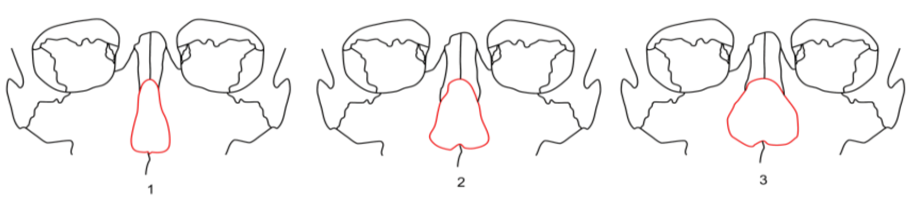

Figure 3.5: OSSA trait: Nasal Aperture Width (image from Macromorphoscopic Databank).

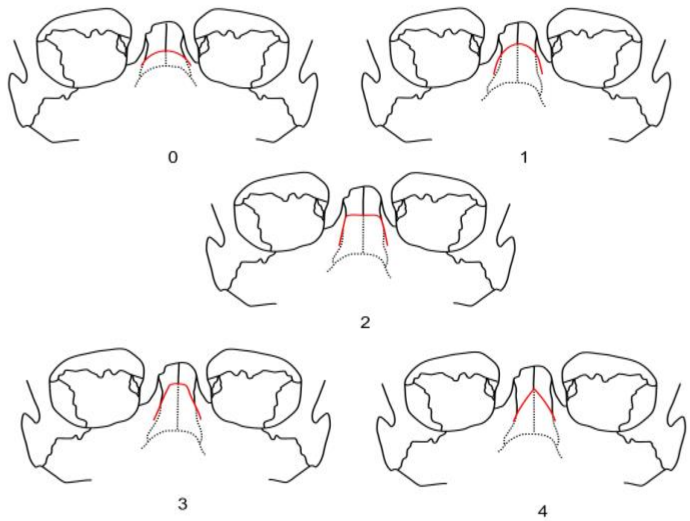

Figure 3.6: OSSA trait: Nasal Bone Contour (image from Macromorphoscopic Databank).

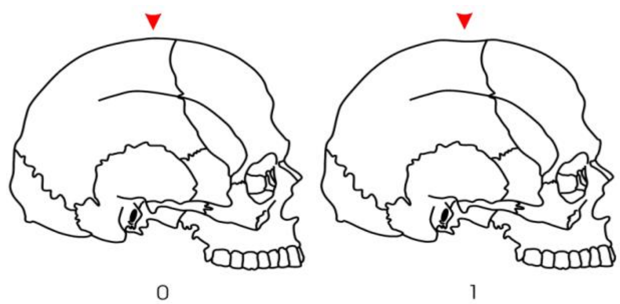

Figure 3.7: OSSA trait: Post-bregmatic Depression (image from Macromorphoscopic Databank).

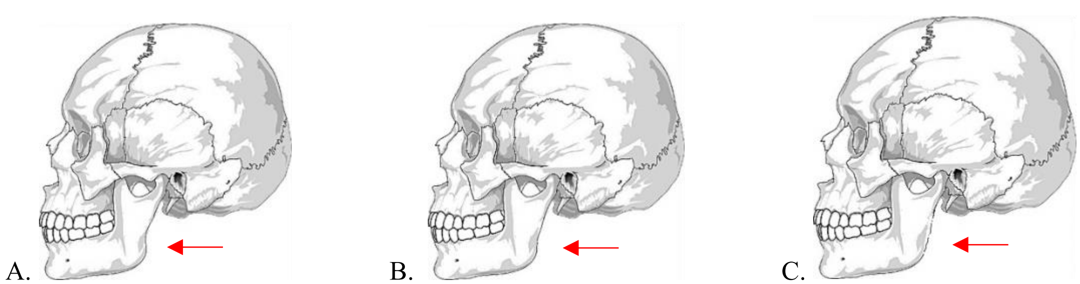

Figure 3.8: Pinching of the Ascending Ramus. A) pinched, B) slightly pinched, C) not pinched (image modified from Moller 2016) (Rhine 1990; 1993).

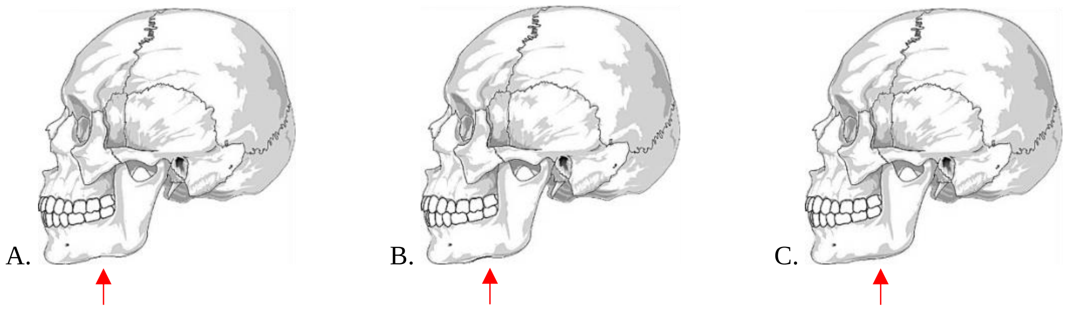

Figure 3.9: Undulation of the Inferior Border. A) undulated, B) slightly undulated, C) not undulated (image modified from Moller 2016) (Rhine 1990; 1993).

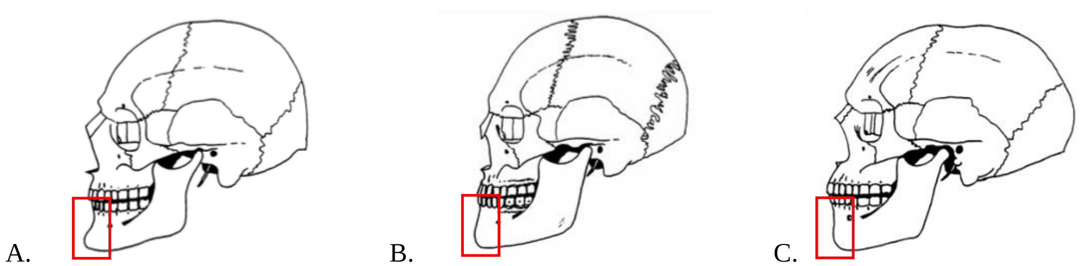

Figure 3.10: Chin Profile. A) protruding, B) vertical, C) receding (image from Rhine 1993: pp.57-62).

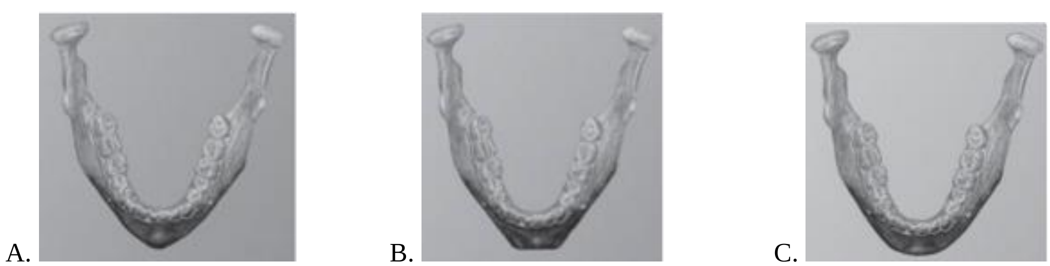

Figure 3.11: Shape of the Inferior Body. A) pointed, B) square (or bilobate), C) rounded (Berg and Ta'ala 2015: p.49).

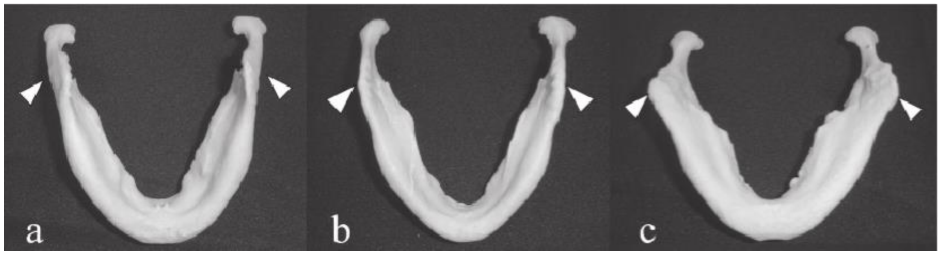

Figure 3.12: Gonial Eversion. A) inverted, B) vertical, C) everted (Ongkana and Sudwan 2010: p.125).

Figure 3.13: Shape of Mandibular (Sigmoid) Notch. A) sloping, B) wide, C) round (Shakya et al. 2013: p.1).

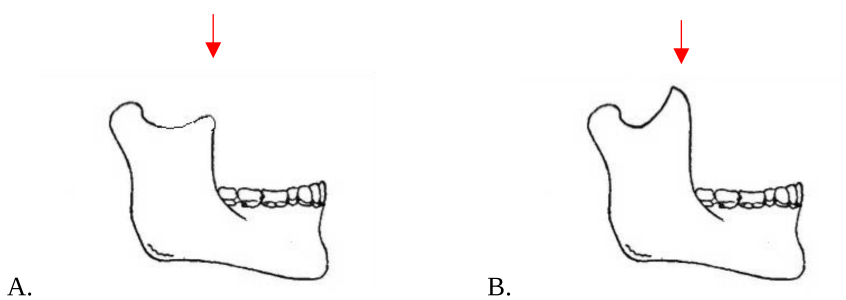

Figure 3.14: Height of the Coronoid Process. A) short, B) tall (image modified from Duckworth 2012: p.4).

