

2017

Sex and age at death estimation from the os pubis: validation of two methods on a modern autopsy sample

<https://hdl.handle.net/2144/23749>

Boston University

BOSTON UNIVERSITY
SCHOOL OF MEDICINE

Thesis

**SEX AND AGE AT DEATH ESTIMATION FROM THE OS PUBIS:
VALIDATION OF TWO METHODS ON A MODERN AUTOPSY SAMPLE**

by

ASHLEY E. CURTIS

B.A., University of Connecticut, 2012

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

2017

© 2017 by
ASHLEY ELIZABETH CURTIS
All rights reserved

Approved by

First Reader

Tara L. Moore, Ph.D.
Associate Professor of Anatomy and Neurobiology
Program in Forensic Anthropology

Second Reader

Jonathan D. Bethard, Ph.D.
Assistant Professor of Anthropology
University of South Florida

ACKNOWLEDGMENTS

The completion of this project would not have been possible without the assistance and guidance of my thesis committee. My first reader, Dr. Jonathan Bethard, who provided mentorship and guidance which were invaluable to the creation and progress of this thesis. Without his help, this project would have never gotten off the ground. Also, to my second reader, Dr. Tara Moore, whom was willing to offer insight into the process of conducting research and scientific writing.

I am also grateful to Dr. Laura Fulginiti, whom so graciously allowed me to utilize her laboratory facilities at the Maricopa County Forensic Science Center in Phoenix, Arizona. Her hospitality and generosity made the data collection a fun and informative experience. Additionally, I would like to thank Dr. Kristen Hartnett-McCann for both assisting on my initial project proposal, and allowing me access to the phenomenal skeletal collection she has curated at the FSC.

Furthermore, I would like to acknowledge the help of Dr. Farzad Mortazavi in completing the statistical analysis on my data. I would also like to give many thanks to my classmate and friend Megan Sharpe, who not only flew to Phoenix to be my second observer, but whose kindred spirit has kept me sane throughout this entire process. Also, my gratitude goes to Boston University School of Medicine for offering the funding to complete this project.

Lastly, I would like to thank my cat, J.B., for only napping on my thesis notes when I really need them, and trying to help with the typing of this document by walking on my laptop keyboard when I was not looking.

**SEX AND AGE AT DEATH ESTIMATION FROM THE OS PUBIS:
VALIDATION OF TWO METHODS ON A MODERN AUTOPSY SAMPLE**

ASHLEY E. CURTIS

ABSTRACT

Estimating sex and age at death are two crucial processes during the creation of a biological profile for a set of skeletal remains. Whether the remains are archaeological or forensic, estimating the sex and age of the individual is necessary for further analysis and interpretation. Specifically, in a medicolegal context, knowing the biological sex and approximate age of the remains assists law enforcement or government agencies in identifying unknown individuals. Since the inception of the field of forensic anthropology, practitioners have been developing methods to perform the aforementioned tasks. It is crucial that these methods be consistent, repeatedly tested, validated, and improved for multiple reasons. Firstly, to conform to *Daubert* (1993) standards, and additionally, to make sure that they are accurate and applicable to modern forensic cases. The present study was performed to validate the efficacy of the method for estimating sex from the *os pubis* originally proposed in Klales *et al.* (2012), as well as the efficacy of the “transition analysis” method for estimating age, originally outlined in Boldsen *et al.* (2002).

Considering the recent popularity of using these methods to create a biological profile for forensic cases, it is necessary to develop error rates on a large, modern, American autopsy sample. These two methods are not only being readily utilized, but are additionally being taught to students in training. The utilization of these models involves

a “logistic regression model” created by Klales *et. al* (2012) to process ordinal scores, and the Bayesian statistics software program “ADBOU” that was created for processing data collected using the method in Boldsen *et. al* (2002). These statistical systems which produced age estimates are relatively young compared to methods developed for the same purpose. The new generation of forensic anthropologists is fully responsible for objectively critiquing and validating these methods that are being disseminated by their professors and senior practitioners. The goal of the present study is to do just that. A skeletal reference sample of 630 pubic bones, all removed from modern autopsy cases and housed at the Maricopa County Forensic Science Center in Phoenix, Arizona, was utilized for data collection in the present study. Each pubic bone was assessed and scored according to the exact instructions outlined in the materials for each method, which was the Klales *et al.* (2012) paper for sex estimation, and the UTK Data Collection Procedures for Forensic Skeletal Material 2.0 for age estimation (Langley *et al.* 2016). Additionally, the observers recorded their “gestalt” estimates for sex using the Phenice (1969) system, as well as Brooks and Suchey (1990) and Hartnett (2010a) phases for each pubis. Demographic information labels were hidden, and the collection demographic information was not viewed until the completion of data collection.

The null hypothesis in the present study is that both methods (the Klales *et al.* method (2012) and “transition analysis” method (Boldsen *et al.* 2002) will perform as well as they did in the original studies. The alternate hypothesis is that they do not result in the same accuracy rates reported in the original studies. Statistical analysis of the data indicates that there is sufficient evidence to reject the null hypothesis as it applies to the

Klares *et al.* (2012) method. The classification accuracies achieved applying the logistic regression equation to the sample of pubic bones was found to be significantly lower than reported in the original study (86.2%), averaging around 70% between observers. The level of both intraobserver and interobserver agreement was only moderate for this method. It was also found that asymmetry occurred in some individuals, producing differing estimates of sex when the left and right pubes were scored separately. When utilizing the Boldsen *et al.* (2012) method and the ADBOU software package on only pubic symphyseal components to estimate age, the method was found to perform reasonably well. The majority (about 82%) of individuals had actual ages at death that fell within the predicted range produced by the statistical analysis. The majority of the symphyseal component scores showed moderate to good levels of interobserver agreement, and the estimated maximum likelihood (point estimate) of age at death predicted by the software package correlated moderately well with the actual age of death of the individual. These methods did not perform as well as reported in the original studies, and they should be further validated and recalibrated to improve their accuracy and reliability.

TABLE OF CONTENTS

TITLE.....	i
COPYRIGHT PAGE.....	ii
READER APPROVAL PAGE.....	iii
ACKNOWLEDGMENTS	iv
ABSTRACT.....	v
TABLE OF CONTENTS.....	viii
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1: INTRODUCTION.....	1
CHAPTER 2: PREVIOUS RESEARCH.....	10
CHAPTER 3: METHODS.....	31
CHAPTER 4: RESULTS.....	42
CHAPTER 5: DISCUSSION.....	57
CHAPTER 6: CONCLUSIONS	73
APPENDIX: EXEMPLARS AND SOFTWARE.....	85
REFERENCES	91
CURRICULUM VITAE.....	97

LIST OF TABLES

Table	Title	Page
3.6.1.	Experimental sample age distribution	40
4.3.1.	Classification accuracies for sex estimation	52
4.3.2.	Interobserver agreement for sex estimation	52
4.3.3.	Intraobserver agreement for sex estimation	52
4.4.2.	Component correlation with age for left pubes	54
4.4.3.	Component correlation with age for right pubes	54
4.4.4.	Interobserver agreement for age estimation	54
4.4.6.	Intraobserver agreement for age estimation	56
5.3.2.	Trait score frequencies for Kiales <i>et al.</i> (2012)	69

LIST OF FIGURES

Figure	Title	Page
2.4.1.	Metric methods for estimating sex	30
3.6.2.	Experimental sample age distribution bar graph	41
4.4.1.	Scatterplot for actual and estimated age	53
4.4.5.	Interobserver agreement scatterplot	55
5.3.1.	Medial aspect of male pubis	68
5.3.3.	Example of male ventral arc on pubis	69
5.3.4.	Comparison of male and female ventral arcs	70
5.4.1.	Ventral symphyseal margin component	71
5.4.2.	Dorsal symphyseal texture component	71

CHAPTER 1: INTRODUCTION

1.1. Forensic Anthropology and the Biological Profile

Human remains are sometimes concealed or deposited in locales where they are not discovered shortly after the time of death, and if this is case, they will eventually skeletonize (Algee-Hewitt 2013). The field of forensic anthropologists was created out of necessity – physical anthropologists were called on by law enforcement to recover and identify these skeletal remains in a modern, medicolegal context (Işcan 1988).

Information such as age, sex, ancestry, and stature, for example, are components of the biological profile which are utilized in order to identify unknown individuals (Brues 1958; Algee-Hewitt 2013). This process of estimating sex and age at death from skeletonized remains has been a task allocated to anthropologists since the early beginnings of the field (Stewart 1948, 1954). Unlike paleodemographers and bioarchaeologists who have more leeway with their estimates and biological profiles (Wright and Yoder 2003), forensic anthropologists face a unique challenge, because their estimates of such demographic data must be as accurate as possible to ensure the deceased individual is properly identified (Brues 1958; Algee-Hewitt 2013).

Additionally, a forensic anthropologist may be called into court to testify about the identity of the individual, and after the *Daubert* (1993) precedent, the methods utilized by expert witness to reach their scientific conclusions must be based on research which is peer-reviewed and widely accepted by the field (Algee-Hewitt 2013).

When creating a biological profile in a medicolegal context, estimating the biological sex and age at death of an individual is crucial (Algee-Hewitt 2013). If no

associated artifacts or belongings are found with the remains, this data must be ascertained solely from the present skeletal elements, even in cases where the remains are not complete (Brues 1958; Algee-Hewitt 2013). DNA analysis is being used in some jurisdictions to make identifications of skeletal remains, but it has several limitations (Komar and Buikstra 2008). Sample degradation, contamination, mixed samples, sample quantity, laboratory backlogs, and high costs for example, can limit the accuracy and plausibility of DNA results (Komar and Buikstra 2008). Additionally, DNA can only provide limited information about an individual if no familial sample exists for comparison (Komar and Buikstra 2008). So while skeletal analyses then seem preferred, the remains which forensic anthropologists must analyze are often fragmentary or incomplete, making some sex and age indicators unavailable (Komar and Buikstra 2008; Langley *et al.* 2016). However, using multivariate analyses that consider as many skeletal elements as possible and available can alleviate some of these concerns, and is becoming the preferred way to estimate sex and age for skeletal biologists (Milner and Boldsen 2012). It follows that methods for estimating sex and age from as many osseous elements as possible should be a part of any forensic anthropologist's repertoire, although they must have known rates of accuracy and reliability in order to be admissible in a court of law if a circumstance should call for scientific testimony (Komar and Buikstra 2008).

1.2 The Role of Sex Estimation

Sex estimation is an important first step in creating a biological profile, because it immediately eliminates half of the population for identification purposes (France 1998).

Various elements of the human skeleton have varying levels of sexual dimorphism, and in turn, are used by forensic anthropologists to estimate sex when they receive remains to analyze (Buikstra and Ubelaker 1994; France 1998). Both metric and morphological methods for estimating sex have been devised, which rely on statistics and measurements, as well as visual observation of the skull (Giles and Elliot 1963; Acsádi and Nemeskéri 1970; Buikstra and Ubelaker 1994; Jantz and Ousley 2005; Walker 2008). Additionally, non-metric morphological characteristics of post-cranial elements (Walker 2005; Vance *et al.* 2011), and postcranial measurements (Jantz and Ousley 2005; Spradley and Jantz 2011, Tise *et al.* 2013) have found to have high classification accuracies when estimating biological sex. However, non-metric, morphological analysis of the pelvis has emerged as the preferred method for determining sex due to its marked sexual dimorphism resulting from its role in childbirth and locomotion (Buikstra and Ubelaker 1994; Komar and Buikstra 2008). The Phenice (1969) system was the first to use macromorphoscopic traits of the pubis for sex estimation, specifically the ventral arc, the subpubic concavity and the subpubic ridge on the medial ischiopubic ramus (see A.1. in Appendix). This system described the various states of pelvic morphology which could be visually assessed as more appearing to be more consistent with female or male morphology. In Buikstra and Ubelaker (1994), these Phenice traits were assigned ordinal scores ranging from one to three, based on whether these morphological areas appeared to be clearly male (a score of three), clearly female (a score of one), or more ambiguous (a score of two).

There has been a shift in the field of forensic anthropology in recent decades towards using more standardized, metrical or statistical methods for creating a biological

profile in order to minimize error and bias, and to improve interobserver reliability, including methods for estimating sex (Işcan 1988; Komar and Buikstra 2008). The Klales *et al.* (2012) system was created for this reason. In this 2012 study, the authors updated the description of the Phenice (1969) traits in an effort to account for a wider range of variation, as well as the scoring system from Buikstra and Ubelaker (2004) which were developed to analyze them. This study was conducted on a sample of 310 individuals, drawn mostly from the Hamman-Todd Collection, with some individuals from the William M. Bass Donated Collections, and re-examined the three macromorphoscopic traits of the pubis as described in Phenice (1969). The original three-score system from Buikstra and Ubelaker (1994) was modified to account for the wider scope of morphological variation in Klales *et al.* (2012), and a new five-score ordinal system was devised to allow scoring of the three pubic traits. In addition, a logistic regression equation was created to process these scores, and produce a posterior probability of male and female group membership. There are additional linear discriminant function equations that can produce posterior probabilities using six different models that have yet to be published (available online at <http://math.mercyhurst.edu/~sousley/Software/>). This method has become popular amongst some practitioners of forensic anthropology, and is actively being taught to students (Klales 2014; Call 2016). Validation studies are therefore required to ensure this method performs well enough to be used in medicolegal cases, considering the well-documented phenomenon of secular change, as well as ancestral differences, which can affect the appearance of these sex indicators and introduce error into forensic anthropological methods (Klales 2014; Call 2016).

Accessing new, modern skeletal collections which better represent the current living population that will become future forensic cases and testing pre-existing methods on them will be important to ensure proper peer review and a better understanding of error rates and what can be improved (Hartnett 2010a; 2010b).

1.3 The Role of Age Estimation

Furthermore, age estimation is a crucial step in the biological profile, because it further narrows the possible group of people from which to make an identification (Algee-Hewitt 2013). It is also considered the most difficult piece of information to determine from the skeleton, because it relies on skeletal metamorphoses which occur with age, and can take place at different times in different individuals (Işcan 1988). Systems for age estimation have been decades in the making, starting most prominently with Todd (1920, 1921), who recognized areas of the skeleton that seemed to build up and break down as a person aged. His research has been expanded and applied to areas throughout the entire skeleton, which include various epiphyses and joints, dentition, cranial sutures, sternal rib ends, and osteoarthritis and osteophyte formation (Buikstra and Ubelaker 1994; Brooks and Suchey 1990; Suchey 2006; Webb and Suchey 1985; Harnett 2010b; DiGangi *et al.* 2009; Brennaman *et al.* 2016). These methods look to correlate a specific morphological appearance of the area of interest with an age at death in a “reference sample”, defined by Bocquet-Appel and Masset (1982) as the collection or sample of skeletons (or skeletal elements) with a known age at death distribution, to which unknown individuals are compared. These reference samples are studied in order

to create a set of useful age ranges from the morphology and known age distribution, in hopes that this information can be applied to future forensic cases (Boldsen *et al.* 2002). Unfortunately, phenomena such as “age mimicry” or “attraction to the middle” plague paleodemographers and skeletal biologists who are interested in estimating the age at death for individuals of unknown age (Boldsen *et al.* 2002). Validating and re-calibrating methods as new skeletal collections become available is important to ensure this problem is compensated for.

The pubic symphysis is a joint that has been widely used estimate age at death (Todd 1920; Brooks and Suchey 1986; Buikstra and Ubelaker 1994), and due to its robust survivability and its seemingly consistent good correlation with age, it persists as the preferred bone to use for estimating sex for bioarchaeologists and forensic anthropologists (citation). The pubic symphysis is a cartilaginous (hyaline and fibrocartilage) joint at which the two pubes of the left and right os coxa meet. The osseous faces of this joint are epiphyses that build up, change, and degenerate with age. Although the timing of changes can vary from person to person, the changes tend to proceed in a predictable, unilineal direction (Boldsen *et al.* 2002). McKern and Stewart (1957) were the first to isolate various “components”, or specific areas of the pubic symphysis and see how they specifically correlated with age, versus using the overall morphology of the pubic symphysis. In the 1980s, Bayes’ theorem, a type of conditional probability that uses known information about a reference sample to offer information about the “real world”, began to be widely used (Hoppa and Vaupel 2002). Forensic anthropologists quickly realized that these Bayesian statistics could be applied to the

components they had isolated for age estimation, and began working on new methods combining the two systems.

It was in Boldsen *et al.* (2002) that Bayes' theorem was first applied to age estimation from skeletal components as originally proposed by McKern and Stewart (1957). The premise of this new method was that it could be applied to any osteological structure that experienced senescent changes that proceed in a predictable sequence (Boldsen *et al.* 2002). The original 2002 study utilized a sample from the Terry skeletal collection housed at the Smithsonian Institution and the Portuguese University of Coimbra collection, and analyzed three groups of osteological structures: the pubic symphysis, the sacroiliac joint of the ilium, and cranial sutures, and their various components. Statistical analyses can be applied to the raw scores for each of these traits, producing an age interval, as well as a point estimate (Langley *et al.* 2016). This new method for age estimation was given the name "transition analysis", meaning a system that uses the timing at which certain skeletal areas "transition" between two senescent stages of morphology to create age ranges (Boldsen *et al.* 2002; Langley *et al.* 2016). The validation of this transition analysis method conducted by Milner and Boldsen (2012) had a sample size of 252 individuals, all drawn from the Bass Donated Collection and Mercyhurst University Forensic Cases, and sought to test the efficacy of the previously developed method of transition analysis. They found that the known age distribution of the population did affect the resulting age estimates, but that the estimates were still more imprecise than those created simply from observational experience. They noted in this study that while these component-based methods for age estimation were very promising,

some features (such as cranial sutures) do not provide useful information, and other skeletal structures should be tested for their utility in a component-based framework. A transition analysis statistical software package which uses Bayesian statistics to produce age ranges from components, called ADBOU, will eventually be integrated into the next version of the FORDISC software (a discriminant function program; Ousley and Jantz 2005) utilized by the majority of the field to predict sex, ancestry, and stature (Langley *et al.* 2016). The research outlined in this document is another important step in the validation of this method which is only recently burgeoning in popularity and is already being utilized by many forensic anthropology practitioners.

1.4 The Current Study

The purpose of the present study is to assess each of the aforementioned methods using a previously untested, modern autopsy sample of pubic bones. The sex estimation method from Klales *et al.* (2012) has shown great promise, but secular change has already shown to affect the expression of those pubic traits utilized in the study (Klales 2014). While component-based methods have great promise, the osteological structures used in transition analysis (specifically the ADBOU software program) were shown to have varying levels of utility for age estimation, especially cranial sutures which produce severely negatively-biased age estimates (Milner and Boldsen 2012). In order to understand the validity and reliability of these methods, the pubic bones from modern autopsy cases at the Maricopa County Forensic Science Center were utilized. These remains represent a modern reference sample upon which to test and observe the

expression of the Phenice (1969) traits and the efficacy of Klales *et al.* (2012). In addition, it provides an opportunity to ascertain whether the components of the pubic symphysis do produce useful age estimates on their own, and whether they should be considered in conjunction with other osteological structures for future component-based methods. Based on recent validation studies (Klales *et al.* 2014; Milner and Boldsen 2012), it is hypothesized that the methods will not perform as well as reported in the original studies.

