Impacts of the Growth Environment on Cortical Bone Plasticity during Childhood.

by

Julia Meyers

Master of Arts, University of Victoria, 2017

Bachelor of Arts and Science, University of Guelph, 2014

Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in the

Department of Archaeology

Faculty of Environment

SIMON FRASER UNIVERSITY

Summer 2023

Declaration of Committee

Name:	**Julia Meyers**
Degree:	**Doctor of Philosophy**
Title:	**Impacts of the Growth Environment on Cortical Bone Plasticity during Childhood.**
Committee:	**Chair: Dana Lepofsky** Professor, Archaeology
	Hugo FV Cardoso Supervisor Professor, Archaeology
	Deborah C. Merrett Committee Member Adjunct Professor, Archaeology
	Lesley Harrington Committee Member Associate Professor, Anthropology University of Alberta
	Dongya Yang Examiner Professor, Archaeology
	Julia Gamble External Examiner Associate Professor, Anthropology University of Manitoba

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Bone growth is plastic in its response to the environment, particularly during development. This research aims to examine the plastic response of cortical bone in the long bones of children via the examination of cross-sections of the femur, tibia, and humerus. The goal of this research is to identify some of the factors of the growth environment that impact cortical bone deposition, like biosocial stress, physical activity, and body mass, and to attempt to pull apart their impact on cortical bone distribution. The first paper is a preliminary study examining the intraobserver error from several rounds of cortical area measurements taken from computed tomography scans of dry and wet bone using both manual and algorithmic segmentation methods. The error rates and reliability coefficients for cross-sectional parameters taken from dry and wet samples are comparable, as are those between the manual and algorithmic methods, indicating that consistency of measurement is similar among both samples and al segmentation methods. The second paper develops body mass estimation formulae based on regression of known weight with several breadth measurements and torsional rigidity (J) of the femur and the tibia based on a modern sample of children (n=77). Body mass estimates created using the formulae developed in the study are then compared with documented weights to examine prediction error. Formulae using J values at the femoral mid-shaft produce the least amount of error and are accurate to ±12.62 kg for the overall sample, but surprisingly the formulae using J from sections of the tibia produce comparable results. The third paper examines the relationship between biosocial stress, mechanical loading, and distribution of cortical bone among 106 children from the Lisbon Documental Skeletal Collection (n=45) and the New Mexico Decedent Image Database (n=61). Results suggest that the amount of cortical bone deposited, and medullary cavity size are most impacted by biosocial stress, while bone strength and cross-sectional shape are more influenced by physical activity and, particularly in the lower limbs, body mass. This research demonstrates the importance of environmental context when examining cortical bone deposition in children. Biosocial stress and mechanical loading demonstrate distinct patterns of cortical bone distribution during development. The research also demonstrates the importance of examining multiple long bones, as different sections, bones, and limbs will have different plastic responses to the growth environment.

Keywords: Cortical bone development; children; long bones; historical context; bone plasticity

Dedication

This one’s for me and P.

Acknowledgements

I would like to thank the Nation Museum of Natural History and Science in Lisbon for allowing me access to the Luís Lopes Documented Skeletal Collection and the Imagens Medicas Integradas Clinic in Lisbon for allowing me access to their CT scanner, and for their time and work in helping to collect the scans. I would also like to thank the Office of the Medical Examiner of New Mexico and the University of New Mexico for access to the New Mexico Decedent Image Database.

I would like to thank my supervisor Hugo Cardoso for his guidance and effort in aiding the completion of this thesis. I would also like to thank the members of my advisory committee Deborah Merrett, and Lesley Harrington for their help in crafting my research over the course of my degree. The members of the JUNO Lab Group have also help me during my time at Simon Fraser University. Their continued support and encouragement have helped keep me going.

My friends and family have always supported my academic work, and I want to thank them for all they have done and continue to do. Mum, Dad, and Ari have always been around when I needed support and guidance and to help inspire me and give me perspective. Logan, Lucas, and Lissy, thank you for being the best siblings in the world. Ellie and Hossein, thank you for letting me crash on your couch once a week. I honestly don't think you'll ever know how much it means to me.

I would also like to acknowledge the lives and remains of those whose remains were used in my research. This research would not have been possible without their bones. Finally, I would like to acknowledgement the traditional territories of the xʷməθkʷəy̓əm (Musqueam), Sḵwx̱wú7mesh (Squamish), səl̓ilw̓ətaʔɬ (Tsleil-Waututh), q̓íc̓əy̓ (Katzie), kʷikʷəƛ̓əm (Kwikwetlem), qiqéyt (Qayqayt), q'ʷa:n̓ƛ'ən̓ (Kwantlen), Səmyámə (Semiahmoo), and sc̓əwaθən (Tsawwassen) Nations. It has been a joy and a privilege to spend time on your land as I completed my degree, and as I live my life.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

Socially, childhood is not a universal experience. The period known as childhood differs depending on the time and place in which an individual lives. Biologically, the period of growth and development in which a sub-adult matures into an adult is more well defined. The impacts of environment on bone growth and development mean that the biosocial experience of growing up will vary. Bone growth itself is plastic in its response to the environment and can be influenced by a number of factors. This research aims to examine the plastic response of cortical bone in the long bones of children via the examination of cross-sections of the femur, tibia, and humerus. This is done using two samples of children: one composed of individuals from early 20th century Lisbon, Portugal, and another composed of individuals from early 21st century New Mexico, United States of America. The goal of this research is to investigate how variation in environmental factors during growth and development influences the measurement of cortical bone distribution.

1.1. Long Bone Development and Plasticity

There are two main types of bone tissue found in the human skeleton, trabecular and cortical bone. Both are found in the long bones, with cortical bone comprising the majority of the diaphyses of long bones and trabecular bone making up most of the epiphyses and parts of the metaphyses (Clarke 2008; Rajamannan 2018). Long bone diaphyses are formed primarily through endochondral formation, which involves the creation of a cartilage model that is gradually replaced with bone, though some intramembranous ossification does occur at the compact outer surfaces (Mackie et al. 2008).

There are two phases of endochondral skeletal development for long bones: morphogenesis and linear growth. Morphogenesis involves the formation of the embryonic skeleton and ends in mid gestation with the establishment of a functional growth plate or metaphysis (Degnin et al. 2010). During the early cellular differentiation stages, embryonic lineage, paraxial mesoderm, and lateral plate mesoderm cells form the axial and appendicular skeleton (Berendsen and Olsen 2015). Next, cells specifically differentiate into osteoblasts and chondrocytes which deposit extracellular matrices

composed of specific cartilage cells (Rajamannan 2018). Invading osteoblast progenitors, osteoclasts, blood vessels, and hematopoietic cells allow for the vascularization and ossification of the cartilage, becoming the primary ossification centre (Berendsen and Olsen 2015). Linear growth continues after in the metaphyses throughout childhood and ending in early adulthood (Degnin et al. 2010).

Long bones begin formation during the fetal period, and then continually grow and remodel to varying degrees throughout an individual's life. Changes to the rate of linear bone growth tend to occur in early childhood, with growth occurring at a higher and more consistent rate during and after the onset of puberty in the adolescent period (MacKelvie et al. 2002; Ruff 2003a). Appositional growth, particularly, among the cortices of the long bones, continues later into life through the process of remodeling. When examining appositional growth of the cortical bone in the diaphysis, the mechanism responsible for bone deposition is a structure called a basic multicellular unit (BMU). Cortical formation and the subsequent remodeling are accomplished by the BMU, containing osteoclasts and osteoblasts, which are responsible for removing old bone and depositing new tissue (Robling et al. 2006). The BMU interacts with different bone components like osteocytes within the bony matrix and lining cells, which cover the surface of the bone to deposit new bone tissue (Sims and Gooi 2008).

The regulation of the growth and development of cortical bone in long bones is heavily influenced by genetic heredity, hormone secretion, nutritional intake, and habitual loading. Individuals have the potential for a certain amount of bone mineral and density growth the is inherited and genetic in origin (Smith e al 1973; Prentice 2001; Cowgill and Hager 2007). During childhood bone growth is regulated by growth hormones and insulin growth factors, and after puberty sex steroids (estradiol and testosterone) increase dramatically and become the primary modulators of skeletal growth (Bass 2000; Devlin 2011). As bone is comprised of inorganic minerals and an organic matrix (in addition to cells and water) the intake of macronutrients and micronutrients during childhood are important for reaching peak bone mass and preventing bone deterioration later in life (Bonjour et al. 2015). Dietary protein is involved in regulating the calcium–phosphate economy and bone metabolism, processes that are mediated by insulin growth factor-I (Bonjour et al. 2015). Additionally, micronutrients like calcium and vitamin D are integral to the development of the inorganic bone component (Weaver et al. 2015; Holick 2015). Loading of the long bones has more localized impacts

on bone growth during later childhood and adolescence and can influence adult bone density and strength (Ruff et al. 1994; Ruff et al. 2006).

Biosocial and mechanical factors are not fixed and can be influenced by different aspects of the growth environment, leading to a plastic response in bone growth. Cortical bone distribution is plastic within individuals as bone is continually remodelled throughout life and thus can be influenced by the environment (Robling et al. 2006). Biological plasticity refers to the ability of an organism to adjust their morphology and behaviour to their environment (Bogin 1999; Cardoso 2005). In this case, biological plasticity helps to explain how cortical bone morphologies reflect energy and nutrient allocation by the body during development, as well as habitual loading history. This research examines how the plastic response of cortical bone during growth is influenced by both *biosocial stress* and *mechanical loading*. Both biosocial stress and mechanical loading are products of the environment in which an individual develops, but each have different effects on cortical bone distribution.

For the purposes of this research, biosocial stress will be considered to be the effects of the growth environment which lead to a nutrient deficiency and/or energetic trade-off that could favour other areas of growth over cortical bone development. Other researchers use the terms nutritional, metabolic, or physiological stress (Garn 1964; Eleazer and Jankauskas 2016; Osipov et al. 2020), but biosocial stress allows for the incorporation of additional environmental factors that may cause stress to the bone growth process, such as disease load or lack of access to medical care. The previously mentioned factors that are integral to normal bone development; genetics, hormonal regulation, and nutritional access, can be subjected to trade-offs if under stress which could result in the disruption of cortical bone deposition.

Factors like genetic make-up and hormonal regulation are less likely to be affected by the growth environment, but there are some exceptions. Within Development Origins of Health and Disease (DOHaD) research there is evidence to demonstrate that stressors (like malnutrition or disease) experienced by the parent can affect the (epi)genetic make up of their gamete (Gowland 2015; Temple 2019). Sex hormones are also integral to the regulation of cortical bone growth in the long bones. During the onset of puberty, sex steroids (estradiol and testosterone) increase dramatically, replacing growth hormones and insulin growth factors as the primary modulators of skeletal growth

(Bass 2000; Devlin 2011). This change in hormone regulation, as well as the difference in timing of puberty between males and females means that the amount and distribution of cortical bone is partially reliant on hormone levels/types present at the time of death (Rogol et al. 2000). The impacts of both genetic capacity and the presence of sex hormones on appositional growth are generally well understood and can be investigated using population histories and sex estimates (epigenetic changes not withstanding). The amount of impact environmental stressors, like malnutrition and disease, affect have on cortical bone development is less easy to identify. It is difficult to measure the influence of stressors within populations, and, as their impacts are largely non-specific, identifying a stressor based on individual bone morphology, is also difficult.

Access to nutrition and extent of disease load can affect the energetic and mineral requirements needed for cortical bone development. Undernutrition and disease are stressors that are heavily influenced by social and environmental contexts, wherein depending on the time period, and/or an individual's social status the impact of these stressor on cortical bone distribution can be highly variable. Thus, the presence and level of these stressors in the environment can be indicative of the amount of biosocial stress an individual experiences. Nutritional stress in the form of protein-caloric deficiencies has been shown to influence cortical bone development, as well as overall attained stature (Garn et al. 1964). Malnutrition refers to either deficiencies or excesses in nutrient intake, meaning that there is some nutrient imbalance occurring. Undernutrition involves an insufficiency in energy and nutrient intake, whereas overnutrition involves excessive intake of nutrients (Mathur and Pillai 2019). It is possible for an individual to be both under and over nourished (e.g., an over abundance of calories, but a lack of micronutrients). Increased disease loads present during development also negatively impact growth. Decreased bone mineral density can be associated with childhood osteoporosis, but more commonly it is observed as a side effect in other chronic childhood diseases (van der Sluis and de Muinck Keizer-Schrama 2001). Long term exposure to pathogens and/or chronic disease leads to an increased inflammatory response by the immune system, which has been demonstrated to cause energetic trade-offs leading to stunting in bone growth and lowered bone mineral density (Gowland 2015; Maratova et al. 2017). In these examples, the energetic and/or nutrient requirements of normal bone growth are not met, and thus cortical bone apposition is adversely affected. Generally, this results in lower levels of total cortical bone area, as

well as an increase in medullary cavity size when compared with normal cortical bone development in long bones. Long term exposure to malnutrition and disease can create systematic disruptions in the bone formation and remodeling processes, unlike mechanical loading, which has more localized effects on bone remodeling in the long run (Eleazer and Jankauskas 2016).

When working with skeletal samples, even well documented ones, it is not possible to identify each individual stressor a person would have experienced during life, but historical context and documentation giving individual factors like age at death, cause of death, and demographic information can be used to create an understanding of the general level of biosocial stress an individual would have experienced within the environment they grew up in. When certain stressors (malnutrition, disease, poor healthcare) would have been present an assumption about the general level of biosocial stress can be made (Goodman and Armelagos 1988; Temple and Goodman 2014)

The other group of factors in the growth environment that can affect cortical bone plasticity are the result of mechanical loading of the diaphysis. The mechanical factors that have been shown to affect cortical bone apposition are bone loading via habitual physical activity and body mass (Ruff et al. 1994; Liebermann et al 2001; Osipov et al. 2016). The theory explaining the impacts of mechanical loading on long bones was drawn from engineering research conducted on hollow beams (Ruff 2003) and Wolff's Law, which states that where bone is needed it will be added, and where it is not it will be lost (Ruff et al. 2006). Therefore, when differential loading forces are applied to the bone (i.e., shear, torsion, compression), the area of the diaphysis where the most force occurs will experience an increase in cortical bone deposition. This can lead to increases in bone strength and rigidity, as well as to more ovoid cross-sectional shapes (Ruff and Hayes 1983; Pomeroy et al. 2018).

Physical activity is one of the mechanical factors that influences an individual's loading history. Increased physical activity leads to increased cortical bone deposition, particularly during growth and development. Hormonal regulation plays a role in cortical bone's plastic response to loading via physical activity, as there are differential growth responses during different hormone periods. The physical activity related growth that occurs prior to puberty (regulated by growth hormones) is particularly responsive in terms of bone density, but puberty is when bone remodeling (regulated by sex

hormones) is at its lifetime peak (Bass 2000). The bone deposition that occurs during late-childhood and adolescence is therefore likely to influence cortical bone distribution and robusticity well into adulthood.

Body mass is the other significant mechanical factor that influences loading history. The same plastic response occurs, wherein as more force is applied to the long bone, cortical bone is deposited in response, but instead of the force being the result of increased physical activity, it is the result of the individual's body mass (Cowgill 2018; Pomeroy et al. 2018). There is also the possibility that lower limb size and shape are constrained by body mass and proportion (Shaw and Stock 2011). The effects of body mass on bone loading are particularly important when it comes to modern populations, many of whom have demonstrated a secular trend towards obesity in children, though there is evidence to support changes in cortical bone deposition coinciding with obesity in adults (Beck 2009; Reeves 2014). The investigation of the relationship between cortical bone development and body mass among children has yet to examine the effect of obesity on cortical bone distribution.

The effects of loading history on cortical bone distribution have been used by many researchers to study patterns of human variation. From foraging patterns among hunter-gatherer (Stock and Pfeiffer 2001; Sládek et al. 2015) to bone strength among professional athletes (Shaw and Stock 2009; Shaw et al. 2012), cortical distribution has given many researchers insight into cortical bone's plastic response to physical activity. The affects of body mass on cortical bone distribution have been used to develop body mass estimation formulae for children (Robbins 2010; Robbins Schug 2013). Whereas the relationship between physical activity and bone strength has been studied in multiple populations spanning thousands of years, very few individuals have been studied when producing weight estimation formulae from cross-sectional cortical bone properties for children. Most equations rely on data from The Denver Growth Cohort, run from 1927 to 1967 (Maresh 1943; Maresh 1955; Maresh 1970). It was conducted using twenty white, middle class, American children who were unlikely to be under any caloric or nutritional stress and would not have been overweight or obese. The utility of these formulae among modern populations of children have been questioned (Yim et al 2020; Spake et al. 2021), and no weight estimation formulae have been developed using a sample containing obese children. Thus, this aspect of cortical plasticity in response to loading via body mass has been overlooked.

While the biological mechanisms (osteoclasts, osteoblasts, BMU) responsible for cortical bone deposition are influenced by both biosocial stress and mechanical loading, this dissertation examines if these factors can be teased out when examining cortical bone distribution in the long bones of children using cross-sectional geometry.

1.2. Cross-Sectional Geometry

Cross-sectional geometry (CSG) quantifies the cortical bone distribution in a transverse cross-section of bone. CSG allows for the measurement of skeletal robusticity, which is often a response to habitual loading undertaken during a person's life (Ruff et al. 1994; Ruff 2007; Ruff 2013; Osipov et al. 2016). A cross-section can be examined for several area measurements: the cortical tissue area (CA) (representing the bone strength), the area of the medullary cavity (MA), and the total area of the cross-section (TA). The maximum and minimum second moments of area, represent the axes of greatest and least bending strength of the bone diaphysis and are dependent on the distribution of the cortical bone around the centroid (I_{max} and I_{min}, respectively) (see Figure 1.1 for a visual representation of these properties). The polar second moment of area (J) indicates the torsional strength of the bone ($I_{max} + I_{min}$), while the I_{max} and I_{min} index (I_{max}/I_{min}) quantifies the shape of the cross-section, while removing size from consideration.

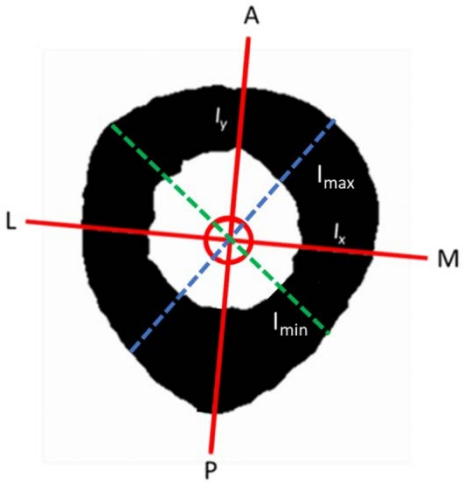


Figure 1.1 Representations of CSG measurements taken from a femoral diaphyseal cross-section. Cortical area (CA) is represented by the black bone surface area. The maximum and minimum second moments of area are indicated in blue (I_{max}) and green (I_{min}).

The use of CSG on long bones was developed using beam theory, a section of engineering that deals with the impact of different types of external forces on hollow, cylindrical objects (Lovejoy et al. 1976; Ruff and Hayes 1983). The axial rigidity (representative of the compressive or tensile forces applied to the bone) is proportional to the CA of the cross-section, while the bending and torsional rigidities (representative of the bending and torsional forces applied to the bone) are proportional to the bending rigidity(I) of a plane of the cross-section, and polar second moment of area (J), respectively (Ruff and Hayes 1983).

The methods used to obtain CSG measures have evolved over the course of the last few decades. Originally, physical cross-sections of the bone were isolated. There were obvious limitations associated with these methods, as cutting cross-section was destructive to the bone itself. Two dimensional radiographs began to be used frequently to examine both external and internal long bone cortical dimensions, but there are some issues in translating one-dimensional breadth measures taken from the x-rays into two-dimensional area measurements (Ruff and Hayes 1983). In recent decades the use of computed tomography (CT) to examine the distribution of cortical bone within cross-sections of long bones has become the method of choice among biological anthropologists looking to visualize both outer and inner bone dimensions (Sumner et al. 1985; O'Neill and Ruff 2004). It is possible to estimate robusticity from outer dimensions alone, but the examination of the internal dimensions of the medullary cavity is an important factor when examining the variable cortical dimensions that tend to be found among juveniles and older adults (Stock and Shaw 2007; Macintosh et al. 2014; Kurki et al. 2022). Also, periosteal (or bone surface) measures seem to be more influenced by loading history, while endosteal measures are more likely to be effected by biosocial stress (Ruff et al. 2013).

Biological anthropologists use CT imaging to examine cortical bone distribution in both bones situated in tissue (wet bone) and archaeological samples of bone (dry bone). There has been little to no research done on the differences between CT images of dry and wet human cortical bone, and none on how these different CT images are processed within ImageJ, a popular, Java-based image processing program. However clinically, different CT protocols are employed for soft tissue and bone reconstructions. There has been some research conducted on the consistency and repeatability of dry bone measurements taken on physical bone and CT bone images, wherein reliability of

variables between physical and virtual bone was around 95% (Corron et al. 2017). It is important to be able to understand how reliability of measurements may change depending on whether the CT images were taken on wet bone or dry bone. Saers and colleagues (2021) examined trabecular bone in low resolution CT scans taken from living individuals and micro-CT scans taken from dry bone specimens and developed a method for comparing the two. Their method relied on the use of the ratio of bone area to total area of the scanned surface, and they found that they were able to compare this variable between the different resolutions, but other variables such as bone mineral density and trabecular patterning could not be compared. They also noted that archaeological specimens were subject to taphonomic changes that would complicate attempts to quantify bone mineral density (Saers et al. 2021). In addition, dry bone literally dries out, losing its periosteal tissue layer and other non-bone, cartilaginous components that could affect how the bone is visualized in a CT image. Wet bone CT images will still contain these components.

1.3. Historical Context of the Samples

One of the more unique aspects of the samples is level of documentation and information on the individuals in each of the collections used in this research. The individuals used in this research come from two separate skeletal collections: the Luís Lopes Documented Skeletal Collection (the Lisbon sample) and the New Mexico Decedent Image Database (NMDID) (the New Mexico sample). Each collection has documentation corresponding to each individual, which gives specific information on date of birth, date of death, cause and manner of death, and biological sex. Additional information is provided depending on the collection. The Lisbon sample is housed in the National Museum of Natural History and Science in Lisbon, Portugal, and is composed of individuals who lived in Lisbon during the early and mid 20th century. Despite being a recent historical skeletal collection, it is comparable to bone archaeological samples (Cardoso 2005). The skeletal remains were previously interred in temporary grave plots, disinterred for secondary burial and eventually entered the museum once unclaimed by family members. The CT scans of individuals in the Lisbon sample were taken at the Imagens Medicas Integradas Clinic in Lisbon. The New Mexico sample is composed of wet bone specimens taken from the pre-autopsy CT scan collection of the Office of the Medical Investigator (OMI) in New Mexico, USA. The OMI partnered with the University

of New Mexico to create the NMDID, where CT scans and supporting documentation are housed.

These collections are uniquely well suited to understanding the growth environments the children lived in and determining the levels of biosocial stress and mechanical loading each child would have experienced. Individual level data available for those in these collections is much larger than what is available for a traditional archaeological sample, as area level population or subpopulation demographic data is often the only contextual information available for these samples. Additionally, the individual level data in archaeological samples are generally estimates, like age and (especially difficult for children) sex. The Lisbon and New Mexico samples allow for the investigation of more individual-level environmental factors, thus addressing similar questions regarding cortical bone development as studies using archaeological samples, but with fewer limitations surrounding individual level data.

The following sections give the historical and social context under which children in each sample would have lived. As previously discussed, the growth environment can have a direct impact on cortical bone distribution as it is plastic in its response to changes in biosocial stress and mechanical loading. The environment in which these individuals developed would have had a direct impact on the distribution of cortical bone in their long bones.

1.3.1. The Lisbon Sample

Portugal makes for an interesting case study in child health in the early 20th century. Conditions in many European nations were poor for children during this time period, but, unlike Great Britain and France, Portugal was not at the forefront of many policy and social changes regarding children. Infant mortality rates can often give a general, population-level indication of child and public health conditions, especially when comparing changes in rates over time. As of 1910 infant mortality rates were around 134 deaths per 1000 births (Moreira and Henriques 2016). In terms of education, Portugal made attempts to legislate a national school system and made primary school compulsory in 1840 (Goulart and Beli 2017), but by 1870 had fallen behind other European nations in the expansion of its primary education (Benavot and Riddle 2009). Schooling suffered from the educational influence of royalists and Catholics, and by

1926 Portugal was considered the least educated nation in Europe (Lannon 1987). Sanitation was another issue, as many Portuguese were moving into crowded urban areas at the start of the 20th century. Children living in urban Portugal during the early 20th century were subjected to poor living conditions, as these areas were overpopulated and unsanitary (Cardoso and Garcia 2009). Conditions would not improve until the later part of the 20th century, as by 1950 43% of working-class families in Lisbon still had no piped water, 69% had no electricity, and 81% had no toilet (Cardoso and Garica 2009). Census data show that between 1900 and 1911, 45% to 48% of children aged 10–19 were engaged in paid work (Goulart and Beli 2017). Children in Lisbon often entered the labour market around 12 years of age and would have been at a high risk of this increased physical activity stunting overall growth (Cardoso and Garcia 2009). Both linear and appositional growth velocities increase during the late juvenile and early adolescent period, leaving these individuals at risk for adverse growth outcomes as excessive physical labour increases (Bogin 1999; Bogin et al. 2007) Even children who were not involved in paid wage labour would have been expected to engage in unpaid labour around the home.

The sample of Portuguese children examined in this research were taken from the Luís Lopes Documented Skeletal Collection, and as such, had several aspects of their identity documented. The documented data includes date of birth and death, cause of death, birth neighbourhood (*freguesias*), father's occupation, and occupation of the child. Cause of death among the Lisbon sample tended to include more "natural" causes, meaning those associated with disease, especially tuberculosis. While an attempt to include other cause of death among the sample was made, it was clear that chronic or acute disease were the most common causes of death in the collection. Mortality bias towards children who were chronically ill therefore may be present within the sample. These data give both individual and area level information on the factors that would have influenced the growth environment of each child in the sample (Cardoso 2005; Cardoso 2006).

1.3.2. The New Mexico Sample

The United States began the 20th century with an infant mortality rate of 100 deaths out of 1000 births and ended it with a rate of 7.2 deaths per 1000 births (Meckel 1990; Hoyert et al. 1999). The establishment of the Children's Bureau in 1912

demonstrated the newfound interest in the lives and health of American children in the 20th century (Meckel 1990). Throughout the wars and the Depression as federal programs grew, aid for child health grew as well (Ashby 1985). The implementation of Medicaid and other federal programs aimed at child health in the 1960s is credited with causing a large decline in infant mortality (Pharoah and Morris 1979). Technological advances and expanding perinatal services caused neonatal mortality to decrease 41% and post-neonatal mortality declined 14% from 1970 to 1979, a trend which slowed during the mid-1980s (Kleinman 1990). A larger cultural shift occurred in the post-baby boom years as well, with women becoming more involved in family planning and contraception, a change which began to cause a flattening in birth rates by the mid 1970's (Dowan 1985). Child mortality rates in New Mexico during the early 21st century stood at around 7.2 deaths per 1000 births (Meckel 1990; Hoyert et al. 1999). In addition to experiencing the benefits of policy change and advancements in medical care, children living in New Mexico during the early 2000's would have been at the end of a century-long positive secular trend in body mass that occurred in the US (Sun et al. 2012; Tylavsky et al. 2019). Over one third of the children within the New Mexico sample fell into the obese weight category (above 95th percentile BMI), which falls in line with larger obesity and trends towards sedentarism in the United States (Troiano et al. 1995; Ogden et al. 2008; Komlos et al. 2009). These children would have been heavier, less active, and less biosocially stressed than the children who lived in early 20th century Lisbon.

As with the Portuguese children, the children selected to represent the 21st century New Mexico sample were part of a documented anatomical collection. The New Mexico sample is comprised of children from the NMDID, meaning each individual has information on date of birth and death, cause of death, height, weight, and death neighbourhood (in the form of zip codes). Cause of death among the New Mexico sample was more likely to reflect "accidents" (motor vehicle accidents, homicides, or suicides), thus mortality bias is less likely to be present among the children in the sample. These data, though not identical to the data recorded as part of the Luís Lopes collection, allow for similar individual- and area-level information to be gathered on the growth environment of each child in the sample. The documented data from both collections can aid in examining the impact that different types of growth environments have on cortical bone distribution.

1.4. Goal and Structure of the Dissertation

This dissertation contains three papers, each with specific goals and research questions, but all related to the development and analysis of the distribution of cortical bone in the long bones of children. They also share the use of diaphyseal cross-sections from the femur, tibia, and humerus, taken from CT scans. This research was conducted in order to examine how the growth environment influences cortical bone deposition during growth and development.

The first paper in this dissertation examines the differences in cortical bone measurements taken from CT scans of wet bone and dry bone. The purposes of this study are to examine reliability of dry and wet bone cortical area measurements through an examination of intraobserver error between three rounds of measurement, and to examine differences in CA calculated using manual and algorithmic methods of differentiating bone from non-bone. The results of this paper build the foundation for any future comparisons between the CT scans taken from the Luís Lopes Documented Skeletal Collection, which consists of dry bone, and CT scans taken from the New Mexico Decedent Image Database, which consist of bone *in situ*.

The second paper develops weight estimation formulae based on sections of the femur and tibia using a documented weight-for-age sample of children taken from the NMDID. Most of the contemporary formulae used by anthropologists to estimate weight among children using skeletal parameters were developed from the same 20 individuals from the Denver Growth Study (Ruff 2007; Robbin et al. 2010; Cowgill 2018). This limitation is mainly due to the availability of data, wherein during the study period only 20 individuals had weights recorded and radiograph taken at 6-month intervals. The consistent use of the same 20 individuals from the Denver Cohort to develop weight estimation formulae for children indicates the lack of and need for childhood skeletal growth data from other sources, particularly when used to estimate weight in modern forensic contexts. The current study develops weight estimation formulae based on classical linear regression of known weight with metaphyseal breadth, epiphyseal breadth, and torsional rigidity measurements taken from the femur and tibia. Weight estimates created using the formulae developed in the study are then compared with known weight to examine error. The formulae are created from a reference sample of 77 children from the New Mexico sample, who are, in general, more overweight and obese

than children from the Denver Growth Cohort. These formulae are uniquely suited for use among modern forensic populations of children and adolescents.

The third paper examines the relationships between biosocial stress, mechanical loading, and the amount and distribution of cortical bone. The study relies on cross-sectional bone data from known sex and age children from the Luís Lopes Documental Skeletal Collection and the New Mexico Decedent Image Database. The New Mexico sample is composed of 45 children ranging in age from birth to 18 years of age (20 male and 25 female), and the Lisbon sample includes 61 children from the same age range (33 male and 28 female). Differences in the growth environment exist between the two samples and among each sample, in the form of differential access to modern health care, socioeconomic status, level of physical activity, and average body mass. To analyze this relationship, the study addresses two main research questions. The first question asks if there is a difference in cortical distribution between individuals in the early 20th century Portuguese sample and the early 21st century New Mexico sample. The second asks if there is a relationship, within each sample, between socioeconomic status and cortical bone distribution. This paper examines the impact of biosocial stress, such as access to health care, nutrition, and disease load, in the growth environment and determines if it is possible to differentiate their influence on cortical bone distribution from those produced by mechanical loading, such as through variations in physical activity level and body mass. The ability to pull apart these different influences would allow for a more nuanced insight into the growth environment of children in archaeological samples that goes beyond just active v. inactive or stressed vs non-stressed.