Scanning and Scoring Procedures

Following common convention within the forensic anthropological literature, cranial and mandibular traits of all specimens were scored on the left side. The author was also blinded to all biological data (e.g. sex, age) so as not to introduce bias into the scoring process. All traits were scored as ordinal or discrete measures, with two to five character states associated with each trait (see Figures 3.1-3.14). All data were collected in an excel spreadsheet and subsequently analyzed in RStudio (version 1.2.5033).

Dry bones and 3D surface scans were each scored twice for a total of four sets of Walker (2008) scores, four sets of OSSA (Hefner and Ousley 2014) scores, and four sets of mandibular trait scores. The second assessment for each format (dry bone or scan) was separated from the first by at least one week to mitigate biases resulting from memory. Dual scoring sessions enabled multiple comparisons of intrarater reliability - between physical and 3D scoring sessions, as well as between repeat physical and 3D scoring sessions - offering greater insight into the consistency of morphological interpretation from dry to virtual bone and, thus, the reliability and utility of virtual models in place of physical elements.

All 3D surface scans were generated using a NextEngine 3D Ultra HD desktop laser scanner (NextEngine Inc., Malibu, CA). This 3D surface scanner uses both laser and white light to digitize objects of interest, as well as cameras to preserve the color and texture of the object's surface (NextEngine Inc., Malibu, CA). Multiple scans were generated for each specimen in different orientations so as to maximize coverage and data capture. All crania were positioned atop a cork ring which sat on the automatic turntable during scanning, and mandibles positioned on a raised platform with detachable metal placement arms during scanning. Three 360° scans were performed for each cranium: one with the cranium resting on its base in a Frankfurt horizontal position, one with the cranium in a norma lateralis position, and the third in a norma basilaris position. Three to four scans were generated for mandibles: one 360° scan with the mandible resting on its inferior border, one bracket (180°) scan of the posterior surface, and one single view scan of the inferior border. If needed, an additional single view scan of the left lateral ramus was generated to fill gaps or provide greater morphological detail. Scans were also generated with 6 to 10 divisions (object rotations of set degree-increments between scanning), with point densities of 2.9-10 thousand points per square inch and in macro or wide mode. Crania were generated with higher densities to capture their more complex anatomy and scanned in wide mode. Mandibles were generated with less dense point clouds and were scanned in macro mode, as crania are larger and necessitated a position farther from the scanner.

After scans were generated for a given specimen, they were trimmed, aligned, and refined using ScanStudio (NextEngine Inc., Malibu, CA), NextEngine's 3D model processing and editing software. The *trim* function was used first to remove background noise and unnecessary objects (e.g. turntable, cork ring). Once all scans for a given specimen were trimmed, images were aligned using approximately six alignment pins, whose placement indicated corresponding points between pairs of scans. ScanStudio (NextEngine Inc., Malibu, CA) automatically labels individual scans with single letters in alphabetical order - as crania required 3 scans, these were typically labeled A, B, and C, as can be seen in Figure 3.15, below. Initially, two of these scans would be aligned (e.g. A and B), followed by alignment of this pair with the third scan (e.g. A+B and C). After all scans for a specimen were aligned, the model would be *refined*, for removal of unnecessary data, and saved as a .PLY file for subsequent viewing and scoring.

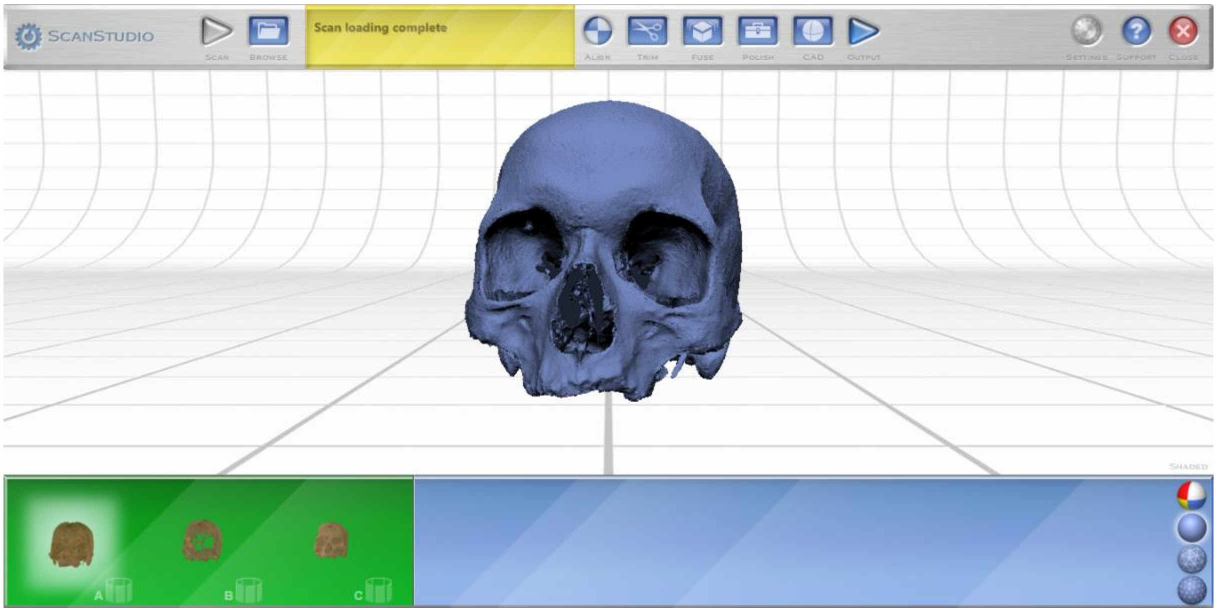

Figure 3.15: Individual scans are merged to create one model in ScanStudio (NextEngine Inc., Malibu, CA)

Scoring of completed models was performed in MeshLab (version 2020.07), an online open source processing and editing software program for use with digital 3D models. After scoring was completed for all specimens, three separate intrarater comparisons were performed using dry and virtual scores: 1) dry scores versus dry scores, 2) 3D scores versus 3D scores, and 3) dry scores versus 3D scores. Further, the interaction of score agreement and experience level was explored through the recruitment of additional participants.

Six additional participants with varying levels of experience were asked to participate by scoring a subset of specimens, in both dry and virtual formats, to glean a better understanding of the reliability of the proposed method and to reveal any interactions that might exist between experience level, scoring consistency, and agreement. A brief training session preceded scoring to provide conspicuous and standard definitions to participants. Definitions and character state illustrations were also provided for use as guides to all scorers during assessments. Participants represented two experience levels, the first *some experience,* a group comprised of graduate students and research specialists of biological anthropology at UTK who have graduate-level forensic anthropological training, and the second, *professional experience,* a group comprised of faculty and staff from the Anthropology Department at UTK. Each group contained three participants who were asked to score 21 crania and 24 mandibles in both dry and virtual formats. Ten of each specimen type scored (i.e. crania or mandibles) were then re-scored, enabling assessment of intrarater reliability at divergent experience levels.

Following the author, all participants scored traits on the left side of all specimens. During participants' first scoring session, 45 specimens (n crania = 21, n mandibles = 24) were scored in both dry and virtual formats, approximately 49 percent of the total sample - 50 percent of the cranial sample and 48 percent of the mandibular sample. Participants were supplied with the same excel sheet used by the author but condensed to include only the necessary specimens stipulated for participant use. The subsamples chosen for participants were randomly selected using a random number generator, however, it was verified that all demographics included in the initial sample were represented. During the second participant scoring session, 10 crania and 10 mandibles were re-scored by participants using the same excel sheet, again condensed to only the necessary specimens. Ten specimens were randomly chosen from each of the 21 crania and 24 mandibles identified from the first participant scoring session using the same random number generator. Once participants completed a scoring session, their completed excel sheets were emailed to the author for compilation and statistical analyses. The second scoring session for all participants was separated from the first by at least three days.

Statistics

Weighted Kappa was the primary statistic used for evaluation of both intrarater and interrater reliability for both the author and participants. The Kappa statistic approximates the strength of agreement between two groups of categorical data, either ordinal or nominal in nature (Fleiss and Cohen 1973; Kassambara 2019), and provides output in the form of a Kappa value. The Kappa value is an approximation of the agreement between corresponding sets of judgements, correcting for instances of chance agreement, and ranges in value from -1 to +1. Negative values indicate poorer than chance agreement, positive values indicate greater than chance agreement, a value of zero indicates agreement by chance alone, and a value of 1 indicates perfect agreement (Cohen 1960; Kassambara 2019; Walrath et al. 2004). Kappa values further correspond to one of five agreement levels, as outlined by Landis and Koch (1977), and were used to interpret the results (see Table 4.1).

In accordance with previous morphoscopic research (Colman et al. 2019; Lewis and Garvin 2016; Telmon et al. 2005; Villa et al. 2013; Walrath et al. 2004), Weighted Kappa was used for all analyses for the author and participants. Weighted Kappa is most appropriate for use with ordinal categorical data as it allows users to ascribe divergent weights to disagreements of varying magnitudes (Kassambara 2019). Linear weights place equally spaced weights on near and non-near disagreements as opposed to quadratic weights, which place less weight on near disagreements and increasing weights on more disparate disagreements (Agresti 2007: p.264; Kassambara 2019: p.21). Linear weights are recommended for use when differences between the first and second categories for scoring have the same importance as differences between second and third categories, and so forth (Kassambara 2019; Lowry 2021). Accordingly, linear weights were used for all statistical analyses herein. Furthermore, all statistical analyses were performed using Weighted Kappa as only two traits (under different conditions) were compared at any one time. To accommodate multiple raters in one analysis, data for individual scorers were organized identically and stacked in an excel sheet (version 16.35), which was subsequently imported and analyzed in RStudio. Cranial data and mandibular data were analyzed independently from one another but in the same manner. All statistical analyses were performed in RStudio (version 1.2.5033).

