University of Tennessee, Knoxville

Trace: Tennessee Research and Creative Exchange

Masters Theses   Graduate School

8-2018

Age-at-death Estimation: Accuracy and Reliability of Age-Reporting Strategies

Christine Ashley Bailey
University of Tennessee, creilly1@vols.utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Christine Ashley Bailey entitled "Age-at-death Estimation: Accuracy and Reliability of Age-Reporting Strategies." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts, with a major in Anthropology.

Dawnie W. Steadman, Giovanna M. Vidoli, Major Professor

We have read this thesis and recommend its acceptance:

Amy Z. Mundorff

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Age-at-death Estimation: Accuracy and Reliability of Age-Reporting Strategies

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Christine Ashley Bailey
August 2018

ACKNOWLEDGEMENTS

This thesis would not be possible without the help and support of my committee. Thank you, Drs. Giovanna Vidoli, Dawnie Steadman, and Amy Mundorff for your advice, encouragement, and multiple drafts of edits. Giovanna, your guidance and patience over the last three years have been instrumental in my growth as scholar and my ability to complete this work. I truly would not be where I am without you.

Thank you to the University of Tennessee faculty and staff members who have served as mentors to me. Drs. Joanne Devlin and Lee Jantz, I have immensely enjoyed the opportunity to work with both of you. Your guidance and encouragement has made be a better osteologist and instructor. Though, I am mostly thankful that your offices were always open to me. Dr. Trish Hepner, you have always encouraged me to think outside the box and to bring my passions of forensic anthropology and DDHR together. I am forever grateful for your confidence in me and I am glad to have become your friend throughout my time here. Mary Davis, you have been a huge help throughout the thesis writing process and I cannot thank you enough for that. I appreciate the times when you listened to me rant, let me bounce ideas off of you, and gave thoughtful advice when it was needed. Thank you for always bringing me back to reality, especially when I underestimated myself.

I cannot express enough gratitude to my friends and family who believed in me, especially during the times when I did not believe in myself. Cory, your unwavering support and love is what has kept me going and lifted me up when I was at my lowest points. You will never know how much I appreciated the times when you had dinner

ready when I got home, picked up the slack around the house, and forced me to take time for myself. All of the sacrifices we have made have brought us to this point and I am so glad I can share this success with you. Susan, you are one of the brightest parts of my life. I am so glad that you have been around throughout my graduate school journey. The countless laughs we have shared, episodes of iZombie we have binged, and moments of support we have shown each other have meant the world to me.

Haley, Lana, and Sam, you have been the best support system and friends anyone could have ever asked for. This journey would not have been the same without you. Thank you for the laughs, the late-night talks, and the many spontaneous adventures, among everything else. No matter where our journeys take us next, I know we will always be there for each other.

Thank you to the many friends I have made throughout my journey at UTK. I will always remember the Friday night happy hours, crazy hat parties, game nights, Halloween extravaganzas, spontaneous Sassy outings, and so much more. You have all made this time extraordinary!

Thank you to the friends I made before graduate school that have kept me sane throughout this process. Jacque, you have been my best friend since middle school and now we find ourselves finishing graduate school at the same time. I am glad our journey has brought us back together. Thank you for the late-night thesis-writing sessions, girl's nights, and the deep conversations. I can say confidently that we will always remain a part of each other's lives. To my Cracker Barrel Friends, thankfully most of us have moved on and are flourishing in our lives. I hope we remain as close as we still are. I

don't know what I would do without football get togethers, card nights, and crazy parties at Chris' house.

Finally, thank you Thelma Gann. You have always been there for me no matter what. You have been the mother that I never had, one that I did not deserve. Your encouragement, blunt opinions, and support in every sense has helped shape who I am today. I hope I have made you proud.

ABSTRACT

One task of a forensic anthropologist is to assist law enforcement in the identification of unknown human skeletal remains by building a biological profile. Age-at-death estimations are an informative aspect of biological profiles as they help law enforcement narrow down potential victim identifications. However, age-at-death estimation continues to be a challenge within forensic anthropology due to the uncertainty regarding method selection and the production of a final estimate for law enforcement.

The purpose of this research is to identify the age-reporting strategies that provide the most accurate and reliable (low inaccuracy and low bias) age-at-death estimations when evaluated by total sample, age-cohort (20-39; 40-59; 60-79), and sex. The age-reporting strategies in this study were derived from six age-at-death estimation methods and tested on 58 adult individuals (31 males and 27 females) from the William M. Bass Donated Skeletal Collection, at the University of Tennessee, Knoxville. An experience-based approach where the observer produced a final estimation using the data collected and their expert judgment was included to assess the appropriateness of experience-based estimations in medico-legal contexts. Two-way repeated measures ANOVAs were conducted to determine if there were significant differences in reliability between the age-reporting strategies.

The results show that the most accurate and reliable age-reporting strategy varied if the sample was evaluated as a whole, by age, or by sex. The most accurate and reliable strategy for the total sample was the experience-based approach. When

the sample was divided by age Suchey-Brooks pubic symphysis performed the best for the 20-39 age-cohort, the experience-based approach for the 40-59 age-cohort, and Buckberry-Chamberlain auricular surface for the 60-79 age-cohort. Finally, when separated by sex, Hartnett pubic symphysis performed the best for males and the experienced-based approach performed the best for females.

While none of the age-reporting strategies evaluated in this study were consistently the most accurate and reliable for all of the sample categories, the experience-based approach performed well in each category. This research helps shed light on the performance of different age-reporting strategies and provides further support to the reliance on multiple aging indicators and professional judgment in developing a final age-at-death estimation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

One primary task of a forensic anthropologist is to assist law enforcement in the identification of unknown human skeletal remains by building a biological profile through the estimation of sex, age, ancestry, stature, and pathology. Age-at-death estimation is a critical part of the biological profile but continues to be a challenge within anthropology due to the uncertainty regarding method selection and the production of a final age-at-death estimation.

The *Daubert v. Merrell Dow Pharmaceuticals (1993)* ruling established that a scientific theory or technique, being presented in federal court, should be judged on whether or not it has been validated, has been subject to peer review and publication, has a known error rate, has established standards, and has gained acceptance within the relevant scientific field. Positive identifications of unknown skeletal remains are often made through DNA, dental records, or through the comparison of antemortem and postmortem data (Parsons 2017). In instances when a positive identification cannot be obtained, presumptive identifications may be constructed through contextual and circumstantial evidence (Parsons 2017; Wiesema et al. 2009). If a case goes to trial with a presumptive identification then the methodologies used to construct the biological profile are subject to the *Daubert* criteria (Wiesema et al. 2009)

Age estimations derived from individual aging methods can meet the *Daubert* criteria, but often provide wide and often unhelpful age ranges. Reporting an age range that does not include the decedent's actual age-at death can prevent a positive

identification and reporting one that is too broad does not help narrow down the victim pool. While many studies have tested the accuracy and reliability of different aging methods, there is still a need for further investigation of how to systematically produce and report accurate age-at-death estimations, especially if information from multiple aging methods/ phases are being considered when developing a final age estimation (Baccino et al. 1999; Martrille et al. 2007; Merritt 2013; Saunders et al. 1992).

A major concern within forensic anthropology is how to arrive at a final age estimation to report to law enforcement in forensic cases. Should practitioners report a range from a single aging method, or should they report age using a combination of methods? How do you arrive at a final age estimation that is both accurate and reliable? While many scholars recommend using the results from multiple aging methods or age indicators when coming to a final age estimation (Baccino et al. 1999; Brooks and Suchey 1990; Lovejoy et al. 1985a; Merritt 2013; Parsons 2017; Saunders et al. 1992; Ubelaker 2010); others have found that statistically sound, multifactorial approaches do *not* perform significantly better than employing a single method (Martrille et al. 2007; Nawrocki 2010; Saunders et al. 1992; Schmitt et al. 2002). The last published standards for aging developed by the Scientific Working Group for Forensic Anthropology (SWGANTH) states that the best practice is to include all available information for a final age estimation and reporting results based on "expert judgement"(SWGANTH 2013). More recently, SWGANTH has been subsumed under Organization of Scientific Area Committees (OSAC) for Forensic Science which is working to strengthen forensic standards, including those regarding skeletal age estimation(NIJ 2017). Unfortunately,

OSAC has yet to publish the approved standards for age-at-death estimation at this time.

Determining how to combine information from multiple methods is challenging because each aging method was developed using different statistical methodologies, which typically cannot be combined in a statistically valid manner (Garvin and Passalacqua 2012; Nawrocki 2010; Uhl and Nawrocki 2010). When final age estimates are derived using information from multiple methods, they are often difficult to interpret or are combined in a way that does not meet any of the recommendations of the *Daubert* ruling. Therefore, determining how to report age in a way that meets forensic standards and aids in positive identification is crucial.

Important Nomenclature

There is nomenclature throughout this thesis that sounds similar but have distinct meanings. These terms are *chronological age, skeletal age, age indicator, age range, final age estimation, and age-reporting strategies*. Understanding the definition of each of these terminologies and being able to distinguish them is crucial for moving forward in this thesis. The definitions of each are provided below and remain consistent throughout this work.

(1) *Chronological Age:* The age of an individual measured in years from birth until death.

(2) *Skeletal Age*: The description of an individual's age based on the development or degeneration of skeletal features. Skeletal age can be influenced by genetics, environment, nutrition, health, etc.

(3) *Age indicator:* a skeletal feature that holds predictive value for aging.

(4) *Age range*: the most likely approximation of age given in a defined span of time.

(5) *Final age estimation:* the age range that would be reported to law enforcement in a forensic context. The final age estimate may be derived from a single method or using information from multiple methods.

(6) *Age-reporting strategies:* the different ways in which one or more age ranges can be reported to produce a final age estimation.

Thesis Layout

The main goals of this study are to identify the most accurate and reliable strategies for reporting age-at-death and to evaluate the appropriateness of using one's professional judgement to produce a final age estimation.

Following the introduction, Chapter Two provides an overview of the literature pertaining to the main themes of this research. The first half of Chapter Two establishes fundamentals of age-at-death estimation within forensic anthropology by providing a discussion of the history of age estimation within the field, giving an overview of the age-at-death estimation methods relevant to this study, and exploring the potential sources of error related to age estimation. The latter half of Chapter Two shifts towards the practical applications of forensic aging by presenting an overview of current practices in skeletal aging and reviewing studies aimed at comparing different aging methods. The final component of Chapter Two details the three research hypotheses driving this study.

Chapter Three outlines the materials and methods used to test the hypotheses presented in Chapter Two, provides a detailed explanation of each age-reporting strategy, and discusses the intra- and interobserver components of this research.

Chapter Four presents the result from the larger study and the intra- and interobserver studies and Chapter Five provides a discussion of all findings.

Finally, Chapter Six provides concluding thoughts, a discussion of project limitations, and future research directions.

CHAPTER TWO

LITERATURE REVIEW

Overview of Skeletal Age Estimation and Aging Methods

Age estimation of adult skeletons has been a key area of study within forensic anthropology since its inception as a discipline. From the earliest studies of age estimation to modern day approaches, researchers have noted the challenges that come with estimating age from skeletal indicators (Buckberry and Chamberlain 2002; Garvin and Passalacqua 2012; Işcan et al. 1984a; Lovejoy et al. 1985b; Martrille et al. 2007; Merritt 2013). Thomas Dwight is considered to be the father of forensic anthropology in the United States, and is one of the first individuals to meaningfully discuss age estimation in a forensic context (Ubelaker 2010). Dwight not only calls attention to the vast amount of variation in the skeleton due to age, especially during the "mature" and "senile" stages of life, but also asserts that age cannot be estimated with a high degree of accuracy due to individual skeletal variation (Dwight 1878). Dwight's book, *The Identification of the Human Skeleton: a Medico-legal Study* (1878) was particularly significant for the field of forensic anthropology because it prompted the development of aging methods aimed at understanding skeletal variation.

As individuals age, there is a decline in the amount of tissue produced in the musculoskeletal system, a thinning of the cellular matrix of the tissues, and decrease function of tissue cells leading to overall skeletal degeneration, particularly at joints (Freemont and Hoyland 2007; Lovejoy et al. 1985b; Todd 1920). Joints are areas within the skeleton where bones articulate with one another via cartilage or ligaments. There

are a variety of joints within the body that are categorized by their functional role in movement and stabilization. The amount of joint movement is dependent on the surface area and shape of the articular surfaces of the two joining bones as well as the ligaments binding them together (White et al. 2012).

Broadly, joints can be divided up into three categories: synovial, cartilaginous, and fibrous (White et al. 2012). Synovial joints are free-moving joints characterized by a joint capsule and synovial fluid which creates a friction-free environment. Synovial joints tend to have a high degree of movement as is seen in the shoulder and knee. Cartilaginous joints are connected with cartilage which restricts their movement. The sternal rib ends and pubic symphysis are both examples of different cartilaginous joints. Finally, fibrous joints restrict movement even further than cartilaginous joints as these are either bound by strong fibrous membranes or ligaments. Cranial sutures are an example of interlocking fibrous joints that exhibit little to no movement.

When developing aging methods anthropologists have focused on joints that are less affected by daily activities. Therefore, many aging methods have largely been restricted to areas of the skeleton such as the pubic symphysis, sternal rib ends, iliac auricular surface, and cranial sutures. These joints have been shown to display a sequence of morphological changes tightly associated with age (Işcan et al. 1984a; Işcan et al. 1984b; Işcan et al. 1985; Lovejoy et al. 1985b; Meindl and Lovejoy 1985; Todd 1920).

In the 20th century, several researchers identified the pubic symphysis as a reliable skeletal age indicator and subsequently published age estimation methods

based on observed pubic changes (Krogman 1939; McKern and Stewart 1957; Stewart 1957; Todd 1920; Todd 1930). These scholars have described the pubic symphyses of younger individuals as marked by "furrows and ridges." The degeneration process is recognized by the development a rim around the pubic symphysis and eventually breakdown of the symphysial face through time. Early pubic symphysis methods, such as those developed by Todd and McKern and Stewart, are not widely used in forensic casework today due to their development on skeletal samples not representative of modern forensic populations. However, their descriptions of age-related changes of the pubic symphysis has persisted and informed current aging methods. It was also during the early 20th century that anthropologists focused on cranial suture closure in developing age-estimation methods (Todd and Lyon 1924; Todd and Lyon 1925). Cranial suture aging relies on the recognition of the stages of suture obliteration. The first aging methods were important because they established the fundamentals of skeletal aging and were invaluable for describing key considerations of age estimation that persist today. These considerations include the notions that method performance may vary by age group (Baccino et al. 1999; Martrille et al. 2007; Merritt 2013), population and sex differences may influence estimations (Franklin 2010; Nawrocki 2010; SWGANTH 2013), and sample composition is important in developing methods and deciding which methods to use in age estimation (Nawrocki 2010).

The late 20th century brought an expansion of aging methodologies focused on diverse regions within the skeleton including the sternal ends of ribs and the iliac auricular surface of the pelvis (Işcan et al. 1984a; Işcan et al. 1984b; Işcan et al. 1985;

Lovejoy et al. 1985b). Like the pubic symphysis, these areas are recognized for their consistent morphological changes associated with age. For example, the sternal rib end starts out flat with billows, forms a pit with a rim, and eventually becomes irregular with age (Işcan et al. 1984a; Işcan et al. 1985). The auricular surface is often evaluated for its physical appearance as well as its texture. In youth, the auricular surface is billowed with fine granularity. Over time, the auricular surface becomes irregular and dense, eventually breaking down completely (Lovejoy et al. 1985b). The first methods developed for sternal rib ends and the iliac auricular surface are still utilized, and their descriptions have been adapted in revised methods focusing on these regions.

Pubic Symphysis Aging

Todd Pubic Symphysis

As discussed, the age-related morphological changes of the pubic symphysis have been recognized since the late 19th and early 20th centuries (Todd 1920). Wingate Todd was the first scholar to systematically describe the gross changes the pubis undergoes with age. In the development of his first pubic symphysis method, Todd observed 306 white, adult male os coxae to describe age related changes. Ten phases of the aging pubis were defined from his initial research. In 1930, Todd expanded his study of the pubic symphysis to include males and females of both African and European ancestry (Todd 1930). The skeletal material used in both studies was a medical sample with associated demographic records that Todd and his colleague Carl Hamann curated and kept at the Anatomical Laboratory in Ohio (Todd 1920). This

collection is now known as the Hamann-Todd collection and is comprised of over 3,000 skeletons that were accessioned between 1912 and 1938.

Pubic symphyses of young individuals are marked by "ridges and furrows", lack a margin around the symphyseal face, do not have a clear delineation of the superior and inferior borders of the face, and do not have a defined dorsal or ventral rampart (Todd 1920). During the "post-adolescent" phases (20-24 years old) the margin begins to form around the pubic symphysis. In later phases (25-39 years old), the ridges and furrows become less apparent and the margin around the pubic symphysis is more distinct due to greater definition of the superior and inferior margin and development of the ventral and dorsal ramparts. There is still no rim around the margin. Todd notes that it is also possible for bony outgrowths to appear during this time. Todd's eighth phase (39-44 years old) marks the completion of the outline around the symphyseal face and the complete smoothing of the ridge and furrow system. The last two phases (45+ years old) are marked by degeneration of the pubic symphysis starting with a breakdown of the rim around the pubic symphysis. The later phases are also characterized by lipping on the ventral and dorsal rampart, erosion of the symphyseal face, erratic osteophytic outgrowths (Todd 1920).

While Todd's original aging system is rarely used in forensic contexts today, his descriptions of age-related changes to the pubic symphysis and the general timeline have been adapted and utilized in the development of subsequent pubic symphysis aging methods.

Suchey-Brooks Pubic Symphysis

The pubic symphysis method developed by Suchey, Brooks, and Katz (1990; 1986; 1988) is a six-phase system that describes the age-related changes to the pubic symphysis of the pelvis. This method is a modification of the Todd (1920; 1930) pubic symphysis aging method described in the previous section. Essentially, Suchey and colleagues reduced the number of phases from ten to six and simplified the definitions developed by Todd. In addition to descriptions of each phase, casts are available that represent the "early" and "late" stages of each phase. The authors also provide separate male and female standards, as well as a mean age, the standard deviation, and a 95% range, for each phase. Brooks and Suchey's (1990) modified method was developed using an autopsy sample collected from the Los Angeles Office of the Chief Medical Examiner, making it the first method to be developed on a modern forensic population. The sample includes a large number of modern males (n=739) and females (n=273) whose ages-at-death range from 14-99 years old, with majority of the individuals comprising the early decades of life. The Suchey-Brooks sample was also inclusive of individuals with diverse ancestral backgrounds.

Hartnett Pubic Symphysis

The pubic symphysis aging method developed by Hartnett (2010a) is a revision of the Suchey-Brooks(1990) method and aims to increase the accuracy and precision of pubic symphysis aging. As with the Suchey-Brooks method, the Hartnett pubic symphysis method was mainly derived from an autopsy sample. Male (n=419) and female (n=211) pubic bones were collected during autopsy at the Maricopa County

Forensic Science Center or from the Barrow Neurological Institute in Phoenix, Arizona in the early 2000s. The sample consists of individuals from different ancestral backgrounds and whose ages-at-death range from 18-99 years old. Unlike the Suchey-Brooks sample, the Hartnett sample has a higher proportion of older individuals.

The major difference. between the Suchey-Brooks pubic symphysis method and the Hartnett pubic symphysis method is the increase from six to seven phases. Hartnett's sample includes a larger proportion of older individuals than Suchey-Brooks', so she added the additional phase to better capture morphological changes associated with later decades of life. Hartnett argues that the addition of a seventh phase helps avoid a non-descriptive category of 50+ years, which is the range for phase six of the Suchey-Brooks method. Hartnett's pubic symphysis method also takes into consideration the texture and weight of the bone for phase assignment. As with the Suchey-Brooks method, Hartnett provides sex-specific phases, mean age, standard deviation, and an age range. Unlike Suchey-Brooks, the age ranges provided by the Hartnett method were determined based on where 100% of the data fell rather than a 96% range. Additionally, Hartnett does not provide supplementary visual materials such as casts.

Auricular Surface Aging

Iliac auricular surface aging is conducted by evaluating different portions of the iliac joint surface as well as the area surrounding it. The iliac joint surface itself is divided into two major aspects; the superior demiface and the inferior demiface, which come together at the apex (Lovejoy et al. 1985b). In addition to the joint surface, the

area just superior and posterior to the iliac auricular surface (retroauricular area) is also evaluated. The auricular surface is aged based on morphological attributes (billowing, striae, porosity, osteophytic growths) as well as its texture (granularity and density) (Lovejoy et al. 1985b).

In general, the auricular surface can be broken divided into four broad categories of age related changes, "young adult phase," "mid adult phase," "early senescent phase," and "breakdown" (Lovejoy et al. 1985b). The young adult phase is characterized by distinct transverse organization, billowing, course granularity and lack of retroauricular bone growth. In the mid adult phase there is a loss of billowing, the surface texture is more fine-grained, and there is slight bony buildup in the retroauricular area. Early senescence is marked by a distinct change in surface granularity, increased porosity and density, and morphological changes at the apex. Finally, the breakdown stage is characterized by destruction of the subchondral bone accompanied by extensive porosity, irregularity, and bony growths in the retroauricular area.

Lovejoy Auricular Surface

The description of auricular surface aging provided above is drawn from the method developed by Lovejoy, Meindl, Pryzbeck, and Mensforth (1985b). The Lovejoy method was developed using individuals from three different sample populations; the Libben archaeological population housed at Kent State University (n=250), the Hamann-Todd collection at the Cleveland Museum of Natural History (n=500), and identified forensic cases from the Cuyahoga County Coroner's office (n=14). For their method, Lovejoy et al. define eight phases based on chronological changes of the auricular

surface and retroauricular area. The eight phases provide 5-10 year age intervals, with

the exception of phase 8, which is simply 60+ years. In addition to the descriptions of

each phase, the authors provide black and white photos for reference. The authors do

not provide any statistical information such as standard deviations or confidence

intervals. Scholars have found that the method consistently underestimates the ages of

older individuals and overestimates the ages of younger individuals (Bedford ME 1993;

Murray and Murray 1991; Saunders et al. 1992).

Buckberry-Chamberlain Auricular Surface

Buckberry and Chamberlain (2002) evaluated and revised the Lovejoy (Lovejoy

et al. 1985b) method due to its consistent use within forensic anthropology despite its

optimistically small age ranges. The authors translated the categories described by

Lovejoy et al. (1985b) into numerical scores based on degrees of expression. For

example, porosity is scored on a 1-3 scale with score 1 indicating no porosity, 2

indicating porosity on only one demiface, and score 3 indicating porosity on both

demifaces. Once each individual component score is added up into a composite score it

is translated into an auricular surface stage. There are seven age stages, and each

have corresponding statistical information such as mean age, standard deviation,

median age, and an age range. Buckberry and Chamberlain tested this new auricular

surface scoring system for observer reliability and correlation to age. Due to the high

observer comparability and high correlations to age, the component system was then

tested on a skeletal sample (n=180) from Christ Church, Spitalfields London. The

Spitalfields collection includes over 900 individuals, with associated perish records,

dating from 1759-1859 AD. The authors also applied the Suchey-Brooks pubic

symphysis method to the Spitalfields sample and found that their new auricular surface

system had a higher correlation with age than Suchey-Brooks.

Sternal Rib End Aging

Age estimation using ribs typically involves observing the morphological changes

the fourth rib end undergoes through time. These changes broadly involve an increase

in pit depth, shape changes to the rib pit and rim, and bone quality changes (Işcan et al.

1984a; Işcan et al. 1984b; Işcan et al. 1985).