The dissertation concludes with a discussion section that briefly reviews the results from the three papers, and then delves deeper into some of the common themes and ideas that arise from the studies, such as the impact of body mass, physical activity, nutrition, and disease on cortical bone distribution. It also compares the results with previous literature on appositional cortical bone development and outlines some of the future directions other researchers may take with the topic. Finally, the significance of the research is discussed within the broader context of development and plasticity of cortical bone in children.

Contemporary research on biosocial stress tends to focus on cortical thickness, while most research on bone loading examines cross-sectional parameters. The research that does use both is often done on archaeological populations (Ruff 2003a; Cowgill and Hager 2007; Cowgill 2010; Cowgill 2014; Harrington and Osipov 2018). As mentioned previously, this means there is a lack of individual level data available for these samples. Less attention is to the impact of body mass on cross-sectional parameters than the effects of physical activity in archaeological populations, likely because there are limited data sets wherein both body mass and long bone cross-sectional properties are known. The documented nature of the samples used in this research allows for novel research into the differences between measuring wet and dry bone from CT images, the use of cross-sectional properties in estimating body mass among children, and on pulling apart the influences of biosocial stress and mechanical loading on cortical bone distribution.

The overarching goal of the research is to unpack and examine some of the factors that influence cortical bone deposition during growth and development. From examining the methods researchers use to quantify differences in cortical bone distribution to the effects of the growth environment on cortical bone deposition. This research attempts to distinguish the different influences of the growth environment on cortical bone distribution using collections of documented skeletal remains with known differences in medical access, socioeconomic status, physical activity level, body mass, malnutrition, and disease load allowing for more individual detail and historical context than have previously been used.

Chapter 2. Assessing error and reliability among wet and dry bone CT images.

2.1. Abstract

There has been very little work examining the differences in bone measurements taken from CT scans of wet bone and dry bone among humans in bioarchaeology or biological anthropology (Corron et al. 2017; Saers et al. 2021). there are two goals of this study; the first is to examine the reliability of dry and wet bone cortical area (CA) measurements through an examination of intraobserver error between three rounds of measurement, and the second is to examine differences in CA calculated using manual and algorithmic methods of differentiating bone from non-bone. Overall, most differences in CA between observer rounds were not significant, and there is no major difference between the CA values calculated by the algorithmic methods and the manual method in either the wet or the dry bone samples. The analysis of intraobserver error of the measurements taken using the algorithms was similar to the error rates produced between rounds when using manual method. For both algorithmic methods and the manual method mean percent technical error was below 0.047% and the mean absolute technical error was between 4.9 and 5.9 mm^2.While the manual method most approximates the researcher differentiating internal and external contours of the cortex from non-bone within ImageJ, the results indicate that reliability and the technical measurement of error percentage were uniform across both the automated methods and the manual method of segmenting (0.999, and <1%, respectively).

2.2. Introduction

The development of computed tomography (CT) has allowed for biological anthropologists to examine the human skeleton in ways that were previously difficult, impossible, or that would have caused damage to the specimen. Nearly half a century after its development for clinical use, CT imaging has aided in the study of palaeopathology, forensics, and human biology. One area in particular where CT scans have been used is in the analysis of cortical bone in long bone cross-sections. Measurements of cross-sectional geometry taken from these scans can be used to approximate the bone loading history of the individual. Many studies using CT images

have focused on differences in cross-sectional geometry associated with physical activity, (Stock 2006; Hind et al. 2012; Shaw and Ryan 2012). senescence (Bouzsein et al. 1994; Agarwal and Grynpas 2009), and malnutrition (Agarwal 2016). While there has been a large amount of research on the distribution of cortical bone using CT imaging, there has been very little work focused on examining the differences in measurements taken from radiographs of human wet bone and dry bone (Elliot 2022). The compatibility of these types of CT images is often taken for granted in biological anthropology research examining differences in cortical bone distribution.

Cross-sectional geometry (CSG) quantifies the cortical bone distribution in a cross-section of bone. In this study computed tomography was used to study slices of bone taken transversely at specific sections of the long bone diaphysis. Measurements of cortical area (CA) were calculated from these cross-sectional slices by the Slice Geometry function in BoneJ (Figure 2.1).

Figure 2.1 The cortical bone area (CA) measurement taken from a femoral diaphyseal cross-section is represented by the black bone surface area.

The methods used to obtain CSG measures have changed as radiograph and CT technologies have evolved. Prior to the use of two-dimensional radiographs to examine both external and internal long bone cortical dimensions, physical cross-sections were removed, leading to irreparable damage of the bone. As CT technology has become more available it has been used to examine the distribution of both outer and inner bone dimensions of the cortex of long bones (Sumner et al. 1985; O'Neill and Ruff 2004). It is possible to estimate robusticity from outer dimensions alone, but the examination of the internal dimensions of the medullary cavity is an important factor

when examining the variable cortical dimensions that tend to be found among juveniles and older adults (Stock and Shaw 2007; Macintosh et al. 2014; Kurki et al. 2022).

Once a cross-section has been isolated, it must be processed in order to produce a quantified CSG measurement. One of the most common programs used to measure cortical CSG is ImageJ (Bourne and Bourne 2010), using the plugin BoneJ (Doube et al. 2010), wherein CSG is calculated using the Slice Geometry function. Prior to using the Slice Geometry function to quantify the CSG, the image must be converted into a binary image, where all cortical bone pixels are represented by either black or white, and all non-cortical bone pixels are represented by the opposite colour. In order to convert the image to binary, a threshold that divides each pixel into either bone or background must be determined. This threshold can be determined manually by the user, or automatically using one of several segmentation algorithms. The density, shape, and distance from the center of the frame of the object in the CT scans can affect which algorithms are best suited to segmenting the image into a binary image.

Bones situated in tissue (wet bone) and archaeological samples of bone (dry bone) are examined using CT imaging, but there has been little attention paid to how the different properties of these tissues might affect the comparability of measurements taken from each. Corron and colleagues (2017) conducted research on measurements of the same landmarks taken on physical and virtual dry bone and found reliability to be over 95% (Corron et al. 2017). Saers and colleagues (2021) examined trabecular bone in low resolution CT scans taken from living individuals and micro-CT scans taken from dry bone specimens and found that variables such as bone mineral density and trabecular patterning were not comparable. They also noted that archaeological specimens were subject to taphonomic changes that would complicate attempts to quantify bone mineral density (Saers et al. 2021). There have been no comparisons of the reliability of cortical bone cross-sectional geometry measurements taken on wet and dry bone. Archaeological specimens are prone to literally drying out, losing the periosteal tissue layer and other non-bone, cartilaginous components, whereas wet bone will still contain these components, in addition to other tissues. These tissues are visualized within a CT as additional grey scale images surrounding the bone that may impact how a researcher (or a segmentation algorithm) chooses what is or is not bone due to the increased range of intermediate values between bone and air. When attempting to quantify cortical robusticity measures in imaging programs like ImageJ, it is possible that

factors like the reliability and replicability may be impacted by the type of bone being analyzed.

The purpose of this study is to examine the reliability of dry and wet bone cross-sectional CA measurements through an examination of intraobserver error, in the form of absolute technical measurement of error (AbTEM), percent technical error (%TEM), and reliability coefficients (CR), between three rounds of measurement using both manual and algorithmic segmentation methods. There are two major questions that arise from this project; 1) does algorithmic segmentation produce cortical data comparable to those from manual measurements, and 2) is there a difference between the intraobserver error among the CA values produced from dry and wet bone CT images? The study used two image segmentation algorithms in the assessment of algorithmic error: the ImageJ default (Ridler and Calvard 1978) and Otsu (1979). The dry bone sample is derived from the Luís Lopes Documented Skeletal Collection from Lisbon, Portugal, and the wet bone sample is derived from a pre-autopsy CT collection from the New Mexico Decedent Image Database (NMDID), USA. The goal of this study is to examine the intraobserver error among algorithmic and manual methods for the dry and wet bone samples, in order to establish the comparability of any cross-sectional measurements taken from each sample.

2.3. Materials and Methods

2.3.1. Materials: Dry and Wet Bone Sample Populations

The dry bone sample was taken from the Luís Lopes Documented Skeletal Collection, housed in the Nation Museum of Natural History and Science in Lisbon, Portugal. The collection consists of disinterred individuals, who lived in Lisbon during the early and mid 20th century. The condition of the collection, despite being a historical one, is like the dry bone that is present in a disinterred archaeological specimen. The wet bone sample was taken from the pre-autopsy CT scan collection of the NMDID, USA.

The femora of twenty individuals (10 from each collection) were selected for the total sample, ranging in age from 1 to 19 years of age (Table 2.1). An attempt was made to match the samples in age composition as closely as possible. To get a consistent area of cortical bone in each section, the cross-sections of the femoral diaphysis at

45.5% of total diaphyseal length were used. This position in the juvenile femur was chosen based on the protocol outlined by Ruff (2003), wherein the midway point of the fully fused adult femur is best approximated at 45.5% from the distal end of the unfused juvenile diaphysis. In cases where the epiphyses were fully fused, the length of the diaphysis was estimated based on an examination of both the minimum and maximum intensity slab projection of the femur, wherein only the lowest and highest values for each voxel of the volume and thickness values for slice integration are displayed. This contrast helps to compare the fusion of the bone at internal and external view of the bone.

Table 2.1 Age and sex distribution across the dry and wet subsamples.

Individual's Collection Number	Dry bone sample (Lisbon) age (years)	Sex	Individual's Collection Number	Wet bone sample (New Mexico) age (years)	Sex
375	0.003	M	254	0.1	M
561	1.5	F	191	0.4	F
1581	3.1	F	143	3.0	M
371	4.2	M	56	4.1	M
1658	7.1	F	276	6.7	F
574	9.8	M	130	9.9	M
1665	11.1	M	296	10.9	M
380	15.1	M	307	12.8	F
1671	17.5	F	376	17.4	F
1132	19.9	M	407	19.6	M
MEAN	8.9			8.5	
SD	6.9			6.8	

The dry bone sample scans were collected at the Imagens Medicas Integradas Clinic in Lisbon, using a Siemens SOMATOM Flash 128-row which had a spatial resolution of 0.33 mm and a temporal resolution of 75 ms. Slice thickness is 0.6 mm, with an overlap of 0.3 mm, each DICOM stack is comprised of roughly 1500 slices, and the field of view is 200 mm. Each DICOM file stack consists of the femur, tibia, humerus,

ulna, the seventh cervical vertebrae, and fifth lumbar vertebrae from two individuals, meaning there were twelve bones scanned in total per stack. The bones were placed on a Styrofoam platform to better mimic the position of bone *in situ*, placing them closer to the center of the field of view. The wet bone sample scans were collected by the Office of the Medical Examiner of New Mexico. They were taken with a Philips Brilliance Big Bore16-slice CT scanner which had a spatial resolution of 0.5 mm and a temporal resolution of 53 ms. Slice thickness is 1 mm with 0.5 mm overlap, and each individual is comprised of roughly 10,000 slices. One DICOM stack consists of the entire body of the deceased individual, and the field of view was 180 mm (300mm for the lower extremities if they were captured separately). Both the Lisbon and New Mexico DICOM stacks have 512 x 512 voxel matrices, both were reconstructed as bone models by the technicians, and both CT scanner's resolution fell at or below 0.5 mm. Despite the slight difference in slice thickness and overlap, the other comparable scan parameters, such as the position of the bones in the scanner and the reconstruction of the bone models, mean the DICOM sample stacks are comparable as not only are these differences minor, but data extracted from both sets of stacks are two dimensional, and this doesn't rely on the frequency and spacing between slices. The dry and wet bone samples are both standard low resolution CT scans and the difference between them should not impact the Hounsfield Units or the voxel size of the scans. Hounsfield units and voxel size affect the quantification of cortical bone, meaning that the composition scans themselves should not impact data that are being segmented to calculate the CA.

2.3.2. Methods for Measuring Cortical Area

The femora were aligned in the software program Dragonfly in accordance with the protocol set out by Spake and colleagues (2020), wherein long bones were positioned along sagittal and coronal planes.. The maximum length of the femoral diaphysis was taken, and transverse cross-sections were exported at 45.5% of the total length, taken from the distal end (Ruff 2003). The cross-sections were then isolated as 16-bit TIFF image files, and processed in the program ImageJ, using the BoneJ plugin.

Slice Geometry function of BoneJ (Doube et al. 2010) was used to calculate the CA of each femoral cross-section. This process involved setting the scale of each image, which was done by matching a line drawn in ImageJ with a 10 mm scale bar in the original TIFF image. The orientation of the cross-sectional image was then set

anteroposterior and mediolateral planes, based on the planes present in the TIFF image that were set in Dragonfly. The BoneJ plugin uses these parameters to similar to the scale function, wherein the program is them able to produce data along mediolateral and anteroposterior planes. In order for CA to calculated by the Slice Function of the BoneJ plugin each image was converted to an 8-bit pixelation and then segmented into a binary, black and white image.

Segmentation of bone from either air or tissue can be created by using algorithms or by manually selecting the bone area. When using the auto-threshold function to transform the images into binary format, an algorithm must be used to differentiate bone from non-bone pixels in the image. The two algorithms that were determined to be best suited to successfully segmenting bone from non-bone were the default (or IsoData) (Ridler and Calvard 1978) and the Otsu (Otsu 1979) discriminant function algorithms (see Appendix A). These algorithms were chosen from 13 other algorithms based on a visual assessment of which best differentiated the white bone from the black background. Previous studies that calculated CA on adult bone using the Slice Geometry function have used the segmentation algorithm of Ridler and Clavard (1978) (i.e., the default algorithm) (Mactinosh et al. 2013; Gosman et al. 2013), The default and Otsu algorithms consistently produced the most visually accurate segmentation of bone from non-bone.

The default algorithm differentiates bone from non-bone by taking an initial threshold from a gray-scale histogram, then averaging the pixels at or below the initial threshold and the pixels above an initial threshold. Then the average of those two values are taken to create a new threshold, and the process is repeated until the resulting threshold is larger than the composite average (Ridler and Calvard 1978). The Otsu algorithm differentiates bone from non-bone (or background from object) by separating the pixels based on a discriminant function derived from a gray-scale histogram, wherein a threshold is determined that will minimize intraclass variation (Otsu 1979). Lastly, the manual segmentation between bone and non-bone was made by selecting the bone parameters using a selection wand to outline the periosteal and endosteal surfaces. In this sense, it is akin to a local segmentation method when compared with the global segmentation used by the algorithmic methods. The wand algorithm is choosing whether adjacent pixels are included in the selection. Within this paper the local segmentation will be referred to as “manual”, as this is as much control the user can exert over

segmentation within the program. When calculating the CA for the images where segmentation was conducted manually the area of the internal endosteal surface was subtracted from the area of the whole cross-section. This was done because two separate masks had to be created to differentiate the bone from the background, due to how BoneJ deals with annular forms. After a threshold was determined, either by algorithmic or manual means, a binary file of the shape was created (Figure 2.2).

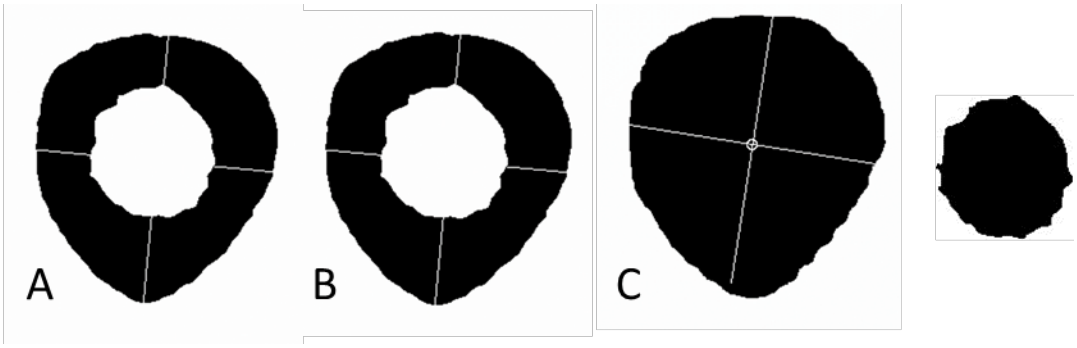


Figure 2.2 Images of cross-sections from the same femur made binary using the default algorithm (a), Otsu algorithm (b), and manual (c) segmenting methods.

After the binary file has been created, CA can be calculated by the BoneJ plugin using the Slice Geometry function.

2.3.3. Intra-Observer Error Analysis

The entire process, from measuring the length of the femoral diaphysis in the CT scan using Dragonfly to segmenting the cross-sectional images in Image J, was conducted a total of three times. It should be noted that user skill with the software likely improved after each round. Additionally, there was a gap of a few months between when the measurements were taken for the first round and when measurements were taken for the second and third rounds, which were taken within a matter of days from each other. Each specimen in the sample had a separate cross-section taken at 45.5% of diaphyseal length in Dragonfly three times. Each of these three cross-sections was then processed in ImageJ, wherein they were each subjected to the default, Otsu, and manual segmenting methods, which differentiated bone from non-bone in the image. CA was then calculated from each binary cross-section. This resulted in the production of 30 CA measurements calculated in total for each round, 10 from each segmenting method.

Intraobserver error comparisons were made between rounds 1 and 2, 2 and 3, and 1 and 3 of the wet and dry bone samples (see Table 2.2). Error and reliability between each segmenting method was assessed by comparing the CA values taken during each round (see Table 2.3).

Table 2.2 Comparisons run between the CA measurements from each round during the intra-observer error analysis. Three rounds of measurements were taken from the same cross-sectional image using each of the three segmenting methods, meaning in total nine measurements were taken per slice image.

Default	Otsu	Manual
Rounds 1 and 2	Rounds 1 and 2	Rounds 1 and 2
Rounds 2 and 3	Rounds 2 and 3	Rounds 2 and 3
Rounds 1 and 3	Rounds 1 and 3	Rounds 1 and 3

Table 2.3 Comparisons run between the CA measurements taken using each segmentation method during the reliability analysis. Three rounds of measurements were taken from the same cross-sectional image using each of the three segmenting methods, meaning in total nine measurements were taken per slice image.

Round 1	Round 2	Round 3
default and Otsu	default and Otsu	default and Otsu
default and manual	default and manual	default and manual
Otsu and manual	Otsu and manual	Otsu and manual

The analysis of the intra-observer error included the calculation of the mean difference between the CA measurements, and the mean absolute difference between the measurements, the technical error of measurement (TEM), and the percent technical error. Technical error of measurement provides an absolute measure of error in mm^2 and is calculated by finding the square root of the summed error values after the sum has been divided by the sample size multiplied by the number of measurement rounds (Ulijaszek and Kerr 1999). Reliability of measurements provides a measure of the between-individual variance, is free of measurement error, and is used to calculate the coefficient of reliability (CR) (Ulijaszek and Kerr 1999). The percent technical error is calculated using the CR. Paired t-tests were used to assess error and reliability for each segmenting method, and to compare the means of each segmentation method within the same observer round.

2.4. Results

When averages were calculated for the CA values collected during each round of testing, generally differences were largest between round 1 and round 3 of testing (Table 2.4). When the default algorithmic method was used, the wet bone sample (New Mexico) had a difference of average CA of 7.58 mm^2 between rounds 1 and 2, 2.04 mm^2 between rounds 2 and 3, and 5.54 mm^2 between rounds 1 and 3. In the dry bone sample (Lisbon), the default method had a difference of average CA of 5.19 mm^2 in between rounds 1 and 2, 3.79 mm^2 in between rounds 2 and 3, and 8.98 mm^2 in between rounds 1 and 3. The average Cas obtained using the Otsu method difference of 7.76 mm2 between rounds 1 and 2, 2.14 mm^2 between rounds 2 and 3, and 5.56 mm^2 between rounds 1 and 3 for the wet bone sample, and had a difference of average CA of 3.01 mm^2 in between rounds 1 and 2, 4.26 mm^2 in between rounds 2 and 3, and 7.27 mm^2 in between rounds 1 and 3 for the dry bone sample. The average Cas for the algorithmic segmenting methods were similar to the manual method averages (Table 2.4). Average Cas obtained using the manual method difference of 4.59 mm^2 between rounds 1 and 2, 5.46 mm^2 between rounds 2 and 3, and 10.05 mm^2 between rounds 1 and 3 for the wet bone sample and had a difference of average CA of 0.7 mm^2 in between rounds 1 and 2, 46.77 mm^2 in between rounds 2 and 3, and 6.07 mm^2 in between rounds 1 and 3 for the dry bone sample. The minimum CA were lower and the maximum CA values were higher in the wet bone samples, resulting in a larger standard deviation in the wet bone samples through all three rounds. Average dry bone areas were smaller than wet for all segmentation methods and standard deviations were lower.

Table 2.4 Cross-sectional CA measurements in mm^2 per rounds 1, 2, and 3 of measurement using the default (D), Otsu (O), and Manual (M) segmenting methods for the dry (Luís Lopes) , and the wet (New Mexico Office of the Medical Examiner) bone samples.

	Age	D round 1	D round 2	D round 3	O round 1	O round 2	O round 3	M round 1	M round 2	M round 3
Dry Sample										
MEAN	8.94	175.59	180.78	184.57	183.42	186.43	190.69	179.53	180.23	173.46
SD	6.94	113.50	117.61	120.20	119.50	120.78	123.47	121.00	120.81	119.16
Wet Sample										
MEAN	8.50	208.37	215.95	213.91	205.55	213.31	211.17	211.25	215.84	221.30
SD	6.81	166.57	167.08	165.89	167.29	167.68	166.30	169.42	173.72	177.30

2.4.1. Intraobserver Error

To examine the reliability of the measurements taken from the dry bone sample and the wet bone sample the intraobserver error was determined by calculating absolute and percent technical measurement of error (AbTEM), percent technical error (TEM%), the coefficient of reliability (CR), the mean difference (MD), and mean absolute difference (MAD) (values were calculated in accordance with Ulijaszek and Kerr 1999). Paired t-tests were also used to determine if there were significant differences between the total cross-sectional area (CA) measurements from each observation round both between rounds, and within rounds.

Table 2.5 Intra-observer error analysis of the CA between 3 rounds of measurements. The absolute technical measurement of error (AbTEM), percent technical measure of error (TEM%), coefficient of reliability (CR), mean difference (MD), mean absolute difference (MAD), and paired t-test p values are given for each method.

Default						
	AbsTEM	TEM%	CR	MD	MAD	P
Wet Bone Sample						
Round 1 and 2	6.982	0.026	0.9997	-7.578	7.578	**0.006**
Round 2 and 3	3.071	0.012	0.9999	2.036	3.076	0.146
Round 1 and 3	5.350	0.020	0.9998	-5.542	6.120	**0.010**
Mean	5.134	0.020	0.9998	-3.695	5.591	
Dry Bone Sample						
Round 1 and 2	4.336	0.034	0.9997	-5.192	7.024	0.052*
Round 2 and 3	3.867	0.029	0.9997	-3.787	4.913	0.126
Round 1 and 3	6.405	0.049	0.9995	-8.979	10.305	**0.016**
Mean	4.869	0.038	0.9996	-5.986	7.414	
Otsu						
	AbsTEM	TEM%	CR	MD	MAD	t-test alpha
Wet Bone Sample						
Round 1 and 2	7.062	0.027	0.9997	-7.759	7.759	**0.005**
Round 2 and 3	3.099	0.012	0.9999	2.134	3.060	0.129
Round 1 and 3	5.445	0.021	0.9997	-5.625	6.417	**0.011**
Mean	5.202	0.020	0.9998	-3.750	5.745	
Dry Bone Sample						
Round 1 and 2	3.562	0.032	0.9997	-3.011	4.877	0.086
Round 2 and 3	5.445	0.047	0.9995	-4.260	5.044	0.078
Round 1 and 3	7.283	0.063	0.9994	-7.271	8.499	**0.018**
Mean	5.430	0.047	0.9996	-4.847	6.140	
Manual						
	AbsTEM	TEM%	CR	MD	MAD	t-test alpha
Wet Bone Sample						
Round 1 and 2	5.157	0.018	0.9998	-4.598	8.222	0.169
Round 2 and 3	5.778	0.020	0.9998	-5.455	9.165	0.143
Round 1 and 3	6.864	0.024	0.9998	-10.053	11.139	**0.010**
Mean	5.933	0.021	0.9998	-6.702	9.509	
Dry Bone Sample						
Round 1 and 2	3.395	0.025	0.9998	-0.699	6.759	0.764
Round 2 and 3	5.679	0.042	0.9996	6.770	9.644	0.053*
Round 1 and 3	6.814	0.050	0.9995	6.071	10.930	0.170
Mean	5.296	0.039	0.9996	4.047	9.111	

Significant results are indicated in bold (α < 0.05)
*Indicates a near significant result

Intraobserver Error Among the Default Measurements

Within the analysis of intra-observer error in the measurements taken using the default algorithm it was found that absolute TEM between rounds for the total sample (dry and wet bone) was low, falling roughly between 4 and 8 mm^2. Percent technical error was also low, between 0.02% and 0.05%. The coefficient of reliability was highest between rounds 2 and 3 (0.9999). The mean differences and absolute mean differences were also lowest between rounds 2 and 3 (Figure 2.3). When paired t-tests were used to determine if significant differences were found between CA measurements in rounds 1 and 2 and rounds 1 and 3, indicating that round 1 measurements were the most different compared with the measurements taken in the other 2 rounds.

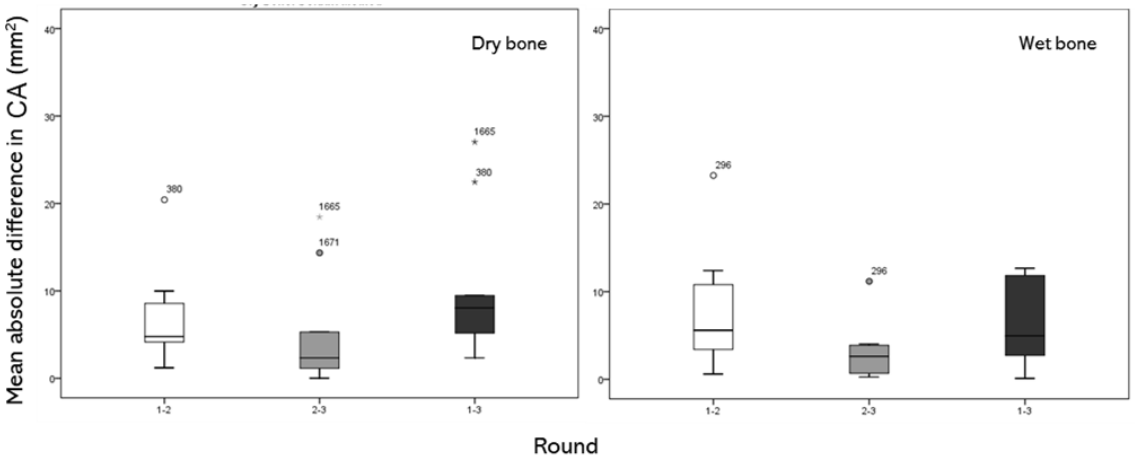


Figure 2.3 Mean absolute difference in CA measurements (mm^2) between rounds 1 and 2 (white), rounds 2 and 3 (grey), and rounds 1 and 3 (black) for the default segmenting algorithm. Dry bone sample is on the left, wet bone sample is on the right. Specimen numbers are listed next to outlier differences.

When examining the dry and wet bone CA measurements taken using the default algorithm, a similar pattern emerges to what was seen in the total sample, where differences are more pronounced between measurements taken in round 1 and 2, and rounds 1 and 3 (Table 2.5). Overall, the mean AbTEM is lower in the dry bone sample, and the mean coefficient of reliability between each round of measurement is higher in the wet bone sample than the dry bone sample, though not by much. Mean differences

and mean absolute differences were comparatively smaller between rounds for the wet bone sample, but there were higher mean differences in the dry bone sample relative to both the total sample and the wet bone sample.

Intraobserver Error Among the Otsu Measurement

The analysis of intraobserver error from the measurements taken using the Otsu algorithm was very similar to the analysis of the default algorithm. Absolute TEM between rounds for the total sample fell roughly between 4 and 7 mm^2. Percent technical error was also low, between 0.01% and 0.07%. The coefficient of reliability was highest between rounds 2 and 3 (0.9998). The mean differences and absolute mean differences were also lowest between rounds 2 and 3 (4.032 mm^2). There were determined to be significant differences between CA measurements in round 1 and 2 and rounds 1 and 3, indicating that round 1 measurements were the most different compared with the measurements taken in rounds 2 and 3.

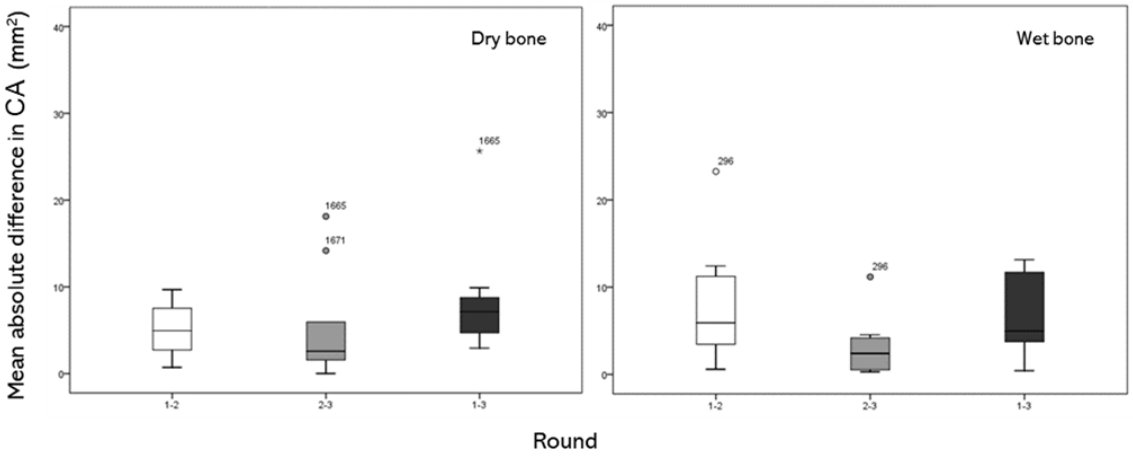


Figure 2.4 Figure 2.4 Mean absolute difference in CA measurements (mm^2) between rounds 1 and 2 (white), rounds 2 and 3 (grey), and rounds 1 and 3 (black) for the Otsu segmenting algorithm. Dry bone sample is on the left, wet bone sample is on the right. Specimen numbers are listed next to outlier differences.

When examining the dry and wet bone CA measurements taken using the Otsu algorithm, a similar pattern emerges to what was seen in the total sample, where differences are more pronounced between measurements taken in round 1 and 2, and rounds 1 and 3 (Figure 2.4). The AbTEM were not consistent across the samples or the

rounds of observation (Table 2.3). Coefficients of reliability were slightly higher in the wet bone sample (0.9997–- 0.9999) than in the dry bone sample (0.9994–- 0.9997). Mean differences and mean absolute difference were comparatively smaller between rounds 1 and 3, and 2 and 3 for the wet bone sample, but were larger between rounds 1 and 2, when compared with the dry bone sample. As with the results when using the default algorithm, the average mean absolute difference was higher in the dry bone sample than in the wet bone sample when using the Otsu algorithm.

Intraobserver Error Among the Manual Measurements

When examining error in the manually segmented measurements, the most variation appears to be between rounds 1 and 2 within the total sample, the wet bone sample, and the dry bone sample (Table 2.5). The AbTEM and %TEM were similar between the wet and the dry bone samples. The coefficients of reliability were slightly higher in the wet bone sample, and the mean differences and mean absolute differences were consistent between the wet and the dry bone sample (Figure 2.5). The only case where there was a significant difference between the CA measurements was in the wet bone sample between rounds 1 and 3.