Following all Kappa tests, Wilcoxon Rank Sum tests were performed to assess statistically significant differences in agreement for all cranial and mandibular traits between the professional experience group and the graduate experience group. Wilcoxon Rank Sum tests were also performed in RStudio (version 1.2.5033).

CHAPTER 4: RESULTS

The results of all Weighted Kappa tests were interpreted following the standards for strength of agreement published by Landis and Koch (1977) and as outlined in Table 4.1. All results tables indicate Kappa values, significance of corresponding p-values, and strength of agreement (through shading).

Kappa results for intraobserver assessment of cranial and mandibular traits are presented in Tables 4.2 and 4.3, respectively. As can be visualized in these tables, agreement for cranial traits was greater between repeat dry and scan comparisons, was most consistent for Walker traits, and was lowest for inter-format comparisons of OSSA traits. Similarly, agreement for repeat dry and scan comparisons amongst mandibular traits was greatest, though modest learning can be seen from the first to the second inter-format comparisons.

Kappa results for all interobserver analyses are presented in Tables 4.4-4.9. Results for pooled experience levels are presented first (Tables 4.4-4.5), followed by results for the *professional* experience group (Tables 4.6-4.7), and finally by the results for the *graduate* experience group (Tables 4.8-4.9). Pooled cranial results demonstrate greatest agreement between repeat dry and scan comparisons and lowest agreement for inter-format comparisons of OSSA traits. Pooled mandibular results demonstrate high consistency at the moderate level across all traits and all four comparisons, suggesting good interpretability of virtual mandibular traits.

Results for the *professional* experience group indicate greatest cranial trait agreement between repeat dry and scan comparisons, though primarily for the OSSA traits. Inter-format agreement, however, is greater for the Walker traits. Mandibular trait agreement is similarly

Table 4.1: Kappa Strength of Agreement (Landis and Koch 1977).

Kappa Statistic	Strength of Agreement
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

greatest for repeat dry and scan comparisons, and agreement decreases from the first to the second inter-format comparison.

Results for the *graduate* experience group indicates greatest cranial trait agreement for repeat dry and scan comparisons, however learning can be seen for the Walker traits from the first to the second inter-format comparisons. Mandibular trait agreement is most consistent for the inter-format comparisons and increases from the first to the second inter-format comparison.

Subsequent to all Kappa tests, Wilcoxon Rank Sum tests were performed for the two inter-method analyses (i.e. Dry1: Scan1 and Dry2: Scan2). These tests compared the mean Kappa scores of the professional group with those of the graduate group for each cranial and mandibular trait. No statistically significant results were found for cranial or mandibular traits during either inter-format comparison, indicating that there were no significant differences in strength of agreement displayed by professionals and graduate students. Results for the Wilcoxon Rank Sum Tests for the first and second inter-method comparisons are displayed below in Tables 4.10 and 4.11, respectively.

Following the Wilcoxon Rank Sum Tests, the agreement found for Walker cranial traits and mandibular traits at greater and lesser experience levels were compared to the results of previous studies. Table 4.12 demonstrates the impact of experience on agreement found in the current study compared to the agreement found in Lewis and Garvin (2016) for the four Walker traits included in this investigation. Similarly, Table 4.13 demonstrates the impact of experience on four of the included mandibular traits as compared to Brynes et al.'s (2017) previous findings.

Finally, the intraobserver evaluations for repeat dry and scan assessments of Walker and OSSA traits, used for Table 4.2, were used to find correct classification accuracies for the total cranial sample used in this investigation. Correct classification accuracies found are listed in Table 4.14, below, and are comparable to correct classification accuracies reported by Walker (2008), Hefner and Ousley (2014), and Kenyhercz et al.'s (2017) OSSA validation study (see Chapter 5 for further discussion).

Table 4.2: Weighted Kappa and strength of agreement for intraobserver analyses of cranial traits.

Walker Traits		**Dry 1: Dry 2**	**Scan 1: Scan 2**	**Dry 1: Scan 1**	**Dry 2: Scan 2**
	Supra Orbital Margin	0.689**	0.648**	0.676**	0.656**
	Supra Orbital Ridge (Gla)	0.705**	0.786**	0.655**	0.766**
	Mastoid Process	0.706**	0.646**	0.570**	0.580**
	Nuchal Crest	0.712**	0.836**	0.707**	0.681**
OSSA Traits	Anterior Nasal Spine	0.71**	0.676**	0.549**	0.637**
	Inferior Nasal Aperture	0.841**	0.580**	0.304*	0.470**
	Interorbital Breadth	0.759**	0.752**	0.501**	0.607**
	Nasal Aperture Width	0.698**	0.653**	0.421**	0.524**
	Nasal Bone Contour	0.483**	0.566**	0.425**	0.481**
	Post-Bregmatic Depression	0.943**	0.687**	0.329	0.314

*Significant with a <0.01 p-value
**Significant with a <0.001 p-value

Fair Moderate Substantial Almost Perfect

Table 4.3: Weighted Kappa and strength of agreement for intraobserver analyses of mandibular traits.

Mandibular Traits	**Dry 1: Dry 2**	**Scan 1: Scan 2**	**Dry 1: Scan 1**	**Dry 2: Scan 2**
Shape of Inferior Border	0.621**	0.323	0.654**	0.323
Undulation of Inferior Body	0.810**	0.812**	0.632**	0.585**
Pinching of Ascending Ramus	0.500*	0.556*	0.389**	0.143
Height of Coronoid Process	0.710**	0.753**	0.481**	0.517*
Chin Profile	0.667**	0.538*	0.170	0.308
Gonial Eversion	0.589**	0.512*	0.582**	0.591**
Shape of Mandibular Notch	0.461*	0.505**	0.432**	0.564**

*Significant with a <0.01 p-value
**Significant with a <0.001 p-value

Slight | Fair | Moderate | Substantial | Almost Perfect

Table 4.4: Weighted Kappa and strength of agreement for combined participant interobserver analyses of cranial traits.

Walker Traits		Dry 1: Dry 2	Scan 1: Scan 2	Dry 1: Scan 1	Dry 2: Scan 2
	Supra Orbital Margin	0.522**	0.520**	0.492**	0.403**
	Supra Orbital Ridge (Gla)	0.773**	0.614**	0.626**	0.719**
	Mastoid Process	0.695**	0.458**	0.320**	0.408**
	Nuchal Crest	0.684**	0.473**	0.475**	0.510**
OSSA Traits	Anterior Nasal Spine	0.581**	0.622**	0.458**	0.642**
	Inferior Nasal Aperture	0.678**	0.594**	0.401**	0.561**
	Interorbital Breadth	0.549**	0.514**	0.425**	0.350**
	Nasal Aperture Width	0.766**	0.645**	0.569**	0.566**
	Nasal Bone Contour	0.397**	0.489**	0.321**	0.260*
	Post-Bregmatic Depression	0.522**	0.706**	0.292**	0.571**

*Significant with a <0.01 p-value
**Significant with a <0.001 p-value

Fair   Moderate   Substantial

Table 4.5: Weighted Kappa and strength of agreement for combined participant interobserver analyses of mandibular traits.

Mandibular Traits	**Dry 1: Dry 2**	**Scan 1: Scan 2**	**Dry 1: Scan 1**	**Dry 2: Scan 2**
Shape of Inferior Border	0.479**	0.392**	0.480**	0.458**
Undulation of Inferior Body	0.506**	0.585**	0.408**	0.616**
Pinching of Ascending Ramus	0.499**	0.547**	0.457**	0.467**
Height of Coronoid Process	0.680**	0.682**	0.681**	0.792**
Chin Profile	0.524**	0.184	0.336**	0.488**
Gonial Eversion	0.567**	0.301*	0.424**	0.386**
Shape of Mandibular Notch	0.555**	0.411**	0.502**	0.501**

*Significant with a <0.01 p-value
**Significant with a <0.001 p-value

Slight	Fair	Moderate	Substantial

Table 4.6: Weighted Kappa and strength of agreement for *professional* interobserver analyses of cranial traits.

Walker Traits		**Dry 1: Dry 2**	**Scan 1: Scan 2**	**Dry 1: Scan 1**	**Dry 2: Scan 2**
	Supra Orbital Margin	0.463**	0.523**	0.544**	0.333*
	Supra Orbital Ridge (Gla)	0.824**	0.658**	0.631**	0.623**
	Mastoid Process	0.724**	0.481**	0.277**	0.406**
	Nuchal Crest	0.750**	0.522**	0.464**	0.419**
OSSA Traits	Anterior Nasal Spine	0.400*	0.593**	0.427**	0.556**
	Inferior Nasal Aperture	0.720**	0.612**	0.513**	0.646**
	Interorbital Breadth	0.628**	0.680**	0.362**	0.309*
	Nasal Aperture Width	0.879**	0.708**	0.478**	0.480**
	Nasal Bone Contour	0.480**	0.669**	0.306**	0.377*
	Post-Bregmatic Depression	0.667**	0.760**	0.440**	0.667**

*Significant with a <0.01 p-value
**Significant with a <0.001 p-value

Fair | Moderate | Substantial | Almost Perfect

Table 4.7: Weighted Kappa and strength of agreement for *professional* interobserver analyses of mandibular traits.

