In presenting this thesis in partial fulfillment of the requirements for an advanced degree at Idaho State University, I agree that the Library shall make it freely available for inspection. I further state that permission for extensive copying of my thesis for scholarly purposes may be granted by the Dean of the Graduate School, Dean of my academic division, or by the University Librarian. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Signature ___________________________________

Date ____

Performance of Pubic Symphysis-Based Adult Age-Estimation Methods Applied to Reconstructed Computed Tomography Scans

by

Kayla Eachus

A thesis

submitted in partial fulfillment

of Master of Science in the Department of Anthropology

Idaho State University

Summer 2021

To the Graduate Faculty:

The members of the committee appointed to examine the thesis of KAYLA EACHUS find it satisfactory and recommend that it be accepted.

Elizabeth Cartwright,
Major Advisor

Katherine Reedy,
Committee Member

Sara Getz,
Major Advisor

Curt Anderson,
Graduate Faculty Representative

ACKNOWLEDGEMENTS

I would first like to express my deepest appreciation to my advisor, Dr. Sara Getz, for her wisdom, guidance, and assistance throughout this project. Without her immeasurable knowledge and patience throughout the unexpected obstacles that have arisen over the last two years, and her willingness to partake despite them, I'm not sure I would have been able to accomplish the task set before me.

I am deeply indebted to my committee members for their valuable advice and feedback, and of course their patience, during the completion of this project. Their unwavering support is appreciated more than they know.

I would also like to thank Dr. Christian Peterson for his moral support not only over the last two years, but the last six, as his supervision of my academic endeavors began long before this project. I owe him a great deal of thanks for the non quantifiable impact he has had on my life.

Lastly, I would like to thank my friends and family, especially my husband, for their encouragement and love, and most of all for pretending to read this paper. I truly am fortunate to have so many incredibly supportive people involved in this project, and look forward to working with them in the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Performance of Pubic Symphysis-Based Adult Age-Estimation Methods Applied to Reconstructed Computed Tomography Scans

Thesis Abstract- Idaho State University (2021)

This study evaluates the performance of commonly applied age-estimation methods when applied to reconstructed computed tomography scans. Three popular methods, Suchey-Brooks, Hartnett, and Boldsen et al. Transition Analysis were applied to a sample of 169 reconstructed CT scans obtained from the New Mexico Deceased Image Database, and reconstructed using an open source software- 3D Slicer. The accuracy and bias of each method was compared to each of the other applied methods, as well as to validation studies conducted by other researchers. Similar trends (decreasing accuracy and increasing bias for advanced age) were found when these three methods were applied to digital bone that have been observed in previous studies applying these methods to dry bone. These trends were likely exaggerated due to the lack of visibility of features more closely associated with advanced age.

Key Words: Age-at-death estimation, Computed Tomography, Digital Bone, Suchey-Brooks, Transition Analysis, Hartnett, Pubic Symphysis

Chapter 1: Introduction and Background

Introduction

An accurate and precise age-at-death estimation is an important piece of information produced by anthropologists using human skeletal remains. In a medicolegal context, age estimation (along with estimation of sex, stature, and ancestry) may aid in positively identifying an unknown individual. Age-at-death estimations also form the foundation for reconstruction of past population demography for most of human history. Formal methods for the estimation of adult age have been around since the 1920s, yet despite a century of method development, the improvement in the ability to generate accurate and precise age-at-death estimations has been unremarkable.

While the progress on improving the ability to estimate adult age has been minimal, there have been astounding developments in the field of medical imaging over the last century. From the first roentenograph (named after Wilhelm Conrad Röntgen- who discovered the x-ray) in 1895 (Roentgenogram, Encyclopedia Britannica 2009) to the modern use of a variety of imaging technologies, the advancement of medical imaging has opened a new realm for medical practitioners and researchers alike (Bradley 2008). The evaluation of internal structures is possible without high-risk surgeries, and with little discomfort to individuals being imaged. Historically medical images, such as computed tomography (CT), lacked the resolution necessary to see small features, such as that on bone, and processing power necessary to analyze the scans were costly. Over the past decade, the costs associated with data storage have become more manageable and CT scanners (as well as other medical images) have become more commonplace in medicolegal contexts (Reid et al. 2008). This development in imaging technology can be of particular use to forensic anthropologists.

Computed tomography (CT) is of particular interest, as it is able to capture characteristics of both cortical and trabecular bone without requiring the destruction of the original sample. In addition, bone can be digitally reconstructed from CT slices, creating a three-dimensional model that anthropologists may be able to evaluate in a manner similar to dry bone without requiring the lengthy process of maceration (i.e., the physical and/or chemical removal of soft tissues). This is especially relevant in cultural or religious settings where alteration of the body is not appropriate. The use of CT also provides an interesting possibility for large, digital skeletal collections, which may provide a more accessible and varied supplement to existing physical collections. In fact, the first large, publicly accessible database composed of CT scans─the New Mexico Deceased Image Database (NMDID)─was released in February of 2020, providing a new critical resource to researchers of various disciplines.

In order to use digital models of bone to their fullest potential, the limitations of digital bone as a substitute for dry bone must be understood. This study applies three age estimation methods developed using physical skeletal samples to approximately 169 digital bone models, reconstructed from Computed Tomography (CT) scans obtained from the NMDID. The sample for this study was collected from NMDID and includes approximately 24 individuals from each decade of life. The scans were reconstructed using the free open source software 3D Slicer

(Kikinis et al. 2014). Age at death was be estimated using two phase-based methods for the pubic symphysis: Brooks and Suchey (1990), Hartnett (2010), and one component-based method for the pubic symphysis and auricular surface .

The sample used was scored in a randomized order three separate times, once for each method. A basic statistical analysis of the results comparing the overall accuracy and bias was performed for each applied method. The accuracy (reported as a percentage) and bias of each age cohort (20-29, 30-39, 40-49, etc.) and each method as a whole were compared to validation studies conducted by Savall et al. (2018), Merritt (2014), and Milner and Boldsen (2012) in order to asses the performance of the Suchey-Brooks method, the Hartnett method, and the pubic symphysis component of Boldsen et al. Transition Analysis (TA) when applied to digital bone models compared to the method performance on dry bone, as originally intended.

As medical imaging becomes more accessible, it will become an imperative tool to researchers of various disciplines (Dirnhofer et al. 2006). In order for this new data source to be used to its full potential, its limitations as a substitute for dry bone must be understood. The aim of this study is to evaluate whether there is a significant difference in accuracy, precision, bias, and error when commonly used age-at-death estimations developed for dry bone are applied to digital bone reconstructed from CT scans.

Background

An Overview of Age Estimation

Skeletal age estimation has been a topic of interest for over a century. In 1920, Todd developed the first formal phase-based method for the pubic symphysis. This system was later criticized for sacrificing accuracy for precision and use of an inadequate sample size for method development (Meindl et al. 1985, Katz et al. 1986, Brooks et al. 1990). In 1986, Katz and Suchey revised the original Todd method by modifying the narrow age ranges of ten phases into a more manageable six and providing standardized images, casts, and written descriptions for each phase to assist observers with correctly applying the revised method (Katz et al. 1986, Brooks et al. 1990).

In 1957, McKern and Stewart published the first component-based method for estimating age from the pubic symphysis (McKern et al., 1957). The sample used was composed of male Americans killed during the Korean War. A similar study with a female sample was later done in 1973 by Gilbert and McKern (Gilbert et al., 1973). Both samples, particularly the sample used by McKern and Stewart, were limited terms of age distribution, resulting in an age estimation method heavily influenced by reference sample mimicry. Because of this, the early component based method developed by McKern and Stewart never gained widespread usage in forensic or archeological communities.

In 2002, Boldsen and colleagues developed a new component-based approach using features of the cranial sutures, pubic symphyses, and iliac auricular surfaces that they called

Transition Analysis (TA). The publication of the new method and the accompanying manual and software demonstrated the importance of incorporating skeletal components from multiple locations in the body. The TA method combines features from different areas of the skeleton into a single estimate, which is then “weighted” based on prior knowledge about age-at-death distributions in human populations. In addition, the Boldsen et al. (2002) method produces confidence intervals for each estimate (Boldsen et al., 2002).

Evaluations of commonly used ate-at-death estimation methods have demonstrated that current practices are limited in their ability to reliably produce both accurate and precise estimates. An ideal method requires a balance of the two (i.e., intervals that are narrow enough to be practically useful, while still being highly accurate), but such a method does not yet exist. One approach to this problem has been the investigation of new methods incorporating multiple components of the body with documented age-related change.The development of these methods may be aided by large, easily accessible skeletal samples. Currently, such physical skeletal collections are limited due to several practical limitations addressed in detail in a later section. The use of digital bone models as a supplemental or substitute sample base provides a reasonable alternative to dry bone in the development and evaluation of various age-estimation methods. This study will evaluate several commonly used methods using digital bone models, in order to assess the potential limitations of digital bone as a sample source for future research.

Criticisms of Existing Approaches

Phase-Based Methods

In forensic anthropology, the Brooks and Suchey pubic symphysis method is the most widely used age estimation method (Garvin & Passalacqua, 2012). The written descriptions and standardized casts are easily accessible. Despite the common use of long-lasting phase-based age estimation methods, like Brooks and Suchey (1990), there are some major criticisms of these methods. Phase-based methods assume multiple characteristics progress as a whole and continue to correlate with age at an equal rate throughout the entire adult lifespan. When characteristics contradict each other, phase-based methods leave it up to the observer to determine which of these characteristics should be most heavily weighted in the determination of phase placement. **(**Shirley et al. 2015, Savall et al. 2018)

This introduces inherent subjectivity to phase based methods. Two different observers may place the same sample into different phases based on their personal hierarchy of characteristics (Shirley et al. 2015). A noted phenomenon in the practice of the Suchey-Brooks method is the increase of inter-observer error between experienced anthropologists. While novice observers tend to score the same sample in the same phase, more experienced observers tend to disagree with each other more frequently because of the difference in their subjective characteristic hierarchies (Shirley et al. 2015, Savall et al. 2018).

Component- Based Methods

The importance of incorporating multiple age indicators into an age estimation method cannot be understated. Some skeletal aspects may degenerate and ossify during different periods of life, ceasing to deliver information after a certain point, while skeletal aspects may not undergo age related change until later in life, revealing more information about latter decades (Getz et al. 2016).

The incorporation of multiple skeletal components is a strength of multi-feature methods, including Transition Analysis (Boldsen et al. 2002), however it can be time consuming to score multiple indicators individually. While this may not be of particular significance when a single individual needs to be evaluated, time can be a major factor when a large sample is in use, or in certain forensic circumstances with a large number of individuals.

Limited sample base for method development

A significant underlying issue in the inability to improve age-at-death estimation is the limited reference samples available to researchers. The same samples are used to create nearly all methods because of the limited number of large human skeletal collections. While there are several collections in various countries around the world, many are relatively limited in the amount of human variation they cover (Forensic Anthropology Society of Europe cite).

To make current methods more applicable to a larger range of human variation, better reference samples are a necessity. However, there are several obstacles to creating larger skeletal samples. The amount of time and personnel needed to process thousands of sets of remains is simply not a feasible task for most collections, which typically relay on one, or a very small number of paid employees, volunteers, and students. Thousands of remains would be tens of thousands of labor hours, as well as costly in terms of personal protective equipment for workers, curations costs, and storage.The amount of storage space necessary to store thousands of sets, as well as the personnel to manage and maintain the collections of remains are yet another set of practical obstacles.There is also the simple fact that in order to process a few thousand sets of remains, a few thousand people would have to die and donate their bodies to skeletal collections, this simply is not a practical expectation. In addition, by the time the number of necessary remains were collected, processed, and accessioned into the collection, they would have already passed the mark of a temporally relevant sample.

Age Estimation Methods Developed for CT Scans

In recent years, several age-estimation methods have been proposed specific to the unique abilities of computed tomography. These methods utilize the various components of CT and its related software to evaluate characteristics of bone not typically visible, such as the quality of cortical and trabecular bone (Villa et al. 2013). CT is also able to measure aspects of bone that are difficult to assess quantitatively. This is particularly useful in age-estimation methods that focus on the Os Coxae because of their non-linear shape. For example, shifts in subpubic contour have been correlated with age and can easily be measured with basic CT processing software (Chiba et al. 2014). Connective tissue is another age indicator seldom explored by forensic anthropologists, however the thickness of connective tissue (i.e. the pubic symphysis) has been negatively correlated with age. The use of CT has allowed researchers to measure and evaluate this relationship, without requiring lengthy maceration processes. (Villa et al., 2013; Wink 2014; Grabherr et al., 2009). Other methods use a more technologically intense approach, focusing on the qualities of CT that are not applicable to visual methods using dry bone. One particularly technologically intense method analyzes pixel grayscale values to create a quantitative evaluation of bone density that is known to be inversely related to age (Alcaraz et al. 2015).

Several attempts have been made to apply methods developed using dry bone to digital models reconstructed from CT scans. Nemerskeri's complex method (Ascádi & Nemerskeri, 1970) has been applied to a small sample with comparable results to the original method study (Grabherr 2009). The Suchey-Brooks methods has been applied to digital bone models as well, again with similar results to studies performed on dry bone (Savall 2018). A new component method based on characteristics of three common phase-based methods was applied to a small digital sample. While only two of the nine components significantly correlated with age, the sample size limits the conclusions that can be drawn from this study (Ferrant et al. 2008). The application of commonly used age estimation methods to digital bone seems to be promising, however the general consensus is that there needs to be more research with larger samples.

Limitations of Digital Bone

When applying methods developed using dry bone to digitally rendered bone, there are several limitations that must be taken into consideration. While detailed reconstructions are possible when images are taken close together with significant overlap, microporosity (small, pin-prick sized holes in the bone surface) generally is not visible in digital reconstructions. This is because of the technological constraints of current computed tomography. As CT scanners are designed to image the living, there are design constraints of the scanners themselves that aim to reduce the negative impact of the radiation used to produce the images of living individuals. This places some limitations on the amount of resolution visible in computed tomography. Most scanners can produce an image overlap up to 50%, but anything more would produce too much radiation in living individuals (Ginat et al. 2014). The scanner used to produce the data in the NMDID collected images 1mm apart with 50% overlap.Though the images were collected from deceased individuals, a medical scanner with design parameters for the living was used.

While there is a significant amount of image resolution, it is impossible to see microporosity because the holes in the bone are smaller in diameter than the 0.5mm image overlap. Generally, microporosity is a trait associated with relatively advanced age. Microporosity can also be indicative of poor bone quality, another characteristic of advanced age. The inability to view microporosity may lead to a tendency to underestimate the age of older individuals (Barrier et al. 2009).

Surface texture is another characteristic unobservable with the use of computed tomography. Texture is by nature a characteristic dependent on the manipulation and tactile feel of the bone. It does not lend itself well to evaluation through digital reconstruction (Barrier et al. 2009). Most reconstructed bone appears relatively smooth in digital reconstructions. If the bone is deteriorating, macroporosity may be visible. Generally, smooth textured bone is associated with the earlier decades of life, while course grained bone and microporosity are associated with the middle ages of life (Boldsen et al. 2002). In CT data bases, course grained bone and microporosity will not be visible, this may lead to an underestimation or overestimation of age-at-death for individuals in the middle decades of life.

A new obstacle unique to the use of reconstructed CT scans is the concept of false macroporosity. Digital data units that convey information in two planes (the x and y), are commonly known as pixels. Each pixel conveys information through an arbitrary numerical value that a computer of choice recognizes as a code for color. Information is converted through CT in much the same manner, however the data collected from a CT scan is in three dimensions (x, y, and z), and must convert information in the same three dimensions. This is a combination of volume and pixel, or *voxel*. Voxels, similar to pixels, have an arbitrary number associated with them that conveys a greyscale value. This value is based on the density of the tissue the cone beams from the CT penetrate (Spin-Neto et al. 2014). In the reconstruction process described below, a threshold value of voxel is selected to isolate bone from its surrounding tissues. False macroporosity occurs when the cortical bone is so thin the voxel density falls below the selected threshold for digital reconstruction, causing porosity in the digital bone that may not be present in the dry bone. However, the thinning of cortical bone is suspected to be the step immediately preceding macroporosity in the aging process. False macroporosity likely does not lead to a significant overestimation of age (Barrier 2009).

Bone weight is another important factor in determining overall bone quality. It is of particular importance in the Hartnett (2010) pubic symphysis method. Phases 5, 6, and 7 of the Hartnett method describe the loss of bone density and a light-weight feeling as a phase placing criterion. It is a major factor in determining whether an individual should be described as a phase 6 or 7. This may lead to an underestimation in age, or a wider age range encompassing the brackets of phase 6 or 7 if the determination cannot be made (Hartnett 2010). Although reduced bone density can be observed in CT data, there is not an easy way to make a correspondence between subjectively classified bone weight and the 3D CT reconstruction.

Methods Used

This thesis evaluates three different age estimation methods: Brooks and Suchey 1990, Hartnett et al. 2010, and Boldsen et al. 2002.

The Suchey-Brooks method is a phase-based method that scores the age related change in the symphyseal surface, adapted from the method published by Todd in 1920. While other similar methods exist (McKern and Stewart, 1957; Nameskéri and Askádi,1970; Harsanyi and Ascádi), this is the most commonly used among anthropologists for age-at-death estimation (Garvin & Passalacqua 2012). The sample used by Suchey and Brooks was collected between 1977 and 1979 in the Los Angeles County Medical Examiner Office with documented information on race, occupation, socioeconomic class, and age at death. The ages ranged from 14-99, however the overall sample was skewed to ages from 18-55 (Brooks et al. 1990, Savall 2018). Individuals are classified into one of six different phases based on criteria described by

Suchey and Brooks. Phases one is relatively narrow, however subsequent phases include age brackets spanning 20-50 years with different age brackets for males and females.

The Hartnett method is comparable to the Suchey-Brooks method in that it is a phase-based method scoring the pubic symphysis. The sample used was more modern, collected in Phoenix, Arizona between January 2005 and June 2006. This method does not use a separate age scale for males and females. The sample was sorted into broad groups base on overall bone quality, and further subdivided using more specific criteria and morphological characteristics. This subdivision resulted in seven groups. Written descriptions were made for each group, and basic statistics were recorded including the actual ages of individuals within each group, the mean, standard deviation, and range. Many of the characteristics described for each phase are similar to the written criteria of the Suchey-Brooks method, particularly in the earlier phases. Hartnett does include additional criteria regarding the weight of the bone beginning in phase 5. Weight is of particular importance in phases 6 and 7, and may in fact be the deciding factor in determining if an individual belongs in phase 6 or 7.

The third method that will be used in this study is Boldsen et al. 2002 Transition Analysis for the iliac auricular surface and pubic symphysis. Boldsen et al. developed this method using Terry and Coimbra skeletal samples. This method of transition analysis examines the degree of cranial suture closure and various components of the face of the pubic symphysis and the iliac auricular surface, totaling to 33 individual components with unique characteristic descriptions. While this method originally included the evaluation of cranial sutures, however cranial sutures cannot be effectively evaluated to the necessary degree using CT scans, and will not be included in this study (Lopez-Alcaras et al. 2015). In addition, cranial sutures have recently been found to contribute a negligible amount of information to the overall age estimation, and are poorly correlated with age as a whole (Boldsen et al. 2012). While the auricular surface was originally going to be used in this study, it was not included for several reasons after observing this feature on the digital reconstructions. The observable characteristics were very few, and not observable on all individuals in the reconstruction due to varying scan quality, affecting the reconstruction. It is the unique combination of the 33 characteristics included in TA that result in more individual age range estimates and maximum likelihood than phase based method, however not all 33 characteristics need to be included in order to use the associated software to generate an age estimation. Rather, by reducing the number of included features- in this study particularly- from the typical 33 to 4 did affect the estimates of each individual in that there are fewer possible combinations of scores, and ultimately less variation in the individual estimations.

Due to the limitations of digital bone previously discussed, a subset of symphyseal face characteristics of the pubic symphysis and the iliac auricular surface have been chosen ahead of time for evaluation:

Pubic Symphysis features (4 of the 5 method traits):

- 1) symphyseal relief
- 2) superior protuberance

- 3) ventral symphyseal margin
- 4) dorsal symphyseal margin.

Evaluation of Methods

There is no official standardized method of evaluating an age estimation method. However, most validation studies report some form of accuracy and bias. To evaluate the Suchey-Brooks method, results from this study were compared to a validation study performed by Savall et al. (2018). This validation study compared method performance applied to the original Suchey-Brooks sample and method performance when applied to a sample of digital bone models reconstructed from multi slice computed tomography (MSCT) scans. While MSCT was not used as the sample source for this study, no other validation studies were found at the time of this paper comparing the accuracy and bias of Suchey-Brooks method performance on dry and digital bone samples. This study was selected for comparison due to the similarities in structure and sample.

The Hartnett method is newer, the original method was published in 2010. Compared to Suchey-Brooks and TA, the Hartnett method is not as widely used. Because of this, there are not many validation studies of the Hartnett method. At the time of this paper, no validation studies were found comparing the performance of the Hartnett method when applied to dry bone versus digital bone. Only one validation study with reported accuracy and bias rates was found, conducted by Catherine E. Merritt (2014). This method was selected for comparing method performance in the form of accuracy and bias.

Milner and Boldsen (2012) conducted a validation study of transition analysis, reporting the accuracy and bias of the age range estimate and maximum likelihood generated by each component individually (the cranial sutures, the bilateral iliac auricular surface, and the bilateral pubic symphysis) and the estimates of all three components in combination. Only the data associated with the pubic symphysis was used for comparison in this study, given that the iliac auricular surface and cranial sutures were not used.