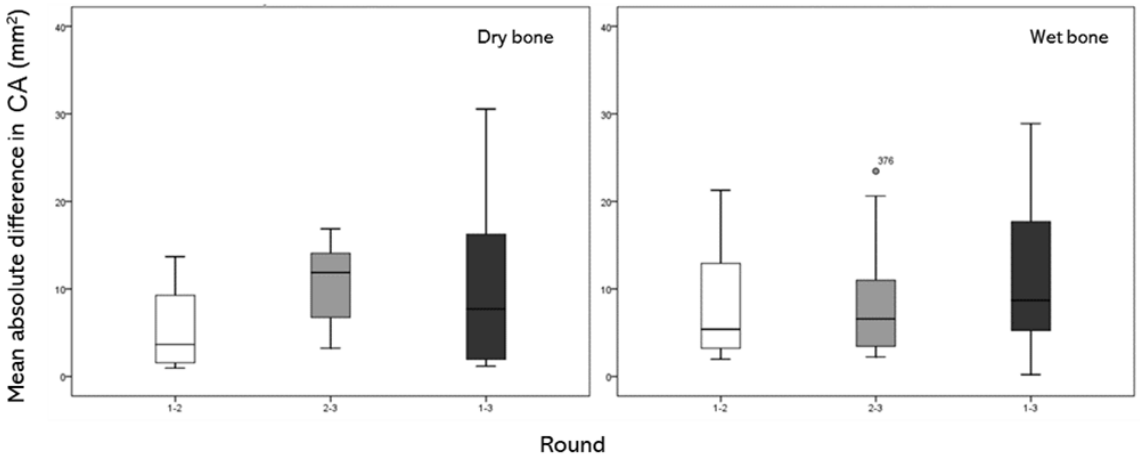


Figure 2.5 Mean absolute difference in CA measurements (mm^2) between rounds 1 and 2 (white), rounds 2 and 3 (grey), and rounds 1 and 3 (black) for the manual segmenting method. Dry bone sample is on the left, wet bone sample is on the right. Specimen numbers are listed next to outlier differences.

2.4.2. Manual Cross-sectional Measurements Compared to Algorithmic Segmentation Measurements

When comparing the CA measurements that were taken using the manual method to those taken using algorithmic methods within each observer round, the measures taken from the wet bone sample appear to be more similar to the manual measurements than the measurements taken from the dry bone sample (Table 2.6). When the samples are combined into the total sample there does not appear to be any significant difference between the manual and the default methods, nor the manual and the Otsu methods. The mean absolute difference between measures is between 5 and 15 mm^2 in each round.

Table 2.6 Comparison of CA values determined using manual and two different algorithmic segmentation methods (default and Otsu). Mean difference (MD), and mean absolute difference (MAD) were calculated, and paired t-tests were conducted to determine if there were significant differences between Cas for each segmentation method. Significant results are bolded.

	Manual v. default			Manual v. Otsu		
	MD	MAD	t-test alpha	MD	MAD	t-test alpha
Wet bone sample						
Round 1	-2.876	7.49	0.357	-5.700	7.543	0.059
Round 2	0.104	7.268	0.976	-2.538	7.072	0.396
Round 3	-7.387	12.721	0.124	-10.127	12.897	**0.036**
MEAN	3.456	9.160		6.122	9.171	
Dry bone sample						
Round 1	-3.943	9.601	0.306	3.886	7.698	0.186
Round 2	0.550	4.130	0.763	6.198	6.198	**0.001**
Round 3	11.107	11.107	**0.001**	17.228	17.228	**0.000**
MEAN	5.200	8.280		9.104	10.374	

Significant results are indicated in bold ($\alpha < 0.05$)

The wet bone sample demonstrates a similar pattern to the total sample. The mean absolute difference is low (between 8 and 12 mm), and there are only significant differences between the manual and Otsu methods in the 3rd rounds of measurements (Figure 2.6).

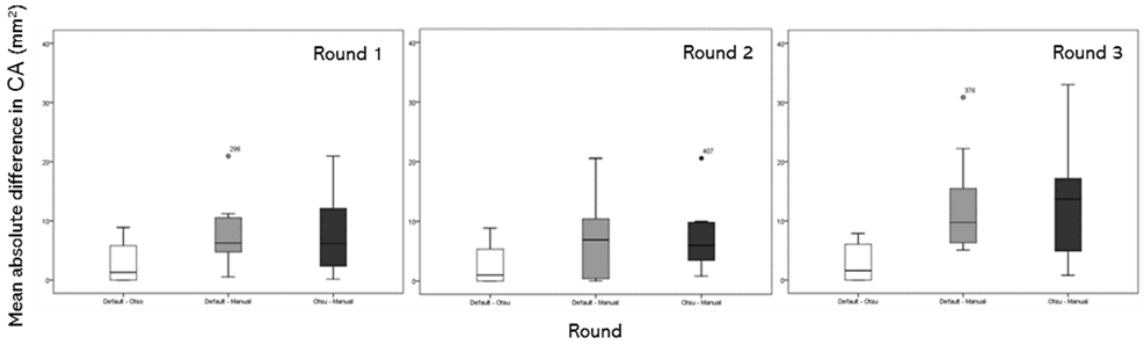


Figure 2.6 Mean absolute difference in CA measurements (mm^2) between default and Otsu (white), rounds default and manual (grey), and rounds Otsu and manual (black) for the wet bone sample. Round 1 is on the left, round 2 is in the middle, and round 3 is on the right. Specimen numbers are listed next to outlier differences.

The dry bone sample also has a low mean absolute difference between manual and algorithmic measurements for each round, but there were multiple cases of significant differences between manual and algorithmic measurements (Figure 2.7). Round 2 of the manual and Otsu, and round 3 of both the manual and default, and manual and Otsu all demonstrated significant differences between the measurements.

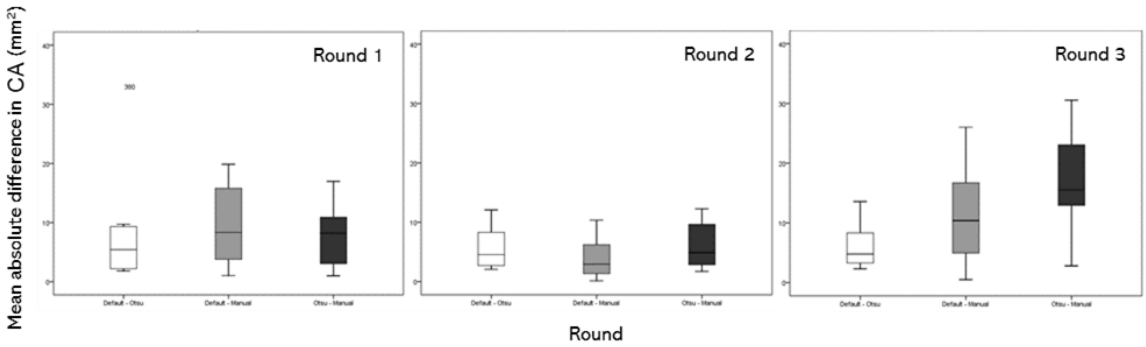


Figure 2.7 Mean absolute difference in CA measurements (mm^2) between default and Otsu (white), rounds default and manual (grey), and rounds Otsu and manual (black) for the dry bone sample. Round 1 is on the left, round 2 is in the middle, and round 3 is on the right. Specimen numbers are listed next to outlier differences.

2.5. Discussion

There is an inherent lack of research on CT imaging protocols and procedures within biological anthropology (Elliot 2022), let alone investigations of the differences

involved in taking measurements from wet and dry bone. This research attempts to examine CT usage for wet and dry bone through the examination of intra-observer error. When considering the first question posed in this study, to determine if data derived from algorithmic methods is comparable to manually segmented data, the results demonstrate that reliability, and TEM percentage were uniform across both automated methods and the manual method (0.999, and <1%, respectively (Table 2.5)).

Determining which of the segmentation algorithms produced CA values most like the manual method can be accomplished by examining the average mean absolute differences (MAD) between each of the methods across the rounds. In the dry bone sample the average MAD between the default and the manual method was roughly 8 mm^2 and the average MAD between the Otsu and the manual method was roughly 10 mm^2, whereas for the wet bone samples the average MAD between each algorithmic method and the manual method was around 9 mm^2. There was one instance of significant differences between the measurements from the manual and default methods and three instances of significant differences between the Otsu and the manual methods. Overall, most differences were not significant and there were no major differences between the CA values calculated by the algorithms and the manual method in either the dry or wet bone samples. Ideally, the gold standard for measuring cross-sectional would be taking physical measurements of the cross-section, but when using BoneJ to calculate CA from CT images, the local, wand tool approximates this process the closest, and thus is referred to as he manual method. The manual method allows the user to make the most specific decisions about which area would be considered cortical bone and which area would be background/non-bone within ImageJ. However, the manual method is highly time consuming, and unrealistic in terms of use for segmenting hundreds of slices. Therein lies the utility of the algorithms, and in accordance with the results found in this study the algorithms (default and Otsu) produce CA measures that are very similar to the manual method, in both the dry and wet bone samples.

The second question was if there is a difference in manual and algorithmic measurements between the dry and wet bone samples. The results suggest that there are more instances of significant differences in intraobserver error among the wet bone samples than the dry bone samples, but fewer instances of significant differences between each segmentation method among the observer rounds in the wet bone

sample. However, all coefficients of reliability (CR) produced were greater than 0.95, and thus fall within an acceptable range for intraobserver error (Ulijaszek and Lourie 1994).

Intraobserver error between the rounds was higher in the wet bone samples, which would potentially align with the increased variability in pixel grey scale associated with a bone sitting in situ, as opposed to a dry bone surrounded by empty space. Error was highest between rounds 1 and 3 of CA measurements, both in terms of MAD in CA and the presence of statistically significant differences in measurements between rounds. The wet bone samples demonstrated statistically significant differences between rounds 1 and 2, in addition to between rounds 1 and 3. This is likely due to the observer becoming more proficient in isolating the scans as the rounds increased.

While differences between the dry and wet bone samples were small, there were still instances of significant differences between rounds of CA measurements taken between the two samples. There are several potential reasons for this difference. The first set are the limiting factors of the sample that was used for the study.

A major limiting factor in this research is that these data are taken from two different dry and wet bone samples, meaning that while attempts were made to create comparable samples based on known age and sex, these are not the same individuals in dry and wet form. This means that there is no way to directly compare the CA values obtained for an individual from a CT scan of wet bone to a CT scan of that same bone dry. In the future it would be beneficial to conduct animal testing (similar to that done on trabecular bone by Saers and colleagues (2021)), wherein animal bone is CT scanned in situ, then de-fleshed, and CT scanned as dry bone in order to remove some of the limiting factors present in the study. In this scenario it would be possible to mimic how an archaeological bone would differ from in situ bone, and how that removal of the flesh impacts how cortical bone is visualized by CT imaging. The timing between rounds of observation and subsequent improvements in observer command of the Dragonfly and ImageJ software should also be taken into consideration, especially when comparing the CA values obtained in the first round of measurement with those from the second and third round. Since the measurements were being retaken in Dragonfly during each round, it is possible that some of the error between rounds is occurring at the measurement stage, and not at the segmentation stage. Thus, it is important to look at error rates from the same round, as well as between rounds of observation. Future

research should also consider conducted interobserver error analyses to examine the effects of multiple observers on the error rates and reliability coefficients for wet and dry bone and should incorporate individuals with differing familiarities with CT imaging software and ImageJ.

There is also a potential possibility for error to be introduced due to samples consisting of subadults, who are still experiencing appositional growth of the cortex. Cortical bone will not have been fully developed and would have been subjected to different levels of bone modelling depending on the stage of growth and development the individual was in (Lampl et al. 1992; Johnson et al. 1996; Swan et al. 2020). This differential development could impact the bone margin on the scan making it more difficult to determine where the cortical bone ends. However, the sections were taken from the mid-shaft, where there is less trabecular bone relative to cortical bone, and not the end of the diaphysis or the metaphyses, where there is more trabecular bone and higher instances of trabecular bone modeling. Therefore, bone modeling should not account for a large amount of variation between the segmentation methods.

There are other potential reasons for the differences observed between wet and dry bone samples that are more related to the dry and wet nature of the bones themselves. The algorithms may have an effect on the CA measurements. As seen when comparing the different segmentation methods within each observer round, there were fewer significant differences between the manual and algorithmic methods within the wet bone samples than the dry bone samples. This means that, on average, the CA measures taken using the manual and algorithmic methods were more similar among the wet bone samples than the dry bone samples. The differences were small but could be due to the nature of dry bone. As mentioned previously the bones in the dry sample have undergone a literal drying out process. Changes to the cortical bone could have occurred both through taphonomic processes that occurred when the individual was buried and the process of being disinterred and stored in a museum. Decomposition and storage in warm, dry locations can make bones more prone to fractures, which could impact observed cortical area (Karr and Outram 2012). The burial and decomposition process can have a disruptive effect to the periosteal surface of the bone, through damage caused by roots or rodents (Pokines 2016), while interment in a museum can result in wear on or damage to the periosteum through shelf wear and constant handling (Pokines et al 2017). Bone in situ contains more water than dry bone, making it less

prone to relaxation (Sasaki and Enyo 1995), but stronger and tougher by comparison (Nyman et a. 2006). Simply put, there may be more variation for the segmentation methods to detect on the dry bone, as the surface may not be as intact as wet bone. Additionally, the process of bone diagenesis could involve drying and warping of the cortex. Alternatively, there is more contrast between the background and the bone in the dry sample, as there is no tissue surrounding the bone like there is in the wet sample, meaning there is less contrast in the wet than dry bone. This could be why there was less intraobserver error in the dry sample, but more differences between the CA measurements calculated using different segmentation methods within the rounds. Future research should also consider conducting interobserver error analyses to examine the effects of multiple observers on the error rates and reliability coefficients for wet and dry bone.

Another aspect to consider when comparing the dry and wet samples is the nature of the endosteal surfaces in each. The boundary between cortical bone and other types of bone and tissue at the endosteal surface is less clear in the wet bone sample, whereas in the dry bone sample any other tissue that is present besides cortical bone, and potentially a small number of trabeculae, have decomposed. The potential error introduced by differences in endosteal surfaces of the dry and wet bone samples, while present, is lessened when using 0.33 to 0.5 mm resolution CT scans, as these types of scans are not generally as useful in visualizing trabecular bone, unlike micro-CT images. Therefore, it is unlikely that preservation would result in large enough differences in cortical bone on the endosteal surfaces to create significant differences between dry and wet bone samples.

These results are likely applicable to other parts of the bone or to other archeological or present-day populations. These sections were taken from 45.5% of the diaphyseal length, which is an area of the femur where the line between the cortical bone and the medullary cavity is mostly clear, whereas when examining sections of the bone closer to the epiphyses, where there is more trabecular bone, the error between the rounds and the different segmentation methods may be even more pronounced. In terms of other bones, similar results might be expected in other long bones, whereas in vertebrae or other skeletal elements, the results may be even more subject to error.

As mentioned previously, it is important to consider the limitations of this study when drawing conclusions about the applicability of the results to other specimens or collections. In the future, researchers should look to create a similar error analysis between dry and wet bone samples, using the same individual specimen *in situ* then de-fleshed and dried out to mimic archaeological conditions. This process would allow for a more definitive analysis of the difference in quantifying cortical bone area in dry and wet bone samples using CT scans in ImageJ.

2.3 Conclusion

This paper examined measurements of cross-sectional cortical area taken from 45.5% of juvenile diaphyseal femur length. The process by which the measurement is calculated from a two-dimensional image can be altered, thus there is potential for error to occur within the creation of the cross-section and the calculation of the cortical area. Error and reliability of measurements were examined between 3 rounds of measurement, and between three segmentation methods in a dry and wet bone sample. There were two questions posed by the study; the first being whether data derived from algorithmic methods is comparable to manually segmented data. The results demonstrate that reliability, and TEM were uniform across both automated methods and the manual method. The second question was is there a difference in manual and algorithmic CA measurements between the dry and wet bone samples? The results suggest that there are more instances of significant differences in intra-observer error among the wet bone samples than the dry bone samples, but fewer instances of significant differences between each method among each observer round in the wet bone sample. The results of the study suggest that the use of automated algorithms is comparable to manual methods of segmentation when creating cross-section images for processing within ImageJ (all coefficients of reliability were above 0.95, in accordance with Ulijaszek and Lourie 1994). The results also suggest that error rates between rounds of measurements taken from dry and wet bone may not differ hugely when taken from comparable low resolution CT scans, but that further research using the same bone samples, both in dry and wet form, is needed to confirm which differences exist between cortical area measurements taken from the same specimen.

Chapter 3. Novel formulae for estimating weight in children using appositional and longitudinal long bone measurements derived from a modern New Mexico sample.

3.1. Abstract

The two most prominently used weight estimation formulae using skeletal parameters for subadult individuals are based on the same 20 individuals from the Denver Growth Cohort (Ruff 2007; Robbin et al. 2010). These formulae use femoral dimensions to estimate body mass among children from many different geographic regions and time periods and are often used in forensic contexts. This sample limitation is mainly due to the availability of data, wherein during the study period only 20 individuals had weights recorded and radiograph taken at 6-month intervals. The consistent use of the data derived from the Denver Cohort indicates the lack of childhood skeletal growth data from other sources, particularly when it comes to estimating weight in obese children (above the 95th percentile for BMI). The following study develops weight estimation formulae from a computed tomography (CT) sample of 21st century subadult individuals (n=77) with known weights and ages at death using metaphyseal breadth, epiphyseal breadth, and torsional rigidity (J). Weight estimates created using the formulae developed in the study are then compared with known weight to examine error. The results suggest that mid-shaft femoral J values and proximal tibial J values correlate highest with body weight, and when used to develop weight estimation formulae produce the least amount of error. The formulae developed in this study are more appropriate when attempting to estimate body mass for 21st century, obese individuals within forensic contexts.

3.2. Introduction

Weight estimation formulae based on skeletal parameters are important tools for anthropologists looking to reconstruct identity in forensic contexts and for developing demographic data in archaeological contexts. There have been many weight-for-age studies of children, but few that are foundational in the development of weight estimation equations. The Denver Growth Study (Maresh 1970) has been the sample most used in

the development of weight estimation formulae based on long bone measurements (Ruff 2007; Robbins et al. 2010; Robbins Schug et al. 2013). The prominence of the Denver Cohort in the development of weight estimation formulae is due to both the thorough nature of the sample (in terms of known weight-for-age and consistent radiographs taken throughout growth) and the lack of other comparable samples of subadult individuals with known weights and skeletal measurements. The application of these weight estimation formulae developed from the Denver Cohort is varied, from modern children in forensic contexts (Sciulli and Blatt 2008), to children found in archaeological contexts (Trinkaus 2002; Cowgill and Trinkaus 2007; Harrington 2010; Cowgill 2018), to hominin sub-adults and small-bodied adult hominins (Ruff 2007; Walker et al. 2018).

The Denver Growth Study was a longitudinal anthropometric study run from 1927 to 1967 (Maresh 1943; Maresh 1955; Maresh 1970). It was conducted using twenty American subadult individuals, ten male and ten female, who were examined every two months from birth to six months of age, and then roughly every half year until around 20 years of age. Radiographs, body mass (weight), stature (or supine length for infants), and bi-iliac (maximum pelvic) breadth were taken during each examination. The cohort consisted of individuals from in and around Denver Colorado; all were white, and most came from middle- and upper-class backgrounds (McCammon 1970). Based on the specific context in which those in the Denver Growth Cohort grew up, it is unlikely they were under caloric or protein stress, nor were there any instance of individuals being overweight or obese (Maresh 1970).

Data from this cohort were used to develop the most widely used equations for estimating weight based on long bone dimensions. One set of formulae were developed by Ruff (2007) using radiographs of ten females and ten males with nearly complete longitudinal datasets from the Denver Growth Study. For the long bones, he created weight estimation formulae based on the femoral distal metaphyseal breadth and the femoral head breadth. Robbins and colleagues (2010) also developed a set of equations for estimation age based on a subsample of the Denver Growth Study, this time focusing on measuring torsional rigidity (J) at the femoral mid-shaft. Both Ruff (2007) and Robbins and colleagues (2010) created age-specific formulas, as the relationship between bone dimensions and body mass changes during growth (Ruff 2003). These formulae rely either on the use of breadth measurements from the metaphysis or

epiphyses of the femur (Ruff 2007), or on measurements of torsional rigidity (J) at the mid-shaft of the femoral diaphysis (Robbins et al. 2010).

Since these formulae were all produced using the same sample population, their application is limited by the specifics of the place and time that those individuals lived; the majority of whom fell within "normal" weight for age for 20th century (meaning that for the time most fell within the 5th to 95th weight-for-age percentile (McCammon 1970)). This results in a potential limitation of the applicability of any formulae derived from these data on modern forensic populations. Additionally, the formulae produced from long bones only examine breadth and J in specific areas of the femur, and there has been little examination of other sections of the femur and none of the tibia. The use of age-specific formulae requires prior knowledge of age or the estimation of age (Cowgill 2018), which introduces further error. There is limited applicability of these formulae in estimating weight from long bones in populations outside of the original population from which the measurements were taken. The use of these formulae to estimate the body mass of children from modern forensic populations, who are more likely to be overweight or obese, has been questioned by research demonstrate a tendency of these formulae to underestimate weight (Spake et al. 2021).

Weight estimation equations developed from the Denver Growth cohort would fail to account for the positive secular trend in weight that has occurred in many populations of children across the world (Sun et al. 2012). Just in the United States, between 1980 and 2000 there was an increase in the prevalence of obese children, and the children who fall on the heavier end are continuing to get heavier (Troiano et al. 1995; Ogden et al.; 2008, Komlos et al. 2009). The utility of this reference sample in producing formulae to estimate the weight of individuals who fall above the "normal" BMI percentiles has yet to be examined. Yim and colleagues (2020) and Spake and colleagues (2021) examined the application of several of the formulae developed by Ruff (2007) and Robbins and colleagues (2010) and found that they consistently underestimated weight when used on a modern (21st century) sample of subadult individuals. Metaphyseal breadth-based formulae underestimated weight more severely than J-based formulae, but in both cases, underestimation increased with age in the sample. This underestimation was explained by a combination of bias introduced when transforming the logged data back into kilograms and the positive secular trend in weight observed over the 20th and 21st century in modern children (Spake et al. 2021).

In order to explore the relationship between weight and bone growth further, this study examines a sample from a modern population of subadult individuals with known weight at death and produces a novel set of weight estimation formulae. This study will produce formulae using several proximal and distal breadth measurements and torsional rigidity values taken from the femur and tibia, based on a modern forensic sample consisting of 77 individuals with a wide range of body mass indices (BMI). There are two main goals of this paper, the first is to create weight estimation formulae for a sample of modern subadult individuals aged 1 to 19 years old. The second goal is to examine the error between the weight estimates produced from the formulae and the known weight of the individuals in the sample. To complete these goals this study uses a 21st century post-mortem juvenile sample of subadult individuals of identified age, height, weight, and sex from the Albuquerque, New Mexico area. The formulae developed in this study are best suited for estimating body mass among obese children in 21st century forensic contexts.

3.3. Materials

The sample is composed of 77 individuals aged between a month old to 19 years of age, from Albuquerque, New Mexico (heretofore referred to as “the New Mexico sample”). Anonymized demographic information, medical histories, and CT scans of the children are stored at the New Mexico Decedent Image Database (NMDID), which is curated by the Office of the Medical Investigator of New Mexico and the University of New Mexico. The children’s birth years span from 1994 to 2015. The children themselves are from diverse backgrounds, including different socioeconomic and racialized groups, which reflect the contemporary population demographics of New Mexico. Cause of death also varied between accidental, natural, suicide, and homicide.

A sample of CT scans of 77 individuals was taken from the larger NMDID database. Individuals were selected based on their BMI, age, manner of death, and sex. An attempt was made to create an equal distribution of age from just after birth to 19 years of age. When selecting for BMI in the sample, individuals with BMI within, above, and below the 5% to 95% percentile were selected to mirror the distribution in the larger sample of subadults within the NMDID. Manner of death was selected to have roughly equal accidental and intentional causes of death. Accidental mainly consisting of car accidents, while intentional was a mix of homicides and suicides. Sex was selected for to

create an equal distribution of individuals assigned as male and as female (hereby referred to as male and female). An attempt was also made to create paired-sex samples with relatively even distributions of BMI. Gender was not a documented variable, so the research is not able to draw any conclusions based on gender identity. Other variables were also documented as part of the NMDID database, including height, weight, racialized group, and Hispanic identified. Racial group was selected from a set list (White, Black, American Indian, Asian) by the next of kin. Hispanic identification was also based on identification by the next of kin. The variety in socioeconomic status, racialized group, and cause of death was not selected for specifically when selecting individuals for the sample but is expected to also vary in relation to BMI. Height and weight were measured by the medical investigator at the time of autopsy. Individuals whose height and weight measurements were impacted by peri- or post-mortem damage were excluded. Demographic information was obtained during the death investigation, and additional information for some individuals was obtained from next-of-kin interviews.

When examining the breakdown of BMI among the sample it appears that the individuals in the sample occupy, on average, a BMI percentile classifying them as overweight or obese. Individuals were split into three categories based on the WHO classification system; underweight (under 5th percentile BMI), normal weight (between 5th and 95th percentile BMI), and obese (above 95th percentile for BMI) (based on the reference data from Kuczmarski et al. 2002). Out of the 77 individuals in the sample, 36 are considered to be obese, while 35 individuals fall within the normal BMI parameters, and only 6 fall into the underweight group (Table 3.1). There is a relatively even distribution of weight statuses across the ages in the sample, except for the obese group, in which most obese individuals were under the age of 5 years old.

Table 3.1 Size and composition of the sample by age, sex, and BMI percentile category. Sample minimum ages, maximum ages, mean ages, and standard deviation of age are also given.

Age	Sex Samples		BMI samples			Total Samples
	Male	Female	<5th	5th – 95th	>95th	
0-1	2	2		3	1	
1-2	2	4		5	1	
2-3	5	2	1	5	2	
3-4	4	1			4	Child subsample
4-5	2				2	24
5-6	3	1		1	3	
6-7	3	1	2	1	3	
7-8	1	3			2	
8-9						
9-10	7	3	1	7	2	Juvenile
10-11	2	1		2	1	subsample
11-12	2	3		4	1	30
12-13		3		2	1	
13-14	1	3		1	3	
14-15	1		1			
15-16	1	2	1	1	1	
16-17	1	1			2	
17-18	2	2		1	3	Adolescent
18-19	2	2		2	2	subsample
19-20	2				2	23
Total	43	34	6	35	36	77
Min	0.09	0.28	2.83	0.09	0.41	0.09
Max	18.51	18.45	14.78	18.45	18.51	18.51
Mean	7.98	8.98	9.18	7.39	9.3	8.42
SD	5.68	5.76	4.67	5.55	5.97	5.7

3.4 Methods

Metaphyseal and epiphyseal breadth measurements and J measurements were taken from CT scans of the femur and tibia (Figure 3.1). The New Mexico sample is composed of wet bone specimens taken from the pre-autopsy CT scan collection of the NMDID. Bone scans were collected and reconstructed by the Office of the Medical Investigator of New Mexico. They were taken with a Philips Brilliance Big Bore 16-slice CT scanner. Slice thickness was 1 mm with 0.5 mm overlap, each individual is

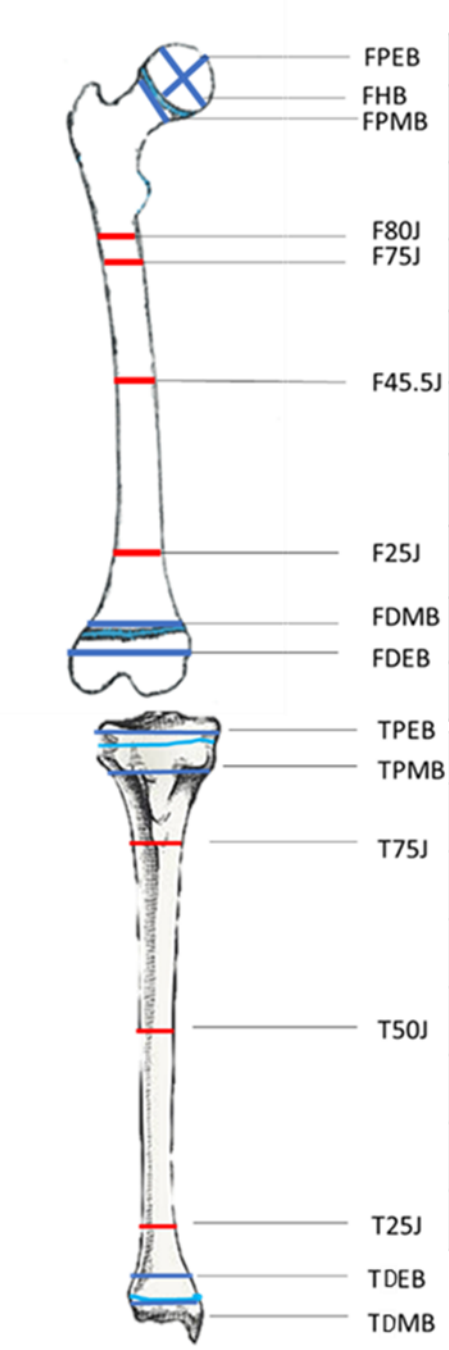
comprised of roughly 10,000 scans, and the special resolution was 0.5 mm. One DICOM file consists of the entire body of the deceased individual and the field of view was 180 mm for whole body scans.

All CT scans were oriented in Dragonfly 2020.1-RC2 software in accordance with the protocols laid out by Spake and colleagues (2020). Long bones were positioned along sagittal and coronal planes, as if they had been placed along an osteometric board. This was accomplished by aligning the coronal plane along the posterior aspect of the distal epiphysis, and the sagittal plane along the lateral aspect of the distal epiphysis. Cross-sectional images from each slice section for the femur and tibia, as well as all length and breadth measurements were taken in Dragonfly.

Cross-sectional slices at each diaphyseal section were captured using the snapshot function in Dragonfly, and then exported as TIFF images. J values were calculated using the Slice Geometry function of the BoneJ plugin in ImageJ software. Segmentation of the images was conducted using the Otsu algorithm (1979). While previous research using BoneJ has mainly used the Ridler and Calvard (1978) algorithm (Mactinosh et al. 2013; Gosman et al. 2013), the results of chapter 1 demonstrate that the Otsu algorithm has similar reliability, and in practise, fewer segmentation issues arose when using that Otsu algorithm on the images of very young individuals. Thus, the Otsu algorithm was used to segment the bone from non-bone. Automated measurements were taken from multiple, sequential cross-sectional images. The 2D bone images were converted to black and white, 8-bit images, oriented along anteroposterior and mediolateral planes, and torsional rigidity (J) was calculated by adding I_{max} and I_{min}. J values were not standardized for weight, as they were used to develop weight estimation formulae, wherein differences in J values based on size were relevant.

Among the femoral diaphyses, J values were calculated from measurements taken at 25%, 45.5%, 75%, and 80% of the total diaphyseal length (Figure 3.1). The 45.5% section was used to approximate the juvenile diaphyseal midsection as calculated by Ruff (2003b), and the 80% was added to account for the subtrochanteric region of the femur. On the tibia, J values were taken from 25%, 50%, and 75% of the diaphyseal length (Figure 3.1). J values were taken at midsections of the diaphysis, as well as towards the proximal and distal end of the diaphysis to explore changes in torsional

rigidity along the bone. Long bone research on adult individuals often uses multiple sections of the diaphysis (Macintosh et al. 2013; Davies and Stock 2014), while many studies focusing on cortical deposition in children tend to focus on the midshaft (Osipov et al. 2016; Harrington and Osipov 2018; Osipov et al. 2020). Breadth measurements were taken at maximum breadths of the proximal and distal femoral and tibial metaphyses as well as at the maximum breadth at the proximal and distal epiphyses of each bone (Ruff 2007). Femoral head breadth was also measured in accordance with Ruff (2007).