Mandibular Traits	Dry 1: Dry 2	Scan 1: Scan 2	Dry 1: Scan 1	Dry 2: Scan 2
Shape of Inferior Border	0.309	0.669**	0.493**	0.308
Undulation of Inferior Body	0.698**	0.626**	0.405**	0.578**
Pinching of Ascending Ramus	0.604**	0.669**	0.533**	0.397*
Height of Coronoid Process	0.259	0.712**	0.673**	0.839**
Chin Profile	0.426*	0.155	0.426	0.153
Gonial Eversion	0.750**	0.302	0.475**	0.328
Shape of Mandibular Notch	0.647**	0.522**	0.583**	0.486**

*Significant with a <0.01 p-value
**Significant with a <0.001 p-value

Slight | Fair | Moderate | Substantial | Almost Perfect

Table 4.8: Weighted Kappa and strength of agreement for *graduate* interobserver analyses of cranial traits.

Walker Traits		Dry 1: Dry 2	Scan 1: Scan 2	Dry 1: Scan 1	Dry 2: Scan 2
	Supra Orbital Margin	0.551**	0.469**	0.409**	0.382*
	Supra Orbital Ridge (Gla)	0.720**	0.562**	0.622**	0.834**
	Mastoid Process	0.632**	0.410**	0.333**	0.383**
	Nuchal Crest	0.623**	0.425**	0.474**	0.585**
OSSA Traits	Anterior Nasal Spine	0.781**	0.650**	0.490**	0.720**
	Inferior Nasal Aperture	0.629**	0.575**	0.281**	0.459**
	Interorbital Breadth	0.464**	0.362*	0.487**	0.338*
	Nasal Aperture Width	0.639**	0.569**	0.688**	0.642**
	Nasal Bone Contour	0.318	0.311	0.328**	0.159
	Post-Bregmatic Depression	0.403	0.661**	0.170	0.485*

*Significant with a <0.01 p-value
**Significant with a <0.001 p-value

Slight	Fair	Moderate	Substantial	Almost Perfect

Table 4.9: Weighted Kappa and strength of agreement for *graduate* interobserver analyses of mandibular traits.

Mandibular Traits	**Dry 1: Dry 2**	**Scan 1: Scan 2**	**Dry 1: Scan 1**	**Dry 2: Scan 2**
Shape of Inferior Border	0.631**	0.136	0.466**	0.555**
Undulation of Inferior Body	0.314	0.541**	0.412**	0.644**
Pinching of Ascending Ramus	0.400*	0.435*	0.381**	0.535**
Height of Coronoid Process	0.634**	0.521*	0.568**	0.609**
Chin Profile	0.514*	0.058	0.426**	0.550**
Gonial Eversion	0.430*	0.302	0.377**	0.473*
Shape of Mandibular Notch	0.467**	0.297	0.415**	0.476**

*Significant with a <0.01 p-value
**Significant with a <0.001 p-value

Slight | Fair | Moderate | Substantial

Table 4.10: Results of Wilcoxon Rank Sum Tests for Dry 1: Scan 1.

	Traits	**P-values**
Walker Traits	Supra Orbital Margin	0.20
	Supra Orbital Ridge (Gla)	0.70
	Mastoid Process	0.70
	Nuchal Crest	1
OSSA Traits	Anterior Nasal Spine	1
	Inferior Nasal Aperture	0.40
	Interorbital Breadth	0.20
	Nasal Aperture Width	0.20
	Nasal Bone Contour	1
	Post-Bregmatic Depression	0.40
Mandibular Traits	Shape of Inferior Border	1
	Undulation of Inferior Body	0.40
	Pinching of Ascending Ramus	0.51
	Height of Coronoid Process	1
	Chin Profile	0.70
	Gonial Eversion	0.70
	Shape of Mandibular Notch	0.10

Table 4.11: Results of Wilcoxon Rank Sum Tests for Dry 2: Scan 2.

	Traits	**P-values**
Walker Traits	Supra Orbital Margin	1
	Supra Orbital Ridge (Gla)	0.10
	Mastoid Process	0.70
	Nuchal Crest	0.20
OSSA Traits	Anterior Nasal Spine	0.70
	Inferior Nasal Aperture	0.40
	Interorbital Breadth	0.51
	Nasal Aperture Width	0.40
	Nasal Bone Contour	0.20
	Post-Bregmatic Depression	0.16
Mandibular Traits	Shape of Inferior Border	0.50
	Undulation of Inferior Body	0.92
	Pinching of Ascending Ramus	0.63
	Height of Coronoid Process	0.38
	Chin Profile	0.08
	Gonial Eversion	0.40
	Shape of Mandibular Notch	0.70

Table 4.12: Impact of Experience on Walker (2008) Cranial Trait Agreement

Traits	Exp (current study)	Inexp (current study)	Exp (Lewis and Garvin 2016)	Inexp (Lewis and Garvin 2016)
Supra-orbital Margin	0.544[1]/0.333[2]	0.409[1]/0.382[2]	0.61	0.79
Supra-orbital Ridge	0.6316[1]/0.623[2]	0.622[1]/0.834[2]	0.78	0.66
Mastoid Process	0.277[1]/0.406[2]	0.333[1]/0.383[2]	0.68	0.60
Nuchal Crest	0.464[1]/0.419[2]	0.474[1]/0.585[2]	0.62	0.70

[1]First inter-format comparison
[2]Second inter-format comparison

Table 4.13: Consistency of Mandibular Morphoscopic Trait Scores

Traits	Pooled Exp (current study)	Exp (current study)	Inexp (current study)	Pooled Exp (Byrnes et al. 2017)*	Exp (Byrnes et al. 2017)*
Shape of Inferior Border	0.480[1]/0.458[2]	0.493[1]/0.308[2]	0.466[1]/0.555[2]	0.685	0.743
Undulation	0.408[1]/0.616[2]	0.405[1]/0.578[2]	0.412[1]/0.644[2]	0.407	0.452
Ascending Ramus	0.457[1]/0.467[2]	0.533[1]/0.397[2]	0.381[1]/0.535[2]	0.457	0.721
Gonial Eversion	0.424[1]/0.386[2]	0.475[1]/0.328[2]	0.377[1]/0.473[2]	0.675	0.684

[1]First inter-format comparison
[2]Second inter-format comparison
*Intra-class correlation coefficient scores

Table 4.14: Classification Accuracies for Walker (2008) and OSSA (Hefner and Ousley 2014) Methods

Walker* Method Dry 1	OSSA** Method Dry 1	Walker* Method Dry 2	OSSA** Method Dry 2	Walker* Method Scan 1	OSSA** Method Scan 1	Walker* Method Scan 2	OSSA** Method Scan 2
90.5%	69%	92.8%	83.3%	76%	80.9%	81%	88%

*Walker (2008)
**OSSA (Hefner and Ousley 2014)

CHAPTER 5: DISCUSSION

Morphoscopic traits are widely implemented by forensic anthropologists for their quick, effective, and versatile nature. However, as most morphoscopic methods implemented by practitioners today were developed for use on dry bone, it is unknown whether these methods can be reliably used to estimate biological parameters on virtual skeletal elements. As such, the primary goals of this research were to assess the interpretability of morphoscopic traits for sex and ancestry estimation from 3D surface scans, specifically traits of the Walker (2008) and OSSA (Hefner and Ousley 2014) methods, and to evaluate the potential of 3D surface scans to serve as proxies for physical bones in the event that the latter are inaccessible or are no longer within an anthropologist's possession.

At the outset of this research, it was hypothesized that greater experience would correspond with greater agreement, that cranial traits would display heightened agreement over mandibular traits, and that the two methods investigated would present similar levels of agreement. It was further predicted that a subset of the morphoscopic traits investigated would be interpretable from 3D surface scans, though not all, with the majority presenting moderate to substantial agreement between formats. Following these hypotheses, the results of this investigation suggest 1) similar patterns of agreement - both in agreement level and frequency of level - for cranial and mandibular traits in pooled interobserver analyses, 2) similar trends in agreement exhibited by graduate and professional-level participants, and 3) heightened agreement for Walker (2008) traits in several intra- and interobserver comparisons. All cranial traits investigated, and abbreviations or acronyms used in subsequent discussion, are listed in Table 5.1, below. Mandibular trait names were used in full.

Table 5.1: Cranial and Mandibular Trait Acronyms and Abbreviations

	Trait Names	**Abbreviations**
Walker Traits	Supra-Orbital Margin	SOM
	Supra-Orbital Ridge	SOR
	Mastoid Process	MP
	Nuchal Crest	NC
OSSA Traits	Anterior Nasal Spine	ANS
	Inferior Nasal Aperture	INA
	Interorbital Breadth	IOB
	Nasal Aperture Width	NAW
	Nasal Bone Contour	NBC
	Post-Bregmatic Depression	PBD

Intraobserver Results

Cranial Traits

Intraobserver analyses of cranial traits (see Table 4.2) revealed greatest agreement between repeat dry bone comparisons, followed closely by repeat scan comparisons. Agreement for the Walker traits was more consistent across the four comparisons, indicating greater inter-format agreement for the Walker traits, with agreement for the OSSA traits decreasing, in some cases notably, for inter-format comparisons. NBC exhibited the lowest agreement across the four comparisons remaining consistently at the moderate level. PBD and INA also displayed only fair agreement during one or both inter-format comparisons. Learning from the first to the second inter-format comparison is also demonstrated by almost all traits suggesting that increased familiarity with virtual skeletal elements leads to increased agreement.