Sex-specific differences in accuracy and bias are often assessed in validation studies as well, and reported in all three validation studies selected for comparison. Accuracy is typically reported as a cumulative number (a percentage reflecting total males/females included in the sample, rather than a percentage of males/females per each included decade- e.g. 90% of 20 year old males, 88% of 40 year old females). To facilitate comparison of method performance in this study and reported data of validation studies, the accuracy and average bias of each method for males and females is included.

Chapter 2: Materials and Methods

The sample used in this study was obtained from records in the New Mexico Decedent Image Database (NMDID) that matched predetermined criteria (e.g., age, ancestry, medical history). Approximately twenty-four individuals were requested from each decade of life (e.g., 20-29, 30-39, 40-49, etc.), twelve male and twelve females, to create an equally distributed sample of 204 individuals. .For each individual, a three dimensional model of the os coxae was reconstructed from the requested CT scan file using an open source software (3D Slicer, http://www.slicer.org, Fedorov et al. 2012) and a standardized reconstruction process developed by the observer using current literature referencing programs similar to 3D slicer and through personal trial and error working with the NMDID data files It should be noted that no standardized reconstruction process for creating 3D models from CT data for purposes analogous to this study currently exists, so the author's best judgement was used in developing this process. Because suitable reconstructions could not be produced from the data from some individuals, not all individuals were included in the final sample (N=169). Three age estimation methods, Suchey-Brooks (Brooks and Suchey 1990), Hartnett (Hartnett 2010), and Boldsen et al. Transition Analysis (Boldsen et al. 2002), were applied to the pubic symphyses of the individuals included in the final sample. The iliac auricular surface was initially also reconstructed for each individual, however, insufficient detail could be seen for age assessment, so this joint was not included in this study.

Sample Selection

New Mexico Deceased Image Database (NMDID)

The study sample was obtained from the New Mexico Decedent Image Database (NMDID)- a free database funded by two National Institute of Justice Research Grants. The first was awarded to the New Mexico Office of the Medical Investigator (OMI) to fund CT scans in cases where the cause of death was suspected to be related to blunt force trauma, ballistics, childhood trauma, or overdose (NIJ 2010-DN-BX-K205, Kurt Nolte, Principal Investigator). CT scans ultimately proved to be useful in determining cause of death, both as supplemental as supplants to traditional autopsy. Because of the success of this research, CT scanning became a routine part of death investigation at OMI, allowing for the development of the NMDID after the award of an additional NIJ grant to compile collected CT scans into a single database (2016-DN-BX-0144, Heather Edgar, Principal Investigator).

Currently, the NMDID is composed of CT scans from 15,000 individuals whose cases were processed by the OMI between 2010 and 2017. All individuals are assigned a random six digit code as their deidentified record number (DID) in order to anonymize personal data, and to accommodate HIPPA regulations. For each individual in the database, next of kin was determined by the OMI. As long as the next of kin was over 18 years of age, and not imprisoned in any institution, a telephone survey was conducted through the OMI in an attempt to gather detailed demographic information about each decedent, Standardized questionnaires regarding sex, gender identification, drug use, socioeconomic status, ancestry, maternal and paternal birth place, decedents' birth place and zip code at time of death, parturition, etc. were offered in both English and Spanish to qualifying next of kin. As this process was labor intensive, next of kin who were not contacted after two attempted phone calls were not pursued further. Demographic information is not available for every individual, though it is present for a large portion of individuals in the NMDID.

The database is accessible to researchers from all fields, as long as they are able to demonstrate an appropriate academic or professional affiliation at the time their account is created. Once the application for an account is accepted, a researcher can enter their sample criteria into the search engine, and view the metadata of individuals that meet their criteria. Specific sets of scans can be added to the 'cart', up to 500 at a time and a brief research proposal is required to officially request the sample. After the data request is approved, the requested sample is available for download and can be saved to the researchers' choice location.

Age

Twenty-four individuals—twelve males and twelve females—from each decade of life were requested to create a test sample with as uniform an age distribution as possible. This age structure facilitates comparisons of accuracy and bias in different portions of the lifespan. Individuals below 20 and over 90 years of age were grouped to form the youngest and oldest age categories (i.e., 15-19 years, and 90+) as there are fewer individuals available in these categories than in other portions of the lifespan (Figure 2.1).

The minimal age of the individuals in this study was set at 15 years. The methods used in this study would not typically be applied to individuals this young, as methods based on the fusion of epiphyses would provide more accurate and precise results in this age category (White et al. 2012). However, individuals of this age were included because it is possible that all three elements of the os coxa may be fused at this time, as fusion between the ilium, ischium, and pubis occurs around puberty, which has been shifting younger in recent decades (Baker et al. 2005, Mendle 2019).

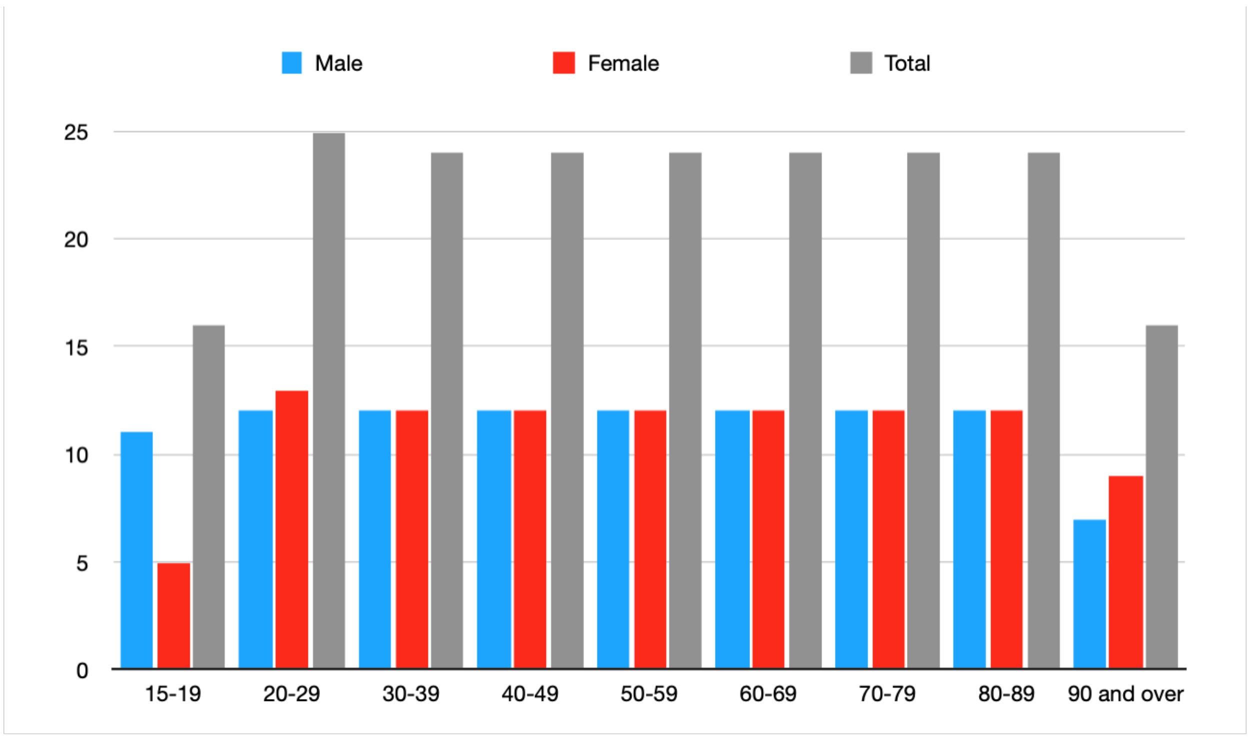

Figure 2.1 *Distribution of the original sample requested from the NMDID including the number of individuals (y-axis) and age cohort (x-axis). Total number of males (blue), total number of females (red) and the combined total (Grey).*

and females. As this thesis is a study of the applicability of dry bone methods to digitally rendered reconstructions, it is logical to include the youngest individuals used to develop the dry-bone methods that are being tested. Individuals who did not demonstrate complete fusion the ilium, ischium, and pubic were not included in the final sample.

Table 2.1 *Individuals not included in the final sample, and the reason for exclusion*

Reason not used	Males	Females	Total
Artifact	5	6	11
Noise or Anomaly	4	5	9
Os Coxa(e): not fused or trauma	3	5	8
Corrupted File	2	2	4
Error in Sequence	2	0	2
Total Eliminated	16	18	34

Familiarization with Methods Applied to Digital Data

Before the primary sample data were collected, 20 randomly selected individuals were scored with each method. This was done to familiarize the observer with how the features would present in CT scans and reduce the likelihood of introducing error or scoring bias not related to differences in how the methods perform in the sample. Individuals in this sample ranged from 23 to 73 years of age and were reincorporated into the primary sample after their evaluation.

Final Sample Distribution

The initial sample requested from the NMDID consisted of 204 individuals; however, a total of 34 individuals were not included in the final sample because of file corruption issues, bony features that could indicate young age or skeletal trauma, the presence of surgical implants, and unusual noise or anomalies in the scan reconstruction. Additionally, this study was meant to include only fleshed individuals who were CT scanned, so several scans of skeletonized individuals were also not included in the final sample (Table 2.1).

While downloading CT series from the NMDID, four data files were corrupted during the decompressing process, and could not be properly converted to a DCM file for processing. Several attempts were made to download this data, which did not resolve the issue. These four individuals were not used in the study.

Any individual demonstrating a separation the pubis from the ilium or ischium was not included in the final sample. This separation was determined by the CT scans, not the reconstruction. Because the observer was blinded to the reported age of all individuals in the sample, detachment of the pubis from the rest of the os coxa could indicate trauma- more likely in adults of advanced age- or a lack of epiphyseal fusion were treated in the same manner, as the observer could not determine if trauma or a subadult age was responsible for the separation looking at the scans alone. This disqualified a total of eight individuals from the final sample.

An additional eleven individuals were not included in the scan due to the presence of bodily artifacts. In most cases, this was a hip replacement, however in a handful of individuals screws through the sacrum or femoral head prevented their reconstruction. All 3D models were created using the same process described later in this section. Part of this process included an established threshold value. Because metal absorbs the X rays used in a CT scan differently than bone, the threshold values used for reconstruction were not effective in these individuals, and no threshold value could be determined that resulted in a useable reconstruction while eliminating excess noise present as a result of the given artifact (Figure 2.3).

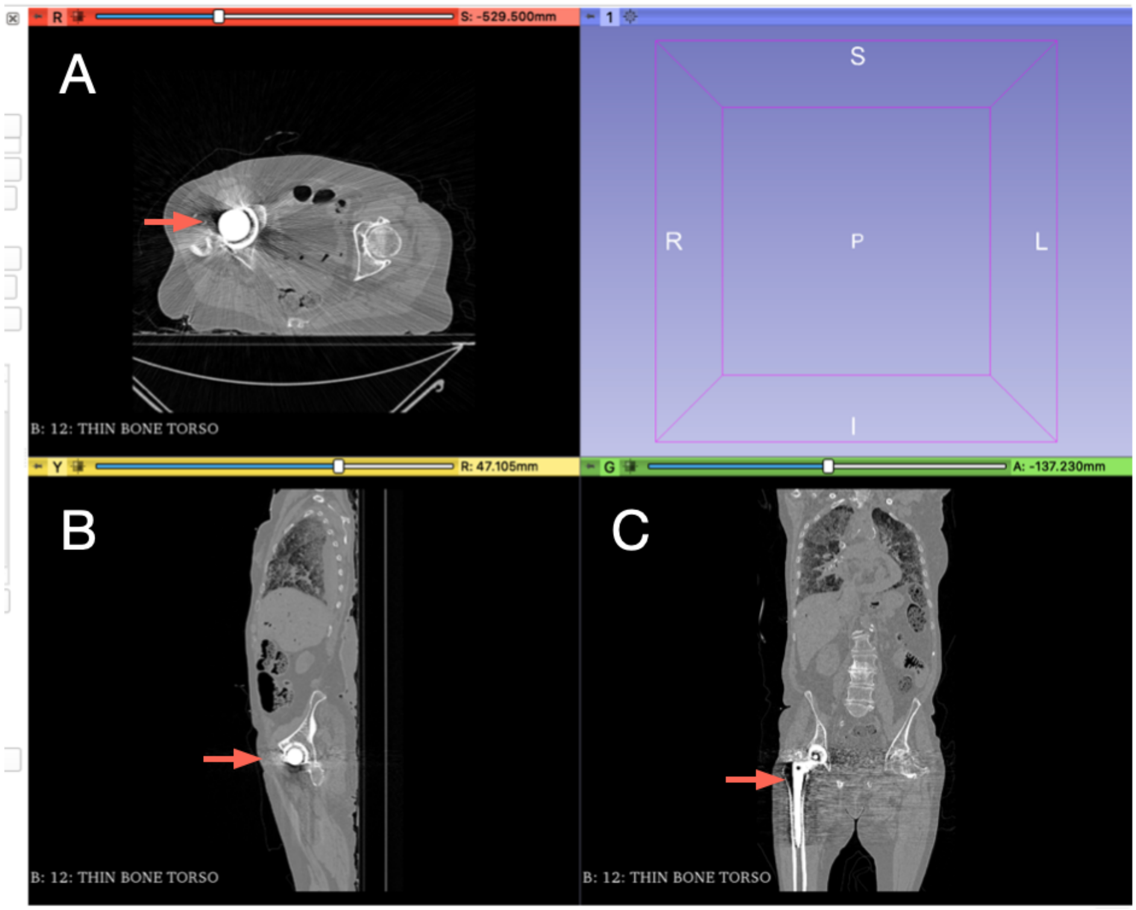

Figure 2.2 *Bodily artifact observable in scan, affecting contrast and reconstruction. Visible in the transverse (A), sagittal (B) and coronal (C) planes.*

Excess noise and anomalies in the scan obscuring the region of interest prevented the reconstruction of nine individuals, who ultimately were not included in the final sample. While in most cases adjusting the contrast of the scan greatly reduced or eliminated the effects of noise in the scan, in several individuals this was not the case and the os coxae could not be found in one or all of the image planes. In several datasets, black anomalies obscured the region of interest, and ultimately prevented reconstruction. These were likely attributable to a buildup of gas in the body, preventing the complete penetration of x rays and imaging of the bone (Figure 2.3).

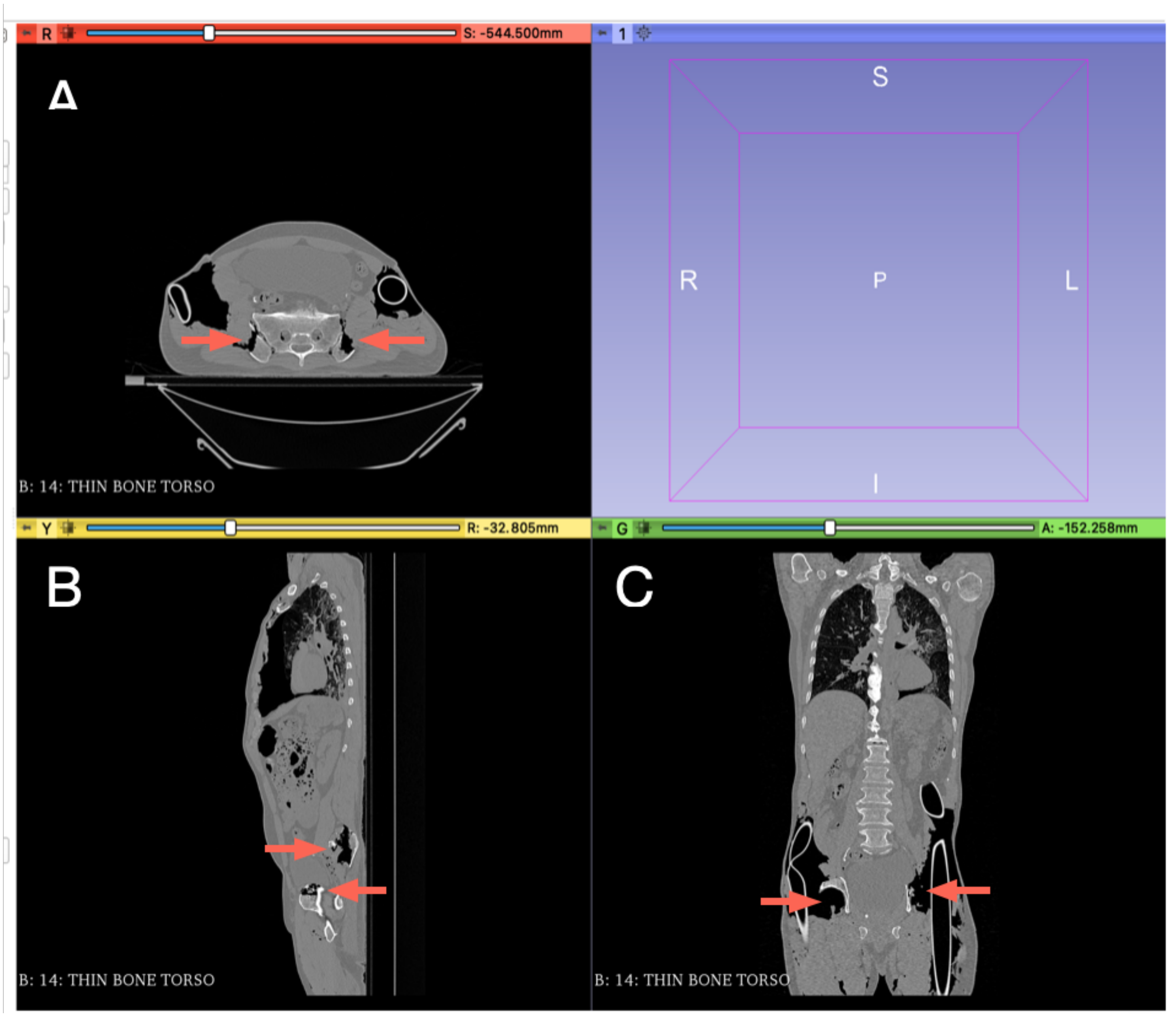

Figure 2.3 *Anomaly in scan obscuring the os coxae in the transverse (A), sagittal (B) and coronal (C) planes*

Lastly, two individuals were not included in the final sample due to an error in the dataset itself. One individual was a "bone only" scan. The decomposition of this individual was so far progressed that all remains were skeletonized. However, this individual was still scanned, as this is standard protocol at OMI. They were not included in this study because all other individuals included in this sample were still fully fleshed. There is not yet enough research to determine the degree of difference in skeletal reconstructions of fully fleshed individuals and skeletal remains in terms of applied methods. The last individual not included did not contain a “thin bone sequence” in the data set.

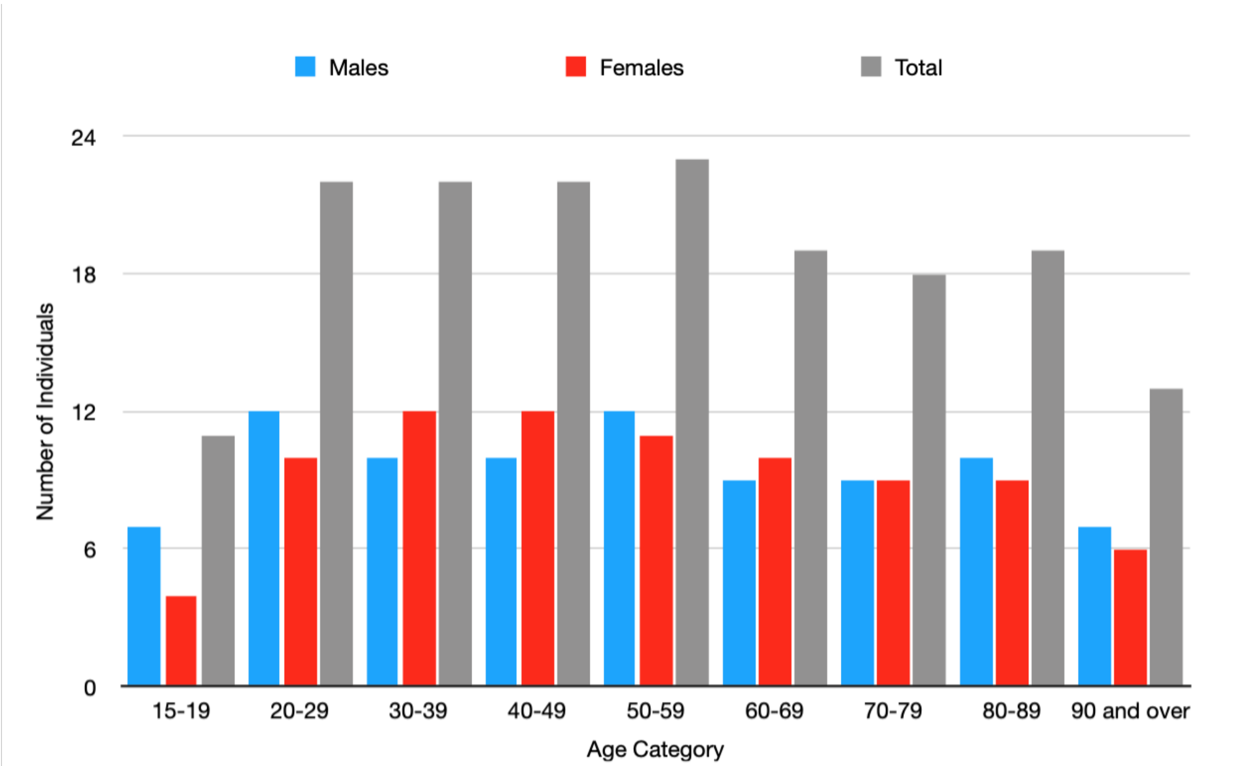

Figure 2.4 *The distribution of the final sample included in this study including the total number of individuals (y-axis) of males (blue), females (red), and combined sexes (grey), distributed according to age cohort (x-axis)*

Ultimately, this resulted in a final sample size of 169 individuals: 86 males and 83 females, distributed according to Figure 2.4.