Landmark	*Weight Estimation Parameter*
FHB	Superio-inferior femoral head breadth
FPEB	Proximal epiphyseal breadth (anterio-posterior)
FPMB	Proximal metaphyseal breadth
F80J	J at 80% of the femoral diaphysis
F75J	J at 75% of the femoral diaphysis
F45.5J	J at 45.5% of the femoral diaphysis
F25J	J at 25% of the femoral diaphysis
FDMB	Distal metaphyseal breadth
FDEB	Distal epiphyseal breadth

Landmark	*Weight Estimation Parameter*
TPEB	Proximal epiphyseal breadth
TPMB	Proximal metaphyseal breadth
T75J	J at 75% of the tibial diaphysis
T50J	J at 50% of the tibial diaphysis
T25J	J at 25% of the tibial diaphysis
TDMB	Distal metaphyseal breadth
TDEB	Distal epiphyseal breadth

Figure 3.1 Sections of the femur and tibia where weight estimation parameters were taken. Breadth measurements are coloured blue and cross-sectional measurements of J are coloured in red.

Two sets of z-scores were created for this study to be able to compare the weights, heights, BMIs, and femoral midshaft J values of individuals in the New Mexico sample to contemporary standards, regardless of age. The first set of z-scores were

calculated for known weight, height, and BMI among the sample based on child growth data collected by the World Health Organization (birth to 2 years old) and the Centre for Disease Control (2 to 19 years old). The second set of z-scores were created to compare the New Mexico J values with J values calculated from the Denver Growth Study, independent of age. Ruff (2021) collected the measurements of anteroposterior and mediolateral cortical breadth from the mid-shaft of the femoral diaphysis based on radiographs procured during the Denver Growth Study. These radiographs were taken three times during the first year of life, and then semi-annually until 20 years of age. The external mediolateral and anteroposterior cortical breadth measurements were then used to estimate J, based on an eccentric elliptical model of the cross-section (O'Neill and Ruff 2004). Z-scores were calculated based on age-specific means and standard deviations based on Ruff's raw data. It is important to note that J values derived from these elliptical models will potentially overestimate J when compared with the polar moments of area calculated by BoneJ.

A linear correlation analysis was conducted to examine the strength of the relationship between the bone breadths or J measurements, and the weight of the individuals in the sample. Two-way ANCOVA tests were conducted to assess differences in each weight estimation parameter between the sexes and weight status. Weight status was assessed for an individual if they fell above, below, or within the 5^{th} to 95^{th} percentile for BMI. An interaction variable was also tested to determine whether differences occurred among one sex group from one weight status group. Before performing the ANCOVA calculations, data were tested for homogeneity of variances using Levene's test for equality of the slopes by including an interaction term of the covariate with sex and weight status. SPSS v.21 was used to complete the statistical analyses.

Weight prediction formulae from metaphyseal and epiphyseal breadth measurements and J values were calculated using classical calibration rather than inverse calibration, as it is generally a more suited method when the weight being estimated cannot be assumed to fall within the range of the reference sample (Konigsberg et al. 1998; Ruff 2007). In classical calibration the weight (independent variable) is regressed on the skeletal measurement (dependent variable) then the terms of the resulting equation are resolved for the independent variable. Classical calibration equations were developed from all the weight estimation parameters (Figure 3.1).

Formulae were produced for the overall samples, for the sexes separately, for the 3 age categories of child (0-5), juvenile (6-12), and adolescent (13-19), and for those who fell below the 95th percentile ,within the 5th to 95th percentile, and above the 95th percentile for age specific BMI. Age categories were modified from developmental stages proposed by Bogin (1999). Bogin's (1999) infant and child categories were combined for the purposes of this study, on account of the smaller sample size of individuals under 5 years of age. These subsamples were developed to examine any differences between sexes in growth trajectories, to account for changes in growth trajectory over childhood, and to examine the effects of BMI on estimating weight. While it is possible to end the juvenile category at 10 years of age for females and 12 years of age for males, the juvenile cut off was kept at 12 for both sex due to the limited sample size in the juvenile and adolescent subsamples.

In case of the epiphyseal and metaphyseal breadth measurement, there was an exponential relationship with weight, and both variables were transformed by calculating using the natural logarithm. Raw breadth measurements were also log transformed because there was a lower limit to the raw breadth measurements that could be used to estimate weight, and once that threshold was passed the formulae would produce a negative value for the weight estimation. The breadth formulae using logged values had to be returned to normal weight in kilograms by de-transforming the weight estimates using the exponent of the natural log. There is bias incurred by transforming and de-transforming these data. J values were not log transformed, as the correlation between weight and J was (more) linear, and the difference in the error produced by the logged and unlogged formulae was small, thus it was decided to exclude the logged J formulae. Therefore, there is no de-transformation bias present in the weight estimates produced using the J formulae.

For each of the formulae, residuals were calculated as the estimated minus the real weight, meaning that a negative residual indicates an underestimation, and a positive residual indicates overestimation. To explore the performance of the formulae, the mean residual (MR) and mean absolute residual (MAR) for each formula were calculated as measures of accuracy and precision, respectively. The MR was then compared against zero with a single sample t-test.

Mean standard error (MSE) was calculated for each formula based on individual standard error (Lucy 2005), calculated as:

$$\sqrt{\sum \left(\Delta_y\right)^2/(N-2)} * \sqrt{1+\left(\frac{1}{100}\right)+\frac{\left(y-\mu_y\right)^2}{\sum (x-\mu_x)^2 * b^2}}$$

wherein y is the known weight, x is the weight parameter measurement, Δ is the difference between the known weight and the estimated weight, and b is the y-intercept from the regression equation. A prediction interval was created using the standard error, and the range of individuals that fell within the predictive range was calculated for each formula for future observations of the weight parameter used in the formulae. Unlike confidence intervals, prediction intervals are estimated from the individual point values rather than from the sample mean, and thus reliability of the prediction interval is dependent on an even variance of errors across the age range of the regression, referred to as homoscedasticity. The MSE values produced from the breadth formulae were detransformed from their logged version, in order to represent the prediction error in kilograms and for comparison of the MSE values produced from the J formulae. A certain amount of bias is introduced when transforming and then de-transforming the weights. Heteroscedasticity was assessed visually by plotting the standardized residuals against the standardized predicted residuals.

3.5 Results

The overall sample contained positive average z-scores for height, weight , and BMI (Table 3.2). Z scores for height were above average (0.43), while z-scores for weight and BMI were more than 1 unit above the WHO averages of 1.83 and 2.00, respectively (WHO 2006). Z scores for J taken from the femur at the mid-shaft were an average of 2.05 in the sample, based on mean for age measurements taken from Ruff's unpublished radiographs measuring cortical thickness (Ruff 2021).

Table 3.2 Maximum, minimum, average, and standard deviation of z-scores for height, weight, BMI, and J at the femoral mid-shaft.

	N	Min	Max	Mean	SD
Height					
Overall	77	-5.35	3.09	0.43	1.42
Female	34	-2.59	3.00	0.44	1.20
Male	43	-5.35	3.09	0.43	1.58
Child	25	-5.35	3.09	0.26	1.72
Juvenile	31	-1.48	3.00	0.55	1.15
Adolescent	21	-2.59	2.29	0.47	1.43
Below 5th	6	-5.35	1.83	-0.47	2.60
Between 5th and 95th	35	-2.83	3.00	0.26	1.34
Above 95th	36	-1.66	3.09	0.75	1.19
Weight					
Overall	77	-3.45	8.00	1.83	2.36
Female	34	-1.33	6.57	1.97	1.85
Male	43	-3.45	8.00	1.72	2.72
Child	25	-3.45	8.00	1.48	2.76
Juvenile	31	-1.34	6.63	1.54	1.84
Adolescent	21	-1.33	6.69	2.68	2.43
Below 5th	6	-3.45	6.00	-0.24	3.23
Between 5th and 95th	35	-3.38	3.56	0.46	1.15
Above 95th	36	-1.50	8.00	3.51	1.96
BMI					
Overall	77	-2.34	11.78	2.00	2.74
Female	34	-1.49	9.51	2.17	2.35
Male	43	-2.34	11.78	1.87	3.03
Child	25	-2.34	7.89	1.60	2.94
Juvenile	31	-1.99	11.78	1.88	2.83
Adolescent	21	-1.49	6.42	2.66	2.32
Below 5th	6	-1.99	9.51	0.36	4.49
Between 5th to 95th	35	-2.34	2.07	0.31	0.95
Above 95th	36	-1.35	11.78	3.92	2.33
Mid-shaft J					
Overall	73	-5.35	12.49	2.05	3.11
Female	33	-3.31	12.49	1.73	3.07
Male	40	-5.35	12.33	2.31	3.16
Child	24	-5.35	12.49	3.11	4.10
Juvenile	30	-0.95	6.72	1.86	2.04
Adolescent	19	-3.31	9.07	1.00	2.80
Below 5th	6	-5.35	2.36	-0.47	2.60
Between 5th and 95th	33	-3.31	6.49	1.42	2.21

Above 95th	34	-2.46	12.49	3.10	3.56

Homogeneity of the slopes between sex and weight status for each weight estimation parameter was assessed visually. Plots of J values against weight demonstrated little overall variation and no heteroscedasticity when comparing male and females, and when comparing those above and below the 95th percentile for BMI. The logged proximal and distal epiphyseal breadths, and the distal metaphyseal breadth demonstrated heteroscedasticity between the sexes when plotted against logged weight. When plotting the logged weights against the logged breadth measurements, the only parameter to demonstrate no heteroscedasticity between weight statuses was the distal tibial metaphyseal breadth. The error associated with the formulae derived from the heteroscedastic parameters must therefore be questioned, along with the prediction intervals.

In testing for homogeneity of the variances prior to the two-way ANCOVA, Levene's tests of equality results show that within the sexes homogeneity of variance was observed to be unequal across weight status groups for the proximal femoral epiphyseal breadth, the femoral head breadth, 25% of the femoral diaphysis, the proximal and distal tibial epiphyses, and 25%, 50%, and 75% of the tibial diaphysis. The tests also show that across the weight status groups homogeneity of variance was observed.

Two-way ANCOVA tests were conducted to assess differences in each weight estimation parameter collected from the tibia and femur between the sexes and the weight status for the breadth parameters (Table 3.3) and the J parameters (Table 3.4. Significant differences in weight status were found among the section of the femur at 25% and 75% of the diaphyseal length, and the tibia at 50%, and 75% of the diaphyseal length. The section of 45.5% of the femoral diaphysis demonstrated both a significant difference among the sexes (F 172.129, p 0.000), and among the interaction between sex and the different weight status groups (F 3.785, p 0.014), but not among the weight status groups alone (F 0.572, p 0.452), indicating that a combined sex sample might mask differences in weight at the femoral midshaft.

Table 3.3 F and p values for the two-way ANCOVA tests examining sex and weight status for the breadth parameters, wherein the weight estimation parameter is the dependent variable, sex and weight status act as the independent variables, and weight acts as a covariate.

		F	P
MEB Proximal Femur			
	Sex	1.175	0.283
	Weight Status	0.21	0.889
	Interaction	0.156	0.926
MEB Distal Femur			
	Sex	0.004	0.952
	Weight Status	0.761	0.52
	Interaction	0.608	0.612
MMB Proximal Femur			
	Sex	0.984	0.325
	Weight Status	2.711	0.052
	Interaction	0.162	0.921
MMB Distal Femur			
	Sex	0.408	0.525
	Weight Status	1.745	0.167
	Interaction	2.232	0.093
Femoral Head Breadth			
	Sex	0.073	0.788
	Weight Status	0.49	0.691
	Interaction	1.108	0.352
MEB Proximal Tibia			
	Sex	0.223	0.638
	Weight Status	0.387	0.762
	Interaction	0.801	0.498
MEB Distal Tibia			
	Sex	0.027	0.87
	Weight Status	0.218	0.884
	Interaction	0.431	0.731
MMB Proximal Tibia			
	Sex	0.263	0.61
	Weight Status	0.96	0.417
	Interaction	1.299	0.282
MMB Distal Tibia			
	Sex	0.12	0.73
	Weight Status	0.462	0.71
	Interaction	1.991	0.124

Table 3.4 F and p values for the two-way ANCOVA tests examining sex and weight status for the breadth parameters, wherein the weight estimation parameter is the dependent variable, sex and weight status act as the independent variables, and weight acts as the covariate.

		F	P
F25J			
	Sex	0.814	0.37
	Weight Status	2.823	**0.045**
	Interaction	0.117	0.95
F45.5J			
	Sex	172.129	**0.000**
	Weight Status	0.572	0.452
	Interaction	3.785	**0.014**
F75J			
	Sex	1.285	0.261
	Weight Status	4.823	**0.004**
	Interaction	0.469	0.705
F80J			
	Sex	1.016	0.317
	Weight Status	2.729	0.051
	Interaction	0.28	0.84
T25J			
	Sex	1.563	0.216
	Weight Status	2.692	0.053
	Interaction	0.493	0.689
T50J			
	Sex	2.502	0.118
	Weight Status	4.001	**0.011**
	Interaction	0.44	0.725
T75J			
	Sex	1.718	0.194
	Weight Status	3.131	**0.031**
	Interaction	0.32	0.811

Significant results are indicated in bold ($\alpha < 0.05$)

Regression analyses were run on raw J values and natural log values for the breadth measurements. The classical calibration models are summarized in Table 3.5 (breadth measurements) and Table 3.6 (J measurements) for the overall sample and the male and female subsamples. Classical calibration models are summarized in Table 3.7 and 3.8 (breadth measurements) and Tables 3.9 and 3.10 (J measurements) for the age

and BMI divided subsamples. Average values for each weight estimation parameter, sample maximum and minimum measurements, sample size, and standard deviation are also provided.

Table 3.5 Regression formulae for estimating weight using the natural log of the femoral and tibial breadth measures for the overall sample and the sample divided by sex. X is the logged weight value (kg), and y is the logged breadth measurement (mm). Average, minimum, and maximum breadth values are given, as well as the standard deviation. FPEB = proximl femoral ephiphyseal breadth, FDEB = distal femoral ephiphyseal breadth, FPMB = proximl femoral metaphyseal breadth, FDMB = distal femoral metaphyseal breadth, FHB = femoral head breadth, TPEB = proximl tibial ephiphyseal breadth, TDEB = distal tibial ephiphyseal breadth, TPMB = proximl tibial metaphyseal breadth, TDMB = distal tibial metaphyseal breadth.

	FPEB	FDEB	FPMB	FDMB	FHB	TPEB	TDEB	TPMB	TDMB
Overall									
N	70	73	76	74	75	75	72	76	75
Eqn	X = y-1.3775/0.5539	X = y-2.0319/0.5497	X = y-1.9965/0.3831	X = y-2.8781/0.3322	X = y-2.0286/0.3942	x=y-1.8675/0.5642	x=y-1.336/0.5987	x=y-2.5933/0.3692	x=y-2.1191/0.3933
M	31.61	57.29	29.28	58.07	31.71	51.61	34.15	50.25	34.07
Max	51.3	91.1	45.2	87.5	51.2	85.3	56.3	79.2	52.6
Min	6.5	12.6	9	19.4	11.7	13.7	5.7	15.3	10.9
SD	0.46	0.5	0.35	0.3	0.36	0.5	0.55	0.34	0.36
Female									
N	32	33	34	33	32	33	32	34	33
Eqn	x=y-0.9327/0.6669	x=y-1.877/0.5893	x=y-1.974/0.384	x=y-2.801/0.346	x=y-2.001/.4	x=y-1.703/0.611	x=y-1.029/0.681	x=y-2.528/0.380	x=y-1.920/0.439
M	31.55	58.51	29.37	57.45	32.54	53.66	35.3	50.27	33.91
Max	46.8	82.1	44.8	74.3	45.4	76.3	85.3	42.2	75.8
Min	6.5	14.2	13.8	28.4	15.7	13.9	5.7	21.2	13.7
SD	0.52	0.48	0.32	0.28	0.33	0.5	0.58	0.32	0.37

Male									
N	38	40	42	41	43	42	40	42	42
Eqn	x=y-1.676/0.4785	x=y-2.126/0.525	x=y-2.005/0.385	x=y-2.915/0.329	x=y-2.0396/0.3927	x=y-1.972/0.533	x=y-1.524/0.547	x=y-2.621/0.368	x=y-2.219/0.372
M	31.66	56.29	29.21	58.58	31.09	49.99	33.22	50.24	34.2
Max	51.3	91.1	45.2	87.5	51.2	85.3	56.3	79.2	52.6
Min	12.3	12.6	9	19.4	11.7	13.7	7	15.3	10.9
SD	0.41	0.51	0.37	0.31	0.37	0.5	0.52	0.35	0.35

Table 3.6 Regression formulae for estimating weight using the J values at diaphyseal cross-sections for the overall sample and the sample divided by sex. X is the weight value (kg), and y is the J value (mm4). Average, minimum, and maximum J values are given, as well as the standard deviation. F25J = J at 25% of the femoral diaphysis, F45.5J = J at 45.5% of the femoral diaphysis, F75J = J at 75% of the femoral diaphysis, F80J = J at 80% of the femoral diaphysis, T25J = J at 25% of the tibial diaphysis, T50J = J at 50% of the tibial diaphysis, and T75J = J at 75% of the tibial diaphysis.

	F25J	F45.5J	F75J	F80J	T25J	T50J	T75J
Overall							
N	75	76	74	76	73	76	76
Eqn	x=y+8317.4/958.81	x=y+6670.6/717.14	x=y+3902.1/751.26	x=y+5796.5/1001.4	x=y+1209.5/337.67	x=y+4525.4/540.91	x=y+9368.2/941.58
M	33911	24769	29468	37729	13884	19169	31761
Max	151789	113613	118993	236322	59330	103468	181041
Min	1314	343	1461	2346	646	372	1077
SD	36287	26882	28527	39467	12889	20405	35358
Female							
N	34	34	34	33	34	34	34
Eqn	x=y+4746.8/850.17	x=y+4476.6/667.9	x=y+3776.4/731.56	x=y+2418.4/895.66	x=y+109.23/298.93	x=y+2360.7/454.34	x=y+4985.9/790.85
M	32596	24860	29008	37606	13021	17595	29751
Max	95541	80279	80132	124323	30903	49305	77982
Min	1913	710	1461	2346	646	898	1151
SD	25683	20453	22342	29003	9069	13592	23399
Male							
N	40	42	43	43	42	42	42

Eqn	x=y+8962.6/997.88	x=y+7493.2/735.35	x=y+3650.1/758.52	x=y+2418.4/895.66	x=y+1381.7/352.39	x=y+4604.6/573.06	x=y+9786.9/996.55
M	43452	24695	29156	40366	14021	20443	33314
Max	151789	113613	118993	236322	59330	103468	181041
Min	1314	343	2037	2976	648	372	1077
SD	43452	31128	32071	46264	15068	24500	39909

Table 3.7 Regression formulae for estimating weight using the natural log of the femoral and tibial breadth measures for the sample divided by age category. X is the natural logged weight value (kg), and y is the logged breadth measurement (mm). Average, minimum, and maximum breadth values are given, as well as the standard deviation. FPEB = proximal femoral epiphyseal breadth, FDEB = distal femoral epiphyseal breadth, FPMB = proximl femoral metaphyseal breadth, FDMB = distal femoral metaphyseal breadth, FHB = femoral head breadth, TPEB = proximl tibial epiphyseal breadth, TDEB = distal tibial ephiphyseal breadth, TPMB = proximl tibial metaphyseal breadth, TDMB = distal tibial metaphyseal breadth.

	FPEB	FDEB	FPMB	FDMB	FHB	TPEB	TDEB	TPMB	TDMB
Child (0-5 years old)									
N	18	22	24	24	24	23	23	25	25
Eqn	x=y-0.745/0.728	x=y-1.315/0.774	x=y-1.828/0.427	x=y-2.663/.4	x=y-2.246/0.283	x=y-1.633/0.6	x=y-0.700/0.792	x=y-2.341/0.447	x=y-2.032/0.403
M	15.79	29	18.83	40.42	19.7	25.4	17.1	33.59	22
Max	23.2	47.7	25.7	53.3	25.4	42.2	29	50.3	30.6
Min	6.5	12.6	9	19.4	11.7	13.7	5.7	15.3	10.9
SD	0.34	0.39	0.25	0.23	0.19	0.32	0.46	0.27	0.26
Juvenile (6-12 years old)									
N	31	31	31	30	30	31	29	30	30
Eqn	x=y-2.186/0.359	x=y-3.3268/0.2276	x=y-2.493/0.260	x=y-3.328/.22	x=y-2.277/0.337	x=y-2.8384/0.3295	x=y-2.247/0.387	x=y-3.383/0.161	x=y-2.525/0.298
M	32.34	62.89	30.66	60.98	32.83	55.73	37.44	52.22	36.13
Max	44.9	82.1	44.8	73	44	75.8	47.1	63.3	46.7
Min	21.5	46.9	22	44.6	21.3	34.1	23	43	23
SD	0.18	0.13	0.14	0.1	0.16	0.18	0.19	0.09	0.15
Adolescent (13-19 years old)									

N	21	20	21	20	21	21	21	20	21
Eqn	x=y-3.2906/0.1113	x=y-3.8084/0.1285	x=y-2.9362/0.1644	x=y-3.5132/0.1813	x=y-3.2401/0.1215	x=y-3.7979/0.1149	x=y-3.1834/0.1595	x=y-3.5344/0.1522	x=y-3.0959/0.1657
M	44.1	79.73	39.2	74.91	43.81	74.23	48.97	67.29	46.08
Max	51.3	91.1	45.2	87.5	51.2	85.3	56.3	79.2	52.6
Min	32.6	68.2	26.8	63.6	35.7	63.3	40.2	57.8	36.1
SD	0.1	0.08	0.13	0.09	0.1	0.08	0.09	0.09	0.1

Table 3.8 Regression formulae for estimating weight using the natural log of the femoral and tibial breadth measures for the sample divided by BMI percentile. X is the natural logged weight value (kg), and y is the logged breadth measurement (mm). Average, minimum, and maximum breadth values are given, as well as the standard deviation. FPEB = proximal femoral ephiphyseal breadth, FDEB = distal femoral ephiphyseal breadth, FPMB = proximl femoral metaphyseal breadth, FDMB = distal femoral metaphyseal breadth, FHB = femoral head breadth, TPEB = proximl tibial ephiphyseal breadth, TDEB = distal tibial ephiphyseal breadth, TPMB = proximl tibial metaphyseal breadth, TDMB = distal tibial metaphyseal breadth.

	FPEB	FDEB	FPMB	FDMB	FHB	TPEB	TDEB	TPMB	TDMB
<95% Percentile BMI									
N	36	40	41	41	41	40	40	41	41
Eqn	x = y - 0.9164/0.7187	x = y - 1.5042/0.7322	x = y - 1.8061/0.4543	x = y - 2.7316/0.3901	x = y - 1.874/0.4603	x = y - 1.4184/0.7282	x = y - 0.7419/0.8041	x = y - 2.3799/0.4513	x = y - 1.9683/0.4593
M	30.33	53.16	27.12	54.84	29.83	48.14	32.07	47.46	32.32
Max	49.3	88.6	45.2	77.4	48.9	80.3	53.1	73	49.3
Min	9	12.6	9	19.4	11.7	13.7	5.7	15.3	10.9
SD	11.88	22.64	8.9	14.69	10.41	21.17	14.72	14.58	10.19
> 95th Percentile BMI									
N	34	33	35	34	35	35	32	35	34
Eqn	x = y - 1.4176/0.5191	x = y - 2.4495/0.4232	x = y - 2.1044/0.3444	x = y - 2.9187/0.3098	x = y - 1.9944/0.387	x= y - 2.0824/0.4858	x = y - 1.8073/0.4541	x = y - 2.683/0.3316	x = y - 2.0738/0.3871
M	32.96	62.31	31.82	62.09	33.86	55.57	36.75	53.53	36.17
Max	51.3	91.1	45.1	87.5	51.2	85.3	56.3	79.2	52.6

Min	6.5	23.1	16.6	30.7	17.3	20	12.3	23.7	15.1
SD	11.79	19.78	8.76	15.51	10.19	19.95	12.99	14.38	11.07

Table 3.9 Regression formulae for estimating weight using the natural log of the femoral and tibial breadth measures for the sample divided by age category. X is the natural logged weight value (kg), and y is the logged breadth measurement (mm). Average, minimum, and maximum breadth values are given, as well as the standard deviation. F25J = J at 25% of the femoral diaphysis, F45.5J = J at 45.5% of the femoral diaphysis, F75J = J at 75% of the femoral diaphysis, F80J = J at 80% of the femoral diaphysis, T25J = J at 25% of the tibial diaphysis, T50J = J at 50% of the tibial diaphysis, and T75J = J at 75% of the tibial diaphysis.

	F25J	F45.5J	F75J	F80J	T25J	T50J	T75J
Child (0-5 years old)							
N	24	24	25	25	25	25	25
Eqn	x=y-1259.8/ 348.44	x=y-156.8/ 229.55	x=y-2615.6/ 281.45	x=y-5921.9/ 235.68	x=y-309.86/ 170.64	x=y-889.74/ 167.21	x=y-883.94/ 288.2
M	6104	3364	6631	9284	2744	3275	4996
Max	12793	7987	18897	20272	6824	8498	9344
Min	1314	343	1461	2346	646	372	1077
SD	3383	2116	4015	4640	1593	1750	2579
Juvenile (6-12 years old)							
N	30	31	31	30	30	30	31
Eqn	x=y+1307.8/ 687.58	x=y+1740.3/ 518.28	x=y-2036.4/ 549.33	x=y-3539.1/ 687.35	x=y-1744/ 255.78	x=y-1233.9/ 382.97	x=y+2131.1/ 716.52
M	24142	17276	22192	28453	11237	15447	24158
Max	63107	47872	45319	52844	23996	32293	58317
Min	6063	6001	6576	9693	3667	3882	6352
SD	12654	8745	9156	11236	4763	6998	12626
Adolescent (13-19 years old)							
N	21	21	21	21	20	21	21
Eqn	x=y+2156/ 924.01	x=y+1704.7/ 687.92	x=y-2862/ 701.42	x=y+5730.2/ 1023.1	x=y-3997.1/ 291.5	x=y+8585.3/ 587.3	x=y+13738/ 999.81
M	79645	59196	64958	84843	29803	43407	74773

Max	151789	113613	118993	236322	59330	103468	181041
Min	27020	15424	25803	38515	14514	6933	22156
SD	36732	26895	28548	46101	11912	22918	38641

Table 3.10 Regression formulae for estimating weight using the natural log of the femoral and tibial breadth measures for the sample divided the sample divided by BMI percentile. X is the natural logged weight value (kg), and y is the logged breadth measurement (mm). Average, minimum, and maximum breadth values are given, as well as the standard deviation. F25J = J at 25% of the femoral diaphysis, F45.5J = J at 45.5% of the femoral diaphysis, F75J = J at 75% of the femoral diaphysis, F80J = J at 80% of the femoral diaphysis, T25J = J at 25% of the tibial diaphysis, T50J = J at 50% of the tibial diaphysis, and T75J = J at 75% of the tibial diaphysis

	F25J	F45.5J	F75J	F80J	T25J	T50J	T75J
<95% Percentile BMI							
N	41	41	41	40	41	41	41
Eqn	x = y +7022/ 1018.9	x = y + 6140.4/ 754.71	x = y + 2086.2/ 746.6	x = y + 1088.4/ 948.99	x = y + 1476.5/ 392.87	x = y + 2485.2/ 518.41	x = y + 6260.3/ 942.07
M	23600	16541	20351	26958	10330	13094	22051
Max	103907	70614	70558	87770	35692	45048	85263
Min	1314	343	1461	2346	646	372	1077
SD	21958	16216	15994	19874	8148	10660	19126
> 95th Percentile BMI							
N	34	35	36	36	35	35	36
Eqn	x = y + 15024/ 1007.5	x = y + 11581/ 755.61	x = y + 7689.2/ 782.76	x = y + 13339/ 1069.9	x = y + 3735.1/ 352.29	x = y + 8000.4/ 572.24	x = y + 15853/ 995.11
M	46345	33749	38429	49697	17372	26285	42776
Max	151789	113613	118993	236322	59330	103468	181041
Min	2082	831	2888	2976	718	898	1853
SD	46009	33945	35743	51597	16051	26465	45403

Correlation coefficients, mean standard error, mean residuals, and mean absolute residuals are listed in Table 3.10 for the overall and sex subsample breadth formulae, and table 3.11 for the overall and sex subsample J formulae. The same values are listed for the age and weight samples for the breadth and J formulae in Table 3.12 and 3.13,

respectively. For the overall sample, the equation using J at 75% of the tibial diaphysis had the lowest overall mean standard error (MSE), 12.61 kgs, though J at 45.5% of the femur had a similar MSE value, and J at 25% and 75% of the femur as well as J at 25% and 50% of the tibia were not much higher (Table 3.11). The MSE values for the logged breadth values were all higher, by between 3 to 12 kgs for the overall sample (Table 3.11). MSE was also lowest for the formulae using J at 75% of the tibia for females (11.34 kgs) and 45.5% of the femur for males (12.14 kgs). There were comparable MSE values found for the formulae using the distal femoral metaphyseal breadth for the female subsample, but otherwise all sex specific breadth formulae produced higher MSEs. For the subsamples as divided by age, the lowest MSE value for the child subsample was made with the formulae using the femoral distal epiphyseal breadth (4.05 kgs), but the other equations using the femoral breadth measurements (except for femoral head breadth) had similar MSE values (Table 3.12). For the juvenile subsample, the formulae using J at 80% of the femoral diaphysis had the lowest MSE (10.98 kgs), with the J at 45.5%, 50%, and 75% of the femur (Table 3.13), and the distal tibial metaphyseal breadth (Table 3.12) formulae producing similar MSE values. The adolescent subsample MSE was lowest in the formula that used J at 45.5% of the femoral diaphysis (12.41 kgs). When examining the BMI specific formulae, the MSE from the formula for J at 75% of the tibia was lowest for the below 95th percentile sample (6.98 kgs), and the J at 45.5% of the femur formulae produced the lowest MSE for the above 95th percentile sample (12.41 kgs) (Table 3.13).

Table 3.11 Coefficients of determination (R^2), mean standard error (MSE), mean residuals (MR), mean absolute residual (MAR), and percentage of individuals whose known weights fall within the 95% prediction interval for the overall sample and sex subsample logged breadth formulae for the femur and tibia. FPEB = proximl femoral ephiphyseal breadth, FDEB = distal femoral ephiphyseal breadth, FPMB = proximl femoral metaphyseal breadth, FDMB = distal femoral metaphyseal breadth, FHB = femoral head breadth, TPEB = proximl tibial ephiphyseal breadth, TDEB = distal tibial ephiphyseal breadth, TPMB = proximl tibial metaphyseal breadth, TDMB = distal tibial metaphyseal breadth.