Mandibular Traits

Intraobserver agreement for analyses of mandibular traits (see Table 4.3) similarly revealed greatest agreement for dry bone comparisons, followed by repeat scan comparisons. Undulation and height of the coronoid process presented the highest and most consistent agreement across the four comparisons. Shape of the inferior border, pinching of the ascending ramus, and chin profile demonstrate the least consistency in agreement, with two of the four comparisons demonstrating only slight or fair agreement. Most findings of low agreement are concentrated in the inter-format comparisons, suggesting these traits are difficult to evaluate in a virtual format.

Interobserver Results

Pooled Participants – Cranial Traits

Pooled interobserver analyses (see Table 4.4) for repeat dry bone and scan comparisons demonstrate similar occurrences of moderate and substantial agreement, though not necessarily for the same traits. OSSA traits present slightly fewer instances of moderate or substantial agreement than do Walker traits during inter-format comparisons, suggesting greater interpretability of these versus OSSA traits. SOR demonstrates the greatest precision with substantial agreement presented during all four comparisons. NBC demonstrates the lowest agreement throughout, presenting only fair agreement in three of the four comparisons. Modest learning is also demonstrated between inter-format sessions, as agreement increased for six of the ten traits from the first to second inter-format comparison.

Pooled Participants – Mandibular Traits

Pooled interobserver analyses of mandibular traits (see Table 4.5) demonstrates surprising consistency in agreement at the moderate level across the four comparisons. Height of the coronoid process, again, demonstrated greatest precision and consistency with substantial agreement across all four comparisons. Chin profile and gonial eversion display the least precision, presenting slight or fair agreement for two of the four comparisons. Modest learning can again be seen as four of the seven traits demonstrate increased agreement during the second inter-format comparison.

Professional versus Graduate Experience – Cranial Traits

Agreement of cranial traits at both the professional and graduate levels (see Tables 4.6 and 4.8) present greatest precision for repeat dry comparisons, followed by repeat scan comparisons. Inter-format comparisons among professionals present frequent agreement at the moderate level, whereas agreement level is less consistent amongst graduate students. Agreement for Walker traits during repeat dry and scan comparisons is similar between professional and graduate experience groups, however, agreement of OSSA traits for repeat dry and scan comparisons is greater and more consistent for the professional group. Among the professional experience group, MP, NAW, and NBC display the least consistency, presenting agreement at three divergent agreement levels. Additionally, IOB and NBC present only fair agreement during both inter-format comparisons amongst the professional group, suggesting low compatibility of these traits between dry and virtual formats.

Graduate students demonstrate lowest scoring consistency for MP, INA, and PBD, again spanning three different agreement levels. Additionally, agreement for NBC is only fair or slight across all four comparisons among graduate students suggesting low agreement, consistency, and, thus, reliability for this trait. MP also presents only fair agreement during the two inter-format comparisons, suggesting that this trait is challenging to score in a virtual setting. Modest learning is demonstrated between inter-format comparisons by both professional and graduate groups, however, is greater for OSSA traits among professionals and Walker traits among graduates.

Professional versus Graduate Experience – Mandibular Traits

Patterns of agreement for mandibular traits during repeat dry and scan sessions are similar for professionals and graduates (see Tables 4.7 and 4.9) though represent different levels of agreement - substantial for professionals and moderate for graduate students. Shape of the inferior border, chin profile, and gonial eversion demonstrate the least precision amongst professionals and pinching of the ascending ramus and gonial eversion demonstrate the least precision amongst graduates, all traits presenting only slight or fair agreement for two of the four comparisons. Low agreement for chin profile is not surprising as anatomical position is challenging to achieve for isolated, free-floating elements. Lower agreement for gonial eversion was unanticipated, however, given the frequent regard for this trait as informative within the sex estimation literature (Berg and Ta'ala 2015; Rhine 1993; Warren et al. 2008). Shape of the inferior border also displayed fair agreement during two comparisons for professionals - one of which was the repeat dry assessment - and during one comparison for graduates. Similar to gonial eversion, this finding was surprising as shape of the inferior border is commonly cited within the literature as useful for sex estimation using the mandible (Berg and Ta'ala 2015; Rhine 1993). Undulation and height of the coronoid process display the greatest agreement across the four comparisons for both experience levels. Graduate students display considerable learning from the first to the second inter-format comparison, as agreement increased for all seven traits. Surprisingly, agreement displayed by professionals decreased for five traits from the first to the second inter-format comparison, contrary to the agreement pattern displayed by professionals for cranial traits.

Discussion of Traits

A few Walker (2008) traits, namely MP and SOM, warrant further discussion. MP is typically scored with the cranium in profile and relative to surrounding hard tissue anatomy (Buikstra and Ubelaker 1994; Walker 2008). As virtual elements are not quite life size - though the virtual 3D scan can be zoomed in or out - this may have made virtual evaluation of this feature more challenging, a notion consistent with findings of lowest precision during inter-format comparisons. Increased familiarity with virtual bones, and using the zoom feature when assessing MP, may help to increase scoring agreement for this trait. Additionally, SOM displayed some disagreement, though less than that of the MP, and less than was expected.

Lower agreement for this trait was anticipated given that this region is often palpated during evaluation and could not be carried out in this way on virtual elements. As all instances of fair agreement for this trait were concentrated during repeat scan or inter-format comparisons, increased familiarity with virtual skeletal elements will likely yield increased agreement.

Further, notable differences in Walker (2008) trait agreement between this study and Lewis and Garvin's (2016) can be seen in Table 4.12, particularly for SOM and MP. Lewis and Garvin (2016), however, assessed all elements in dry format only. As such, lower agreement displayed in the current study is, understandable given that reported precision from this study in Table 4.12 is for dry bone and virtual element comparisons.

The OSSA trait NBC demonstrated the least precision across all four comparisons for all intra- and interobserver analyses. This finding suggests that replicability is low for this trait and that methods for its interpretation may require further attention. As contour gauges - the recommended instrument for evaluation of this trait (Hefner 2009; Hefner 2003) - cannot be used to assess virtual specimens, lower agreement for inter-format comparisons is not surprising. Instances of only fair agreement for repeat dry comparisons, however, suggest use of this tool may not be easy or intuitive and may result in less consistent interpretations, producing the opposite of the desired effect. Such findings are also reasonable given that sources recommending use of this instrument for evaluation of NBC do not provide clear instructions or recommendations on how to interpret the output offered by the gauge - in other words, should the positive or negative of the feature be interpreted. Further clarification, pictorial guides featuring gauge output (rather than features or specimens), or disuse of this instrument altogether, may result in heightened scoring consistency of this trait.

The OSSA (Hefner and Ousley 2014) traits INA, IOB, and PBD may also require further consideration prior to implementation in virtual forensic analyses. Complete morphology of the INA was difficult to capture, and on several occasions the floor of the nasal aperture was incomplete in finished models due to obscuration of this region by the vertical maxillae and due to the stationary procedures required the NextEngine scanner and of object being scanned (NextEngine Inc., Malibu, CA). An additional scan of this region, or alternative cranium orientations during scanning, may resolve this issue. Agreement for this trait, however, was similar to or greater than previous reports by Hefner (2009) and Klales and Kenyhercz (2015) (see section Further Findings), suggesting that this trait can be reliably assessed from 3D surface scans. IOB also demonstrated several instances of only fair agreement in interobserver analyses. Difficulties or inconsistencies in evaluation of this trait may pertain to challenges in visualizing the fine frontomaxillary and nasomaxillary sutures of the nasal region, as these may serve as useful indicators of breadth of the interorbital region on dry bone. Increasing scan quality may allow observers to better visualize these features, though this may come at the cost of a greater scan time.

PBD also surprisingly demonstrated several instances of slight or fair agreement. This finding is consistent with interobserver agreement between dry bone comparisons reported by Hefner (2009) however, contrasts with results reported by Klales and Kenyhercz (2015). As capture of fine detail or morphology is not required for evaluation of this trait, these findings were unexpected. Similar to MP, however, this trait is typically evaluated in profile and smaller-than-life-size virtual elements may have impacted evaluation, as lower agreement was concentrated during inter-format comparisons. A more hands-on approach may also be used when assessing this trait on dry bone, similar to SOM, and inability to carry-out evaluations in this way may have resulted in lower agreement.

In regard to the mandibular traits, chin profile, followed by gonial eversion, and shape of inferior border, displayed lower score agreement and consistency throughout most observer analyses than did other traits. Agreement for chin profile is particularly low for repeat scan and inter-format comparisons, though does not achieve agreement above the moderate level during dry comparisons, suggesting that virtual format alone does not make this trait challenging to score. Gonial eversion also indicated lower score agreement amongst participants during repeat scan and inter-format comparisons. Repeat dry comparisons of gonial eversion present moderate or substantial agreement across observers, suggesting that this trait is difficult to score in a virtual format. Shape of the inferior border also warrants mentioning, as this trait exhibited only slight or fair agreement during at least one comparison across all observers. Lower agreement for this trait was unanticipated, given its frequent discussion within the sex estimation literature on the mandible (Berg and Ta'ala 2015; Rhine 1993). Increased familiarity with virtual elements, however, will likely increase agreement, as was seen for several traits from the first to the second inter-format comparisons for mandibular traits.