CHAPTER 2: PREVIOUS RESEARCH

2.1 An Overview of Sex Estimation

Sexual differences in the skeleton have been studied for decades, with early physical anthropologists being intently focused on cranial morphology and its correlation with race and sex. Practitioners such as Hrdlička, Stewart, and Krogman claimed they could estimate sex accurately 90 percent of the time by just visually assessing cranial morphology (Stewart 1948; Krogman 1962; Giles and Elliot 1963). However, much of this early work relied on a few experienced observers using their own knowledge to produce their best guesses (Stewart 1954; Giles and Elliot 1963). This focus on the skull has persisted, but has become more standardized, and the cranium is still utilized by the majority of anthropologists for sex estimation whenever it is available (Buikstra and Ubelaker 1994; France 1998; Konigsberg and Hens 1998; Rogers 2005; Walker 2008; Spradley and Jantz 2011). Morphological traits of the skull and their sexual differences have been validated time and time again, and the most commonly utilized ones in modern cases include the glabella, mastoid process, supraorbital margin, mental eminence, and nuchal crest (Acsádi and Nemeskéri 1970; Buikstra and Ubelaker 1994; Walker 2008). Males tend to have more robust cranial morphology, and females more gracile; in males, the glabellar region is more prominent, the mastoid process longer and wider, the supraorbital margin rounder, the mental eminence wider, and the nuchal crest more pronounced (Acsádi and Nemeskéri 1970). An ordinal scoring system has been established (Buikstra and Ubelaker 1994; Walker 2008) for these traits, with clear female morphology scored a one, and clear male morphology scored a five, and absent or

unobservable traits scored a zero (see A.2. in Appendix) (Buikstra and Ubelaker 1994). Walker (2008) applied discriminant function analysis to these ordinal scores used on the skull to provide statistical probabilities of male or female group membership.

Originating with the influential work of Giles and Elliot (1963), metric data collected from the cranium has also been used to estimate sex in bioanthropological contexts. The Giles and Elliot method (1963) utilized 11 interlandmark distances from the cranium to create various multivariate linear discriminant function equations. These equations produce a posterior probability of a group membership when the standardized craniometrics from an unknown individual are plugged in. Desktop statistical packages have made this discriminant function approach more efficient, and therefore a preferred method for many biological anthropologists (Ousley and Jantz 2005). Reference samples of various populations are available in a software program called FORDISC, and measurements taken from the crania of unknown individuals can be entered into the software, resulting in a statistically-based classification of biological sex, using posterior probabilities (Ousley and Jantz 2005). More recent studies into geometric morphometrics and three-dimensional CT scanning have investigated their application in sex estimation, and show great promise, although the technology is still expensive and not readily available to most practitioners (Franklin *et al.* 2005; Ramsthele *et al.* 2010).

The postcranial skeleton has also been studied in depth for indicators of sex (see Figure 2.3.1.) (France 1998; Komar and Buikstra 2008). Nonmetric methods using visual assessment of the morphology of certain postcranial areas, and comparing them to male or female exemplars (France 1998). One such method using the distal

humerus has been established, which includes assessing the trochlear shape, symmetry, and angle (Rogers 1999). Rogers (1999) found that females have more spool-shaped, symmetrical, and less angled trochlea than do male individuals. Another study used a South African reference sample to assign ordinal categorical scores to the shape of the olecranon fossa, the angle of the medial epicondyle, and the degree of trochlear extension (Vance *et al.* 2011). These data are then summed to give an estimate of sex, with a score of three to eight indicating male morphology, nine indicating ambiguous morphology, and 10 to 15 indicating female morphology. A study by Rogers *et al.* (2000) investigated a morphological structure called the rhomboid fossa, which is a pit for the attachment of the costoclavicular ligament on the inferior clavicle. This study found that the rhomboid fossa is more consistent with male morphology and is not usually present in females; statistical analyses of the reference sample in this study indicate that a fossa on the right clavicle is indicative of a male with 81.7percent probability; a fossa on the left is indicative of a male with 92.2 percent probability (Rogers *et al.* 2000). Suchey *et al.* (1979) looked at a sample of 486 females to see how common dorsal parturition pits on the pubis were, and found that they are quite common in females over the age of 30 years who have carried a full term pregnancy. More recent studies are looking to validate if this is a good indicator of sex, and have found that males do not present with deep, dorsal parturition pits as females do (McArthur *et al.* 2016).

There have also been statistical approaches developed to analyze standardized measurements of the postcranial elements (Buikstra and Ubelaker 1994) for the purpose of estimating biological sex. Standardized postcranial metrics have been demonstrated to

produce accurate sex estimation results when entered into the FORDISC software package, which compares them to various reference populations (or samples) and will also provide a posterior probability of male and female group membership (Ousley and Jantz 2005). Simpler metric methods for the postcranial skeleton exist as well, such as those introduced in Albanese *et al.* (2005), Spradley and Jantz (2011), Tise *et al.* (2013), and Moore *et al.* (2016). Univariate methods like these that involve simply calculating an overall average (combining males and females) of a single postcranial metric from a reference sample of at least 30 individuals can produce a “sectioning point”, or a threshold by which to determine the sex of an individual (Albanese *et al.* 2005; Spradley and Jantz 2011; Tise *et al.* 2013; Moore *et al.* 2016). If the measurement of an individual falls below the sectioning point, they can be classified as female, and if it falls above, they can be classified as males (Albanese 2005). Those with measurements that are equal to the sectioning point would be classified as sexually ambiguous morphology (Albanese *et al.* 2005; Spradley and Jantz 2011). This method has been particularly useful in estimating sex in populations that have not been studied in depth due to lack of documented skeletal collections (i.e. Latin American populations), because it requires a relatively small reference sample and no statistical analysis or complicated software (Albanese *et al.* 2005; Spradley *et al.* 2008; Spradley and Jantz 2011; Tise *et al.* 2013). The classification accuracies for these methods have been high enough, some exceeding 95 percent, for them to be useful in estimating biological sex (Spradley and Jantz 2011).

Although the skull and postcranial skeleton can be utilized for sex estimation, the pelvis has long been established as having the best morphological indicators of sex

(Buikstra and Ubelaker 1994; Phenice 1969; France 1998; Kjellstrom 2004; González *et al.* 2007). The pelvis is a very sexually dimorphic structure in humans because of its role in childbirth (Komar and Buikstra 2008). During the pubertal growth spurt, the os coxae of females grows in a way which allows the pelvic inlet to expand, creating a very dimorphic morphology that is easily distinguished from males (Buikstra and Ubelaker 1994; Komar and Buikstra 2008). The morphology of the greater sciatic notch of the pelvis has been studied in depth and has been found to correlate with sex due to these differences in pubertal growth, with females tending to have much wider notches than males (Buikstra and Ubelaker 1994; Walker 2005). In addition, a feature called the preauricular sulcus has been thoroughly investigated for its correlation with biological sex (Milner 1992; Buikstra and Ubelaker 1994). It appears as a groove or furrow inferior to the auricular surface of the ilium, and its presence is consistent with female morphology while males very rarely have a sulcus (Milner 1992; Buikstra and Ubelaker 1994).

T.W. Phenice (1969) published the first study that recognized the value of three specific traits of the os pubis as sex indicators: the ventral arc, the subpubic concavity and the medial aspect of the ischiopubic ramus (see A.1. in Appendix). The ventral arc is a diagonal ligamentous outgrowth that occurs on the ventral surface of the pubis and comes off of the upper margin of the symphyseal face. Phenice (1969) noted that women tended to have prominent ventral arcs had very small ones, or completely lacked one altogether. In The subpubic concavity is a drastic lateral recurve of the ischiopubic ramus below the inferior rim of the symphyseal face of the pubis. Females have this recurve while males

usually do not (Phenice 1969; Buikstra and Ubelaker 1994). The medial aspect of the ischiopubic ramus is the trait that should be given the least weight according to Phenice (1969). Males more commonly present with a broad ramus compared to their symphyseal face, while females have thin, pinched rami. Additionally, women more often have a raised ridge of bone extending off the inferior rim of the symphyseal face and down the ramus, while males do not (Phenice 1969; Buikstra and Ubelaker 1994). Using this system, sex would be assessed based on whether or not the three areas had morphology that more closely resembled the female exemplars or the male exemplars. This landmark study, while having important implications for sex estimation, relied entirely on visual assessment with no quantitative or statistical weight. Buikstra and Ubelaker (1994) added an ordinal scoring system to the Phenice (1969) traits in order to make them slightly less subjective: a score of one representing clear female morphology, two being ambiguous morphology, and three being clearly male morphology. All three scores would then be considered together to make an overall sex assessment for the individual (Buikstra and Ubelaker 1994). The Klales *et al.* (2012) study updated the Phenice (1969) system and Buikstra and Ubelaker (1994) scoring system by creating five ordinal scores for these same traits on the os pubis, accounting for more morphological variation (see A.3. in Appendix). This newer study utilized a sample of 310 individuals from the Hamman-Todd collection (Klales *et al.* 2012) to develop its updated morphological descriptions of the three Phenice (1969) traits. The publication was accompanied by a user-friendly logistic regression equation in the form of a Microsoft Excel spreadsheet, which allows the observer to enter their scores for each trait and generate a posterior probability of

either male or female group membership (Klales *et al.* 2012). The results reported in this study indicate that the cross-validated classification rate for all observers was 94.5 percent, and when tested on an independent validation sample, the classification accuracy was 86.2 percent (Klales *et al.* 2012).

The validation studies of the Klales *et al.* (2012) method have been limited, and further research has been published investigating the effects of secular change on the Phenice (1969) traits, acknowledging that older, historical samples will have different distributions of trait scores than more modern samples (Klales 2014). Klales (2014) used a sample of 170 individuals from the Hamman-Todd Collection (a historical sample) and 129 individuals from the William Bass donated skeletal collection (a modern donated sample) to re-test the Klales *et al.* (2012) method. When the method was applied to this different sample, the classification accuracy was only 68.7 percent, much lower than reported in the original study (Klales 2014). A more recent validation study produced different results when utilizing the Klales *et al.* (2012) method. Call (2016) validated the method on a sample of 204 individuals from the Bass collection, and found that the method had a 93.6 percent classification accuracy. That classification accuracy jumped to 99 percent when the original logistic regression equation was re-calibrated using the new sample from the Bass collection (Call 2016). Another validation study tested the method against a Mexican sample of 203 left os coxae (Gómez-Valdés *et al.* 2017). The original Klales *et al.* (2012) method was found to have a sex bias, with females being correctly classified 100 percent of the time, and males only being correctly classified 86 to 92 percent of the time (Gómez-Valdés *et al.* 2017). The authors also reported that after the

logistic regression was re-calibrated with the new sample, it produced 100 percent classification accuracy (Gómez-Valdés *et al.* 2017). These validation studies that have been performed were performed on small sample sizes, and many using the Bass collection (Klaes 2014; Call 2016; Gómez-Valdés *et al.* 2017).

2.2 An Overview of Age Estimation

Estimating chronological age at death from skeletal remains is another crucial part of creating a biological profile, especially for forensic anthropologists working on forensic cases, and biological anthropologists attempting to answer paleodemographical questions (Konigsberg and Frankenberg 1992; Kemkes-Grottenthaler 2002; Algee-Hewitt 2013). In forensic contexts, age estimates provided by forensic anthropologists from skeletal markers aid law enforcement and government agencies in making identifications of unknown individuals (Algee-Hewitt 2013). Estimating the age of juvenile and subadult individuals is able to be done with better accuracy than adults, due to the much narrower age ranges within which important skeletal metamorphoses occur (Komar and Buikstra 2008). However, estimating chronological age for adult individuals is still important, and can be done with some degree of accuracy by experienced practitioners (Algee-Hewitt 2013). Although individual life history, environmental stress, genetics, behavior, and other influencing forces can shape how a skeleton senesces, scientists are still able to estimate chronological age at death from a skeleton with reasonable accuracy, although biological age does not always match chronological age (Kemkes-Grottenthaler 2002). Derived estimates of age using the skeleton represent “subjective assessments” of an

individual's physiological state, and are based on tenuous correlations with senescent skeletal changes and the calendar years that have passed since an individual's birth (Algee-Hewitt 2013). Age estimation is also a contentious issue in skeletal biology and its related fields because of the subjective nature of creating age ranges from skeletal degeneration, as there exists marked heterogeneity in the timing and appearance of these changes in adults (Algee-Hewitt 2013). In fact, in their 1982 paper entitled "Farewell to Paleodemography, Bocquet-Appel and Masset went so far as to claim that estimating age from the skeleton was not able to be done with enough accuracy, and the practice should be abandoned altogether. Forensic anthropologists would disagree, because age at death can be invaluable to making an identification (Komar and Buikstra 2008; Algee-Hewitt 2013), and these issues of bias and validity can be accounted for and improved by recalibration of statistical methods (Konigsberg *et al.* 1997). It has been suggested that a multivariate approach to estimate age using as many skeletal markers as possible is the most thorough and accurate procedure (Komar and Buikstra 2008). The issues that exist within the bevy of age estimation methods continue to be reliability (interobserver and intraobserver agreement) and validity (correlation between the age indicator and the actual age at death), but those the two issues are something that can be improved through continual research and improvement of the methods we use (Konigsberg *et al.* 1997; Kemkes-Grottenthaler 2002).

Although other anatomists had previously written of observations of age-related degenerative changes to skeletal joints, T. Wingate Todd was the first to conceptualize a prescribed method for estimating biological age at death from skeletal elements (Todd

1920; Todd 1921; Kemkes-Grottenthaler 2002; İşcan and Loth 1989). He studied the symphyseal face of the pubis in American Black males, and American White males and females, noticing that it went through a “metamorphosis” as a person aged (Todd 1921). He treated it as an epiphysis that builds up and breaks down with age, and created ten phases based on these morphological appearances and at what ages they tended to occur (Todd 1920, Todd 1921; Algee-Hewitt 2013). The first phase found in the youngest adults was characterized by solid, good quality bone, no osseous rim or rampart formed, and billowing on the symphyseal face. The last phase in older adults was characterized by loss of bone density, porosity, depression of the symphyseal face, and an eroding rim (Todd 1921). It was therefore conceptualized by Todd that an individual of an unknown age could be assessed and given a phase using the morphological categories he had described, and therefore be assigned to that age range (Todd 1920).

This principle has been applied to various other skeletal elements in hopes that age can be estimated even when the pubic symphysis is not available (Komar and Buikstra 2008). These more recent age estimation methods can be sorted into two categories: phase-based methods (like Todd’s 1921 system) and component-based methods (Shirley and Ramirez Montes 2015). Phase-based methods classify the morphology of the entire area being considered, and a phase is assigned (usually an ordinal score) which has an attached age range (Todd 1921; Shirley and Ramirez Montes 2015). Component-based methods consider various, discrete morphological aspects of the overall area being considered and scores them separately, so they can then be considered in conjunction with each other to produce an age estimate (McKern and Stewart 1957;

Boldsen and Milner 2002; Kemkes-Grottenthaler 2002; DiGangi *et al.* 2009; Shirley and Ramirez Montes 2015). Each type of system has advantages and disadvantages.

Component-based methods are more time-consuming and complex than are phase-based methods, but allow a wider amount of variation in senescent changes to be accounted for than phase-based methods (McKern and Stewart 1957; Boldsen *et al.* 2002; Kemkes-Grottenthaler 2002).

In younger individuals, age estimation methods tend to produce narrower and more accurate age ranges, because the timing of growth and skeletal changes is more predictable – reaching full skeletal maturity occurs in less time than does the breakdown and degeneration of bony tissue (Buikstra and Ubelaker 1994). The appearance of primary ossification centers, as well as the measurement of long bones is a useful method to apply to prenatal and neonatal individuals, as the timing of those growth events tend to occur within a narrow range of a few weeks in the majority of individuals (Scheuer and Black 2009). In infants and juveniles, the lack or presence of secondary ossification centers (or epiphyses), as well as the state of their fusion to the metaphysis of the bone can be very useful in estimating narrow age ranges for subadults (Scheuer and Black 2009). In juvenile and subadult individuals, an epiphysis will either not be present, present but not fused, in early stages of fusion, actively fusing, almost completely fused, completely fused with fusion scars, or completely fused with complete obliteration of the fusion site (Buikstra and Ubelaker 1994; Scheuer and Black 2000; Scheuer and Black 2009). In juveniles and older sub-adults when epiphyses have already ossified, the fusion becomes crucial to age estimation. There are established methods and age ranges for

epiphyseal fusion for many areas of the skeleton in juveniles and younger adults, including the anterior iliac crest of the os coxa (Webb and Suchey 1985), the medial clavicle (Webb and Suchey 1985, Langley-Shirley and Jantz 2010), the ring epiphysis of the thoracic and lumbar vertebral bodies (Albert and Maples 1995), the first sacral segment (Scheuer and Black 2009), the spheno-occipital synchondrosis (Shirley and Jantz 2011), and the long bones (Suchey 2006). The appearance and formation of the dentition in juveniles is also useful in estimating age, with very narrow age ranges (Buikstra and Ubelaker 1994). Methods have been developed that can age juveniles based on what dentition has erupted or not erupted (Buikstra and Ubelaker 1994), or by radiographically viewing the roots of the third molar and seeing how complete they are (Mincer *et al.* 1993).

In adults, after the major epiphyses have fused and all permanent dentition has erupted, age at death becomes more difficult to predict, and the age ranges can get wider (Algee-Hewitt 2013). Once a juvenile becomes a fully grown adult, age estimates rely on degenerative changes to the epiphyses, joints, dentition, and other areas of the body (Komar and Buikstra 2008). Cranial sutures are the places where the various cranial bones meet during craniofacial growth and development (Acsádi and Nemeskéri 1970; Meindl and Lovejoy 1985). Due to the focus on the skull in early physical and forensic anthropology in the early part of the 20th century, these sutures some of the first skeletal traits to be assessed for age related changes (Todd and Lyon 1924; Montagu 1938). The closure of these sutures and their eventual obliteration were investigated for a correlation with age first by Todd and Lyon (1924). Several studies since have continued

investigating if the timing of the closure and obliteration of these sutures can correlate with age in adults, specifically focusing on the sutures of the cranial vault and the maxilla (Brooks 1955; McKern and Stewart 1957; Meindl and Lovejoy 1985; Mann *et al.* 1991; Komar and Buikstra 2008). Meindl and Lovejoy (1985) created a component-based system by which the anterior-lateral and vault sutures of the cranium would be scored based on their level of closure and added together for a composite score which was associated with an age range. Mann *et al.* (1987) applied an identical scoring system to the sutures of the maxillary palate. In a study by Milner and Boldsen (2012), it was found that for the purpose of estimating age at death, cranial sutures performed “poorly in terms of both age estimate accuracy and precision”. Cranial suture closure is considered to have a tenuous, at best, correlation with age, but they are still utilized by many forensic anthropologists (Komar and Buikstra 2008).