Sternal rib ends of young adults typically have flat, billowing surface, a regular

rim, and a dense, smooth bone texture (Işcan et al. 1984a; Işcan et al. 1984b; Işcan et

al. 1985). Over time, the pit deepens and the cross section of the pit takes on a "V"

shape which eventually widens into a "U" shape. As the pit deepens, the rim

surrounding the sternal rib end also changes. At first the rim is rounded and smooth but

will become scalloped in appearance. Eventually, the rim becomes sharp and irregular

with the cartilage around the sternal rib end ossifying in some instances. Finally, the

overall quality of the rib as well as the quality of bone within the pit deteriorates through

time. In youth the bone is smooth, dense, and strong, but with age the bone becomes

thin and brittle and exhibits porosity within the pit and on the outer cortical layer (Işcan

et al. 1984a; Işcan et al. 1984b; Işcan et al. 1985).

Işcan Rib

The sternal rib method developed by Işcan, Loth, and Wright (1984a; 1984b; 1985) is a macroscopic evaluation of the sternal end of the fourth rib. The morphological qualities being evaluated include pit formation, pit depth and shape, wall configurations, and bone texture and quality of the bone as a whole. While Işcan's (1984b) first method for sternal end aging was a component-based system, his subsequent methods were phase-based and are the ones included within this study (Işcan et al. 1984a; Işcan et al. 1985). Işcan and colleagues developed their sternal rib end methods using an autopsy sample from the Broward County Medical Examiner's Office in Fort Lauderdale, Florida. Separate standards were developed for males and females. In both methods, only white individuals with known demographic information were selected. The male sample included ribs from 118 individuals ranging in age from 17-85 years and the female sample included ribs from 86 individuals ranging in age from 14-90 years. The average age of individuals included was 41 years for the males and 48 years for the females, with the majority of the individuals between 20-40 years old.

The Işcan sternal rib end method involves the evaluation of the right, fourth rib and includes nine phases ranging from 0-8, with 0 containing more youthful qualities and 8 more degenerative qualities. The general process of age-related changes to the rib can be referenced in the previous section. Again, there are sex-specific descriptions for each phase. The statistical information associated with each phase includes a mean, standard deviation, standard error, 95% confidence interval, and an age range. Casts

and photographs illustrating variations within each phase are also available to use along with the written descriptions.

Hartnett Rib

The Hartnett (2010b) rib method is a modification of the Işcan (1984a; 1984b) method and is aimed at increasing the accuracy and precision of rib aging. The sample Hartnett used in the development of the sternal rib end is the same sample she used to develop the modified pubic symphysis method. Male (n=419) and female (n=211) fourth ribs were collected during autopsy at the Maricopa County Forensic Science Center or from the Barrow Neurological Institute in Phoenix, Arizona in the early 2000s. To reiterate, the sample consists of individuals from different ancestral backgrounds and whose ages-at-death range from 18-99 years old.

One of the major differences between the two rib methods is that Hartnett (2010b) reduces the number of phases from eight to seven. Additionally, more emphasis is placed on the bone weight and quality for phase assignment in the later decades of life. Other minor changes include clarifying language of phase descriptions and adjusting the age ranges and means per phase to reflect the phase composition of Hartnett's sample. Hartnett provides descriptive statistics for both males and females that include the mean, standard deviation, and age range based on 100% of individuals. Unlike Işcan, Hartnett does not provide supplementary visual aids for phase assignment.

Error in Aging

In addition to establishing methods, aging studies have helped anthropologists become more aware of the variation that different skeletal indicators can display as a result of the aging process. Moreover, aging studies have contributed to a greater consciousness of the potential sources of error when estimating age-at-death. Understanding the error involved in age estimation is necessary when developing new aging methods and for validating current methods. Nawrocki (2010) discusses many sources of error that anthropologists face with age estimation, including non-age related skeletal variation, the error caused by transformative processes of skeletal aging, and the "trajectory effect." Non-age related skeletal variation is the observable differences in skeletal indicators that are not accounted for by age. Variation not associated with age can be a result of the individual's sex, ancestral background, activity level, disease history, etc. In some instances, non-age related variation can be controlled for in the development of aging methods which can help reduce the overall error of the methodology. The transformative process of skeletal aging refers to the process of transcribing the appearance of skeletal indicators into a chronological age range. The trajectory effect of aging is the concept that the variation in skeletal indicators will increase with chronological age, and subsequently, associated error intervals will also increase (Figure 1). This increase in error with age is because there is not a one-to-one correlation between chronological age and skeletal age. As per the trajectory effect, as one ages, the less correlated skeletal age and chronological age become. Familiarity with this phenomenon is pertinent for reducing error associated with age-at-death

estimation. While small final age rages might be appropriate for younger individuals, the trajectory effect illustrates the inappropriateness of providing narrow age estimate for individuals of more advanced age.

As previously mentioned, skeletal age estimation relies upon recording perceived skeletal age, based on visual observation, and translating it into a chronological age range. Since skeletal indicators and chronological age are not perfectly correlated with one another due to individual life histories, there is error inherently involved in this process. It is difficult to control all of the variables that affect skeletal morphology, including sex and population differences, activity level, disease, etc. Even if these variables could be controlled, their actual effect on the aging process is difficult to assess (Nawrocki 2010). Further complicating age estimation is that skeletal regions do not always age consistently with respect to each other (Franklin 2010; Nawrocki 2010). Because different areas of the skeleton are not always analogous in their degeneration process, Nawrocki (2010) states that no single method can account for more than 50% of variability associated with aging. Because individual aging methods are limited in their ability to capture the age-related variation observed in isolated skeletal regions, anthropologists may try to combine results from multiple methods and/or used their expert judgement to arrive at a final estimation. This is problematic because age estimations arrived in this way have the potential to not meet the evidentiary admissibility guidelines set forth by the *Daubert* ruling.
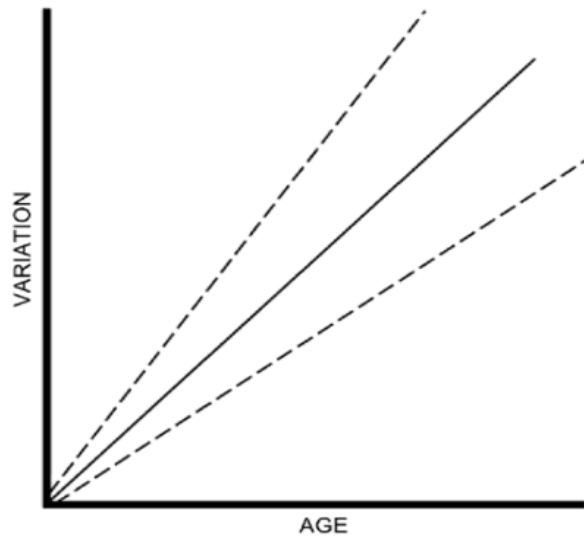
Figure 1: Trajectory effect in age estimation (Nawrocki 2010).

When anthropologists attempt to combine information from multiple aging indicators and methods, they introduce more error into the age estimation process. While multifactorial approaches that utilize principal component analysis have been shown to moderately increase accuracy of age estimation (Lovejoy et al. 1985a; Martrille et al. 2007; Nawrocki 2010; Saunders et al. 1992), they do not perform significantly better than single indicators or simply averaging the results of each aging indicator observed (Martrille et al. 2007; Nawrocki 2010; Saunders et al. 1992). Additionally, multifactorial approaches, which involve attributing different weights to calculated point estimates via principal component analysis (PCA), are arduous and statistically difficult to employ (Martrille et al. 2007; Saunders et al. 1992). Nawrocki (2010) suggests that a better approach might be to rely on the skeletal indicator with the lowest error rate and use it exclusively. This suggestion should be evaluated and tested for its accuracy and reliability compared to other age-reporting strategies within anthropology. If using the best skeletal indicator with lowest error rate is the most accurate and reliable way to arrive at a final age estimation, then this should be the standard for reporting age within forensic anthropology. However, if the accuracy of methods is contingent on large age intervals, this may not be the most pragmatic approach in a forensic setting.

Current Practices in Skeletal Aging

Due to the large body of literature concerning age estimation, Garvin and Passalacqua (2012) administered a survey to 145 members of the Physical Anthropology section of the American Association of Forensic Sciences to gain an

understanding of common practices in age estimation. The major goals of their survey were to assess how anthropologists make decisions regarding which skeletal region to evaluate, which method(s) to use, how to report statistical information, and how information from different methods were being translated into a final age estimation.

The results of Garvin and Passalacqua's survey revealed that the pubic symphysis is the most preferred region to evaluate for age estimation, followed by the sternal rib ends and the auricular surface (Garvin and Passalacqua 2012). In contrast, respondents least preferred evaluating cranial suture and dental wear for assessing age. Participants were also asked to identify the specific methods that they typically used for estimating age. Unsurprisingly, the favored methods reflected the preferred skeletal regions: Suchey-Brooks (1990) (pubic symphysis), İşcan (1984b; 1985) (sternal rib ends), and Lovejoy et al. (1985b) (auricular surface). Garvin and Passalacqua also note that anthropologists prefer to use the studies developed in 1980s-1990s because they are often included within edited volumes, are the most commonly used methods, and do not require learning new/different methodologies (Garvin and Passalacqua 2012). These results are further supported by Parsons (2017) who examined the accuracy of the biological profile in casework across three different Medical Examiner's offices.

When reporting the results from a single method, the preferred strategy is using the full age range provided by the original study (Garvin and Passalacqua 2012). There is variation in how age ranges are developed among aging studies. For example, Katz and Suchey (1986) utilize a 95% range; Hartnett (2010a; 2010b) developed ranges on

the basis of where 100% of where her data fell; and Lovejoy et al. (1985b) arbitrarily

chose to report 5-year age ranges.

An experience-based aging approach was the next most utilized age-reporting

strategy, especially when a skeletal indicator does not fit nicely into one of the described

phases because it displays characteristics from multiple phases (Garvin and

Passalacqua 2012). In these instances, all but one respondent reported considering

descriptive information from multiple phases within a single method. The survey showed

that there was no consensus regarding how to combine information from multiple

phases. While some participants responded that they reported the overlap of the

phases, others indicated that they reported the entire range of multiple phases or would

use their expertise to produce a narrower age range. Finally, when asked how they

combine information from different skeletal regions or methods, respondents gave

variable responses, but many indicated that experience was a deciding factor in their

final determination of a final age estimation (Garvin and Passalacqua 2012). Survey

respondents expressed concern towards the subjectivity and statistical invalidity of the

experience-based age estimations. Because experience is relied upon to narrow ranges

and/or provide final age estimations, it is important for studies to evaluate estimations

derived from experience for accuracy and reliability.

Garvin and Passalacqua's (2012) survey is important because it sheds light on

current practices and problems affecting age-at-death estimations. The survey

highlights the most commonly evaluated skeletal regions, the most relied upon

methods, and the overall lack of standardization, particularly when combining

information from multiple methods. Understanding which age-reporting strategies produce the most accurate and reliable age estimations is fundamental for informing decisions regarding method selection and improving aging results.

While multifactorial approaches to age estimation, such as transition analysis (Milner and Boldsen 2012), have helped alleviate some of the statistical challenges of combining multiple indicators for age estimation, only a few respondents of Garvin and Passalacqua's survey reported using transition analysis or other Bayesian approaches. Additionally, some of the aging literature demonstrate that multifactorial approaches do not fare significantly better than single indicators (Martrille et al. 2007; Saunders et al. 1992). Milner and colleagues (2016) are currently working on a more comprehensive aging program which aims to allow anthropologists to combine different aging indicators in a statistically valid manner. Until this new program is released and implemented broadly, its contribution to skeletal age estimations remains unclear.

To understand how current methodologies can affect age estimations, several anthropologists have conducted comparative studies. The goals of these studies are to assess which skeletal regions and methods perform the best with respect to accuracy and reliability (inaccuracy and reliability). Additionally, the studies comparing multiple aging methods explore strategies that combine information (Baccino et al. 1999; Martrille et al. 2007; Merritt 2013; Saunders et al. 1992).

Studies Comparing Aging Methods

Studies aimed at comparing different aging methods typically evaluate method performance by their accuracy and reliability (Garvin and Passalacqua 2012; Martrille et

al. 2007; Merritt 2013; Miranker 2016; Saunders et al. 1992). Accuracy is the

assignment of an individual to a category that includes their age-at-death. (Garvin and

Passalacqua 2012; Merritt 2013; Miranker 2016). Overall method accuracy is calculated

by assessing the number of individuals whose ages are included within their assigned

age range. For example, if 10 out of 10 individuals' ages-at-death fall somewhere within

the age range of the phase ascribed by the researcher, then the accuracy of the method

is 100%. Accuracy does not take into account whether the individual could fit into

multiple phases described by the method. Therefore, methods that have very large age

ranges and/or phases with overlapping ranges tend to be more "accurate."

Reliability assesses how far an estimate is from the actual age and whether a

method has the tendency to over- or underestimate certain cohorts. Reliability is

determined by calculating inaccuracy and bias (Merritt 2013). Inaccuracy is defined as

the *absolute* distance of the actual age from the mean of the range the individual was

ascribed (Merritt 2013; Nawrocki 2010). Inaccuracy does not take into consideration

over- or underestimation, but rather total distance from the mean. Conversely, bias is

defined as the tendency of a method to under/overestimate an individual's actual age

(Merritt 2013). Methods are often considered reliable if they have low inaccuracies and

bias scores approaching zero.

In order to better understand the performance of aging methods on samples

outside of those used to develop the method, Saunders et al. (1992) tested four

traditional aging methods and one multifactorial aging approach on individuals within an

archaeological sample. Saunders and colleagues sample size ranged between 27-49

25

individuals depending on preservation. The individuals were excavated from the St. Thomas Anglican church in Belleville, Ontario, which was in use from 1821-1874. The researchers only included individuals with known ages. The selected methods for their study include the Suchey, Brooks, and Katz (1986; 1988) pubic symphysis, Lovejoy (1985b) auricular surface, Meindl and Lovejoy (1985) ectocranial suture, and İşcan, Loth, and Wright (1984a; Işcan et al. 1984b; 1985) sternal rib end. In addition to these four aging methods, the authors employed a multifactorial aging approach which they modeled after a previous study conducted by Lovejoy and colleagues (1985a). For this multifactorial approach, all aging indicators observed are applied independently and used to generate an intercorrelation matrix which is then subject to a principle components analysis (PCA) (Saunders et al. 1992). The final age estimate is calculated using the weighted averages of each skeletal indicator. Saunders and colleagues used the reported means for each method and calculated means for the Lovejoy method (1985b) which only provides 5-10 year ranges for each phase. A simple average was also calculated and compared to the multifactorial approach to see if one approach has greater aging potential over the other.

The value of each individual age indicator was assessed based on the difference between predicted and actual age and bias of each method, which is defined above as reliability. The results of the study led the authors to assert that no skeletal indicator of age is likely to encompass all of the variation of chronological age, and reliance on a single method for age estimation is cautioned (Saunders et al. 1992). The findings of this study also indicate that the multifactorial approach did not predict age much better

than a simple mathematical average of the estimates derived from each method. Regardless of the poor performance of the multifactorial method, the authors recommend that anthropologists continue to utilize all available information and rely on their "professional judgment" to develop a final age estimation (Saunders et al. 1992). Again, while age estimations derived using one's professional experience are problematic within the contexts of *Daubert*, they should not be discounted if they are both accurate and reliable.

Like Saunders (1992), Baccino and colleagues (1999) evaluated individual aging methods for their accuracy and reliability. The skeletal elements for their study were collected at autopsy from 19 European individuals (15 males and 4 females) ranging in age from 19-54 years. The methods included the study were Lamendin (1992) single rooted tooth, İşcan (1984a; 1985) sternal rib ends, Suchey-Brooks (1986) pubic symphysis, and the Kerley (1978) microscopic cortical bone thickness. In addition to evaluating individual aging methods, Baccino et al. (1999) were interested in evaluating age-reporting strategies that consider the results of multiple methods and age indicators, which are referred to as "comprehensive methods" by the authors.

Baccino and colleagues (Baccino et al. 2014; Baccino et al. 1999) were the first to implement a strategy for systematically selecting an aging method based on preliminarily age assessment of the skeleton. This strategy for method selection is known as the two-step procedure (TSP). The TSP carries two assumptions: (1) no single aging method is appropriate for the entire lifespan and (2) methods developed using age cohorts similar to the unknown skeleton will produce more accurate age

estimations (Baccino et al. 2014; Baccino et al. 1999). For the Baccino et al. (1999) study, the TSP helped the researches choose between the Suchey-Brooks and Lamendin method since Suchey-Brooks is more accurate for young individuals and Lamendin is more accurate for older individuals. If the unknown skeleton fell within the first three phases of the Suchey-Brooks method then the Suchey-Brooks age range was reported as the final age estimate, but if the individual fell within phases four-six of the Suchey-Brooks method, Lamendin method was reported as the final age estimate (Baccino et al. 1999).

In addition to the TSP, the researchers also produced age estimations using a "global approach" (Baccino et al. 1999). Essentially, the researchers were able to include or exclude the results of individual methods and rely on their professional experience to produce the age estimation they deemed most appropriate. Two observers conducted each method mentioned above and interobserver error was calculated for observer comparability, bias, and accuracy.

The results of the Baccino et al. (1999) study revealed that the TSP had the best overall observer comparability, the Lamendin method had the highest correlation of bias scores, and the global method had the smallest mean inaccuracy difference.

The Baccino et al. (1999) study showed that the standard errors were lower for all comprehensive methods than for single methods. Additionally, the study revealed that strategies of age estimation that consider the results of multiple methods produce better estimations than those relying on single methods, which supports the findings of Saunders and colleagues (1992). It is important to note that the sample size for the

Baccino et al. study was small (n=19) meaning the results could be an artifact of sample bias. Further comparisons of age estimates derived from individual ranges and those derived from the results of multiple methods are necessary to determine if one is actually superior to the other.

Martrille and colleagues (2007) also tested skeletal aging methods for their accuracy and reliability and employed a multifactorial approach to combine aging indicators. The goal of their study was to determine if single aging methods are more reliable than a combination of methods for estimating age. The four methods included within the study were Suchey, Brooks, and Katz (1986; 1988) (Suchey-Brooks) pubic symphysis, Loveyjoy (1985b) auricular surface, the Lamendin (1992) single rooted tooth, and the İşcan (1984a; 1984b; 1985) fourth rib. The study assesses the inaccuracy and bias for each age indicator. Similarly to Saunders et al. (1992), PCA was used in order to combine the four methods. The sample for the Martrille et al. (2007) study consisted of 218 black and white individuals (115 males and 103 females) ranging in age from 25-90 years old from the Terry collection. The Terry collection is comprised of over 1,700 individuals who were born between 1828 and 1943. The researchers analyzed results for the entire sample and then by ancestry, sex, and age cohort (25-40; 41-60; 60+).

When broken down by age, Suchey-Brooks was more accurate for young adults and İşcan for older adults. For the combined sample, Suchey-Brooks and İşcan methods were more accurate than the Lamendin and auricular surface methods. Additionally, the results revealed that PCA provided the best overall age estimation with

regards to mean inaccuracy. However, like Saunders et al. (1992), Martrille et al. (2007) found that the multifactorial method did not perform significantly better than the individual methods. The results were similar when the sample was subdivided ancestry and sex. Their research also support the two-step procedure described by Baccino and colleagues (1999). The authors suggest that the preliminary assessment of skeletal indicators to inform which method to use for aging will yield the most accurate results. Again, the suggestion of preliminary assessment of a skeletal indicator gives validity to the notion that aging methods constructed using age cohorts that are similar to that of the unidentified remains will reduce the error of the final the age estimation (Baccino et al. 2014; Martrille et al. 2007; Merritt 2013).

Finally, Merritt (2013) examined different age estimation methods, but focused on comparing five original aging methods to six new or revised methods. The original methods examined were Kunos et al. (1999) first rib, İşcan, Loth and Wright (1985) sternal fourth rib, Lovejoy (1985b) auricular surface, Todd (1920) pubic symphysis, and Suchey-Brook (1990) pubic symphysis. The new and revised methods were Digangi et al.(2009) first rib, Hartnett (2010a; 2010b) pubic symphysis and fourth rib methods, Passalacqua (2009) sacrum, Buckberry and Chamberlain (2002) auricular surface, and Rougé-Maillart et al. (2007) acetabulum and auricular surface. The sample for Merritt's study included 20 European Males between the ages of 29-85 years old from the University of Toronto J.C.B Grant Collection. The J.C.B Grant Collection consists of 202 adult skeletons that were received by the anatomy department from 1920s-1950s. All

individuals have associated records with their names, sex, age-at-death, and cause of death.

The accuracy and reliability of each method were compared to one another. The results of the study revealed that new and revised methods tend to produce more accurate age estimations with lower biases, overall. Like Saunders and colleagues (1992), Merritt also assessed methods by age interval and found that original methods performed better for younger individuals and newer/revised methods were more accurate for older individuals. This is unsurprising given that original methods were developed on samples comprising younger individuals and new methods are often developed on samples comprising older individuals.

Along the lines of Baccino's (2014; 1999) two-step procedure, and addressing Martrille's (2007) suggestion to make a preliminary assessment of age before choosing aging methods, Merritt (2013) suggests considering the relative age of the skeleton and using original methods if the skeleton is likely younger (<40 years old) and revised methods if it is likely older (>40 years old). The sample composition for Merritt's study was also biased, as the number of individuals was limited to 20 European males, most over the age of 60.

Age estimation research has focused on testing the accuracy and reliability of current aging methodologies and exploring ways to combine the results from different methods and skeletal indicators to produce a final age estimation. While a variety of age-reporting strategies have been suggested, they have not been systematically tested for their accuracy, reliability, and practicality in a forensic setting.

Hypotheses

As evident by the literature, there are differing viewpoints regarding which age-reporting strategies are most appropriate to arrive at a final age estimation. When relying on a single aging method to produce a final age estimation, one could simply report the age range provided. However, if multiple methods or indicators are relied upon for estimating age it is often difficult to decide which of the method's ranges would be best to report. Nawrocki (2010) suggests that it is statistically best to rely on the indicator with the lowest standard error when choosing between age ranges. This may be helpful for determining which age range to report when there are multiple methods and indictors, or multiple methods and one age-indicator being evaluated. Alternatively, several researchers (Baccino et al. 2014; Baccino et al. 1999; Martrille et al. 2007; Merritt 2013). suggest using a two-step strategy where the most appropriate method is determined based on a preliminary assessment of skeletal morphology. Finally, some anthropologists favor comprehensive age-reporting strategies such as the overlap of multiple ranges or using professional judgement to combine information from multiple methods/indicators (Baccino et al. 2014; Garvin and Passalacqua 2012; Lovejoy et al. 1985a; Saunders et al. 1992).

This study aims to assess age-reporting strategies for their ability to produce accurate and reliable final age estimations. Age-reporting strategy accuracy and reliability will be assessed for the total sample, by age-cohort, and by sex. Therefore, there are three hypotheses for this study:

32

(1) Final age estimations derived using one's experience will be the most

accurate and reliable overall.

(2) The most accurate and reliable age-reporting strategy will differ when the

sample is divided by age.

2a. The two-step strategies will result in the most accurate and reliable

final age-estimations for the young and old age cohort.

(3) The most accurate and reliable age-reporting strategy will differ when the

sample is divided by sex.

3a. Final age-estimations derived from sex-specific methods will be more

accurate and reliable than those derived from non sex-specific methods

when the sample is evaluated by sex.