Contrastingly, investigations of the replicability of mandibular morphoscopic traits are scant. Byrnes et al. (2017) is seemingly the only published study that evaluates interobserver consistency for mandibular morphoscopic traits, albeit through divergent statistical methods and a divergent set of traits. Byrnes et al. (2017) used intra-class correlation (ICC) to measure the consistency of scores amongst observers at divergent experience levels. According to Fleiss and Cohen (1973), under certain conditions, weighted agreement (i.e. Weighted Kappa) for ordinal categorical variables can be interpreted similarly through ICC. Thus, the results from the two studies are compared below.

Consistency for traits that were addressed in both studies are listed in Table 4.13. Scores for experienced participants in the present study frequently fall below those for experienced observers in Byrnes et al. (2017); however, scores for pooled-experience observers for both studies are similar, with consistency for two of the four traits in the present study, shape of the inferior border and gonial eversion, falling below scores presented in Byrnes et al. (2017). Byrnes et al. (2017) did not report results for inexperienced observers, however, results from the current study are also presented in Table 4.13 for comparative purposes.

Wilcoxon Rank Sum Tests

Wilcoxon Rank Sum tests were also performed for all cranial and mandibular traits during both inter-format comparisons (i.e. Dry1:Scan1 and Dry2:Scan2) to assess correlations between experience and agreement. Wilcoxon Rank Sun tests evaluated differences in agreement (mean Kappa values) displayed by professionals and graduates in inter-format comparison 1 and inter-format comparison 2. No statistically significant differences in strength of agreement between graduates and professionals were found for cranial or mandibular traits during either inter-format comparison. In other words, the professional experience group did not exhibit agreement that was statistically significantly above or below that of graduate students. While these tests did not assess the consistency of ordinal scores attributed at the two levels, they did assess consistency of agreement at the two levels and indicate that differences in agreement are minimal. These findings are in accordance with previous research regarding the impact of experience on the assessment of morphoscopic traits, which suggests that, while greater experience may see greater precision (consistency), differences are negligible and those with lesser experience can mitigate its implications by becoming intimately familiar with the methods, traits, and trait definitions implemented in morphoscopic analyses (Byrnes et al. 2017; Lewis and Garvin 2016; Walker 2008).

Further Findings

Intraobserver evaluations for repeat dry and virtual assessments were also used to estimate sex and ancestry of the cranial sample (n = 42) following the procedures set forth by Walker (2008) and Hefner and Ousley (2014). Results of these estimations can be found in Table 4.14. Only two of Walker's logistic discriminant functions for American/English populations could be employed, as the remaining four functions required scores for the mental eminence (2008: p.47). These classification accuracies demonstrate good accuracy and at rates similar to those achieved in studies using the same methods on dry bone alone (Walker 2008; Kenyhercz et al. 2017). Walker reports accuracy rates of 85.4% and 82.9% for males and females, respectively, using the glabella and mastoid discriminant function, and rates of 76.8% and 82.9% for males and females, respectively, using the nuchal and mastoid discriminant function (2008: p.47). Accuracy rates for the current study are for pooled sexes and range from 90-92.8% for dry bone and 76-88% for virtual bones.

OSSA classification rates for the current study were similarly pooled for sexes and ancestries, in the current study, and ranged from 69-83.3% for dry bone evaluations and 80.9-88% for virtual evaluations. These classification accuracies are comparable to those reported by Hefner and Ousley (2014) at 86.1%, and Kenyhercz et al. (2017) at 50.9% for their pooled U.S. Black sample and 89.2% for their pooled U.S. White sample (using a sectioning point of 3 for both).

As there is currently no traditional suite of traits or established method for assessing sex and ancestry from the mandible, sex and ancestry were not estimated for this portion of the sample.

Study Limitations

One limitation of this study is the small sample size utilized - the total sample was comprised of 42 crania and 50 mandibles, representative of both males and females and wide age ranges, though only Black and White Americans. It is worth mentioning, however, that most published investigations that exceed this sample size are retrospective studies implementing CT scans in various method validation endeavors. Investigations more closely resembling the present one - comparing virtual and dry bone specimens - are closer in sample size.

Additionally, participants representing each of the two experience levels were at their minimum capacities for significance testing, with three individuals in each group. As such, the limited number of participants should be considered when applying this study to future investigations.

A further limitation is the exclusion of the Walker (2008) trait, mental eminence. This trait was excluded for reasons mentioned previously (see Chapter 3: Materials and Methods), however, it is typically used in forensic casework when the Walker (2008) method is implemented. Several studies report low agreement for this trait and difficulties in assessing the morphology of this region, as the recommended character state guide (Walker 2008) does not represent the true range of variation and expression of this trait, providing clear support for the exclusion of this trait in the present investigation (Lewis and Garvin 2016). In a test between two observers, one with graduate level experience and one professional experience, and using similar Weighted Kappa statistical procedures to those implemented herein, Lewis and Garvin (2016) report intraobserver agreement for mental eminence at 0.30-0.73, and interobserver agreement at 0.03-0.08. Further, Stull et al. (2013) report less sexual dimorphism exhibited in the region of the mental eminence than in other cranial or pelvic regions, results also echoed by Lewis and Garvin (2016). Moreover, the latter investigation cautions researchers against the use of this trait and recommends that in the future, researchers consider dividing the extensive morphology of this region into multiple traits, as current schematics cannot and do not adequately capture the range of variation or patterns of sexual dimorphism observed in this region (Lewis and Garvin 2016).

Additionally, technology limitations were identified during this research. As the NextEngine 3D laser scanner (NextEngine Inc., Malibu, CA) is a stationary desktop surface scanner, operating by the emission of lasers, oddly shaped spaces or deep crevices were difficult to capture, as mentioned previously in regard to INA. Aside from rotating a segmented number of degrees between scans via automated turntable, both the scanner and the object being scanned must remain immobile during scanning. Jockeying of a hand-held scanner to obtain optimum coverage is a draw-back of this particular scanner. Shape of the element itself can also pose challenges to the scanning process. In several instances, there were holes on the posterior mandibular body (lingual surface) as the lasers could not reach this surface. Given this bone's horseshoe shape, which tends to be deeper and narrower, rather than shallow and wide, complete capture of this surface was difficult, and can be further hindered by the position of the ascending rami. To the best of the author's knowledge, the best way to mitigate this is to increase the number of divisions (automated, degree-based rotations between scans), ensuring fuller coverage of the object. Scans in multiple orientations are also helpful however, overlap of regions is essential (as in photogrammetry) to ensure that multiple scans can be aligned and fused properly.

Research Significance and Future Directions

This study appears to be the first VA investigation of Walker (2008) traits for sex estimation and OSSA (Hefner and Ousley 2014) traits for ancestry estimation, as well as for morphoscopic traits of the mandible. As has been recognized by previous authors, surface scanning for use within forensic anthropology remains in its infancy (Garvin and Stock 2016). Three-dimensional surface scans can be used for the collection of precise metric data from human skeletal elements (Algee-Hewitt and Wheat 2016; Friend and Stock 2016; Sholts et al. 2011; Sholts and Wärmländer 2012), and, thus, to increase sample sizes in research; to preserve evidence in the event that a forensic case goes to trial; and to preserve remains that are to be repatriated, inhumed, or cremated. However, it has not yet been ascertained whether surface scans can be used for the collection of precise morphoscopic data. As such, this study sought to investigate this question and demonstrates that, while few traits are not well suited-for virtual evaluation, several traits do present moderate, substantial, or almost perfect agreement between dry and virtual formats, suggesting that precise morphoscopic data can in fact be obtained from 3D surface scans. Those traits that proved challenging to assess virtually should be modified to better suit virtual evaluation, or methods including such traits should be modified to better assess biological parameters within a virtual format. Until further virtual method validation, virtual evaluation of the biological profile should be performed with much caution and supplemented with metric or physical analyses, as well.

Editing tools within MeshLab should also be investigated in future research, as there may exist tools more conducive to trait evaluation outside of observation alone - for example cross-sectional tools may enlighten evaluation of the SOM or NBC. MeshLab alternatives should also be considered as research has found that software can impact score consistency (Colman et al. 2019) and should be designated as part of the standardized methods for virtual bone evaluation.

Furthermore, standards for the construction of 3D surface scans (Garvin and Stock 2016) and acceptable agreement thresholds currently do not exist for forensic anthropology. As mitigating biases and subjectivity is crucial to the development and application of morphoscopic methods, devising standards not only for evaluating traits from 3D surface scans, but for generating scans themselves, should also be considered. Moreover, as VA is concerned with the agreement (precision) between multiple formats, and several previous studies aimed at answering similar questions of precision have used the Weighted Kappa statistic, it should be made clear what an acceptable threshold for agreement is, as there does not appear to be a defined strength-of-agreement threshold within forensic anthropology. Several studies make mention of agreement greater than 0.60 as an acceptable range within forensic anthropology (Bertoglio et al. 2020; Lewis and Garvin 2016; Klales and Kenyhercz 2015), corresponding to substantial (0.61-0.80) or almost perfect (0.81-1.0) agreement (Landis and Koch 1977), however, as this threshold is typically much higher within various medical fields (McHugh 2012), and agreement may differ in meaning and import across fields, this standard should be explicitly defined for forensic anthropology. A lower threshold may in fact be acceptable as forensic anthropology is not concerned with living human subjects, however, the implications of such a threshold (0.61 and above) on correct classification accuracies and positive identification rates should be explored in future research.

Finally, validation studies are also needed to address previously investigated methods, traits, and skeletal elements to progress the emergent virtual forensic anthropology, for which there has been little validation thus far.