Other epiphyses have been repeatedly studied for their correlation with, and use to create both phase and component-based methods for age estimation (Boldsen *et al.* 2002; Komar and Buikstra 2008). The ribs and their articular surfaces are one such area (Işcan 1989a; Işcan 1989a; DiGangi *et al.* 2009; Hartnett 2010b). Sternal rib ends, which articulate with costal cartilage in the medial area of the anterior thorax, were previously noted to undergo age-related generation, but were first recognized as valuable age at death indicators in Işcan *et al.* (1984, 1985). The authors devised a phase-based system (with 10 phases total) for estimating age from the sternal end of the fourth rib, using a sample of White males (Işcan *et al.* 1984) and White females (Işcan *et al.* 1985). This system by Işcan *et al.* (1984; 1985) was updated using a larger, more modern sample

(630 individuals from forensic autopsy cases) by Hartnett (2010b). The results of this study indicated that the best correlations with age found using the sternal end of the fourth rib were the depth of the pit, the regularity of the rim edges, and the bone quality (Hartnett 2010b). Another recently developed method utilizes a component-based approach, involving different articular areas of the first rib (DiGangi *et al.* 2009). Developed on a Balkan male sample, the components used for age estimation in this study included the rib head, the costal face, and the tubercle facet, and each is scored individually, then a composite score is created which has an associated age range (DiGangi *et al.* 2009). The use of the sternal end of ribs especially is still a popular method amongst forensic anthropologists (Garvin and Passalacqua 2011).

The auricular surface of the ilium has also been tested in multiple studies for its validity in estimating chronological age at death (Lovejoy *et al.* 1985; Buckberry and Chamberlain 2002; Osborne 2004; Milner and Boldsen 2012). Lovejoy *et al.* (1985) authored the first study in which the correlation of age with degenerative and morphological changes to the auricular surface of the ilium was tested. These changes included: grain and density, macroporosity, billowing, striations, apex, retroauricular area, and transverse organization (Lovejoy *et al.* 1985). Later studies sought to create more accurate age ranges than the ones described in Lovejoy *et al.* (1985). Buckberry and Chamberlain (2002) re-examined the components originally described by Lovejoy *et al.* (1985) and created a system by which they could be assigned scores based on their appearance, and then added for a composite score which would have an associated age range. Their method was found to have lower levels of interobserver and intraobserver

error, as well as more accurate age ranges (Buckberry and Chamberlain 2002). Osborne (2004) modified the work of Lovejoy *et al.* (1985) into a phase-based system that focused on the overall morphology of the auricular surface, but the age ranges it produced, while valid, were very wide and of limited applicability. Milner and Boldsen (2012) applied Bayesian analysis to components of the sacroiliac surface and found that it tended to underestimate age in people who were at least 60 years old at the time of their death. Despite varying validity of methods, the auricular surface remains an often-used area for age estimation amongst forensic anthropologists (Algee-Hewitt 2013).

In addition to the aforementioned methods, some systems for age at death estimation have been developed based on adult dentition (Algee-Hewitt 2013). Gustafson (1950) was the first study to note six stages of attritional change to permanent adult teeth which progressed with age. Brothwell (1989) actually created a phase-based method using observations of such attrition in an adult human sample. The system generally considered maxillary and mandibular teeth together, had “tentative” age ranges, and was based on a Neolithic and Middle Age sample, which limits its utility in modern forensic cases (Brothwell 1989). The concept of aging from the dentition has been re-visited and modified by other scientists (Lamendin *et al.* 2002; Prince *et al.* 2008; Prince and Konigsberg 2008b). Lamendin *et al.* (2002) used the translucency of the apical root dentin and periodontal recession in single-rooted teeth to estimate age at death. This concept was later modified, and Bayes’ theorem was applied to the root dentin translucency, and new formula to estimate age were created (Prince and Konigsberg 2008). The application of Bayesian statistics to root translucency and alveolar resorption

produced lower error rates, less bias, and overall higher correlation with age than the original Lamendin *et al.* (2002) study (Prince and Konigsberg 2008b). Bayes' theorem was also applied to dental attrition in an attempt to create a useful formula for estimating age at death (Prince *et al.* 2008). The formula utilized phase scores for the wear on the incisors and canines, as well as the premolars, which calculated a probability that an individual had reached a certain "transition" point in their degenerative, age-related morphology (Prince *et al.* 2008).

However, in the decades following Todd's (1920; 1921) original studies of degeneration to the pubic symphysis, this particular skeletal area has continued to be the preferred one for estimating age (see A.4. in Appendix) (Buikstra and Ubelaker 1994; Garvin and Passalacqua 2011; Algee Hewitt 2013). The initial 10 phases Todd (1920) created were tested over the next several decades against various other skeletal collections and found to perform reasonably well for age estimation (McKern and Stewart 1957). Brooks (1955) used the same collection as Todd (1920), and noticed different age-related patterns to the pubic symphysis. She collaborated with a statistician to predict the age at which a person would transition between these morphologies (Brooks 1955). McKern and Stewart (1957) built off this previous research by analyzing a sample of Korean War dead, and attempting to create formulae from the nine "features" of the pubic symphysis identified by Todd (1920) with the goal of producing more accurate age ranges for individuals over 20 years of age. These features were: ridges and furrows, the dorsal margin, ventral beveling, the lower extremity, the superior ossific nodule, the upper extremity, the ventral rampart, the dorsal plateau, and the symphyseal

rim (Todd 1920; McKern and Stewart 1957). McKern and Stewart (1957) combined some of the features that were more easily distinguishable into new categories (which they called “components”) to utilize for their new formulae; these were: the dorsal plateau, the ventral rampart, and the symphyseal rim. Each component would be given an ordinal score from zero (0; meaning absent) to five (5; the most degenerative state), and those scores would be used to create formulae based on Bayes’ theorem. The result was a set of 21 conditional probability equations that could be used to estimate age in unknown individuals (McKern and Stewart 1957). The same methodology was applied to a female sample in a later study by Gilbert and McKern (1973), because it was noted that female populations had a much different mortality distribution than did males.

In 1990, Suchey and Brooks used a large sample of autopsy cases to combine the original 10 phases proposed by Todd (1920) into six phases, each attached to an estimated age range (at the 95 percent confidence level). Instead of using various “components” of the pubic symphysis and scoring them separately, this method was highly visual, and considered the overall morphology of the pubis and which phase it most closely resembled (Brook and Suchey 1990). This method has persisted as the most commonly used by forensic anthropology practitioners since its publication (Garvin and Passalacqua 2011). In Hartnett (2010), the Brooks and Suchey (1990) six-phase system was modified into a seven-phase system which accounted for more variation in the expression and timing of this senescent changes to the pubic symphysis and surrounding area, each phase having a narrower age interval than the original study. This study was part of Hartnett’s dissertation, and involved the curation of a sample of 630 pubic bones

which were removed from autopsy cases performed at the Maricopa County Forensic Science Center (Hartnett 2010a). Although the original Brook and Suchey (1990) method is still widely used, the Hartnett (2010a) method has been adopted for casework by many forensic anthropologists because of its inclusion of an extra category for much older individuals (Hartnett 2010a; Garvin and Passalacqua 2011; Algee-Hewitt 2013).

In 2002, Boldsen and colleagues published a new Bayesian approach to age estimation; based on the conceptualized components originally introduced by McKern and Stewart (1957), it was given the name “transition analysis”. Though Boldsen and colleagues (2002) developed the method for the sacroiliac surface, pubic symphysis, and cranial sutures, the statistical framework is potentially applicable to any skeletal trait that transitions through a unilineal set of senescent changes (Boldsen *et al.* 2002) (an example of one such trait in the new “transition analysis” can be seen in A.5. in the Appendix). Although the timing of these changes can differ from individual to individual, the sequence tends to be fixed and linear, and age estimates can be procured by making inferences based on these timing of these transitions (Boldsen *et al.* 2002). Other age estimation methods utilizing various “components” of other osteological structures and Bayesian statistics (transition analysis) have since been developed (DiGangi *et al.* 2009; Kimmerle *et al.* 2008; Shirley-Langley and Jantz 2010; Brennaman *et al.* 2016). These methods have shown extreme promise, excepting the use of cranial sutures, which consistently produce negatively-biased age estimates (Milner and Boldsen 2012, Konigsberg 2015).

Milner and Boldsen (2012) validated their original study (Boldsen *et al.* 2002) on a sample of 252 individuals from the Bass donated skeletal collection and forensic cases from Mercyhurst University. Their results indicated that out of the three components considered in their method (cranial sutures, sacroiliac surface, and pubic symphysis), the pubic symphysis performed the best on its own (Milner and Boldsen 2012). The authors concluded that transition analysis formulae were too reliant on “commonly used” structures and that the estimates from their software did not outperform “experience-based estimates” (Milner and Boldsen 2012). They noted that the age at death distribution of the reference sample affects the resulting age estimates, which become less accurate as individuals get older (Milner and Boldsen 2012). The authors ultimately called for continued research into transition analysis methods and components that could be used in conjunction with each other to create more accurate age estimates (Milner and Boldsen 2012).

2.3 Research Goals

The purpose of the present study is to contribute to the existing body of data pertaining to the determination of sex and age in the field of forensic anthropology. Validation of previously existing methods is crucial, considering the well-documented phenomenon of secular change, and the fact that the most popular skeletal reference collections are quickly becoming historical (Algee-Hewitt 2013; Klaes 2014). Validation and peer review of newer methods is just as important, and should be performed thoroughly before these methods are widely disseminated as viable for forensic casework

(Algee-Hewitt 2013). As a student of forensic anthropology, the Klales *et al.* (2012) and Boldsen *et al.* (2012) methods for sex and age estimation were part of the curriculum as viable methods to be used in casework. However, considering the relative youth of these methods, the results of recent validation studies (Klales 2014; Milner and Boldsen 2012), and the inherent problems that come with age estimation methods in adults (Kemkes-Grottenhaler 2002), it is the obligation of new researchers to think critically about the methods they are taught, and to objectively test these methods on modern samples whenever they are available (Algee-Hewitt 2013).

2.4 Figures

BOX 5.12
The estimation of sex through measurement

Sample	Element(s)	% Correct	Reference
Terry and Coimbra	Os coxae and femur	91–98.5	Albanese, 2003
Chinese	Femur	82.3–87.2	Wu, 1989
Croatian	Femur	91.3–94.4	Slaus et al., 2003
Spanish	Femur	85.6–99.1	Trancho et al., 1997
Thai	Femur	85.6–94.2	King et al., 1998
Dart (South African black)	Calcaneus	79–86	Bidmos and Asala, 2004
Dart (South African black)	Talus	80–89	Bidmos and Dayal, 2004
Terry	Talus/calcaneus	79–89	Steele, 1976
South Italy	Calcaneus	76.3–85	Introna et al., 1998
Dart (South African black)	Patella	78.3–85	Dayal and Bidmos, 2005
Japanese	Tibia	80.3–87.2	Iscan et al., 1994
Terry	Tibia	77.2–87.3	Iscan and Miller-Shaivitz, 1984 a, b
Japanese	Tibia	95.7–96	Hanihara, 1958
Terry	Humerus	85.5–93.5	France, 1983
FDB (Tennessee)	Humerus	86–93	France, 1998
FDB (Tennessee)	Scapula	94	France, 1998
FDB (Tennessee)	Clavicle	88–96	France, 1998
FDB (Tennessee)	Radius	86–99	France, 1998
Japanese	Radius	97	Hanihara, 1958
Japanese	Ulna	88.9–90.5	Hanihara, 1958
FDB (Tennessee)	Ulna	91	France, 1998
British	Metacarpals and hand phalanges	74–94	Scheuer and Elkington, 1993
UNM documented	Metacarpals	75–90	Stojanowski, 1999
Pennsylvania	Metacarpals	63–95.7	Burrows et al., 2003
Terry, UNM documented and Royal Free Medical Society of West Africa	Metacarpals	84.4–92	Falsetti, 1995
	4th rib	74–80	Wiredu et al., 1999
Florida	4th rib	82–89	Iscan, 1985
Turkey	4th rib	86–90	Cologlu et al., 1998
Taiwan	Skull	100	Hsiao et al., 1996
Terry	Cranial base	70–85	Holland, 1986
Terry	Petrous temporal	66–74	Kalmey and Rathbun, 1996
Terry	Skull	82.9–87.9	Giles and Elliot, 1962, 1963
Japanese	Skull	83.1–86.4	Hanihara, 1959

Figure 2.4.1. A list of several different methods developed to estimate sex through some type of measurement of the skeleton – including which skeletal sample they were developed on, the element used, the accuracy rates, and the study authors (taken from Komar and Buikstra 2008).

CHAPTER 3: METHODS

3.1 Sample Description

The Maricopa County Forensic Science Center located in Phoenix, Arizona houses a collection of sternal rib ends and pubic bones. The collection was compiled in 2010 by Dr. Kristen Hartnett-McCann for her dissertation project on phase-based age estimation from the pubic symphysis and the sternal end of the fourth rib (Hartnett 2010a; Hartnett 2010b). The pubic bones were removed from modern autopsy cases that were processed at the Forensic Science Center (FSC) with permission from the next of kin (Hartnett 2010a). The collection contains 605 individuals – 409 males (67.6 percent) and 196 females (32.4 percent). The ratio is approximately 2.1:1 for males to females. Only 599 individuals had documented ages at death, which ranged from 18 to 99 years. The mean age of the sample (N=599) is 54.25 years, the median age is 52.00 years, and the mode is 48.00 years. The 25th percentile is represented by 40.00 years, and the 75th percentile by 70.75 years (see Table 3.6.1. and Figure 3.6.2. for age distribution of sample). The majority of the pubes are in excellent condition, and most have their identification number labeled on the dorsal surface of the bone.

Out of the 600 individuals with documented demographic information in the collection, 537 were assessed for sex estimation, and 519 were assessed for age estimation. Some individuals had only one pubis available, so only that side was scored and considered in analysis. Pubes that displayed pathology or severe postmortem damage were excluded from analysis. The pubes and ribs were removed during autopsy, so some bones were cut within the ischiopubic ramus. Therefore, not all of the specimens were

sectioned low enough that the subpubic contour was observable. These individuals did not have their subpubic contour scored and were excluded from any statistical analyses involving those traits and the logistic regression equation (following Lovell 1989). Some individuals additionally were sectioned just below the pubic symphyseal face, so the medial aspect of the ischiopubic ramus could not be scored. These were also excluded from analyses as mentioned previously. The location of sectioning excluded some individuals from the sex estimation portion of this project, however the majority of the specimens were able to be scored following the procedure outlined in Boldsen *et al.* (2002). Some bones had postmortem damage to the symphyseal face that obscured some of the components, but observable areas were scored. All data were collected blind to demography.

A sub-sample of 50 individuals was set aside to re-score and test interobserver agreement. These individuals were selected randomly from the entire skeletal collection and included 19 females (38 percent of the sub-sample) and 31 males (62 percent of the sub-sample). This sub-sample had ages ranging from 18 to 67 years, with a mean age of 44.22 years and a median age of 43 years. The second observer was also a graduate student with some experience using both methods being validated. Another randomly selected sub-sample of 70 individuals was set aside to be re-scored for the sake of calculating intraobserver agreement: 27 females (38.6 percent of the sub-sample) and 43 males (61.4 percent of the sub-sample). The range of ages in this sub-sample were from 18 years to 85 years, the mean was 47.62 years, and the median was 47.5 years.

3.2 Data Collection Procedure

Data collection was completed over the course of two business weeks in the forensic anthropology laboratory of Dr. Laura Fulginiti at the Maricopa County Forensic Science Center in Phoenix, which is where the skeletal collection is housed. Data collection by the primary observer took place from June 6th, 2016 until June 17th, 2016. The second observer spent June 15th, 2016 at the laboratory collecting their data in a separate room from the primary observer. Each observer was provided with a well-lit area to perform their data collection procedure, as well as an overhead lamp to use if necessary. The observers made use of physical copies of each reference material necessary for the data collection, including Klaes *et al.* 2012 and Phenice (1969) for the sex estimation portion, and Boldsen *et al.* (2002), the Data Collection Procedures for Forensic Skeletal Material 2.0 manual (Langley *et al.* 2016) (Section 2, pages 21-34), and the Brooks and Suchey (1990) and Hartnett (2010a) papers for the age estimation portion. This was to ensure that the observers were using exclusively the images and descriptions in the reference materials to obtain their ordinal scores for both portions of the study. The materials (including written descriptions, drawings or diagrams, and images) were thoroughly consulted for the scoring of every trait on every pubis analyzed to ensure as little bias as possible.

The observer had a Microsoft Excel file open which contained two sheets: one for sex estimation and one for age estimation. For each individual in the sample, the specimen catalogue or identification number was recorded in the first column of the spreadsheet. The left pubis would then be observed, and using the Klaes *et al.* (2012)

reference, the observer would score the ventral arc (one through five), the subpubic contour (one through five), and the medial aspect (one through five). These scores would be recorded in their respective columns. The observer would then use the Phenice (1969) reference to assign the individual a “gestalt” score. The “gestalt” estimates are defined as creating an overall assessment of sex (male or female) based on the overall appearance of the pubis and the basic Phenice (1969) descriptions and exemplars for each other the three aforementioned morphological traits (i.e. clearly male, ambiguous, or clearly female). No individual scores from Buikstra and Ubelaker (1994) were recorded due to time constraints and the overall simplicity of the system.

The left pubis would then have its five transition analysis traits (symphyseal relief, dorsal symphyseal texture, superior protuberance, ventral symphyseal margin, and dorsal symphyseal margin) assessed and scored (each having a different range of possible values per Langley *et al.* 2016). Each component score was recorded in its respective column in the spreadsheet. The observer would then assign the pubis a phase-score using the Brooks and Suchey (1990) and Hartnett (2010a) references, which was also recorded in the spreadsheet. Once all data was collected for the left pubis, the same procedure was repeated for the right pubis. The left and right pubes were scored separately to avoid biasing the observer’s estimates. Each individual was assessed completely for a sex and age estimate before moving onto the next. Any observations about the pubis (i.e. possible pathology, trauma, parturition pits, postmortem damage, or absent pubes) were recorded in a spreadsheet column for observations. Upon completion of the data collection,

photographs were taken of some of the sample (with permission from the collection curators).

In order to assess the inherent interobserver error of these two methods, a second visitor was invited to assess a portion of the sample. The second observer was Megan Sharpe, another Master of Science student who has experience utilizing both the Klales *et al.* (2012) and Boldsen *et al.* (2002) methods. She spent one business day at the collection scoring the random sub-sample of 50 individuals set aside by the first observer. The sample size of 50 individuals was chosen because it is large enough to be statistically significant, but enough that it could be thoroughly analyzed and scored by the second observer in the one day she was visiting the laboratory. She followed the same procedure as the first observer, however she worked in a separate space than the first observer and her assessments were recorded by her in separate Microsoft Excel spreadsheet. The second observer did not have access to any of the demographic information, and the only conversation between observers before data collection was about the procedure for data collection. Upon total completion of the data collection, the second observer relayed their data to the primary observer for statistical analysis. To analyze intraobserver error, the another sub-sample of 70 randomly selected individuals were re-scored by the primary observer. The first observer followed the original procedure exactly, assessing the left pubis for sex and age, then moving to the right. The data was recorded in an entirely different Microsoft Excel document to avoid bias by seeing the original scores.