CHAPTER THREE
MATERIALS AND METHODS

Sample

The sample for this study includes 58 adult individuals (31 males and 27 females) from the William M. Bass Donated Skeletal Collection, at the University of Tennessee, Knoxville. The Bass Skeletal Collection contains nearly 1,500 skeletons of individuals born after 1900 and with known demographic information. Table 1 provides a summary of the sample demographics for this study. The ages-at-death range from 21-79 years old, with roughly 10 individuals representing each decade of life. The Bass Collection has an underrepresentation of individuals in the younger age categories, and many of the young adults within the collection exhibit trauma making them unobservable for this study. The second decade of life was delineated as the lower cutoff for this study because it is during this time that skeletal maturation is completed and reliance on growth and development markers is no longer feasible. The upper age limit was 79-years-old due to the challenge of procuring adequate representation of the eighth, ninth, and tenth decades of life. Only individuals of European ancestry were included within this study to avoid low sample sizes. Excluding the 20-30 age group, sex was equally distributed for each decade. The sample was selected from a list of donors that fit the age criteria listed above. Other than the 20-30 age group, five males and five females were randomly selected for each age group. All the individuals that fit the criteria in the 20-30 age group were included in this study.

sTable 1: Sample demographics.

Age Group	Females	Males	Total
20-30	2	6	8
31-39	5	5	10
40-49	5	5	10
50-59	5	5	10
60-69	5	5	10
70-79	5	5	10
Total	27	31	58

Age Estimation Methods

Six adult age estimation methods (three original methods and three revised methods) were independently conducted following the original publication descriptions. The original methods included in this study are Suchey-Brooks (1990), pubic symphysis (SBPS); Lovejoy (1985b) auricular surface (LJAS); and İşcan (1984a; 1985) fourth rib (ISR). The revised methods are Hartnett (2010a) pubic symphysis (HNPS); Buckberry-Chamberlain (2002) auricular surface (BCAS) and Hartnett (2010b) (2010b) fourth rib (HNR). A summary of each method can be found in Table 2.

All of the methods in this study are phase-based except for Buckberry-Chamberlain, which is a component-based method. Both phase and component-based aging methods involve the evaluation of several different bony morphological traits on a skeletal indicator. The major difference between the two systems is that phase based methods group several traits together in broad phases that occur throughout the aging process, while component methods allow the observer to score traits independently from one another (Shirley and Ramirez Montes 2015). Phase-based methods operate under the assumption that age-related changes alter the overall appearance of the indicator, while component-based methods assume that different traits have independent correlations to age.

Method Implementation

Because the results of one method has the potential to bias the results of another, only one method at a time was applied to all individuals in the sample before moving on to the next method.

Table 2: Age estimation methods utilized in this study.

Skeletal Indicator	Method	Method Abbreviation	Original/Revised	Phase/ component
Pubic symphysis	Suchey-Brooks (1990)	SBPS	Original	Phase
Pubic symphysis	Hartnett (2010a)	HNPS	Revised	Phase
Auricular surface	Lovejoy (1985b)	LJAS	Original	Phase
Auricular surface	Buckberry-Chamberlain (2002)	BCAS	Revised	Component
Sternal rib end	İşcan (1984a; 1985)	ISR	Original	Phase
Sternal rib end	Hartnett (2010b)	HNR	Revised	Phase

Donor ages were hidden from the researcher until the completion of data collection.

Each skeletal element was assigned to a phase or score, according to the method being

utilized. The left os coxa and fourth rib from each individual was examined, unless the

element from the left side was missing, pathological, or fragmented, in which case, the

right element was examined. If the fourth rib was not observable for either side, but the

third or fifth rib was, then this rib was used for age estimation. A study by Yoder and

colleagues (2001) found that there is not a significant difference between the scores of

right and left ribs 3-9. Therefore, the substitution of the right rib four with left rib four and

the potential substitution of rib four with three or five was justified. Casts were

referenced when conducting the methods for which they were available. Data collection

sheets on which all notes and age ranges were recorded were developed for each

method. At the end of data collection, all the sheets were reassociated by individual and

all the data were transcribed into an Excel sheet.

Age Reporting Strategies

Sixteen final age estimations were derived through age-reporting strategies

(method range, lowest error, two-step, overlap, and experience) (Figure 2). Each final

age estimation was informed by the ranges of the six age-at-death methods conducted

for this study. A brief overview of each age-reporting strategy is provided here, but each

strategy is elaborated upon in the subsequent sections.

First, an age estimate based on the phase/score derived from each method was

recorded. Thus, six ranges were provided, one for each method.
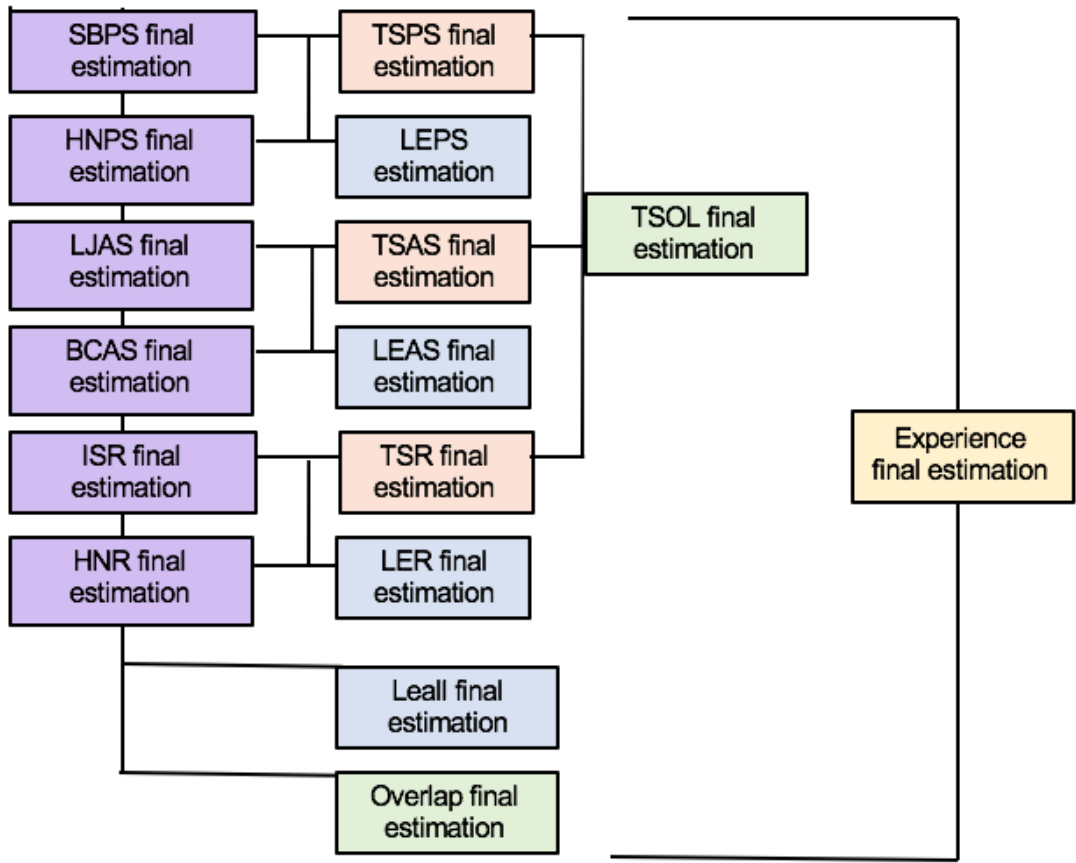
Figure 2: Final age estimations derived from age-reporting strategies.

Next, the range with the lowest standard error when considering all methods together was recorded. Then the range with the lowest error by skeletal region was identified for each individual. Four ranges were produced using the lowest error strategy. Next, estimations were derived using a two-step strategy where a preliminary assessment of each skeletal indicator was conducted to determine which method results would be recorded. The two-step was conducted between the two methods for each skeletal region. If the individual was assigned to a lower phase using an original method, than that method's results were recorded, but if the individual was assigned to a higher phase using an original method, than the revised method's score was recorded. Using the two-step strategy three ranges were provided for each individual.

Finally, due to the propensity of forensic anthropologists to include data from multiple methods in their final age estimations, two comprehensive strategies for reporting age were executed: overlap approach and an experience-based approach. While no one approach for combining data from multiple methods was preferred among respondents in Garvin and Passalacqua's (2012) study, overlap of results and experience-based estimations were among the top responses given and why these approaches, in particular, were included in this study.

For the overlap approach, the investigator chose a range based on the overlap of the six ranges derived from the aging methods included in this study. Additionally, an estimation was produced using the overlap of the three methods derived from the two-step strategy. The experience-based approach was a subjectively derived range based on the data from all methods and the observer's professional experience. For the

experience-based approach, the researcher had the ability to delineate any range they felt appropriate for the individual given the results of the aging methods and their understanding of human variation. Table 3 displays a list of the age-estimation strategies and their abbreviations. The individual method abbreviations are provided in Table 2 and are not included here. As mentioned, further descriptions of each age-reporting strategy are provided in the following sections.

Lowest Standard Error

After reporting the full range for each method, a single method for each individual was chosen based on the method/phase with the lowest standard error, following Nawrocki's (2010) advice. Of the six aging methods included in this study, only Işcan's (1984a; 1984b; 1985) rib methods provide the standard error value for each phase. Therefore, standard error was calculated for the remaining methods using the formula:

$$SE\underline{x} = \frac{S}{\sqrt{n}}$$

where SE= standard error, S=standard deviation, and n=sample size. The standard error calculations for the six methods are found in Table 4. Lovejoy (1985b) did not provide enough information to calculate standard error and was not an option for this portion of the study. The phase with the lowest standard error when taking all methods into consideration was identified for each individual and the resulting range was recorded.

Table 3: Age-estimation strategies and abbreviations.

Strategy	Abbreviation
Lowest Error-All methods	Leall
Lowest Error- Ribs	LER
Lowest Error- Pubic Symphysis	LEPS
Lowest Error- Auricular Surface	LEAS
Two-Step- Ribs	TSR
Two-Step- Pubic Symphysis	TSPS
Two-Step- Auricular Surface	TSAS
Overlap	Overlap
Overlap: Two-Step	TSOL
Experience	Experience

Table 4: standard error of each method phase.

Phase	SBPS-Males	SBPS -Females	HNPS- Males	HNPS- Females	LJAS	BCAS
1	0.2	0.38	0.52	0.59	-	0.88
2	0.42	0.71	0.5	1.06	-	2.74
3	0.91	1.22	1.11	1.02	-	2.79
4	0.72	1.75	1.01	1.03	-	2.56
5	0.9	2.2	0.89	0.7	-	1.62
6	0.87	1.74	1.38	1.24	-	1.86
7	-	-	0.95	0.99	-	3.67
8	-	-	-	-	-	-
Phase	ISR-Males	ISR-Females	HNR-Males	HNR-Females		
1	0.25	-	0.32	0.63		
2	0.59	0.68	0.38	0.44		
3	0.85	0.74	0.71	0.68		
4	1.11	1.46	0.43	0.67		
5	1.93	2.96	0.4	0.59		
6	2.71	3.52	0.45	0.8		
7	2.31	2.81	0.76	0.82		
8	2.97	2.66	-	-		

In addition, individual skeletal regions were evaluated separately, and an age estimate was recorded based on the method with the lowest standard error for each region. When determining the method with the lowest error for auricular surface, Buckberry-Chamberlain was the only option since it was impossible to calculate standard error for the Lovejoy method.

Two-Step Strategy

Several researchers have identified a two-step stagey as a reasonable way to estimate age because it takes into consideration the development of original methods on younger sample populations and revised methods on older sample populations (Baccino et al. 1999; Martrille et al. 2007; Merritt 2013). The two-step strategy described by Martrille (2007) and Merritt (2013) involves making a preliminary assessment of the skeleton before choosing which method results to report as a final age range. If the skeletal indicators suggest the individual is "young" analysis should continue using original methods. Conversely, if the morphology indicates that it is an older individual, further analysis should include revised methods (Martrille et al. 2007; Merritt 2013). Following Merritt's (2013) study, forty years of age was used to differentiate young from old. For example, if the initial analysis with the Suchey-Brooks pubic symphysis method placed an individual within phases I-III, then the result from Suchey-Brooks was recorded; however, if the age estimate was within Suchey-Brooks phases IV-VI, then the Hartnett estimation was recorded. For both the auricular surface and ribs, phase four of the original studies was specified as the cutoff between young and old as that cutoff roughly corresponds to forty years of age for those methods.

Overlap Approach

To incorporate all the data provided from different methods, practitioners often report an age estimate where methods overlap (Garvin and Passalacqua 2012). For this study, two overlap strategies were employed: (1) overlap of all six age ranges (2) overlap of the three age ranges derived from the two-step strategy. For each strategy, the ranges of the six aging methods were "mapped" on a piece of paper, as depicted in Figure 3. The region where all, or most, of the ranges overlapped was visually identified and recorded as the overlap age estimation. For continuity, the start and end of the overlap age estimations include existing points from the ranges derived from the aging methods.

Experience-based approach

The experienced-based approach for this study is the same as the global approach that is outlined in Baccino's (1999) study. As such, the examiner was able to utilize any and all notes and results derived from the other aging methods and re-examine skeletal structures to develop a comprehensive age estimate. As with applying the individual aging methods, the observer developed their experienced-base range blindly. The experience-based approach was a completely subjective approach as the observer did not have any parameters when constructing the age range for each individual.
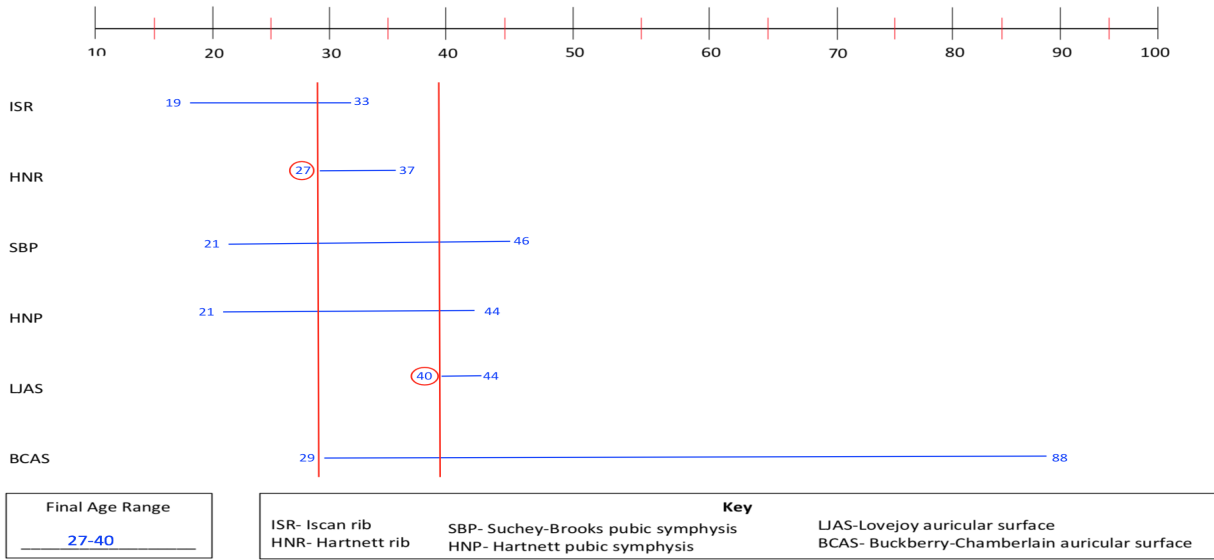
Figure 3: Example of how overlap ranges were determined.

Statistical Methodologies

A paired-samples *t*-test was conducted to determine if there was a significant difference between estimated age and actual age for all age-reporting strategies used in this study, including those derived from the six aging methods. In order to compare mean estimated age to mean actual age, the mean of the final age estimation was used. The null hypothesis of the paired-samples *t*-test is that the mean difference between estimated age and actual age is equal to zero. The null hypothesis assumes that any observable differences that are present are due to random variation. The alternative hypothesis is that the mean difference between the paired samples is not equal to zero and that something besides random variation is contributing to the difference. Samples were considered significantly different if the *p*-value, included in the results of the *t*-test, was less than 0.05.

All age-reporting strategies were evaluated for their accuracy and reliability. Accuracy and reliability were calculated for the sample as a whole, by age cohort, and by sex. In order to increase the sample size for the age assessment, age cohorts were expanded from six, ~10-year ranges to three, ~20-year ranges. The three age-cohorts represent "young" (20-39), "middle-age" (40-59), and "old-age" (60-79) individuals in the sample. The accuracy of an age-reporting strategy was calculated by assessing the number of individuals who were correctly assigned to a range that included their age-at-death. Accuracy was calculated as follows:

$$Accuracy(\%) = \frac{\#Correct}{\#Total}$$

47

The thresholds for accuracy and reliability were arbitrarily delineated for this study due to the lack of standards for these measures. The goal of this study was to identify thresholds that would be rigorous enough to distinguish the best performing age-reporting strategies, but not too restrictive that none of the strategies met the standards. Therefore, a final age-reporting strategy was considered accurate if 80% of the individuals in the sample were correctly assigned to a range that included their age-at-death. An 80% threshold ensured that the vast majority of individuals were correctly assigned to a range that included their age. Reliability was calculated by evaluating the inaccuracy and bias of each age estimation, as described by Meindl and colleagues (1985). Inaccuracy assesses the absolute difference of estimated age and actual age without considering under-/over-estimation and is calculated using the following equation:

$$\frac{\Sigma[estimated\ age - actual\ age]}{n}$$

where inaccuracy is the sum of the absolute value of estimated age minus the actual age divided by the number of individuals within the sample.

Bias is the mean over- or under-prediction of the individual's age and is calculated using the following equation:

$$\frac{\Sigma(estimated\ age\ -\ actual\ age)}{n}$$

where bias is the sum of estimated age minus actual age divided by the number of individuals within the sample. If the bias score is positive, then the age-reporting strategy overestimated age. If the bias score is negative, the age-reporting strategy underestimated age.

An age-reporting strategy was considered reliable if it had a low inaccuracy score and a bias score close to zero. Again, the thresholds for reliability were arbitrarily assigned. For this study, low inaccuracy was a mean difference of less than 10 years and minimal bias was a mean difference of greater than -1 year but less than 1 year. Age is often discussed within ten-year increments; therefore, one decade was used as the standard threshold for inaccuracy in this study. The bias threshold was particularly rigorous to exclude age-reporting strategies with gross systematic errors in either direction. All calculations for inaccuracy and bias were conducted in Excel. The inaccuracies and biases of the age-reporting strategies were further explored through a two-way repeated measures ANOVA in SPSS version 24 (IBM 2016). A repeated measures ANOVA is used to compare two or more group means where the participants are the same in each group (Girden 1992). For this study, there are two between-subject factors (sex and age-cohort) and 16 within-subject factors (the age-reporting strategies). Because all age-reporting strategies were applied to the same 58 individuals, a two-way repeated measures ANOVA was used to compare the age-reporting strategies for the total sample, by age-cohort, and by sex. To ensure that the assumptions of ANOVA tests were met, normality tests and tests of homogeneity of variance were conducted.

Four two-way repeated measures ANOVAs were conducted for this study. One ANOVA assessed the differences in bias of the age-reporting strategies with age-cohort as the between-subject factor, one assessed the differences in inaccuracy with age-cohort as the between-subject factor, one assessed bias with sex as the between-subject factor, and the last ANOVA assessed inaccuracy with sex as the between-subject factor. If the ANOVA was significant for any of the within and between-subject factors, pairwise comparison tables with a Bonferroni adjustment were referenced to see where the differences were. Finally, if the ANOVA showed a significant interaction between the within and between-subject factor, the data were split by the factor of interest (sex or age cohort), and a univariate ANOVA was conducted to better understand the interaction.

The null hypothesis for repeated measures ANOVA is that the mean of the variable being tested is the same for all groups (Emden 2008). The alternative hypothesis is that the mean of the variable being tested is not the same for all groups. For this study, the variables being tested are bias and inaccuracy. The F statistic was reported for each ANOVA conducted. The F statistic is the variance ratio produced by the ANOVA comparisons and signifies whether the effects of the experimental treatments are greater than the chance residual variation (Emden 2008). If the F value is less than 1.00, it indicates that the effects of the experimental treatment is less than the variation that would occur by chance.

Interobserver Error

Thirteen observers participated in an interobserver study for this research. Four undergraduate students, five graduate students, and four professional anthropologists were tasked with estimating age for two randomly selected skeletons from the original sample. The prerequisite for participation in the interobserver study was successful completion of an introduction to forensic anthropology class. All participants were provided with a packet containing the data collection forms used in this study and each had access to a binder containing the original publications of the aging methods being conducted.

The observers provided eight final age estimations: one for each of the six methods, one derived using the overlap approach, and one derived using the experienced-based approach. Observer and age-reporting strategy accuracy were assessed in the interobserver error study. Observer accuracy is how well individual observers were able to estimate the age of the skeletons using the different age-reporting strategies. For example, if an observer was only able to estimate age accurately using four of the eight strategies, then that observer's accuracy was 50%. Age-reporting strategy accuracy, in contrasts, assesses the percentage of people who accurately estimated age using each strategy. For example, if all 13 of the observers estimated age correctly with Suchey-Brooks, then this strategy was considered 100% accurate.

Observer experience on age estimation was also evaluated by comparing the three groups using a Chi-Square analysis in SPSS. Chi-Square tests are helpful for determining if there is an association between variables (Emden 2008).

Cronbach's alpha reliability coefficient was used to determine interobserver reliability. Cronbach's alpha is typically used to test reliability of survey questions, particularly how well the questions measure the variable of interest (Bonett and Wright 2015). Here, Cronbach's alpha is used to test observer consistency in assigning phases for each method. For example, if all observers selected Suchey-Brooks phase IV for skeleton one, this would demonstrate high reliability between observers. The formula for Cronbach's alpha is as follows:

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N - 1) \cdot \bar{c}}$$

Where N=the number of items, $\bar{c}$= the average covariance between item-pairs, and $\bar{v}$= the average variance. Cronbach's alpha was calculated in SPSS version 24.0 (IBM 2016). Cronbach's alpha is measured on a scale from 0-1, with one indicating perfect reliability. A Cronbach's alpha score of 0.7-0.8 indicates acceptable agreement, 0.8-09 indicates good agreement, and greater than 0.9 signifies excellent observer agreement (Goforth 2015).

Intraobserver Error

Ten skeletons were randomly chosen from the original sample to address intraobserver error. Included in the intraobserver sample were four females and six males whose ages-at-death ranged from 35-71 years old with a mean of 52 years old.

Each method was conducted as previously described and final age estimations were derived with the age-reporting strategies. The previously recorded data were compared to the new data using a paired-samples *t*-test in SPSS. Paired-samples *t*-test was selected since a subsample of the original sample was reevaluated using the same methodology (Emden 2008). Age ranges derived from the six methods were compared separately from the age ranges derived from the other ten age-reporting strategies. The results for both the methods and strategies were initially pooled and then individual methods/strategies were compared to each another.

CHAPTER FOUR
RESULTS

The goal of this study is to determine which age-reporting strategies produce the most accurate and reliable age-at-death estimations. The final age estimations derived from the age-reporting strategies were compared to each other and to the actual age of the skeleton. Accuracy and reliability were assessed for the total sample, by age-cohort, and by sex.

T-test Results Comparing Actual Age to Estimated Age

The results from the paired-samples *t*-test comparing actual age and estimated age are found in Table 5. There are significant differences in the mean values between actual age and estimated age for Lovejoy (LJAS) (t=4.107, df=57, *p*<0.05) Buckberry-Chamberlain (BCAS) (t=-3.485, df=57, *p*<0.05) and Least Error Auricular Surface (LEAS) (t=-3.485, df=57, *p*<0.05). Because only BCAS contributed to LEAS estimations, their results are identical and will be discussed together.