Ancestry

As this study is intended to be a test of the applicability of dry bone methods to digitally reconstructed bone, only self-identified white individuals were requested. Various studies over the last several decades have emphasized the population specific trends in age progression. This has been a major criticism of the Suchey-Brooks method (Hartnett 2010), especially because the original sample is no longer available for further research use and the sample used to develop the Hartnett method was also composed of predominantly white individuals (598/630). To reduce the potential error resulting from differences in the reference sample composition of the methods applied and other validation studies, only white individuals were requested from the NMDID. While there are many factors that may create differences between samples- life history, nutrition, socioeconomic status, exposure to environmental pollutants, time period during life of the individuals, developments in medical technologies, etc.- creating a sample with no idiosyncratic variation is not feasible. Therefore, efforts were taken, in so far as possible, to reduce variation between test samples that has been documented by other studies in the reference to the age estimation methods applied in this study.

CT Scan Information

Scanner and data series

Each adult individual in the NMDID underwent three full body scans in three planes- coronal, sagittal, and transverse- taken using a Philips Brilliance Big Bore 16 Ct scanner. CT parameters were 1.0 mm slice width with 0.5 mm of overlap (tube voltage 120 kVp, effective mAs: 300). The data file for each individual has approximately 18 different series of scans in which all three planes are visible. If anomalies or pathologies were present, more scans focusing on the region of interest were generated. Each scan series was categorized by region of interest (i.e. head/neck, upper extremities, lower extremities, torso), tissue of focus (i.e. brain, bone, lung, general soft tissue, etc.), and number of slices (ex. 3x3 scans).

The torso data series was of primary interest for this study because it contained both areas of interest (pubic symphysis and iliac auricular surface). For each individual the 'thin bone torso' dataset was selected for reconstruction to optimize the amount of detail visible in the bone model. Images taken began just above the clavicle and ended just below the ischial tuberosities. Several other data series containing the pelvis were available, including 'soft tissue' datasets. However, the algorithm used in the thin bone sequences differed from the soft tissue scans. The parameters for 'bone' sequences set by the OMI better highlighted bone compared to other tissues. A '3x3' torso sequence including the pelvis was also available. However, for the same region (just above the clavicle to just below the ischial tuberosities) a total of 300 images with 3 millimeters spacing and no slice overlap were taken, compared to the total of 1,600 images with .5 millimeter overlap present in the 'thin bone sequence'. Overall, the thin bone sequence was determined to have the most available resolution for reconstructions (Edgar et al. 2020).

CT Scan Processing

All scans were downloaded from the NMDID as .dcm (Digital Imaging and Communication) files to an external hard drive. The scans were then uploaded to 3D slicer, an open-source software package intended for viewing and reconstructing medical images for research purposes (http://www.slicer.org, Fedorov et al. 2012), and the 'thin bone torso' data sequence was selected from file options. A step wise process was followed for reconstructing all os coxae to minimize variation in the reconstruction and expedite processing time.

1. If necessary, the contrast was adjusted in the DICOM module (the opening screen of 3D slicer) to clearly highlight the bone in the coronal, transverse, and sagittal views (Figure 2.5).
2. The scan was cropped around the region of interest, the left and right os coxae, in the coronal, transverse, and sagittal views (Figure 2.6).
3. The threshold effect in the segment editor module was adjusted to a minimum voxel value within the range of 220-230 (an experimentally determined threshold in the first dataset of 20 individuals). This value range consistently eliminates the soft tissues from the reconstruction and the fluctuation of 'noise'- a grainy appearance in the image due to an unwanted change in voxel values (Verdun et al. 2013) (Figure 2.7, 2.8).
4. The threshold effect is used for masking. This function applies a minimum threshold to the entirety of the scan, dramatically reducing processing time. Using the threshold tool, larger areas of the scan can be 'painted over' in the next step, telling the program which tissues (in this case bone) are of interest, without including soft tissues, ultimately allowing for segmentation of the right and left os coxa (Figure 2.8, 2.9).
5. The paint tool is used with a 10% sphere brush, which penetrates multiple slices at a time and helps to reduce processing time. Fine details are painted with a 3-5% sphere brush. The right and left side were segmented separately using different pain 'colors', telling the

program to generate separate models (one for the right half and one for the left) rather than a single model including both os coxae (Figure 2.9).

6. The right and left sides are segmented using the segment editor module (Figure 2.10).
7. The 3D models for the right and left halves were generated and saved under the decedent image database number (the anonymized number assigned to each individual by the NMDID) to an external hard drive (Figure 2.11).

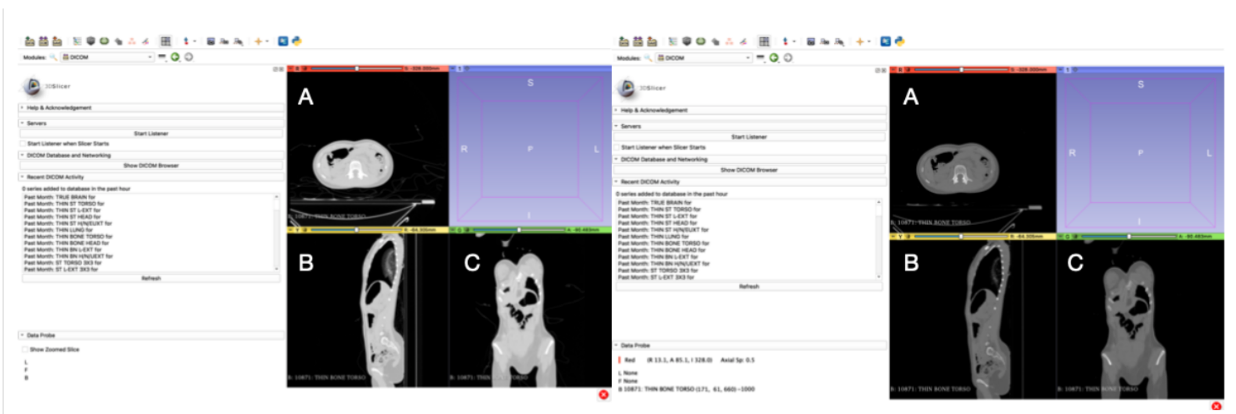

Figure 2.5 *Step 1: Contrast adjustment to better highlight bone in the transverse (A), sagittal (B), and coronal (C) planes*

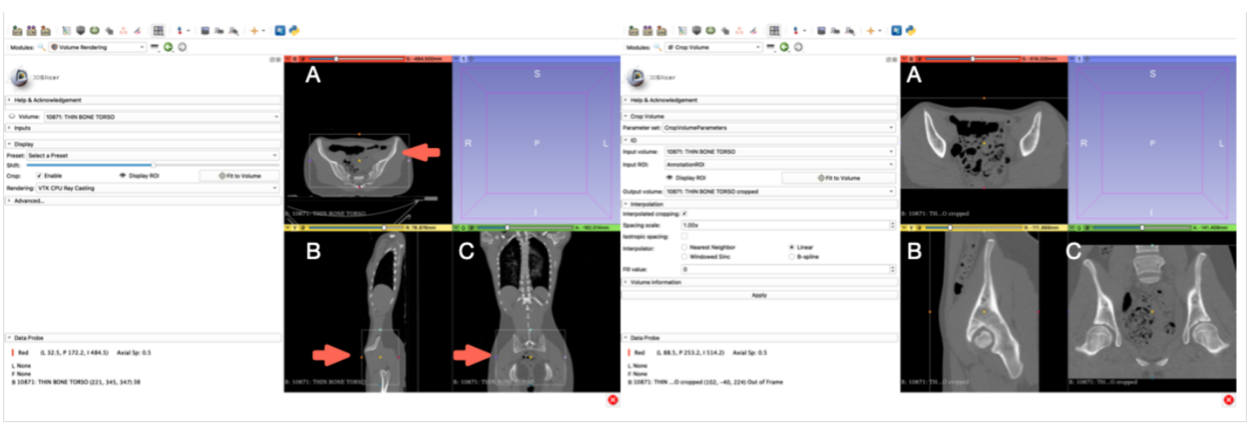

Figure 2.6 *Step 2: Image is cropped around the region of interest 9the os coxae) in the transverse (A), sagittal (B), and coronal (C) planes*

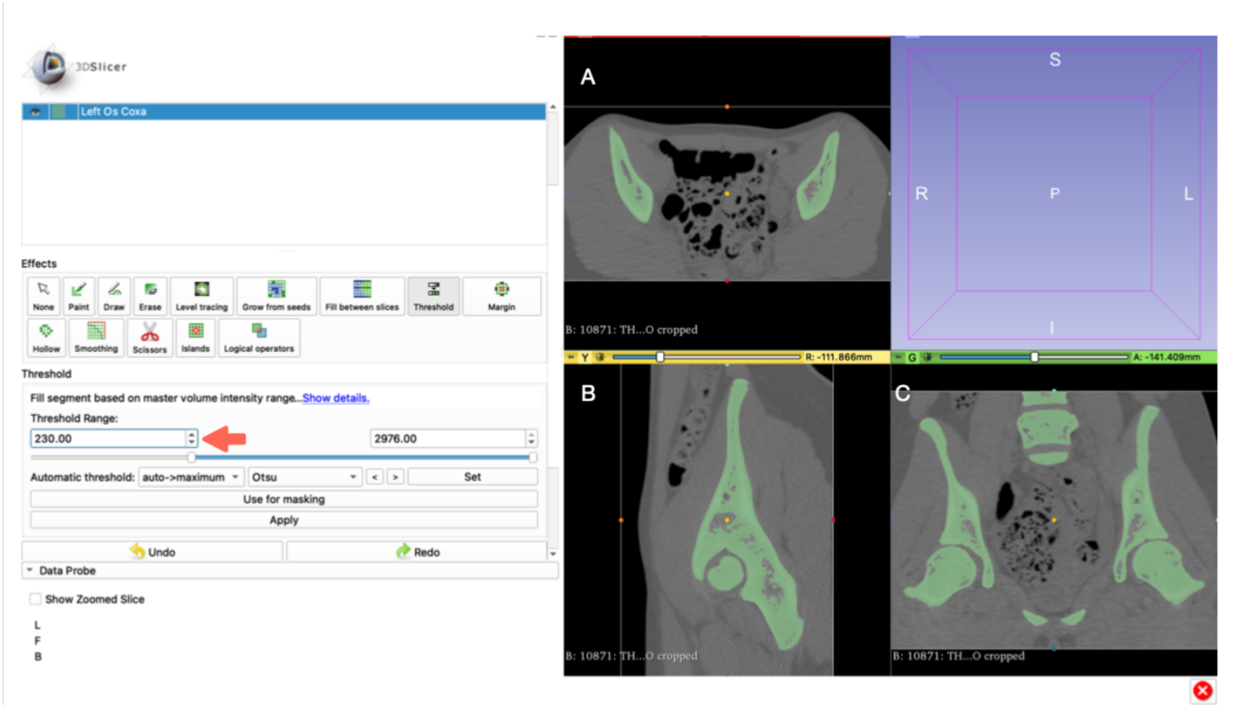

Figure 2.7 *Steps 3 and 4: Setting the threshold value (red arrow) used for masking, and seen highlighting the bone in the transverse (A), sagittal (B), and coronal (C) views. Note the 'Use for masking', allowing for segmentation of right and left sides*

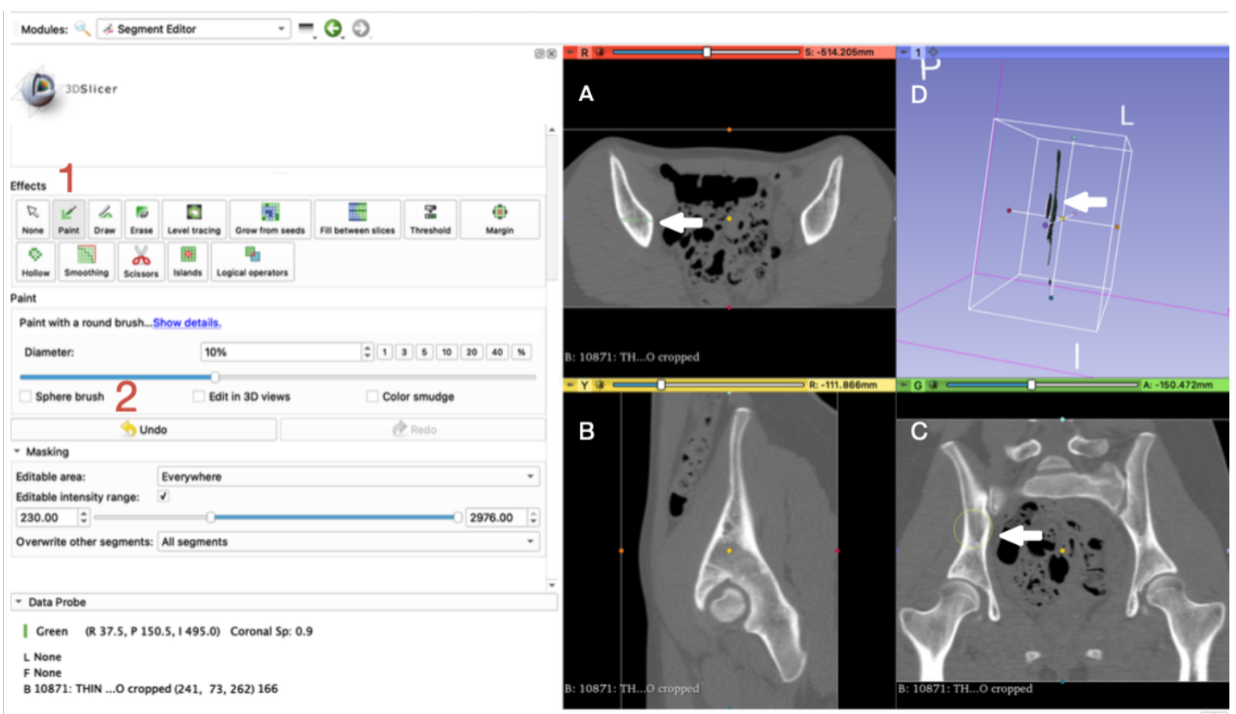

Figure 2.8 *Steps 4 and 5: Mask is applied, area that was previously green can now be segmented using the paint tool. Sphere brush not currently in use, visible in the transverse (A) and 3D (D) views. Compare to figure 2.9*

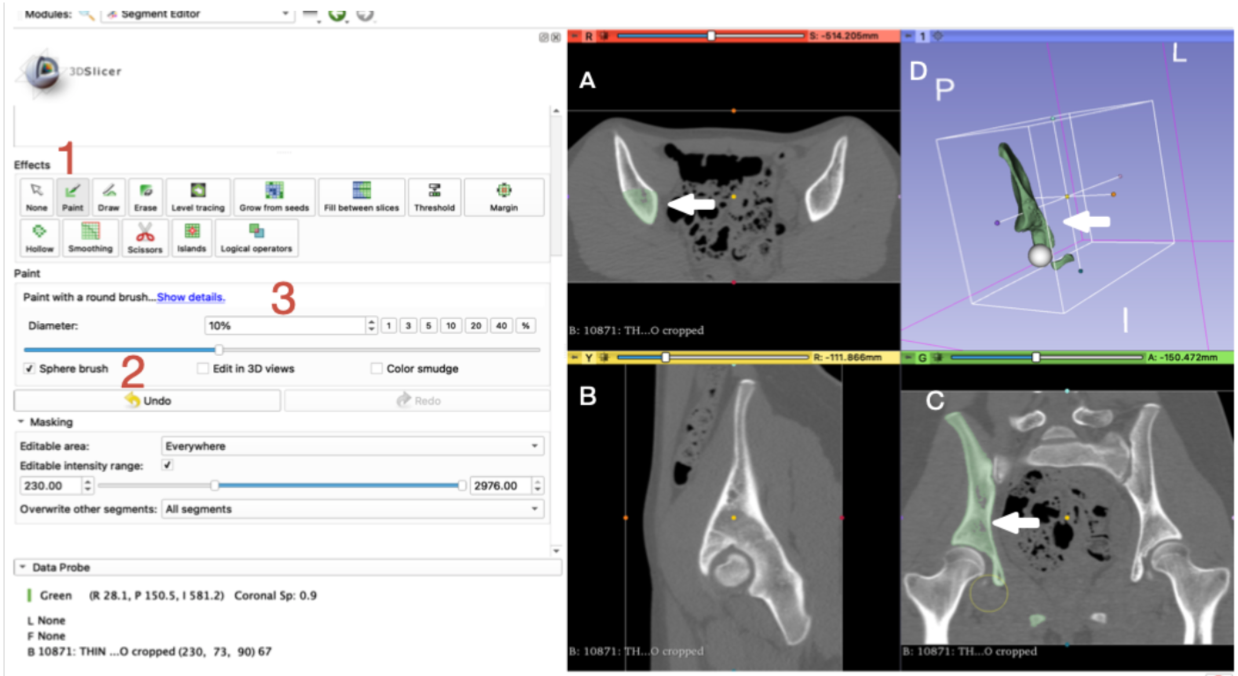

Figure 2.9 *Steps 4 and 5 The sphere brush (2) is used, visible by the penetration of multiple slices in the transverse (A) and 3D (D) views by painting with a 10% sphere brush*

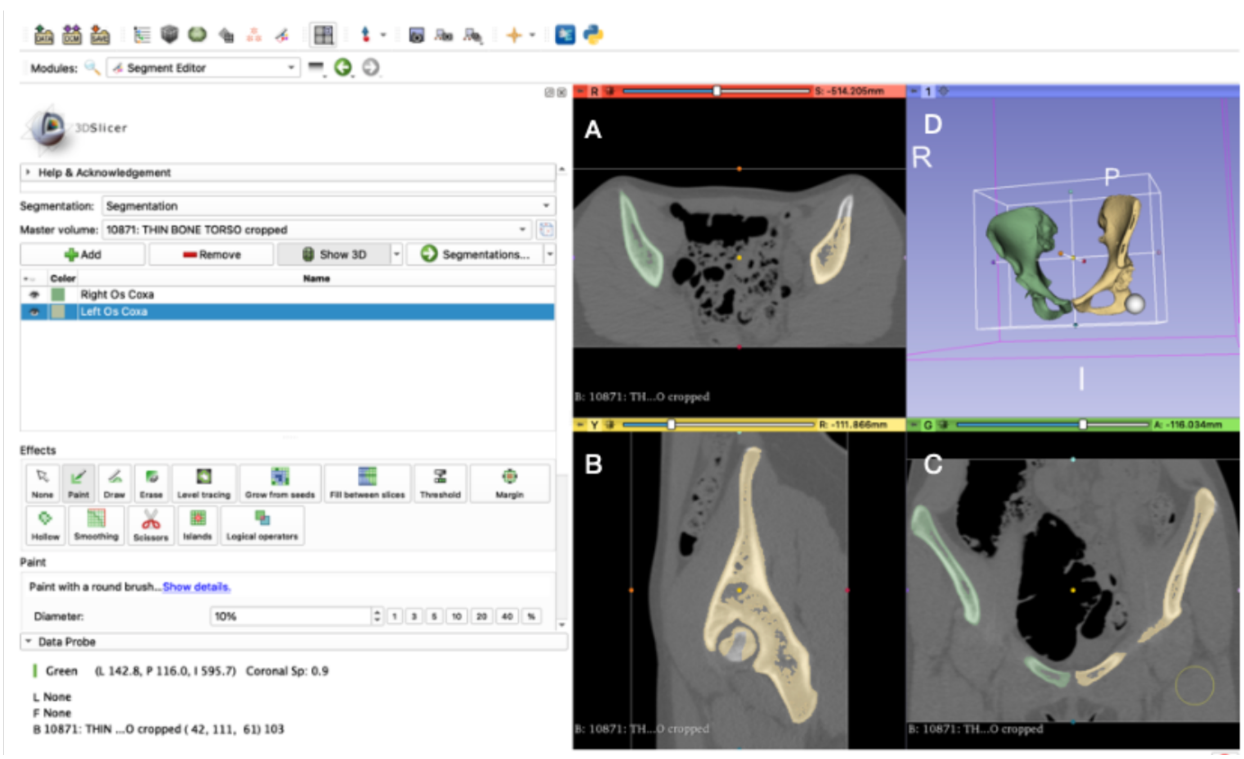

Figure 2.10 *Step 6: the right and left sides of the os coxae are segmented by using separate arbitrary colors, indicating to the program that the models will be discrete rather than a single unit. Models visible in the 3D view (D)*

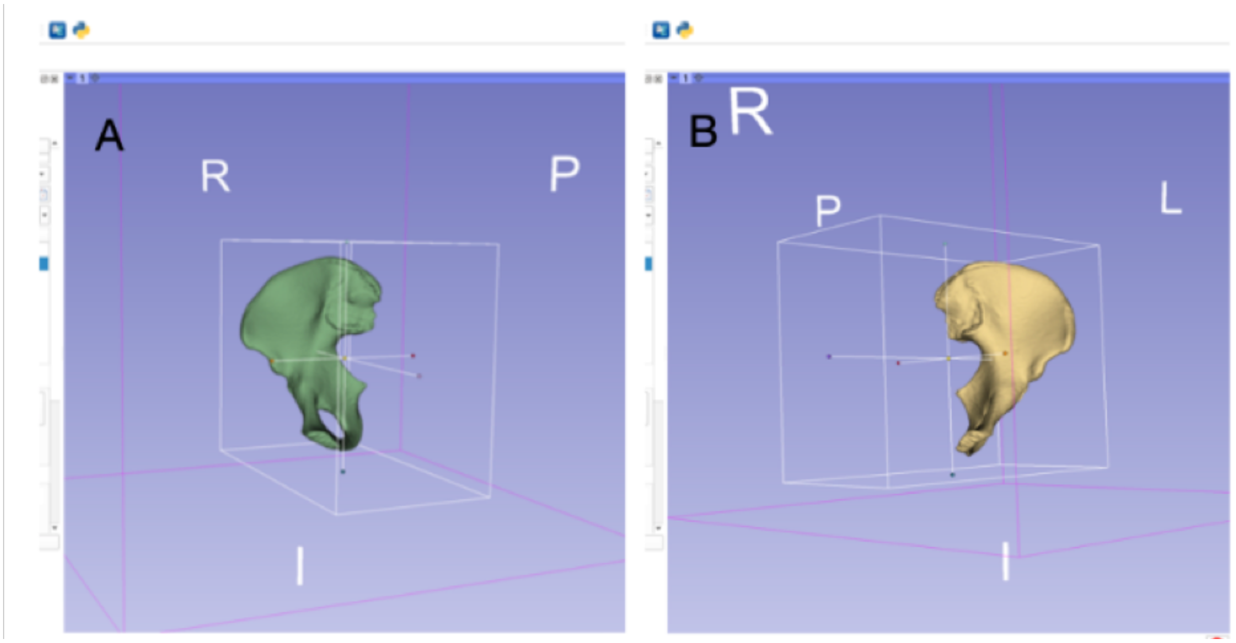

Figure 2.11 *Step 7: The final models of the right (A) and left (B) sides, exported as models, saved and viewed using an stl viewer for scoring.*

Application of Standard Age-at-Death Estimation Methods

Methods Applied

The Suchey-Brooks pubic symphysis method is the most commonly applied age-estimation method among both forensic anthropologists and bioarchaeologists (Falys and Lewis, 2010; Garvin and Passalacqua 2012). Originally a revision of the Todd 1920 method for the pubic symphysis, the Suchey-Brooks method scores individuals into one of six phases with sex-specific 95% confidence intervals and an estimated mean calculated directly from the original reference sample. Written descriptions, casts, and images serve as standardized examples of each phase, allowing an observer to compare a given individual to a 'type specimen' of each phase.