3.3 Software, Data Processing, and Statistical Analysis

All data collected in the present study were stored in Microsoft Excel spreadsheets. The demographic information of the skeletal collection, including age at death, race, and sex was sent to the primary observer by the curator in a Microsoft Excel spreadsheet after conclusion of the data collection. The logistic regression equation (a Microsoft Excel spreadsheet) from Klales *et al.* (2012) and the ADBOU 2.1.044 (herein referred to as “ADBOU”) software package designed for use with the Langley *et al.* (2016) manual were downloaded from the Mercyhurst University Forensic Anthropology web page (<http://math.mercyhurst.edu/~sousley/Software/>). The software package IBM SPSS Statistics 20 was used for all statistical analyses of the raw data and creation of charts and plots.

3.4 Data Analysis: Sex Estimation

The actual sexes of the individuals were entered into a separate column in the spreadsheet for further analysis and comparison to the estimates. The raw ordinal scores collected by the first observer were entered into the logistic regression equation created for the Klales *et al.* (2012) method (see A.6. in Appendix). This spreadsheet can only utilize one pubis at a time (left or right), and all three scores must be entered for the equation to work properly (pubes with missing or unobservable scores cannot be used). The left pubes and right pubes for each individual were entered, and the probabilities and estimates were recorded for each side. The logistic regression produces a posterior

probability of both male and female group membership. The higher posterior probability would indicate the sex the individual is more likely to be.

The estimates provided by the logistic regression were compared to the actual sex of the individual, and those that were incorrectly classified were tallied. Overall classification accuracies were calculated for the Klales *et al.* (2012) method with the first data set by dividing the total number of individuals correctly classified by the logistic regression to the total number of individuals scored. This was performed separately for the left and right sides so that two classification accuracies were calculated (treating the pubes separately), in order to understand if the right or left pubes performed differently. Classification accuracies were also calculated for the left and right sides by comparing the individuals' actual sexes to the "gestalt" sex estimated using the Phenice (1969) traits. Individuals who were incorrectly sexed were tallied, and the number of correctly sexed individuals were divided by the total number of individuals in the scored sample.

Statistical tests were performed to analyze the performance of the Klales *et al.* (2012) method. The trait scores for the three defined area of the pubis were compared between the left and right sides (using paired sample t-tests) to investigate if there was asymmetry that would have a significant effect on the outcome of the logistic regression. The classification accuracies between the left and right sides were compared to see if they were substantially different. The classification accuracies for the Klales *et al.* (2012) method and the Phenice (1969) method were compared to see which performed better overall. The number of individuals that were sexed differently with their left pubis than their right was tallied to surmise if intraobserver error or asymmetry could lead to

different sex estimates using only one pubis or the other. The overall frequency of trait scores was calculated for both males and females.

To assess interobserver error, the data collected by the second observer on the sub-sample of 50 individuals was statistically compared to the original data collected by the first observer. The scores for the left sides and right sides were analyzed for asymmetry using paired sample t-test. The second observer's Kappa's weighted coefficient was calculated for each trait score from the first observer's data and the second observer's data to find the level of interobserver agreement. The overall classification accuracies for the second observer's estimates (both Klales *et al.* 2012 and the "gestalt" estimate) were calculated and compared to those of the first observer. To test intraobserver error, the weighted Kappa's coefficient was calculated for each of the three trait scores between the first observer's two trials. Overall classification accuracies for the Klales *et al.* (2012) method and the "gestalt" estimate were calculated for this second trial and compared to those achieved in the first.

3.5 Data Analysis: Age Estimation

The actual ages of the individuals were entered into a separate column in the spreadsheet for further analysis and comparison to the estimates. The raw ordinal scores for the pubic components collected by the first observer were entered into the ADBOU software package per Langley *et al.* (2016) (see A5 in Appendix). The software requires that the user select a sex (male, female, or unknown), ancestry (White, Black, or unknown), and case type (forensic or archaeological) before the data can be analyzed.

The actual sex of the individual was used, but ancestry was selected as “unknown”. All of the data was run using the forensic case reference sample. The ADBOU software allows the observer to enter a “minimum” and “maximum” score for each component, and the left and right pubes can be scored differently and still be used in the Bayesian analysis. Pubes with some missing or unobservable components were still used in the software. The left pubes and right pubes for each individual were entered (or just for one side if a pubis was missing), and the resulting age range (at the 95 percent confidence interval) and maximum likelihood (a point estimate) for the age of the individual were recorded in the spreadsheet.

The estimates provided by the ADBOU software package were compared to the actual ages of the individuals, and those that were incorrectly classified were tallied. Incorrect classification for this method means that the individual’s actual age at death did not fall within the age range produced by the ADBOU software package. Correct classification means that the individual’s actual age at death fell inside the age range produced by the ADBOU software package. Overall classification accuracies were calculated for the Boldsen *et al.* (2002) method with the first data set by dividing the total number of individuals correctly classified by the logistic regression to the total number of individuals scored. Bias and inaccuracy were calculated using the actual ages and maximum likelihood age estimates produced by the software. A scatter plot was constructed to visually represent how actual age at death correlated with the estimated age at death (using the actual age at death as the x-axis and the maximum likelihood estimate as the y-axis). Each pubic component was also tested to see how strongly it

correlated with actual age at death by calculating Pearson's and Spearman's correlation coefficients.

To test interobserver agreement, weighted Cohen's kappa's coefficient was calculated for each component score between the first and second observers' data sets. The Spearman's correlation coefficient was also calculated to see how well the two sets of scores correlated with each other. The overall classification accuracy using the ADBOU software package was calculated for the second observer's data set, as defined above. Each pubic component was also tested to see how strongly it correlated with actual age at death (by calculating Spearman's correlation coefficient). A scatter plot was constructed to visual represent how actual age at death correlated with the estimated age at death (using the maximum likelihood).

3.6 Tables

Age groups (in years)	Number of individuals in age group
18-19	12
20-29	70
30-39	62
40-49	118
50-59	104
60-69	76
70-79	73
80-89	67
90-99	17
Total	599

Table 3.6.1. A breakdown of the number of individuals in the Forensic Science Center pubic bone the sample in each decadal age cohort.

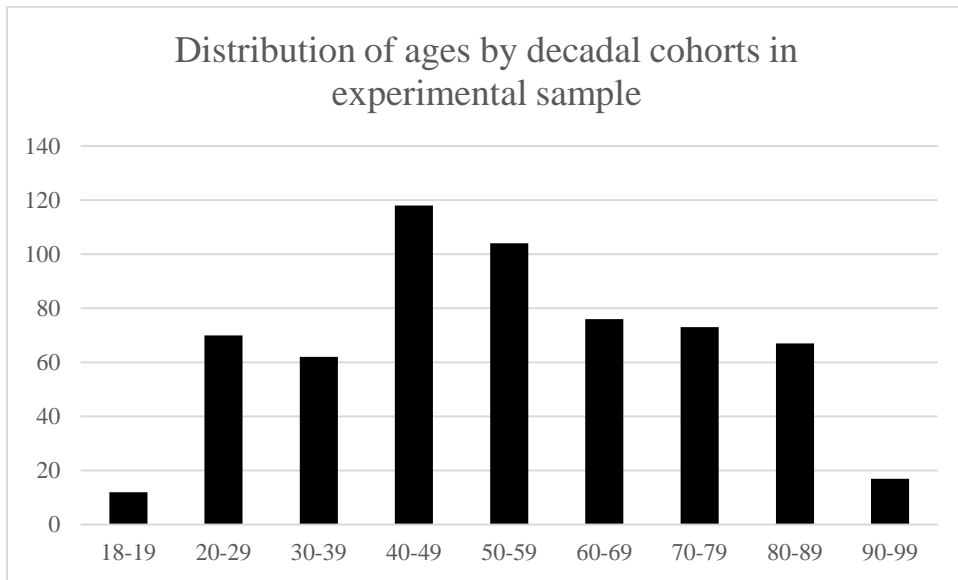


Figure 3.6.2. A bar graph showing the distribution of ages by decadal cohorts (years) in experimental sample.

CHAPTER 4: RESULTS

4.1 Results of Sex Estimation Validation

Some pubes received one or more trait scores which differed between the left and the right sides in a single individual (bilateral asymmetry). Out of the 537 individuals that were assessed for sex estimation by the first observer, 56 had one morphological trait that was scored differently between the left and the right, and five individuals had two traits scored differently between the left and right sides. This bilateral asymmetry resulted in the Klales *et al.* (2012) logistic regression equation producing a different sex estimate for the left and right sides in nine individuals. For the purpose of the t-tests, the null hypothesis was defined as: the average difference in the trait scores between the left and right side of the same individual is zero. The alternative hypothesis was defined as: this same average difference does not equal zero. A paired sample t-test was performed in IBM SPSS to compare the difference in the means of the ordinal scores (one through five) for each of the pubic morphological traits (ventral arc, subpubic contour, and medial aspect) on the left and the right sides. For the ventral arc, the sample size where both traits were observable for comparison was 519. The resulting p value of 0.01 ($t = -2.596$, $SD = 0.339$) means there is sufficient evidence to reject the null hypothesis and accept the alternative. Therefore, the asymmetry in the ventral arc for each individual is statistically significant. For the subpubic contour, 447 pubes were analyzed. The paired t-test resulted in a p value of 0.912 ($t = -0.111$, $SD = 0.426$), therefore, the mean difference between the scores on the left and right side is not significant. In regards to the medial aspect of the ischiopubic ramus, 456 pubes were analyzed. The paired samples t-test

produced a p value of 0.536 ($t = 0.620$, $SD = 0.387$), indicating the mean difference between MA scores for the left and right sides of are not statistically significant. When running the logistic regression on the left and right pubes for each individual, 13 individuals were sexed as both male and female from each of their pubes, meaning the left and right side were asymmetrical enough to produce different overall sex estimates depending on which was entered into the logistic regression.

A sample of 500 pubes for the left side and 492 for the right side were used to calculate the overall classification accuracies for each method, because that is how many pubes had all traits available and were used in the logistic regression equation. The first observer's "gestalt" (using the broader Phenice (1969) categories to choose an overall estimate of "male", "ambiguous", or "female") classification accuracy rate for the left side pubes was 98.4 percent and 98.6 percent for the right side pubes. The logistic regression equation (Klales *et al.* 2012) produced overall classification accuracies for the same sample of 72.4 percent and 75.2 percent for the left and right side pubes respectively (see Table 4.3.1.). In the case of the logistic regression, there were 138 individuals that were incorrectly sexed (using only pubes from the left side). Of those 138, only two were actually females that were incorrectly sexed as male. The remainder were actually males that were incorrectly sexed as females. For the observer's "gestalt" estimate of the same pubes, five of the eight incorrectly sexed individuals were actually females that were scored as male, and the other three were actually males incorrectly sexed as females.

4.1.1 Sex Estimation: Interobserver Agreement

The second observer assessed a random sample of 50 pubes. The sub-samples utilized in the Klales *et al.* (2012) logistic regression (based on all three traits being present and observable), were 46 and 43 respectively for the left and right sides. Using the left pubes of each individual, her “gestalt” classification accuracy was 87.0 percent. Using the right pubes, the classification accuracy was 88.4 percent (see Table 4.3.1.). Using the same samples for each side, the classification accuracies provided by the logistic regression from Klales *et al.* (2012) were 80.4 percent and 76.7 percent respectively for the left and right sides (see Table 4.3.1.). In regards to each morphological trait, the ventral arc scores for the left and right sides from 49 individuals were compared using a paired sample t-test. The resulting p-value of 0.554 ($t = 0.597$, $SD = 0.719$) is greater than the significance level of 0.05, indicated that in this sample, the asymmetry between the left and right ventral arcs is not statistically significant. Performing the same test on the subpubic contour produced a p-value of 0.109 ($t = -1.636$, $SD = 0.638$), which is higher than the significance level of 0.05. Again, the null hypothesis cannot be rejected. The same test procedure was again repeated for the medial aspect of the ischiopubic ramus, and the resulting p-value was 0.80, again exceeding the significance level (0.05). The null hypothesis cannot be rejected in this instance either. Out of the 44 individuals that had both pubes present, seven were given different sex estimates using either the left or right pubis, meaning the asymmetry of scores was substantial enough between sides to sex the individual as both male and female depending on which pubis was used in the regression equation, although the asymmetry

between the left and right side trait scores was not found to be statistically significant. All of the individuals that were incorrectly sexed by the logistic regression equation (considering the left and right sides separately), were actually males that were misclassified as males. During the observer's "gestalt" estimations, she only misclassified one individual as a male, although the actual sex was female. The other individuals incorrectly sexed using the Phenice (1969) method were all actual males sexed as females.

Kappa's coefficient was calculated to surmise the amount of interobserver agreement for each of the three trait scores assigned using the Klales *et al.* (2012) system, using only the pubes from the left side (see Table 4.3.2.) (see A.7. in Appendix for information on interpretation of Kappa values). For the ventral arc, the k-coefficient was 0.397 for a valid sample of 50 individuals. For the subpubic contour, the k-coefficient was 0.362 for a valid sample of 44 individuals. For the medial aspect of the ischiopubic ramus, the k-coefficient was 0.589 for a valid sample of 44 individuals. All kappa values were found to be significant at the 95 percent confidence level. Of the sample of 50 individuals that were scored by the second observer, there were five individuals that were assigned a different sex by the first and second observer using the Phenice (1969) system. All other sex estimates given for each individual by the two observers were the same (although the estimates were not necessarily correct). This produced a Kappa's coefficient of 0.946 (significant at the 0.05 level) for the overall sex estimates. In the subsample of 46 individuals that were scored with the Klales *et al.* (2012) system, eight of the individuals were given different sex estimates based on the scores assigned by each

observer (although these estimates were not necessarily correct). This produced a Kappa's coefficient of 0.760 (significant at the 0.05 level). The "gestalt" estimates each observer arrived at using the basic Phenice (1969) system had a higher level of agreement than those found using the Klales *et al.* (2012) logistic regression equation.

4.1.2 Sex Estimation: Intraobserver Agreement

To investigate intraobserver error, the first observer re-scored a random sample of 70 individuals. Only the left pubes were used in this portion of the statistical analysis. A sub-sample of 59 individuals was used in the logistic regression (having all traits observable). The "gestalt" classification accuracy rate for the observer's second trial when using the Phenice (1969) system was 98.3 percent and 98.2 percent for the left and right pubes respectively (see Table 4.3.1.). The classification accuracies for the left and right pubes in this trial were 69.5 percent and 73.7 percent respectively using the Klales *et al.* (2012) logistic regression equation (see Table 4.3.1.). Kappa's coefficient was calculated for each of the three morphological trait scores to surmise the level of agreement between the two trials by the same observer (see Table 4.3.3.). For the ventral arc (n=63), the k-coefficient value calculated was 0.398, the Spearman's correlation coefficient was 0.869, and the Pearson's correlation coefficient was 0.857. For the subpubic contour (n=63), the Kappa's coefficient calculated was 0.619, the Spearman's correlation coefficient was 0.901, and the Pearson's correlation coefficient was 0.897. For the medial aspect of the ischiopubic ramus (n=63), the k-coefficient calculated was 0.569, the Spearman's correlation coefficient was 0.892, and the Pearson's correlation

coefficient was 0.895. The Kappa's coefficient calculated for the "gestalt" estimate using Phenice (1969) criteria between the first observer's two trials was 0.946 (n=68). The Kappa's coefficient calculated for the sex estimates produced by the logistic regression equation between the first observer's two trials was 0.502 (n=57).

4.2 Results of Age Estimation Validation

For the statistical analyses of the age estimation method, 517 individuals were considered. Out of these 517 individuals considered, 97 were incorrectly classified, meaning their actual age at death was not included in the age range produced by ADBOU. This was 18.76 percent of the sample, which means that 81.24 percent were correctly classified, or that their actual age at death fell inside the age range produced by the software. Bias and inaccuracy were also calculated. Bias is the differences between the estimated ages and the actual ages being summed, and then divided by the total number of individuals; the value for the present study (n=517) being -0.0201 years. Inaccuracy is the differences between the estimated ages and the actual ages having their absolute values summed, and then divided by the total number of individuals. In the present study inaccuracy was calculated as 12.4901 years (n=517). There were only three individuals of Asian ancestry represented in the sample, and all three of them were misclassified. Of the 19 Black individuals in the sample, five were misclassified, and three of the Hispanic individuals (out of seven total) were misclassified. Only one of the seven Native American individuals was misclassified. There were 33 individuals that were missing one or more scores, due to either postmortem damage or because one pubis

was missing. Out of these 33 individuals, eight of them were incorrectly classified. If we remove these individuals from the analysis (N= 484), then 18.39 percent of individuals were misclassified.

Statistical analyses were used to compare the actual age at death of each individual with the “maximum likelihood” or best point estimate of age that was produced with the ADBOU software. The Pearson’s correlation coefficient for the actual age and point estimate age estimate was calculated, since the data is scale integers and theoretically they should be perfectly correlated. The resulting Pearson’s *R* is 0.628 (Spearman’s *R* = 0.631), indicating a moderate to good correlation (see Figure 4.4.1 for scatterplot of data). The mean estimated age at death (using the point estimate) of this sample was 47.15 years, while the mean actual age at death of the sample was 53.08 years, so the mean difference between the estimated age and actual age (maximum likelihood) was -5.92 years, meaning age estimates from ADBOU were overall negatively biased. Each component score was compared to the actual ages at death to see which was most highly correlated with age. For the left side pubes, from highest to lowest correlation, the components were: ventral symphyseal margin, dorsal symphyseal margin, symphyseal relief, superior protuberance, and dorsal symphyseal texture (see Table 4.4.2.). Correlation tests were repeated using the component scores for the right side pubes, and from highest to lowest correlation, the components were: ventral symphyseal margin, dorsal symphyseal margin, symphyseal relief, superior protuberance, and dorsal symphyseal texture (see Table 4.4.3.). Although the Spearman’s *R* varied

between the left and right sides, the components ranked in the same order for correlation between the two sides.

4.2.1 Age Estimation: Interobserver Agreement

Out of the 50 individuals scored by the second observer, 10 of them (20 percent of the sample) had actual ages at death that did not fall inside the age range produced by the ADBOU software; six of these individuals were male and four were female. This means that 80 percent of the sample had actual ages at death that fell inside the age range ADBOU produced. The correlation between actual age at death of the sample and estimated age at death (maximum likelihood “point estimate”) was calculated for this sample. The resulting Pearson’s R is 0.716, and Spearman’s R is 0.762, indicating a strong positive correlation. To investigate interobserver agreement between the scoring of pubic symphyseal components, Kappa’s coefficient was calculated for the first and second observers’ samples. In order from highest to lowest interobserver agreement, the pubic component scores ranked as followed: symphyseal relief, dorsal symphyseal relief, dorsal symphyseal margin, superior protuberance, and ventral symphyseal margin (see Table 4.4.4.). The ages that were estimated for the same sample by the different observers were compared.