Table 6 shows the accuracy and reliability of the six skeletal aging methods by age cohort, sex, and total sample and Table 7 shows the accuracy and reliability of ten additional age-reporting strategies by age cohort, sex, and total sample. Because each estimation produced using the six aging methods is also considered an age-reporting strategy, they are included in the analyses comparing results of the different age-reporting strategies.

Table 5: Paired-samples *t*-test comparing actual and estimated age. Significant values are highlighted.

Paired Samples Test									
		Paired Differences							
				Std. Error Mean	95% Confidence Interval of the Difference				
		Mean	SD		Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Age - ISR	1.2500	11.6590	1.5309	-1.8156	4.3156	.817	57	.418
Pair 2	Age - HNR	-2.7328	12.4405	1.6335	-6.0038	.5383	-1.673	57	.100
Pair 3	Age - SBPS	2.0000	15.4056	2.0229	-2.0507	6.0507	.989	57	.327
Pair 4	Age - LJAS	7.6810	14.2425	1.8701	3.9362	11.4259	4.107	57	.000
Pair 5	Age - BCAS	-6.4397	14.0728	1.8479	-10.1399	-2.7394	-3.485	57	.001
Pair 6	Age - LEall	-1.1810	11.1145	1.4594	-4.1034	1.7414	-.809	57	.422
Pair 7	Age - LER	-2.7672	12.8838	1.6917	-6.1549	.6204	-1.636	57	.107
Pair 8	Age - LEPS	-.6552	13.7883	1.8105	-4.2806	2.9703	-.362	57	.719
Pair 9	Age - LEAS	-6.4397	14.0728	1.8479	-10.1399	-2.7394	-3.485	57	.001
Pair 10	Age - TSR	-.4138	12.5164	1.6435	-3.7048	2.8772	-.252	57	.802
Pair 11	Age - TSPS	-1.2586	14.3812	1.8883	-5.0400	2.5227	-.667	57	.508
Pair 12	Age - TSAS	-2.6638	14.9556	1.9638	-6.5962	1.2686	-1.356	57	.180
Pair 13	Age - TSOL	-.0086	10.1914	1.3382	-2.6883	2.6711	-.006	57	.995
Pair 14	Age - overlap	.1552	10.4748	1.3754	-2.5990	2.9094	.113	57	.911
Pair 15	Age - experience	.6293	8.2921	1.0888	-1.5510	2.8096	.578	57	.566
Pair 16	Age - HNPS	-2.0690	12.3702	1.6243	-5.3215	1.1836	-1.274	57	.208

Table 6: Method accuracy and reliability (inaccuracy and bias).

	İşcan Ribs	Hartnett Ribs	Suchey-Brooks Pubic Symphysis	Hartnett pubic symphysis	Lovejoy Auricular Surface	Buckberry-Chamberlain Auricular Surface
20-39 N=18 Accuracy Inaccuracy Bias	77.8% 7.00 2.06	38.9% 8.92 6.53	88.9% 8.81 0.47	66.7% 9.53 8.36	22.2% 8.89 5.33	77.8% 19.00 16.06
40-59 N=20 Accuracy Inaccuracy Bias	70% 10.08 1.68	35% 10.78 2.73	95% 14.10 7.35	80% 11.25 4.60	25% 7.95 -6.15	85% 12.07 9.48
60-79 N=20 Accuracy Inaccuracy Bias	90% 10.00 -7.15	50% 9.18 -0.68	70% 15.58 10.68	85% 9.28 -6.13	15% 20.93 -20.93	100% 7.85 -5.25
Males N=31 Accuracy Inaccuracy Bias	77.4% 8.10 -4.97	41.9% 9.82 4.08	74.2% 12.50 4.02	80.7% 10.58 -1.03	29.0% 12.58 -5.54	83.9% 13.18 9.24
Females N=27 Accuracy Inaccuracy Bias	81.5% 10.24 3.02	40.7% 9.44 1.19	96.3% 13.50 9.06	74.1% 9.41 5.63	11.1% 12.87 -10.13	92.59% 12.30 3.30
Total N=58 Accuracy Inaccuracy Bias	79.3% 9.10 -1.25	41.4% 9.65 2.73	84.4% 12.97 6.36	77.6% 10.03 2.07	20.7% 13.75 -8.72	87.9% 12.77 6.44

Table 7: Age-reporting strategy accuracy and reliability (inaccuracy and bias).

	Lowest Error All	Lowest Error Ribs	Lowest Error Pubic Symphysis	Lowest Error Auricular Surface	Two-step Ribs	Two-Step Pubic Symphysis	Two-step Auricular Surface	Overlap	Overlap Two-step	Experience
20-39 N=18										
Accuracy	44.4%	38.9%	83.3%	77.8%	66.7%	77.8%	55.6%	83.3%	72.2%	77.8%
Inaccuracy	7.36	9.83	8.44	19.00	6.61	10.19	14.56	5.72	6.64	5.42
Bias	4.92	7.56	8.22	16.06	0.33	9.36	11.00	4.00	4.03	4.58
40-59 N=20										
Accuracy	35%	30%	80%	85%	30%	75%	60%	65%	85%	85%
Inaccuracy	8.70	10.20	10.68	12.08	10.53	12.20	14.43	8.05	6.28	5.63
Bias	1.80	3.30	3.98	9.48	2.98	3.55	4.48	0.05	1.78	-0.08
60-79 N=20										
Accuracy	40%	45%	80%	100%	45%	80%	95%	50%	65%	75%
Inaccuracy	10.70	9.98	12.08	7.85	9.98	11.48	8.95	10.45	9.68	7.28
Bias	-2.80	-2.08	-9.48	-5.25	-2.08	-8.33	-6.65	-4.10	-5.38	-5.88
Males N=31										
Accuracy	41.9%	42%	87.1%	83.9%	48.4%	77.42%	61.29%	61.30%	71.0%	71.0%
Inaccuracy	8.58	10.02	10.82	13.18	9.45	12.68	13.5	8.86	8.42	7.19
Bias	1.84	3.37	-3.47	9.24	0.58	-2.87	4.82	-4.68	-0.61	-0.84
Females N=27										
Accuracy	37.0%	33.3%	74.1%	92.6%	44.4%	77.8%	81.5%	70.4%	77.78%	88.9%
Inaccuracy	9.43	10.0	10.06	12.30	8.74	9.78	11.52	7.35	6.57	4.91
Bias	0.43	2.07	5.39	3.30	0.22	6.0	0.19	0.20	0.72	-0.39
Total (N=58)										
Accuracy	39.7%	37.9%	81.0%	87.9%	46.6%	77.6%	70.7%	65.5%	74.1%	79.5%
Inaccuracy	8.97	10.01	10.47	12.77	9.12	11.33	12.58	8.16	7.56	6.13
Bias	1.18	2.77	0.66	6.44	0.41	1.26	2.66	-0.16	0.01	-0.63

Total Sample Results

When considering the total sample (n=58), the most accurate age estimations
are produced using Buckberry-Chamberlain (BCAS)/ Least Error Auricular Surface
(LEAS) (87.9%), Suchey-Brooks (SBPS) (84.5%), Least Error Pubic Symphysis (LEPS)
(81.03%) and experience (79.5%). The least accurate age ranges are those produced
using Lovejoy (LJAS) (20.7%), Least Error Rib (LER) (37.93%) and Least Error-All
(Leall) (39.7%). The accuracies of the other age-reporting strategies range from 41.4%-
79.3%.

Figure 4 is a chart of the mean biases of each age-reporting strategy for the total
sample. The age-reporting strategies with the lowest mean biases ($\bar{x} \geq -1, \bar{x} \leq 1$) are
Overlap: Two-Step (TSOL) (0.01 years), overlap (-0.16 years), Two-Step Rib (TSR)
(0.41 years), experience (-0.63 years), and LEPS (0.66 years). The strategies with the
highest biases are LJAS (-8.72 years), BCAS/ LEAS (6.67 years). LJAS, SBPS, İşcan
Rib (ISR), experience, and overlap tend to underestimate age and all other strategies
tend to overestimate age.

The two-way repeated measures ANOVA comparing the mean biases of each
age-reporting strategy had one between-subject factor (age cohort) and one within-
subject factor (age-reporting strategies). The between-subject factor contained three
levels (young, middle, and old) and the within-subject factor had 16 levels, one
representing each age-reporting strategy in this study. The results (Table 8) indicate
that there is a significant difference in mean bias between at least two age-reporting
strategies [$F(15, 348.46)=7.41$, $p<0.01$] and that there is a significant difference in bias
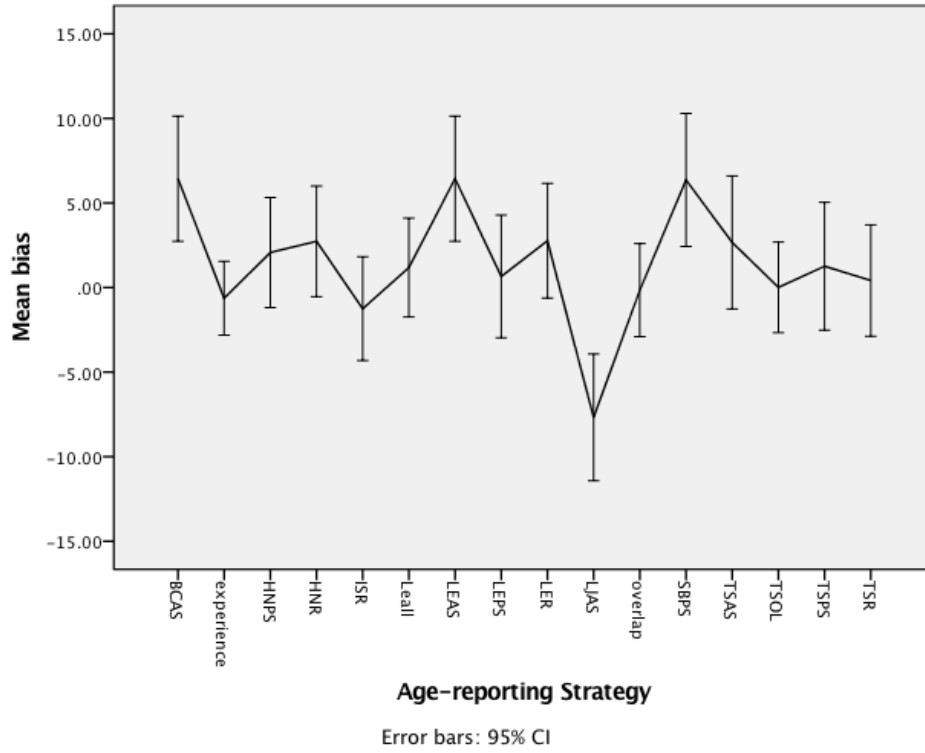
Figure 4: Mean bias of each age-reporting strategy for total sample.

Table 8: Results of the two-way ANOVA comparing age-reporting strategy biases with age-cohort as the between-subject factor.

Type III Tests of Fixed Effects[a]				
Source	Numerator df	Denominator df	F	Sig.
Intercept	1	49.321	6.065	.017
reporting_strategies	15	348.457	7.409	.000
agegroup	2	49.321	28.139	.000
reporting_strategies * agegroup	30	348.457	4.380	.000
a. Dependent Variable: bias.				

between at least two age-cohorts [$F_{(2, 49.32)}$=28.14, $p<0.01$]. Additionally, the ANOVA shows that there is a significant interaction between the reporting strategies and age-group [$F_{(30, 348.46)}$=4.38, $p<0.01$].

LJAS has the greatest amount of bias (Table 6) and the results of the post-hoc test show that it is significantly different ($p<0.01$) than all other strategies excluding ISR ($p=0.09$). BCAS/ LEAS has significantly different ($p<0.05$) bias scores than the experience-based approach, ISR, LJAS, overlap, and TSOL age-reporting strategies. The overlap age-reporting strategy only has significantly different ($p<0.01$) bias scores from LJAS.

The pairwise comparison table (Table 9) indicates that there is a significant difference ($p<0.05$) in the bias scores between the young and old cohort and between the middle and old cohort. There is not a significant difference ($p=0.10$) between the bias scores of the young and middle-age cohorts. Because the results show a significant difference in the bias scores between age groups (Table 9) and a significant interaction between reporting-strategies and age-cohort [$F_{(30, 348.46)}$=4.39, $p<0.01$] (Table 8), the data were split by age groupings and a univariate ANOVA was conducted. The results from univariate ANOVA (Table 10) shows that there is a significant difference between in the mean biases of at least two of the age-reporting strategies for all three age cohorts. Figure 5 displays the mean biases of each age-reporting strategy by age-cohort.

Table 9: Differences in age-reporting strategy mean bias between the different age-cohorts.

						95% Confidence Interval for Difference[c]	
(I) age group	(J) age group	Mean Difference (I-J)	Std. Error	df	Sig.[c]	Lower Bound	Upper Bound
Young	Middle age	3.618	1.639	49.321	.096	-.444	7.679
	Old age	11.894[*]	1.639	49.321	.000	7.833	15.956
Middle age	Young	-3.618	1.639	49.321	.096	-7.679	.444
	Old age	8.277[*]	1.595	49.321	.000	4.323	12.230
Old age	young	-11.894[*]	1.639	49.321	.000	-15.956	-7.833
	Middle age	-8.277[*]	1.595	49.321	.000	-12.230	-4.323
Based on estimated marginal means							
*. The mean difference is significant at the .05 level.							
a. Dependent Variable: bias.							
c. Adjustment for multiple comparisons: Bonferroni.							

Title row above table: **Pairwise Comparisons[a]**

Table 10: Results of univariate ANOVA comparing biases of age-reporting strategies by age-cohort.

| Age group	Source	Type III Sum of Squares	df	Mean Square	F	Sig.
		Tests of Between-Subjects Effects				
	Dependent Variable: bias					
young	Corrected Model	5960.513[a]	15	397.368	3.397	.000
	Intercept	13332.084	1	13332.084	113.976	.000
	reporting_strategies	5960.513	15	397.368	3.397	.000
	Error	31816.653	272	116.973		
	Total	51109.250	288			
	Corrected Total	37777.166	287			
middle-age	Corrected Model	4299.049[b]	15	286.603	1.910	.022
	Intercept	3248.063	1	3248.063	21.644	.000
	reporting_strategies	4299.049	15	286.603	1.910	.022
	Error	45621.138	304	150.070		
	Total	53168.250	320			
	Corrected Total	49920.187	319			
old-age	Corrected Model	11627.397[c]	15	775.160	6.315	.000
	Intercept	8292.628	1	8292.628	67.557	.000
	reporting_strategies	11627.397	15	775.160	6.315	.000
	Error	37315.975	304	122.750		
	Total	57236.000	320			
	Corrected Total	48943.372	319			
a. R Squared = .158 (Adjusted R Squared = .111)						
b. R Squared = .086 (Adjusted R Squared = .041)						
c. R Squared = .238 (Adjusted R Squared = .200)						
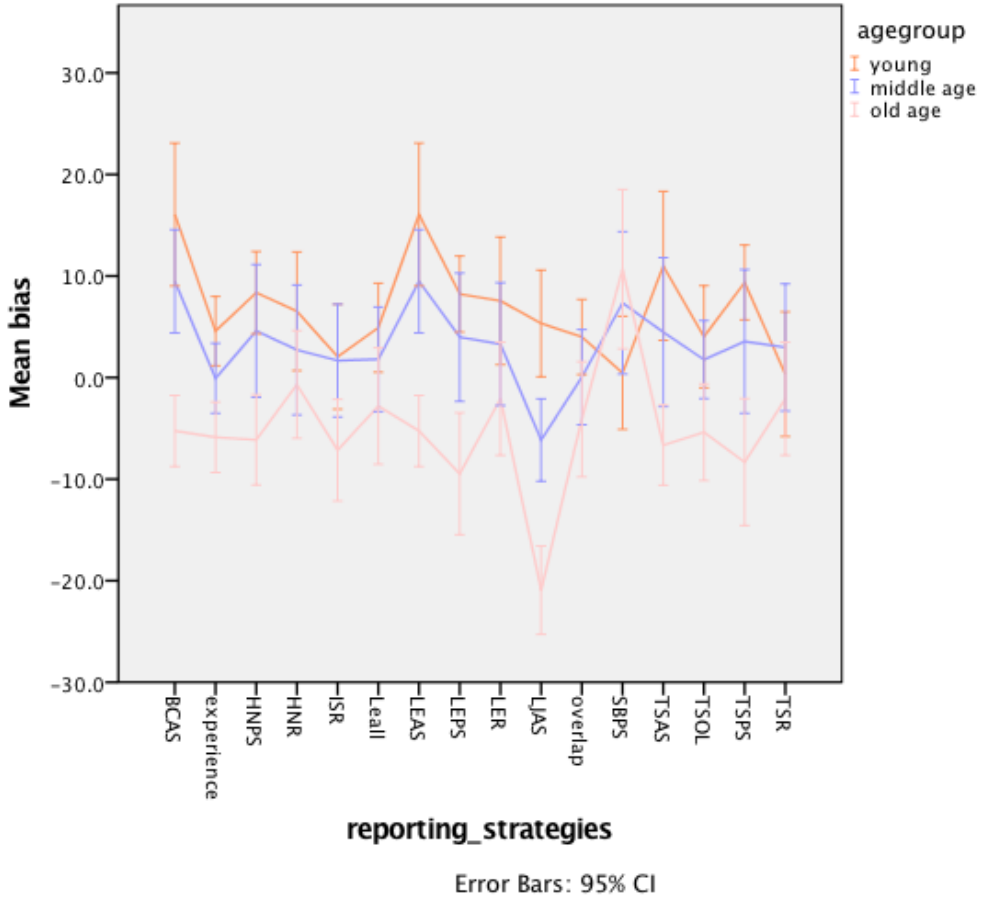
Figure 5: Age-reporting strategy mean bias by age cohort.

The results of the ANOVA comparing bias with sex as the between-subject factor shows that there are no significant differences in the bias results between the sexes but that there is a significant interaction between age-reporting strategies and sex (Table 11). Figure 6 displays the mean bias of each age-reporting strategy by sex. Because there is a significant interaction between sex and age-reporting strategy, the data were split by sex and a univariate ANOVA was conducted. The results of the ANOVA comparing biases of age-reporting strategies by sex (Table 12) shows that there is a significant difference in the mean bias of age-reporting strategies for both males $F(1,15)=3.50$, $p<0.05$ and females $F(1,15)=3.38$, $p<0.05$.

The mean inaccuracies of all age-reporting strategies for the total sample are represented in Figure 7. The strategies with the lowest mean inaccuracies ($\bar{x}$ <10 years) are experience-based approach (6.13 years), TSOL (7.56 years), overlap (8.16 years), Leall (8.97 years), ISR (9.09 years), Two-Step Rib (TSR) (9.12 years), and Hartnett Rib (HNR) (9.65 years). The strategy with the highest amount of inaccuracy is LJAS (13.75 years). All other method inaccuracies range between 10.03-12.78 years. None of the age-reporting strategies meet all the criteria for accuracy and reliability for the total sample, however, the experience-based approach meets both criteria for reliability (inaccuracy and bias) and is only 0.5% away from meeting the criteria for accuracy.

The results of the two-way ANOVA comparing age-reporting strategy inaccuracies with age-cohort as the between-subject factor (Table 13) show that there is a significant difference [$F(15, 347.44)=5.09$, $p<0.01$] in the mean inaccuracy between at least two age-reporting strategies, but not between the three age-cohorts

Table 11: Results of the two-way ANOVA comparing age-reporting strategy biases with sex as the between-subject factor.

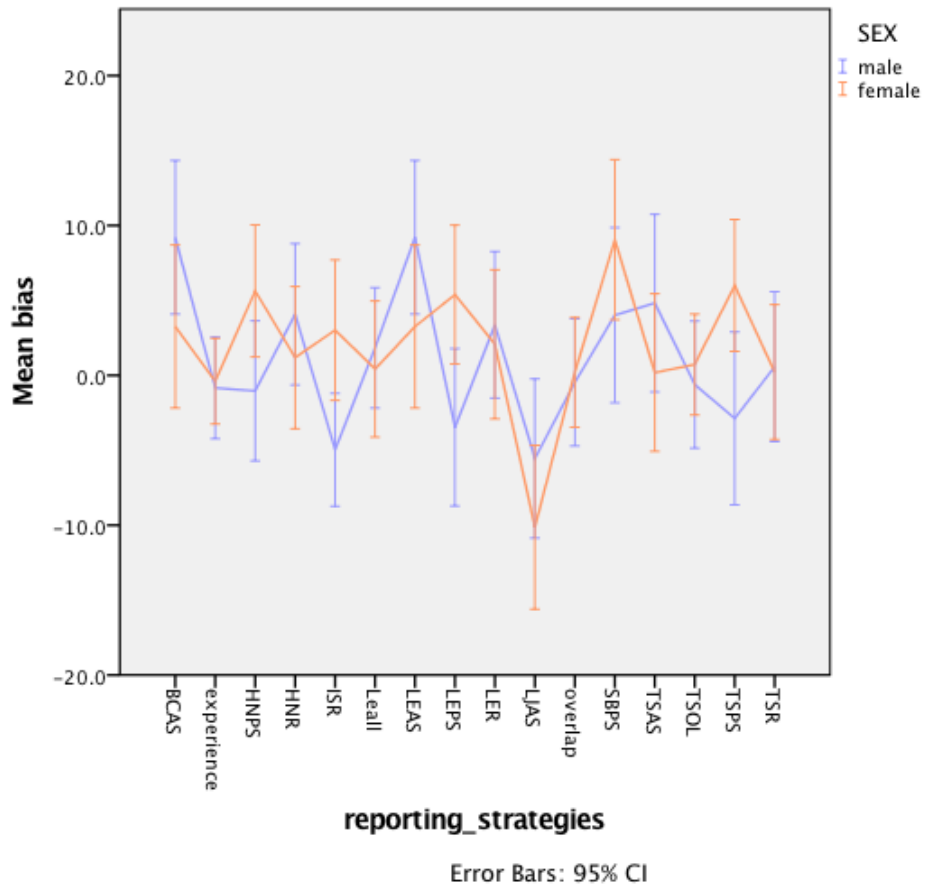| Type III Tests of Fixed Effects[a]				
Source	Numerator df	Denominator df	F	Sig.
Intercept	1	56.012	2.195	.144
reporting_strategies	15	431.072	7.045	.000
SEX	1	56.012	.160	.691
reporting_strategies * SEX	15	431.072	4.202	.000
a. Dependent Variable: bias.				

Figure 6: Age-reporting strategy mean bias by sex.