The Hartnett method is a phase method based on Suchey-Brooks, but with an additional phase to accommodate individuals in older age (Hartnett 2010). The method was developed largely in response to major criticisms of the Suchey-Brooks method, including high intra-observer error rates, large age ranges, and population specificity (Hartnett 2010). Using written descriptions provided by Hartnett, an observer can estimate an individual into one of seven phases with associated age ranges with a 95% confidence interval, and an estimated mean. This method is less widely used at this time, largely due to its recent development.

Boldsen et al. (2002) Transition Analysis is unlike the Suchey-Brooks or Hartnett methods in that it is based on components, rather than phases, and combines them using a more sophisticated statistical approach. This method uses five cranial suture segments, five bilateral components of the pubic symphysis, and nine bilateral components of the iliac auricular surface. Each component can be in one of several phases, specific to that particular component. Using the Transition Analysis Manual and accompanying software, each independently scored component is used to generate transition probabilities. The various potential combinations of scores of each component for each individual result in specific estimated maximum likelihood estimates and associated 95% confidence intervals, dependent on the specific combination of component scores.

Although this method includes the cranial sutures, iliac auricular surfaces, and pubic symphyses, only the pubic symphysis components were evaluated in this study. Several studies have found that cranial sutures (both ecto- and encocranial) are not rendered accurately or reliably using CT scans. This is because these are extremely narrow and the typical CT slice overlap does not depict cranial sutures in sufficient detail for them to be scored for any age-estimation method, including transition analysis (Grabherr et al. 2009). Further, the iliac auricular surfaces were not scored in this study due to the lack of observable features in the generated models. There were no components consistently observable in all individuals.

Scoring Procedure

An initial subset of twenty individuals was scored to familiarize the observer with the application of the methods used in this study on CT data. This was done to reduce the error introduced by novice scoring and provide an equal starting point for the application of all methods.

This initial subset was scored using the Suchey-Brooks pubic symphysis method, the Hartnett pubic-symphysis method, and the pubic symphysis component of the Boldsen et al. Transition Analysis method (TA). These individuals were later reincorporated into the final sample and scored randomly without knowledge that they were part of this original training sample.

The sample of 169 individuals (including the reincorporated subset of 20 individuals) were put in a randomized order using excel. The sex of each individual was documented using the metadata from the NMDID. Each individual maintained their assigned DID number. All individuals were scored using the written description, casts, and images provided by and associated with the 1990 Suchey-Brooks methodology. The entire sample was randomized a second time one week after the completion of the Suchey-Brooks scoring, and scored using the written descriptions provided by Hartnett (2010). The time delay was again an attempt to reduce observer error by minimizing the influence of previous scoring sessions using another method. A week after completing the Hartnett scoring session, the sample was randomized a third and final time, and scored using the pubic symphysis component of the TA manual, accompanying photographs, and the necessary software (Boldsen et al. 2016).

After the sample was scored using all three age estimation methods, the reported age at death was collected from the NMDID metadata. If the documented range fell between the upper and lower ends of the age range estimated by each method, the individual was considered to be scored accurately. Bias was calculated using the phase specific reported mean by each method and the documented age, the absolute value of the difference was reported as error. The average error, bias, and percent accuracy for each decade of life for each method was then calculated.

Quantitative Analysis of Data

After the sample was scored with all three applied age estimation methods, the metadata collected from the NMDID was reviewed. The age range estimated were compared to the reported age documented in the NMDID. If the reported age was included in the estimated age range, the individual was considered 'accurately' estimated. The mean for each phase of the Suchey-Brooks and Hartnett methods was used as a point estimate of age, while the maximum likelihood was used as a point estimated for TA- as is standard practice in the field. This was compared to the reported age of each individual, and any difference between the two determined the degree of bias. Analysis was reported in age cohorts (e.g., 15-19, 20-29, 30-39, etc.) in order to facilitate comparison of method accuracy and bias for each age group, and to better observe trends in the data. In addition to analyzing each age cohort, sex specific differences in method performance were reported as well, as this was a common aspect of analysis in the literature, and sex was a factor in determining the estimated age range and mean for each applied method.

It is important to note that the sample used in this study is relatively small. This small sample size precludes assessing statistical significance in terms of methodological comparisons to each other, or to the aforementioned validation studies. However, accuracy and bias are still reported in relation to the performance of each of these methods applied to the NMDID sample used in this study, and to make comparisons with previously conducted studies, as these evaluations are still helpful, though their statistical significance is beyond the scope of this study.

Intra-observer Reliability

To test intra-observer reliability, sixty-one (61) individuals were selected at random from the larger, final sample to test intra-observer reliability. These individuals were tested using only the Hartnett method due to time constraints. A phase-based method was chosen over TA because of the time-consuming nature of this particular component based method. Hartnett was chosen over Suchey-Brooks to test intraobserver error because of the extra phase in the Hartnett method. Additionally, as that method was adapted from the Suchey-Brooks descriptions, there is no reason to believe that intra-observer error rates would different substantially between these two methods. Individuals scored as a Phase 6 using the Suchey-Brooks method would have likely been scored as a Phase 6 or Phase 7 using the Hartnett method, enabling the comparison of the initial scores from the Suchey-Brooks and Hartnett methods to a second round of scoring using just the Hartnett method. The two rounds of scoring were then compared to determine the frequency of agreement by the observer.

Comparison of Method Performance- dry bone and digital bone

The results of the phase-based methods were compared to other validation studies that used a similar sample. For the Suchey-Brooks method, a study conducted by Savall et al. (2018) was used for comparison. This study examined the performance of the Suchey-Brooks method when applied to a digital sample of os coxae reconstructed from multi-slice computed tomography collected from several French hospitals, to the "Suchey-Brooks original sample" (Savall et al. 2018). The accuracy and bias of the Suchey-Brooks method when applied to the French and Suchey-Brooks samples were compared to the accuracy and bias results when applied to the reconstructed NMDID sample.

The Hartnett method is less widely evaluated, as it is a newer method and not as commonly applied as the Suchey-Brooks method. Validation studies for the Hartnett method are more difficult to find, no validation studies comparing a digital and dry bone sample were found for comparison. However, one study comparing the Suchey-Brooks and Hartnett methods for the pubic symphysis and the Isçan (1984) and Hartnett (2010) methods for the sternal rib ends was found and used in this study for comparison. Because sternal rib ends were not used at all in this study, only the data for the pubic symphyses methods were used. This study was conducted by Catherine Merritt in 2014 using a dry bone sample of 322 individuals from the Bass Donated Collection. This study analyzed the average bias and accuracy in totality rather than by decade. Because of this, only the total accuracy rate of the Hartnett method when applied to the NMDID sample and the total average bias for all age cohorts was used for comparing method performance.

In *Transition Analysis: A Validation Study with Known-Age Modern American Skeletons* Boldsen and Milner (2012) use a sample of 239 individuals for this validation study, primarily sourced from the Bass Donated Collection- all identified as white individuals. This validation study is composed of multiple analyses, first looking at the pubic symphysis, iliac auricular surface, and cranial sutures and discrete components. The estimated age range generated using only the features of that particular component is charted with reported age, demonstrating the bias and accuracy of each component for each individual. The accuracy and bias of the method when all three components (the pubic symphyses, iliac auricular surfaces, and cranial sutures) for all individuals in the validation sample are then charted together, in order to demonstrate the increase accuracy and decrease bias when all three features are used in combination. Only the data for the pubic symphysis component was used for comparison, as this was the only component scored using the NMDID sample.

While the reported rates of accuracy and the magnitude and direction of bias for these validation studies were compared to those collected in this study, these comparisons are not statistically significant as the sample in this study is relatively small. However, these evaluations were still useful and reveal trends in methods applied to digital bone that are similar to those observed in dry bone, as discussed in Chapter 4.

Chapter 3: Results

After all individuals determined to be suitable for the final sample were reconstructed in 3D slicer and saved as a model, the entire data set was randomized a total of three times prior to the application of each age estimation method used in this study─Boldsen et al. (2002) Transition Analysis (TA), the Hartnett method for the pubic symphysis (Harnett 2010), and the Suchey-Brooks method for the pubic symphysis (Brooks and Suchey 1990).

Accuracy

In this study, method accuracy was assessed based on the number of individuals whose reported age fell within the estimated age range determined by their phase or component score. Several trends in accuracy were common to all three methods were observable and are summarized in Figure 3.1 and Table 3.1. All three methods performed the best when scoring individuals under the age of fifty years. All three methods scored the most individuals accurately between the ages of thirty and thirty-nine. There was a decline in the rate of accuracy for all three methods from after the age of forty-nine, and a second more dramatic drop in the rate of accuracy after the age of 59 for both phase-based methods (i.e., Suchey-Brooks and Hartnett).

In the youngest age cohort, 15-19 years, the Suchey-Brooks method performed the best overall, estimating 91% of individuals (10/11) accurately. This was followed by the Hartnett method, scoring 72% of individuals (8/11) accurately. TA was the least accurate for this age cohort, scoring 64% of individuals (7/11) accurately.

Table 3.1 ***Accuracy of each of the methods (n, %) applied to the same sample of individuals.*** *The total number of individuals included in each age cohort and total number of individuals accurately estimated using each method.*

Test Sample		Method Accuracy					
		Suchey-Brooks		Hartnett		TA	
Age Group (Years)	N	n	%	n	%	n	%
15-19	11	10	90.9	8	72.7	7	63.6
20-29	22	19	86.4	14	63.6	19	86.4
30-39	22	21	95.5	20	90.9	19	86.4
40-49	22	20	90.9	18	81.8	15	68.2
50-59	23	17	73.9	17	73.9	12	52.5
60-69	19	7	36.8	5	26.3	9	47.4
70-79	19	5	26.3	5	26.3	7	36.8
80-89	18	4	22.2	2	11.1	7	38.9
90 and above	13	0	0	2	15.4	4	30.8
Total	169	103	60.9	91	53.8	99	58.6

Compared to the youngest individuals (15-19 years), both phase-based methods were less accurate overall for the 20-29 age cohort, while TA was more accurate. Suchey-Brooks and Hartnett aged 86% (19/22) and 64% (14/22) of individuals accurately, respectively, while TA estimated 86% (19/22) of individuals accurately.

All three methods were the most accurate for the 30-39-year age cohort, with both phase-based methods performing better than TA. Using the Suchey-Brooks method, 95% (21/22) individuals were estimated accurately, while 91% (20/22) individuals were estimated accurately using Hartnett. TA was equally accurate for the 30-39 year-old age cohort (86%, 19/22) as for individuals in the previous decade.

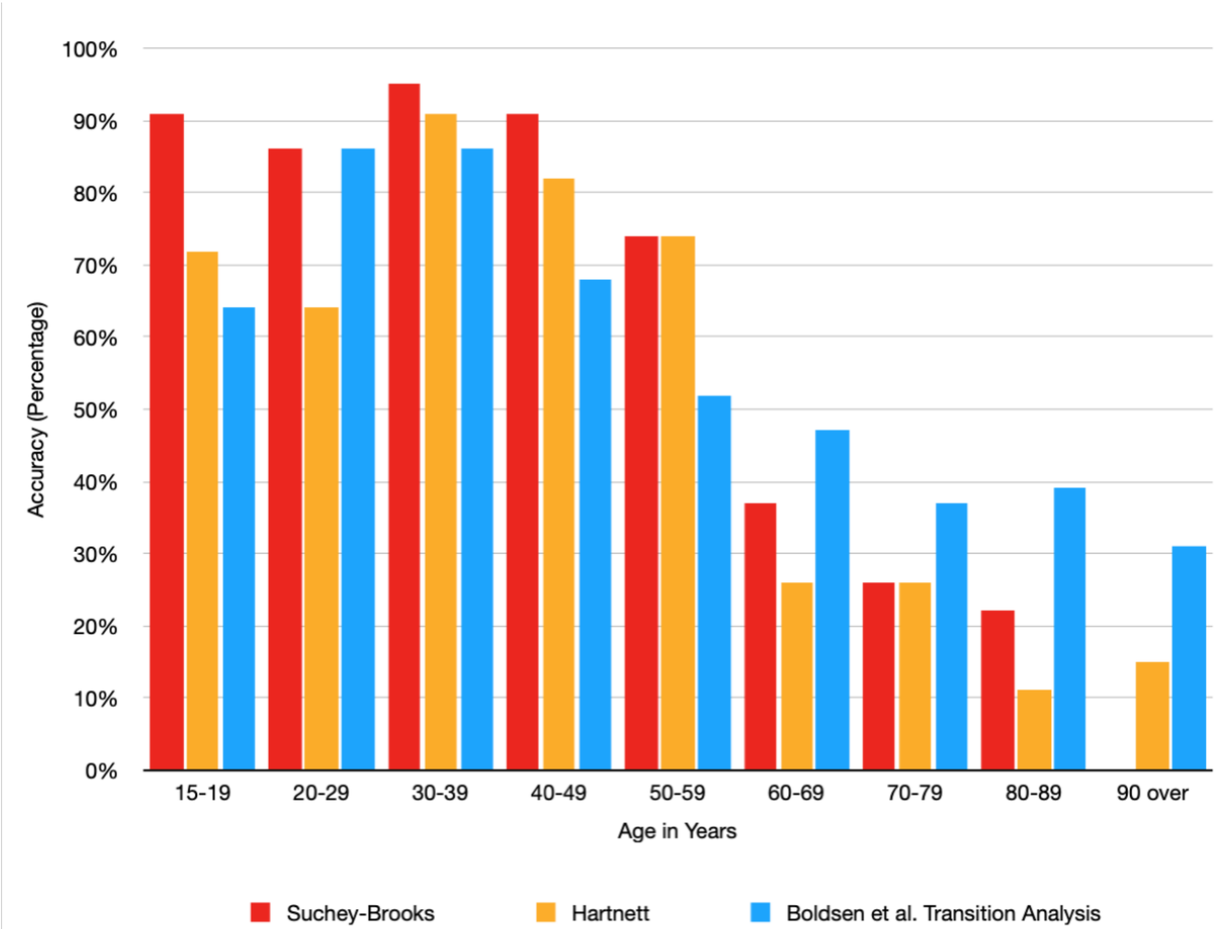

Figure 3.1 A comparison of the overall accuracy of each method *with accuracy as a percentage (y-axis) per age cohort (x-axis)*

Both the Suchey-Brooks and Hartnett methods maintained a relatively high degree of accuracy for individuals in their 40s, estimating 91% (20/22) and 82% (18/22) of individuals accurately, respectively. TA, on the other hand, estimated significantly fewer individuals accurately when compared to younger age cohorts. 68% (15/22) individuals' reported ages fell within the interval estimated by the method.

In general, beginning in the 40-49 year age category, the accuracy of all methods began to decline. While the decline was not as noticeable between the 40-49 and 50-59 age cohorts, there was a stark difference between the 50-59 and 60-69 age categories.The Suchey-Brooks method estimated 74% of individuals (17/23) accurately in their 50s, but only 37% of individuals (7/19) in their 60s. The Hartnett method was equally as dramatic in this decline, estimating 74% of individuals (17/23) accurately in there 50s, compared to 26% (5/19) in their 60s. TA declined as well, though less significantly when compared to the phase-based methods. Using TA 52% of individuals (12/23) in their 50s were estimated accurately, while 47% (9/19) were estimated accurately in their 60s. This trend is observed in validation studies of these methods performed on dry bone (Merritt 2014, Savall et al. 2018, Milner and Boldsen 2012), and it was expected to carry over to digital bone model to at least the same degree.

TA essentially maintained this relatively low rate of accuracy with minimal fluctuation for the remaining decades included in this sample, accurately estimating individuals in their 70s, 80s, and 90s at rates of 37% (7/19), 39% (7/18), and 31% (4/13) respectively. Both phased based methods continued to be negatively correlated with age, estimating fewer adults accurately in each progressive age cohort. The Suchey-Brooks method estimated 28% (5/19) of individuals in their 70s accurately, 22% (4/18) of individuals in their 80s, and 0% (0/13) of individuals age 90 and over. Using the Hartnett method, 26% (5/19) of seventy-year-olds were estimated accurately. 11% (2/18) of eighty-year-olds were estimated accurately, and lastly 15% (2/13) of individuals aged 90 and over were estimated accurately (Figure 3.1, Table 3.1).

Bias

Bias in age estimates produced by a method is structured directional error. If reported (i.e., known age) falls below the estimated mean or maximum likelihood, the individual has been overestimated, while the inverse, reported age falls above estimated age, means that that individual's age has been underestimated. Bias for each individual was calculated by subtracting their known age from the central tendency of age estimated by each method (i.e., bias = estimated age – known age). Average bias was then calculated for individuals in each age decade-based age cohort [average bias = Σ(estimated age – documented age) /cohort sample size]. The trends in bias for each method are displayed in Figure 3.2 and Table 3.2

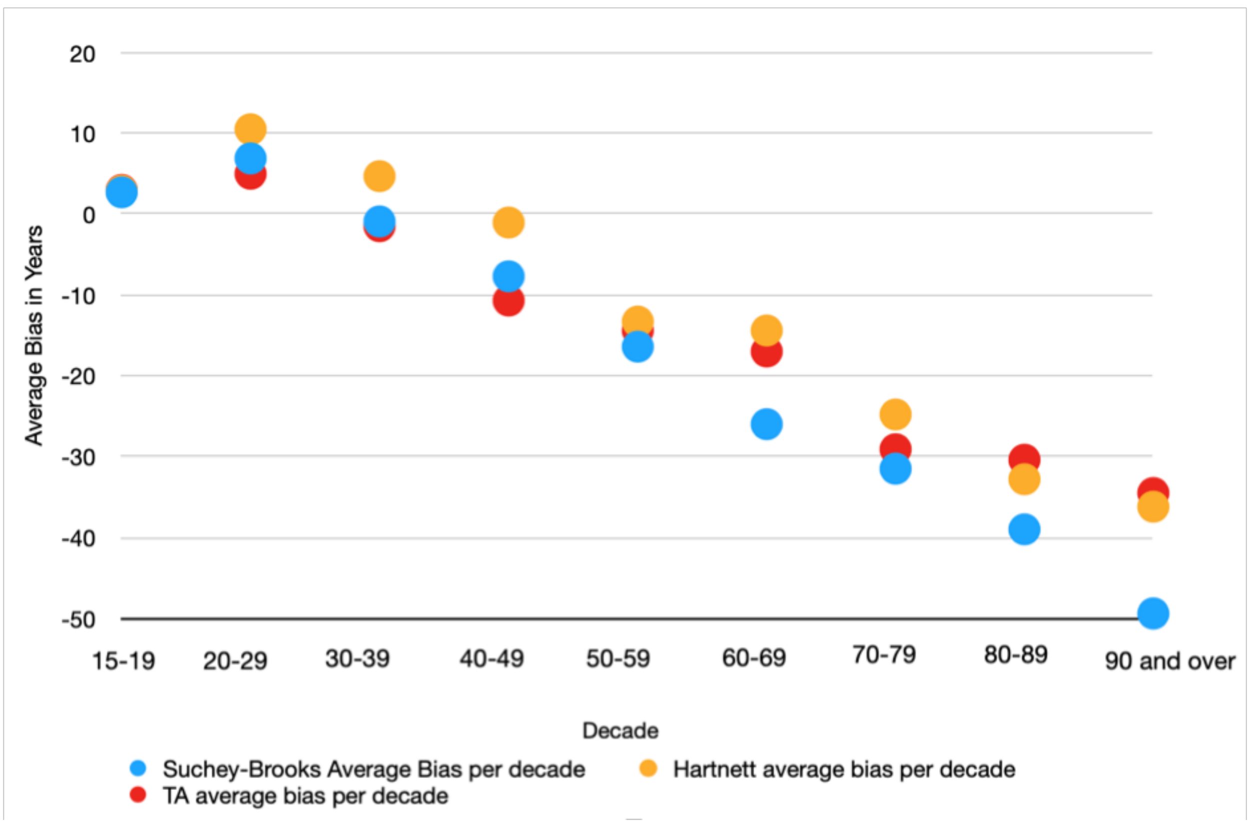

***Figure 3.2 A comparison of the average bias per decade of life for each applied method** according to age cohort (x-axis) and displayed as an integer (y-axis) reflecting the average direction and magnitude of bias for the given age cohort.*

Table 3.2 Bias of each method (n, Bias (years)) applied to the same sample of individuals*. The total number of individuals included in each age cohort and average bias for that age cohort.*

Test Sample		**Method Bias**		
		Suchey-Brooks	**Hartnett**	**TA**
Age Group (Years)	**N**	**Bias** (years)	**Bias** (years)	**Bias** (years)
15-19	11	2.7	2.9	3.0
20-29	22	6.9	10.5	5.0
30-39	22	-0.9	4.7	-1.5
40-49	22	-7.7	-1.04	-10.7
50-59	23	-16.4	-13.3	-14.4
60-69	19	-26.0	-14.4	-17.0
70-79	19	-31.5	-24.8	-29.1
80-89	18	-39.0	-32.8	-30.4
90 and above	13	-49.4	-36.2	-34.5
Total	169	-17.9	-11.6	-14.4

Accuracy of Point Estimates

Another way of assessing the performance of each method, is by comparing the accuracy of the point estimates (i.e., mean or maximum likelihood estimate) or age generated for each method compared to the known age of each individual (i.e., the distance of each point estimate from known age). For forensic cases, the accuracy of the associated interval is of highest importance and is what must be reported in the case report. However, trends in the accuracy of point estimates can provide a clearer picture of the trends of under- and over-estimation of age with each method. Assessment of accuracy in this way is also important to understand the possible ramifications of using each technique for paleodemographic work, where point estimates, rather than estimated intervals, are typically used.