The mean age estimated by observer one (using the ADBOU software package) was 45.92 years on the sample sub-sample as the second observer. The mean age estimated by observer two on this sample was 38.82 years. The actual mean age of the sample ($n=50$) was 44.22 years. The first observer also scored individuals in a way that

produced higher age estimates than the second observer (with the mean difference in their estimates being 7.1 years). The mean of the actual age at death (44.22 years) and the mean of observer one's estimates (45.92 years) had a difference of -1.69 years. The same procedure was carried out for the second observer's data set, and it was found that the difference between the mean of actual age at death and the mean of their estimates was +5.40 years. Observer one produced positively-biased age estimates from their component scores, while observer two produced negatively-biased age estimates from their component scores. A scatterplot was created to visually compare the first observer's age estimates with the second observer's (See Figure 4.4.5.). The correlation coefficient between actual age at death and the first observer's estimates was calculated as 0.783, and the same was done for the second observer, producing a correlation coefficient of 0.716, which are very similar levels of correlations.

4.2.2 Age Estimation: Intraobserver Agreement

Intraobserver agreement was also assessed when the first observer re-scored a sub-sample of 70 individuals. Out of a viable sample of 68 that could be used for statistical analysis, 13.24 percent of the sample was scored in such a way that their actual age at death did not fall within the age range produced by the ADBOU software package. Therefore, 86.76 percent of the sample was scored in a way that their actual age at death did fall within the estimated range. Spearman's correlation coefficient was calculated using the maximum likelihood estimate from the first round of estimates and the second round of estimates by observer one, and it was found to equal 0.929 (indicating an almost

perfect positive correlation). Correlation and intraobserver agreement was also calculated for each of the five component scores. In order from highest to lowest agreement, they ranked as follows: symphyseal relief, dorsal symphyseal margin, ventral symphyseal margin, dorsal symphyseal texture, and superior protuberance (see Table 4.4.6.). All component scores had good to excellent positive correlation with each other between trials by observer one (see Table 4.4.6). Cohen's weighted Kappa was calculated, using the maximum likelihood estimates from the primary observer's first and second trials, and was found to be 0.255. This indicates a fair level of agreement for the actual age estimates produced by ADBOU.

The mean of the maximum likelihood (point estimates) of the age estimates from the ADBOU software package were compared between the first observer's two trials. The difference between the mean age estimate of the first trial and second trial was calculated as +1.67 years, indicating that the second trial produced overall lower age estimates than did the first trial. Three paired sample t-tests were conducted to examine if there was a substantial difference between the various trials of estimates. The first t-test compared the actual ages at death of the individuals in the sample to the age estimates (maximum likelihood) produced by ADBOU in the first trial. The results indicated that the difference in the means of the actual age and the estimated age was -0.86 years (for a sample of 66 individuals), with a *p* value of 0.618. The second t-test compared the actual ages at death to the estimates from the first observer's second trial. The difference in the means of the two groups was found to be +0.81 years, with a *p* value of 0.606. The third t-test compared the point estimates from the first and second trial to each other. The

difference in the means of the two trials was found to be +1.67 years, with a p value of 0.076. Therefore, none of the differences in the means were statistically significant.

4.3 Sex Estimation: Tables

	Observer 1		Observer 2		Observer 1 (2 nd trial)	
	Left	Right	Left	Right	Left	Right
Klales <i>et al.</i> (2012)	72.4%	75.2%	80.4%	76.7%	69.5%	73.7%
Phenice (1969)	98.4%	98.6%	87.0%	88.4%	98.3%	98.2%

Table 4.3.1. The various classification accuracies for the left and right side pubes for sex estimation using the two different methods.

	Ventral arc	Subpubic contour	Medial aspect	Klales <i>et. al</i> sex estimate	Gestalt estimate
Kappa value	0.397	0.362	0.589	0.760	0.946

Table 4.3.2. The weighted Cohen’s kappa values calculated to assess interobserver agreement for the Klales *et al.* (2012) ordinal scores and sex estimates.

	Ventral arc	Subpubic contour	Medial aspect	Klales <i>et. al</i> sex estimate	Gestalt estimate
Kappa value	0.398	0.619	0.569	0.502	0.946

Table 4.3.3. The weighted Cohen’s kappa values calculated to assess intraobserver agreement for the Klales *et al.* (2012) ordinal scores and sex estimates.

4.4 Age Estimation: Figures and Tables



Figure 4.4.1. A scatterplot showing the actual ages at death (x-axis) compared to their corresponding age estimates (using maximum likelihood point estimate; y-axis). The red line is indicative of perfect correlation.

	Spearman's R	Pearson's R
Symphyseal relief	0.455	0.318
Dorsal symphyseal texture	0.379	0.349
Superior protuberance	0.435	0.469
Ventral symphyseal margin	0.623	0.608
Dorsal symphyseal margin	0.561	0.571

Figure 4.4.2. The correlation coefficients calculated for each component score from the left pubes (n=516) and actual age at death.

	Spearman's R	Pearson's R
Symphyseal relief	0.504	0.528
Dorsal symphyseal texture	0.440	0.461
Superior protuberance	0.449	0.476
Ventral symphyseal margin	0.583	0.569
Dorsal symphyseal margin	0.511	0.526

Figure 4.4.3. The correlation coefficients calculated for each component score from the right pubes (n=503) and actual age at death.

	Spearman's R	Pearson's R	Kappa value
Symphyseal relief	0.897	0.883	0.665
Dorsal symphyseal texture	0.701	0.746	0.523
Superior protuberance	0.782	0.801	0.464
Ventral symphyseal margin	0.694	0.793	0.287
Dorsal symphyseal margin	0.798	0.819	0.476

Figure 4.4.4. The correlation coefficients and kappa values calculated to assess interobserver agreement between each set of component scores given by the two observers using the left pubes (n=49).

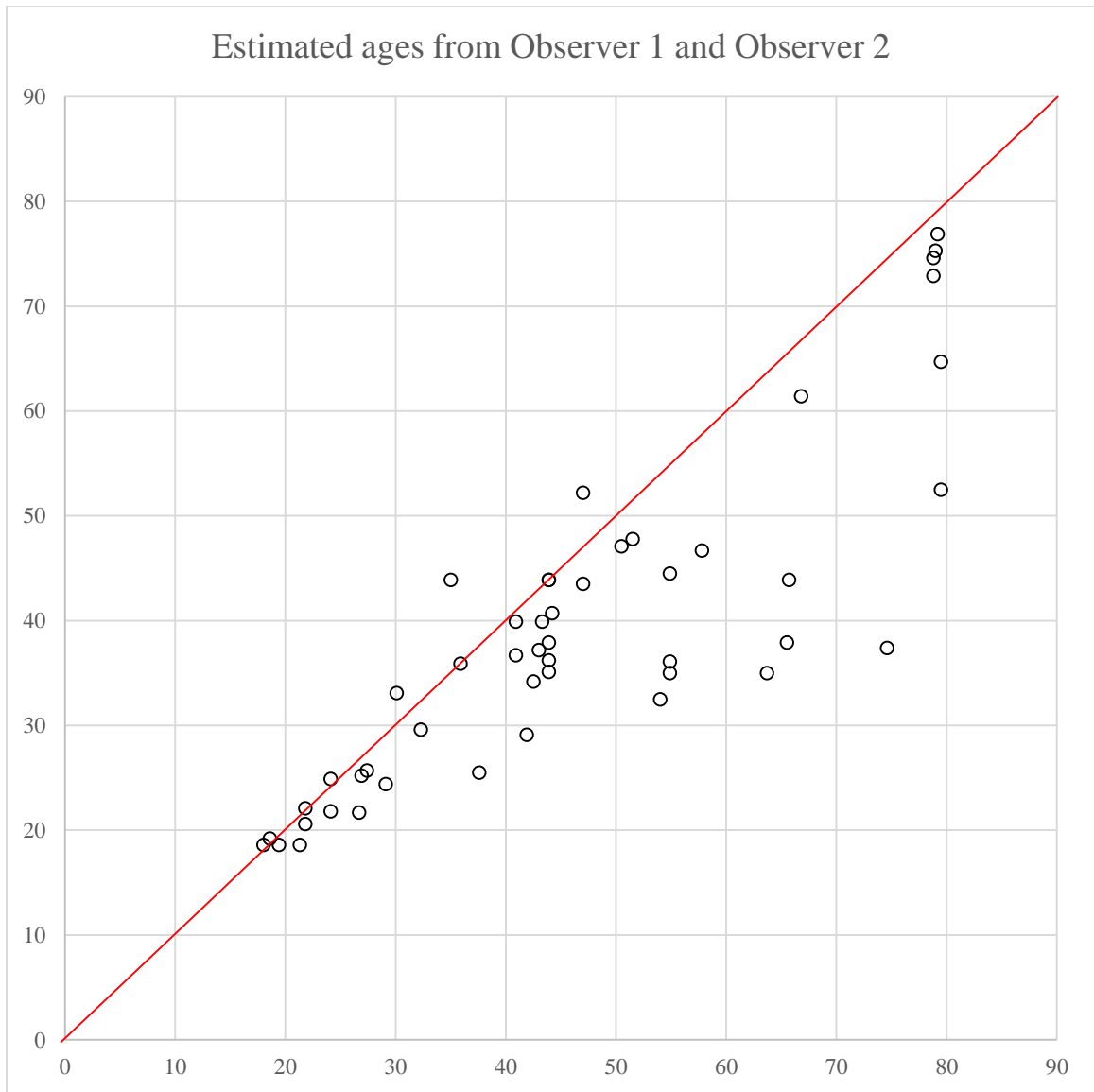


Figure 4.4.5. The age estimations from the first observer (x-axis) plotted versus the age estimates from the second observer (y-axis). The red line represents perfect correlation.

	Spearman's R	Pearson's R	Kappa value
Symphyseal relief	0.854	0.913	0.696
Dorsal symphyseal texture	0.820	0.837	0.573
Superior protuberance	0.744	0.811	0.535
Ventral symphyseal margin	0.848	0.834	0.573
Dorsal symphyseal margin	0.844	0.868	0.659

Table 4.4.6. The correlation coefficients and Kappa values calculated to assess intraobserver agreement for each set of component scores using the left pubes (n=68).

CHAPTER 5: DISCUSSION

5.1 Sex Estimation

The Klales *et al.* (2012) method did correctly classify the majority of the individuals in the present study, but none of the trials produced classification accuracies exceeding 78.5 percent. This is indeed a significantly lower classification accuracy than was predicted in the original study, which was 86.2 percent. There were 138 individuals in the sample that were incorrectly sexed during the first trial, and almost all (excepting two individuals) were males misclassified as females. The classification accuracies were considered separately for the left and right side pubes, because the logistic regression equation can only consider one pubis at a time. The classification rate of 72.2 percent (from the left pubes) was less than the classification accuracy of the first observer (98.4 percent) using basic Phenice criteria and observational experience. The second observer had a classification accuracy of 87.0 percent using basic Phenice criteria and observational experience, while the Klales *et al.* (2012) method produced a classification accuracy of 80.43 percent for the left pubes. For the right pubes, the first observer had a classification accuracy of 98.6 percent using Phenice criteria, and 75.0 percent using the Klales *et al.* (2012) method. The second observer had a classification accuracy of 88.4 percent using Phenice (1969) criteria, and 76.7 percent using the Klales *et al.* (2012) method. Although both methods performed relatively well, the utilization of basic Phenice criteria (a strict dichotomy of clearly male morphology or clearly female morphology), did produce higher classification accuracies than did Klales *et al.* (2012) logistic regression (up to 98.6 percent). In addition, the utilization of the Klales *et al.*

(2012) methodology is much more time consuming than is the utilization of Phenice (1969). There appears to be no advantage to utilizing the equation, excepting that it produces quantified statistical probabilities, which Phenice (1969) does not. Phenice (1969) is not *Daubert* (1993) compliant at this time, but it seems to be much more accurate for sex estimation than is the Klales *et al.* (2012) method, an issue that should be addressed in further research.

Additionally, it is apparent that asymmetry between the left and right pubes can affect the overall sex estimate. Of all three morphological traits, only the asymmetry between the left and right ventral arc was found to produce statistically significant results. The asymmetry between the other two traits (the subpubic contour and medial aspect) did not produce statistically significant results. It is problematic, however, that the ventral arc is the most heavily weighted of the three traits in the logistic regression equation and also shows statistically significant asymmetry. If a ventral arc score between one and three is given, then the individual will almost always be estimated as female, even if the other two traits receive higher “male” scores. This bilateral symmetry of at least one trait was noted in 56 individuals from the sample, and five of those individuals actually had different scores for two traits. Out of these 56 individuals, nine were incorrectly sexed using at least one pubis. This would indicate that in a real case, an individual could be incorrectly sexed if both pubes are not considered, or if one pubis is not recovered, especially if the ventral arc is the most variable trait that carries the heaviest statistical weight.

In a debriefing after the collection of data was concluded, the two observers discussed issues they encountered in the utilization of the Klales *et al.* (2012) method. The most significant grievance was with the variation in the morphology of the medial aspect of the ischiopubic ramus that was not accounted for in the literature (see Figure 5.3.1.). In some females, the ramus was quite wide relative to the symphyseal face, enough for a score of three, four, or even five, but had a prominent and sometimes sharp ridge simultaneously. Some females had rami, both wide and narrow, with no ridge at all. And some male individuals had rami with ridges as well, some rounder and some sharper. This made it difficult to correctly assign a score to the trait, because the presence of a ridge on a narrow or pinch rami is a score of one or two, while there is no score to assign a ramus that is wide but also has a ridge. The observers noted that in this instance, they would more heavily weight the width of the ramus in relation to the pubic symphyseal face for a score rather than the ridge. Although no conscious bias was considered when encountering these pubes, it is possible that bias caused the observers to choose a score based on what the other traits had scored, or whether or not the pubis “seemed” male or female.

Trait frequencies were tabulated amongst males and females (see Table 5.3.2), and the results indicated that males could have trait scores that would be considered more female by the descriptions given in the original study. The majority of males had a ventral arc. Ventral arc scores of two and three were actually more common among males than females (see Figure 5.3.3. and 5.3.4.). This greatly skewed the sex estimates because of the heavy statistical weight on the ventral arc, and the tendency of the equation to

consider any score of three or below as more likely to be female. Both observers noted, however, that females will almost always have a ventral arc that is scored a one or two, rarely a three or four, and never a five (excepting pathological specimens in which the ventral arc was obscured). Conversely, males can have a ventral arc of any score, although scores of one and five were the rarest among them. These results are similar to Walker's 2005 study of the greater sciatic notch and the ordinal scores used to estimate sex from its morphology. He found that scores of two and three were more common among males than females, while females usually had a score of one (Walker 2005). Similar results were found for the other pubic traits – males can have scores ranging anywhere from one to five (although one is the rarest score), while most females tend to have scores of one or two for all traits.

The level of interobserver agreement was found to be moderate to fair for the Klales *et al.* (2012) method with the Kappa values falling around 0.4 for every trait score. The level of intraobserver agreement was moderate to good with Kappa values around 0.5. Although this indicates some level of agreement between trials, it would appear that each time a pubis is being considered, it is probable it will be assigned an entirely different trait score, whether by two different individuals, or by the same individual at different times. This is problematic, because an ordinal score differing by even one for just a single pubic trait could produce a completely different sex estimate. So although interobserver agreement was fair to moderate, any difference in scoring can introduce error into the sex estimates. The resulting classification accuracies indicated that overall, males are far more likely to be incorrectly sexed than are females. Almost all of the

individuals that were misclassified were actual males being labeled as females by the logistic regression. Both observers hypothesize that this is due to the ventral arc being heavily weighted. A ventral arc score of three is more likely to be indicative of male morphology, and yet the logistic regression weights that as that female morphology, which would greatly influence the posterior probability.

5.2 Age Estimation

Although the ADBOU software does include the auricular surface of the innominate and cranial sutures, it does allow for Bayesian analysis to be performed even if not all of that data is present. The purpose of the present study was to see how well the components of just the pubis performed on their own. Previous Bayesian analysis of cranial sutures and the sacroiliac surface has shown that they tend to produce negatively biased age estimates, while the pubis performs relatively well on its own (Milner and Boldsen 2012). It is therefore beneficial to validate the performance of just the pubis. In the present study, the pubic symphysis did perform relatively well on its own. For the sample of 517 individuals that were viable for age estimation and subsequent statistical analysis in the current study, 81.2 percent were given component scores that allowed the ADBOU software package to produce an age range that included the individual's actual age at death, even in cases where some components were unobservable (i.e. damaged postmortem) or if only one pubis was available for analysis. This classification accuracy is relatively high considering only the pubic symphysis data was used. Out of the 97 individuals that were incorrectly classified, 12 were not White, which is 12.37 percent of

the group of incorrectly classified individuals. This is to be expected, since most reference samples on which these methods are developed contain mostly White individuals. It is also apparent then that some younger individuals have pubes that appear “older”, while some older individuals maintain pubes with a more “youthful” morphology. The Spearman’s R value between actual age and estimated age was 0.762, indicating a strong positive correlation between the two. Considering that the maximum likelihood was used for the purpose of correlation, this indicates that the method performs relatively well using just pubic components, and that the actual ages are, in most cases, very close to the point estimate predicted by the ADBOU software package.

The resulting Pearson’s correlation coefficient for the actual age at death and the estimated ages from ADBOU is 0.628 (Spearman’s $R = 0.631$), which indicative of a moderate to strong correlation. The mean difference between the actual age and estimated age (maximum likelihood) was -5.92 years, an overall negatively biased age estimate. This indicates that the pubes were scored in a such a way that the age at death was underestimated, by almost 6 years on average. This is not unexpected, considering the known difficulty of aging older individuals who have reached full skeletal maturity (Boldsen and Milner 2012; Algee-Hewitt 2013). Out of the 97 individuals who were aged incorrectly, 63 had their ages underestimated, and 34 had their ages overestimated. Additionally, 68 of those 97 individuals who were aged incorrectly were over the age of 40. These results are consistent with what other methods have reported: that the older an individual becomes, the more inaccurate age estimates from their skeletal remains become. The data also could be explained by the phenomenon of “attraction to the

middle” – predicted ages of unknown samples tending towards the middle-age categories due to uncompensated errors in the much older age categories (Bocquet-Appel and Masset 1982). Also, it is known that “age mimicry” of the reference skeletal sample upon which the method is based can complicate age estimation methods, and that also seems to be the case with this method (Algee-Hewitt 2013). Essentially if more individuals of younger ages were used to create the method, then older individuals in the validation sample will tend to have their ages underestimated, due to the “mimicking” of the reference sample (Algee-Hewitt 2013). Transition analysis using pubic components is no exemption to these problems that have plagued all other previously devised age estimation methods (Bocquet-Appel and Masset 1982).