Table 12: Results of univariate ANOVA comparing biases of age-reporting strategies by sex.

|						
Tests of Between-Subjects Effects						
Dependent Variable: bias						
SEX	Source	Type III Sum of Squares	df	Mean Square	F	Sig.
male	Corrected Model	9302.966[a]	15	620.198	3.499	.000
	Intercept	581.389	1	581.389	3.280	.071
	reporting_strategies	9302.966	15	620.198	3.499	.000
	Error	85086.145	480	177.263		
	Total	94970.500	496			
	Corrected Total	94389.111	495			
female	Corrected Model	7066.461[b]	15	471.097	3.382	.000
	Intercept	1530.021	1	1530.021	10.984	.001
	reporting_strategies	7066.461	15	471.097	3.382	.000
	Error	57946.519	416	139.295		
	Total	66543.000	432			
	Corrected Total	65012.979	431			
a. R Squared = .099 (Adjusted R Squared = .070)						
b. R Squared = .109 (Adjusted R Squared = .077)						
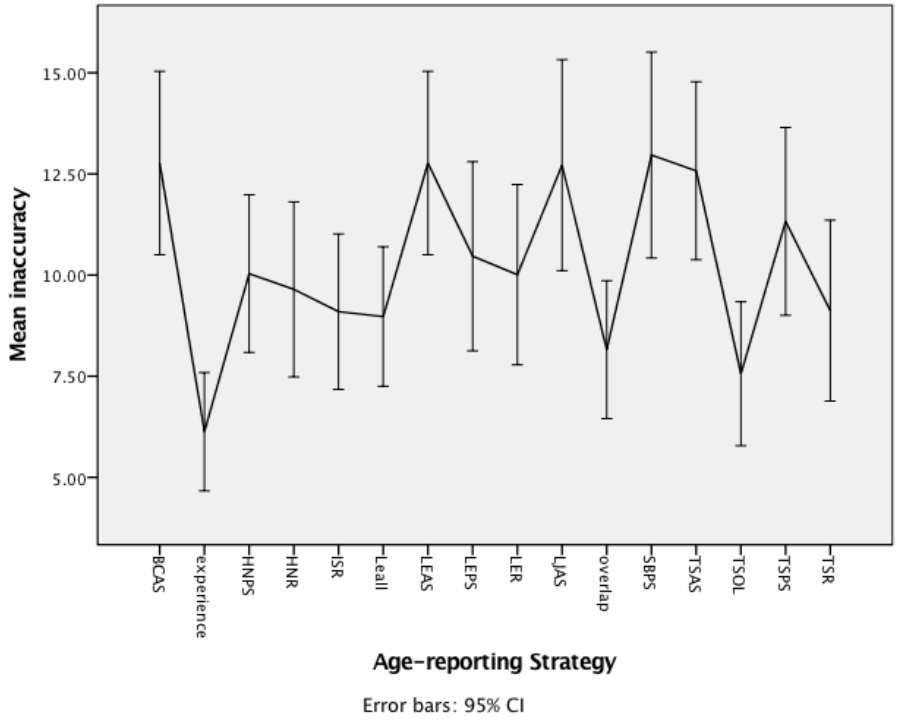
Figure 7: Mean inaccuracies of each age-reporting strategy for total sample.

Table 13: Results of the two-way ANOVA comparing age-reporting strategy inaccuracies with age-cohort as the between-subject factor.

Type III Tests of Fixed Effects[a]				
Source	Numerator df	Denominator df	F	Sig.
Intercept	1	41.387	615.460	.000
reporting_strategies	15	347.437	5.093	.000
agegroup	2	41.387	.440	.647
reporting_strategies * agegroup	30	347.437	4.161	.000
a. Dependent Variable: inaccuracy.				

[$F_{(2, 41.39)}$=0.44, p=0.65]. The results also indicate that there is a significant interaction [$F_{(30, 347.44)}$= 4.16, p<0.01] between the age-reporting strategies and age-cohorts. The pairwise comparison table for the total sample shows that the inaccuracies of the experience-based approach are significantly different (p<0.05) than the inaccuracies of BCAS/LEAS, LJAS, SBPS, TSAS, and TSPS. The TSOL strategy has significantly different (p<0.05) inaccuracies from BCAS/LEAS, SBPS, and TSAS. Finally, the overlap age-reporting strategy is significantly different (p<0.05) from BCAS/LEAS, and SBPS.

Table 14 shows the comparison of inaccuracies by age-cohort. There are no significant differences (p=1.00) between any of the three age-cohorts. However, due to the significant interaction of age-reporting strategies and age-cohort, the data were split by age group and a univariate ANOVA was conducted to understand the nature of this interaction. The results of the univariate ANOVA (Table 15) show that there is a significant difference between age-reporting strategy inaccuracies for all age-cohorts. Figure 8 illustrates the mean inaccuracies of each age-reporting strategy by age-cohort. A pairwise comparison table was used to determine the differences between age-reporting strategy inaccuracies by age-cohort.

Finally, the results of a two-way ANOVA comparing inaccuracies with sex as the between-subject factor (Table 16) shows that there is not a significant difference in the inaccuracies between the sexes [$F_{(1, 42.59)}$=0.74, p=0.39] nor a significant interaction [$F_{(15, 347.99)}$=0.45, p=0.96]. Because the ANOVA comparing inaccuracies by sex did not yield significant results, no further tests were conducted. Figure 9 shows the mean inaccuracies of each age-reporting strategy by sex.

Table 14: Differences in age-reporting strategy mean inaccuracy between the different age-cohorts.

						95% Confidence Interval for Difference[b]	
(I) agegroup	(J) agegroup	Mean Difference (I-J)	Std. Error	df	Sig.[b]	Lower Bound	Upper Bound
young	middle-age	-.566	1.021	41.387	1.000	-3.114	1.982
	old-age	-.955	1.021	41.387	1.000	-3.503	1.593
middle-age	young	.566	1.021	41.387	1.000	-1.982	3.114
	old-age	-.389	.994	41.387	1.000	-2.869	2.091
old-age	young	.955	1.021	41.387	1.000	-1.593	3.503
	middle-age	.389	.994	41.387	1.000	-2.091	2.869

Pairwise Comparisons[a]

Based on estimated marginal means

a. Dependent Variable: inaccuracy.

b. Adjustment for multiple comparisons: Bonferroni.

Table 15: Results of univariate ANOVA comparing inaccuracies of age-reporting strategies by age-cohort.

|						
Tests of Between-Subjects Effects						
Dependent Variable: inaccuracy						
Age group	Source	Type III Sum of Squares	df	Mean Square	F	Sig.
young	Corrected Model	4793.506[a]	15	319.567	4.583	.000
	Intercept	27348.758	1	27348.758	392.201	.000
	reporting_strategies	4793.506	15	319.567	4.583	.000
	Error	18966.986	272	69.732		
	Total	51109.250	288			
	Corrected Total	23760.492	287			
middle-age	Corrected Model	1878.924[b]	15	125.262	2.205	.006
	Intercept	34020.938	1	34020.938	598.919	.000
	reporting_strategies	1878.924	15	125.262	2.205	.006
	Error	17268.388	304	56.804		
	Total	53168.250	320			
	Corrected Total	19147.312	319			
old-age	Corrected Model	3377.125[c]	15	225.142	3.974	.000
	Intercept	36636.800	1	36636.800	646.704	.000
	reporting_strategies	3377.125	15	225.142	3.974	.000
	Error	17222.075	304	56.652		
	Total	57236.000	320			
	Corrected Total	20599.200	319			
a. R Squared = .202 (Adjusted R Squared = .158)						
b. R Squared = .098 (Adjusted R Squared = .054)						
c. R Squared = .164 (Adjusted R Squared = .123)						
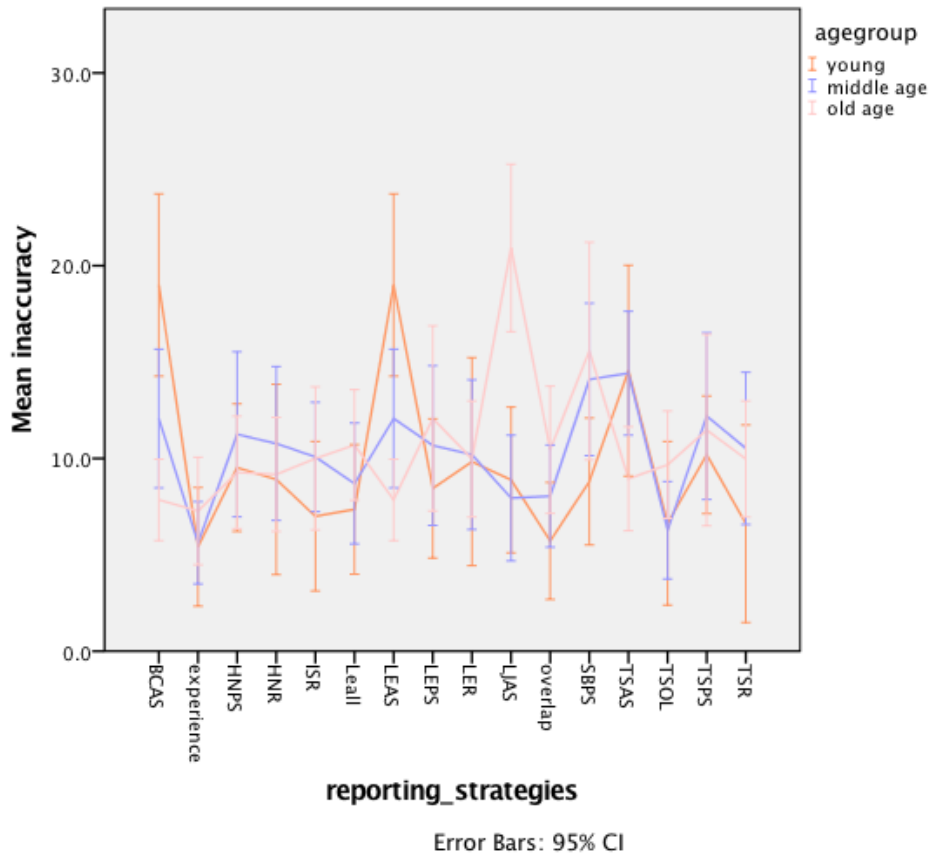
Figure 8: Mean inaccuracies of age-reporting strategies by age cohort.

Table 16: Results of the two-way ANOVA comparing age-reporting strategy inaccuracies with sex as the between-subject factor.

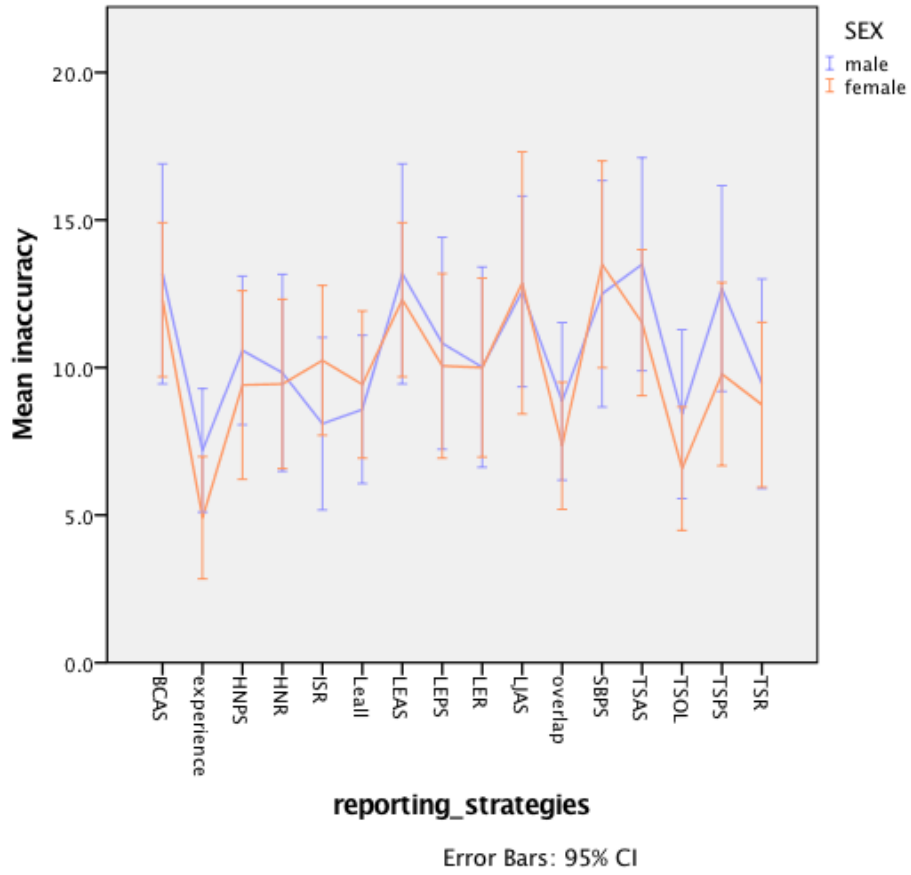| Type III Tests of Fixed Effects[a]				
Source	Numerator df	Denominato r df	F	Sig.
Intercept	1	42.588	655.246	.000
reporting_strategies	15	347.992	4.321	.000
SEX	1	42.588	.744	.393
reporting_strategies * SEX	15	347.992	.451	.962
a. Dependent Variable: inaccuracy.				

Figure 9: Mean inaccuracies of age-reporting strategies by sex.

Results by age-cohort

The age-cohorts in this study are defined as young (20-39 years old), middle-age (40-59 years old), and old-age (60-79 years old). All accuracy and reliability (inaccuracy and bias) scores by age-cohort are found in Tables 5 and 6. If the ANOVA results indicated that there are differences in the mean bias or inaccuracies of age-reporting strategies based on age-cohort, the pairwise comparison chart was consulted to see where the differences are.

Young (20-39 years old) Cohort Results

The age-reporting strategies that resulted in the most accurate estimations for the young cohort (n=18) are Suchey-Brooks (SBPS) (88.9%), overlap (83.3%), and Least Error Pubic Symphysis (LEPS) (83.3%). The least accurate age-reporting strategies are Lovejoy (LJAS) (22.2%), Hartnett Rib (HNR) (38.9%), and Least Error Rib (LER) (38.9%). All other age-reporting strategies have accuracies ranging from 77.8%-44.4%.

Figure 10 displays the mean bias of each age-reporting strategy for the 20-39 age cohort. The age-reporting strategies with minimal bias ($\bar{x} \geq -1, \bar{x} \leq 1$) are Two-Step Rib (TSR) (0.33 years) and SBPS (0.47 years). The strategies with the most amounts of bias are Least Error Auricular Surface (LEAS) (16.06 years) and BCAS (16.06 years). Because the univariate ANOVA comparing biases of age-reporting strategies by age group yielded significant results for the young age-cohort (Table 10), the pairwise comparison table was referenced to see which strategies have significantly different biases from one another.
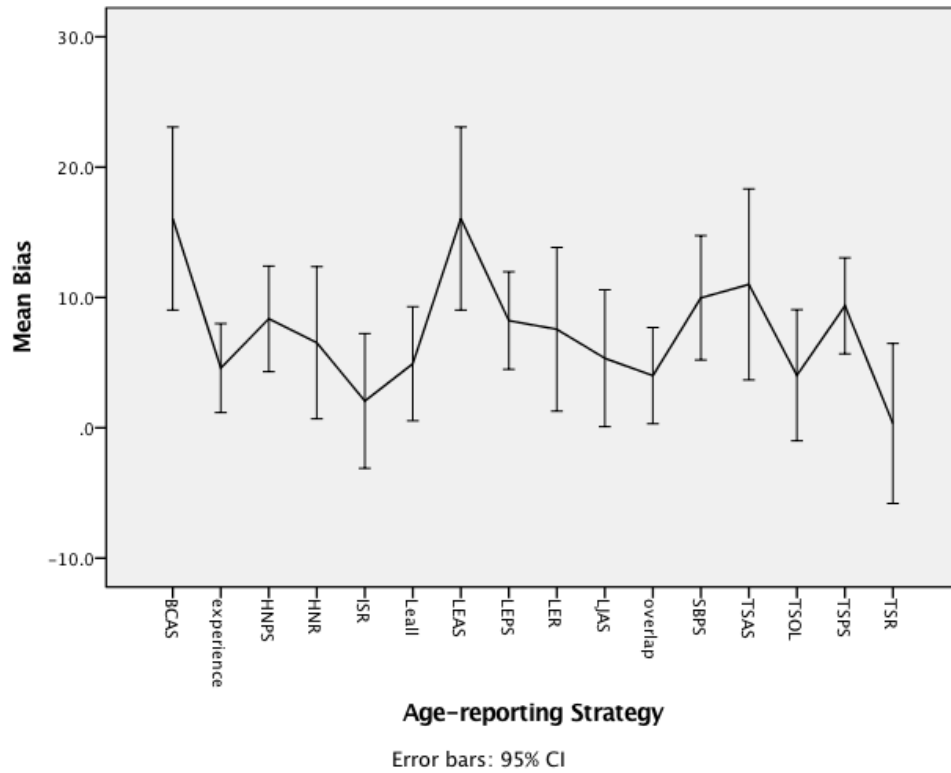
Figure 10: Mean bias of each age-reporting strategy for age cohort (20-39 years old).

The results show that both SBPS and TSR have significantly different ($p<0.01$) bias scores from BCAS/LEAS. BCAS /LEAS are also significantly different ($p<0.05$) than Işcan (ISR) with regards to bias.

Figure 11 illustrates the mean inaccuracy of each age-reporting strategy for the 20-39 age cohort. The strategies with the lowest inaccuracies ($\bar{x}$ <10 years) are experience (5.42 years), overlap (5.72 years), TSR (6.61 years), Overlap: Two-Step (TSOL) (6.64 years), ISR (7 years), Least Error- All (Leall) (7.36 years), LEPS (8.44 years), SBPS (8.81 years), LJAS (8.89 years), HNR (8.92 years), Hartnett Pubic Symphysis (HNPS) (9.53 years), and LER (9.83 years). The strategies with the highest inaccuracies are BCAS/ LEAS (19 years). The other age-reporting strategy, Two Step Pubic Symphysis (TSPS), have an inaccuracy of 10.19 years for the young cohort.

Again, the results of the univariate ANOVA indicateds that the inaccuracy scores of at least two strategies are significantly different for the young cohort (Table 15). The pairwise comparison table shows that BCAS/ LEAS has significantly different ($p<0.05$) inaccuracies from experience, HNR, ISR, LEPS, LJAS, overlap, SBPS, TSOL, and TSR. Only SBPS meets all the criteria for accuracy and reliability for the young age cohort.

Middle-age (40-59 years old) Cohort Results

The middle-age cohort included 20 individuals. The SBPS method is the most accurate age estimation for the middle-age cohort with 95% accuracy. Experience, TSOL, BCAS, and LEAS are 85% accurate and HNPS and LEPS are 80% accurate.
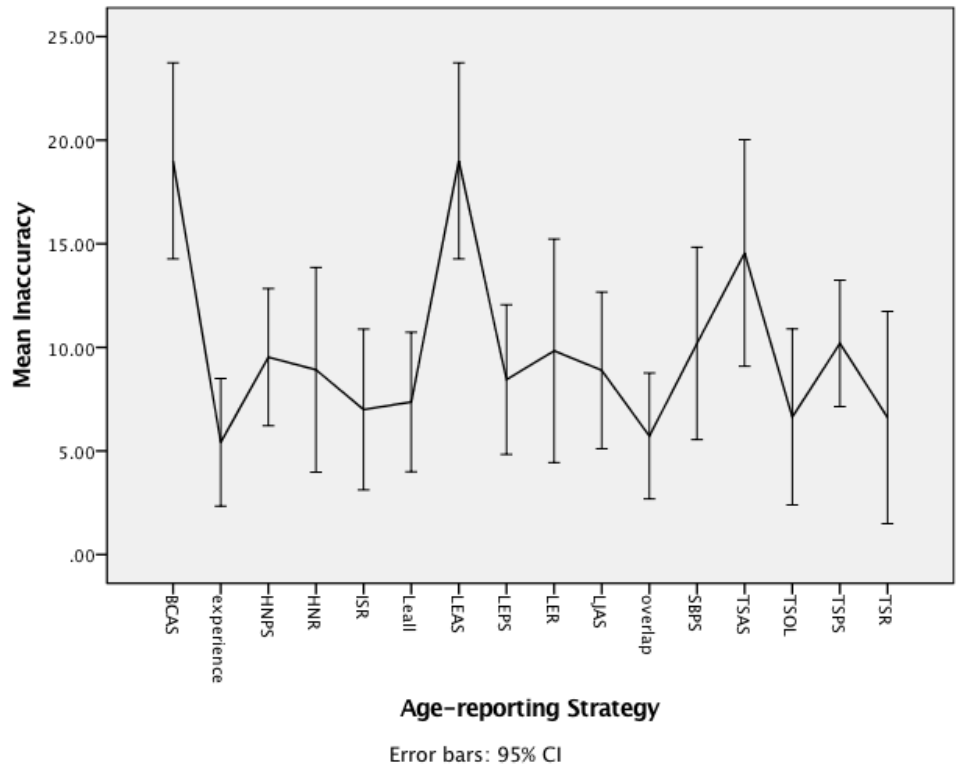
Figure 11: Age-reporting strategy inaccuracies for young cohort (20-39 years old).

The least accurate age-reporting strategy is LJAS with 25% accuracy followed by LER and TSR with 30% accuracy and Leall and HNR with 35% accuracy. All other strategies have accuracies ranging from 60%-75%.

The mean bias of each strategy is shown in Figure 12. The strategies with the lowest biases ($\bar{x} \geq -1, \bar{x} \leq 1$) for the 40-59 age-cohort are the overlap (0.05 years) and experience (-0.08 years) strategies. The strategies with the highest mean biases are BCAS /LEAS (9.48 years). The univariate ANOVA (Table 9) comparing bias scores of by age-cohort shows that at least two strategies have significantly different biases. The pairwise comparison table revealed that only BCAS/LEAS has significantly different biases (p<0.05) from LJAS.

Figure 13 displays the mean inaccuracies for each age-reporting strategy for the middle-age cohort. The age-reporting strategies with the lowest inaccuracies ($\bar{x}$ <10 years) for the middle-age cohort are experience (5.63 years), TSOL (6.28 years), and LJAS (7.95 years). The strategies with the highest inaccuracies for the middle-age cohort are TSAS (14.43 years) and SBPS (14.10 years). All other age-reporting strategies have inaccuracies ranging from 10.08-12.20 years.