CHAPTER 6: CONCLUSIONS

This study sought to investigate the interpretability of morphoscopic data from 3D surface scans by examining the applicability of morphoscopic methods designed for use on dry bone to virtual skeletal models and the impact of experience on agreement of traits between dry and virtual formats. This study revealed that there is often moderate to substantial inter-format agreement for cranial and mandibular traits, though more often for Walker (2008) cranial traits than OSSA (Hefner and Ousley 2014) cranial traits, and that differences in agreement based on experience (i.e. for those with graduate experience versus those with professional experience) were minimal and not statistically significantly different.

This study supports continued investigation of VA methods, however, also reveals limitations regarding the interpretability of some traits and, thus, the applicability of the methods investigated, to virtual bone. Low agreement exhibited by various cranial and mandibular traits likely results from a lack of familiarity with 3D surface scans or virtual models in general, and will likely increase with greater exposure for individuals at all experience levels. Discrepancies may also result from a lack of familiarity with the extensive morphological variation exhibited by individuals both within and between age, sex, and ancestral populations, as well as from a lack of familiarity with the traits in general, particularly the mandibular traits. Increased exposure to variation and mandibular traits will, again, likely yield increased agreement.

This is believed to be the first VA investigation of Walker (2008) and OSSA (Hefner and Ousley 2014) cranial traits, and mandibular morphoscopic traits in general, and highlights the potential for virtual construction of the biological profile. Future investigations should consider the traits and methods evaluated herein when devising new or amending established methods, as this investigation considered two of the most widely used morphoscopic methods for sex and ancestry estimation from the cranium.

The utility of 3D surface scans to serve as proxies to dry bone may enable forensic anthropologists and bioarchaeologists to increase sample sizes in research and may facilitate the construction of widely accessible digital skeletal collections, resources that would provide research opportunities for those with limited access to physical skeletal collections. They can also serve as contemporary comparative collections for those without access to population-specific comparative collections for religious or other reasons (Colman et al 2019; Johnstone-Belford et al. 2018; Hisham et al. 2019). Digital models can further serve as permanent documentation for remains that are too fragile to be handled or are to be inhumed, cremated, or repatriated (Algee-Hewitt and What 2016; Decker et al. 2011; Freiss 2012; Villa et al. 2013).

Given the ever-expanding reaches of VA, standards do not exist for the construction of virtual models or biological parameters, rendering these methods inadmissible under the Daubert guidelines for scientific testimony (Daubert v. Merrell Dow Pharmaceuticals). As such, continued research is necessary to understand error rates, achieve wide acceptance, and ultimately understand the utility of these models.

REFERENCES

Abdel Fatah EE, Shirley NR, Jantz RL, Mahfouz MR. 2014. Improving Sex Estimation from Crania Using a Novel Three-dimensional Quantitative Method. Journal of Forensic Sciences 59:590–600.

Acsádi G, Nemeskéri J. History of human life span and mortality. Budapest: Akademiai Kiado, 1970.

Agresti A. 2007. Cram 101 textbook outlines to accompany An introduction to categorical data analysis, 1st edition. Ventura, CA: Academic Internet Publishers.

Algee-Hewitt BFB, Wheat AD. 2016. The reality of virtual anthropology: Comparing digitizer and laser scan data collection methods for the quantitative assessment of the cranium. American Journal of Physical Anthropology 160:148–155.

Ali Z, Cox C, Stock MK, Zandee vanRilland EE, Rubio A, Fowler DR. 2018. Estimating Sex Using Metric Analysis of the Scapula by Postmortem Computed Tomography. Journal of Forensic Sciences 63:1346–1349.

Daubert v Merrell Dow Pharmaceuticals. 1993. US Supreme Court 509. U.S.579,113S.Ct.2786, 125L, Ed.2d 469.

Bass WM. 2005. Human osteology: a laboratory and field manual. 5th ed. Columbia: Missouri Archaeological Society.

Berg GE, Ta'ala SC. 2015. Biological affinity in forensic identification of human skeletal remains beyond black and white. Boca Raton, FL: CRC Press.

Bertoglio B, Corradin S, Cappella A, Mazzarelli D, Biehler-Gomez L, Messina C, Pozzi G, Sconfienza LM, Sardanelli F, Sforza C, De Angelis D, Cattaneo C. 2020. Pitfalls of Computed Tomography 3D Reconstruction Models in Cranial Nonmetric Analysis*. Journal of Forensic Sciences 65:2098–2107.

Bethard JD, DiGangi EA. 2020. Letter to the Editor—Moving Beyond a Lost Cause: Forensic Anthropology and Ancestry Estimates in the United States. Journal of Forensic Sciences 65:1791–1792.

Brooks S, Suchey JM. 1990. Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. Human Evolution 5:227–238.

Buckberry JL, Chamberlain AT. 2002. Age estimation from the auricular surface of the ilium: A revised method. American Journal of Physical Anthropology 119:231–239.

Buikstra JE, Ubelaker DH. 1994. Standards for data collection from human skeletal remains: proceedings of a Seminar at The Field Museum of Natural History. Fayetteville: Arkansas Archeological Survey Publications.

Byrnes JF, Kenyhercz MW, Berg GE. 2017. Examining Interobserver Reliability of Metric and Morphoscopic Characteristics of the Mandible. Journal of Forensic Sciences 62:981–985.

Cohen J. 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20:37–46.

Colman KL, de Boer HH, Dobbe JGG, Liberton NPTJ, Stull KE, van Eijnatten M, Streekstra GJ, Oostra R-J, van Rijn RR, van der Merwe AE. 2019. Virtual forensic anthropology: The accuracy of osteometric analysis of 3D bone models derived from clinical computed tomography (CT) scans. Forensic Science International 304:109963.

Cooke SB, Terhune CE. 2015. Form, Function, and Geometric Morphometrics. The Anatomical Record 298:5–28.

Dabbs GR, Moore-Jansen PH. 2010. A Method for Estimating Sex Using Metric Analysis of the Scapula. Journal of Forensic Sciences 55:149–152.

Decker SJ, Davy-Jow SL, Ford JM, Hilbelink DR. 2011. Virtual Determination of Sex: Metric and Nonmetric Traits of the Adult Pelvis from 3D Computed Tomography Models*,†. Journal of Forensic Sciences 56:1107–1114.

Dedouit F, Savall P, Mokrane FZ, Rousseau H, Crubezy E, Rouge D, Telmon N. 2013. Forensic Radiology Special Feature: Review Article Virtual Anthropology and Forensic Identification Using Multidetector CT. The British Institute of Radiology:1–12.

Dedouit F, Telmon N, Guilbeau-Frugier C, Gainza D, Otal P, Joffre F, Rougé D. 2007. Virtual Autopsy and Forensic Identification?Practical Application: A Report of One Case. Journal of Forensic Sciences 52:960–964.

DiGangi EA, Bethard JD. 2021. Uncloaking a Lost Cause: Decolonizing ancestry estimation in the United States. American Journal of Physical Anthropology.

Duckworth HWL. 2012. Prehistoric man. The Cambridge Manuals of Science and Literature. New York: Cambridge University Press.

Errickson D, Thompson TJU, Rankin BWJ. 2014. The application of 3D visualization of osteological trauma for the courtroom: A critical review. Journal of Forensic Radiology and Imaging 2:132–137.

Fleiss JL, Cohen J. 1973. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. Educational and Psychological Measurement 33:613–619.

Freiss M. 2012. Scratching the Surface? The Use of Surface Scanning I Physical and Paleoanthropology. Journal of Anthropological Sciences 90:7–31.

Garvin HM, Stock MK. 2016. The utility of advanced imaging in forensic anthropology. Academic Forensic Pathology, 6:499-516. doi:10.23907/2016.050

Gill GW, Rhine S. 1990. Nonmetric Skull Racing. In: Skeletal attribution of race. Albuquerque, NM: Maxwell Museum of Anthropology. p 9–20.

Grabherr S, Cooper C, Ulrich-Bochsler S, Uldin T, Ross S, Oesterhelweg L, Bolliger S, Christe A, Schnyder P, Mangin P, Thali MJ. 2008. Estimation of sex and age of “virtual skeletons”–a feasibility study. European Radiology 19:419–429.

Hefner JT, Linde KC. 2018. Atlas of Human Cranial Macromorphoscopic Traits. San Diego: Elsevier Science & Technology.

Hefner JT, Ousley SD. 2014. Statistical Classification Methods for Estimating Ancestry Using Morphoscopic Traits. Journal of Forensic Sciences 59:883–890.

Hefner JT. 2003. Assessing Nonmetric Cranial Traits Currently Used in Forensic Anthropology.

Hefner JT. 2009. Cranial Nonmetric Variation and Estimating Ancestry. Journal of Forensic Sciences 54:985–995.

Hefner JT. Naive Bayesian Classifier. Macromorphoscopic Data [Internet]. Available from: https://macromorphoscopic.com/databank/

Hisham S, Abdullah N, Mohamad Noor MH, Franklin D. 2019. Quantification of Pubic Symphysis Metamorphosis Based on the Analysis of Clinical MDCT Scans in a Contemporary Malaysian Population. Journal of Forensic Sciences 64:1803–1811.

Johnstone-Belford E, Flavel A, Franklin D. 2018. Morphoscopic observations in clinical pelvic MDCT scans: Assessing the accuracy of the Phenice traits for sex estimation in a Western Australian population. Journal of Forensic Radiology and Imaging 12:5–10.

Kassambara A. 2019. Inter-Rater Reliability Essentials: A Practical Guide in R. Datanovia.