In a test of correlation, the ventral symphyseal margin was ranked as the most strongly correlated with age (see Figure 5.4.1), while the dorsal symphyseal texture had the weakest correlation (see Figure 5.4.2.). The ventral symphyseal margin has seven possible component scores to describe the state of the ventral border of the pubic symphysis. The youngest individuals will have sloping bone with serrated edges and no rampart formation, while the oldest individuals should have fully formed rims which are in the process of breaking down or becoming disfigured (Langley *et al.* 2016). This particular area on the pubic symphysis seems, statistically, to be the most diagnostic of age. The ventral rampart has been utilized in other age estimation methods from the pubic symphysis (Todd 1920; McKern and Stewart 1957; Phenice 1969), because it builds up predictably with age. The correlation coefficient of 0.623 indicates this trait especially should be considered for future component-based methods for age estimation involving

the pubis. The dorsal symphyseal texture was the component that was least correlated with age, with a Spearman's R value of 0.379. This component has four possible scores, with the younger individuals have smooth, fine-grained bone and the older individuals having marked macroporosity. This indicates a weak relationship between the variables. Throughout the course of data collection for the present study, it was noted by both observers that macroporosity (sufficient enough to give a component score of four) was not a common occurrence, even in much older individuals. Further research is needed to surmise how common macroporosity truly is, and to understand if this component is helpful in predicting age at death. It is also possible that the lower number of scores to describe the appearance of this component could mean that some variation in older age cohorts is not being accounted for, which could lead to a lower correlation with age. Although these various pubic components showed slight to moderate correlation with age, none of the correlations exceed 0.7, which according to Boquet-Appel and Masset (1982), is not good enough to be a sex indicator. Others may disagree with this (Konigsberg and Frankenberg 1992), but it should certainly be the case that the age indicators we use in forensic cases should correlate as strongly with age as possible.

Interobserver agreement ranged from fair to substantial for the transition analysis component scores. The symphyseal relief showed the highest level of interobserver agreement, with a weighted Kappa value of 0.665, while the ventral symphyseal margin had the lowest interobserver agreement, with a weighted Kappa value of 0.287. This is potentially problematic considering that the ventral symphyseal margin is also the component that correlates the best with age. If it is therefore not reliably scored between

observers, but is the best predictor of age, it can lead to the production of very different age estimates if not scored correctly. Observer one tended to give out higher component scores for this trait, while observer two gave out lower component scores. The Langley *et al.* (2016) manual contains very detailed descriptions and photos to illustrate every component score, so it is likely that the difference between observers was caused by the amount of experience with assessing whether the ventral rim is forming or breaking down. However, every pubic component had weight Cohen's kappa values of at least 0.5, and the values indicated that the components had fair agreement to substantial interobserver agreement which is promising for the continued use of the method.

The second observer's data set had similar levels of accuracy to the first observer's: 80 percent of the second observer's sample had actual ages that fell within the range produced by ADBOU. The mean age estimated by observer one (using the ADBOU software package) was 45.92 years. The mean age estimated by observer two was 38.82 years. Therefore, the first observer tended to score individuals in a way that produced higher age estimates than did the second observer (with the mean difference in their estimates being 7.1 years). During both the first trial for observer one and the trial for observer two, pubic data produced negatively biased age estimates when entered into the ADBOU software. It was also noted by both observers that parturition pits on the dorsal pubis can obscure the dorsal symphyseal margin or make it hard to score. They often disfigure the margin or rim to the point where it can resemble breakdown. Despite that, females tended to have their age more accurately estimated than males. Only 29.9 percent of the individuals who had their ages estimated wrong were females, while the

other 70.1 percent were males. Considering that females make up 32.4 percent of the reference sample, they therefore had their ages predicted correctly at a higher proportion than did males.

Intraobserver agreement was also assessed, and in this second trial, it was found that 13.24 percent of the sample was scored in such a way that their actual age at death did not fall within the age range produced by the ADBOU software package. Near perfect correlation was found between the first round of age estimates and the second round of age estimates. The component with the highest level of intra observer agreement was symphyseal relief, which does not have a particular good correlation with age. The superior protuberance component, which also had only a moderate correlation with age, had the lowest level of interobserver agreement. Overall, the level of intraobserver agreement was moderate for the Boldsen et al. (2002) method. These results are promising, but agreement was not any higher than substantial (no Kappa values exceed 0.7; Landis and Koch 1977). The intraobserver reliability has much room for improvement with this method. Varying levels of experience or misinterpretation of the Langley *et al.* (2016) manual may have been responsible for only this moderate agreement. The mean difference between the age estimates of the first trial and second trial was calculated as +1.67 years, indicating that the second trial produced overall lower age estimates than did the first trial. Three paired sample t-tests were conducted to examine if there was a substantial difference between the various trials of estimates. The first t-test compared the actual ages at death of the individuals in the sample to the age estimates (maximum likelihood) produced by ADBOU in the first trial. The statistical

results indicated that there was no statistically significant difference in the means between the age estimates from each trial and the actual ages at death, nor between the age estimates of the first trial and the age estimates of the second trial. This indicates that the Boldsen *et al.* (2002) method has a design which allows observers to produce relatively similar age estimates between different trials on the same sample. So although the component scores had only moderate agreement on average, it still allowed for similar age estimates to be produced.

5.3 Sex Estimation: Figures



Figure 5.3.1. A male with a very wide ischiopubic ramus that also has a sharp ridge extending off the inferior symphyseal face. This variation is not accounted for in the Klaes *et al.* (2012) literature. The scale is in centimeters.

Trait score frequencies (Klaes <i>et al.</i> 2012 method)							
	Score	1	2	3	4	5	Total
Ventral Arc	Males	8	83	116	119	32	358
	Females	146	17	2	1	0	166
Subpubic Contour	Males	2	20	140	115	39	316
	Females	124	28	5	3	0	160
Medial Aspect	Males	7	35	103	133	45	323
	Females	54	80	24	2	0	160

Figure 5.3.2. A chart showing score frequencies for each pubic trait among males and females using the left pubes and the Klaes *et al.* (2012) method.

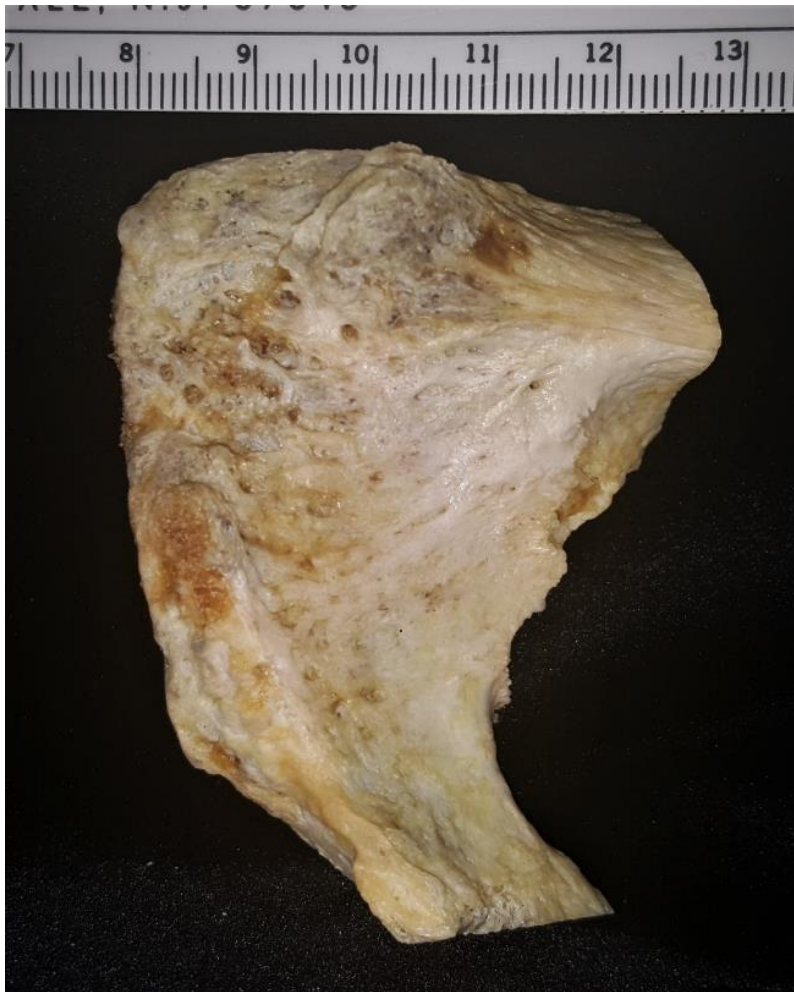


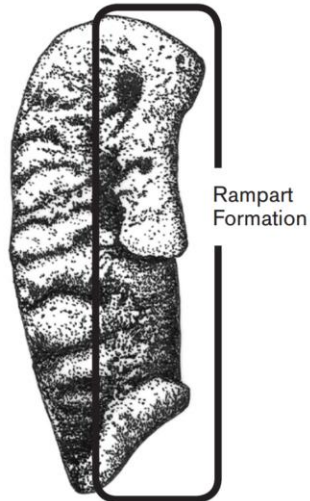
Figure 5.3.3. A male pubis with a prominent ventral arc that would be scored a two per Klaes *et al.* (2012). The scale is in centimeters.



Figure 5.3.4. A male pubis (on the left) that has a more prominent ventral arc than a female pubis (on the right). The scale is in centimeters.

5.4 Age Estimation: Figures

IV. Ventral Symphyseal Margin



LOCATION

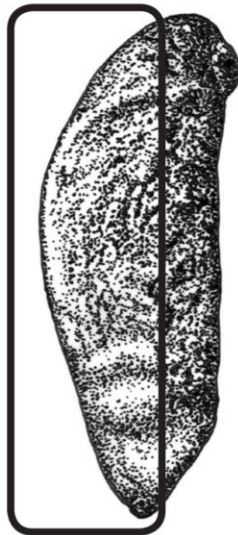
The ventral margin of the pubic symphysis is scored separately from the rest of the bone (Figure 2.18).

CHARACTERISTICS [STAGES]

1. Serrated
2. Beveled
3. Rampart formation
4. Rampart completion with anterior sulcus
5. Rampart completion without sulcus
6. Rim
7. Breakdown

Figure 5.4.1. The ventral symphyseal margin component (from Langley *et al.* 2016), which had the strongest correlation with age.

II. Dorsal Symphyseal Texture



LOCATION

The dorsal portion of the surface is examined (Figure 2.8).

CHARACTERISTICS [STAGES]

1. Smooth (fine-grained)
2. Coarse-grained
3. Microporosity
4. Macroporosity

Figure 5.4.2. The dorsal symphyseal texture component (from Langley *et al.* 2016), which had the weakest correlation with age.



Figure 5.4.3. A female who was 75 years old at the time of her death, who had her age predicted by ADBOU within 3 years. The disfigured ventral rim and pronounced macroporosity were likely helpful in producing an accurate age estimate for a much older individual. The scale is in centimeters.

CHAPTER 6: CONCLUSIONS

6.1 Sex Estimation

The Klales *et al.* (2012) method for estimating sex from the pubis did not perform as well as reported in the original study. The sample upon which the present study was conducted was larger, and more modern than the sample upon which the method was created in the original study. Secular change undoubtedly played a role in the poor performance of the method (Klales *et al.* 2014), with an overall classification accuracy over 10 percent lower than the original study, which was 86.2 percent (Klales *et al.* 2012). The majority of the individuals who were incorrectly sexed were males, misclassified as females. This was likely a result of the presence of a ventral arc or similar appearing ligamentous outgrowth on the majority of male pubes. Examining the left pubes of 328 males, only 32 of them had ventral arcs which were clearly a score of five. This indicates that the majority of males (296 of 328 males) had some type of ventral arc, even if it was slight. The ventral arc is the most heavily weighted trait in the logistic regression, because it is classically considered a female trait (Phenice 1969; Klales *et al.* 2012). The logistic regression would benefit from re-calibration, taking into account that a ventral arc is not exclusively a female trait in modern American populations. The method has already been recalibrated on a Mexican sample, and was shown to have produced much higher classification accuracies after using data from the new sample to adjust the logistic regression (Gómez-Valdés *et al.* 2017). More samples should be tested and then used to recalibrate the logistic regression. Although Phenice (1969) produced better classification accuracies, it is not *Daubert* (1993) compliant and

has no statistical weight to the sex estimations. Improving the Kiales *et al.* (2012) method so it has a level of accuracy similar to the Phenice (1969) method could alleviate these issues.

Additionally, the descriptions of each possible ordinal score in the Kiales *et al.* (2012) system should be updated to reflect this new data. Several males in the sample had ventral arc scores of three or two. Males actually proportionally had more ventral arc scores of two or three than did females. This would produce a high posterior probability of female group membership using the logistic regression, even if the pubis still appears more male in morphology per Phenice (1969). In his 2005 paper, Walker decided to validate the use of the greater sciatic notch for sex estimation. The original scoring system from Buikstra and Ubelaker (1994) scores the greater sciatic notch from one (very wide; clearly female) to five (very narrow; clearly male). Walker (2005) found that a much higher percentage of males had a greater sciatic notch score of two and three than did females. Therefore, what would normally be considered clear female morphology (a score of two) or ambiguous morphology (score of three) was actually more common in males. Similar results were found in the present study. In modern populations, a prominent ventral arc (that may even be scored a two) is not an uncommon trait in males. The most common ventral arc score in males was a three, which due to the heavy statistical weight in the logistic regression, would push the posterior probability toward female group membership. More research should be done to see what range of ventral arcs occur in modern populations, and how their morphology differs between males and

females. If males from modern populations have a variety of possible ventral arc scores, it could render the logistic regression unusable.

Another issue that was encountered with using both the descriptions, images, and drawings provided in the Klales *et al.* (2012) study is that there were some common morphologies of the medial aspect of the ischiopubic ramus that were not accounted for. Both Phenice (1969) and Klales *et al.* (2012) discuss that females will have narrower, more “pinched” ischiopubic rami with a prominent, sometimes sharp ridge of bone projecting from the bottom of the symphyseal face. The studies also state that males tend to have wider ischiopubic rami with no ridges of bone. Both observers noted several types of morphological variation during the course of the present study which introduced complication and possible bias into the study. Firstly, it was observed that it is not uncommon for females to have wider ischiopubic rami relative to the width of the symphyseal face (i.e. scores of three, and sometimes four). Females with narrower rami sometimes had flat medial aspects that lacked any ridge of bone at all. Secondly, while most males did have wider ischiopubic rami, several were found to have very clear ridges of bone on the medial aspect, projecting off of the inferior symphyseal face. This has classically been attributed to female morphology, but it is not an uncommon male variant. More skeletal samples accumulated from modern American populations should be studied to see how wide this range of variation is for the ischiopubic ramus and its medial ridge. The Klales *et al.* (2012) descriptions and exemplars could be updated to include these possible morphological variations and how to score them so observers are not left to make the decision for themselves. In the present study, the observers chose to weight the

overall width of the ischiopubic ramus relative to the symphyseal face over the presence of a ridge. This seems to be a better correlate with sex than the presence of a ridge on the medial aspect, but could have also been a factor introducing error and bias into the sex estimates.

Despite the difficulty in scoring, the medial aspect of the ischiopubic ramus had the highest level of interobserver agreement (although it was just moderate). The subpubic concavity and ventral arc scores had poor levels of interobserver agreement. This is likely due to varied levels of experience using the system, but also interpreting the descriptions and exemplars from the Klales *et al.* (2012) study. While the authors were attempting to cover a wider range of variation by expanding the number of available traits scores, they were inadvertently introducing an increased possibility for interobserver error. One observer could score a ventral arc a four, and a different observer could score the same ventral arc a three. Although those two scores are only one ordinal interval from each other, they would lead to two completely different sex estimates when entered into the logistic regression equation (barring extreme scores for both the other traits). Although the trait scores from Klales *et al.* (2012) had only fair to moderate levels of interobserver agreement, the two observers achieved similar overall classification accuracies using the logistic regression equation. This indicates that there are larger issues with the system than just reliability, although it certainly should be addressed. Phenice (1969) sex estimates have a much higher level of interobserver agreement because the method involves using the overall morphology of the three traits in conjunction with each other. So while the Klales *et al.* (2012) method gives us statistical

weights to our sex estimates, the scores by which they are achieved are not scored in the same way by different observers. A useful study to perform would be to apply a similar logistic regression approach to the scoring system from Buikstra and Ubelaker (1994) to see if less trait scores leads to higher levels of reliability when estimating sex from the three pubic traits. If there are only three scores to choose from rather than five, it could greatly increase the reliability of the method while simultaneously having known statistical probabilities and being *Daubert* (1993) compliant.

The next step in the validation research for the Klales *et al.* (2012) will be to test the method on other modern samples that are representative of the forensic cases upon which it may be used. The results and samples should be compared between validation studies to understand what issues are leading to low accuracy and only fair reliability. Any modern skeletal samples available to researchers should be investigated to see how secular change is affecting the morphology of pelvic traits and how wide the range of variation really is for the three Phenice (1969) traits. The original study was created using the Hamman-Todd skeletal collection, which is a historical collection (Klales *et al.* 2012), and validated using the Bass donated skeletal collection. These two skeletal collections are utilized to create a large percentage of sex estimation methods because of their accessibility and demographic information. Our methods are only as good as the skeletal collections upon which they are based, therefore newer ones must be curated and made accessible to researchers to continue improving forensic anthropological techniques. Additionally, we know that the original method performed poorly on a Mexican sample, so samples that are not mostly White individuals should be tested as

well. To improve the Klales *et al.* (2012) method, new populations should be analyzed and special attention should be paid to how the ventral arc appears on the males in the sample, and how common a bony ridge on the medial aspect of the ischiopubic ramus is in males. New descriptions and exemplars should be published to improve the interobserver reliability of scoring the ventral arc and the subpubic contour. The logistic regression should be continually improved and re-calibrated once these new patterns are studied. The ventral arc score should be adjusted to have lower statistical weight in the logistic regression, or scores of three should be weighted as ambiguous rather than more female. The ultimate goal should be a logistic regression that can produce similar or higher classification accuracies to other systems, especially the Phenice (1969) method upon which the Klales *et al.* (2012) study is based. This author proposes another follow-up study which should be performed: a table could be set up at an American Academy of Forensic Sciences or American Association of Physical Anthropologists conference, and any willing attendees could score a small sample of pubes using the Klales *et al.* (2012) method. A much larger scale interobserver study like this could allow inferences to be made about which specific portions of the method are causing its less than substantial reliability, and the low accuracy rates.