The results of the univariate ANOVA comparing age-reporting strategy inaccuracy by age-cohort (Table 15) indicates that there are significant differences between at least two strategies for the middle-age cohort. The pairwise comparison table revealed that the experience strategy has significantly different bias scores from TSAS (p<0.05) and SBPS (p=0.05). Only the experience-based approach meets the accuracy and reliability criteria for the middle-age cohort
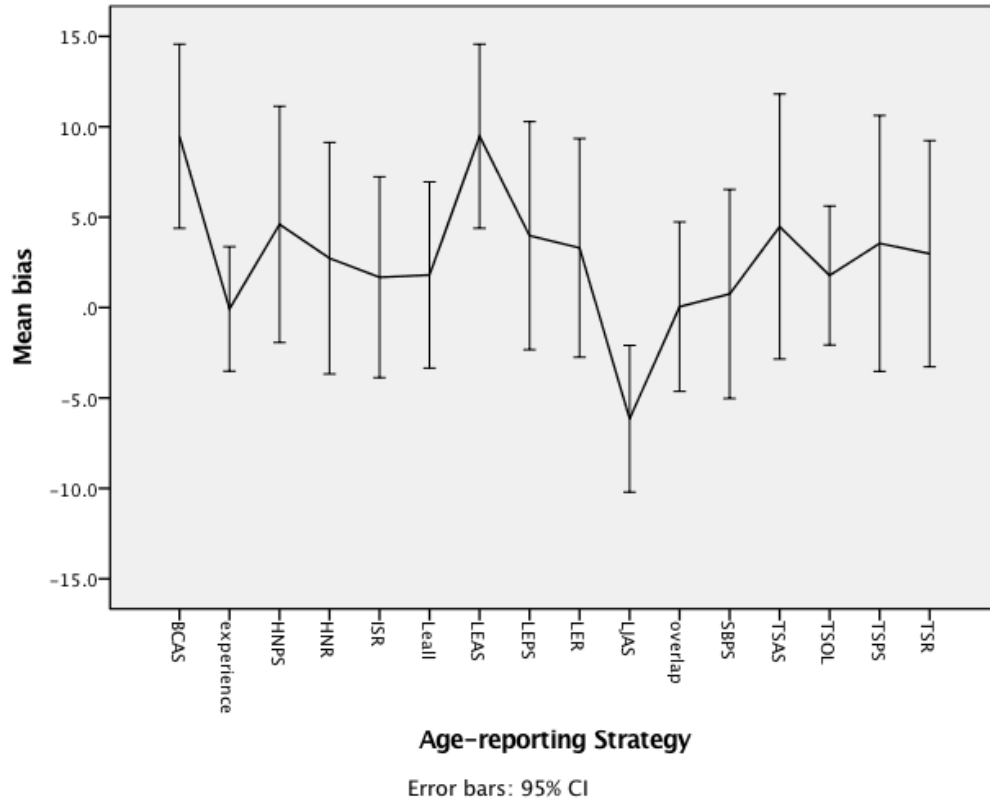
Figure 12: Age-reporting strategy biases for middle-age cohort (40-59 years old).
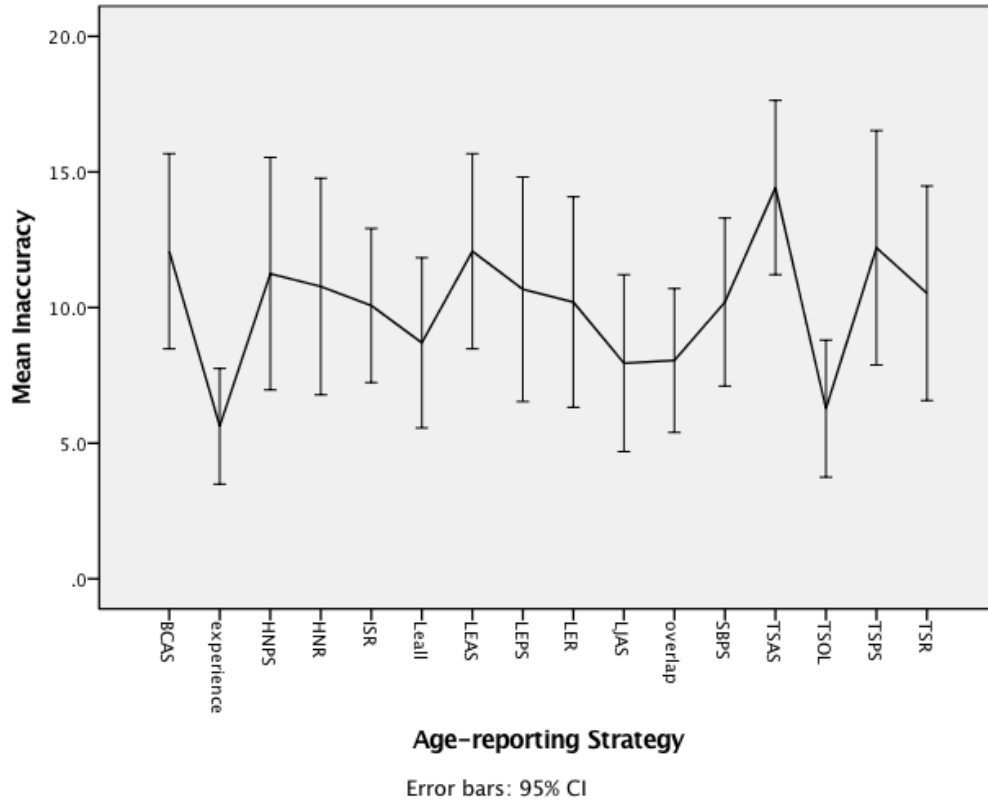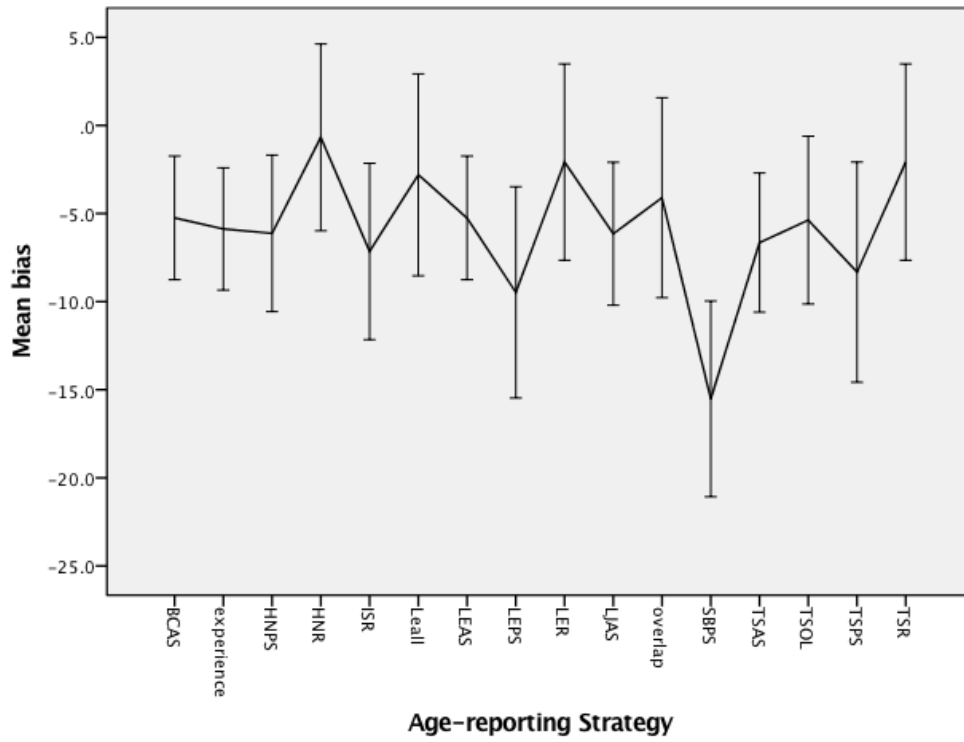
Figure 13: Age-reporting strategy inaccuracies for middle-age cohort (40-59 years old).

Old-age (60-79 years old) Cohort Results

The most accurate age-estimation strategies for the old-age cohort (n=20) are BCAS/ LEAS with 100% accuracy followed by TSAS (90%), HNPS (85%), LEPS (80%), and TSPS (80%). The least accurate method is LJAS with 15% accuracy. All other age-reporting strategies have accuracies ranging from 40-75%.

The mean bias scores for each age-reporting strategy for the old-age cohort are displayed in Figure14. The only age reporting strategy with a minimal bias ($\bar{x} \geq -1, \bar{x} \leq 1$) is HNR (-0.68 years). The age-reporting strategy with the highest bias is LJAS (-20.93 years). All age-reporting strategies excluding SBPS (10.68 years) tend to underestimate age for individuals in the 60-79 year age-cohort. The univariate ANOVA comparing bias scores of age-reporting strategies by age-cohort (Table 10) shows that at least two age-reporting strategy biases are significantly different. According to the pairwise comparison table, LJAS differed significantly ($p<0.05$) from all age-reporting strategies excluding LEPS. SBPS also has significantly different ($p<0.05$) biases from all age-reporting strategies except for HNR.

The mean inaccuracy of each age-reporting strategy for the old-age cohort are represented in Figure 15. The age-reporting strategies with the lowest inaccuracy scores ($\bar{x}$ <10 years) are experience (7.28 years), BCAS/ LEAS (7.85 years), TSAS (8.95 years), HNR (9.18 years), HNPS (9.28 years), TSOL (9.68 years). TSR (9.98 years) and LER (9.98 years). The age-reporting strategy with the highest inaccuracy scores is LJAS (20.92 years). All other age-reporting strategies have inaccuracies ranging from 10-15.58 years.

Figure 14: Age-reporting strategy biases for old-age cohort (60-79 years old).
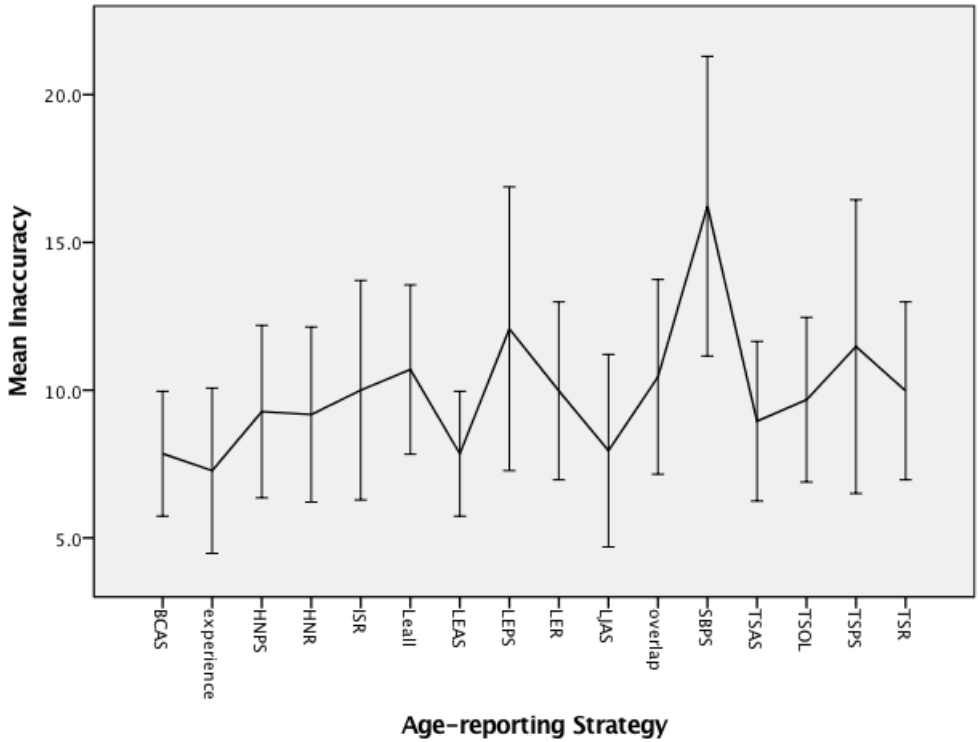
Figure 15: Age-reporting strategy inaccuracies for old-age cohort (60-79 years old).

The univariate ANOVA comparing inaccuracies of age-estimation strategies by age-cohort (Table 15) indicates that there are differences between at least two strategies.

The pairwise comparison table revealed that LJAS differed significantly ($p<0.05$) from all strategies except SBPS. No other age-reporting strategies differed significantly from one another with respects to inaccuracy. None of the age-reporting strategies meet both criteria for accuracy and reliability for the old-age cohort.

Results by Sex

All accuracy and reliability (inaccuracy and bias) scores for males and females are found in Tables 5 and 6. If the two-way ANOVAs with sex as the between-subject factor (Tables 11 and 16) indicates that there are differences in the mean bias or inaccuracies by sex, pairwise comparison tables were consulted to see where the differences are.

Females

The results show that the most accurate age-reporting strategies for females (n=27) are SBPS (96.3%), BCAS/ LEAS (92.6%), experience (88.9%), TSAS (81.5%) and ISR (81.5%). The least accurate age-reporting strategies for females is LJAS (11.1%). All other age-reporting strategies ranged in accuracy from 33.3%-70.4%.

Figure 16 displays the mean bias for each age-reporting strategy for females. The age-reporting strategies with the lowest mean biases ($\bar{x} \geq -1, \bar{x} \leq 1$) are TSAS (0.19 years), overlap (0.20), TSR (0.22 years), experience (-0.39 years), Leall (0.43 years) and TSOL (0.72). The age-reporting strategy with the largest mean bias is LJAS (-10.13 years).
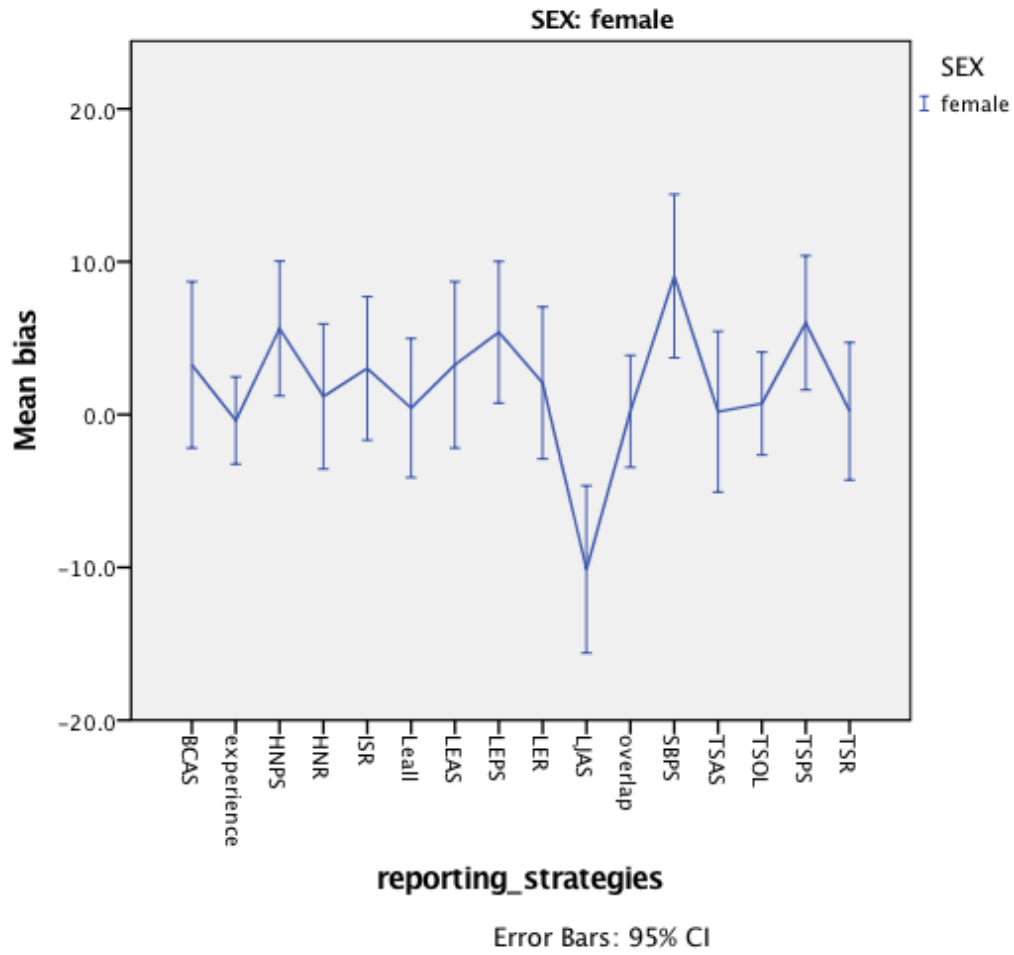
Figure 16: Mean biases of age-reporting strategies for females.

The univariate ANOVA comparing biases of age-reporting by sex (Table 12) indicates that at least two age-reporting strategies differed significantly with regards to their bias scores for females. The pairwise comparison table showed that LJAS was significantly different ($p<0.05$) from BCAS, HNPS, ISR, LEAS, LEPS, LER, SBPS, and TSPS.

The mean inaccuracies of age-reporting strategy for females are displayed in Figure 17. The age-reporting strategies with the lowest mean inaccuracies ($\bar{x}$ <10 years) are experience (4.91 years), TSOL (6.58 years), overlap (7.35 years), TSR (8.74 years), HNPS (9.41 years), HNR (9.44 years), and TSPS (9.78 years). The age-reporting strategy with the highest inaccuracy was SBPS (13.5 years). All other strategies have inaccuracies ranging from 10-12.8 years. The results of the two-way repeated measures ANOVA comparing age-reporting strategy inaccuracies with sex as the between-subject factor (Table 16) indicates that there is not a significant difference in method inaccuracies between the sexes. Further, the ANOVA shows that there is not a significant interaction between sex and age-reporting strategies so no further analyses were conducted for sex. Only the experience-based approach meets all the criteria for accuracy and reliability for the female cohort.

Males

The most accurate age-reporting strategies for males (n=31) are LEPS (87.1%), BCAS (83.9%), LEAS (83.9%), HNPS (80.7%). The least accurate strategy is LJAS (29%). All other age-reporting strategies have accuracies ranging from 41.9%-79.3%.
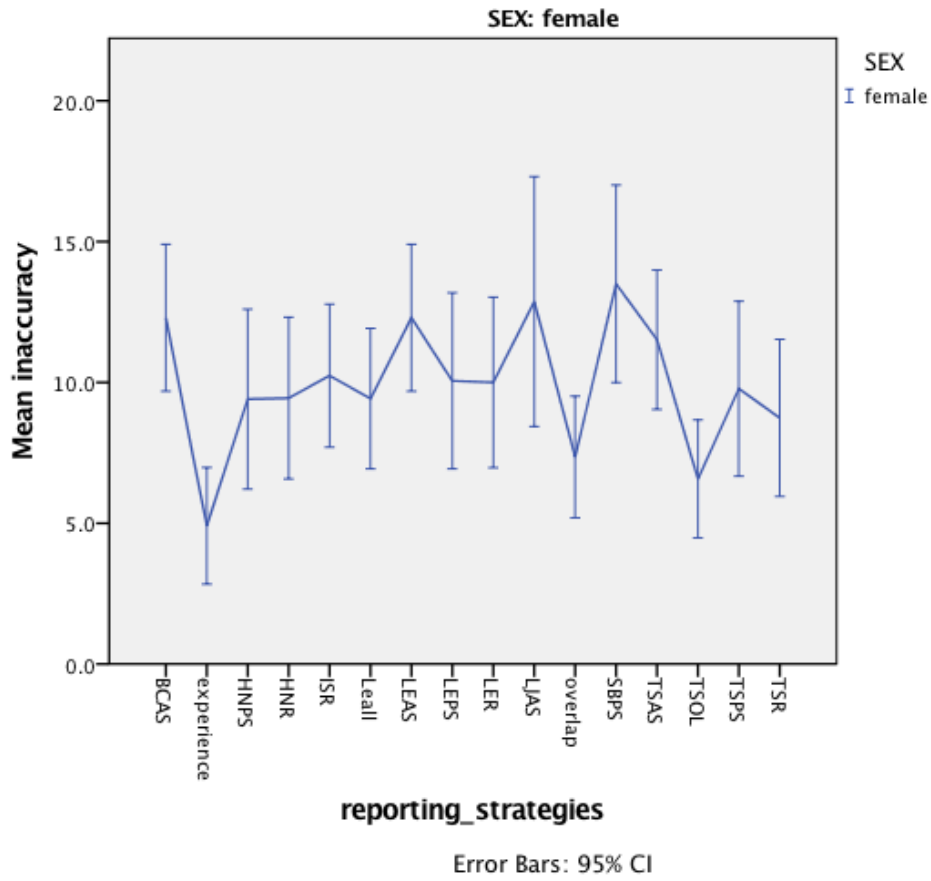
Figure 17: Mean inaccuracies of age-reporting strategies for females.

Figure 18 displays the mean bias for each age-reporting strategy for males. The age-reporting strategies with the least amount of bias ($\bar{x} \geq -1, \bar{x} \leq 1$) are overlap (0.47 years), TSR (0.58 years), TSOL (-0.61 years), and experience (-0.84 years). The age-reporting strategies with the most amount of bias are BCAS/ LEAS with 9.2 years. The univariate ANOVA comparing biases of age-reporting by sex (Table 12) reveals that at least two age-reporting strategies have significantly bias scores for males. The pairwise comparison table shows that BCAS/ LEAS are significantly different ($p<0.05$) from ISR, LEPS, LJAS, and TSPS.

The mean inaccuracies of each age-reporting strategy for males is found in Figure 19. The age-reporting strategies with the lowest mean inaccuracies ($\bar{x}$ <10 years) for males are experience (7.19 years), ISR (8.01 years), TSOL (8.42 years), Leall (8.58 years), overlap (8.86 years), TSR (9.45 years), and HNR (9.82 years). The strategy with the highest mean inaccuracy is TSAS (13.5 years). All other strategies have inaccuracies ranging from 10.02-13.18 years. None of the age-reporting strategies meet both the accuracy and reliability criteria for the male cohort

Interobserver Error Results

As previously mentioned, Cronbach's alpha is a measure of reliability and is used in this study to determine the consistency of phase assignment between 13 observers. The results demonstrate that there is high observer agreement regarding phase assignment, α=.98. Additionally, the correlation matrix confirms that the responses of the 13 observers are highly correlated.
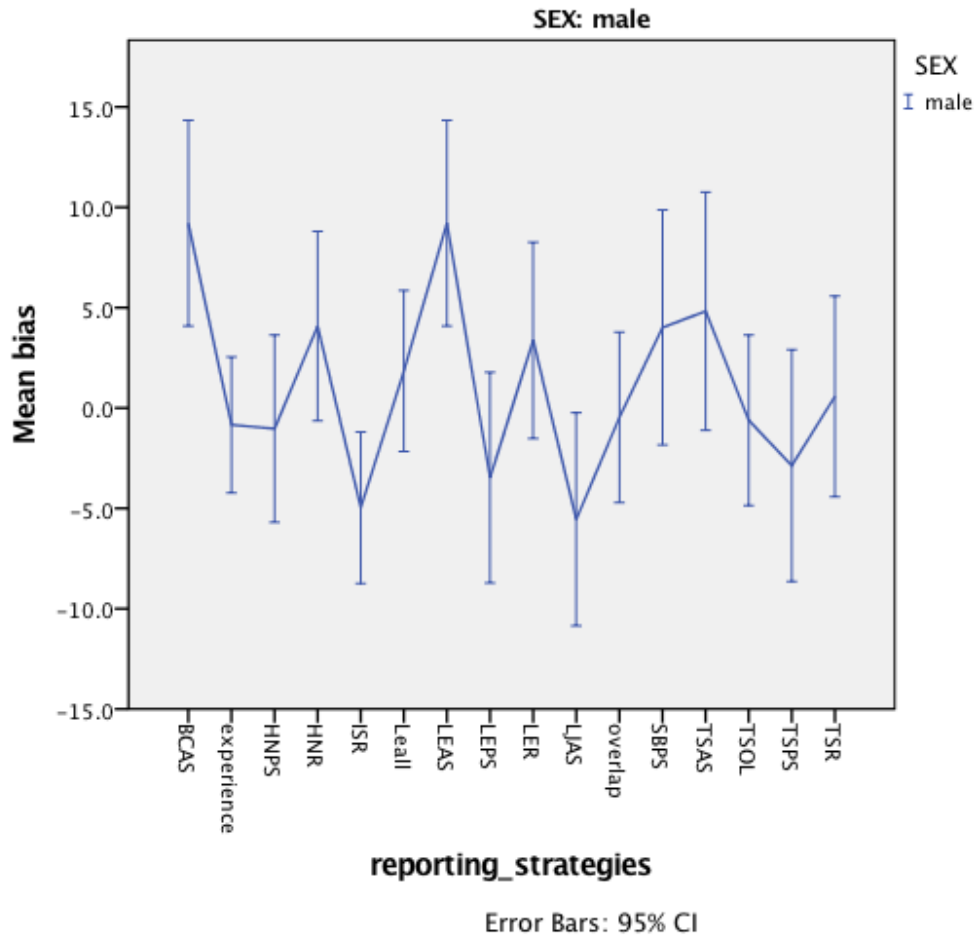
Figure 18: Mean biases of age-reporting strategies for males.
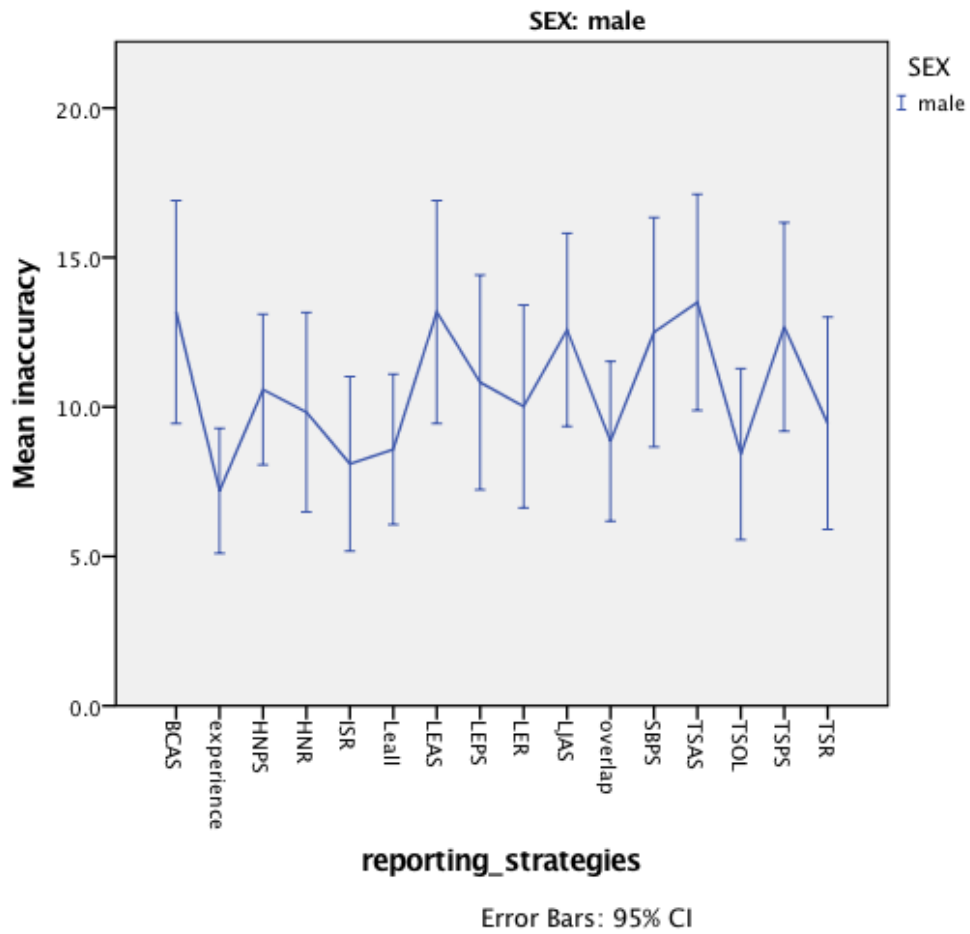
Figure 19: Mean inaccuracies of age-reporting strategies for males.