Kenyhercz MW, Klales AR, Rainwater CW, Fredette SM. 2017. The Optimized Summed Scored Attributes Method for the Classification of U.S. Blacks and Whites: A Validation Study. Journal of Forensic Sciences 62:174–180.

Kimmerle EH, Ross A, Slice D. 2008. Sexual Dimorphism in America: Geometric Morphometric Analysis of the Craniofacial Region. Journal of Forensic Sciences 53:54–57.

Klales AR, Kenyhercz MW. 2015. Morphological Assessment of Ancestry using Cranial Macromorphoscopics. Journal of Forensic Sciences 60:13–20.

Klales AR, Ousley SD, Vollner JM. 2012. A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. American Journal of Physical Anthropology 149:104–114.

Klales AR. 2020. MorphoPASSE: Morphological Pelvis and Skull Sex Estimation Program. In: Sex estimation of the human skeleton: history, methods, and emerging techniques. London: Academic Press, an Imprint of Elsevier. p 271–278.

Klales AR. Current practices in forensic anthropology for sex estimationin unidentified, adult individuals. Proceedings of the 65th Annual Meet-ing of the American Academy of Forensic Sciences; 2013 Feb 18-23; Washington, DC. Colorado Springs, CO: American Academy of ForensicSciences, 2013;19(H81):439–40

Landis JR, Koch GG. 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics 33:159–174.

Langley NR, Dudzik B. 2016a. Sex Estimation. Oxford Bibliographies.

Langley NR, Jantz LM, Ousley SD, Jantz RL, Milner G. 2016b. Data Collection Procedures for Forensic Skeletal Material 2.0. Knoxville, TN: University of Tennessee.

Langley NR, Meadows Jantz L, McNulty S, Maijanen H, Ousley SD, Jantz RL. 2018. Error quantification of osteometric data in forensic anthropology. Forensic Science International 287:183–189.

Lewis CJ, Garvin HM. 2016. Reliability of the Walker Cranial Nonmetric Method and Implications for Sex Estimation. Journal of Forensic Sciences 61:743–751.

Lowry R. 2021. Sidebar: Kappa [Internet]. Available from: http://www.vassarstats.net/kappaexp.html

Moller, N. 2016. TMI or TMJ problems? How to identify & Relive Jaw Pain. Available from http://www.livingwellfamilychiropractic.com/blog/2016/3/9/tmi-or-tmj-problems-how-to-identify-and-relieve-jaw-pain.

Nuzzolese E, Quinney PR, Quinney JR, Vella GD. 2019. Geometric Morphometric Analysis in Sexual Dimorphism of the Mandible from Panoramic X-Ray Images. Journal of Odonto-Stomatology 37:35–44.

Ongkana N, Sudwan P. 2010. Morphologic Indicators of Sex in Thai Mandibles. Chaing Mai Med J 49:123–128.

Ousley RD, Jantz RL. 1998. The forensic data bank: Documenting skeletal trends in the United States. In: Forensic osteology: advances in the identification of human remains. Springfield, Ill, U.S.A.: Charles C. Thomas. p 441–458.

Petaros A, Garvin HM, Sholts SB, Schlager S, Wärmländer SKTS. 2017. Sexual dimorphism and regional variation in human frontal bone inclination measured via digital 3D models. Legal Medicine 29:53–61.

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. American Journal of Physical Anthropology 30:297–301.

Pink CM, Maier C, Pilloud MA, Hefner JT. 2016. Cranial Nonmetric and Morphoscopic Data Sets. In: Biological distance analysis - forensic and bioarchaeological perspectives. Elsevier Science Publishing Co. p 91–107.

Ramsthaler F, Kettner M, Gehl A, Verhoff MA. 2010. Digital forensic osteology: Morphological sexing of skeletal remains using volume-rendered cranial CT scans. Forensic Science International 195:148–152.

Rhine S. 1993. Skeletal Criteria for Racial Attribution. Race, Ethnicity, and Applied Bioanthropology:54–67.

Robedizo BP. 2016. The Identifiability of Osteological Traits on 3D Models of Human Skeletal Remains. 1–106.

Rogers TL. 2005. Determining the Sex of Human Remains Through Cranial Morphology. Journal of Forensic Sciences 50:1–8.

Savall F, Rérolle C, Hérin F, Dédouit F, Rougé D, Telmon N, Saint-Martin P. 2016. Reliability of the Suchey-Brooks method for a French contemporary population. Forensic Science International 266.

Shakya S, Nagraj SK. 2013. Morphology of Coronoid Process and Sigmoid Notch in Orthopantomograms of South Indian Population. World Journal of Dentistry 4:1–3.

Sholts SB, Walker PL, Kuzminsky SC, Miller KWP, Wärmländer SKTS. 2011. Identification of Group Affinity from Cross-sectional Contours of the Human Midfacial Skeleton Using Digital Morphometrics and 3D Laser Scanning Technology. Journal of Forensic Sciences 56:333–338.

Sholts SB, Wärmländer SKTS. 2012. Zygomaticomaxillary suture shape analyzed with digital morphometrics: Reassessing patterns of variation in American Indian and European populations. Forensic Science International 217.

Slice D E, Algee-Hewitt B F. 2015. Modeling bone surface morphology: A fully quantitative method for age-at-death estimation using the pubic symphysis. Journal of Forensic Sciences, 60: 835-843. doi:10.1111/1556-4029.12778

Spradley MK, Jantz RL. 2011. Sex Estimation in Forensic Anthropology: Skull Versus Postcranial Elements. Journal of Forensic Sciences 56:289–296.

Stull KE, Bartelink EJ, Klales AR, Berg GE, Kenyhercz MW, L'Abbé EN, Go MC, McCormick K, Mariscal C. 2020. Commentary on: Bethard JD, DiGangi EA. Letter to the Editor—Moving beyond a lost cause: Forensic anthropology and ancestry estimates in the United States. J Forensic Sci. 2020;65(5):1791–2. doi: 10.1111/1556-4029.14513. Journal of Forensic Sciences 66:417–420.

Stull KE, Tise ML, Ali Z, Fowler DR. 2014. Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. Forensic Science International 238:133–140.

Stull KE, Kenyhercz MW, L'Abbe EN. 2013. Non-metric cranial and pelvic traits as a measure of sexual dimorphism in a modern South African population. Proceedings of the 82nd Meeting of the American Association for Physical Anthropology; Apr 9-12; Knoxville, TN. Am J Phys Anthropol 2013; 150: 266.

Suchey J.M.; Brooks S.T.; Katz D. 1988. Instructions for use of the Suchey-Brooks system for age determination of the female os pubis. Instructional materials accompanying female pubic symphyseal models of the Suchey-Brooks system. Distributed by France Casting, Colorado.

Tallman SD, Go MC. 2017. Application of the Optimized Summed Scored Attributes Method to Sex Estimation in Asian Crania. Journal of Forensic Sciences 63:809–814.

Telmon N, Gaston A, Chemla P, Blanc A, Joffre F, Rougé D. 2005. Application of the Suchey-Brooks Method to Three-dimensional Imaging of the Pubic Symphysis. Journal of Forensic Sciences 50:1–6.

Villa C, Buckberry J, Cattaneo C, Lynnerup N. 2013. Technical Note: Reliability of suchey-brooks and buckberry-chamberlain methods on 3D visualizations from CT and laser scans. American Journal of Physical Anthropology 151:158–163.

Walker PL. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. American Journal of Physical Anthropology 136:39–50.

Walrath DE, Turner P, Bruzek J. 2004. Reliability test of the visual assessment of cranial traits for sex determination. American Journal of Physical Anthropology 125:132–137.

Warren MW, Parr NM, Zambrano CJ, Skorpinski KE. 2011. Bare bones: a survey of forensic anthropology. Dubuque: Kendall Hunt Pub Co.

Weber GW, Bookstein FL. 2011. Virtual anthropology: a guide to a new interdisciplinary field. Vienna: Springer.

Weber GW. 2015. Virtual Anthropology. Yearbook of Physical Anthropology 156:22–42.

Williams BA, Rogers TL. 2006. Evaluating the Accuracy and Precision of Cranial Morphological Traits for Sex Determination. Journal of Forensic Sciences 51:729–735.

Yaşar İşcan M, Loth SR, Wright RK. 1985. Age Estimation from the Rib by Phase Analysis: White Females. Journal of Forensic Sciences 30.

APPENDICES

Appendix A

Table A.1: Mandibular Trait Scoring Guide

Trait	Character State		
Shape of Inferior Border	Pointed 0	Square/Bilobate 1	Rounded 2
Undulation	Not Undulated 0	Slightly Undulated 1	Undulated 2
Pinching of Ramus	Not Pinched 0	Slightly Pinched 1	Pinched 2
Height of Coronoid Process	Short 0	Tall 1	- -
Chin Profile	Receding 0	Vertical 1	Protruding 2
Gonial Eversion	Inverted 0	Vertical 1	Everted 2
Shape of Mandibular Notch	Sloping 0	Wide 1	Round 2

Vita

Originally from Orlando, Florida, Sarah attended the University of Florida in Gainesville, Florida where she obtained dual bachelor's degrees in Psychology, with a specialization in Behavioral and Cognitive Neuroscience, and Anthropology. While there, she discovered a passion for physical and forensic anthropology, specifically human skeletal variation, and looked to pursue this passion in graduate school. Following the completion of her Master's program in biological anthropology at UTK, she hopes to pursue a Doctorate degree in biological anthropology at UTK, with the ultimate goal of teaching, conducting research, and consulting on casework within an academic setting.