Figures 3.3-3.5, compare the estimated mean or maximum likelihood (y-axis) to the reported age documented in the metadata from the NMDID (x-axis). In each figure, the identity line – the line of perfect agreement between estimated and documented age – is shown to facilitate comparison [i.e., . if the reported and estimated ages had been equal for each individual, all points would fall along the identity line (slope =1)].

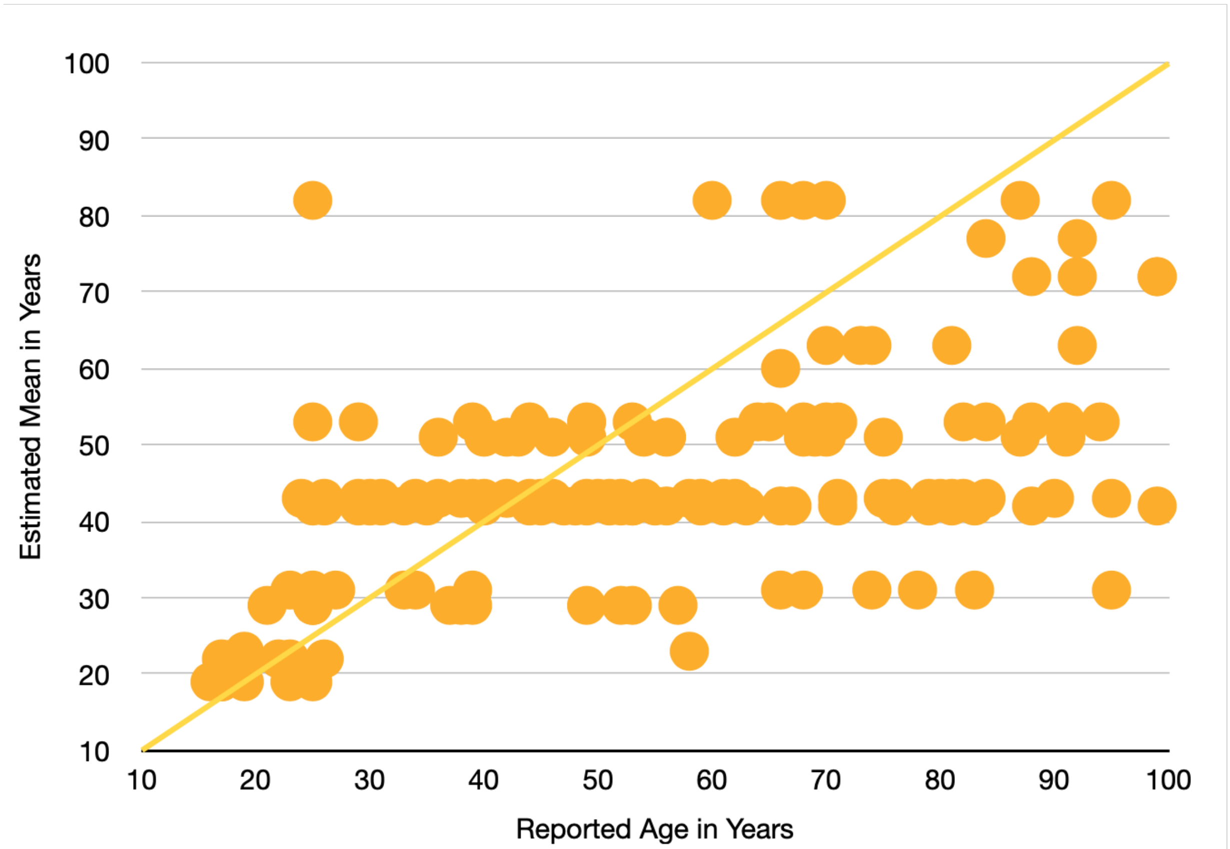

Figure 3.3 The reported age of each individual (x-axis) compared to the estimated age (y-axis) using the Hartnett method. *Identity line (slope=1) demonstrates where points would lie if there were perfect agreement between these data.*

For all three methods, there was a general tendency to overestimate younger individuals (i.e., those under 50 years), and to underestimate older individuals (i.e., those over the age of fifty). This trend was generally expected, as it is observable in all three methods when applied to dry bone as well (Merritt 2010, Savall et al. 2018, Milner and Boldsen 2012). The degree to which individuals were under-or over- estimated increases with age as well, a trend observed in the results of all three methods.

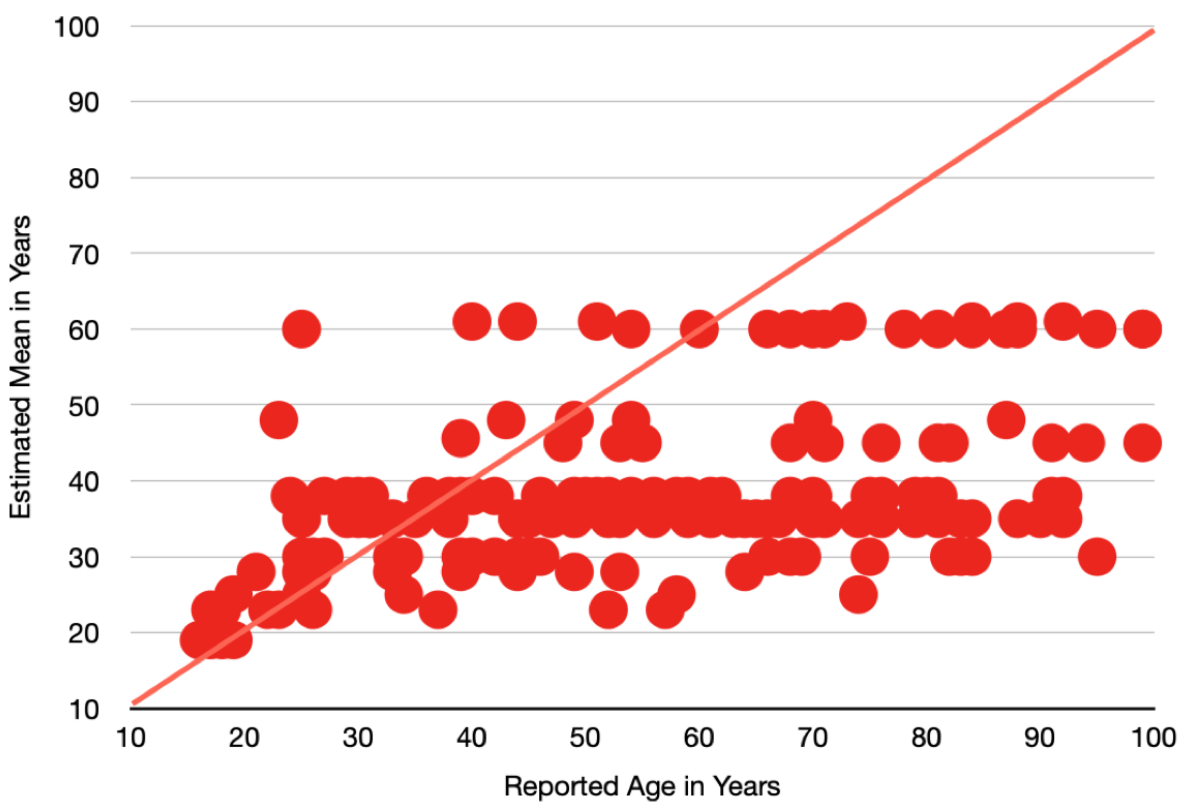

Figure 3.4 The reported age of each individual (x-axis) compared to the estimated mean (y-axis) using the Suchey0Brooks method*. Identity line (slope=1) demonstrates where points would lie if there were perfect agreement between these data.*

For the 15-19 age cohort, Suchey-Brooks, Hartnett, and TA overestimated individuals by an average of 2.7. 2.9, and 3.0 years respectively. No individuals in this age cohort were underestimated. This trend held true for the 20-29 age cohort, with average overestimates of 6.8, 10.5, and 5.0 years, respectively.

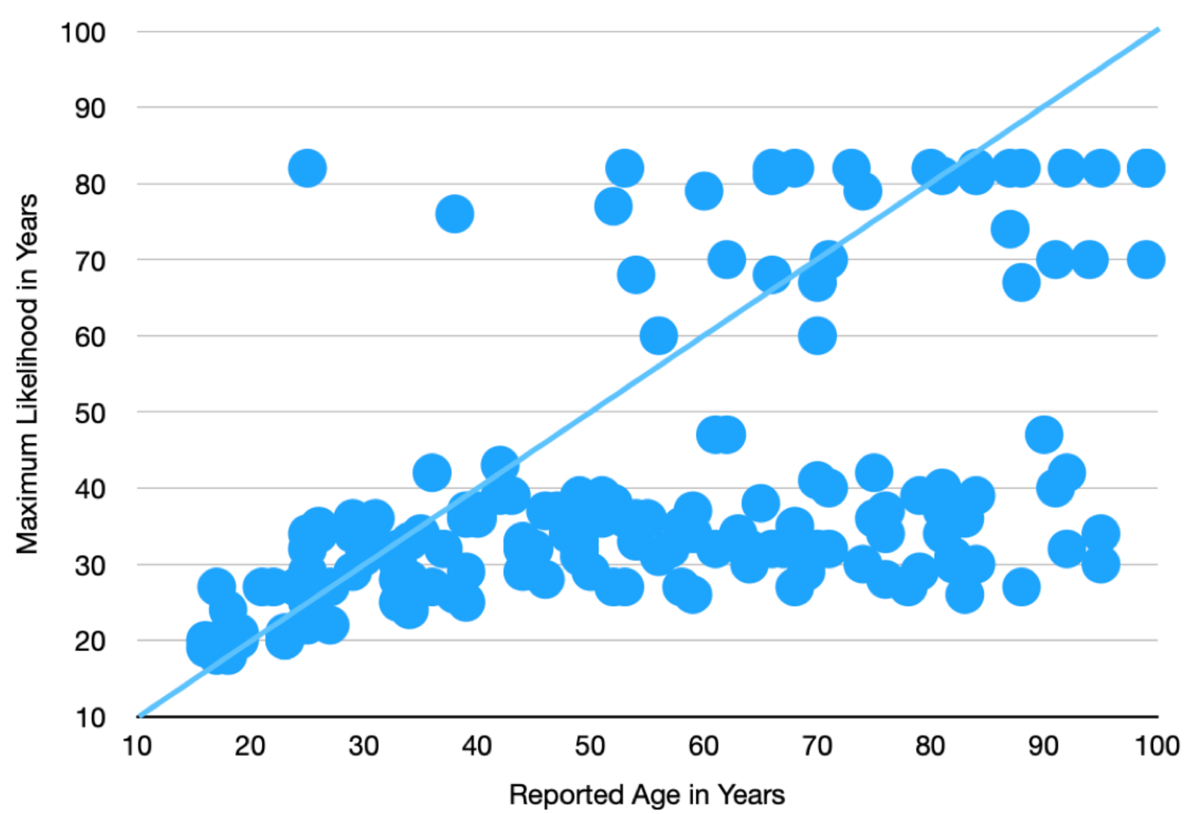

Figure 3.5 The reported age of each individual (x-axis) compared to the estimated maximum likelihood (y-axis) using TA*. Identity line (slope=1) demonstrates where the points would lie if there were perfect agreement between these data.*

There was slightly more variation in the 30-39 age cohort, though the degree of over- or underestimation remained small. Suchey-Brooks and TA both underestimated by -0.9 and -1.5 years, respectively Hartnett overestimated by an average of 4.7 years.

Beginning at age 40, the general trend for all methods was to underestimate, which is consistent with published studies of these methods. While a handful of individuals in this age cohort were overestimated by each method, the overall tendency was to underestimate, with averages of -7.7, -10.7, and -1.2 years, respectively.

The degree of underestimation increased significantly for the 50-59 age cohort for all three methods. Suchey-Brooks, Hartnett, and TA underestimated by -16.4, 13.3, and -14.4 years respectively. This trend of overall underestimation continued for all subsequent age cohorts.

Hartnett underestimated the least for the 60-69 and 70-79 year olds with an average of -14.4 and -24.6, respectively. TA underestimated individuals in the 60-69 age cohort by -17.0 years, and the 70-79 year olds by –29.1 years. Suchey-Brooks underestimated these two groups by the most, -26.0 for the individuals in their 60s, and -31.5 years for those in their 70s.

Suchey-Brooks continued to underestimate by the most on average for those above 80 years of age, with an average of -39.0 years for those in their 80s -49.4 years for those in their 90s. The Hartnett method similarly underestimated all individuals in these two age cohorts by -32.8 years and -36.2 years, respectively. While TA still underestimated individuals in the 80-89 and 90+ age cohorts by a significant amount, it was slightly less compared to the other two, averaging -30.4 years and -35.5 years respectively.

Sex-Specific Differences in Accuracy and Bias

Sex-specific differences were noticed in all three age estimation methods. For both TA and Suchey-Brooks, females were estimated accurately at a slightly higher frequency than males with less bias when averaged across all age categories. TA accurately estimated 64% (53/83) of females accurately with an overall average bias of -11.8 years, while males were estimated accurately 53% (46/86) of the time with an average bias of -15.6 years (Table 3.3).

Using the Suchey-Brooks methods, 69% (57/83) of females were estimated accurately and underestimated by an average of -15.8 years, while 53% (46/86) of males were estimated accurately with an average bias of -17.4 years.

The Hartnett method estimated more males than females accurately. 57% (49/86) of males' reported age fell into their estimated age range, with an average bias of -10.8 years. Females were estimated accurately at a rate of 49% (41/86) and underestimated by -9.6 years on average.

Table 3.3 Accuracy and bias of each of the applied methods (Bias (years), %) applied to the same sample of males and females. *The number of individuals estimated accurately, and cumulative bias for each sex using each method.*

Test Sample and Sex		Method Accuracy and Bias					
		Suchey-Brooks		Hartnett		TA	
Females	N	Bias (years)	%	Bias (years)	%	Bias (years)	%
	83	-15.8	68.7	-9.6	49.4	-11.8	63.8
Males	N						
	86	-17.4	53.4	-10.8	57.0	-15.6	53.4

Intra-observer Reliability

A randomly selected sample of 61 individuals was scored a second time using the Hartnett method approximately two weeks after the initial round of scoring, and compared to their original Hartnett score. The degree of intra-observer agreement was determined by the number of individuals scored in the same phase both rounds of scoring.

In comparing the initial round of scoring using the Hartnett method to the second round of Hartnett scoring, there was perfect intraobserver agreement on 66% (40/61) of individuals. Of the 21 individuals where the scores disagreed between the two rounds of scoring, 95.2% (20/21) individuals were scored within one phase of the original score.

A total of eight individuals were scored in a more advanced phase in the second round of scoring (i.e. scored as a Phase 3 in the initial round, compared to a phase 4 in the second round). Twelve individuals were scored as a more advanced phase in the initial round of scoring (i.e. scored as a Phase 4 in the initial round, compared to a Phase 3 in the second round).

Comparison of Method Performance

Published studies were used for comparative data regarding accuracy and bias of all three methods. In addition to relying of samples of varying size and composition, each study reports this data in a slightly different way, making direct comparison of accuracy and bias rates difficult. Because of this, comparisons of method performance focus on general reported trends for each of the three methods and compare rates of bias and accuracy when possible.

The Suchey-Brooks method is consistently reported to be population-specific, particularly for later phases (Djuric et al. 2007). This has resulted in varying reports of accuracy and bias, dependent on the sample population for a given validation study (e.g. when applied to Balkan populations, the Suchey-Brooks method classified 89.74% of males and 72.0% of females accurately [Djuric et al. 2007] When applied to a modern population in Queensland, Australia, the Suchey-Brooks accurately estimated 63.9 and 69.% of male and females, respectively [Lottering et al. 2013].) Comparison of the performance of the Suchey-Brooks method in this study was difficult because of the abundance of other validation studies, all with differing reports of accuracy and bias. However, a study conducted by Savall et al. (2018) compared data from the collected using the dry bone sample originally used for the Suchey-Brooks method and data collected by applying the Suchey-Books method to a digital sample, reconstructed using CT scans from two French Hospitals. Data collected from this study is compared to both the dry bone data and data from the French sample. According to the Savall study, when applied to a dry bone sample Suchey-Brooks accurately estimated 62.3% of males and 45.4% of females accurately. When applied to a digital French sample, Suchey-Brooks accurately estimated 60.4% of males and 68.2% of females accurately. When applied to the digital sample reconstructed from the NMDID, this method estimated 69% of females and 53% of males accurately. In all three applications (i.e., the two from the published study and the application in this research), individuals were scored accurately more frequently below the age of 40 and tended to be overestimated below the age of 40. After the age of 40 the rate of accuracy decreased dramatically, and individuals tended to be underestimated.

The Hartnett method has not been evaluated by as many publications as Suchey-Brooks, as it is not yet as widely applied in bioarchaeological or forensic settings. In a validation study of the Hartnett method by Merritt (2014), the method was tested for accuracy rates as a percentage on a skeletal sample of 322 individuals from the William M. Bass Donated Collection – a modern collection, primarily composed of donated adults from the regions surrounding Tennessee. Using the written descriptions provided by Hartnett for the pubic symphysis, individuals were scored into one of the seven phases. Individuals estimated with an age range that included their reported age were considered to be scored accurately. 58% (187/322) of individuals were estimated accurately using the Hartnett method in this validation study (Merritt 2014). In this study, the Hartnett method performed similarly, estimating 53.8% (91/169) individuals accurately using the same definition 'accuracy'. The Merritt study reported that generally, individuals below the reported age of 40 were estimated accurately more often than individuals above the age of 40. Trends in bias were also mentioned in comparison to the Suchey-Brooks method, which was also applied in the Merritt study. While the Hartnett method was less accurate overall, it demonstrated less bias compared to the Suchey-Brooks method. When applied to the sample of digital models from the NMDID, the Hartnett method was overall most accurate for individuals aged 40 and younger. Compared to the Suchey-Brooks method, there was less cumulative bias when comparing the average bias for all included age cohorts.

Comparisons for Boldsen et al. Transition Analysis (2002) were made using only data for the pubic symphysis from a validation study conducted by several of the authors of the original method (Milner and Boldsen 2012).

Milner and Boldsen (2012) published the accuracy and precision for each of the three skeletal regions included in TA (cranial sutures, iliac auricular surfaces, and pubic symphyses) were discussed separately, and in combination. Several different samples were used for comparison. The sample collected from the William M. Bass Donated Collection was most similar to the sample used in this study, as it was composed of 246 self-identified white individuals. The data collected from this sample was used for comparison. Since the pubic symphysis was the only feature used in this study, this was the only data used for this aspect of analysis. Of the 246 individuals used in this aspect of the Boldsen et al. (2002) study, 87.8% (216/246) were scored with an estimated age range that included their reported age using only the pubic symphysis (data collected from the graphs published by Boldsen et al. [2002]). Up to the reported age of 40, TA was remarkably accurate, partially attributable to the wide age range estimated generated using only the pubic symphysis. Individuals over the age of 40 tended to be underestimated, and the rate of inaccuracy increased, particularly after the age of 50.

The rate of accuracy for TA when applied to 3D models was 57% (99/169). However, the same general trend of increased accuracy and overestimation prior to the age of 40, and decreased accuracy and underestimation post-40 years upheld.

Overall, Suchey-Brooks overall was the most accurate when comparing the cumulative accuracy rate, estimating 60% (101/169) with an age range that included the reported age. TA was the second most accurate method, while Hartnett was the least accurate. In terms of bias, Hartnett performed with the least amount of bias on average, displayed in Figure Y (figure comparing all three methods in each age cohort with different colored dots). Suchey-Brooks tended to estimate with the most bias, while TA generally fell somewhere in the middle for most age cohorts.