	FPEB	FDEB	FPMB	FDMB	FHB	TPEB	TDEB	TPMB	TDMB
Overall									
N	70	73	76	74	75	75	72	76	75
R^2	0.88	0.89	0.92	0.93	0.92	0.91	0.88	0.92	0.91
MSE	20.32	20.45	18.62	14.53	17.44	19.46	20.64	15.82	16.12
MR	-0.12	-0.91	0.16	-0.73	0.95	-0.36	-0.77	-0.47	-0.19
MAR	13.79	14.12	12.67	10.18	11.71	13.39	13.83	11.06	11.23
% Range	91.43	93.15	90.79	93.24	93.33	93.33	94.44	92.11	93.33
Female									
N	32	33	34	33	32	33	32	34	33
R^2	0.89	0.92	0.92	0.92	0.92	0.92	0.89	0.90	0.90
MSE	17.56	15.84	18.42	13.47	17.23	15.79	17.92	14.15	14.97
MR	0.17	-0.37	1.45	0.27	1.09	0.13	-0.51	0.55	1.16
MAR	10.66	11.41	12.31	9.44	10.41	11.31	12.55	10.56	10.62
% Range	93.75	90.91	94.12	87.88	96.85	93.94	93.75	97.06	96.97
Male									
N	38	40	42	41	43	42	40	42	42
R^2	0.89	0.87	0.92	0.94	0.92	0.89	0.88	0.93	0.94
MSE	23.20	23.76	21.26	16.83	18.99	21.53	23.11	18.34	17.80
MR	-0.02	-1.12	-0.96	-2.37	0.85	-0.50	-0.66	-1.82	-1.40
MAR	10.66	11.41	12.31	9.44	10.41	11.31	12.55	10.56	10.62
% Range	97.37	90.00	90.48	90.24	90.70	90.48	90.00	90.48	92.86

Table 3.12 Coefficients of determination (R^2), mean standard error (MSE), mean residuals (MR), mean absolute residual (MAR), and percentage of individuals whose known weights fall within the 95% prediction interval for the overall sample and sex subsample J formulae for the femur and tibia. F25J = J at 25% of the femoral diaphysis, F45.5J = J at 45.5% of the femoral diaphysis, F75J = J at 75% of the femoral diaphysis, F80J = J at 80% of the femoral diaphysis, T25J = J at 25% of the tibial diaphysis, T50J = J at 50% of the tibial diaphysis, and T75J = J at 75% of the tibial diaphysis.

	F25J	F45.5J	F75J	F80J	T25J	T50J	T75J
Overall							
N	75	76	74	76	73	76	76
R^2	0.94	0.95	0.94	0.9	0.94	0.94	0.94
MSE	13.16	12.62	13.42	17.51	13.56	13.28	12.61
MR	0	0	0	0	0	0	0
MAR	8.76	8.55	8.52	9.79	8.85	8.77	7.69
% Range	89.33	90.79	90.54	94.74	90.41	93.42	93.42
Female							
N	34	34	34	33	34	34	34
R^2	0.91	0.92	0.9	0.86	0.91	0.92	0.93
MSE	12.98	13.99	13.77	17.06	13.31	12.24	11.34
MR	-0.04	0	0.12	0	0	0	0
MAR	8.83	9.73	8.75	10.6	9.88	9.02	8.11
% Range	88.24	94.12	88.24	93.94	94.12	94.12	94.12
Male							
N	40	42	43	43	42	42	42
R^2	0.95	0.96	0.95	0.91	0.95	0.95	0.96
MSE	13.8	12.14	13.09	21.5	13.53	13.51	13.07
MR	0	0	0	0	0	0	0
MAR	9.17	7.83	8.06	11.45	8.61	8.53	7.81
% Range	87.5	88.1	93.02	95.35	90.48	92.86	95.24

The mean residuals (MR) for the weight estimates derived for J measurement formulae ranged between -0.04 and 0.012 kgs (Tables 3.11 and 3.13). All non-zero mean residuals resulted from minor variations in the rounding of regression coefficients. The fact that the J MR values are indistinguishable from zero indicates no bias in estimating weight using the non-logarithmic weight estimation formulae. The logged breadth formulae produced mean residuals ranging from -2.37 to 20.34 kgs. (Table 3.11

and 3.12). The fact that there are such large mean residuals, the majority of which are positive and are produced by the adolescent subsample, would indicate that the logged equations have a bias towards overestimating the body mass in the logged formulae developed from this sample. The log formulae with the lowest MSE in the overall sample and subsamples generally had mean residual values that were near zero, though most still were biased in overestimating weights, with the exception of the below 95^{th} percentile subsamples, which tended to underestimate the weights.

Table 3.13 Coefficients of determination (R^2), mean standard Error (MSE), mean residuals (MR), mean absolute residual (MAR), and percentage of individuals whose known weights fall within the 95% prediction interval for the age and BMI subsample logged breadth formulae for the femur and tibia. FPEB = proximl femoral ephiphyseal breadth, FDEB = distal femoral ephiphyseal breadth, FPMB = proximl femoral metaphyseal breadth, FDMB = distal femoral metaphyseal breadth, FHB = femoral head breadth, TPEB = proximl tibial ephiphyseal breadth, TDEB = distal tibial ephiphyseal breadth, TPMB = proximl tibial metaphyseal breadth, TDMB = distal tibial metaphyseal breadth.

	FPEB	FDEB	FPMB	FDMB	FHB	TPEB	TDEB	TPMB	TDMB
Child (0-5 years old)									
N	18	22	24	24	24	23	23	25	25
R^2	0.77	0.87	0.85	0.83	0.73	0.82	0.77	0.81	0.76
MSE	4.11	4.05	4.68	4.73	6.8	5.92	5.4	5.83	6.35
MR	0.51	0.34	0.42	0.45	1.21	0.67	0.71	0.67	0.98
MAR	2.66	2.75	3.5	3.32	5.02	3.87	3.52	4.02	4.68
% Range	94.44	90.91	95.83	95.83	95.83	86.96	95.65	96	100
Juvenile (6-12 years old)									
N	31	31	31	30	30	31	29	30	30
R^2	0.68	0.61	0.67	0.73	0.69	0.66	0.7	0.63	0.7
MSE	14.54	17.93	21.44	11.66	13.17	14.51	12.58	21.03	11.06
MR	2.61	3.63	3.28	1.74	2.06	2.56	1.91	3.54	1.73
MAR	10.3	12.37	11.52	8.31	9.81	10.11	9.97	13.62	8.36
% Range	93.55	90.32	96.77	96.67	96.67	93.55	100	96.67	96.67
Adolescent (13-19 years old)									
N	21	20	21	20	21	21	20	21	20
R^2	0.43*	0.63	0.51	0.8	0.5	0.57	0.72	0.69	0.71
MSE	93.36	54.45	58.95	33.96	87.53	61.36	46.52	46.12	37.35
MR	26.34	11.26	14.5	4.84	21.93	14.27	6.97	8.26	6.11
MAR	60.3	38.54	39.48	23.12	56.25	40.19	29.61	30.97	24.8
% Range	95.24	95	95.24	95	100	95.24	95	95.24	100
<95% Percentile BMI									
N	36	40	41	41	41	40	40	41	41
R^2	0.86	0.89	0.88	0.91	0.88	0.92	0.86	0.92	0.92
MSE	9.31	8.53	10.23	7.11	9.18	8.1	8.89	7.07	7.42
MR	0.28	-0.16	0.47	-0.39	0.98	-0.05	-0.01	-0.24	-0.14
MAR	6.32	5.93	6.12	5	6.04	5.48	6.22	4.68	5.22
% Range	91.67	92.5	90.24	92.68	95.12	90	92.5	90.24	95.12
> 95th Percentile BMI									
N	34	33	35	34	35	35	32	35	34
R^2	0.9	0.88	0.91	0.92	0.94	0.89	0.89	0.9	0.9

MSE	17.44	17.81	21.53	15.05	14.76	17.2	15.85	14.7	16.02
MR	-0.37	0.74	1.64	0.87	1.05	0.9	0.79	1.48	0.74
MAR	12.11	13.74	13.95	10.15	10.22	13.23	11.6	10.65	11.59
% Range	94.12	93.94	97.14	94.12	94.29	94.29	93.75	94.29	94.12

Table 3.14 Coefficients of determination (R^2), mean standard Error (MSE), mean residuals (MR), mean absolute residual (MAR), and percentage of individuals whose known weights fall within the 95% prediction interval for the age and BMI subsample J formulae for the femur and tibia. F25J = J at 25% of the femoral diaphysis, F45.5J = J at 45.5% of the femoral diaphysis, F75J = J at 75% of the femoral diaphysis, F80J = J at 80% of the femoral diaphysis, T25J = J at 25% of the tibial diaphysis, T50J = J at 50% of the tibial diaphysis, and T75J = J at 75% of the tibial diaphysis.

	F25J	F45.5J	F75J	F80J	T25J	T50J	T75J
Child (0-5 years old)							
N	24	24	25	25	25	25	25
R^2	0.65	0.70	0.45	0.33*	0.69	0.62	0.72
MSE	7.72	6.93	13.33	19.49	7.05	8.62	6.48
MR	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAR	5.72	5.09	8.39	13.36	4.59	5.43	5.22
% Range	95.83	91.67	92.00	96.00	92.00	92.00	96.00
Juvenile (6-12 years old)							
N	30	31	31	30	30	30	31
R^2	0.68	0.74	0.75	0.76	0.67	0.68	0.71
MSE	13.96	11.76	11.44	10.98	14.36	13.86	12.89
MR	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAR	9.31	7.21	8.27	7.42	10.01	9.61	9.16
% Range	93.33	93.55	93.55	93.33	93.33	96.67	96.77
Adolescent (13-19 years old)							
N	21	21	21	21	21	21	21
R^2	0.86	0.87	0.84	0.76	0.83	0.87	0.88
MSE	21.69	20.32	23.57	31.19	23.89	20.16	19.34
MR	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAR	16.98	16.48	18.65	21.06	17.74	15.19	12.57
% Range	100.00	100.00	100.00	95.24	95.24	95.24	95.24
<95% Percentile BMI							
N	38	38	36	37	38	38	39
R^2	0.79	0.79	0.80	0.83	0.85	0.86	0.89
MSE	10.14	10.00	9.86	8.79	8.21	7.74	6.98
MR	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAR	6.40	6.03	6.02	6.21	5.13	5.07	4.25
% Range	92.68	90.24	90.24	90.00	90.24	90.24	95.12

> 95th Percentile BMI							
N	34	35	36	36	35	35	36
R^2	0.95	0.96	0.94	0.89	0.95	0.94	0.95
MSE	14.01	12.41	15.35	22.04	14.41	16.53	15.24
MR	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MAR	10.02	9.39	9.99	12.01	9.70	10.71	9.45
% Range	94.12	100.00	91.67	94.44	91.43	97.14	97.22

* Correlation was not significant at the 0.05 level

For the overall sample mean absolute residual (MAR) values ranged between 10.18 – 14.12 kgs in the breadth measurements (Table 3.11), and between 7.65—- 9.79 kgs for the J formulae (Table 3.12), while the percent of individuals that fell inside the 95% confidence interval range from 89.33 to 94.4%. Of the MAR among the sex subsamples the females ranged from 9.44 – 12.90 kgs for the breadth formulae (Table 3.11) and 8.11 – 10.06 kgs for the J formulae (Table 3.12). Among the males, MAR ranged from 9.44—- 12.55 kgs for the breadth formulae (Table 3.11) and 7.81 – 11.45 kgs for the J formulae (Table 3.12). For the sex samples the percent of individuals within the 95% prediction interval were 88.24 – 96.96% for the females and 87.5 – 97.37% for the males. Among the child subsample the MAR ranged from 2.06 – 5.02 kgs for the breadth formulae (Table 3.13) and from 4.59 – 13.36 kgs for the J formulae (Table 3.14). The juvenile subsample had MARs ranging from 8.36 – 13.2kgs for the breadth formulae (Table 3.13) and 7.21 – 10.01kgs for the J formulae (Table 3.14). In the adolescent subsample had MARs ranging from 23.01 – 60.30 kgs for the breadth formulae (Table 3.13) and 12.57 – 21.06 kgs for the J formulae (Table 3.14). Percent of individuals outside the 95% prediction interval were between 86.96 – 100% among the age subsamples. For the below 95th percentile BMI subsample MARs ranged from 4.68 – 6.32 among the breadth formulae (wherein between 90% to 95.12% fell within the prediction interval) (Table 3.13), and 4.25 – 6.40 kgs for the J formulae (with 90 to 95.15% falling within the prediction interval) (Table 3.14). For those that fell above the 95th percentile for BMI MARs ranged from 10.15 to 13.95 kgs among the breadth formulae (93.15 – 97.14% falling within the prediction interval) (Table 3.13), and 9.39 to 12.01 kgs among the J formulae (with 91.43 – 100% falling within the prediction interval) (Table 3.14).

When examining the raw residuals produced by the J measurement formulae, generally the younger individuals' weights are underestimated, while older individuals

are overestimated. Each parameter began to overestimate weight between the ages of 7 and 10(Figure 3.2). The child sample begins to overestimate at age 2, with the exception of the formulae that used J at 75% of the femoral diaphysis, which begins to underestimate weight at age 2. The juvenile formulae begin to overestimate weight between age 8 and 9, while the adolescent formulae begin to overestimate weight between 15 and 16, except for 50% of the tibial diaphysis formulae, which begins underestimating weight at age 16.

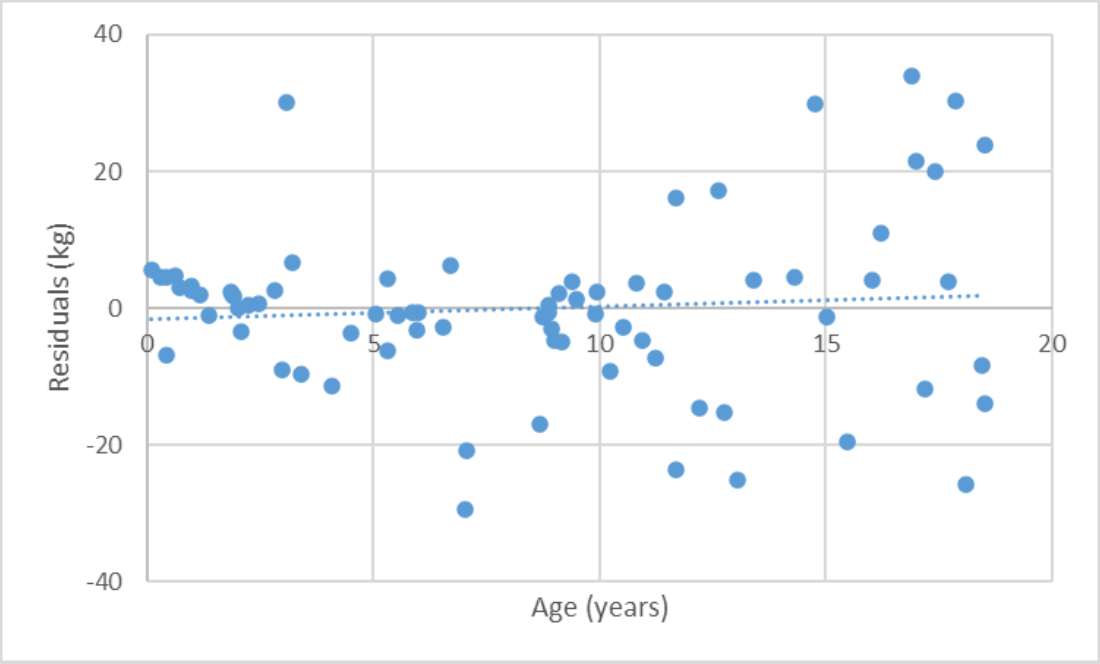


Figure 3.2 Raw residuals (kg) plotted against age produced from the formulae using J at 45% of the femoral diaphyseal length.

The J formulae began to overestimate weight when J values exceeded between 15,000 mm4 to 40,000 mm^4, depending on the parameter used in the formulae (Table 3.15). The J value threshold for the male formulae tended to be slightly higher than the overall sample, while the female threshold tended to be lower. The formulae for children consistently had the lowest J value threshold, followed by juvenile formulae, while the adolescent formulae consistently had the highest J value thresholds wherein weight began to be overestimated. The over 95th percentile formulae consistently had a higher J threshold than the under 95th percentile formulae (Figure 3.3). None of the residuals produced using the J formulae differed significantly from zero, but several of the breadth residuals did (Table 3.15).

Table 3.15 J measurement (mm4) where the formulae begin to demonstrate a bias towards over or under estimating weight for each J parameter and sample, based on visual assessment of residual plots. All values are rounded to the nearest year for age and the nearest 500 mm^4.

	F25J	F45J	F75J	F80J	T25J	T50J	T75J
Overall	40000	30000	30000	40000	15000	20000	30000
Female	35000	25000	20000	35000	12500	20000	30000
Male	41000	30000	35000	45000	17000	20000	30000
Child	6000	3000	7000	9000	3000	3500	5500
Juvenile	25000	15000	23000	25000	27000	12000	27000
Adolescent	79000	61000	65000	850000	79000	43000	65000
<95th Percentile	25000	18000	30000	30000	12500	15000	25000
>95th Percentile	50000	39000	38000	50000	15000	29000	40000

Table 3.16 P values from the single sample t-tests for the breadth formulae residuals.

	Overall	Female	Male	Child	Juvenile	Adolescent
LogFPEB	**0.008**	0.360	0.079	0.733	**0.023**	**0.000**
LogFDEB	**0.002**	0.797	**0.000**	**0.032**	**0.001**	**0.002**
LogFPMB	0.065	0.355	0.215	0.299	0.054	**0.000**
LogFDMB	0.647	0.660	0.452	0.064	0.352	0.584
LogFHB	**0.031**	0.398	0.056	**0.000**	0.095	**0.001**
LogTPEB	**0.025**	0.834	**0.010**	0.626	**0.005**	**0.001**
LogTDEB	**0.000**	0.097	**0.000**	**0.043**	0.166	**0.020**
LogTPMB	**0.032**	**0.025**	0.991	**0.049**	**0.004**	**0.012**
LogTPMB	**0.008**	**0.017**	0.640	**0.001**	0.071	**0.005**

Significant results are indicated in bold ($\alpha < 0.05$)

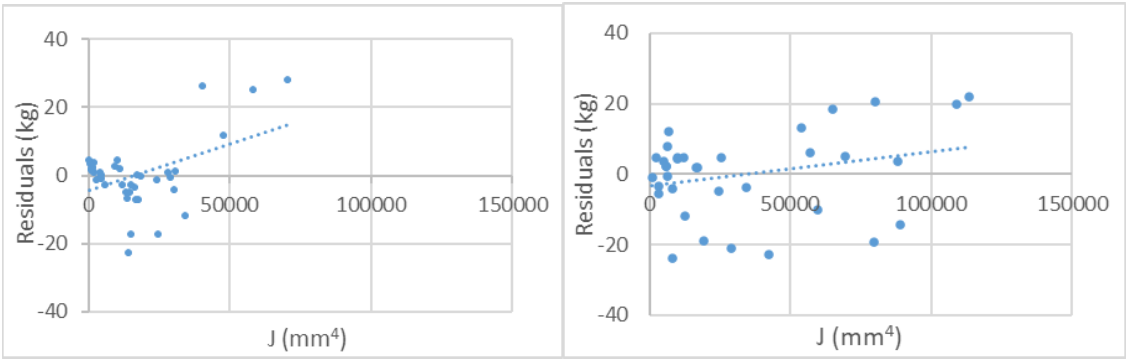


Figure 3.3 Weight residuals from the J at 45.5% formulae plotted against J values at 45.5% of the femoral diaphysis (mm^4) for the below 95th percentile sample (left) and the above 95th percentile sample (right).

The logged breadth formulae demonstrated a more complex pattern in terms of biases in weight estimation. When examining estimates by age there were roughly equal numbers of formulae that underestimated younger individuals and overestimated older individuals as there were formulae that overestimated younger individuals and underestimated older individuals (Table 3.15). There were also several formulae that simply overestimated weight for all individuals in the sample (Figure 3.5). Most formulae tended to overestimate weight in younger individuals and underestimate it in older individuals, except for when the sample was broken down by age category, in which case the older individuals were overestimated, and the younger individuals were underestimated. The only parameter that was the exception to this was femoral head breadth, wherein formulae for all samples overestimated weight in older individuals and underestimated it in younger individuals.

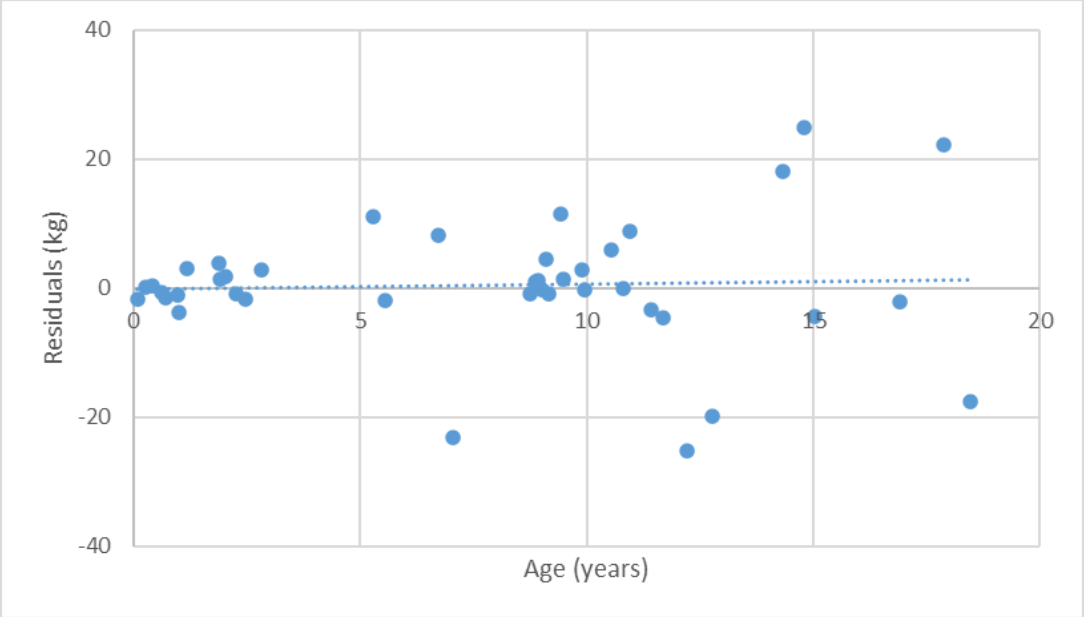


Figure 3.4 De-transformed residuals (kg) plotted against age produced from the below 95th percentile BMI sample formulae using the proximal femoral metaphyseal.

The de-transformed breadth parameters wherein the weights are over or underestimated are similarly complex, though it was more common for formulae to underestimate weight with smaller breadth measurements and to overestimate weight with larger measurements. As with the residuals plotted against age, the femoral head breadth-based formulae, and the formulae produced from the age category samples tended to underestimate weight for the smaller breadths and overestimate weight for the

larger breadths. Among the formulae derived from the BMI percentile subsamples, the above 95th percentile formulae followed a similar trend or overestimated weight in all individuals in the sample, apart from the femoral proximal epiphyseal formulae which overestimated weight among smaller breadth and underestimated weight among larger breadths (Figure 3.5). The below 95th percentile formulae were a mix between the two patterns of bias. Female threshold breadths were slightly smaller than males, and both were similar to the overall breadth threshold. Unsurprisingly, the child sample had the smallest breadth threshold, followed by the juvenile sample, while the adolescent sample has the largest breadth threshold.

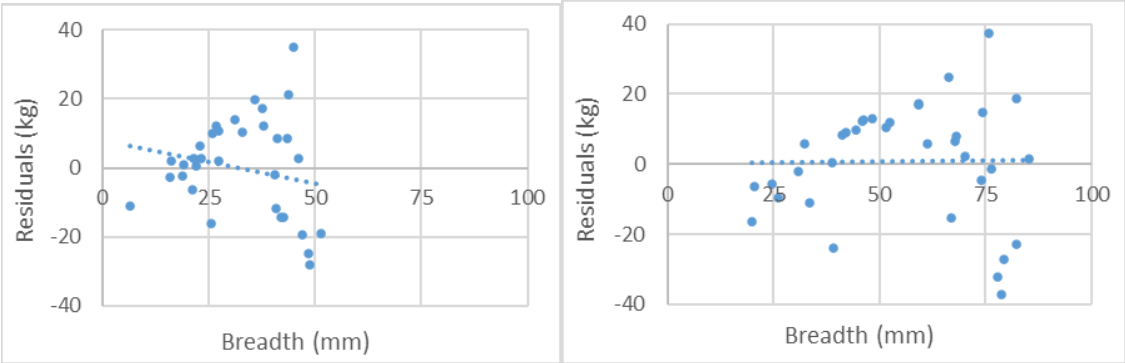


Figure 3.5 De-transformed residuals (kg) plotted against weight estimates (kg) produced from the above 95th percentile BMI formulae using proximal femoral epiphyseal breadth (left) and proximal tibial epiphyseal breadth (right).

When single sample t-tests were conducted in order to determine if the residuals differed significantly from zero, none of the J value formulae produced residuals that were significant for any of the samples ($\alpha < 0.05$). Several of the breadth formulae produced residuals that differed significantly from zero (Table 3.15). The only breadth formula that produced residuals that did not differ significantly from zero for all the samples was the one that used distal femoral metaphyseal measurements.

3.6 Discussion

Certain weight estimation parameters produced more accurate weight estimates for the overall sample and for each of the subsamples. The mean standard error (MSE) can be used to calculate 95% prediction intervals for future observations of breadth and cortical J measurements and the mean absolute residuals (MAR) can give an idea about

the precision of the formulae. Based on the MSEs and MARs for each of the formulae, it appears as though certain bone parameters such as J at 45.5% of the femur, and J at 75% of the tibia perform better at producing weight estimates within the overall sample. The next section will deal with the most appropriate formulae to use for each subsample.

The formulae using J measurements produced lower MSE values for the majority of the subsamples. This is likely because cross-sectional parameters like J are more sensitive to changes in loading of the body throughout an individual's life, which can be reflected in their weight (Ruff et al 1993; Ruff 2000; Spake et al. 2022). Breadth parameters of the joint are less responsive to changes in loading once the epiphyses are fused and are generally more constrained by biomechanical factors, such as the movement of the joint, while torsional rigidity grows in response to loading and can continue to remodel into adulthood (Ruff et al. 2013). Additionally, the heteroscedasticity present between the breadth and weight values indicates that they may be less reliable predictors of weight than the J values. Formulae using J were generally uniform in the MSE produced for each formula, with the exception of formulae that used J at 80% of the femoral diaphyseal length. The J at 80% formulae generally produced a higher MSE (except for the juvenile and >95th percentile samples) than the other cross-sections. A similar pattern is present among the MAR, wherein the formulae using J at 80% produce the largest mean absolute residuals in each sample (with the exception of the juvenile sample). This is likely due to the cross-section being closest to the proximal end of the femur, wherein cross-sectional J could be influenced by other factors beyond the torsional force applied, such as the positioning of the subtrochanteric region which is often heavily impacted by body proportion in adults (Macintosh et al. 2013; Davie and Stock 2014).

When examining the J formulae for the age and weight subsamples, MSE and MAR values were lowest in the child and below 95th percentile subsamples. This is likely the result of these individuals being the lightest of the overall sample, as lower weight will allow for smaller margins of estimation error. Along those same lines, the adolescent sample consistently produced the highest MSE and MAR values. Interestingly, when comparing the MSE and MARs produced using the below 95th and above BMI 95th percentile, and the samples broken down by age category, the values were generally lower (or comparable) in the BMI samples, indicating it may be more useful to produce torsional rigidity formulae for specific BMI or weight groups, rather than age groups, for

children. The range of maximum and minimum J values presented for each formula should be consulted when determining which formulae is best suited for estimating weight.

Among the formulae produced using breadth measurements the femoral metaphyseal breadth generally produced the lowest MSE and MAR values across all the subsamples, but surprisingly, formulae developed using the distal metaphyseal breadth and the proximal metaphyseal breadth of the femur produced similar values. In the tibia, the metaphyseal breadth-based formulae produced smaller MSE and MAR values than the epiphyseal formulae, likely due to the difference in formation periods between these parts of the bone. There is also a higher potential for measurement error of the epiphyses, as these measurements were taken from CT images of bone *in situ*, in which the maximum breadth of the epiphyses can be more difficult to visualize. The femoral breadth formulae were less consistent, except for the distal femoral breadth which, as mentioned previously, generally produced the lowest MSE and MAR values.

The MSE and MAR values produced using the breadth formulae for the sex, age, and weight subsamples demonstrate a similar pattern as those produced by the J formulae. The MSE and MAR values are generally consistent across the overall sample, and male and female samples. The MSE and MAR values increase as the age category sample increase. From smaller to larger along with the age category samples, The MSE and MAR values are also smaller in the <95th percentile sample and higher in >95th percentile sample. Where the patterns diverge is in comparing the age category MSE and MAR values to those produced by the BMI sample formulae. The formulae for the youngest age category produce MSE and MAR values much lower than those from the below 95th percentile sample (and much lower than those produced in the J formulae from the child sample or the below 95th percentile sample). Conversely, the juvenile and adolescent breadth formulae produce MSE and MAR values higher (in the case of the adolescent formulae, much higher) than either the below 95th percentile or the above 95th percentile formulae. The juvenile and adolescent breadth formulae also produce much higher values than the J-based formulae for the same subsamples. All this could indicate that the breadth measurements are useful in developing weight estimations in young individuals, but after infancy and early childhood, J measurements become better estimators of weight. These results also potentially indicate that developing age-specific formulae is more beneficial in younger individuals, but in older individuals BMI based

categories, using either breadth or torsional rigidity measurements may produce more precise estimates. The tendency of J values to act as better body mass predictors of age is consistent with the findings of Spake and colleagues (2021), when they tested the viability of J and breadth formulae (Robbins-Schugg et al. 2013). Spake and colleagues found that these formulae consistently underestimated the weight, but the J formulae were less biased towards this, particularly in older children (2021).

The mean residuals (MR) give a better understanding of any bias towards over or under estimating weight associated with the formulae and the sample. The bias associated with the logged formulae suggests that the formulae may not be suitable for individuals that fall outside the normal distribution of the breadth measurement being used. If an individual's breadth measurement does fall in the normal distribution, the benefit to the logged breadth formulae is that correlation values between weight and breadth are higher and MARs are lower than those produced by unlogged breadth formulae. Formulae that produced residuals significantly different from zero should be avoided for each sample (Table 3.16). The tendency for lighter individuals to be underestimated and heavier individuals to be overestimated should be kept in mind when using the log formulae.

In the ANCOVA analysis, sex did not appear to differ significantly among the weight estimation parameters except for J at 80% of the femoral diaphysis. However, the interaction between sex and weight status differed significantly for J values taken from 45.5% of the femoral diaphysis. As a result of these differences, it would be beneficial to use the sex-specific formulae, if using the J at 80% of femoral diaphysis, though these formulae generally produced the highest MSE and MR values. Additionally, as growth trajectories differ between males and females after puberty, using the general formulae for adolescents may have additional error. The J formulae will likely continue to be more sensitive to loading via body weight for both males and females when compared with formulae using breadth measurements. Future research should examine whether differences in sex impact the accuracy of J based formulae in subadults after puberty. Weight status did appear to differ among many of the J measurements, including J at 25% and 75% of the femoral diaphysis and at 50% and 75% of the tibial diaphysis. Based on the ANCOVA results, as well as the MSE, MR, and MAR values, the J value formulae for 45.5% of the femur and 75% of the tibia may be more effective when used on individuals who fall above the 95th percentile of BMI, though most J values should be

more sensitive to changes in body weight when compared with metaphyseal or epiphyseal breadths. In terms of individuals that fell within the 95% percentile for BMI the tibia at 75% produced the lowest error among the J formulae, and demonstrated no bias in the MRs, unlike all the breadth formulae.