Observer accuracy (Table 17) as well as method accuracy (Table 18) were evaluated in the interobserver study. The actual age-at-death of Skeleton 1 was 21 years old and Skeleton 2 was 67 years old. Observer accuracy ranges from 0-88% when estimating age of Skeleton 1 and from 13-88% when estimating age of Skeleton 2.

When estimating age of Skeleton 1, Suchey Brooks pubic symphysis (92%), Buckberry Chamberlain auricular surface (92%), and the experienced-based age estimates (85%) are the most accurate. The least accurate strategies are Lovejoy auricular surface (23%), and Hartnett rib (23%). For Skeleton 2, Hartnett pubic symphysis (92%) and BCAS (85%) resulted in the most accurate age estimations while SBPS (15%) and HNR (15%) resulted in the least accurate age estimations. When taking both skeletons into consideration, observers are least accurate when estimating age using the two rib methods and most accurate at estimating age using BCAS and HNPS.

The association between observer experience and age estimation was also considered within this study. This relationship was assessed by conducting a Chi-Square test and evaluating the results of the Cross-tabs table. Additionally, the proportion of attempted and correct responses were graphed using SPSS (Figures 20 and 21). The Chi-Square for Skeleton 1 indicates that there is a significant association between accurate age estimations and observer experience level X^2(2, N=102) =8.01, p<0.05. As such, each cohort was directly compared to each other using a follow-up Chi-Square.

Table 17: Interobserver study: Observer accuracy.

Skeleton ID	observer 1	observer 2	observer 3	observer 4	observer 5
Skeleton 1	75%	88%	88%	63%	75%
Skeleton 2	38%	38%	63%	50%	25%
	observer 6	observer 7	observer 8	observer 9	observer 10
Skeleton 1	50%	75%	50%	50%	0%
Skeleton 2	75%	63%	13%	25%	88%
	observer 11	observer 12	observer 13		
Skeleton 1	50%	63%	75%		
Skeleton 2	50%	75%	88%		

Table 18: Interobserver study: Method accuracy.

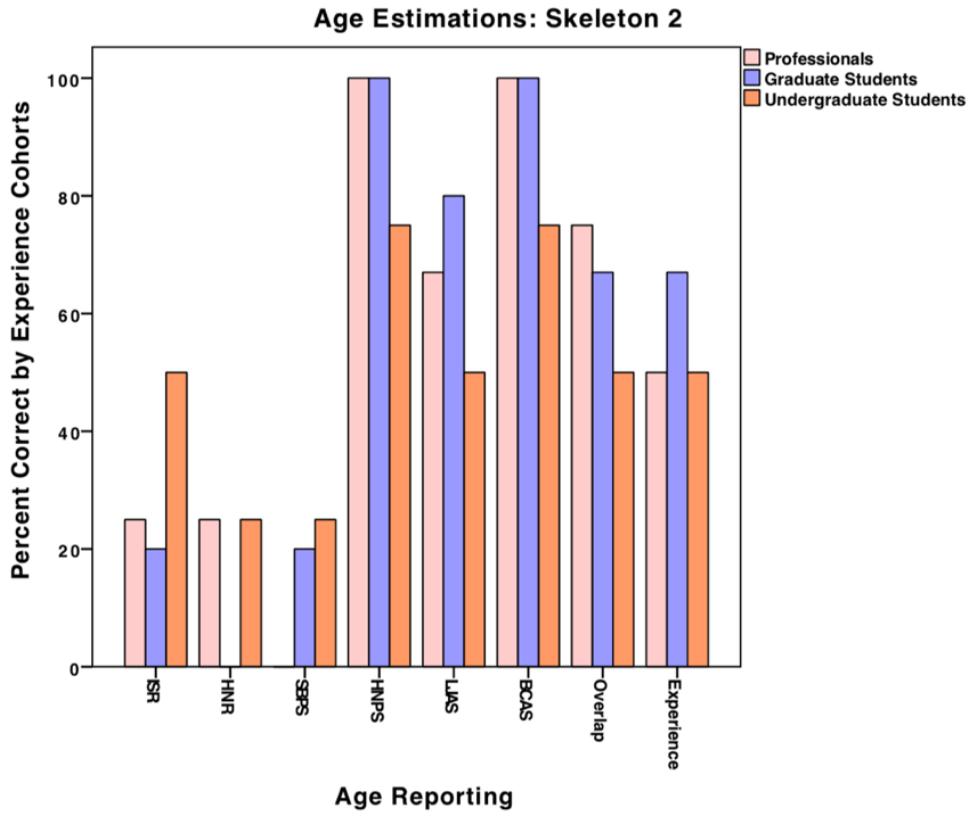| Skeleton ID	ISR	HNR	SBPS	HNPS
Skeleton 1	31%	23%	92%	77%
Skeleton 2	31%	15%	15%	92%
	LJAS	BCAS	Overlap	experience
Skeleton 1	23%	92%	85%	69%
Skeleton 2	62%	85%	69%	54%

Figure 20: Bar chart illustrating the role of experience on age estimation-Skeleton 1.
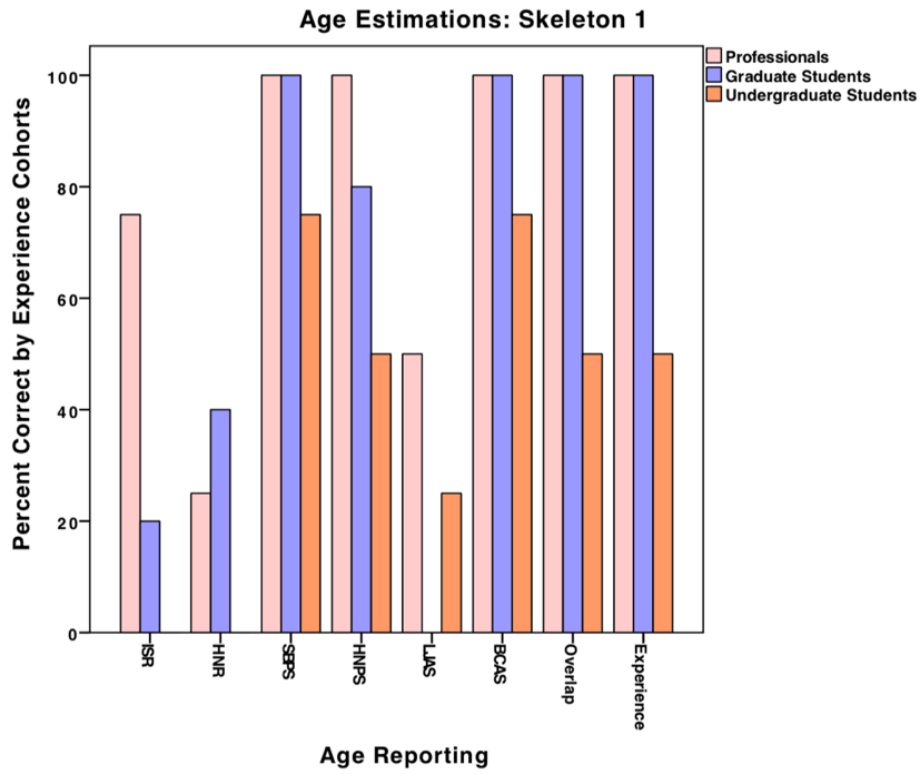
Figure 21: Bar chart illustrating the role of experience on age estimation-Skeleton 2.

These results show no association between correct responses and experience level when comparing the professionals and graduate students X^2 (1, N=74) =4.00, p=0.75 nor between graduate students and undergraduate students X^2 (1, N=70) =1.61, p=0.33. However, there is a significant association between correct estimations and experience level when comparing professionals and undergraduate students X^2 (1, N=60) =7.96, p<0.05.

The Chi-Square for skeleton 2 revealed that there was no association between correct age estimation and experience level X^2 (2, N=98) =0.21, p=.94. Additionally, the Crosstab table shows that overall accuracy of each cohort is similar (professionals=53.3%, graduate students=55.6%, and undergraduate students= 50%). While not significantly different, professionals and graduate students are proportionately more accurate at estimating age than undergraduate students for Skeleton 2 (Figure 21).

Intraobserver Error Results

To assess internal reliability an intraobserver error study was conducted using a paired-samples t-test. The data from the initial observations will be referred to as Observation 1 and the data from the intraobserver sample will be referred to as Observation 2.

There is not a significant difference in the means of the six aging methods between Observation 1 (M=53.63, SD=13.16) and Observation 2 (M=53.65, SD=13.05); t(59)=-0.02,p=0.99. The is a strong correlation between the two observations (r=0.80, p=0.00). Similarly, there is no significant difference in the scores of the ten additional

reporting strategies between Observation 1 (M=55.52, SD=13.61) and Observation 2 (M=55.75, SD=13.35); t(99)=-0.24, p=0.81. There is also a strong positive correlation between the observations for these strategies (r=0.75, p=0.00).

Comparing the individual methods from Observation 1 and Observation 2 illustrates that none of the methods differed significantly (p>0.32). Additionally, five of the six methods display strong, positive correlations (r>.74, p<.02). Only the Hartnett rib method has a moderate correlation (r=.63, p=.05) between the two observations. The results are similar when comparing the ten other age-reporting strategies to one another. None of the age-reporting strategies are significantly different between Observation 1 and Observation 2 (p>.07). However, there are more differences in the correlations of age-reporting strategies between the two observations. LEA, LEPS, LEAS, TSPS, Overlap, TSOL and Experience all have strong, positive correlations (r>0.71, p<0.02). LER and TSR only have moderate, positive correlations between Observation 1 and Observation 2 (r=0.63, p=0.05). The TSAS strategy has a weak, positive correlation between the two observations (r=0.41, p=0.24).

CHAPTER FIVE
DISCUSSION

This study assessed the accuracy and reliability of different age-reporting strategies. The age-reporting strategies included in this study were selected based on the current practices of forensic anthropologists (Garvin and Passalacqua 2012) and suggestions made by scholars within the field (Baccino et al. 1999; Merritt 2013; Nawrocki 2010). Sixteen final age estimations were derived from different age-reporting strategies. There are three main hypotheses for this study. Hypothesis 1 states that the experience-based approach will produce the most accurate and reliable age-at-death estimations overall. If this hypothesis is accepted, it supports using one's professional judgement when producing final age estimations in a forensic setting. Hypothesis 2 states that the most accurate and reliable age-reporting strategy will differ when the sample is divided by age. If there is a difference when divided by age, it is further hypothesized that the two-step strategies will be the most accurate and reliable for the young and old-age cohorts. If hypothesis 2 is accepted, there is support for choosing methods based on preliminary assessments of morphological age. Hypotheses 3 states that the most accurate and reliable age-reporting strategy will differ when the sample is divided by sex. If there is a difference based on sex it is hypothesized that estimations derived from sex-specific results will be the most accurate and reliable. Acceptance of hypothesis 3 would support choosing aging methods based on the sex of the skeleton.

The results of this study reveal that only a few of the age-reporting strategies tested in this study were both accurate and reliable for any single sample category (total sample, young cohort, males, etc.) and none of the strategies were both accurate and

reliable for all categories. It is important to reiterate that the thresholds for accuracy and reliability used in this study were determined by the author. While the parameters set for this study were thought to be rather rigorous, 80% accuracy still leaves room for 20% of all cases to be estimated incorrectly. However, if 90% was set as the accuracy threshold, most of the strategies would not have met this standard for the sample used in this study. This indicates that age-reporting strategies must be further scrutinized for their ability to produce accurate estimations. Additionally, standards for comparing aging methods/ age-reporting strategies should be defined.

Intraobserver Error

The results from the intraobserver study indicate that the observer was consistent in estimating age for the 10 skeletons during two separate observation periods. Further, none of the individual age-reporting strategies were significantly different between Observation 1 and Observation 2. These results support the ability of the researcher to consistently recognize and categorize age-related skeletal morphology.

Performance by Total Sample

Age estimations derived from the Buckberry-Chamberlain (BCAS)/Least Error Auricular Surface (LEAS), Suchey-Brooks (SBPS) and Least Error Pubic Symphysis (LEPS) were the most accurate estimations for the total sample followed by experience-based strategy. BCAS/LEAS and SBPS were accurate but failed to meet either of the reliability criteria for this study. Additionally, BCAS and SBPS both provide age ranges

that averaged over 30 years. As previously discussed, methods that are accurate but unreliable due to large age ranges are not helpful in a forensic setting. The LEPS strategy was accurate and had minimal bias, but just missed the threshold for low inaccuracy (10.47 years). The experience-based approach was the only strategy to meet the accuracy and reliability criteria for the total sample. For the experience-based approach, the analyst was able to consider multiple lines of evidence including multiple method results and overall condition of the skeleton. This contributed to a greater approximation of age as more of the skeletal variation could be captured. Further, the experience-based approach inaccuracy was significantly different from BCAS/ LEAS, Lovejoy (LJAS), SBPS, Two-Step Auricular Surface (TSAS) and Two-Step Pubic Symphysis (TSPS) meaning it performed significantly better than these methods with respect to inaccuracy.

LJAS was the least accurate and least reliable age-reporting strategy for the total sample. It is not surprising that LJAS was not accurate as its ranges are small, about 5-10 years each, and not overlapping (i.e. 25-29, 30-34, etc.). The restrictiveness of the age ranges leads to a greater chance of not including the decedents actual age of death. Additionally, the poor performance of reliability was also expected as many scholars have recognized the LJAS method for its propensity to overestimate the age of younger individuals and underestimate the age older individuals (Bedford ME 1993; Merritt 2013; Murray and Murray 1991; Osborne 2004; Saunders et al. 1992; Schmitt 2004). This study supports the findings of previous studies as LJAS performed significantly worse than all methods except Işcan ribs (ISR) with regards to bias and

worse than experience with regards to inaccuracy. The results of this study support the argument that Lovejoy auricular surface is not an adequate method for estimating age of skeletal remains in a forensic context (Murray and Murray 1991; Osborne 2004). Even if the final age is not reported using the Lovejoy auricular surface result, inclusion in the evaluation of age can negatively impact the analysts final age estimation since it has the tendency to drastically underestimate age.

When estimating age without considering the relative age (young/old) or sex of a skeleton, Least Error Pubic Symphysis (LEPS) and the experience-based approach provided the most accurate and reliable age-at-death estimations and should be considered above all other age-reporting strategies in forensic contexts. Because LEPS provides a final range from an established aging method (SBPS or HNPS), the error can be calculated in the final report presented to law enforcement. Further, LEPS may be preferable in forensic investigations as it meets the *Daubert (1993)* standard and is more likely to hold up in a *Daubert* challenge than an age-estimation that was constructed from experience.

Performance by Age Cohort

The accuracy and reliability of age-reporting strategies varied by age-cohort in this study. SBPS was the most accurate and reliable age-reporting strategy for the young age-cohort. The superior performance Suchey-Brooks method is likely due to its development on a young sample cohort (1990). These results compare with those found by Martrille et al. (2007) who conclude that SBPS was the most accurate method for aging individuals in their young age-cohort (25-40 years old). Merritt (2013) also shows

104

that SBPS had the lowest bias scores and among the lowest inaccuracies for the young age-cohort (20-39 years old) in her study. However, the results of this study partially conflict with those obtained by Saunders et al. (1992). Saunders and colleagues found that SBPS performed poorly in all age categories excluding their 30-39 age range. While 30-39 corresponds to the young cohort in this study (20-39 years old) and is congruent with the results of this study, SBPS performed poorly in the 17-29 age category of the Saunders et al. study, which encompasses the first half of our young cohort, representing conflicting results. Other scholars have also recognized SBPS for its superior performance in estimating the age of young adults which has led to its inclusion, as the "young" option, in two-step strategies (Baccino et al. 1999; Martrille et al. 2007; Merritt 2013). This research evaluates the two-step approach by skeletal indicator. While the Two-Step Pubic Symphysis (TSPS) had high accuracy (83.3%) and low inaccuracy (8.44 years) for the young cohort, its bias was high (8.22 years). The high positive bias demonstrates that age-estimations derived from the TSPS tended to overestimate age. LEPS and overlap were also accurate for the young cohort but neither met the criteria for reliability. Conversely, the Two-Step Rib (TSR) was reliable for the young cohort but the accuracy was poor (66.7%). The least accurate strategy for the young cohort was LJAS and the least reliable strategies were BCAS/LEAS. Not only were BCAS/LEAS the least reliable strategies, their inaccuracy and bias scores were significantly worse than SBPS and TSR. It is also interesting to note that twelve of the sixteen age estimations had low inaccuracies for the young cohort. This provides further

evidence that there is a stronger correlation between chronological and skeletal age in the younger years of life, even into adulthood (Nawrocki 2010; Osborne 2004).

The experience-based approach performed the best for the middle-age cohort as its results were both accurate and reliable for individuals between 40-59 years old. As previously mentioned, individual aging indicators are limited in their ability to adequately capture the variation in skeletal indicators as one ages (Nawrocki 2010). Because experience-based estimations are generated using multiple lines of evidence, they are able to capture the skeletal variation of middle adulthood better than estimations that were derived from a single method's results. The reliability results for the middle cohort showed that the experience-based inaccuracies were significantly lower than Two-Step Auricular Surface (TSAS) and SBPS. SBPS actually achieved the most accurate results for the middle-age cohort but had poor reliability. Additionally, the overlap approach was shown to be reliable but had poor accuracy. The overlap approach for this study involved constructing an age range that overlapped all six ranges produced by the aging methods. It is likely that this methodology produced poor accuracy results due to its inclusion of the LJAS range, which was significantly more inaccurate than all strategies. As with the young cohort, the least accurate strategy for the middle-age cohort was LJAS and the least reliable strategies were BCAS/LEAS.

None of the age-reporting strategies were both accurate and reliable for the old-age cohort. This finding supports the assertion that aging indicators become less accurate and reliable as chronological and skeletal age become less correlated (Nawrocki 2010). BCAS/LEAS met the criteria for accuracy (100%) and inaccuracy

(7.85 years) but not bias (-5.25 years). Therefore, BCAS is considered the most accurate and reliable strategy for the old-age cohort. These results correspond with Merritt (2013) results which showed that BCAS was the most accurate and reliable of all the methods for the 60+ age category. It should be noted that five of the six BCAS ranges extend into the sixth decade of life and average 50 years, which contributes to its accuracy.

The results of this study also showed that age-reporting biases were significantly different between the old-age cohort and the other two age-cohorts in this study. This is likely due to the fact that most of the age-reporting strategies had a tendency to underestimate age for the old-age cohort and overestimate age for both the young and middle-age cohorts. LJAS was the least accurate and reliable strategy for the old-age cohort, with significantly worse bias scores than all methods excluding LEPS, and significantly higher inaccuracy scores than all methods except SBPS.

A few trends were recognized when evaluating the age-reporting strategies by age-cohorts. Regardless of the age of the skeleton being evaluated, LJAS is more likely to provide an incorrect estimation than a correct estimation. Again, this calls into question the continued use of LJAS in forensic anthropology. The most accurate and reliable age-estimation strategies for the young and old-age cohorts were the results of an individual aging method but the most accurate and reliable age-estimation strategy for the middle-age cohort was the experience-based approach. This suggests that individual aging indicators alone may not be able to adequately capture the variation of the middle decades of life.

The purpose of a two-step strategy is to choose the method that is most appropriate for the skeleton being evaluated. If the skeleton is likely younger, then methods that were developed with a younger sample composition should be used and if the skeleton is likely older, methods developed with an older sample composition should be used. A surprising finding of this study was that the two-step approach did not provide the most accurate and reliable age estimations for the young and old-age cohorts. In this research, the two-step strategy was used to choose between two methods that were applied to the same skeletal indicator. This is different from previous studies which use the two-step strategy to choose between two methods applied to different skeletal indicators (Baccino et al. 2014; Baccino et al. 1999; Martrille et al. 2007). Perhaps the two-step strategy would have been more successful in this study if it was used to choose between methods evaluating different skeletal indicators. Future studies should further explore the utility of two-step estimations.

The hypothesis that the most accurate and reliable age estimations will vary when the sample is divided by age is partially accepted. Experience was the most accurate and reliable age-reporting strategy for the total sample and for the middle-age cohort, but not for the young and old-age cohorts. So, while the most accurate and reliable strategy was consistent between the total sample and the middle-age cohort, it differed between the total sample and the young and old-age cohorts. Because none of the two-step strategies were the most accurate and reliable for the young or old-age cohort, hypothesis 2a is rejected.

Performance by Sex

The most accurate and reliable age-reporting strategies differed between males and females. The only age-reporting that was both accurate and reliable for the female cohort was the experience-based approach with 88.9% accuracy, 4.91 years inaccuracy, and -0.39 years bias. The Two-Step Auricular Surface (TSAS) also had a high accuracy (81.5%) and low bias (0.19 years) but did not meet the criteria for inaccuracy with 11.52 years. LJAS was the least accurate strategy and had the highest bias scores. SBPS was the strategy with the highest inaccuracy for females.