Chapter 4: Discussion

Accuracy and Bias by Age Cohort

15-19 Years

For the youngest age cohort of the sample (15-19 years), Suchey-Brooks was the most accurate (10/11, 91%), followed by Hartnett (8/11, 72%), and then TA (7/11, 64%). Overall, the bias for this age cohort was relatively minimal. The average bias for this age cohort using the Suchey-Brooks, Hartnett, and TA methods was 2.7. 2.9, and 3.0 years, respectively. A similar trend – a relatively small, positive average bias for this cohort and for individuals under the age of 40 years –is also observed when these methods are applied to dry bone as well (Milner and Boldsen 2012, Savall et al. 2018, Merritt 2010). Further, when the data for the individuals within this age group are evaluated individually (i.e., the bias in the age estimates is diaggregated) all but 4 individuals belonging to this age cohort were slightly overestimated by a small magnitude using all three methods. For these four individuals- one individual using TA, two individuals using Suchey-Brooks, and one individual using Hartnett- was equal to the maximum likelihood or mean used as a point estimate.

The Hartnett and TA methods, however, distinguish between billow depths, with sharper, deeper billows suggesting a very young individual (teens or early 20s), and more rounded, shallower billows suggesting a slightly older individual (mid 20s-mid 30s).Overall features indicative of a younger age–billows, a lack of or an early superior protuberance, and a lack of ventral rampart formation or filling in of billows–are relatively easy to observe in reconstructions resulting in fairly high accuracy rates compared to later decades of life.

Determining the difference between these two billow types proved difficult in reconstruction evaluation, as this is often a tactile evaluation using dry bone (i.e, in other words, billows are both seen and felt). Further, the degree of overlap of the CT slices collected by the NMDID is 0.5 mm. This degree of overlap may not necessarily capture the difference between a sharp and rounded billow that would be observable in dry bone. Additionally, even if the shape of the billows is captured in the raw data, the overlap between adjacent scans used for reconstruction can alter the contour and size of the features. Ultimately this difficulty in distinguishing between billow height and contour resulted in the overestimation of some individuals using the TA and Hartnett methods because of the difference in phase and component scores associated with shallower and more rounded billows. However, a sharp or rounded billow for this age cohort is typical anatomical variation. While the overestimation may have been attributable to the nature of CT scan collection and reconstruction, this cannot be determined definitively without a study directly comparing the dry bone and CT scans of the same individuals, which was beyond the scope of this study.

The TA and Hartnett methods also use bone quality (texture) to distinguish between phases or component scored indicative of slightly different age cohorts. Individuals in their teens or early 20s typically exhibit fine-grained or smooth bone, while slightly older individuals in their mid-20s to potentially early 40s would often exhibit coarser bone transitioning into a surface with microporosity present.

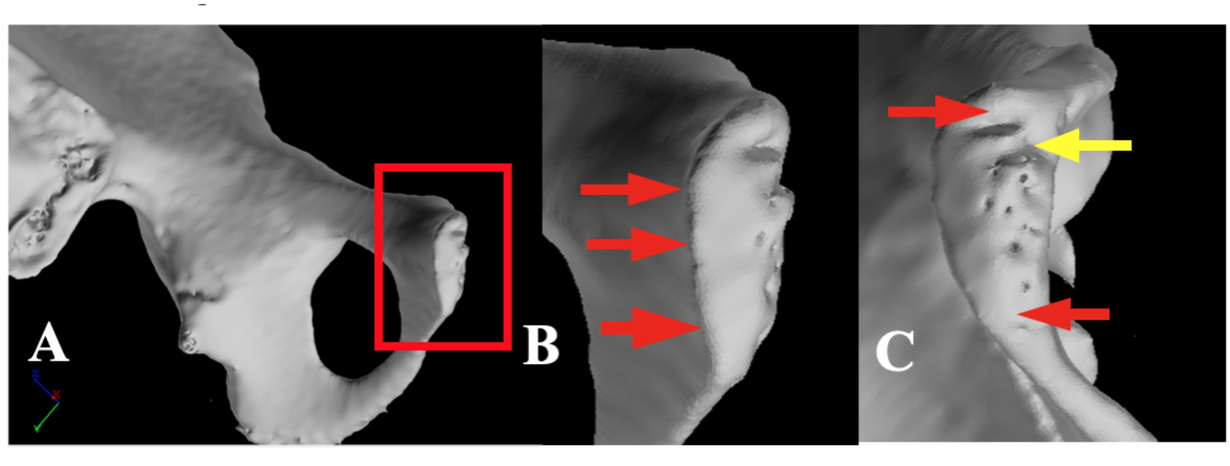

Figure 4.1 Female (17 years) scored as a Suchey-Brooks Phase 2, Hartnett Phase 2, and TA component scores of 3 for all scored features (symphyseal relief, superior protuberance, and dorsal and ventral margins). *(A) Left os coxa reconstructed from CT scan data- pubic bone enlarge in images B and C (red box), (B) pubic bone from red box in A. Red arrow indicates dorsal half of the symphyseal face with complete flattening indicated by a lack of billows cutting across the face and into the margin . (C) Red arrows indicate region of initial rampart formation along the ventral edge (red arrow) with some observable rounded billows still present in the superior portion (yellow arrow)*

Two individuals exemplify the difference between sharp, deep billows, and rounded shallower billows observed as potentially causing the overestimation several individuals whose true age was not captured by the phase-specific age interval.

A single female individual (Figure 4.1) whose reported age (17 years) was not captured by the estimated age interval using the Suchey-Brooks method was scored as a Phase 2, which has an associated age interval of 19-40 for females. This individual was scored as a Phase 2 using the Hartnett method as well. Using TA, the symphyseal relief, superior protuberance, ventral margin, and dorsal margin were all scored as 3's (soft, shallow billowing, late protuberance, rampart formation, and flattening complete, respectively). While billows are observable, they do not cut across the entire symphyseal face. Rather they have begun filling in, particularly noticeable along the dorsal half of the bone. Billows are still observable along the ventral edge of the bone, but they appear shallow and rounded. The superior protuberance is not observable, suggesting it has either not yet formed or it has been integrated to the symphyseal face. Given the degree of flattening that has commenced on the dorsal half of the bone, it is more likely that the occipital protuberance is mostly integrated, and contributing to the formation of the ventral rampart.

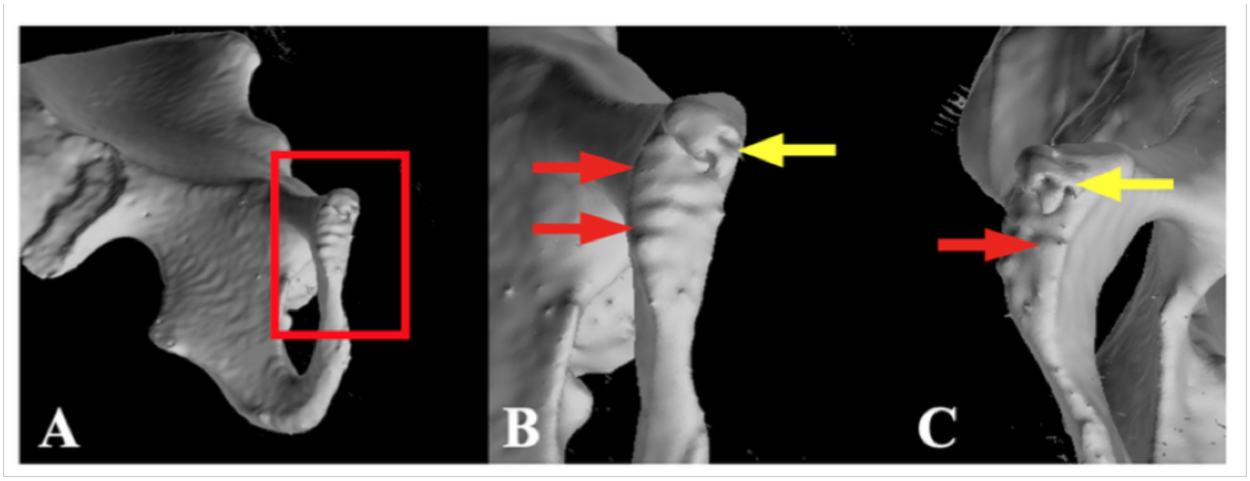

Figure 4.2 Male (18 years) scored as a Suchey-Brooks Phase 1, Hartnett Phase 1, and TA component scores of 1, 2, 2, and 1 for scored features (symphyseal relief, superior protuberance, and dorsal and ventral margins). *(A) Left of coxa reconstructed from CT scan data- pubic bone enlarged in images B and C (red box). (B) Pubic bone from red box in A. Red arrow indicates rides and furrow system across the entire symphyseal face, cutting into the dorsal edge. Yellow arrow indicating early superior protuberance. (C) Red arrow indicates the lack of ventral rampart development and continuation of the ridge and furrow system cutting into the ventral edge. Yellow arrow indicating early superior protuberance.*

For comparison, a male individual estimated with an age interval that included his reported age (18 years) by all three methods is depicted in Figure 4.2. This individual was scored as a Phase 1 using both the Suchey-Brooks and Hartnett methods, and assigned TA component scores of 1 (sharp billowing), 2 (early protuberance), 2 (beveled), and 1 (serrated) for the symphyseal relief, superior protuberance, ventral margin, and dorsal margin, respectively. The billows were distinct and cut across the entire symphyseal surface, including the ventral and dorsal edges. The superior protuberance is obvious, and not integrated into the symphyseal face.

While surface texture is not observable in the digital reconstructions, smooth or fine grained bone is typically observable in individuals displaying features such as this, and would have likely correlated with the other component and phase scores. However with this particular combination of features typically only seen in very young individuals, further collaboration of an addition feature only observable in very young individuals (smooth bone) is not imperative to an accurate age estimation, and would likely not have reduce the magnitude of bias, which is already fairly small compared to other age cohorts.

20-29 Years

TA and Suchey-Brooks both accurately aged 86% (19/22) of the individuals in this age category with an average bias of 5.0 and 6.8 years, respectively. While the Hartnett method was more accurate for this age cohort than for the 15-19 year old group, it still underperformed compared to the Suchey-Brooks method and TA, accurately estimating the age of only 64% (14/22) individuals, and overestimating individuals with an average bias of 10.5 years.

While accuracy as a percentage is the most typical way to report an age estimation interval that includes the reported age of a given sample, it can be misleading when divided into specific age cohorts with different sample sizes. While efforts were taken to reduce the impact of fluctuating sample distribution, the 15-19 age cohort includes a total of 11 individuals, as fewer individuals of this age are present in decedent samples, including donated skeletal collections and medical examiner cases.

In the 20-29 age cohort, a total of 22 individuals are included. While the accuracy rate as a percentage seems to decline for the Hartnett and Suchey-Brooks methods between the 15-19 and 20-29 age cohorts, this is not necessarily an accurate representation of the accuracy of these methods when applied to these age cohorts due to the differing sample size between these two groups. While the average bias for this age cohort was positive, the trend for each individual was less consistent compared to the 15-19 age cohort. In other words while *most* individuals were overestimated, there was some fluctuation in all three methods and a handful of individuals were also underestimated. This variation in bias direction and magnitude correlates with the wide range of features observed. While all individuals in the 15-19 age cohort displayed billowing, and most displayed some degree of superior protuberance, both features suggestive of a young age, this was not the case for individuals in this age cohort. Billows and a superior protuberance are typical of this age cohort, but so is a wider range of features, including varying degrees of ventral rampart formation, billow flattening, and in some extreme cases early rim development.

The billowing discussed previously continued to influence age estimation intervals for the Hartnett and TA methods. While sharp, deep billows are typically indicative of a very young individual, likely in the 15-19 age group or in the earlier half of this age group, more rounded billows would be indicative of a slightly older age. However, distinguishing the exact contour of the billows proved to be difficult in the reconstructions, potentially contributing to a slight overestimation of the youngest individuals, or a slight underestimation of individuals belonging to the older half of this age cohort (25+).

Texture is an additional feature used to distinguish between the young (mid 20s to early 40s) and the very young (teens to early 20s). For the younger half of this age cohort, smooth or fine-grained bone would be expected if it had been observable. However, course-grained bone would have been more suggestive of an individual in the latter half of this age cohort or older. The inability to observe this feature may have contributed to the greater degree of fluctuation in bias observed in this age cohort.

It is worth noting that the Hartnett method was less accurate overall than either TA or Suchey-Brooks. While a direct comparison between Hartnett and TA is difficult, given that one is a phase-based method and the other is component based, a comparison between Suchey-Brooks and Hartnett is informative regarding this discrepancy. Most individuals whose reported age did not fall within the estimated age interval using the Hartnett method were estimated with the same phase score using the Suchey-Brooks method. This discrepancy is not necessarily surprising when the method specific age intervals are examined, as the intervals of the Hartnett method differ slightly in distribution and are more precise (i.e. narrower) than the intervals of the Suchey-Brooks method.

As an example of typical variation for this age cohort, and of an individual estimated accurately by the Suchey-Brooks method but *not* the Hartnett method, a 29 year old female scored as a Phase 4 in both methods is depicted by Figure 4.3. The estimated age interval associated with a Phase 4 using the Suchey-Brooks method was 26-70 (95% CI), while the estimated age interval for a Phase 4 using the Hartnett method was 31-55 (95% CI). This individual displayed typical characteristics of a Phase 4 for both methods- a complete ventral rampart formation, a flat face, no symphyseal face depression, and a totally integrated superior protuberance.

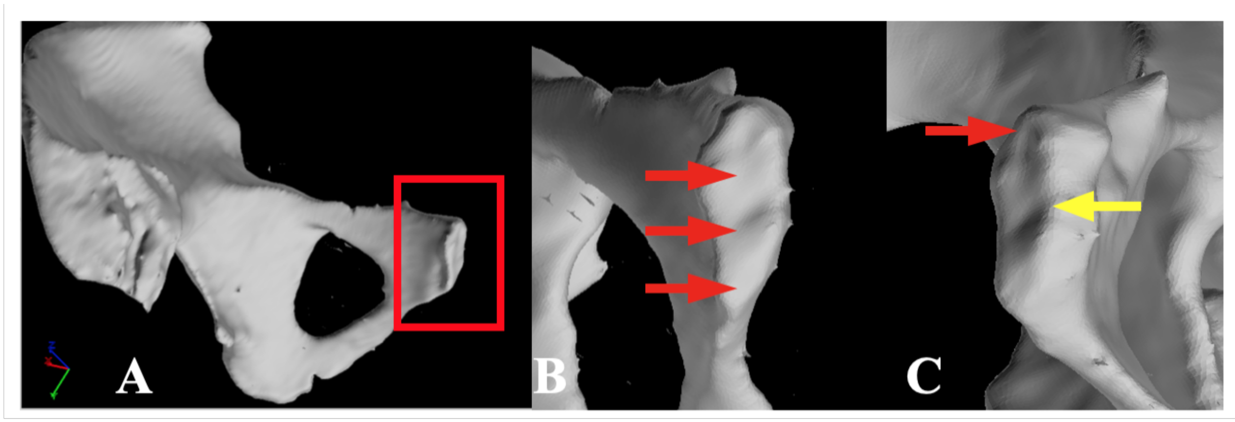

Figure 4.3 Female (29 years) scores as a Suchey-Brooks Phase 4, and Hartnett Phase 4. *(A) Left os coxa reconstructed from CT scan data- pubic bone enlarged in images B and C (red box). (B) Pubic bone from red box in A. Red arrows indicate a flat symphyseal face indicative of a Phase 4 in both the Suchey-Brooks and Hartnett methods. (C) Red arrow indicates region of complete integration of the superior protuberance and a lack of rim development, yellow arrow indicating complete ventral rampart format and demarcation from the pubic body.*

30-39 Years, 40-49 years

For the middle decades of life, especially the 30s and 40s, variation of the observable features in the pubic symphysis is expected and typical. Features indicative of younger age cohorts may still be present, however traits suggestive of older decades may also be observable. This ultimately results in significant variation of assigned phase scores for the Suchey-Brooks and Hartnett methods, as well as a wide variety of component scores using TA. Because of the similarity in trait variation typical in these two age cohorts, as well as the degree of similarity in average bias and rate of accuracy, the discussion of these two age cohorts has been consolidated into a single section here.

For two of the three methods–Suchey-Brooks and TA–there was the least amount of observable bias for the 30-39 age cohort at -0.9 and -1.5 years on average, respectively. Hartnett overestimated the 30-39 age cohort by 4.7 years on average. This mixture of over and under-estimation is typical in the middle decades of life, as there is a range of observable features typical for individuals in this age cohort. The small magnitude of bias suggests that on average, the estimated means and maximum likelihood were very similar to the reported age for any given individual. In terms of accuracy, all three methods performed the best in this age cohort. Suchey-Brooks was the most accurate, estimating 95 % (21/22) individuals accurately. Hartnett was the next most accurate, estimating 91% (20/22) individuals accurately. TA was the least accurate of the three applied methods estimating 86% (19/22) of individuals accurately.

There was no overlap between the inaccurately scored individuals using any of the methods. That is, inaccurately estimated individuals using the Suchey-Brooks method differed from those using the Hartnett method, and both were different from those inaccurately estimated using TA.

For the 40-49 age cohort, the Suchey-Brooks method estimated 91% (20/22) of individuals correctly, Hartnett declined to an accuracy of 82% (18/22), and TA, 68% (15/22). Bias shifted toward an underestimation for this age cohort; the Suchey-Brooks, Hartnett, and TA performed with an average bias of -7.7, -1.2, and -10.7 years respectively. While there was no consistent trend for individuals (some were overestimated while others were underestimated), this variation in magnitude and direction of bias reflects the broad range of symphyseal variation typical for this age cohort. There was minimal overlap in individuals whose reported age was not included in their estimated age interval. Of the 22 individuals included in this age cohort, only two were inaccurately estimated by all three methods.

This suggests that the inaccurate estimations for both the 30-39 and 40-49 age cohorts were not attributable to scan quality or the lack of observability of certain features (namely bone texture), but rather to a difference in how these individuals fit into the method specific phase and component descriptions, and differences in the method specific estimated age intervals. This pattern of variation, with resulting age-estimation error, is also seen in dry bone and is not unexpected in this portion of the lifespan.

A wide range of features are considered typical anatomical variation for the pubic symphysis in these two age cohorts. The ventral rampart may be in various stages of ossification for individuals in their 30s and 40s: Ventral rampart formation (Figure 4.4), rampart completion (Figure 4.5), and rim development (Figure 4.6). The symphyseal surface may display a range of traits as well, though residual billows or flattening is the most frequently observed. While the texture of bone is not observable in the digital reconstructions, microporosity may be observed in individuals in these age cohorts (particularly in their 40s) in dry bone samples. Features suggestive of more advanced age (e.g., microporosity, macroporosity, and rim breakdown) that seemed to have a more significant impact on later age cohorts did not have the same degree of impact for these middle decades. Ultimately not being able to observe these features had minimal impact in these decades, reflected by the small magnitude of bias and the high rates of accuracy.