When comparing the formulae produced from this sample with those produced by Ruff (2007) and Robbins and colleagues (2010) from the Denver Growth Sample it is difficult to draw a direct comparison, as the formulae produced by both were age specific, and only examined three sections of the femur, whereas the nature of the New Mexico collection meant that sample size would not be large enough to create formulae for each age, but the CT scans allowed for formulae to be produced using nine portions of the femur (4 J-based and 5 breadth-based) and seven sections of the tibia (3 J-based and 4 breadth-based). The use of a modern CT collection also allowed for the sampling of a wide variety of weights for age, whereas children in the Denver sample tended to fall within the 5^{th} to 95^{th} percentile for weight. Additionally, inverse calibration regression was used in both studies, meaning mean standard error must be compared with standard error of estimate measures Standard error estimates are based on the predicted mean for the entire sample, rather than mean standard error which is produced based on the prediction error of individual future estimates. Standard error of the estimate was calculated by Ruff (2007) and Robbins and colleagues (2010) as they used inversely calibrated regression models, whereas this research uses classically calibrated regression models. The difference lies in directly solving the formulae for body mass (inverse) or rearranging the equation to solve for body mass (classical) (Ruff 2007). Thus, when calculating error, SEE represents the dependent variable as body mass, whereas MSE represents the independent variable as body mass. Standard error of estimate (SEE) produced by Ruff's (2007) and Robbins and colleagues' (2010) formulae were smaller than those produced by the formulae developed in this study. Ruff's SEE ranged from 0.64 to 8.74 kgs for ages 1 to 13 when using formulae developed from the distal femoral metaphyseal breadth, and 1.41 to 7.01 kgs for ages 7 to 17 using formulae developed from the femoral head breadth. Robbins and colleagues' J formulae produced SEE ranging from 0.27 to 7.84 for ages 1 to 17 years old. Comparatively, MSE from formulae produced from the modern sample were much higher, the lowest of which was 4.77 kgs for the child subsample. The larger MSE values are the result of larger weight ranges being included in each subsample, mostly as a product of the age ranges

covered and the changes in weight that occur therein, but also due to the BMI of the modern sample.

When comparing the samples that were used to create each set of equations, a difference can be observed in the body size between the sample of the Denver Growth Cohort and the individuals that comprised the modern New Mexico sample. According to Ruff (2007), individuals within the subsample used in the formulae were within the 95th percentile of BMI for white children from their time-period (Must et al. 1991). The individuals who are part of the New Mexico sample were specifically chosen to fall above, below, and within the 5th to 95th percentile of BMI, which reflects the prevalence of obesity within modern North American populations (Troiano et al. 1995; Ogden et al. 2008; Komlos et al. 2009). The higher MSE produced by this study's formulae most likely result from the use of age aggregate formulae rather than age dependent formulae, wherein there was increased variation in weights present in the age aggregate subsamples. For example, there may be individuals weighing between 3 and 20 kgs in the child subsample, whereas for the equations generated by Ruff (2007) there is a formula for each year of 1-5, reflecting the weight variation present only in one year. For a 5-year-old using the femoral metaphyseal breadth measurement to predict weight Ruff's equations produce a SEE of 1.08 kgs, whereas the formulae from this study produce a MSE 4.73 kg. However, 42% of the 24 children in the child subsample are above the 95th percentile for BMI, therefore it is likely that this difference may also be reflective of an increased variation in BMI present among this sample when compared with the Denver Growth Cohort.

Additionally, the Denver Growth data used by both Ruff (2007) and Robbins and colleagues (2010) are from a longitudinal cohort study, which may be beneficial in that the same individuals were measured continuously over a set time-period, but there are also disadvantages. The use of the same 20 individuals over the course of birth to 18 years old means that there is only a small amount of variation in growth trajectories captured within the sample. All error then captured within the analyses is error that has occurred within estimating the weight of the same 20 individuals, as opposed to the larger sample of 77 used in this study. Thus, the higher MSE values produced by the formulae from the New Mexico sample are likely a more realistic error prediction when the formulae are used to estimate the weight of a random individual, who is more likely to be represented in the variation of data present in the modern sample.

A novel aspect of this research is the utility of the tibia in developing weight estimation formulae for subadult individuals. The development of formulae based on the weight estimation parameters outlined in this study also allows for the examination and comparison of the performance of different sections of the lower limb and across different types of growth (breadth vs. J values) in weight estimation. While the midsection of the femur is the most precise weight estimator, torsion rigidity of the tibia is comparable in terms of the weight estimates produced; a novel result, as other studies have yet to examine the use of tibial J values in estimating weight in children. The consistency of estimates across the lower limb is surprising, considering the expectation that the more distal elements of the limb will demonstrate more growth plasticity. Proximal tibial metaphyseal breadth formulae also produced similar MSE values to the distal femoral metaphyseal breadth in the overall, female, male, and child samples. Both these connections would suggest a relationship between the development of the distal tibia, either in response to loading of the limb or as a result of the biomechanics required to walk. As Ruff notes (2007) the relationship between femoral metaphyseal breadth and weight develops most during the period where walking is initiated, then begins to fall off during late childhood and adolescence, a pattern that is present in the increasing standard error values between the child, juvenile, and adolescent subsamples. The developments in the distal femur are mirrored in the proximal tibia (with partial and complete fusion occurring in adolescence, though slightly later in the tibia) (Ottow et al. 2017), which could explain the observed similarities in correlation and MSE between the two regions.

Throughout growth and development, the diaphyses of both the femur and the tibia develop in response to both biological and mechanical constraints of growth. Biological constraints, where growth is influenced by periosteal development and endosteal resorption, and mechanical constraints, where cortical distribution begins to reflect loading of the bone (Carter and Beaupré 2001). Periosteal surfaces are more sensitive to loading than endosteal surfaces, and thus measures like torsional rigidity (J) are likely to reflect an individual's loading history, which includes their weight (Ruff and Hayes 1983). Robbins and colleagues (2010) developed their method for estimating body mass using the mid-shaft of the femoral diaphysis, but there have been no formulae for estimating body mass for children that have used tibia cross-sectional properties. The results of this study suggest that the tibia can act as a similar predictor of

weight as to the femur. This is likely due to both bones undergoing similar biological and mechanical constraints with regard to body mass and physical activity level.

Though the formulae developed from the modern New Mexico sample have a larger MSE associated with their body mass estimates than those put forth by Ruff (2007) and Robbins and colleagues (2010), there are several advantages to the utility of these formulae. Firstly, both previous methods require the estimation of age from dentition prior to the use of the formulae, whereas this method allows for no age estimation, or for a broader estimate of the individual as being a child, juvenile or adolescent. While accuracy is likely reduced with the use of larger age categories precision of the estimate is larger. Secondly these formulae can be used on different cross-sectional areas of the femur and tibia, meaning that if the midsection or distal metaphyseal of the femur is missing, it would be possible to estimate age using another parameter. It is suggested that the mean standard error associated with each formula and any bias in the residuals are considered when selecting a formula, and with this in mind most of the adolescent formulae derived from breadth measurements are not suitable for use. There is potential for use of these formulae in forensic populations where preservation and completion of the individual or even the bone are not consistent. These formulae are also especially applicable for use in modern forensic cases in North America, as the sample from which the formulae were derived came from a forensic collection, and are more inline with contemporary height, weight, and BMI among individuals than the Denver Growth study.

3.6.1 Conclusion

This study aimed to create a series of weight estimation formulae based on a modern, 21st century population of subadult individuals, based on several breadth and J measurements taken from the femur and the tibia for use in modern forensic contexts. The nature of the sample indicates that the formulae would be most appropriate for estimating weight among populations of modern subadult individuals. Also, these formulae would be useful in instances where the femur is not present or where sections like the full head breadth or distal metaphyseal breadth are absent. Formulae based on nine measurements from the femur and seven measurements from the tibia were produced for the overall sample, and several sex, age-category, and BMI subsamples. Formulae developed using J values generally had the lowest MSE and MAR values and

demonstrated an increased utility for individuals with BMI falling above the 95th percentile than the breadth formulae. These formulae generally underestimated weight in younger and lighter individuals, and overestimated weight in older and heavier individuals. The formulae produced using the breadth measurements were not as precise, and demonstrated a bias in mean residuals, likely resulting from the heteroscedasticity between the sex and age categories. This bias is also likely the result of de-transforming these data. Based on these results it is suggested that the J formulae are more useful when it comes to accurately estimating weight in subadult individuals. Estimating weight is an important aspect of identifying an individual based on skeletal remains, particularly within forensic contexts.

In terms of the new cross-sections examined within this study, the tibia performed similarly to the femoral midsection when it came to acting as a weight estimation parameter. While the femoral mid-shaft remained generally the most precise weight estimation parameter, tibial cross-section formulae could be used to estimate weight if the femoral midpoint were not available. Additionally, formulae using several sections of the femur demonstrated comparable MSE, MAR, and MR values to those produced by femoral mid-shaft formulae, and thus could be used if the mid-shaft was not available or was damaged. Therefore, for subadult diaphyses, it appears that body size is well reflected in torsional rigidity across the entire shaft, not just at the midshaft.

Chapter 4. Cortical Bone Distribution, Biosocial Stress, and Mechanical Loading during Growth and Development

4.1. Abstract

This study examines the relationship between biosocial stress, mechanical loading, and amount and distribution of cortical bone. To analyze this relationship, the study addresses two main research questions. The first question asks if there is a difference in cortical distribution between individuals in an early 20th century Portuguese sample and an early 21st century New Mexico sample. The second asks if there is a relationship, within each sample, between socioeconomic status and cortical bone distribution. The study relies on cross-sectional bone data from known sex and age children from two documented reference collections; the Luís Lopes Documented Skeletal Collection in Portugal and the New Mexico Decedent Image Database. The New Mexico sample is comprised of 45 children ranging in age from birth to 19 years of age (20 male and 25 female), and the Lisbon sample includes 61 children from the same age range (33 male and 28 female). The overall cross-sectional size of elements from the children in the New Mexico sample was significantly larger than those in the Lisbon sample. When standardized for size, it was determined that the children in the Lisbon sample had thinner cortices and larger medullary cavities relative to the children in the New Mexico sample. The more physically active children of the Lisbon sample demonstrated more rigid and ovoid-shape bones in the upper limb when standardized for size than the more sedentary New Mexico children. The results from the lower limb were more mixed, likely due to the children from the New Mexico sample being heavier than those from the Lisbon sample. When examining the relationship between socioeconomic status (SES) and cortical bone patterning, the majority of high SES individuals had larger cortical areas than the low SES individuals, though most differences were not significant. The size-standardized results were less clear cut, and each sample demonstrated a different pattern. The findings presented in this paper suggest a complex relationship is present in the development of the cortices of long bones of children. Specific bones and sections of the diaphysis appear to react uniformly in response to biosocial stress, but variably with regard to differences in mechanical loading.

4.2. Introduction

The study of bone growth and development within biological anthropology can be conducted along several different trajectories of ontogeny. Researchers have examined well established markers of growth among children in the form of height, weight, and dental development (Green et al. 1946; Bass 1979; Scheuer and MacLaughlin-Black 1994; AlQahtani et al. 2010; Cunningham et al. 2016). Forms of growth based on longitudinal bone growth like long bone length and dental development tend to fall into predictable growth patterns, either in relationship to chronological or developmental age. For example, it is possible to estimate the chronological age of a child to within a small margin of error based on measurements of tooth or femur length (Cardoso et al. 2014; Cardoso et al. 2019).

Another area of bone growth that has been examined in children is appositional growth of the cortices in long bone diaphyses. Long bone cortical distribution in adults is often studied by examining the cross-section of a bone at different intervals along the diaphysis (Macintosh et al. 2013; Davie and Stock 2014). Cross-sections can give researchers information on bone size, shape, and robusticity based on the amount and distribution of cortical bone along the diaphysis. The development of the cortex can be influenced by a multitude of environmental factors including nutrition, disease load, physical activity level, and body mass. Factors like high disease load and undernutrition tend to result in decreased cortical development and increased medullary cavity size (Garn et al. 1964; van der Sluis and de Muinck Keizer-Schrama 2001), whereas increased bone loading through physical activity and body mass will result in more robust, ovoid shaped long bones (Ruff et al. 1994; Liebermann et al 2001; Osipov et al. 2016). How these factors interact with, or counteract, each other to produce the cortical distribution present in a long bone cross-section is not well understood, especially in children. Additionally, while the impacts of nutrition and disease stressors on cortical bone development are well known (Garn et al. 1964; Garn et al. 1969; Gowland 2015; Maratova et al. 2017) whether the socioeconomic status of an individual can be examined by looking at cortical bone has been suggested (Mays et al. 2009), but is still less well understood than the well-established relationship between socioeconomic status and long bone growth (Bogin 1988; Pinhasi et al. 2006).

The continued remodeling of cortical bone during an individual's lifetime means that the distribution is plastic and influenced by environmental factors (Robling et al. 2006). Lasker (1969) proposed the idea of individual-level, biological plasticity in response to the environment during growth and development as being an additional form of adaption; adaptations resulting from genetic evolution and acclimatization being the other two. The earliest studies examining human biological plasticity examined the effects of temperature, climate, latitude, altitude, and environmental pollutants on growth and development (Schell and Rousham 2022) while later studies in corporate more social factors like migration, socioeconomic status, and obesity (Schell et al. 2009). Biological plasticity refers to an organism's ability to adjust its morphology and behaviour to its environments (Bogin 1999; Cardoso 2005). In this case, biological plasticity helps to explain how cortical bone morphologies reflect the body's energy allocation during development. During ontogeny if an energetic trade-off is present, bone development may be altered in favour of uninterrupted growth in other areas of the body (Pomeroy et al. 2012). Development Origins of Health and Disease (DOHaD) theory explains that stressors (like malnutrition or disease) experienced early in life can set developmental trajectories well into adulthood (Kuzawa 2005; Gowland 2015; Temple 2019).

Several factors have been demonstrated to influence the amount and distribution of cortical bone in human long bone diaphyses. For the purposes of this paper, the factors will be separated into those pertaining to biosocial stress and mechanical loading. When an individual is under biosocial stress, the energetic requirements of normal bone growth may not be met, and cortical bone apposition is adversely affected. Protein-caloric deficiencies have been demonstrated to negatively influence overall attained stature and have been shown to similarly impact cortical bone deposition (Garn et al. 1964; Huss-Ashmore 1981; Brenton and Paine 2007). However, cortices in the long bones will continue to grow for a longer period than bone length and articular dimensions (Ruff 2004a; Humphrey 1998). Cortical mass loss during growth due to nutritional deficiencies is the result of bone resorption along the endosteal surface (Eleazer and Jankauskas 2016). Increased disease loads present during development also negatively impact growth. Long term exposure to pathogens and/or chronic disease leads to an increased inflammatory response by the immune system, which has been demonstrated to cause energetic trade-offs leading to stunting in bone growth or lowered bone mineral density (Gowland 2015; Maratova et al. 2017). Decreased bone mineral

density can be a side effect of chronic childhood diseases (van der Sluis and de Muinck Keizer-Schrama 2001). Long term biosocial stress can create systemic disruptions in bone formation and remodeling processes, unlike mechanical loading, which has local effects on bone remodeling (Eleazer and Jankauskas 2016).

Cortical bone deposition can be impacted through habitual mechanical loading of the long bones. Plasticity in periosteal remodelling and overall cortical bone growth are influenced by mechanical loading during childhood and early adulthood (Ruff et al. 2013). The factors that have been shown to affect cortical bone during adolescence are mechanical loading through habitual physical activity and via body mass (Ruff et al. 1994, Liebermann et al 2001, Osipov et al. 2016). The theory behind the plastic response to mechanical loading of long bones was drawn from engineering research conducted on hollow beams (Ruff 2003a) and Wolff's Law, which states that where bone is needed it will be added, and where it is not it will be lost (Ruff et al. 2006). Therefore, when different habitual patterns apply different loading forces on the bone (i.e., shear, torsion, compression), the portion of the diaphysis where the most force occurs will experience an increase in cortical bone deposition. This can lead to increases in overall robusticity, which is defined as a measure of specific and quantifiable patterns of bone shape and geometry reflecting the direction and magnitude of applied forces (Ruff and Hayes 1983, Pomeroy et al. 2018). The mechanical stress and strain that influence bone development will therefore be called 'mechanical loading', for the purposes of this study.

Low socioeconomic status can impact both biosocial stress and mechanical loading. Factors like poverty and marginalization can create biosocial stress via limited access to nutrition or health care, and increased disease load. Low socioeconomic status has been shown to be associated with thinner cortices and an increase in the relative size of the medullary cavity (Mays et al. 2009; Newman and Gowland 2017). It is also possible that lower socioeconomic status can affect mechanical loading, as depending on when and where the child grows up, the amount of physical activity they undergo could be influenced by their social status. In some contexts, this could mean a poorer child may be more active as they need to work for the family, or it could mean a more affluent child would be more active as they would have more opportunity to engage in extracurricular physical activity. Mays and colleagues (2009) examined the impact of low socioeconomic status on cortical bone growth in 19th century English children from Birmingham. They found that endochondral growth of the cortical bone was reduced in

this stressed population when compared to individuals from a higher socioeconomic status group. Newman and Gowland (2017) also examined several areas of bone growth, including cortical thickness, among 18-19th century children from London. While they found that cortical development was negatively affected in children from low socioeconomic status groups, they also observed that the conditions of child health were poor enough to negatively affect higher socioeconomic status individuals from the same time period. Additionally, appositional cortical growth was significantly lower among children from 18th -19th century London compared with modern cortical thickness standards (Newman and Gowland 2017).

As bone is influenced by both biosocial stress and mechanical loading during development it is important to understand the relationship between the two. Mechanical loading tends to lead to bone deposition on the periosteal surface in order to strengthen bone against habitual loading, whereas biosocial stress tends to impact the endosteal surface, leading to greater cortical resorption, and increases the size of the medullary cavity (Eleazer and Jankauskas 2016). In children, there is also the possibility of metabolic factors, like undernutrition, decreasing the amount of cortical bone formed in the first place (Devlin et al. 2010). The differential affects of biosocial stress and mechanical loading on developing cortical bone are still largely unclear, especially when compared with linear bone development. Additionally, the research that has been done on comparing modern and historical children with known differences in socioeconomic status, has not examined differences in physical activity level, and body mass via cross-sectional geometry (Mays et al. 2009; Newman and Gowland 2017). This research attempts to tease apart the effects of biosocial stress and mechanical loading on cortical bone plasticity in the long bones. The unique nature of the skeletal collections used allows for research that is more informed on the growth environment than what is available for archaeological samples, where individual level is rarely available. However, the information gained can help to further examine the relationship between the growth environment and cortical bone distribution in archaeological populations.

The goal of this research is to explore the relationship between biosocial stress, mechanical loading, and amount and distribution of cortical bone in the femur, tibia, and humerus. The study examines two main research questions; the first question asks if there is a difference in cortical bone distribution between individuals in the early 20th century Portuguese sample and the early 21st century New Mexico sample; second asks

if there is a relationship within each sample between socioeconomic status and cortical bone distribution. Cortical bone proportion and distribution will be measured via cross-sectional geometric variables, including cortical bone area, medullary cavity area, torsional rigidity; and cross-sectional shape, and socioeconomic status will be measured using multiple demographic variables. Answers to these two research questions will aid in understanding the relationship between stress, loading, and cortical bone plasticity in children. Understanding how the environmental factor impact cortical bone distribution among children in this sample can lead to increased knowledge the environmental factors that would impact the cortical bone of children in past archaeological populations.

To examine the two research questions this study tests two main hypotheses, each corresponding to one of the questions. The first hypothesis is that individuals in the historic Lisbon sample, coming from a relatively more biosocially stressed, and more physically active population, will exhibit proportionally lower cortical area values, proportionally higher medullary cavity area values, higher torsional rigidity values, and will have a more ovoid cross-sectional shape. Conversely, individuals in the less active, less biosocially stressed, modern New Mexico sample should demonstrate proportionally higher total cortical area values, proportionally lower medullary cavity area values, lower torsional rigidity values, and a more circular-cross-sectional shape. Though children in the historic sample were likely more active than the modern sample, the modern children are on average heavier for their age, thus it is important to consider the effect this increase in body mass will have on mechanical loading, particularly in the lower limbs. It is possible then, that the above pattern will be more evident in the upper limb than in the lower limb, as the effect of increased body mass will have less mechanical impact on the upper limb. The second hypothesis is that socioeconomic status (SES), measured through demographic data, will correlate with cortical distribution within each sample. In this scenario high SES individuals in the New Mexico sample are taken to be less biosocially stressed and more physically active and are expected to demonstrate higher torsional rigidity values, higher cortical area values, and lower medullary cavity area values, whereas low SES individuals in the New Mexico sample are expected to be more stressed and less activity and exhibit the opposite patterns. The opposite pattern for torsional rigidity is expected to be observed in the Lisbon sample, wherein higher SES individuals would be less physically active and less stressed than their low SES counterparts, meaning that they would have comparatively higher cortical area values,

lower medullary cavity areas, lower torsional rigidity, and less ovoid-shaped cross-sections. The examination of socioeconomic status based on individual level data and cortical bone distribution would allow for another layer understanding of the nuance present within samples, as opposed to only examining the differences between them.

4.3. Materials and Methods

4.3.1. Historical Context

This study relies on cross-sectional bone data from known sex and age children from two documented reference collections; the Luís Lopes Documental Skeletal Collection and the New Mexico Decedent Image Database. Since this researcher is interested in the relationship between biosocial stress, mechanical loading, and cortical bone distribution, it is important to establish the potential environmental (i.e., biosocial, and mechanical) factors present in both samples. The sample populations are spatially and temporally distinct, and there will necessarily be different sources and levels of biosocial stress and mechanical loading present in each. Historical context and documented data from individuals will be used to establish the relative biosocial stress and mechanical loading for each sample.

The Luís Lopes Documented Skeletal Collection is housed in the National Museum of Natural History and Science in Lisbon, Portugal. The collection consists of individuals who lived in Lisbon during the early and mid 20th century. Birth years range from between 1904 and 1962. Despite being a recent historical skeletal collection, it is comparable to bone archaeological samples in that the sample consists of individuals who would not have had access to modern medical care, vaccines, or antibiotics. The New Mexico sample is composed of wet bone specimens taken from the pre-autopsy CT scan collection of the Office of the Medical Investigator (OMI) in New Mexico, USA. The OMI partnered with the University of New Mexico to create the New Mexico Descendent Image Database (NMDID), where CT scans and supporting documentation are housed. The individuals in the Lisbon and New Mexico collections have associated documentation that provides researchers with age at death, biological sex, cause of death, and additional biographic information such as demographic information that informs the level of biosocial stress and mechanical loading individuals in each sample experienced during life.

There are distinctive social factors that characterize each population, many of which can impact childhood growth and development. Unlike those in the New Mexico sample, children living in early 20th century Lisbon would not have had access to modern medical care or antibiotics (Gooderham et al. 2020).. Infant mortality rates in Portugal were around 134 deaths per 1000 births as of 1910, compared with 5.5 deaths per 1000 births in 2000 (Guardado Moreira and de Castro Henriques 2016), which reflects the general increase in health of children over the 20th century. Sanitation was another issue, as many Portuguese people were moving into crowded urban areas at the start of the 20th century. Children living in Lisbon during the early 20th century were subjected to poor living conditions in overpopulated and unsanitary areas (Cardoso and Garcia 2009). Conditions would not improve until the later part of the 20th century, as by 1950 43% of working-class families in Lisbon still had no piped water, 69% had no electricity, and 81% had no toilet (Moreira 1950). In terms of education, by 1926 Portugal was considered the least educated nation in Europe (Lannon 1987). Census data show that between 1900 and 1911, 45% to 48% of Portuguese children in the 10–19 age group were engaged in paid work (Goulart and Beli 2017). Many of the children in the Lisbon sample were working as paid domestic servants and apprentices. Children in urban centres like Lisbon, often entered the labour market around 12 years of age and would have been at high risk of their increased physical activity leading to growth deficits of the long bones (Cardoso and Garcia 2009). Additionally, even if they were not employed in wage labour, these children would have been expected to engage in a fair amount of household labour (Cardoso and Garcia 2009). These children would have also been walking Lisbon's hilly terrain, which influences torsional rigidity (Harrington and Osipov 2018). Studies examining the Luís Lopes Documented Skeletal Collection have noted that the children tend to exhibit stunting in the long bones (Cardoso 2005), which is consistent with undernutrition as well as possible disease loads. Overall, the children who comprise the Lisbon sample likely faced relatively high levels of biosocial stress and high levels of mechanical loading via a higher level of physical activity.

Children born in late 20th century and early 21st century New Mexico would have faced very different challenges to growth. The impact of modern medical care and access to antibiotic drugs cannot be overstated, as the mortality rates of children dramatically changed over the course of the 20th century. The United States began the 20th century with an infant mortality rate of 100 deaths out of 1000 births and ended it

with a rate of 7.2 deaths per 1000 births (Meckel 1990; Hoyert et al. 1999). The decline in infant mortality rates during the 20th century is due to the introduction of modern medical practices and the economic changes that occurred. Despite advancements in child health throughout the 20th century, an increasing secular trend of obesity acts as a major determinant of health outcomes (Sun et al. 2012). The children from the NMDID do not appear to demonstrate stunting amongst their long bones, indicating that they are unlikely to be undernourished (Spake 2020). Children living in contemporary US populations are generally heavier and less physically active than children living in each 20th century Lisbon. Over one third of children in the New Mexico sample were above the 95th percentile for BMI, classifying them as obese. Additionally, there were a few children who were classified as underweight. The presence of children in these extreme weight categories is likely the result of the forensic nature of the sample (Spake and Cardoso 2019). It is important to note that the standard of healthcare in the US could differ between low and high SES groups, as there is no universal healthcare. All those in the New Mexico sample would have been less likely to be involved in waged and non-waged forms of child labour. Overall, when compared with the children in the Lisbon sample, those in the New Mexico sample would have experienced low biosocial stress and low mechanical loading via decreased physical activity, but an increased mechanical load due to heavier weight-for-age.

4.3.2. Socioeconomic Status

Establishing a quantitative metric for measuring socioeconomic status (SES) was required in order to compare with cross-sectional measurements. The assumption was made that an individual qualified as being lower SES was more biosocially stressed and those classified as high SES were less stressed, but that mechanical loading may vary based on social context. The variables chosen for each sample were dependent on the documentation available for each population. In some cases, multiple variables were combined to create a proxy variable for socioeconomic status. Two sets of SES variables were created, one for each sample. Both were reliant on area-level measures of socioeconomic status. The variables being used for each sample are described below.

For the Lisbon sample, parental occupation, cause of death, and birth *freguesias* (parishes of birth) information were collected for everyone. The birth *freguesias* variable

was adapted from the demographic information collected by Cardoso (2005) to categorize individuals from the Lisbon sample into either low or high SES is based on place of birth (Cardoso 2005, Cardoso 2007; Conceição and Cardoso 2011). Demographic information for birth *freguesias* were grouped for identification of representative socioeconomic dimensions and then the new variables (factor scores) were subjected to cluster analysis and *freguesias* were grouped into high or low SES groups (Cardoso 2005). This classification was based on four demographic variables available for each *freguesia*: “illiteracy rate”, “mortality due to diarrhoea and enteritis in children under 2 years of age”, “mortality due to pulmonary tuberculosis”, and “mortality due to violent events” (Cardoso 2005: 112). Within the Lisbon sample there were 37 individuals that were categorized into either low (n 10) or high (n 27) socioeconomic status (Table 4.1). Since there are nearly three times as many individuals in the high SES group, it is possible that more variation is captured in the high SES group than the low group.

Table 4.1 Sample composition of the Lisbon and New Mexico samples divided by age category, sex, and socioeconomic status group.

	Lisbon			New Mexico		
	Age	***Sex***	***SES***	***Age***	***Sex***	***SES***
	>6	*Females*	*Low SES*	*>6*	*Females*	*Low SES*
	8	22	10	20	15	19
	6 – 12	*Males*	*High SES*	*6 – 12*	*Males*	*High SES*
	9	15	27	17	25	21
	>12			*>12*		
	20			3		
Total	37			40		

Each individual from the New Mexico sample have a partial zip code on file which allows for the collection of detailed census data for zip code areas. Most individuals had an accessible 5-digit zip code, allowing for several variables pertaining to the socioeconomic status of the zip code area to be collected from the US census. The census statistics ‘Percent of individuals living below the poverty line’, ‘median household income’, and ‘percent of individuals with no high school diploma’ were collected using zip codes of all but two individuals in the New Mexico Sample. The SES classification used to sort individuals into either low or high SES in the New Mexico sample was based on

the value of an individual's zip code data falling below the median percent of individuals living below the poverty line (17.35%), the median household income ($51,228), and the median percent of individuals with no high school diploma (9.66%) calculated for the sample. For example, if the demographic data associated with Case000407 stated that the median income for their zip code was $18,000 per year, they would then be classified as low SES for the median income census variable. If an individual fell below the median value for a minimum of two of the three census variables, they were categorized as low SES overall and if they fell above the median value for two of the three census variables they were categorised as high SES. The majority of individuals fell either above or below the median value for all three census variables. Within the New Mexico sample, 40 individuals were categorized into either low (n 19) or high (n 21) socioeconomic status.

4.3.3. Computed Tomography (CT) Scans

Measurements were collected from CT scans of each individual, one set from individuals in the Lisbon sample and another from the individuals in the New Mexico sample. The New Mexico sample is composed of 45 children ranging in age from birth to 18 years of age (20 male and 25 female), and the Lisbon sample includes 61 children from the same age range (33 male and 28 female). In constructing the sample, efforts were made to match for sex and age in each sample. The New Mexico sample bone scans were collected by the Office of the Medical Investigator of New Mexico. They were taken with a Philips Brilliance Big Bore 16-slice CT scanner. Slice thickness is 1 mm with 0.5 mm overlap, each individual is composed of roughly 10,000 scans, and CT scans had a spatial resolution of 0.5 mm and a temporal resolution of 53 ms. One DICOM file consists of the entire body of the deceased individual. The Lisbon sample bone scans were collected at the Imagens Médicas Integradas Clinic in Lisbon, using a Siemens SOMATOM Flash 128-row. Slice thickness is 0.6 mm, with an overlap of 0.3 mm, and each set of DICOM files are comprised of roughly 1500 scans. The Lisbon sample CT scans have a spatial resolution of 0.33 mm and a temporal resolution of 75 ms. Each DICOM file consists of one femur, tibia, humerus, and other selected bones from two individuals, meaning there was a minimum of six long bones scanned per file. Both sets of DICOM images were reconstructed as standard bone models by the technician, meaning that despite one set being *in situ*, and the other being of dry bone, both sets

were reconstructed with the bone being the primary object in the scans. Voxel size was 512 for both sets of DICOM images.

DICOM stacks from each individual were uploaded into Dragonfly 2020.1-RC2 DICOM imaging software to isolate the cross-sections. To successfully isolate a cross-section and take measurements from a stack, several alignments needed to be made in Dragonfly. Long bones were positioned along sagittal and coronal planes, as if they had been placed along an osteometric board. This was accomplished by aligning the coronal plane along the posterior aspect of the distal epiphysis, and the sagittal plane along the lateral aspect of the distal epiphysis (see Spake et al. (2020) for a full explanation of how to align CT scans of long bones). Once accomplished, measurements of total maximum length of the diaphysis were taken, and from there, measurements of the percentage of the diaphyseal bone length (e.g., 20%, 35%, 50%, or 80%) were made. In cases where the diaphysis and epiphyses were fused in older individuals, the fusion line was used to approximate the maximum length of the diaphyses. This was not just an external estimate, as the CT scan allowed for the internal dimensions of the fusion point to be examined as well. Cross-sectional images were collected at each slice section for the femur, tibia, and humerus.