6.2 Age Estimation

There are many advantages to using the ADBOU software package. It is very user-friendly and easy to use and understand. It allows an age estimate to be derived even if some components are missing or cannot be scored. Choosing a component score is less

subjective than assigning a phase, because more variation can be accounted for by scoring each component separately, and the Bayesian formulae can consider all available scores together to create an age estimate instead of requiring assignment of a “best-fit” phase (i.e. Brooks and Suchey 1990). The accompanying Langley *et al.* (2016) manual has excellent descriptions and images to help with the scoring of the components. However, the Boldsen *et al.* (2002) method is time consuming. Choosing the score for each component includes a “minimum” and “maximum” value, as well as assigning them to both the left and right sides if they are available. The manual must be consulted every time to ensure the correct score is being chosen. However, the relatively good performance of the system makes this a minor consideration. An accuracy rate of 82 percent with and only a slight negative bias (about -0.02) indicate the method shows promise for future forensic case usage, considering that only the pubic components were used (and not cranial sutures or the sacroiliac surface).

Bayesian methods for estimating age have become more popular in recent years, and seem to have promise as far as accurate age at death estimations (DiGangi *et al.* 2009; Shirley-Langley and Jantz 2010; Milner and Boldsen 2012). However, considering that the majority of the often-used methods in forensic cases since McKern and Stewart (1957) have been phase-based, (Işcan 1984; Işcan 1985; Brooks and Suchey 1990; Osborne 2004; Hartnett 2010a), these newer component-based methods need repeated validation to ensure they work for modern forensic cases. This includes validation on newer skeletal samples which have similar age at death distributions to the living modern American population, and making sure that individuals who are not White or of European

ancestry are observed whenever possible. The ADBOU software package is able to produce an age estimate if ancestry is unknown, but it the only two populations it has available for known ancestry are American White and Blacks. While this is an issue caused solely by the availability of other populations in donated skeletal collections, it is something that still needs to be considered and noted when applying these methods to individuals who are not White.

Each of the areas that are utilized in the ADBOU software package should be tested independently, as done with the pubis in this study, to see how well they perform on their own. The validation study by Milner and Boldsen (2012), which tested each of the three areas separately, should be repeated on larger samples. If enough validation studies are performed analyzing each area separately, then the more “useful” areas that correlate strongly with age can be identified and included in the transition analysis software. The ones that are not useful for estimating age can be eliminated. Since cranial sutures seem to greatly underestimate age (Milner and Boldsen 2012), eliminating them from ADBOU could be beneficial to the success of the method. The sacroiliac surface should be revisited and scored on a newer, larger sample as well to see how each of its components correlates with age. Considering that the ADBOU software program is now the official age estimation technique in the Langley *et al.* (2016) manual for forensic anthropology standards, it should be subjected to much more validation in the future. This author proposes the next step in this research would be to find a large, modern sample of crania and os coxa, and repeat the procedure in the present study on those areas. Once we understand how well the various elements in ADBOU perform on their

own, we can begin to eliminate the areas with poor correlation with age and begin testing new areas that may perform better.

Additionally, we know that the pubis performs relatively well for age estimation on its own (Milner and Boldsen 2012) considering past studies and the results from the present study. Other skeletal samples could be observed to see how common certain component scores really are, and at what age the transitions of morphology tend to occur, and this could be used to recalibrate the ADBOU software formulae. Since the Boldsen *et al.* (2002) method is based on Bayes' theorem which involves multi-variable conditional probabilities, having new individuals to include in the statistical model will hopefully help eliminate some of the inherent inaccuracy by accounting for more morphological variation. Even if age ranges cannot become narrower, they can hopefully at least be made more accurate. Also, identifying other components of the pubis that may correlate strongly with age could be useful. Phase-based methods such as the system in Hartnett (2010a) include age indicators like the general shape of the symphyseal face, the quality of the ventral and dorsal bone, and the present of ligamentous outgrowths or exostoses. These traits clearly correlate with age as they have been used in a relatively successful phase-based age estimation system, so they could be subjected to Bayesian analysis to see if they can be scored as components and used in a transition analysis system. The results of the present study indicate that using just the pubis in ADBOU produces negatively biased age estimates by about six years on average. Older individuals commonly have their ages underestimated. If more component scores can be added, or different

components could be considered that better correlate with age in older individuals, then that would be a beneficial piece of research as well.

A Bayesian approach to aging from the sternal rib ends may have promise and should be considered in future studies. They have been used for aging in phase-based methods for years (Işcan 1984; Işcan 1985; Hartnett 2010b), and the first rib has been used in a component-based method (DiGangi *et al.* 2009). These previous studies have considered the depth of the sternal pit, the bone quality and translucency, and the bone quality instead the sternal pit (Işcan 1984; Hartnett 2010b). These traits that are traditionally used to identify the rib as a certain phase could be scored as components that can be the basis for a Bayesian statistical model. A new Bayesian approach to aging from sternal rib ends would require the identification of components and the creation of component scores that could be assigned to them. These components would then have to be tested to see how they correlate with age. Theoretically, transition analysis is a framework for age estimation that can be applied to any area of the skeleton that experiences degenerative changes that are somewhat predictable (Boldsen *et al.* 2002). It is then very likely that applying Bayesian statistics to sternal rib ends could have some applicability in age estimation. The ultimate goal for forensic anthropologists should be to have method to estimate age from as many elements as possible. The ADBOU software which accompanies the Boldsen *et al.* (2002) method consolidates the various age estimation processes (at least, from three skeletal areas) into one estimated age range. Since there seems to be no firm agreement between practitioners on how they produce their final age range to put in their report to law enforcement (Garvin and Passalacqua

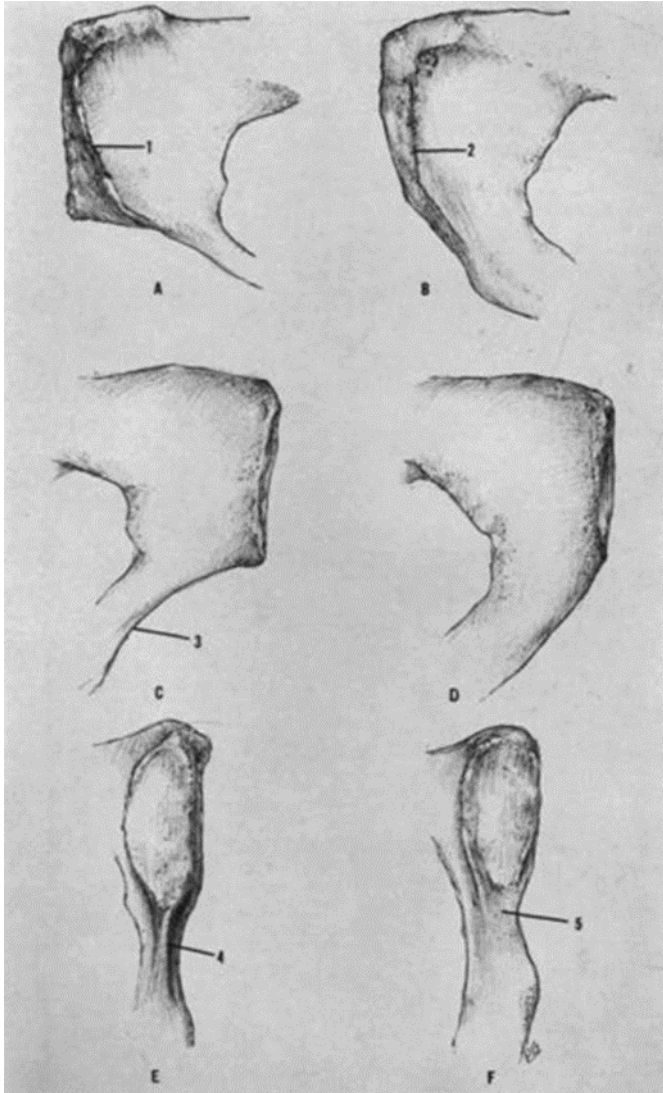
2011), utilizing an age estimation software in the future could eradicate the equivocality in this portion of the biological profile.

6.3 Summary of Conclusions and Implications

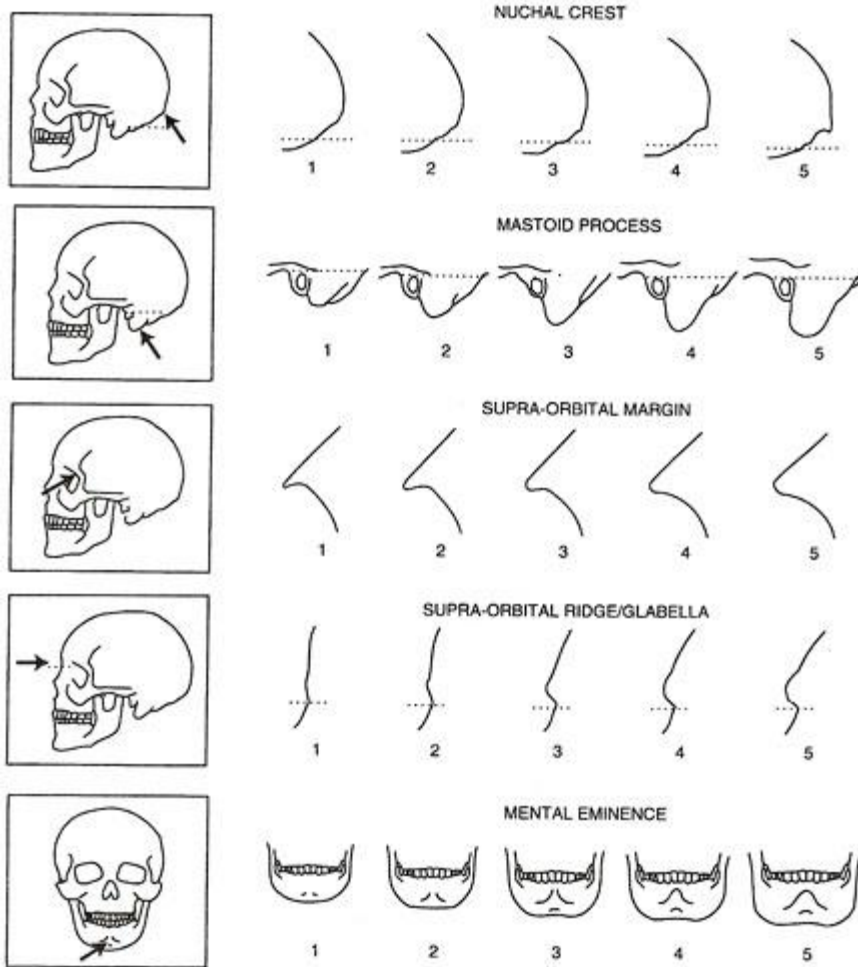
In conclusion, validation studies are crucial to the field of forensic anthropology. While every researcher seems to be keen on creating and publishing new methods for producing sex and age estimates, there seems to be considerably less interest in validating existing methods. As the modern American population evolves, so should the way we conduct research. The limited availability of skeletal collections upon which to base our research can lead to bias and unreliability when these methods are used on actual forensic cases, because resulting estimations tend to mimic the demographics of the collection they are created from (Boldsen *et al.* 2002; Algee-Hewitt 2013). The field of forensic anthropology is moving towards more standardized, statistical methods as the uniform approach to creating a biological profile (Algee-Hewitt 2013), in an effort to make results more reproducible and less subjective. The only way to achieve such desired results is to continue performing validation studies on any new methods that are being used to create biological profiles in a medicolegal context. If we as a field wish to continue compliance with *Daubert* (1993) standards, these methods must be widely peer reviewed and accepted by the field. The Klales *et al.* (2012) methods and Boldsen *et al.* (2002) methods are being taught to new practitioners and even utilized in actual forensic cases. The present study used a modern, autopsy sample that was very recently curated to validate methods which have been created on more historical samples (Hamman-Todd and Terry collections, which are historical and archaeological collections). This is necessary to see

how secular change can affect the results of these more standardized, statistical methods. This author concludes that while validation studies are not as glamorous and exciting as creating entirely new methods, they are infinitely more necessary to ensure the validity and soundness of the field of forensic anthropology and its body of work.

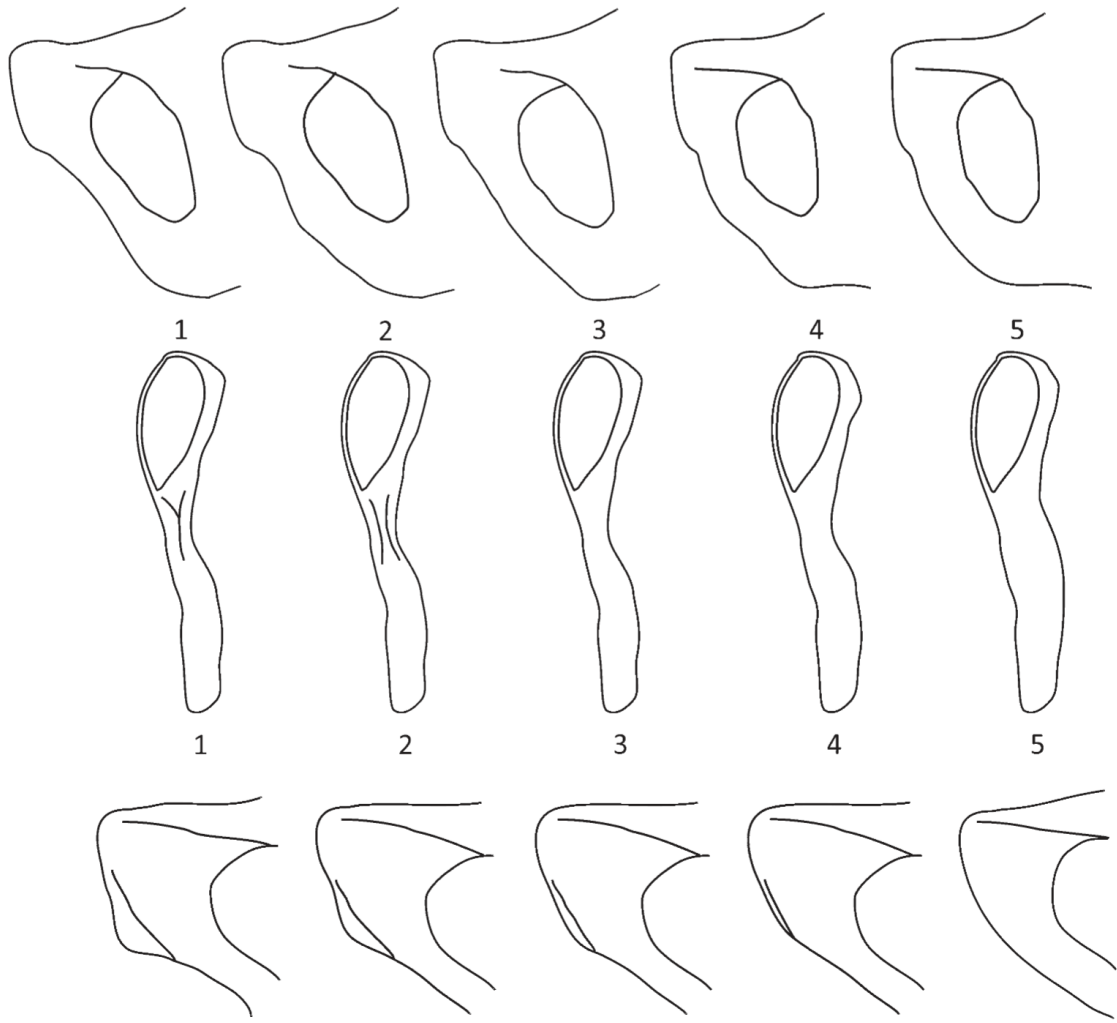
APPENDIX: EXEMPLARS AND SOFTWARE



A.1. The original drawn exemplars of the three pubic traits from the Phenice (1969) study. The top row shows the ventral arc, the middle row shows the subpubic concavity, and the bottom row shows the medial aspect of the ischiopubic ramus. The left column is typical female morphology; the right column is typical male morphology.



A.2. The original drawn exemplars from Buikstra and Ubelaker (1994) showing the various morphological traits of the skull used for sex estimation and their possible ordinal scores.



A.3. The original drawn exemplars from Klales *et al.* (2012). The top row shows the subpubic contour, the middle row shows the medial aspect of the ischiopubic ramus, and the bottom row shows the ventral arc. A score of 1 is indicative of clear female morphology; a score of 5 is indicative of clear male morphology.

Answer Options	1	2	3	4	5	Average Rank	<i>n</i>
Pubic symphysis	103	23	4	2	0	1.28	132
Sternal rib ends	16	63	38	10	6	2.45	133
Auricular surface	9	42	61	14	4	2.71	130
Cranial sutures	0	3	11	57	58	4.32	129
Dental wear	4	1	14	46	65	4.28	130

Question presented: please rank the following skeletal regions according to your personal preference and their reliability in adult age at death estimation, with 1 being the most, and 5 being the least preferred and reliable.

A.4. A chart from Garvin and Passalacqua (2012), indicating survey answers for preferred areas to estimate sex from forensic anthropology practitioners.

I. Symphyseal Relief

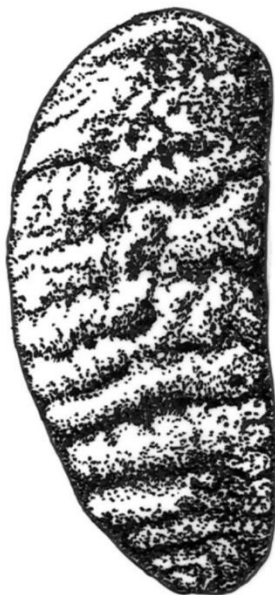


Fig. 2.1 **Symphyseal Relief**
(Note: dorsal is on the left side, ventral on the right in all pubic symphysis figures)

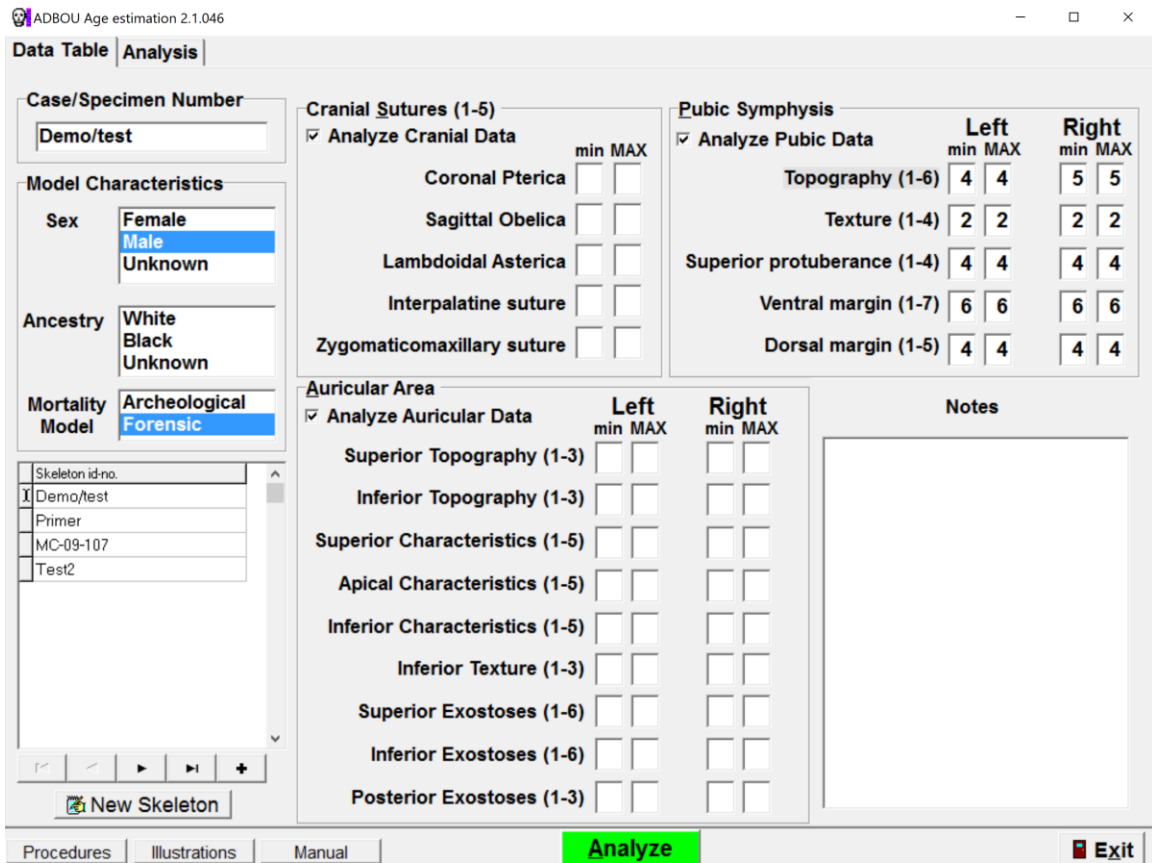
LOCATION

The terms for this feature generally follow those of McKern and Stewart (1957). Although the entire face should be considered, the billowing is typically most clearly seen in the dorsal half of the symphyseal face (Figure 2.1). In fact, low ridges of bone, the billowing, can be entirely absent from the ventral symphyseal face, beginning as early as the ventral beveling stage.