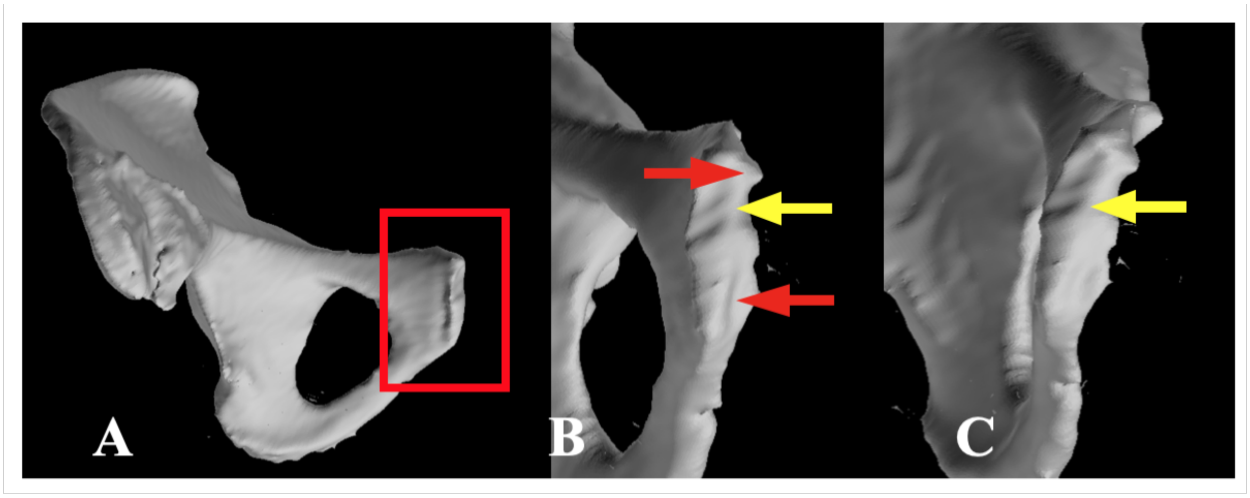

Figure 4.4 Female (34 years) scored as a Suchey-Brooks Phase 3, Hartnett Phase 3, and with component scores of 5, 3, 3, and 3 for all TA features (symphyseal relief, superior protuberance, and dorsal and ventral margins). *(A) Left os coxa reconstructed from CT scan data- pubic bone enlarged in images B and C (red box). (B) Pubic bone from red box in A. Red arrows indicate the initial development of the central rampart in the upper and lower extremities. Yellow arrow indicates the lack of vernal rampart development along the middle of the ventral margin. (C) Yellow arrow indicates the lack of ventral rampart development along the middle of the central margin.*

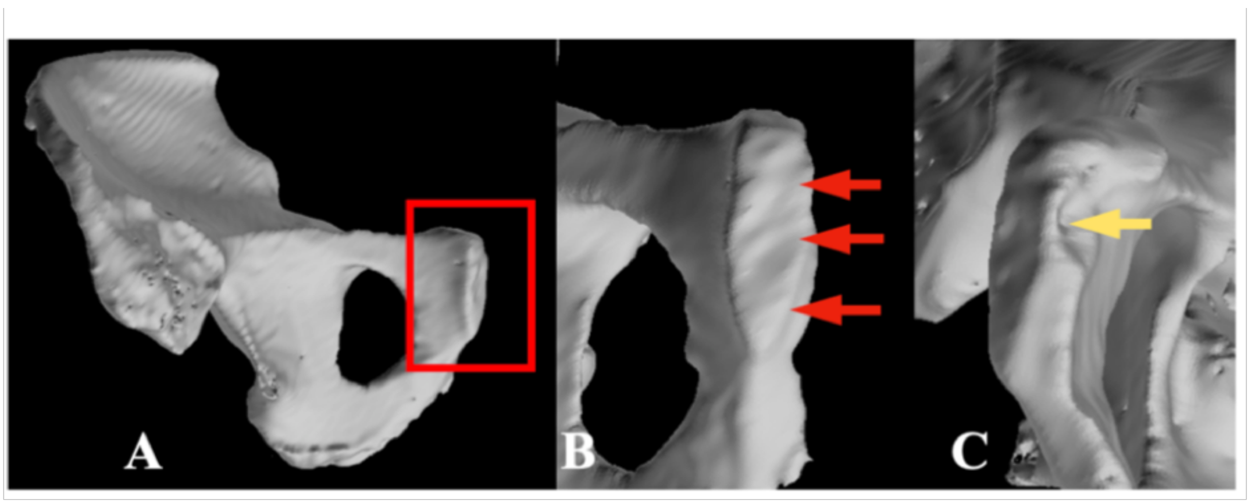

Figure 4.5 Male (43 years) scored as a Suchey-Brooks Phase 4, Hartnett Phase 4, and with component scores of 4, 3, 5, and 3 for all scored TA features (symphyseal relief, superior protuberance, ventral margin, and dorsal *margin). (A) Left os coxa reconstructed from CT scan data- pubic bone enlarge in images B and C (red box). (B) pubic bone from red box in A. Red arrows indicate the presence of a completed ventral rampart. (C) Yellow arrow indicates the complete vernal rampart along the middle of the ventral margin.*

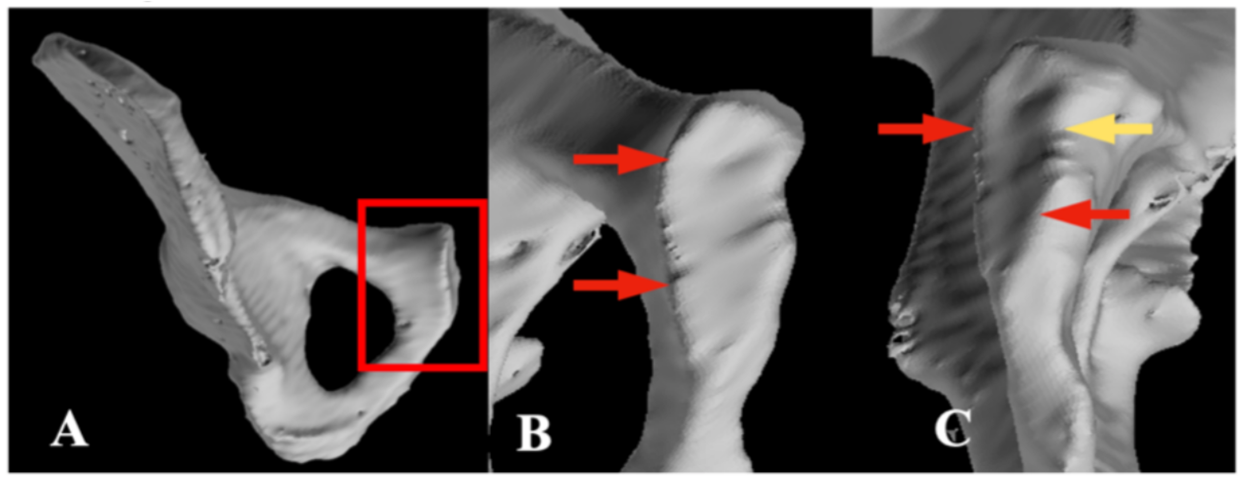

Figure 4.6 Male (47 years) scored as a Suchey-Brooks Phase 5, Hartnett Phase 5, and with component scores of 5, 4, 5, and 4 for all scored TA features (symphyseal relief, superior protuberance, ventral margin, and dorsal margin). *(A) Left os coxa reconstructed from CT scans, enlarged in images B and C (red box). (B) Pubic bone from red box in A. Red arrows indicate the presence of rim along the dorsal margin. (C) Yellow arrow indicates a hiatus in the ventral rampart. Red arrows indicate the presence of rim along the dorsal and ventral margins.*

50-59 Years

For the 50-59 age cohort, Suchey-Brooks, Hartnett, and TA performed with comparable degrees of bias in both direction and magnitude by -16.4, -13.3, and -14.3 years, respectively. In other words, individuals in this age cohort were underestimated on average, suggesting that observable features tended to be indicative of a younger phase or component score, and ultimately a younger age cohort.

The Suchey-Brooks and Hartnett methods both estimated 74% (17/23) individuals in the 50-59 age cohort accurately. TA, on the other hand, estimated 53% (12/23) of individuals with an estimated age interval that included their reported age.

While variation is expected for this age cohort, the same degree of variation that is typical for individuals in their 30s and 40s is not expected for individuals in their 50s. Typically individuals in this age cohort will have a completed ventral rampart (with or without a hiatus), and at least some degree of rim development on either or both the dorsal and ventral margins. The reconstruction process introduces some difficulty, noticeable beginning in this age cohort, in relation to a ventral hiatus.

The ventral hiatus is relatively obvious on dry bone samples. The surrounding perimeter of the symphyseal face will be fairly flat or rimmed, and the gap in the ventral rampart bares little resemblance to a ventral rampart in the process of completion. However, a novel issue unique to CT scan reconstructions introduce some error in deciphering a ventral rampart hiatus from a ventral rampart in the process of completion due to the presence of reconstruction billows along the ventral margin.

CT scans are taken in individual slices, and during the reconstruction process these individual slices are stitched together, essentially stacking until a three dimensional image is created. Software programs that build these three dimensional models all use preset algorithms to stitch these images together, and render them in a three dimensional fashion on a two dimension screen. Using the 3D slicer software, this stitching process often results in a billowed type appearance, most obvious on the iliac blade and occasionally on the rami of the os coxae.

Occasionally, this billowy appearance is evident on the symphyseal face. These reconstruction billows are typically diagonal (Figure 4.7), and distinguishable from residual billows, or billows indicative of younger individuals. However, on occasion these billows are *not* observable on the symphyseal face at all, but the ventral edge appears beveled. In other words, the reconstruction billows are *only* visible along the ventral edge and nowhere else along the symphyseal face (Figure 4.7), and it is impossible to tell if this beveled appearance is attributable to the presence of symphyseal billows (attributable to young age) or reconstruction billows.

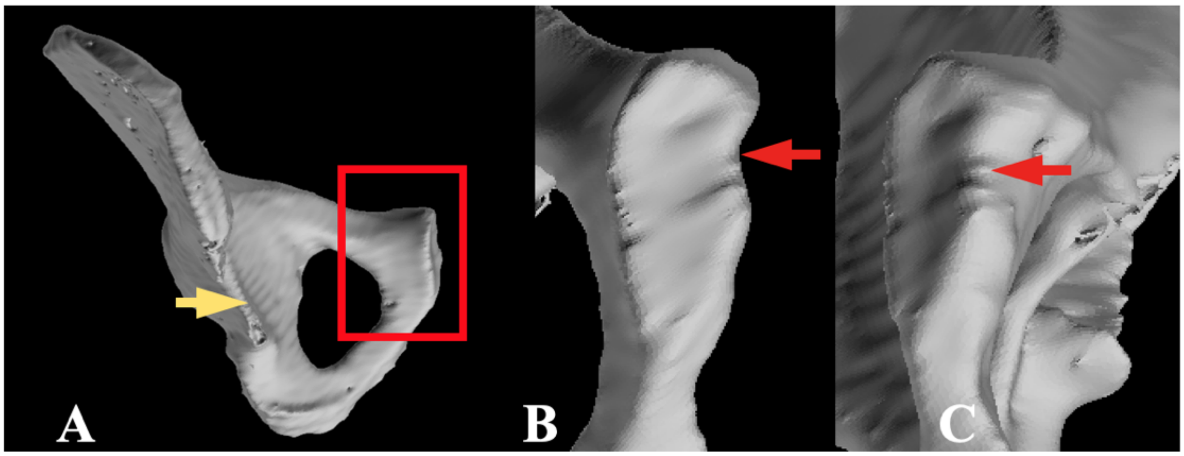

Figure 4.7 Male (53 years) scored as a Suchey-Brooks Phase 3, Hartnett Phase 3, and with component scores of 4, 3, 3, and 3 for all scored TA features (symphyseal relief, superior protuberance, ventral margin, and dorsal margin). *(A) Let os coxa reconstructed using CT scan data- pubic bone enlarged in images B and C (red box). Yellow arrow indicating reconstruction billows. (B) pubic bone from red box in A. Red arrows indicate a ventral rampart hiatus with reconstruction billows. (C) Red arrow indicates the reconstruction billows that make distinguishing a ventral rampart hiatus from ventral rampart formation difficult in reconstructions.*

In the case of a ventral rampart hiatus, the superior and inferior extremities are observable. However, the ruffled appearance along the middle of the rampart that would be a clearly observable hiatus in dry bone becomes less obvious in the reconstruction, attributable to the remnant presence of reconstruction billows. In other words, this ruffled appearance can give the impression of ventral rampart formation rather than a ventral rampart hiatus. For comparison, a ventral rampart hiatus without the reconstruction billows is depicted in Figure 4.8.

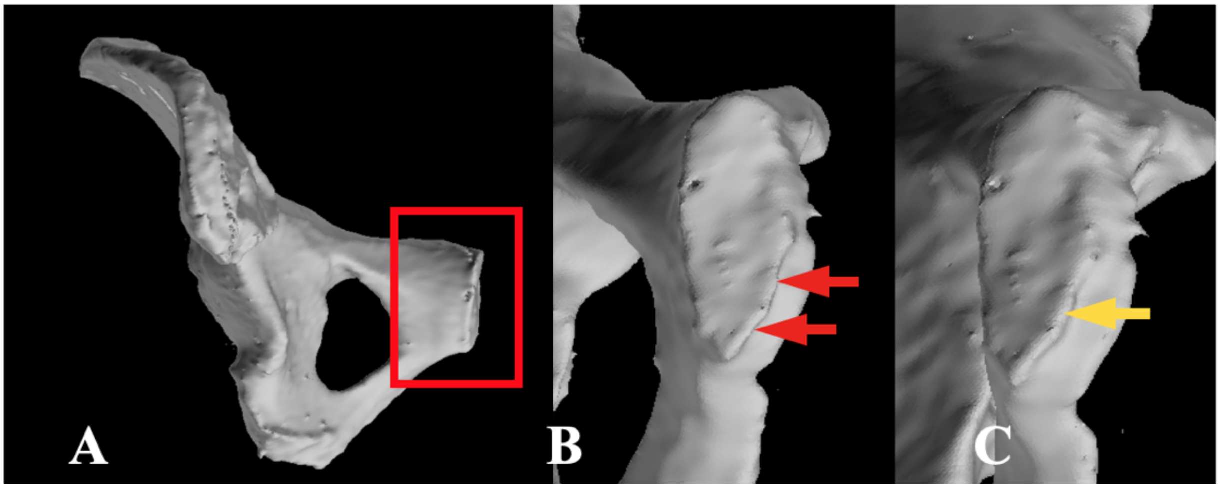

Figure 4.8 Female (54 years) scored as a Suchey-Brooks Phase 5, Hartnett Phase 4, and with component scores of 5, 4, 4, and 3 for all scored TA features (symphyseal relief, superior protuberance, ventral margin, and dorsal margin). *(A) Left os coxa reconstructed using CT scan data- pubic bone enlarged in images B and C (red box). (B) Pubic bone from red box in A. Red arrows indicate the presence of rim along the ventral margin. (C) Yellow arrow indicates the presence of rim along the ventral margin.*

60-69 Years, 70-79 Years

The features that were observable for the 60-69 and 70-79 age cohorts were essentially the same. Additionally, several features that are expected to be observed in these upper age cohorts were not visible due to the reconstruction process (discussed in this section). While the degree of bias increases in magnitude between the 60-69 and 70-79 age cohort, this shift in bias and the decreasing rates of accuracy reveal a trend that is of interest in relation to the observable features and the increasing likelihood of observing these features on individuals in these age cohorts. For this reason, discussion of these two age cohorts has been consolidated into one section.

On average for the 60-69 age cohort, Suchey-Brooks, Hartnett, and TA performed with a greater magnitude of -26.0 years, -14.4 years, and -17.0 years and an accuracy rate of 37% (7/19), 26% (5/19), and 47% (9/19) respectively. The 70-79 age cohort was underestimated on average as well by -24.9 years using the Hartnett method, -31.5 years using Suchey-Brooks, and -29.1 years using TA. Both the Hartnett and Suchey-Brooks methods accurately estimated 26% (5/190 of individuals, while TA estimated 37% (7/19) individuals accurately. A similar trend between the methods, with TA performing better than either Suchey-Brooks or Hartnett, is observed when these methods are applied to dry bone, although the rate of accuracy is still very low.

Individuals in these age cohorts would typically fall into more advanced phase categories using the Suchey-Brooks and Hartnett methods, and more advanced component scores using TA. These scores would be appropriate if micro- or macroporosity, rim, and/or breakdown were observable, in addition to a lack of features suggesting a younger age (e.g. billows, superior protuberance, lack of or formation of a ventral rampart). However, these features were difficult to see in the reconstruction on any individual.

Microporosity and macroporosity were not observable in any individual. These are features typically associated with more advanced age. Microporosity in dry bone appears as small pin pricks in the surface of the bone less than a millimeter in diameter, while macroporosity is slightly larger (typically >0.5mm), but does vary in diameter. Both features typically increase in frequency with advancing age. Given that the CT slices used to reconstruct this sample were taken a millimeter apart with 0.5 mm of overlap, neither micro- or macro-porosity was not expected to be seen as the diameter of these pores was smaller than the overlap of the CT scans.

This reconstruction issue made rim difficult to observe as well. Rim surrounding the symphyseal face can be thin on dry bone, and vary significantly in height in relation to the symphyseal face. Rim may be rounded and low, or sharp and narrow, as long as it is an observable raised ridge along the perimeter of the face (Boldsen et al. 2016)**.** With only a half millimeter overlap between CT slices, the accuracy of the reconstruction of rim around the symphyseal face likely varied a great deal, as the rim itself can vary significantly in appearance and height. Ultimately any rim that was less than a half millimeter tall or half a millimeter wide would not be visible in the reconstruction, as it would not be detected by the CT scan with only half a millimeter overlap.

A further consequence of the inability to observe rim was the inability to observe breakdown of that rim. Rim breakdown is highly indicative of advanced age, contributing to the underestimation of individuals in later age cohorts. There was a further complication with this, however, that was not expected. While the rim and its breakdown was not observable as it would typically be on dry bone, breakdown along the perimeter of the face was often observable as macroporosity. Without the surrounding rim of these areas, this often gave the appearance of serration or beveling rather than breakdown, causing a dramatic underestimation for individuals in later decades in all three applied methods (Figure 4.9).

The inability to observe these features indicative of advanced age contributed to the decreasing accuracy of age estimation for individuals in older age cohorts. Further, the tendency to underestimate individuals in these age cohorts through the assignment of younger phases and component scores with lower associated means and maximum likelihoods, affecting the calculation of the total average bias for each age cohort. When these methods are applied to dry bone, the same trend in decreasing accuracy for upper age cohorts is observed, even when all expected features are observable. While the trend is not novel to digital bone models, the inability to observe some of these features associated with advanced age likely worsened the issue.

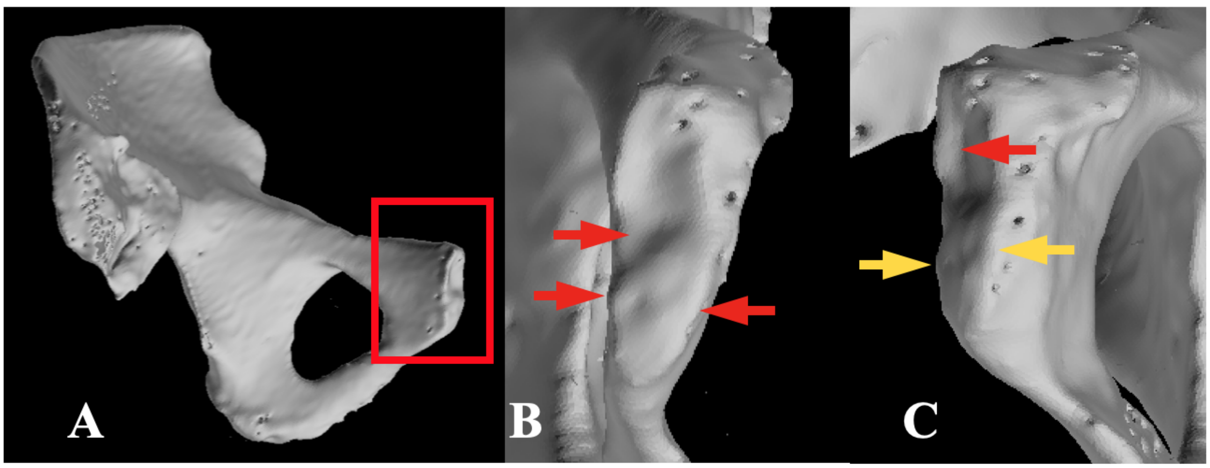

Figure 4.9 Female (68 years) scored as a Suchey-Brooks Phase 3, Hartnett Phase 5, and with component scores of 5, 4, 5, and 4 for all scored TA features (symphyseal relief, superior protuberance, ventral margin, and dorsal margin). *(A) Left os coxa reconstructed using CT scan data- pubic bone enlarged in images B and C (red box). (B) Pubic bone from red box in A. Red arrows indicate a slight beveled appearance along the dorsal and ventral margins. (C) Red arrow indicates the presence of rim along the superior dorsal margin. Yellow arrows indicate a slight beveled appearance along the dorsal and ventral margins.*

80+ Years

For the 80-89 and 90+ cohorts, all three methods performed similarly in that all individuals were underestimated using phase-based methods, and most individuals were underestimated using TA (although to a lesser degree than with the phase-based methods). The degree of bias was at least -30 years for both age cohorts using any of the three methods. For this reason, discussion of the 80-89 and 90+ age cohorts have been combined as one section.

TA was to be the most accurate for individuals in their 80s and 90s, estimating an age interval encompassing the reported age of 39% of individuals (7/18) in the former age cohort, and 31% (4/13) accurately for the latter. Additionally, TA demonstrated the least amount of bias for both the 80-89 and 90+ age cohort, underestimating by -30.4 and -34.5 years, respectively. Suchey-Brooks method, 22% (4/18) individuals in the 80-89 age cohort were estimated accurately, with an average bias of -32.8 years. For the 90+ age cohort, no individuals (0%; 0/13) were accurately estimated, as the upper limit, and thus the oldest individuals in the sample used to develop this method, is 87 years old for females, and 86 years old for males. On average the bias using Suchey-Brooks for the 90+ age cohort was -49.4 years. The Hartnett method was the least accurate for the 80-89 age cohort, accurately estimating 11% (2/18) individuals, and performed with an average bias of -32.8 years. For the 90+ age cohort, 13% (2 of 13) individuals were estimated accurately using the Hartnett method, and the average observed bias was -36.2 years.

At this age, the degeneration of bone would be typical. Rim breakdown, macroporosity, and microporosity as well as a loss of bone quality (e.g. lightweight bone) would suggest an individual belonging to these age cohorts when observing dry bone. However, none of these features were observable reliably in these age cohorts for the same reasons discussed in the 60-69 and 70-79 age cohorts. This is reflected in the continued decrease in accuracy, and the increasing degree of average bias. It is worth noting that no individuals in the 80-89 or 90+ age cohorts were overestimated. This trend is similar to that seen when these methods are applied to dry bone, though it was likely worsened by the inability to see the expected features.

Sex-Specific Differences in Accuracy and Bias

Sex specific differences in both accuracy and bias were present among all three age estimation methods tested. Suchey-Brooks and TA both scored more females accurately than males. TA estimated 64% of females accurately with an average bias of -11.8 years, and 53% of males accurately with an average bias of -15.6 years; however, it should be noted that aggregated accuracy and bias values mask important trends in the performance throughout the adult lifespan. Importantly, the Suchey-Brooks and Hartnett methods are phase-based, with fixed point ages and intervals that were calculated directly from each reference sample. In other words, for an individual of a specific sex, there are only six or seven possible point ages with their associated intervals (depending on method) that can be reported. In contrast, TA calculates a maximum likelihood estimate of age using sex-specific probabilities from the traits present. The width of the confidence interval will vary based on which traits are present and how well they agree with each other.

An interesting pattern arises when taking into account the features of bone used to distinguish between phase or component scores. More women than men were estimated accurately using the TA and Suchey-Brooks method. However, using the Hartnett method, the opposite was true–more men were estimated accurately.

While bone quality (texture) is a component intended to be used in TA, it is not necessarily needed. An age estimation can be performed without the presence of this feature.

Micro- and macro-porosity are listed in the Suchey-Brooks method as characteristics of later age phases, however they are listed with other features that can help place an individual in a given phase, as is characteristic of a phased based method. However, Hartnett differs from either of these two methods. Bone weight and quality is a key factor in determining which phase an individual should be assigned. More emphasis is placed on bone quality in age estimation using the Hartnett method than either the TA or Suchey-Brooks methods. This may become particularly relevant to populations who are susceptible to poor bone quality, as this is not easily evaluated in digital bone models.