4.3.4. Measuring Cortical Bone Patterning

Cross-sectional geometry of the long bones can be taken at any percentage of the total diaphysis length, starting at the distal end (0%) and moving towards the proximal end (100%). There are standard biomechanical lengths along the maximum fused length of any long bone from which cross-sectional measures are often taken from adults (20%, 35%, 50%, 65%, and 80%) (Ruff and Hayes 1983; O'Neill and Ruff 2004b). In children, similar measures have been used, both in fully fused and unfused long bones, except for the distally unfused femur and humerus, where 45.5% and 41%, respectively, of the unfused diaphyseal length is measured (Cowgill 2010; Harrington 2010; Ospiov 2018; Ruff 2021). Capturing these specific cross-sections along the diaphysis of each long bone allows for comparisons of standardized lengths across anthropological biomechanics literature and quantifies intra-diaphyseal mechanical variation within the largest elements (i.e., the femur, tibia, and humerus) (Ruff 2003a; Cowgill 2010; Harrington 2010; Cowgill 2014; Osipov 2018). The sections of the

humerus were examined to be able to look at mechanical loading history without the influence of weight-based loading.

In this study, the sections at 25%, midshaft (41%. 45.5%, or 50%), and 75% were chosen to be the minimum number and position of sections taken from each bone. The femur and the humerus had other sections where measurements were taken, firstly as an attempt to approximate the juvenile diaphyseal midsection (45.5% of the femur and 41% of the humerus), as calculated by Ruff (2003b, 2021). Secondly, another section was added to each to account for the position of the subtrochanteric section of the femur (80%) and the deltoid tuberosity in the humerus (35%). Once captured in Dragonfly, these cross-sections were exported as TIFF images.

The BoneJ plug-in for ImageJ 1.53f51 (Doube et al. 2010; Rasband 2018) was used to process bone images and calculate cross-sectional geometry. Automated measurements were taken from multiple, sequential cross-sectional images. The 2D bone images were converted to black and white, 8-bit images, and aligned along anteroposterior and mediolateral planes. Cross-sectional measurements, including cortical area (CA), medullary cavity area (MA), an index of cross-sectional shape (I_{max}/I_{min}), and torsional rigidity (J) were then calculated. Medullary cavity area was calculated by subtracting the CA from the total cross-sectional area. Total cross-sectional area was calculated by taking an area measurement of the entire cross-section, including the cortex and the medullary cavity (see Figure 2.1). J was calculated by adding I_{max} and I_{min}.

4.5.3. The Development of Standardized Residuals

Standardized residuals were also created for cortical area (STCA), medullary cavity area (STMA), and torsional rigidity (STJ) measurements to allow for comparisons among and between the samples independent of cross-sectional size. As the I_{max}/I_{min} index is already independent of size, a standardized residual was not needed. Area residuals were produced by regressing the $\log_{10}$ of maximum femoral head breadth against the $\log_{10}$ of the CA and MA measurement from each section, and J residuals were produced by regressing the $\log_{10}$ of maximum femoral head breadth multiplied by bone length squared against the $\log_{10}$ of J.

Often standardized residuals of cross-sectional measurements are calculated using a body mass (or a body mass estimate) (Auerbach and Ruff 2004; Ruff 2007; Osipov et al. 2016), but in this case the maximum femoral head breadth was used in the standardization process to account for differences in body size. Both the bending and torsional rigidity of the femur and humerus scale similarly with body mass when it comes to size-standardization, thus the same standardization factor was used in upper and lower limb (Ruff 2000).

The use of femoral head breadth was based on weight estimation formulae developed by Ruff (2007). Though Ruff favours the use of distal metaphyseal breadth for children under eight, regression coefficients were higher among the samples between each measure (CA, MA, J) and femoral head breadth. Additionally, the results of chapter 3 indicate that the mean standard error for weight estimation formulae using the femoral head breadth was lower or comparable with metaphyseal breadth, even among the youngest age group for the New Mexico sample (under 5) (see Chapter 3). The use of a body mass estimate will always introduce some amount of error into the standardization process (Spake et al 2021; Meyers et al 2023), even when the formulae are age specific (Ruff 2007). This error is not present when using a direct measurement that reflects body mass, like femoral head breadth.

It would have been ideal to use a non-mechanical, morphometric measurement, such as bi-iliac breadth since maximal femoral head breadth is influenced by the mechanical history and loading of the bone (Auerbach and Ruff 2004; Ruff 2007). However, bi-iliac breadth can only reasonably be calculated from bones that are still *in situ*, posing a challenge to the measurement of it among archaeological or historical skeletal samples. Femoral head breadth scales with body mass in a similar way to bi-iliac breadth in adults (Auerbach and Ruff 2004), thus despite the mechanical influence on its breadth, femoral head bread was used.

The use of an estimated weight in standardizing the CSG data also suffers from the limitation inherent in Ruff's (2007) equations. These equations were developed using an affluent group of 20 children who had fairly lean body masses but were not malnourished (Maresh 1970; McCammon 1970). The measurements themselves represent little of the variation present in growth, as they only account for these 20 individuals and were taken at specific annual and semi-annual increments throughout

their development. Finally, least squares regression rather than panel regression was used to produce the equations, a method that does not account for the low variation present in the data. In the future it is recommended that researchers use a known measurement alone, preferably a morphometric variable, when standardizing CSG variables, unless body mass is known. This is especially important for cases where both age and body mass are estimated, as the lower amount of error present in an age estimate is further compounded in the standardized residual when combined with the higher (though more variable) error present in the body mass estimation.

The individuals in each sample were not grouped according to Bogin's (1999) life history stages (i.e., child, juvenile, adolescent), in part to due to sample size restrictions, but in part to examine growth across the sub adult period between two samples that may have different developmental trajectories. In the future it would be of interest to compare cortical patterning between the samples among individuals from the same life history stage. Size differences in the cross-sections of individuals in different developmental periods are accounted for via the standardization of CSG measurements.

4.3.6 Statistical Analyses

Descriptive statistics were established, wherein the mean, standard deviation, maximum, and minimum values were calculated for all the size-standardized cross-sectional residuals, as well as for all demographic and socioeconomic variables. This was done so that size-standardized residuals could be compared, particularly in instances where significant differences were found among the cross-sectional measurements. All statistical analyses were conducted in SPSS 24.

Two sets of analyses were conducted to examine the two research questions presented in this study. The first set of analyses focused on examining the differences in CA, MA, J, and I_{max}/I_{min} between individuals in the historic Lisbon sample and those in the modern New Mexico sample. In order to explore these differences several analyses of variants tests (ANOVAs) were conducted on all the size-standardized residuals and I_{max}/I_{min} indices to determine if significant differences were present between the samples for any of the measurements at any of the cross-sections in each bone. In total 52 ANOVA tests were conducted on the size-standardized residuals. This first set of analyses also included the conducting of several analysis of covariance tests

(ANCOVAs) on the raw CA, MA, and J values to examine significant differences between samples and sexes, with age as a covariate. ANCOVA tests were used for the raw values since they were not size-standardized, and thus the potential impact of sex and age differences in the values had to be considered. A total of 39 ANCOVA tests were conducted on the raw values.

The second set of analyses pertained to the second research question, and thus focused on examining differences in socioeconomic status among the individuals from each sample. Here again, ANOVAs were used to look for significant differences in the STCA, STMA, and STJ residuals, as well as I_{max}/I_{min} between low and high SES groups within each sample. A total of 104 ANOVA tests were conducted on the standardized values (52 for the Lisbon and 52 for the New Mexico sample). ANCOVAs were also used to look for significant differences between raw CA, MA, and J values, this time with SES and sex used as fixed factors in the analyses and age and samples used as the covariates. A total of 39 ANOVA tests were conducted on the raw values.

4.4. Results

4.1.1. Differences in Cross-Sectional Variables Between Samples

Four measures of cross-sectional geometry (CSG) were taken from each bone section from each of the femur, tibia, and humerus examined. ,. The CSG variables examined were cortical section area (CA), the medullary cavity area (MA), the polar second moment of area (J), and an index of shape (I_{max}/I_{min}). The sample-specific means, standard deviation, the maximum and minimum values for each of the standardized CSG residuals are listed in Table 4.2 and the results of the ANOVA analyses or the standardized residuals are listed in Table 4.3. Table 4.4 contained the sample-specific means, standard deviation, the maximum and minimum values for the I_{max}/I_{min} ratios, as well as the ANOVA results. The results for the ANCOVA tests comparing the raw CA, MA, and J values between samples are provided for each bone in Tables 4.5, 4.6, and 4.7, respectively.

For the ANOVAs, when assessing homogeneity of variance among the standardized residuals, plots of CGS values against age demonstrated a low amount variation and heteroscedasticity when comparing the samples. Prior to ANOVAs being performed

there were no instances where a section's values did not meet the assumption of homogeneity of variance.

Table 4.2 Sample-specific Mean, standard deviation, maximum, and minimum for the standardized residuals of the cortical section area (STCA), the medullary cavity area (STMA), and the polar second moment of area (STJ) for each section from the femur, tibia, and humerus.

		STCA				STMA				STJ			
		Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
Femur													
F25	Lisbon	-0.352	0.142	-0.023	0.087	-0.300	0.250	0.004	0.125	-0.622	0.254	-0.030	0.161
	New Mexico	-0.057	0.351	0.031	0.092	-0.260	0.400	-0.006	0.117	-0.260	0.545	0.039	0.164
F45.5	Lisbon	-0.329	0.200	-0.002	0.096	-0.360	0.300	0.005	0.142	-0.614	0.532	0.006	0.172
	New Mexico	-0.102	0.405	0.002	0.094	-0.340	0.300	-0.006	0.131	-0.234	0.348	-0.009	0.145
F75	Lisbon	-0.303	0.148	-0.010	0.091	-0.390	0.190	0.015	0.112	-0.445	0.233	0.003	0.130
	New Mexico	-0.117	0.399	0.014	0.094	-0.210	0.280	-0.020	0.111	-0.319	0.271	-0.004	0.150
F80	Lisbon	-0.342	0.272	-0.019	0.109	-0.710	0.320	0.001	0.137	-0.334	0.289	-0.009	0.148
	New Mexico	-0.101	0.371	0.029	0.093	-0.240	0.390	-0.001	0.133	-0.257	0.341	0.012	0.148
Tibia													
T25	Lisbon	-0.237	0.137	-0.027	0.082	-0.260	0.210	0.016	0.119	-0.803	0.552	0.010	0.239
	New Mexico	-0.094	0.403	0.038	0.090	-0.220	0.310	-0.022	0.109	-0.296	0.355	-0.013	0.120
T50	Lisbon	-0.323	0.136	-0.022	0.093	-0.350	0.260	0.028	0.141	-0.783	0.718	0.006	0.259
	New Mexico	-0.146	0.432	0.031	0.106	-0.540	0.230	-0.039	0.131	-0.320	0.356	-0.007	0.140
T75	Lisbon	-0.319	0.148	-0.029	0.085	-0.330	0.260	0.023	0.132	-0.802	0.534	-0.004	0.240
	New Mexico	-0.088	0.371	0.041	0.087	-0.200	0.340	-0.032	0.102	-0.268	0.216	0.005	0.113
Humerus													
H25	Lisbon	-0.199	0.217	0.010	0.100	-0.830	0.250	0.025	0.169	-0.362	0.577	0.060	0.181
	New Mexico	-0.585	0.333	-0.014	0.199	-0.710	0.350	-0.034	0.232	-1.218	0.386	-0.080	0.377
H35	Lisbon	-0.232	0.188	-0.008	0.090	-0.360	0.220	0.004	0.144	-0.354	0.551	0.048	0.180
	New Mexico	-0.456	0.311	0.011	0.130	-0.560	0.350	-0.006	0.154	-1.097	0.398	-0.065	0.251
H41	Lisbon	-0.352	0.153	-0.015	0.091	-0.680	0.230	0.004	0.160	-0.464	0.452	0.003	0.201
	New Mexico	-0.430	0.340	0.021	0.126	-0.420	0.330	-0.006	0.139	-1.081	0.414	-0.005	0.283
H50	Lisbon	-0.483	0.164	-0.019	0.102	-0.330	0.250	0.010	0.137	-0.481	0.573	0.018	0.205
	New Mexico	-0.414	0.312	0.026	0.106	-0.470	0.350	-0.014	0.143	-0.708	0.461	-0.024	0.197
H75	Lisbon	-0.287	0.157	-0.025	0.086	-0.420	0.310	-0.020	0.178	-0.505	0.436	-0.023	0.208
	New Mexico	-0.369	0.287	0.033	0.096	-0.250	0.390	0.027	0.146	-0.369	0.251	0.031	0.150

The standardized residuals that have been developed from each CSG variable demonstrate a pattern where, in general, STCA residuals were higher among the New Mexico sample, STMA residuals were higher among the Lisbon sample, and a more mixed pattern was apparent among the standardized J residuals (Table 4.2). The only STCA section wherein residuals were higher among the Lisbon sample was at the 25% section of the humerus, a difference that was not significant (Table 4.3). The New Mexico sample STCA values were significantly higher at 25% and 75% of the femur, 50% and 75% of the humerus, and all the sections of the tibia (Figure 4.1). All figures are showing the 50% section of the tibia, so as to demonstrate different measurements and the relationship between the sample at the same section.

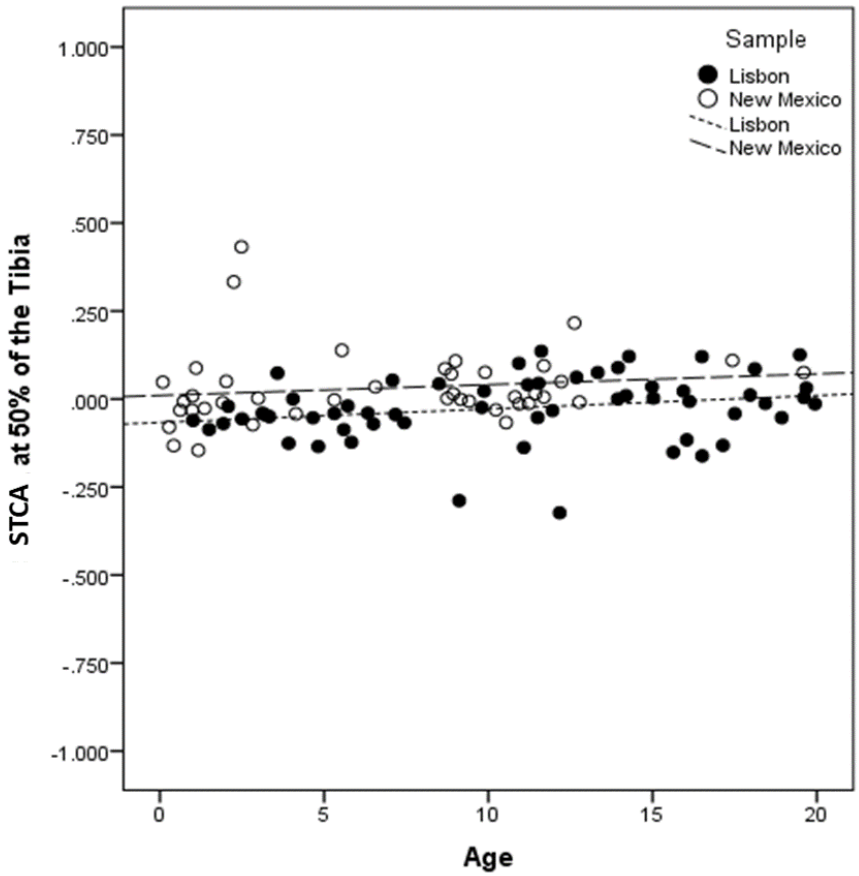


Figure 4.1 Standardized CA residuals of individuals from the Lisbon and the New Mexico samples 50% of the tibial diaphysis plotted against age at death.The New Mexico sample has, on average, higher standardized cortical area than the Lisbon sample.

The STMA residuals demonstrated an opposite pattern to the STCA values. Most bone sections demonstrated a higher average among the Lisbon sample (Figure 4.2);

the only section where the New Mexico average was higher was at the 75% section of the humerus, but this difference was significant. The sections where the Lisbon STMA residuals were significantly higher were at 25%, and 45.5% of the femur.

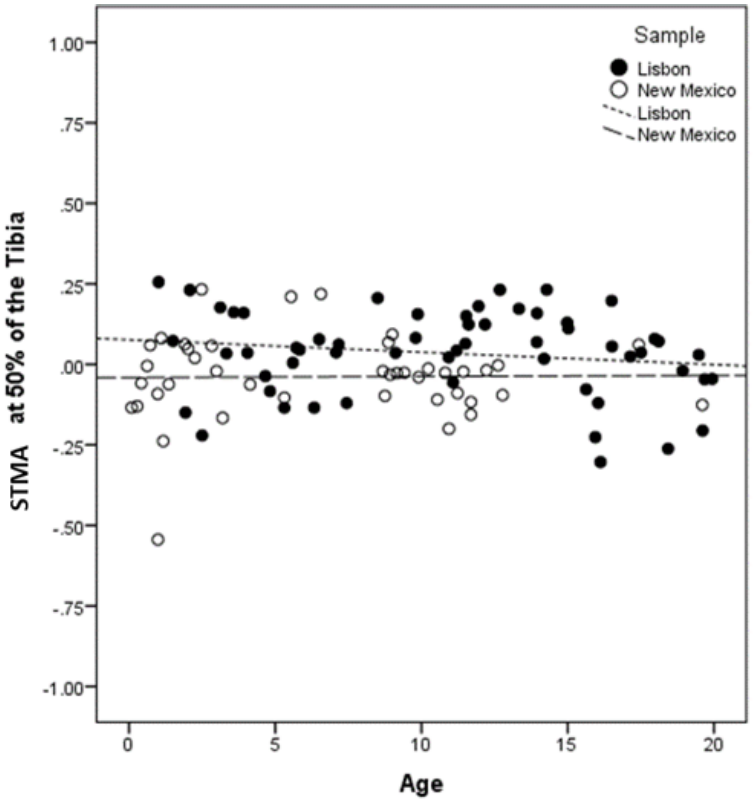


Figure 4.2 Standardized MA residuals of individuals from the Lisbon and the New Mexico samples 50% of the tibial diaphysis plotted against age at death. The Lisbon sample has, on average, higher standardized medullary cavity area than the New Mexico sample.

The STJ values demonstrate a mixed pattern. Among the lower limb bone sections J values were higher for the Lisbon sample at 45.5%, and 75% of the femur, as well as at 25% and 50% of the tibia (Figure 4.3), while the remaining sections demonstrated higher STJ means among the New Mexico sample. Among the upper limb bones the Lisbon sample mean STJ values are highest in all sections except for at 75% of the humerus. None of the differences were significant (Table 4.3).

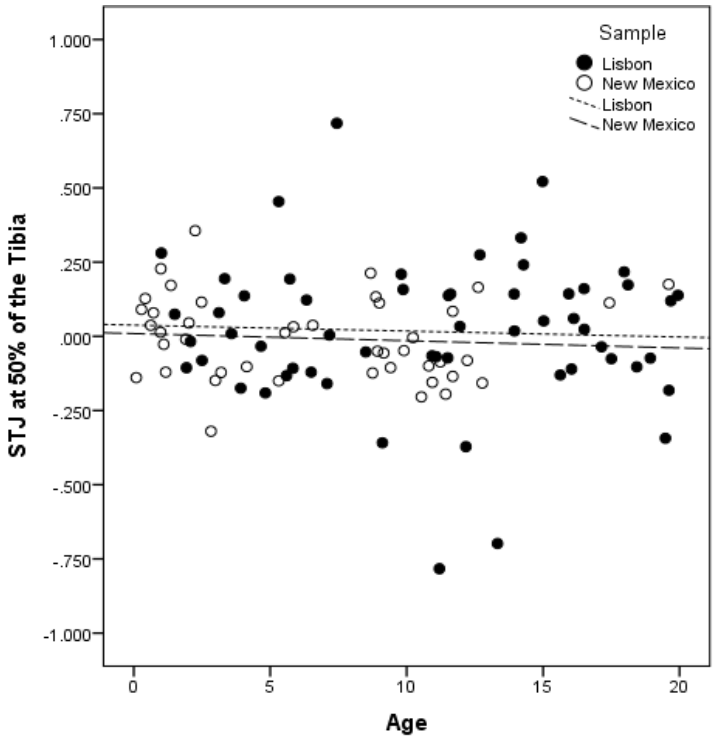


Figure 4.3 Standardized J residuals of individuals from the Lisbon and New Mexico samples at 50% of the tibial diaphysis plotted against age at death. The Lisbon sample has, on average, higher standardized torsional rigidity values than the New Mexico sample.

Table 4.3 ANOVA results for comparing the standardized residuals of the cortical section area (STCA), the medullary cavity area (STMA), and the polar second moment of area (STJ) between the Lisbon and New Mexico samples for each section from the femur, tibia, and humerus.

	STCA		STMA		STJ	
	F	P	F	P	F	p

Femur						
F25	8.86	**0.004**	0.16	0.688	2.45	0.122
F45.5	0.04	0.850	0.15	0.698	0.33	0.567
F75	1.70	0.196	2.43	0.122	0.31	0.579
F80	5.45	**0.022**	0.01	0.935	0.78	0.382
Tibia						
T25	14.32	**0.000**	2.75	0.100	0.12	0.726
T50	7.07	**0.009**	5.78	**0.018**	0.07	0.795
T75	16.03	**0.000**	5.12	**0.026**	0.00	0.967
Humerus						
H25	0.67	0.415	2.17	0.144	3.77	0.056
H35	0.75	0.387	0.12	0.735	0.02	0.884
H41	2.78	0.099	0.11	0.740	0.01	0.937
H50	4.70	**0.032**	0.71	0.401	0.45	0.507
H75	10.21	**0.002**	2.04	0.157	0.24	0.627

Significant results are indicated in bold ($\alpha < 0.05$)

The I_{max}/I_{min} index results, which give an indication of the shape of the cross-section irrespective of size differences, demonstrated a similar pattern as observed in the STJ values (Table 4.4). The lower limb sections were mixed, wherein sections F45.5%, T50%, and T75% demonstrated larger ratio means among the New Mexico sample, while the remainder of the sections had larger mean ratios among the Lisbon sample. There were few instances of significant differences, but they occurred at 75%, and 80% of the femur and 75% of the tibia. The upper limb sections demonstrated no significant differences ($\alpha < 0.05$), and the sections where the indices were higher among the Lisbon sample were at 25%, 35%, 50%, and 75% of the humerus.

Table 4.4 Sample-specific mean, standard deviation, maximum, minimum, and ANOVA results for comparing the I_{max}/I_{min} ratios between the Lisbon and New Mexico samples for each section from the femur, tibia, and humerus.

		Mean	Std. Deviation	Minimum	Maximum	F	Sig.
F25	Lisbon	1.618	0.524	1.072	3.696	0.847	0.360
	New Mexico	1.532	0.389	1.055	2.506		

F45.5	Lisbon	1.413	0.377	1.076	2.657	0.007	0.932
	New Mexico	1.419	0.263	1.021	2.209		
F75	Lisbon	1.565	0.329	1.021	2.530	10.097	**0.002**
	New Mexico	1.392	0.184	1.029	1.802		
F80	Lisbon	1.548	0.297	1.020	2.474	14.707	**0.000**
	New Mexico	1.351	0.201	1.011	1.849		
T25	Lisbon	1.569	0.363	1.023	2.821	0.537	0.465
	New Mexico	1.523	0.247	1.110	2.227		
T50	Lisbon	1.685	0.462	1.060	3.330	2.423	0.123
	New Mexico	1.816	0.380	1.130	2.530		
T75	Lisbon	1.661	0.404	1.120	2.880	7.347	**0.008**
	New Mexico	1.886	0.448	1.100	3.020		
H25	Lisbon	1.715	0.448	1.035	2.909	0.984	0.323
	New Mexico	1.631	0.412	1.019	2.720		
H35	Lisbon	1.599	0.366	1.071	2.723	0.505	0.479
	New Mexico	1.544	0.420	1.127	2.960		
H41	Lisbon	1.579	0.391	1.057	3.281	1.043	0.309
	New Mexico	1.662	0.444	1.009	2.803		
H50	Lisbon	1.611	0.379	1.006	2.952	0.443	0.507
	New Mexico	1.562	0.372	1.039	2.602		
H75	Lisbon	1.839	0.550	1.078	3.006	3.149	0.079
	New Mexico	1.656	0.485	1.053	2.928		

Significant results are indicated in bold ($\alpha < 0.05$)

When examining the ANCOVA results for the raw CA measurements they appear to show a similar pattern to the ANOVA comparisons using standardized residuals, except with more instances of significant differences between samples (Table 4.5). All sections, except for 25% of the humerus, demonstrated significantly higher CA values in the New Mexico sample (Figure 4.4), when sample and sex were defined as fixed factors, with age as a cofactor. No differences among the sexes were determined to be significant, but there were several cases where the interaction between sex and sample demonstrated significant differences between the females in each sample, all in the lower limb. In these cases, the Lisbon males had much higher mean CA values than the

females, but conversely, the New Mexico females had higher mean CA values than the males in the sample. This pattern continued in the upper limb, but differences between the females were no longer significant.

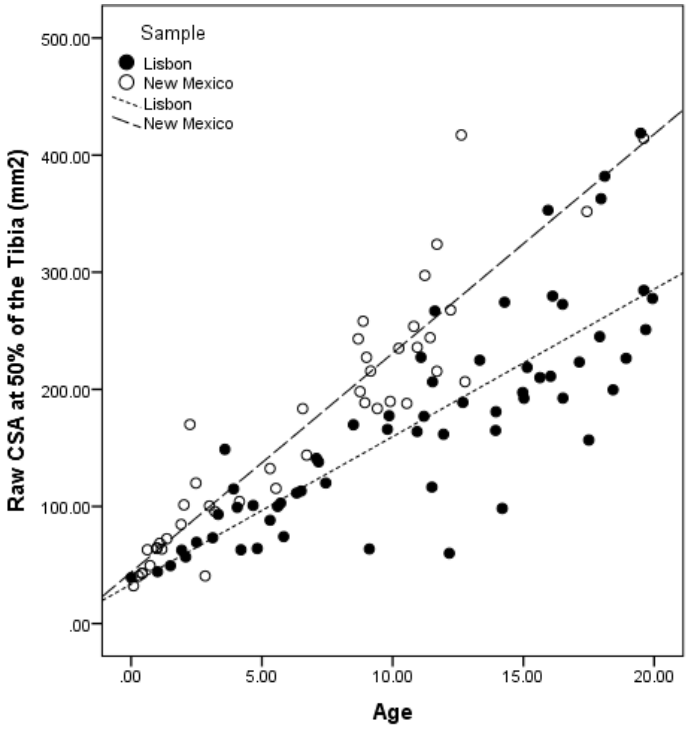


Figure 4.4 Raw CA values (mm2) for individuals from the Lisbon and New Mexico samples at 50% of the tibial diaphysis plotted against age at death.

Table 4.5 ACNOVA results for comparing the raw measures of the cortical section area (CA) between the Lisbon and New Mexico samples for each section from the femur, tibia, and humerus. Sample and sex are the independent variables, the dependent variable is the C*A* at each section, and age is the cofactor in the ANCOVA analysis.

	CA (mm^2)					
	Sample		Sex		Sample*Sex Interaction	
	F	P	F	P	F	P
Femur						
F25	27.30	**0.000**	1.26	0.264	1.66	0.201
F45.5	16.51	**0.000**	0.92	0.340	7.53	**0.007**
F75	20.44	**0.000**	0.06	0.804	7.00	**0.009**
F80	26.88	**0.000**	0.00	0.964	6.68	**0.011**
Tibia						
T25	43.20	**0.000**	0.95	0.332	4.55	**0.035**

T50	33.50	**0.000**	0.52	0.474	4.83	**0.030**
T75	38.81	**0.000**	0.27	0.603	5.68	**0.019**
Humerus						
H25	1.16	0.284	2.76	0.100	0.42	0.518
H35	31.08	**0.000**	1.09	0.299	2.16	0.145
H41	44.36	**0.000**	0.63	0.431	3.06	0.083
H50	40.50	**0.000**	0.33	0.569	2.48	0.118
H75	54.06	**0.000**	0.63	0.430	3.03	0.085

Conversely, the raw MA values demonstrate an opposite pattern as the STMA residuals (Table 4.6). The raw MA values are mostly larger among the New Mexico sample than the Lisbon sample, a difference that appeared significant at 25%, 45.5%, and 50% of the femur, and 75% of the humerus (Figure 4.5). Sex differences wherein males were significantly larger than females were found at 75% of the femur and 75% of the tibia. There were several instances of significant differences for the sample sex interaction, wherein females from the New Mexico sample were significantly larger than females among the Lisbon sample at 25%, 45.5%, and 50% of the femur, and 50% and 75% of the tibia. Like the raw CA values, the Lisbon males had higher mean MA values than the females and the New Mexico females had higher mean MA values than the males in the sample in both the upper and lower limbs.

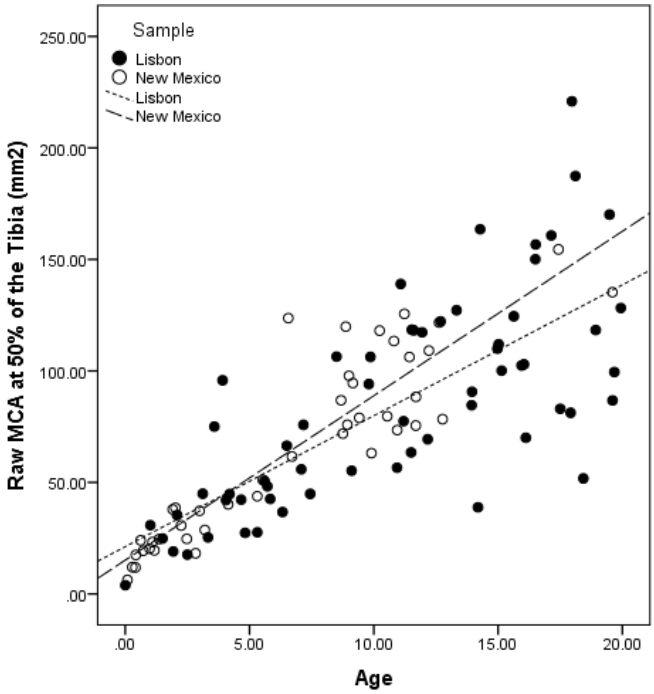


Figure 4.5 Raw MA values (mm2) for individuals from the Lisbon and New Mexico samples at 50% of the tibial diaphysis plotted against age at death.

Table 4.6 ANCOVA results for comparing the raw measures of the medullary cavity area (MA) between the Lisbon and New Mexico samples for each section from the femur, tibia, and humerus. Sample and sex are the independent variables, the dependent variable is MA at each section, and age is the cofactor in the ANCOVA analysis.

	MA (mm^2)					
	Sample		Sex		Sample*Sex Interaction	
	F	P	F	P	F	P
Femur						
F25	5.71	**0.019**	0.00	0.997	6.80	**0.010**
F45.5	8.45	**0.004**	0.00	0.977	6.93	**0.010**
F75	0.03	0.875	5.99	**0.016**	0.29	0.592
F80	1.49	0.225	3.73	0.056	0.05	0.818
Tibia						
T25	0.75	0.388	0.00	0.983	2.00	0.160
T50	0.83	0.365	3.06	0.083	11.92	**0.001**
T75	0.72	0.399	4.38	**0.039**	8.06	**0.005**
Humerus						
H25	0.91	0.343	3.69	0.057	0.11	0.745
H35	3.24	0.075	2.21	0.141	0.43	0.516
H41	1.39	0.241	0.85	0.358	0.91	0.342
H50	1.62	0.206	0.49	0.487	1.29	0.259
H75	6.21	**0.014**	2.45	0.121	3.41	0.068

A similar phenomenon appears among the raw J values as with the raw CA and MA values, wherein the New Mexico sample demonstrates higher values than the Lisbon sample, differences which are significant at all the femur sections, 25% and 75% of the tibia, and at 41%, and 75% of the humerus (Figure 4.6). However, this is decidedly different than the mixed results that appear in the STJ values, wherein the New Mexico sample STJ values are only higher than the Lisbon sample STJ values at 25% and 80% of the femur, as well as at 75% of the tibia and 75% of the humerus, but none of them were significant. There was one instance of males being significantly larger than females, at 25% of the humerus (Table 4.7). Instances where significant differences were present between the females in each sample were present at 45%, 75%, and 80% of the femur. As with the raw CA and MA values, the Lisbon males had higher mean J values than the females in the upper and lower limbs. The New Mexico females had

higher mean J values than the males in the lower limb, but this pattern was not apparent in the upper limb. In the upper limb, mean male J values were higher than mean female J values, though not significantly.