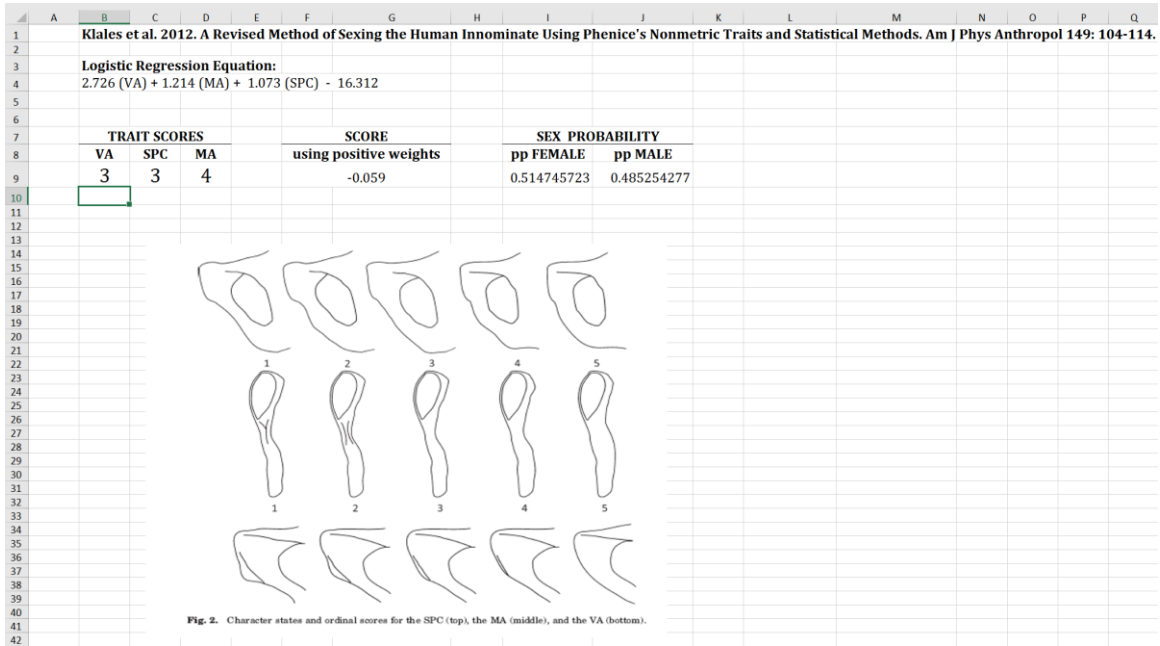
CHARACTERISTICS [STAGES]

1. Sharp billowing
2. Soft, deep billowing
3. Soft, shallow billowing
4. Residual billowing
5. Flat
6. Irregular

A.5. A drawn example of a component (symphyseal relief) and its possible component scores with their descriptions (characteristics 1-6) taken from the Langley *et al.* (2016) manual used for the ADBOU software.



A.6. The ADBOU 2.1.046 transition analysis software package home screen, where the component scores and known information (i.e. sex, ancestry, mortality model) are entered.



A.7. The Microsoft Excel document containing the logistic regression equation in which to enter ordinal scores, and the exemplars, for the Klaes *et al.* (2012) method.

Table 3 Interpretation of Kappa values given by Landis and Koch (1977)

Kappa	Agreement
< 0	Less than chance agreement
0.01 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 0.99	Almost perfect agreement

A.8. A chart showing how to interpret Kappa values, from Landis and Koch (1977).

REFERENCES

- Acsádi, G., Nemeskéri, I. (1970). *History of Human Life Span and Mortality*. Budapest, Hungary: Akadémiai Kiado.
- Albanese, J., Cardoso, H.F.V., Saunders, S.R. (2005). Universal methodology for developing univariate sample-specific sex determination methods: an example using the epicondylar breadth of the humerus. *Journal of Forensic Sciences*, 32,143–152.
- Albert, A.M., Maples, W.R. (1995). Stages of epiphyseal union for thoracic and lumbar vertebral centra as a method of age determination for teenage and young adult skeletons. *Journal of Forensic Sciences*, 40, 623–633.
- Algee-Hewitt, B. (2013). Age estimation in modern forensic anthropology. In M.A. Tersigni-Tarrant & N.R. Shirley (Eds.), *Forensic anthropology: An introduction* (pp. 181–230). Portland, OR: CRC Press.
- Boldsen, J.L., Milner, G.R., Konigsberg, L.W., Wood, J.W. (2002). Transition analysis: a new method for estimating age from skeletons. In R.D. Hoppa & J.W. Vaupel (Eds.), *Paleodemography: age distributions from skeletal samples* (pp. 73–106). Cambridge, UK: Cambridge University Press.
- Brennaman, A.L., Love, K.R., Bethard, J.D., Pokines, J.T. (2016). A Bayesian approach to age at-death estimation from osteoarthritis of the shoulder in modern North Americans. *Journal of Forensic Sciences*, doi:10.1111/1556-4029.13327.
- Brooks, S.T. (1955). Skeletal age at death: the reliability of cranial and pubic age indicators. *American Journal of Physical Anthropology*, 13, 567–589.
- Brooks, S., Suchey, J.M. (1990). Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. *Human Evolution*, 5, 227–238.
- Brothwell, D. (1989). The relationship of tooth wear to aging. In M.Y. Işcan (Ed.), *Age Markers in the Human Skeleton* (pp. 303–316). Springfield, MA: Charles C. Thomas.
- Brues, A.M. (1958). Identification of skeletal remains. *The Journal of Criminal Law, Criminology, and Police Science*, 48, 551–563.

- Buckberry, J.L., Chamberlain, A.T. (2002). Age estimation from the auricular surface of the ilium: a revised method. *American Journal of Physical Anthropology*, 119, 231–239.
- Buikstra, J.E., Ubelaker, D.H. (1994). *Standards for data collection from human skeletal remains*. Fayetteville, AR: Arkansas Archeological Survey.
- Call S. 2016. Sex estimation in forensic anthropology: A test of the Klales *et al.* (2012) method with implications of asymmetry. Master's thesis. University of Alaska, Fairbanks.
- Daubert v. Merrell Dow Pharmaceuticals, Incorporated*, 509 US 579, 1993.
- DiGangi, E.A., Bethard, J.D., Kimmerle, E.H., Konigsberg L.W. (2009). A new method for estimating age-at-death from the first rib. *American Journal of Physical Anthropology*, 138, 164–176.
- France, D.L. (1998). Observation and metric analysis of sex in the skeleton. In K.L. (Ed.), *Forensic osteology: advances in the identification of human remains* (pp. 163–186). Springfield, IL: Charles C. Thomas.
- Franklin, D., Freedman, L., Milne, N. (2005). Three-dimensional technology for linear morphological studies: a re-examination of cranial variation in four southern African indigenous populations. *Homo – Journal of Comparative Human Biology*, 56, 17–34.
- Garvin, H.M., Passalacqua, N. (2011). Current practices by forensic anthropologists in adult skeletal age estimation. *Journal of Forensic Sciences*, 57, 427–433.
- Gilbert, B.M., McKern, T.W. (1973). A method for aging the female os pubis. *American Journal of Physical Anthropology*, 38, 31–38.
- Giles, E., Elliot, O. (1963). Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology*, 21, 53–68.
- Gómez-Valdés, J.A., Garmendia, A.M., García-Barzola, L., Sánchez-Mejorada, G., Karam, C., Baraybar, J.P., Klales, A. (2017). Recalibration of the Klales *et al.* (2012) method of sexing the human innominate for Mexican populations. *American Journal of Physical Anthropology*, 162, 600–604.
- González, P.N., Bernal, V., Pérez, S.I., Barrientos, G. (2007). Analysis of dimorphic structures of the human pelvis: its implications for sex estimation in samples without reference collections. *Journal of Archaeological Science*, 34, 1720–1730.

- Gustafson, G. (1950). Age determination on teeth. *Journal of the American Dental Association*, 41, 45–55.
- Hartnett, K.M. (2010a). Analysis of age-at-death estimation using data from a new, modern autopsy sample—Part I: pubic bone. *Journal of Forensic Sciences*, 55, 1145–1152.
- Hartnett, K.M. (2010b). Analysis of age-at-death estimation using data from a new, modern autopsy sample —Part II: Sternal end of the fourth rib. *Journal of Forensic Sciences*, 55, 1152–1156.
- Hoppa, R.D., Vaupel, J.W. (2002) The Rostock Manifesto for paleodemography: the way from stage to age. In R.D. Hoppa & J.W. Vaupel (Eds.), *Paleodemography: age distributions from skeletal samples* (pp. 1–8). Cambridge, UK: Cambridge University Press.
- Işcan, M.Y. (1988). Rise of Forensic Anthropology. *Yearbook of Physical Anthropology*, 31, 203–230.
- Işcan, M.Y, Loth, S.R. (1989). Osteological manifestations of age in the adult. In M.Y. Işcan and A.R. Kennedy (Eds.), *Reconstruction of Life from the Skeleton* (pp. 23–40). New York, NY: Alan R. Liss.
- Kemkes-Grottenthaler A. (2002). Aging through the ages: historical perspectives on age indicator methods. In R.D. Hoppa & J.W. Vaupel (Eds.), *Paleodemography: age distributions from skeletal samples* (pp. 48–72). Cambridge, UK: Cambridge University Press.
- Kimmerle, E.H., Konigsberg, L.W., Jantz, R.L., Baraybar, J.P. (2008). Analysis of age at-death estimation through the use of pubic symphyseal data. *Journal of Forensic Sciences*, 53, 558–568.
- Kjellstrom, A. (2004). Evaluations of sex assessment using weighted traits on incomplete skeletal remains. *International Journal of Osteoarchaeology*, 14, 360–373.
- Klales, A.R. (2014). Secular change in morphological pelvic traits used for sex estimation. *Journal of Forensic Sciences*, 61, 295–301.
- Klales, A.R., Ousley, S.D., Vollner, J.M. (2012). A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *American Journal of Physical Anthropology*, 149, 104–114.
- Konigsberg, L.W. (2015). Multivariate cumulative probit for age estimation using ordinal categorical data. *Annals of Human Biology*, 42, 368–378.

- Konigsberg, L.W., Frankenberg, S.R. (1992). Estimation of Age Structure in Anthropological Demography. *American Journal of Physical Anthropology*, 89, 235–256.
- Konigsberg, L.W., Frankenberg, S.R., Walker, R.B. (1997). Regress what on what? Paleodemographic age estimation as a calibration problem. In R.R. Paine (Ed.), *Integrating archaeological demography* (pp. 64–88). Center for Archaeological Investigations: Southern Illinois University.
- Konigsberg, L.W., Hens, S.M. (1998). Use of ordinal categorical variables in skeletal assessment of sex from the cranium. *American Journal of Physical Anthropology*, 107, 97–112.
- Komar, D.A., Buikstra, J.E. (2008). *Forensic anthropology: Contemporary theory and practice*. Oxford, UK Oxford University Press.
- Lamendin, H., Baccino, E., Humbert, J.F., Tavernier, J.C., Nossintchouk, R.M., Zerilli, A. (1992). A simple technique for age estimation in adult corpses: the two criteria dental method. *Journal of Forensic Sciences*, 37, 1373–1379.
- Landis, J., Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Langley, N.R., Meadows Jantz, L., Ousley, S.D., Jantz, R.L., Milner, G. (2016). *Data collection procedures for forensic skeletal material 2.0*. Knoxville, TN: Forensic Anthropology Center.
- Langley-Shirley, N., Jantz, R.L. (2010). A Bayesian approach to age estimation in modern Americans from the clavicle. *Journal of Forensic Sciences*, 55, 57–583.
- Lovejoy, C.O., Meindl, R.S., Pryzbeck, T.R., Mensforth, R.P. (1985). Chronological metamorphosis of the auricular surface of the ilium: A new method for the determination of adult skeletal age at death. *American Journal of Physical Anthropology*, 68, 15–28.
- Lovell, N.M. (1989). Test of Phenice's technique for determining sex from the os pubis. *American Journal of Physical Anthropology*, 79, 117–120.
- Mann, R.W., Jantz, R.L., Bass, W.M., Willey, P.S. (1991). Maxillary suture obliteration: A visual method for estimating skeletal age. *Journal of Forensic Sciences*, 36, 781–791.
- McArthur, T.A., Meyer, I., Jackson, B., Pitt, M.J., Larrison, M.C. (2016). Parturition pit: the bony imprint of vaginal birth. *Skeletal Radiology*, 45(9), 1263–1267.

- McKern, T.W., Stewart, T.D. (1957). Skeletal age changes in young American males. *Technical Report EP-45*. Natick, MA: Quartermaster Research and Development Command.
- Meindl, R.S., Lovejoy, C.O. (1985). Ectocranial suture closure: A revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *American Journal of Physical Anthropology*, 68, 57–66.
- Milner, G.R. (1992). *Determination of skeletal age and sex: A manual prepared for the Dickson Mounds Reburial Team*. Lewiston, IL: Dickson Mounds Museum.
- Milner, G.R., Boldsen, J.L. (2012). Transition analysis: A validation study with known age modern American skeletons. *American Journal of Physical Anthropology*, 148, 98–110.
- Mincer, H.H., Harris, E.F., Berryman, H.E. (1993). The A.B.F.O. study of third molar development and its use as an estimator of chronological age. *Journal of Forensic Sciences*, 38, 379–390.
- Montagu, M.F.A. (1938). Aging of the skull. *American Journal of Physical Anthropology*, 23, 355–375.
- Osborne, D.L., Simmons, T.L., Nawrocki, S.P. (2004). Reconsidering the auricular surface as an indicator of age at death. *Journal of Forensic Sciences*, 49, 905–911.
- Phenice, T.W. (1969). A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology*, 30, 297–302.
- Prince, D.A., Kimmerle, E.H., Konigsberg, L.W. (2008). A Bayesian approach to estimate skeletal age-at-death utilizing dental wear. *Journal of Forensic Sciences*, 53, 588–593.
- Prince, D.A., Konigsberg, L.W. (2008). New formulae for estimating age-at-death in the Balkans utilizing Lamendin's dental technique and Bayesian analysis. *Journal of Forensic Sciences*, 53, 578–587.
- Ramsthaler, F., Kettner, M., Gehl, A., Verhoff, M.A. (2010). Digital forensic osteology: Morphological sexing of skeletal remains using volume-rendered cranial CT scans. *Forensic Science International*, 195(1-3), 148–152.
- Rogers, T.L. (1999). A visual method of determining the sex of skeletal remains using the distal humerus. *Journal of Forensic Sciences*, 44, 57–60.

- Rogers, T.L. (2005). Determining the sex of human remains through cranial morphology. *Journal of Forensic Sciences*, 50, 493–500.
- Rogers, N.L., Flournoy, L.E., McCormick, W.F. (2000). The rhomboid fossa of the clavicle as a sex and age estimator. *Journal of Forensic Sciences*, 45, 61–67.
- Scheuer, L., Black, S. (2000). *Developmental Juvenile Osteology*. Oxford, UK: Elsevier.
- Shirley, N.R., Jantz, R.L. (2011). Spheno-occipital synchondrosis fusion in modern Americans. *Journal of Forensic Sciences*, 56, 580–585.
- Spradley, M.K., Jantz, R.L., Robinson, A., Peccerelli, F. (2008). Demographic change and forensic identification: problems in metric identification of Hispanic skeletons. *Journal of Forensic Sciences*, 53, 21–28.
- Stewart, T.D. (1948). Medico-legal Aspects of the skeleton; Age, sex, race and stature. *American Journal of Physical Anthropology*, 6, 315–21.
- Stewart, T.D. (1954). Sex determination of the skeleton by guess and by measurement. *American Journal of Physical Anthropology*, 12, 385–389.
- Suchey, J.M., Wiseley, D.V., Green, R.F., Noguchi, T.T. (1979). Analysis of dorsal pitting in the os pubis in an extensive sample of modern American females. *American Journal of Physical Anthropology*, 51, 517–540.
- Todd, T.W. (1920). Age changes in the pubic bone. I. The male white pubis. *American Journal of Physical Anthropology*, 3, 285–334.
- Todd, T.W. (1921). Age changes in the pubic bone. *American Journal of Physical Anthropology*, 4, 1–70.
- Todd, T.W., Lyon Jr., D.W. (1924). Endocranial suture closure. Part I: Adult males of white stock. *American Journal of Physical Anthropology*, 7, 325–384.
- Vance, V.L., Steyn, M., L'Abbé, E.N. (2011). Nonmetric sex determination from the distal and posterior humerus in Black and White South Africans. *Journal of Forensic Sciences*, 56, 710–714.
- Walker, P.L. (2005). Greater sciatic notch morphology: Sex, age, and population differences. *American Journal of Physical Anthropology*, 127, 385–391.
- Webb, P.A., Suchey, J.M. (1985). Epiphyseal union of the anterior iliac crest and medial clavicle in a modern multiracial sample of American males and females. *American Journal of Physical Anthropology*, 68, 457–466.

CURRICULUM VITAE