According to the National Osteoporosis Foundation, over 20% age 50 or over have some degree of osteoporosis, while half of women over the age of 50 demonstrate some degree of low bone mass(National Osteoporosis Foundation, March 2021) While the written descriptions for each phase provided by Suchey-Brooks, as well as the images and casts for each phase, focus solely on the symphyseal face, it is likely that other aspects of the bone have some degree of influence on a given observers evaluation of the bone, and which phase would be the most appropriate. Bone quality, particularly weight and texture, are listed explicitly as features distinguishing between phases for the Hartnett method, and are features indicative of internal bone quality. An individual with bone degeneration, like osteoporosis, is likely to have low bone weight and high bone porosity. However, these are not observable in digital bone models that only show surface features, ultimately affecting the utility of this method when these features are not observable. Thus, if current methods are applied to digital bone samples as they are to dry bone samples, using methods of age estimation that do not include evaluations of bone density through weight would be most appropriate.

Intra-observer Reliability

From the total sample of reconstructed os coxa suitable for evaluation (N=169), sixty-one (n=61) individuals were selected at random and scored a second time using the Hartnett method to test intra-observer reliability. Perfect intra-observer agree (i.e., the same score was assigned in both rounds), occurred in 66% (40/61) of the sample. This degree of intra-observer agreement is less than anticipated. However, the Hartnett method was new to the observer. While they had previous experience using both TA and Suchey-Brooks on dry bone samples, the Hartnett method was new to the observer and only applied on a digital bone sample- there was no previous experience scoring a 'typical' dry bone sample with this method. This had some impact on the intra-observer error. Additionally, because the observer had relatively little experience with these methods, some degree of experiential learning, and an associated shift in understanding of method definitions, may have occurred. Although a trial scoring sample was used to mitigate this effect, it is likely that it was not entirely eliminated.

While the overall agreement was less than anticipated, it is not outside the realm of normal. Conflicting evidence regarding the intra-observer error has been reported by multiple sources, some reporting high levels of agreement of 86.5% (Kim et al. 2019), while others report disagreement up to 56% (Hartnett 2010).

Comparison of Method Performance

The results of this study were compared to published data regarding the accuracy and bias of the three methods applied. Each publication uses a different sample size with a non-standard age distribution and reports these data in a slightly different way, which complicates the direct comparison of accuracy and bias among studies. Because of this, the comparisons of method performance described here focus on general reported trends for each of the three methods, and detail specific rates of bias and accuracy when possible.

When compared to both the dry bone and digital sample data documented by Savall et al., the performance of the Suchey-Brooks method was relatively comparable in terms of general trends. Applied to the NMDID sample, the Suchey-Brooks method estimated 69% (57/83) of females accurately, and 53% (46/86) of males. The Savall et al. study was unique in that it applied the Suchey-Brooks method to both a dry bone and digital bone sample. When Savall et al. applied the Suchey-Brooks method to a dry bone sample, 62.3% of males and 45.5.% of females were estimated accurately. Using Suchey-Brooks on a digital sample, 60.4% and 68.2% of males and females, respectively, were estimated accurately. Ultimately the performance of the Suchey-Brooks method in this study was comparable to the Savall et al. validation study using digital and dry bone samples. The overall accuracy, including both males and females, was not reported by Savall et al., so sex specific accuracy rates are used for the sake of comparison in this discussion of the Suchey-Brooks method performance. A cumulative total of 60% (101/169) of individuals (this includes both males and females) were estimated accurately in this study.

The Hartnett method has not been evaluated by as many studies as Suchey-Brooks. One validation study compared the Suchey-Brooks method for the pubic symphysis and the Isçan method (Isćan et al. 1984) for the fourth sternal rib ends to the Hartnett modification of these methods using a sample of 322 dry bone specimens from the Bass Donated Collection (Merritt 2010). For this discussion, only the data for Hartnett modification is included and compared to the digital sample used for this study. In the Merritt evaluation study, 58% (187/322) of individuals were estimated accurately using the Hartnett method. When applied to the digital sample reconstructed from the NMDID, the Hartnett method accurately estimated a total of 54.7% (92/169) of individuals. When compared to Suchey-Brooks, the associated age range of each of the seven phases is more narrow.

The Suchey-Brooks method is quite accurate, as would be expected when intervals span thirty to forty or more years. The probability that an individual falls somewhere in that large of a range is relatively high by chance alone, and only increases even if the amount of additional information is relatively small. The Hartnett age ranges are more precise (i.e., narrower), and ultimately estimate fewer individuals accurately. This is demonstrated by the significant number of individuals who were not considered to be scored accurately because they fell just above or below the estimated age range. Merritt observed a similar trend in her validation study, as the presence of a seventh phase compared to the Suchey-Brooks method reduced the degree of bias for individuals in later decades of life, who are likely to be scored as a Phase 6 or Phase 7 using the Hartnett method, whereas all individuals of advanced age are likely to be scored as a Phase 6 using the Suchey-Brooks method. In other words when using the Suchey-Brooks method there are only six points per sex that can be used for comparison for any given individual, whereas with the Hartnett method there are seven potential points per sex.

In the validation of Boldsen et al. Transition Analysis by Milner and Boldsen (2012), accuracy and precision for each of the three skeletal elements typically included in TA–cranial sutures, the iliac auricular surface, and the pubic symphysis–were discussed separately, and as an integrated method- how they are typically used. Since the pubic symphysis was the only feature used in this study, this was the only data used for this aspect of analysis. The pubic symphysis, according to the original publication, estimated 87.8% (216/246) of individuals from the Bass Donated Collection accurately. Age ranges generated using only the pubic symphysis were varied and broad, including several decades of life. All five features of the pubic symphysis- symphyseal relief, dorsal symphyseal texture, superior protuberance, ventral margin, and the dorsal symphyseal margin- were all used for this estimation in the original publication. TA applied to the digital models reconstructed using data from the NMDID estimated 57% (86/169) of individuals accurately.

This is a marked difference in rate of accuracy (87.8% in dry bone from the Bass Collection versus 57% 3D reconstructions from NMDID CT data), which may be attributable to several factors. First, while the original method evaluates five features of the pubic symphysis, only four were evaluated in this study. The dorsal symphyseal texture was not evaluated, a determination that was made prior to beginning this study. The possible scores for the dorsal symphyseal texture range 1-4 (smooth, coarse-grained, microporosity, and macroporosity), and refer to the tactile quality of the bone that, according to the TA manual, is not easily scored through pictures. The determination was made in advance that if the features of the dorsal symphyseal texture were based on tactile quality, it would not be possible to evaluate them in a CT reconstruction.

Not including this feature in the scoring process had some impact on the estimates generated by the TA software program, which relies on the presence of values for as many of the total 33 components as possible. It is also worth noting that this difference in accuracy may be attributable to more than just the difficulty in observing certain features in digital reconstructions, as the test samples for the validation study conducted by Boldsen et al. and the sample for this study were drawn from markedly different populations.

The observable features that were included are all reported to have transition ages at a relatively young age. For example, the completion of the ventral rampart, scored anywhere from a 3-5 (ventral rampart formation, complete with anterior sulcus, complete without anterior sulcus) for the ventral margin takes place relatively early. The breakdown of rim along the ventral and dorsal margin is documented as an ‘older trait’- particularly when present along both margins and in combination with an irregular face (Boldsen et al. 2002). However, the formation of rim was relatively difficult to observe in the digital reconstructions. Bony rim can be very thin, or detectable only in a tactile sense (ex. Distinguishing rim from face with a fingernail). Since rim can be so thin in width, the overlap of .5mm between scans was not always enough to portray that rim in the reconstruction. There is also the possibility that therim was simply not dense enough to be modeled in the reconstruction due to the selected threshold value, though the prior is more likely. The lack of observable rim, and ultimately observable breakdown of that rim, were contributors to the decrease in accuracy compared to TA applied to dry bone, and the overall tendency to underestimate.

Chapter 5: Conclusions

The pubic symphysis is among the most commonly used skeletal feature for adult age estimation around the world for both forensic and bioarchaeological applications. Dozens of validation studies focusing on the wide variety of methods used to age this region of the skeleton exist, and facilitate comparisons between various populations. Of the features commonly used for age estimation, this one is perhaps the easiest to access in the case of autopsy, reducing processing time to some degree. Ultimately the abundance of research of age estimation methods using the pubic symphysis and the wide spread use of these methods make this feature ideal in preliminary evaluations of the performance of commonly applied age estimation methods on digital bone models.

While the majority of methods for evaluating this feature were developed and validated on dry bone, medical imaging is becoming increasingly popular to analyze forensic cases without the need for maceration and to document archaeological remains. This is because medical imaging allows for an easier method of sharing information among professionals without having to physically transport samples or personnel to a second location. It also eliminates a significant degree of expense and liability related to transportation and the processing required to remove tissues prior to analysis. This also mitigates the issues of altering remains in cases where significant cultural or religious beliefs dictate that the remains of a loved one must remain intact. Additionally, it may reduce the degree of subjectivity currently prevalent in our field by allowing collaboration between experts (Grabherr 2009). This increased reliance on medical imaging to document and analyses forensic cases and archaeological remains has called into question whether methods commonly used on dry bone are suitable for use on digital bone models.

This study evaluates the performance of commonly applied age estimation methods to a digital sample - specifically three dimensional models of the adult pelvis reconstructed from computer tomography scans. Three commonly used methods for evaluating the pubic symphysis were included in this study to assess differences in the performance of each technique on the digital models. By comparing the results from each method to each other and to published validation studies, these data inform our understanding of possible systematic differences introduced by the digital documentation and subsequent reconstruction process.

All three methods performed the best on individuals under the age of fifty, maintaining a relatively high rate of accuracy (> 70%). Methods that did not include a component of bone density or quality (Suchey-Brooks (1990) and Boldsen et al. (2002) Transition Analysis) ─classified more individuals accurately than the Hartnett method, which used bone quality to distinguish between phases in older ages. All three methods underperformed in terms of accuracy when compared to their reported rates using a dry bone sample. However, this difference was less notable for the phase-based methods (Hartnett (2010) and Suchey-Brooks)─ than for the component-based method (TA). While it is possible that differences in sample size and composition, as well as experience levels of the observers applying these techniques, may have introduced differences between this study and published work, several important trends were still evident.

Overall, the Suchey-Brooks method classified the most individuals accurately, estimating a total of 60% of individuals with an estimated age interval that included their reported age. Phase-based methods generally have a wider estimated age range per phase, increasing the likelihood that a given age would all within the estimated age range associated with the assigned phase. Given that of the two phase-based methods tested in this study, Suchey-Brooks has fewer phases with wider age ranges associated with them, higher accuracy than other methods is not surprising. In addition, the assumption that bone ages as a cohesive unit, which may reduce the accuracy of phase based methods when applied to a dry bone sample, is beneficial when several age-related features (e.g. microporosity) are not observable in digital bone models reconstructed from CT scans. In other words, because the bone is scored as a unit, features that are visible in the 3D reconstructions can drive the assignment to a phase, even if some of the features typical of that phase are not visible. In contrast, component-based methods that score each feature of the surface individually are less accurate when features cannot be observed because each trait individually contributes to the age estimate. In light of the results of this study, if a digital sample is going to be estimated for age-at-death, phase-based methods are recommended for use. However, the aforementioned faults recognized in the field of forensic and biological anthropology associated with phase-based methods (e.g. wide age intervals, poor performance with older populations, and a lack of statistical validation) are still relevant when applying these methods to a digital sample, and likely exaggerated due to the features that are difficult to observe in reconstructions, and typically associated with advanced age. Further, the development of a component-based method specific to CT scans that incorporates the visible features of reconstructions and the unique aspects of medical imagining would be beneficial if medical imaging is to become a widely useable resource for anthropologists.

This research would be beneficial for researchers studying human variation within biological anthropology or bioarchaeology. Comparing methods that incorporate different parts of the skeleton may be beneficial, as age related change occurs at different points throughout the body. Thus, incorporating more skeletal features would likely aid in estimating more individuals accurately with a smaller degree of bias not only using digital bone samples, but dry bone as well. Additionally, method development specific to CT scans would be a worthwhile endeavor. All three applied methods in this study used age-related features that were not reliably observable, particularly for older age cohorts. The inability to observe features related to advanced age likely exaggerated the trend of underestimation that is already observed in dry bone. Further development of a method specific to CT scans that does not rely as heavily on features like macroporosity and breakdown, and instead focuses on CT-specific advantages (e.g. voxel density, connective tissue thickness, trabecular bone structure (Alcaraz et al. year) would be beneficial.

As medical technology continues to advance, this preliminary evaluation of commonly applied pubic symphysis-based age estimation methods suggests that digital bone samples may be able to supplant dry bone, as long as appropriate methods are applied. Methods applied in this study performed similarly on a digital bone sample when compared to their performance on dry bone samples. However, digital bone is not with out its limitations. When selecting an age estimation method to apply to a digital bone sample, researchers should avoid methods that place a heavy emphasis on tactile components of bone, like density and weight (e.g. Suchey-Brooks and TA). They should also select methods that do not place a heavy emphasis on bone texture, as this is not observable on digital bone. Further research evaluating the degree to which these features *actually* impact the accuracy of an age estimation would be beneficial, as this study did not include a large enough sample to draw statistically significant conclusions regarding the performance of these methods, or the contribution of specific features of the symphyseal face on overall age estimation. Overall, the trends observed in this study differed from previously conducted validation studies using a dry bone sample, particularly for the applied component-based method, TA. If age-estimation is going to be conducted using a digital sample, phase-based methods should be applied, and the limitations of these methods should be taken into consideration. Ultimately, the development of a CT specific component-based method would be the best approach to using digital bone models for age estimation.

REFERENCES

Baker, B. J., Dupras, T. L., & Tocheri, M. W. (2005). The Osteology fo Infants and Children. In (pp. 75-92). Texas A&M University Press.

Barrier, P., Dedouit, F., & Braga, J. (2009). Age at death estimation using multiplied computed tomography reconstructions of the posterior pelvis. In (Vol. 54, pp. 773-778). Journal of Forensic Sciences.

Berry, S., & Edgar, H. (2019). Standardizing Data from the Dead. In (Vol. 264, pp. 773-778). Studies in Health Technology and Informatics.

Bocquet-Appel, J. P., & Masset, C. (1996). Paleodemography: Expectancy and False Hope. In (Vol. 99). American Journal of Physical Anthropology.

Boldsen, J., Milner, G., Konigsberg, L., & Wood, J. (2002). Transition Analysis: a new method for estimating age from skeletons. In (pp. 73-106): Cambridge University Press.

Boldsen JL, Milner GR, Hylleberg R, and Ousley D. 2016. ADBOU Ver. 2.1.046

Bradley, W. G. (2008). History of Medical Imaging. In (Vol. 152, pp. 349-361): Proceedings of the American Philosophical Society.

Brooks, S., & Suchey, J. M. (1990). Skeletal age determination based on the os pubis: a comparison of the Acsadi-Nemeskeri and Suchey-Brooks Methods. In (Vol. 5, pp. 227-238). Human Evolution.

Brooks, S. T. (1955). Skeletal age at death: the reliability of cranial and pubic age indicators. In (Vol. 13, pp. 567-597). American Journal of Physical Anthropology.

Buckberry, J. (2015). The (mis)use of adult age estimates in osteology. In (Vol. 42, pp. 323-331). Annals of Human Biology.

Chiba, F., Makino, Y., & Motomura, A. (2014). Age estimation by quantitative features of pubic symphysis using multi detector computed tomography. In (Vol. 128, pp. 667-673). International Journal of Legal Medicine.

Dirnhofer, R., Jackowski, C., Vock, P., Potter, K., & Thali, M. (2006). Virtopsy: Minimally Invasive, Imaging-guided Virtual Autopsy. In (Vol. 26, pp. 1305-1333). RadioGraphics.

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

Falys, C. G., & Lewis, M. E. (2010). Proposing a Way Forward: A Review of Standardisation in the Use of Age Categories and Aging Techniques in Osteological Analysis (2004-2009). In (Vol. 21, pp. 704-716). International Journal of Osteoarchaeology.

Fedorov A., Beichel R., Kalpathy-Cramer J., Finet J., Fillion-Robin J-C., Pujol S., Bauer C., Jennings D., Fennessy F., Sonka M., Buatti J., Aylward S.R., Miller J.V., Pieper S., Kikinis R. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. Magnetic Resonance Imaging. 2012 Nov;30(9):1323-41. PMID: 22770690.

Ferrant, O., Rouge-Maillart, C., & Guittet, L. (2009). Age at death estimation of adult males using octal bone and CT scan: A preliminary study. In (Vol. 186, pp. 14-21). Forensic Science International.

Garvin, H., M., & Passalacqua, N. V. (2012). Current Practices by Forensic Anthropologists in Adult Skeletal Age Estimation. In (Vol. 57, pp. 427-433). Journal of Forensic Sciences.

Ísçan, Y., Loth, S. R., & Wright, R. K. (1984). Metamorphosis at the sternal rib end: A new method to estimate age at death in white males. *American Journal of Physical Anthropology*, *65*(2), 147–156.

Getz, S. (2020). The use of transition analysis in skeletal age estimation. In (pp. 1-11). SIREs Forensic Science.

Gilbert, B. M., & McKern, T. W. (1973). A method fo aging the female Os pubis. In (Vol. 38, pp. 31-38): American Journal of Physical Anthropology.

Ginat, D. T., & Rajiv, G. (2014). Advances in Computed Tomography Imaging Technology. In (Vol. 16, pp. 431-453).

Grabherr, S., Cooper, C., & Ulrich-Bochsler, S. (2009). Estimation of sex and age of "virtual skeletons"- A feasibility study. In (Vol. 19, pp. 419-429). European Radiology.

Gárate, M. (2017). Voxel Datacubes for 3D Visualization in Blender. In (Vol. 129, pp. 1-11). Publications of the Astronomical Society of the Pacific: Astronomical Society of the Pacific.

Hartnett, K. (2010). Analysis of age-at-death estimation using data from a new, modern autopsy sample- Part I: Pubic bone. In (Vol. 55, pp. 1145-1151). Journal of Forensic Sciences.

Katada, K. (2001). Characteristics of Multislice CT. In (Vol. 125, pp. 1772-1775). Journal of the Japan Medical Association.

Katz, D. and Suchey, J.M. (1986), Age determination of the male Os pubis. Am. J. Phys. Anthropol., 69: 427-435. doi:10.1002/ajpa.1330690402

Lopez-Alcaraz, M., Gonzalez, P., & Aguilera, I. (2015). Image analysis of pubic bone for age estimation in a computed tomography sample. In (Vol. 129, pp. 335-346). International Journal of Legal Medicine.

Martin, J. A., & Buckwalter, J. A. (2001). Roles of Articular Cartilage Aging and Chondrocyte Senescence in the Pathogenesis of Osteoarthritis. *The Iowa Orthopaedic Journal*, *21*, 1–7.

McKern, T. W., Stewart, T. D., & Army, U. S. (1957). Skeletal age changes in young American males. In. Quartermaster Research and Engineering Command.

Meindl, R. S., Lovejoy, C. O., Mensforth, R. P., & Walker, R. A. (1985). A Revised Method of Age Determination Using the Os Pubis, With a Review an dTests of Accuracy of Other Current Methods of Pubic Symphyseal Aging. In (Vol. 68, pp. 29-45). American Journal of Physical Anthropology.

Mendle, J., Beltz, A. M., Carter, R., & Dorn, L. D. (2019). Understanding Puberty and Its Measurement: Ideas for Research in a New Generation. In (Vol. 29, pp. 82-95). Society for Research on Adolescence.

Merritt, C. E. (2014). A Test of Hartnett's Revisions to the Pubic Symphysis and Fourth Rib Methods on a Modern Sample. *Journal of Forensic Sciences*, *59*(3), 703–711. https://doi.org/10.1111/1556-4029.12380

Moore-Jansen, P., Ousley, S., & Jantz, R. (1994). Data Collection Procedures for Forensic Skeletal Material. Report of Investigations. In (Vol. 48). International Journal of Legal Medicine.

Reid, J., & Odle, T. (2008). Computed Tomography in the 21st Century: Changing Practice for Medical Imaging and Radiation Therapy Professionals. In. American Society of Radiologic Technologists.

Savall, F., Herin, F., & Peyron, P. (2018). Age estimation at death using pubic bone analysis of a virtual reference sample. In (Vol. 132, pp. 609-615). International Journal of Legal Medicine.

Spin-Neto, R., Gotfredsen, E., & Wenzel, A. (2014). Variation in voxel value distribution and effect of time between exposures in six CBCT units. In (Vol. 43). DentoMaxilloFacial Radiology.

The Map of Identified Osteological Collections. Forensic Anthropology Society of Europe (FASE). https://www.google.com/maps/d/edit?mid=162_ElRDZuDCJfM10jCkPpRSFPSw&ll=31.412645801177902%2C0&z=2

Ubelaker, D., & Khosrowshahi, H. (2019). Estimation of age in forensic anthropology: historical perspective and recent methodological advances. In (Vol. 4, pp. 1-9): Forensic Sciences Research.

Verdun, F. R., Racine, D., Ott, J. G., Tapiovaara, M. J., Toroi, P., Bochud, F. O., . . . Edyvean, S. (2015). Image quality in CT: From physical measurements to model observers. In (Vol. 31, pp. 823-843): Physical Medica.

Villa, C., Hansen, M. N., Buckberry, J., Cattaneo, C., & Lynnerup, N. (2013). Forensic age estimation based on the trabecular bone changes of the pelvic bone using post-mortem CT. In (Vol. 233, pp. 393-402).

White, T. D., Black, M. T., & Folkens, P. A. (2012). *Human osteology*. Academic Press.

Wink, A., E. (2014). Pubic Symphyseal Age Estimation from Three-Dimensional Reconstructions fo Pelvic CT Scans of Live Individuals. In (Vol. 59, pp. 696-702): Journal of Forensic Sciences.