None of the age-reporting strategies were both accurate and reliable for males, however, the results from HNPS came close to meeting all the criteria for accuracy and reliability (accuracy=85%; inaccuracy= 10.58 years; bias= -1.03 years). Therefore, HNPS is considered the most accurate and reliable strategy for males. The Hartnett pubic symphysis method was developed on a sample that included over 400 males which may explain its ability to provide accurate and reliable estimations for the males in this study (Hartnett 2010a). Consistent with previous results, LJAS was the least accurate age-reporting strategy. BCAS/LEAS were the least reliable age-reporting strategies for males.

The hypothesis that the most accurate and reliable age-reporting strategy will differ by sex is partially accepted. While the most accurate and reliable age-reporting strategy differed between males and total sample, experience was the most accurate and reliable strategy for the total sample and for the female cohort. Additionally, hypothesis 3a is also partially accepted. Neither of the best performing strategies for

females (experience and TSAS) had estimations that were derived directly from sex-specific methods. In contrast, the best performing strategy for males (HNPS) provides an age-estimation that was developed with sex-specific standards. It was unexpected that sex-specific methods did not provide the most accurate and reliable estimations for both males and females. Many scholars have argued that males undergo a more consistent and predictable trajectory of aging than females (Gilbert 1973; Gilbert and McKern 1973; Klepinger et al. 1992; Suchey 1979). Perhaps this consistency of aging contributed to the ability of the HNPS to better capture age-related changes in the male pubic symphysis. The stated goals of the Hartnett (2010a) study were to clarify confusing language and to improve upon age estimations derived from the pubic symphysis. The results of this study indicate that Hartnett was successful in meeting this goal, particularly for males. Because female pubic symphyses do not age as consistently as males, more variation has to be accounted for in method phases and descriptions. As with the old age cohort, the age estimation derived using experience provided the most accurate and reliable estimate for the female cohort. This suggests that experience plays a positive role with age estimation when a high degree of variation is expected within an aging indicator or from a specific sample category.

Overall Trends

While none of the strategies were accurate for all sample categories, Buckberry-Chamberlain (BCAS/ Least Error Auricular Surface (LEAS) were accurate for all except the young cohort. Merritt (2013) also found BCAS to be the most accurate of methods

she tested with 100% of all individuals being correctly aged with this method. Least Error Pubic Symphysis (LEPS) was accurate for all except the female cohort.

While BCAS/LEAS had the highest accuracies across all sample categories, the paired samples *t*-test revealed that the average mean of these strategies were significantly different from the mean actual age of the skeletons. Further, these strategies had high inaccuracies and biases for all sample categories. This shows that the results from the BCAS aging method are accurate, but not reliable. As previously mentioned, accuracy is important in age-estimation so that the age of the individual is not erroneously excluded, however, wide age ranges reduce the probative value of age estimations. The average phase range for BCAS is 39 years with many of the ranges encompassing the majority of adulthood (e.g. 16-65 years old). Therefore, while the results from BCAS are likely to include a decedents actual age at death, they are not beneficial for narrowing down potential matches in a forensic context.

Besides BCAS/LEAS, Lovejoy (LJAS) was the only other age-reporting strategy where mean estimated age was significantly different from the mean actual age of the skeletons. Further, the age estimations derived from LJAS were neither accurate nor reliable for any age categories. In fact, LJAS had the poorest accuracy results of all the strategies tested in this study, with correct estimations ranging from only 11.1%-29%. Comparable to these results, Martrille et al. (2007) found that LJAS was the least accurate of the methods they tested when all ages were pooled. Merritt (2013) also found that Lovejoy was one of the least correct original methods that she tested. While Saunders and colleagues (1992) obtained good accuracies with the Lovejoy method,

they found it to have high levels of bias and conclude that it becomes less reliable for individuals past the third decade of life. This study also found LJAS to have sizable levels of bias, underestimating age in all sample cohorts excluding the 20-39 age cohort.

Age estimations that do not include the decedents age-at-death can greatly hinder the potential of positive identification. Not only was LJAS not accurate, it was also largely unreliable. Therefore, reporting an age-range based on the Lovejoy method in a forensic context is irresponsible as the estimation is likely to be incorrect.

Similarly, to the accuracy results, none of the age-reporting strategies were reliable (i.e. low inaccuracy *and* bias scores) for all sample categories. The experience-based approach did present low inaccuracy scores for all sample categories, but higher bias scores for the young and old-age cohorts. While the experience-based strategy did not meet the accuracy criteria (80%) for all of the sample categories, its accuracies were above 70% for all groups. These results are consistent with the results of the Baccino et al. (1999) study as the two observers both achieved high accuracies using the "global approach." Additionally, Parsons (2017) found that age estimations documented in resolved case reports were 92% accurate and contributed this success to practitioners' reliance on multiple methods. The experience-based strategy was also the most accurate and reliable for total sample with an accuracy of 79.5%, inaccuracy of 6.13 years, and bias of -0.63. Therefore, the hypothesis that final age estimations derived from the experience-based approach will be the most accurate and reliable overall, is accepted.

Interobserver Error

When assigning skeletal indicators to method phases, observers were highly consistent with one another as evident by the high Cronbach's alpha rating and correlations. These results indicate that observers are responding similarly when choosing phases. Observer 10 had lower correlations than the rest of the observers, suggesting that this individual's responses were less consistent with the other participants.

The accuracies of the age-reporting strategies differed between the two skeletons in the interobserver study. For the young, female individual (Skeleton 1), Suchey-Brooks (SBPS) was the most accurate strategy and Lovejoy (LJAS) and Hartnett rib (HNR) were the least accurate strategies. These results are similar to the results obtained in the larger study when comparing the age-reporting strategies by age-cohort. SBPS was the most accurate for the young cohort and LJAS was the least accurate. HNR also had low accuracies for the young cohort. In contrast, the intraobserver results for the young female are not similar to the results of the larger study when comparing the age-reporting strategies by sex. In the larger study, the most accurate strategy for females was the experience-based approach.

HNPS was the most accurate age-estimation strategy for the older, male individual (Skeleton 2) and SBPS and HNR were the least accurate. These results are not consistent with those obtained when comparing strategies by age-cohorts in the larger study. Rather, the most accurate strategy for the old-age cohort was Buckberry-Chamberlain (BCAS) and the least accurate was LJAS.

The results of the interboserver study demonstrate that observers were more accurate estimating age using the pubic symphysis for both of the skeletons. Many of the participants indicated a greater familiarity with and preference for pubic symphysis aging methods which may have contributed to their success using these methods.

As far as observer accuracy is concerned, performance ranged from 0-88% for Skeleton 1 and 13-88% for Skeleton 2. The majority of the observers (8/13) were more successful when estimating the age of Skeleton 1. Observers 8, 9, and 11 were the least successful with estimating the age of both skeletons. None of these observers were professionals and two were undergraduate students. The poor accuracy of these participants is likely due to inexperience with the age-at-death estimation methods included within this study and/or unfamiliarity with the human age variation. The greater accuracy with estimating the age of Skeleton 1 could be due to its age since younger individuals tend to have skeletal morphology more consistent with their chronological age (Nawrocki 2010).

The result comparing the effects of observer experience indicates that professionals were more likely to estimate age correctly for Skeleton 1 than undergraduate students. Although the result comparing the graduate students to the other two groups were not significant, graduate students had method accuracies more similar to professionals, especially when estimating age of the younger individual. Therefore, professionals and graduate students were more successful at estimating age than undergraduate students for Skeleton 1. None of the experience-cohorts were significantly more likely than the others to estimate age correctly for Skeleton 2.

114

However, professionals and graduate students were generally more successful at estimating the age of Skeleton 2 than undergraduate students. The fact that professionals were significantly better at estimating the age of Skeleton 1 than undergraduate students lends support to the argument that experience plays a role in one's ability to estimate age accurately.

CHAPTER SIX
CONCLUSION

This study assessed different age-reporting strategies for their accuracy and reliability. This research is valuable as it provides forensic anthropologist with insights regarding the efficacy of some of the strategies currently used to produce final age-at-death estimations. The results of this study show that the most accurate and reliable age-reporting strategy varied if the sample was evaluated as a whole, by age, or by sex. While none of the strategies were consistently the most accurate and reliable for all of the sample categories, the experience-based approach performed well in each category. The experience-based strategy allowed the researcher to use the results of the individual aging methods and professional judgment to arrive at a final age estimation. While age estimations derived from experience do not meet the *Daubert* criteria they can provide better approximations of age since they are based on the results of multiple aging indicators. The purpose of a biological profile is to narrow down potential identifications for eventual positive identifications. In this effort, it is more important to provide an age-at-death estimation that takes into account the results from accurate and reliable methods as well as the analyst's expert judgement.

The results of this study also call into question the value of auricular surface age-reporting strategies derived from Lovejoy and Buckberry-Chamberlain auricular surface methods. Both Lovejoy and Buckberry-Chamberlain provided age estimations that were significantly different than the actual age of the skeleton. Further, the estimations produced using the auricular surface ranges were either too large to provide exclusionary power in a forensic case or did not produce accurate age-at-death

estimations. It is recommended that auricular surface aging be avoided in forensic casework as it is more likely to do harm than good.

There are several limitations of this study that should be noted. First, only three age indicators were evaluated in this study, despite the availability of aging methods focused on other regions of the skeleton such as the teeth (Lamendin et al. 1992), cranial sutures (Meindl and Lovejoy 1985), acetabulum (San-Millan et al. 2017) and sacrum (Passalacqua 2009). Understanding the accuracy and reliability of reporting strategies that produced estimations from multiple areas of the skeleton is beneficial for deciding how to report age if certain elements are not recovered in a forensic situation. To address this, future studies evaluating reporting strategies should diversify and/or expand the number of methods included. Secondly, this study did not include transition analysis as an age-reporting strategy, despite its ability to combine aging indicators in a way that is statistically valid. Transition analysis was excluded from this study because it is not widely used in forensic practice (Garvin and Passalacqua 2012; Parsons 2017). However, it has been shown to perform well in validation studies (Milner and Boldsen 2012) and has been included in the most updated version of the University of Tennessee's Data Collection Procedures manual (2016). It is possible that transition analysis can provide accurate and reliable age estimations that also meet the *Daubert* standards. Finally, there were major limitations associated with the interobserver error component of this study, specifically the sample size of the skeletons and the experience-cohorts. Because only two skeletons were evaluated in the interobserver study, it is not appropriate to draw definitive conclusions regarding strategy accuracy by

117

sex and age-cohort. Additionally, the results assessing the effects observer experience level on age estimation should be considered cautiously as each experience-cohort was only represented by four or five observers. However, since the results showed that experience level may play a role in one's ability to accurately estimate age-at-death, future research should specifically explore this assertion.

Future studies evaluating age-reporting strategies for forensic contexts could involve a greater representation of age-at-death methods, different two-step strategies, and a stricter evaluation of the role of experience in producing final age estimations. Specifically, future research designs should incorporate methods that estimate age using indicators for the skull and teeth (Lamendin et al. 1992; Meindl and Lovejoy 1985), include transition analysis as an aging method and age-reporting strategy (Milner and Boldsen 2012), and test variations of the two-step strategy described by Baccino and colleagues (2014; 1999). Finally, future research designs should include at least one junior and one senior observer in order to better assess how experience levels may affect age-at-death estimations.

It is crucial to understand how to report age-at-death in a manner that is both accurate and reliable. This study was able to shed light on the performance of different age-reporting strategies and provide further support to the reliance on multiple aging indicators in developing a final age estimation. Ultimately, many factors contribute to how final age estimations are produced, all of which cannot be included within a single research design. Therefore, studies like this one can help with the pursuit of better age-at-death estimations, and ultimately more identifications of unknown skeletal remains.

LIST OF REFERENCES

1993. Daubert v. Merrell Dow Pharmaceuticals, Inc. Circuit USCoAftN. p 579-601.

Baccino E, Sinfield L, Colomb S, Baum TP, and Martrille L. 2014. Technical note: The two step procedure (TSP) for the determination of age at death of adult human remains in forensic cases. Forensic Sci Int 244:247-251.

Baccino E, Ubelaker DH, Hayek LC, and Zerilli A. 1999. Evaluation of Seven Methods of Estimating Age at Death from Mature Human Skeletal Remains. J Forensic Sci 44(5):931-936.

Bedford ME RK, Lovejoy CO, Meindl RS, Simpson SW, Stuart-MacAdam PL. 1993. Test of the Multifactorial Aging Method Using Skeletons With Known Ages-at-Death From the Grant Collection. American Journal of Physical Anthropology 91:10.

Bonett DG, and Wright TA. 2015. Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. Journal of Organizational Behavior 36(1):3-15.

Brooks S, and Suchey JM. 1990. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. Human Evolution 5(3):227-238.

Buckberry JL, and Chamberlain AT. 2002. Age estimation from the auricular surface of the ilium: a revised method. Am J Phys Anthropol 119(3):231-239.

Charles A. Kunos, Scott W. SImpson, Katherine F. Russell, and Hershkovitz I. 1999. First Rib Metamorphosis: Its Possible Utility for Human Age-at-Death Estimation. Am J Phys Anthropol 110:303-323.

DiGangi EA, Bethard JD, Kimmerle EH, and Konigsberg LW. 2009. A new method for estimating age-at-death from the first rib. Am J Phys Anthropol 138(2):164-176.

Dwight T. 1878. The Identification of the Human Skeleton: A Medico Legal Study Boston, D. Clapp.

Emden Hv. 2008. Statistics for terrified biologists: Blackwell Publishing.

ER. K, and DH. U. 1978. Revisions in the microscopic method of estimating age at death in human cortical bone. Am J Phys Anthropol 49:545-546.

Franklin D. 2010. Forensic age estimation in human skeletal remains: current concepts and future directions. Leg Med (Tokyo) 12(1):1-7.

Freemont AJ, and Hoyland JA. 2007. Morphology, mechanisms and pathology of musculoskeletal ageing. J Pathol 211(2):252-259.

Garvin HM, and Passalacqua NV. 2012. Current practices by forensic anthropologists in adult skeletal age estimation. J Forensic Sci 57(2):427-433.

Gilbert M. 1973. Misapplication to Females of the Standard for Aging the Male *Os pubis*. Am J Phys Anthropol(38):19-40.

Gilbert M, and McKern T. 1973. A Method for Aging the Female *Os pubis*. Am J Phys Anthropol 38:31-38.

Girden E. 1992. Anova.

Goforth C. 2015. Using and Interpreting Cronbach's Alpha.

Hartnett KM. 2010a. Analysis of age-at-death estimation using data from a new, modern autopsy sample--part I: pubic bone. J Forensic Sci 55(5):1145-1151.

Hartnett KM. 2010b. Analysis of age-at-death estimation using data from a new, modern autopsy sample--part II: sternal end of the fourth rib. J Forensic Sci 55(5):1152-1156.

IBM. 2016. SPSS Statistics for Macintosh, Version 24.0. Armonk, NY: IBM Corp. .

Işcan MY, Loth SR, and Wright RK. 1984a. Age Estimation from the Rib by Phase Analysis: White Males. J Forensic Sci 29(4):1094-1104.

Işcan MY, Loth SR, and Wright RK. 1984b. Metamorphosis at the Sternal Rib End: A New Method to Estimate Age at Death in White Males. Am J Phys Anthropol 65:147-156.

Işcan MY, Loth SR, and Wright RK. 1985. Age Estimation from the Rib by Phase Analysis: White Females. Journal of Forensic Sciences 30(3):853-863.

Katz D, and Suchey JM. 1986. Age determination of the male os pubis. Am J Phys Anthropol 69:427-435.

Klepinger LL, Katz D, Micozzi MS, and Carroll L. 1992. Evaluation of Cast Methods for Estimating Age from the *Os Pubis*. J Forensic Sci 37(3):763-770.

Krogman M. 1939. A Guide to the Identification of Human Skeletal Material. FBI Law Enforcement Bulletin 8(8):1-29.

Lamendin H, Baccino E, Hubert F, Tavernier JC, Nossintchouk RM, and Zerilli A. 1992. A Simple Technique for Age Estimation in Adult Corpses: The Two Criteria Dental Method. J Forensic Sci 37(5):1373-1379.

Lovejoy CO, Meindl RS, Mensforth RP, and Barton TJ. 1985a. Multifactorial Determination of Skeletal Age at Death: A Method and Blind Tests of Its Accuracy. Am J Phys Anthropol 68:1-14.

Lovejoy CO, Meindl RS, Pryzbeck TR, and Mensforth RP. 1985b. Chronological Metamorphosis of the Auricular Surface of the Ilium: A New Method for the Determination of Adult Skeletal Age at Death. Am J Phys Anthropol 68:15-28.

Martrille L, Ubelaker DH, Cattaneo C, Seguret F, Tremblay M, and Baccino E. 2007. Comparison of four skeletal methods for the estimation of age at death on white and black adults. J Forensic Sci 52(2):302-307.

McKern T, and Stewart TD. 1957. Skeletal Age changes in young American Males. In: Command USAHQRaD, editor. Natick, Massachusetts.

Meindl R, Lovejoy, C., Mensforth, R., & Walker, R. 1985. A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. American Journal of Physical Anthropology 68(1):29-45.

Meindl RS, and Lovejoy CO. 1985. Ectocranial Suture Closure: A Revised Method for the Determination of Skeletal Age at Death Based on the Lateral-Anterior Sutures. Am J Phys Anthropol 68:57-66.

Merritt CE. 2013. Testing the accuracy of adult skeletal age estimation methods: original methods versus revised and newer methods. Explorations in Anthropology 12(1):102-119.

Milner GR, and Boldsen JL. 2012. Transition analysis: a validation study with known-age modern American skeletons. Am J Phys Anthropol 148(1):98-110.

Milner GR, Boldsen JL, Ousley SD, Getz SM, Weise S, and Tarp P. 2016. Estimating Age from Adult Skeletons: New Directions in Transition Analysis Using a Wide Array of Traits. American Association of Physicla Anthorpologists. Atlanta Georgia

Miranker M. 2016. A Comparison of Different Age Estimation Methods of the Adult Pelvis. J Forensic Sci 61(5):1173-1179.

Murray KA, and Murray T. 1991. A Test of the Auricular Surface Aging Technique. J Forensic Sci 36(4):1162-1169.

Nawrocki S. 2010. The Nature and Sources of Error in the Estimation of Age at Death from the Skeleton. Age Estimation of the Human Skeleton Charles C Thomas. p 79-101

NIJ. 2017. Scientific Working Groups.

Osborne DLS, Tal L.; Nawrocki, Stephen P. . 2004. Reconsidering the Auricular Surface as an Indicator of Age at Death*. J Forensic Sci 49(5).

Parsons HR. 2017. THE ACCURACY OF THE BIOLOGICAL PROFILE IN CASEWORK: AN ANALYSIS OF FORENSIC ANTHROPOLOGY REPORTS IN THREE MEDICAL EXAMINERS' OFFICES: Univeristy of Tennessee.

Passalacqua NV. 2009. Forensic age-at-death estimation from the human sacrum. J Forensic Sci 54(2):255-262.

Rouge-Maillart C, Jousset N, Vielle B, Gaudin A, and Telmon N. 2007. Contribution of the study of acetabulum for the estimation of adult subjects. Forensic Sci Int 171(2-3):103-110.

San-Millan M, Rissech C, and Turbon D. 2017. New approach to age estimation of male and female adult skeletons based on the morphological characteristics of the acetabulum. Int J Legal Med 131(2):501-525.

Saunders SR, Fitzgerald C, Rogers T, Dudar C, and McKillop H. 1992. A Test of Several Methods of Skeletal Age Estimation using a Documented Archaeological Sample. Canadian Society of Forensic Science Journal 25(2):97-118.

Schmitt A. 2004. Age-at-death assessment using the os pubis and the auricular surface of the ilium: a test on an identified Asian sample. International Journal of Osteoarchaeology 14(1):1-6.

Schmitt A, Murail P, Cunha E, and Rougé D. 2002. Variability of the Pattern of Aging on the Human Skeleton: Evidence from Bone Indicators and Implications on Age at Death Estimation. J Forensic Sci 47(6).

Shirley NR, and Ramirez Montes PA. 2015. Age estimation in forensic anthropology: quantification of observer error in phase versus component-based methods. J Forensic Sci 60(1):107-111.

Stewart TD. 1957. Distortion of hte Pubic Symphyseal Surface in Female and its Effect on Age Determination. Am J Phys Anthropol 15:9-18.

Suchey JM. 1979. Problems in the Aging of Females Using the *Os pubis*. Am J Phys Anthropol 51:467-470.

Suchey JM, Brooks ST, and Katz D. 1988. Instructions for Use of hte Suchey-Brooks System for Age Determination of the Female Os Pubis. In: Casting F, editor. Fort Collins, Colo.

SWGANTH. 2013. Age Estimation.

Todd TW, and Lyon DW. 1924. Cranial suture closure, its progress and age relationship. Part I. Adult males of white stock. Am J Phys Anthropol:325-345.

Todd TW, and Lyon DW. 1925. Cranial suture closure, its progress and age relationship. Part II. Ectocranial closure in adult males of white stock. . Am J Phys Anthropol:23-45.

Todd WT. 1920. Age Changes in the Pubic Bone. Am J Phys Anthropol 3(3):285-339.

Todd WT. 1930. Age Changes in the Pubic Bone: VIII. Roentgenographic Differentiation. Am J Phys Anthropol 14(2):255-271.

Ubelaker DH. 2010. A History of Methodology in the Estimation of Age at Death for the Skeleton. Age Estimation of the Human Skeleton Charles C Thomas. p   xvii-XXV

Uhl N, and Nawrocki S. 2010. Multifactorial Estimation of Age at Death from the Human Skeleton. Age Estimation of the Human Skeleton. p 243-261.

White T, Black M, and Folkens P. 2012. Human Osteology Third Edition Elsevier Academic Press.

Wiesema J, Love J, and Naul L. 2009. The Influence of the Daubert Guidelines on Anthorpological Mehtods of Scientific Identification in the Medical Examiner Setting. Hard Evidence: Case Studies in Forensic Anthropology New Jersey Prentice Hall. p 80-90.

Yoder C, Ubelaker DH, and Powell JF. 2001. Examination of Variation in Sternal Rib End Morphology Revelevant to Age Assessment. J Forensic Sci 46(2):223-227.

VITA

Christine Ashley Bailey née Reilly was born in the great state of New Jersey on May 31, 1992 to parents Paul and Wendy Reilly. She moved with her family to Strawberry Plains, Tennessee in 2005. Christine was first introduced to Forensic Anthropology through the University of Tennessee's Upward Bound program while in high school. Following her high school graduation in 2011, she pursued her interest in anthropology by attending the University of Tennessee, Knoxville. While completing her degree, Christine took guardianship of her younger sibling, Susan, and married her best friend, Cory Bailey. Christine completed her Bachelor of Arts degree in Anthropology with a concentration in Disasters, Displacement, and Human Rights with Suma Cum Laude and as the outstanding graduate in Anthropology in 2015. She continued her studies at UTK pursuing a Master of Arts in Anthropology under Giovanna M. Vidoli and will graduate in August 2018. During her three years as a graduate student, Christine served as the Volunteer Coordinator for the Forensic Anthropology Center and was the teaching assistant for the undergraduate human osteology course. Christine was also the forensic anthropology instructor and mentor for the Upward Bound program for three summers. During her final year, Christine was selected for the steering committee of the 2018 Conference in Disasters, Displacements, and Human Rights. In spring of 2018, Christine was offered a position as an instructor and the forensic anthropology facilities curator at Western Carolina University beginning fall 2018.