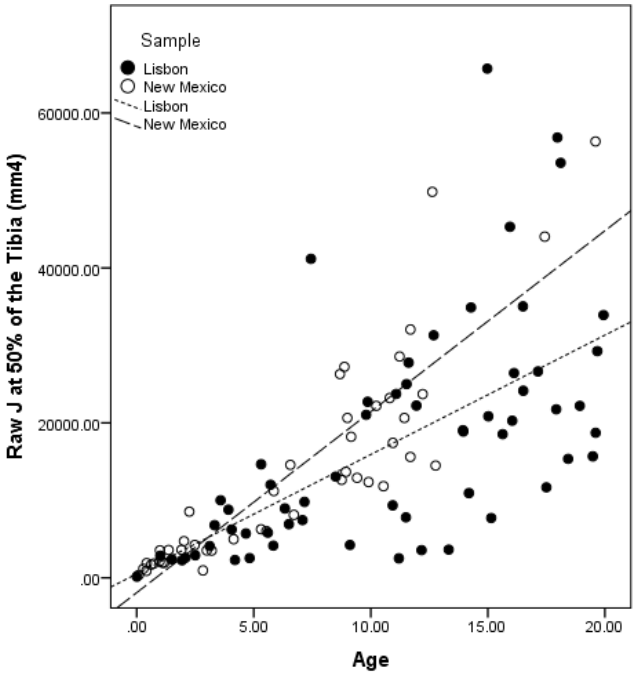


Figure 4.6 Raw J values (mm4) for individuals from the Lisbon and New Mexico samples at 50% of the tibial diaphysis plotted against age at death.

Table 4.7 ANCOVA results for comparing the raw measures of the torsional rigidity (J) between the Lisbon and New Mexico samples for each section from the femur, tibia, and humerus. Sample and sex are the independent variables, the dependent variable is the J at each section, and age is the cofactor in the ANCOVA analysis.

	J (mm^4)					
	Sample		Sex		Sample*Sex Interaction	
	F	P	F	P	F	P
Femur						
F25	14.40	**0.000**	0.56	0.457	2.52	0.116
F45.5	9.52	**0.003**	0.61	0.437	8.92	**0.004**
F75	7.83	**0.006**	1.30	0.257	8.17	**0.005**
F80	9.90	**0.002**	0.86	0.355	6.87	**0.010**
Tibia						
T25	5.52	**0.021**	0.00	0.975	1.93	0.168
T50	3.44	0.067	0.36	0.550	3.30	0.072
T75	5.81	**0.018**	0.06	0.810	3.23	0.075
Humerus						
H25	0.62	0.432	7.06	**0.009**	0.64	0.425
H35	1.18	0.279	1.31	0.254	2.20	0.141
H41	8.63	**0.004**	0.52	0.474	0.15	0.704
H50	3.84	0.053	0.20	0.654	1.05	0.308
H75	8.66	**0.004**	2.12	0.149	1.48	0.227

4.4.2. Cross-Sectional Variables and Socioeconomic Status Among the Samples

Means, standard deviations, minimum and maximum values of the standardized CA (Table 4.8), MA (Table 4.9), and J (Table 4.10) residuals for the low and high SES groups are presented for each sample. Standardized residuals were calculated based on combined samples, so means presented that deviate from zero represent the Lisbon sample mean relative to the combined sample mean of zero. Shape differences between SES groups in the form of size-independent I_{max}/I_{min} ratios among both samples were also calculated (Table 4.11).

ANOVA tests were conducted separately on the Lisbon and New Mexico samples to examine differences of the standardized CSG residuals in accordance with

SES(Table 4.12). Plots of CSG values against age demonstrated a fair amount of variation and heteroscedasticity when comparing the samples. In testing for homogeneity of the variances prior to the ANOVA there were several instances where a section's values did not meet the assumption, mostly among the STCA values from the Lisbon section. The Lisbon STCA values at 25% and 80% of the femur, 25% and 50% of the tibia, 25%, 35%, 50%, and 75% of the humerus did not meet the assumption of homogeneity of variance. Among the STMA values the femur at 25% from the Lisbon sample and the humerus at 25% and 35% from the New Mexico sample did not meet the assumption. Finally, among the STJ values only the femur at 25% and 80% from the Lisbon sample did not meet the assumption of homogeneity of variance.

Table 4.8 Mean, standard deviation, maximum, and minimum for the standardized residuals of the cortical bone area (STCA) for each section from the femur, tibia, and humerus.

		Lisbon STCA				New Mexico STCA			
		Min	Max	Mean	SD	Min	Max	Mean	SD
Femur									
F25	Low SES	-0.078	0.109	0.002	0.074	-0.057	0.351	0.035	0.098
	High SES	-0.352	0.142	-0.031	0.105	-0.057	0.344	0.028	0.092
F45.5	Low SES	-0.083	0.152	0.020	0.085	-0.077	0.405	0.015	0.110
	High SES	-0.329	0.147	-0.017	0.107	-0.078	0.193	-0.009	0.076
F75	Low SES	-0.095	0.148	0.035	0.086	-0.117	0.399	0.016	0.113
	High SES	-0.303	0.116	-0.020	0.096	-0.105	0.229	0.007	0.075
F80	Low SES	-0.123	0.130	0.034	0.086	-0.101	0.371	0.031	0.109
	High SES	-0.342	0.133	-0.035	0.118	-0.097	0.237	0.021	0.074
Tibia									
T25	Low SES	-0.100	0.094	-0.009	0.077	-0.036	0.403	0.057	0.117
	High SES	-0.237	0.135	-0.036	0.095	-0.094	0.159	0.021	0.062
T50	Low SES	-0.132	0.120	-0.008	0.084	-0.073	0.432	0.046	0.133
	High SES	-0.323	0.136	-0.035	0.109	-0.132	0.216	0.021	0.079
T75	Low SES	-0.138	0.077	-0.027	0.079	-0.045	0.371	0.049	0.104
	High SES	-0.319	0.148	-0.030	0.103	-0.088	0.240	0.035	0.075
Humerus									
H25	Low SES	-0.092	0.191	0.074	0.092	-0.344	0.333	0.058	0.139
	High SES	-0.199	0.164	-0.002	0.095	-0.585	0.126	-0.057	0.213
H35	Low SES	-0.109	0.152	0.042	0.082	-0.384	0.311	0.025	0.142
	High SES	-0.232	0.188	-0.017	0.085	-0.456	0.186	-0.006	0.122
H41	Low SES	-0.092	0.120	0.037	0.081	-0.391	0.340	0.026	0.146
	High SES	-0.352	0.153	-0.025	0.093	-0.430	0.189	0.004	0.117
H50	Low SES	-0.110	0.159	0.031	0.087	-0.414	0.312	0.017	0.141
	High SES	-0.483	0.164	-0.034	0.118	-0.067	0.158	0.023	0.069
H75	Low SES	-0.083	0.120	0.028	0.075	-0.369	0.287	0.032	0.129
	High SES	-0.287	0.157	-0.037	0.091	-0.070	0.166	0.027	0.058

The STCA residuals are higher among the lower SES group for most bone cross-sections in both samples (Table 4.8). The only bone sections where the high SES CA means were higher were among the humeri at 50% in the New Mexico sample. The only significant difference among the STCA samples was at 25% of the humerus in the Lisbon sample, wherein the low SES group was significantly higher than the high SES group (Table 4.12).

Table 4.9 Mean, standard deviation, maximum, and minimum for the standardized residuals of the medullary cavity area (STMA) for each section from the femur, tibia, and humerus.

		Lisbon STMA				New Mexico STMA			
		Min	Max	Mean	SD	Min	Max	Mean	SD
Femur									
F25	Low SES	-0.140	0.200	0.036	0.105	-0.260	0.400	-0.001	0.143
	High SES	-0.300	0.230	-0.004	0.148	-0.160	0.190	-0.023	0.093
F45.5	Low SES	-0.130	0.210	0.007	0.120	-0.340	0.300	-0.019	0.147
	High SES	-0.360	0.240	0.006	0.151	-0.190	0.160	-0.017	0.102
F75	Low SES	-0.120	0.160	0.005	0.101	-0.210	0.280	-0.023	0.136
	High SES	-0.210	0.180	0.011	0.104	-0.130	0.250	-0.020	0.099
F80	Low SES	-0.190	0.120	-0.011	0.094	-0.240	0.390	0.004	0.165
	High SES	-0.190	0.180	0.000	0.097	-0.150	0.280	-0.006	0.118
Tibia									
T25	Low SES	-0.170	0.180	0.033	0.132	-0.220	0.310	-0.008	0.127
	High SES	-0.260	0.160	-0.002	0.125	-0.200	0.120	-0.041	0.090
T50	Low SES	-0.210	0.200	0.000	0.143	-0.540	0.230	-0.064	0.156
	High SES	-0.300	0.260	0.032	0.141	-0.200	0.210	-0.027	0.090
T75	Low SES	-0.220	0.170	0.017	0.131	-0.200	0.340	-0.027	0.131
	High SES	-0.290	0.260	0.014	0.143	-0.160	0.070	-0.047	0.068
Humerus									
H25	Low SES	-0.170	0.200	0.015	0.122	-0.690	0.350	0.007	0.216
	High SES	-0.200	0.250	0.048	0.141	-0.710	0.230	-0.066	0.263
H35	Low SES	-0.200	0.200	-0.004	0.140	-0.560	0.350	0.002	0.188
	High SES	-0.310	0.220	0.024	0.150	-0.200	0.190	-0.015	0.122
H41	Low SES	-0.230	0.160	-0.006	0.132	-0.420	0.330	-0.009	0.176
	High SES	-0.290	0.230	0.028	0.133	-0.240	0.190	-0.020	0.111
H50	Low SES	-0.300	0.170	-0.018	0.142	-0.470	0.350	-0.020	0.178
	High SES	-0.330	0.220	0.028	0.139	-0.240	0.240	-0.023	0.114
H75	Low SES	-0.250	0.180	-0.004	0.130	-0.250	0.390	0.031	0.177
	High SES	-0.370	0.250	-0.017	0.179	-0.230	0.240	0.003	0.125

The STMA residuals did not form a clear pattern when looking at the low SES and high SES groups within each sample (Table 4.9). The only bone that exhibited a general pattern was the humerus, wherein the high SES group had higher STMA means in the Lisbon sample, but the lower SES group had a higher STMA means in the New Mexico sample. There were no significant differences between the low and higher SES groups in either sample (Table 4.12).

Table 4.10 Mean, standard deviation, maximum, and minimum for the standardized residuals of the polar second moment of area (STJ) for each section from the femur, tibia, and humerus.

		Lisbon STJ				New Mexico STJ			
		Min	Max	Mean	SD	Min	Max	Mean	SD
Femur									
F25	Low SES	-0.127	0.123	0.026	0.090	-0.260	0.239	0.045	0.148
	High SES	-0.622	0.239	-0.052	0.187	-0.243	0.545	0.011	0.176
F45.5	Low SES	-0.164	0.209	0.009	0.118	-0.227	0.228	0.005	0.126
	High SES	-0.614	0.532	-0.003	0.224	-0.203	0.305	-0.037	0.142
F75	Low SES	-0.206	0.157	0.040	0.116	-0.319	0.271	-0.014	0.168
	High SES	-0.445	0.233	-0.019	0.143	-0.240	0.251	-0.011	0.142
F80	Low SES	-0.215	0.281	0.046	0.137	-0.257	0.341	-0.004	0.164
	High SES	-0.334	0.259	-0.038	0.151	-0.251	0.260	0.006	0.139
Tibia									
T25	Low SES	-0.803	0.428	-0.038	0.335	-0.296	0.355	-0.004	0.157
	High SES	-0.774	0.305	-0.034	0.236	-0.184	0.128	-0.031	0.091
T50	Low SES	-0.698	0.522	-0.020	0.328	-0.320	0.356	-0.012	0.163
	High SES	-0.783	0.454	-0.004	0.264	-0.195	0.213	0.004	0.129
T75	Low SES	-0.802	0.508	-0.074	0.345	-0.268	0.210	0.018	0.126
	High SES	-0.637	0.342	-0.013	0.234	-0.174	0.216	-0.017	0.102
Humerus									
H25	Low SES	-0.222	0.402	0.084	0.181	-0.663	0.386	0.063	0.249
	High SES	-0.362	0.264	0.020	0.166	-1.218	0.274	-0.180	0.431
H35	Low SES	-0.202	0.366	0.046	0.180	-0.722	0.225	-0.054	0.227
	High SES	-0.354	0.313	0.032	0.158	-1.097	0.220	-0.108	0.275
H41	Low SES	-0.200	0.272	0.021	0.155	-0.688	0.383	-0.041	0.293
	High SES	-0.414	0.448	-0.037	0.201	-1.081	0.281	-0.010	0.292
H50	Low SES	-0.330	0.305	0.018	0.178	-0.708	0.209	-0.082	0.206
	High SES	-0.481	0.277	-0.015	0.202	-0.313	0.287	-0.020	0.172
H75	Low SES	-0.361	0.116	-0.049	0.137	-0.369	0.251	0.030	0.180
	High SES	-0.505	0.436	-0.047	0.222	-0.310	0.213	0.020	0.129

The standardized J residuals demonstrate a pattern wherein the low SES groups have higher mean STJ residuals in both samples (Table 4.10). However, there was only one significant difference between the SES groups. At the 25% humerus section the low SES mean STJ residuals are significantly higher than the higher SES group among the individuals in the New Mexico sample (Table 4.12).

Table 4.11 Mean, standard deviation, maximum, minimum, and ANOVA results for comparing the I_{max}/I_{min} ratios between the high and low SES samples for each section from the femur, tibia, and humerus for each sample.

		Mean	Std. Deviation	Minimum	Maximum	F	Sig.
Lisbon							
F25	Low SES	1.365	0.182	1.104	1.700	2.066	0.159
	High SES	1.608	0.520	1.107	3.696		
F45.5	Low SES	1.229	0.138	1.076	1.515	2.891	0.098
	High SES	1.435	0.371	1.091	2.621		
F75	Low SES	1.457	0.287	1.021	2.009	0.577	0.453
	High SES	1.540	0.299	1.121	2.233		
F80	Low SES	1.568	0.309	1.217	2.128	0.258	0.614
	High SES	1.515	0.273	1.020	2.110		
T25	Low SES	1.335	0.166	1.023	1.587	3.935	0.055
	High SES	1.569	0.356	1.060	2.625		
T50	Low SES	1.689	0.555	1.070	2.630	0.000	0.985
	High SES	1.693	0.480	1.220	3.330		
T75	Low SES	1.616	0.298	1.240	2.080	0.068	0.796
	High SES	1.656	0.446	1.120	2.650		
H25	Low SES	1.474	0.307	1.035	1.899	2.636	0.113
	High SES	1.690	0.377	1.166	2.518		
H35	Low SES	1.506	0.206	1.226	1.808	0.249	0.621
	High SES	1.574	0.407	1.071	2.723		
H41	Low SES	1.576	0.385	1.143	2.292	0.126	0.724
	High SES	1.531	0.329	1.087	2.156		
H50	Low SES	1.571	0.335	1.345	2.298	0.016	0.902
	High SES	1.556	0.333	1.006	2.163		
H75	Low SES	1.540	0.379	1.114	2.406	4.214	**0.048**
	High SES	1.908	0.515	1.103	2.861		
New Mexico							
F25	Low SES	1.590	0.411	1.055	2.345	2.180	0.148
	High SES	1.428	0.263	1.124	2.060		
F45.5	Low SES	1.460	0.295	1.085	2.209	1.742	0.195

	High SES	1.352	0.222	1.021	1.912		
F75	Low SES	1.444	0.174	1.193	1.771	2.400	0.130
	High SES	1.363	0.160	1.115	1.655		
F80	Low SES	1.350	0.227	1.067	1.849	0.259	0.614
	High SES	1.318	0.168	1.011	1.619		
T25	Low SES	1.426	0.210	1.136	1.891	5.863	**0.020**
	High SES	1.606	0.256	1.110	2.227		
T50	Low SES	1.810	0.390	1.230	2.530	0.000	1.000
	High SES	1.810	0.363	1.130	2.490		
T75	Low SES	1.821	0.362	1.290	2.590	0.886	0.353
	High SES	1.949	0.486	1.100	3.020		
H25	Low SES	1.703	0.432	1.067	2.720	0.342	0.562
	High SES	1.624	0.424	1.019	2.417		
H35	Low SES	1.528	0.256	1.185	2.030	0.144	0.707
	High SES	1.580	0.550	1.127	2.960		
H41	Low SES	1.670	0.426	1.009	2.715	0.036	0.850
	High SES	1.644	0.430	1.017	2.509		
H50	Low SES	1.469	0.263	1.039	2.090	1.518	0.225
	High SES	1.617	0.458	1.052	2.602		
H75	Low SES	1.614	0.489	1.078	2.774	0.045	0.834
	High SES	1.647	0.475	1.053	2.928		

Significant results are indicated in bold ($\alpha < 0.05$)

The I_{max}/I_{min} ratio results demonstrate different patterns within the New Mexico and Lisbon samples. Within the Lisbon sample most sections have a higher mean ratio among the high SES group (Table 4.11). The only sections where the low SES group had higher mean ratios were the femur at 80%, and 41% and 50% of the humerus. The only significant difference between the means occurred at 75% of the humerus. The New Mexico sample had a more mixed pattern between the high and low SES (Table 4.11). All the femoral sections had higher ratio means among the low SES group, while the tibia had higher mean ratios for the high SES group at 25% and 75% of the diaphysis (the means at 50% were even). Only the mean ratio at 25% of the tibia were significantly higher for the high SES group. At 25% and 41% of the humerus the low SES had higher

mean ratios, while the means at 35%, 50%, and 75% of the humerus were higher among the high SES group. None of the differences were significant.

Table 4.12 ANOVA results for the standardized residuals of the cortical section area (STCA), the medullary cavity area (STMA), and the polar second moment of area (STJ) for each section from the femur, tibia, and humerus.

	Lisbon STCA		NewMex STCA		Lisbon STMA		NewMex STMA		Lisbon STJ		NewMex STJ	
	F	p	F	p	F	P	F	P	F	p	F	p
Femur												
F25	0.754	0.391	0.06	0.808	0.574	0.454	0.317	0.577	1.417	0.242	0.395	0.534
F45.5	0.854	0.362	0.581	0.451	0.001	0.978	0.003	0.955	0.024	0.879	0.884	0.353
F75	2.284	0.14	0.09	0.766	0.028	0.869	0.005	0.943	1.237	0.274	0.003	0.955
F80	2.597	0.117	0.101	0.753	0.086	0.772	0.048	0.828	2.177	0.15	0.041	0.841
Tibia												
T25	0.582	0.451	1.372	0.249	0.46	0.503	0.799	0.378	0.002	0.965	0.431	0.516
T50	0.431	0.516	0.475	0.495	0.34	0.564	0.789	0.381	0.02	0.888	0.108	0.744
T75	0.005	0.941	0.221	0.641	0.003	0.956	0.366	0.549	0.36	0.553	0.911	0.346
Humerus												
H25	4.351	**0.045**	3.736	0.061	0.412	0.525	0.867	0.358	0.937	0.34	4.385	**0.043**
H35	3.212	0.082	0.541	0.467	0.242	0.626	0.117	0.735	0.05	0.824	0.419	0.522
H41	3.148	0.085	0.253	0.618	0.451	0.507	0.057	0.813	0.611	0.44	0.109	0.743
H50	2.35	0.135	0.025	0.874	0.739	0.396	0.005	0.945	0.192	0.664	1.005	0.323
H75	3.793	0.06	0.021	0.886	0.04	0.842	0.32	0.575	0	0.985	0.037	0.849

Significant results are indicated in bold ($\alpha < 0.05$)

ANCOVA tests were conducted on the combined sample of raw CA, MA, and J values to examine differences in the raw CSG residuals in accordance with SES and sex, with age and sample as cofactors. In testing for homogeneity of the variances prior to the ANOVA, plots of CSG values against age demonstrated a fair amount of variation and heteroscedasticity when comparing the samples, and Levene's tests for inequality of variances found unequal variances for the CA values at 75% of the humerus, for MA at 25% of the tibia, and for J at 45.5% of the femur.

The raw CA values were significantly higher among the high SES groups at 25% of the femur, and 25% and 75% of the tibia (Table 4.13). No significant differences were found between the sexes or at the interaction between each sample and SES. The

ANOVA results (Table 4.12) only demonstrate significantly higher STCA values at 25% of the humerus among the low SES group.

Table 4.13 ANCOVA results for comparing the raw measures of the cortical cross-sectional area (CA) between the Lisbon and New Mexico samples for each section from the femur, tibia, and humerus. SES and sex are the independent variables, the dependent variable is the CA at each section, and age and sample the cofactor in the ANCOVA analysis.

	CA (mm^2)					
	SES		Sex		Sample*SES Interaction	
	F	P	F	P	F	p
Femur						
F25	4.20	**0.044**	0.64	0.425	0.95	0.333
F45.5	3.64	0.061	0.23	0.630	1.85	0.178
F75	1.01	0.319	0.00	0.977	0.21	0.649
F80	1.27	0.265	0.02	0.884	0.22	0.643
Tibia						
T25	4.82	**0.031**	0.98	0.325	2.91	0.092
T50	3.85	0.054	0.18	0.675	1.83	0.180
T75	7.01	**0.010**	0.04	0.842	2.75	0.102
Humerus						
H25	0.50	0.480	1.84	0.179	0.23	0.633
H35	0.65	0.422	1.17	0.283	0.28	0.597
H41	0.54	0.463	0.88	0.351	0.01	0.927
H50	0.50	0.484	0.54	0.463	0.08	0.772
H75	0.30	0.584	1.09	0.300	0.13	0.715

Significant results are indicated in bold ($\alpha < 0.05$)

There were no instances wherein the raw MA values were significantly different between the SES groups, but unlike the STMA values, the majority of the raw MA means were higher in the high SES group. There were several instances of males demonstrating significantly higher MA values than females, at 75% and 80% of the femur, 50% and 75% of the tibia, and 25% and 35% of the humerus (Table 4.14). No significant differences were found at the interaction between each sample and socioeconomic status.

Table 4.14 ANCOVA results for comparing the raw measures of the medullary cavity area (MA) between the Lisbon and New Mexico samples for each section from the femur, tibia, and humerus. SES and sex are the independent variables, the dependent variable is the MA at each section, and age and sample the cofactor in the ANCOVA analysis.

	MA (mm^2)					
	SES		Sex		Sample*SES Interaction	
	F	P	F	P	F	P
Femur						
F25	0.30	0.585	0.06	0.814	0.25	0.619
F45.5	1.09	0.301	0.74	0.393	0.38	0.538
F75	1.26	0.266	9.58	**0.003**	0.27	0.608
F80	1.40	0.241	6.76	**0.011**	0.68	0.411
Tibia						
T25	0.00	0.969	0.03	0.865	0.10	0.753
T50	1.80	0.185	6.12	**0.016**	0.59	0.446
T75	1.72	0.193	4.35	**0.041**	1.90	0.173
Humerus						
H25	0.83	0.367	4.75	**0.033**	2.05	0.157
H35	1.79	0.185	4.88	**0.031**	2.55	0.115
H41	1.84	0.180	3.14	0.081	2.50	0.119
H50	3.38	0.070	2.03	0.159	2.64	0.109
H75	3.44	0.068	4.45	0.038	3.22	0.077

Significant results are indicated in bold ($\alpha < 0.05$)

There were no instances where the raw J values were significantly different between SES groups, though the majority of sections demonstrated higher mean J values among the high SES group. There were two instances of males demonstrating significantly higher J values, 25% and 35% of the humerus (Table 4.15). No significant differences were found at the interaction between each sample and socioeconomic status.

Table 4.15 ANCOVA results for comparing the raw measures of the torsional rigidity (J) between the Lisbon and New Mexico samples for each section from the femur, tibia, and humerus. Sample and sex are the independent variables, the dependent variable is the J at each section, and age is the cofactor in the ANCOVA analysis.

	J (mm^4)					
	SES		Sex		Sample*SES Interaction	
	F	p	F	P	F	P
Femur						
F25	2.89	0.094	0.33	0.569	0.65	0.423
F45.5	3.55	0.064	0.60	0.441	1.68	0.200
F75	1.03	0.314	1.46	0.231	0.17	0.678
F80	0.84	0.361	1.25	0.267	0.06	0.801
Tibia						
T25	0.25	0.619	0.49	0.487	0.24	0.627
T50	0.40	0.527	0.03	0.869	0.05	0.825
T75	1.60	0.210	0.13	0.724	0.94	0.335
Humerus						
H25	0.16	0.688	6.56	**0.013**	0.81	0.371
H35	0.29	0.591	6.29	**0.014**	0.45	0.503
H41	0.33	0.570	1.54	0.219	0.30	0.587
H50	0.01	0.927	0.56	0.458	0.06	0.810
H75	3.21	0.078	2.78	0.100	3.43	0.068

Significant results are indicated in bold ($\alpha < 0.05$)

4.5. Discussion

This study investigated two main research questions; 1) is there a difference in cortical patterning between individuals in the early 20th century Portuguese sample and the early 21st century New Mexico sample, and 2) is there a relationship, within each sample, between socioeconomic status and cortical bone patterning. Answering these questions will aid in understanding how cortical bone deposition during growth and development is impacted by the larger growth environment. In order to examine these questions, four variables of cortical bone distribution from several long bone cross-sections were quantified: cortical bone area (CA), medullary cavity area (MA), torsional rigidity (J), and cross-sectional shape (I_{max}/I_{min} index). Standardized residuals were also created (STCA, STMA, STJ) to remove the size differences between the cross-sections so that the relative cortical distributions could be examined.

The initial expectations of the cortical patterning from each sample were that individuals in the higher stressed but more active sample (the Lisbon sample) would demonstrate, lower CA values, higher MA values, higher J values, and higher I_{max}/I_{min} ratios, while those in the lower stressed, less active sample (the New Mexico sample) would likely produce higher CA values, lower MA values, lower J values, and lower I_{max}/I_{min} ratios. Individuals in the Lisbon sample have more rigid, ovoid shaped diaphyses, but thinner bones with larger medullary cavities. The results do not strictly adhere to this pattern, but a similar trend emerges in most bones.

In the femur, several statistically significant differences were observed between the samples across the various CSG variables examined. A general pattern can be observed. In the raw measurements, wherein the CA, MA, and J sample means are higher among the New Mexico sample, all of the differences were significant (Tables 4.5, 4.6, and 4.7). When compared with the size-standardized residuals an interesting pattern begins to emerge. The standardized CA residual means are all still larger among the New Mexico sample (significantly so at 25% and 80% of the femur), but the standardized MA residual means are higher among the Lisbon sample (significantly so at 25%, and 45.5% of the femur) (Table 4.3). Standardized J means appear to be more mixed, where J at 25% and 80% of the femur were higher among the New Mexico sample, while the 45.5%, and 75% STJ residual means were higher in the Lisbon sample. None were significant (Table 4.3). All the femoral cross-sections were more ovoid among the Lisbon sample, significantly at 75%, and 80% of the diaphysis (Table 4.4). At first it may appear that the raw and size-standardized measurements are portraying different, contradictory patterns of cortical development, but what is being shown is the impact of overall size differences between the sample. The total cross-sectional areas of the New Mexico samples are so much bigger than those from the Lisbon sample, that the raw CA, MA, and J values are larger. But when accounting for size differences among the cross-sections, the medullary cavity areas are proportionately larger in the Lisbon sample than in the New Mexico sample. These results are in line with the hypothesized cortical patterning of the less biosocially stressed modern, New Mexico sample, but the STJ values and shape indices are less clear cut. Instead of the more active population producing unilaterally higher STJ values there is a mix among the femoral sections when size differences are removed. This could be the result of an increasing secular trend in weight gain among children over the

last 100 years. It may be that while children now are less active than they were historically, the larger amount of loading occurring as a result of increased body weights means that the lower limb cross-sectional are still heavily loaded in the New Mexico sample.

The tibiae produced results similar to those observed in the femur, wherein CA and MA values were in line with the initial hypotheses. The raw CA and MA means were all significantly higher among those in the New Mexico sample, and the raw J means were also higher among individuals in the New Mexico sample (though not significantly) (Tables 4.5., 4.6., and 4.7). As with the femur, the size-standardized residual means of the CA values are still higher among the New Mexico sample, and the MA values are higher among the Lisbon sample (Table 4.3). The standardized J means also appear to show a mix, wherein the means are higher among the New Mexico sample at 75% of the tibia, but the Lisbon J means are higher at 25% and 50% of the tibia (Table 4.3). Unlike the femur, 50% and 75% of the tibia are more ovoid among the New Mexico sample, while 25% is more ovoid among the Lisbon sample (Table 4.4). As with the femur, this mixed pattern could be the effect of an increased body mass in loading the lower limb in modern children. If this is true, then it appears that with regards to torsional rigidity and strength, the areas towards the ends of the diaphysis near the hip and the knee in both the femur and the tibia are more responsive to the effect of body mass. Whereas the mid section of both bones and the most distal element of the tibia are more responsive to differences in physical activity.

The humeri demonstrated a similar pattern to that of the initial hypothesis. The raw CA, MA, and J residual means were higher in the New Mexico sample in all the sections of the bone, several of which were significant (Tables 4.5, 4.6, and 4.7). Like the bones of the lower limbs, the size-standardized CA and MA residuals were higher in the Lisbon sample (though not significantly so), but unlike the lower limb all but one section of the humeri produced higher STJ means and more ovoid shapes (Table 4.3). The STJ values were higher in the New Mexico sample at 75% of the humerus, but all other sections had higher STJ values among the Lisbon sample, though again none were significant. Again here, it is possible that the portion of the diaphysis closest to the shoulder joint is impacting J values, but not due to body mass. There were no significant differences among the I_{max}/I_{min} indices, but all except 41% of the humeri were more ovoid among the Lisbon sample (Table 4.4). The distal elements of the humeri all produce

results consistent with the initial hypothesis that the individuals in the Lisbon sample would have lower cross-sectional areas, higher medullary cavity areas, higher torsional rigidity, and more ovoid cross-sections.

The humerus is not subject to the degree of loading via body mass that the bones of the lower limb are. This may help explain why the STJ values and I_{max}/I_{min} indices from the sections of the humeri are less mixed and higher in the more active Lisbon sample, and the less active New Mexico sample had lower J values. Again, it can be observed that the medullary cavity is larger in the more biosocially stressed Lisbon sample, while cortical bone areas are smaller when both are standardized for size. When comparing this pattern with the raw CA, MA, and J values it appears that the impact of biosocial stress on the cross-sections greatly reduces the overall size of the cross-section and the relative size of the torsional rigidity. But when these size differences are removed medullary cavity size is still highly affected by biosocial stress, whereas torsional rigidity and ovoid shape are still higher among the more active population (when looking at the elements of the body that are not subject to the effects of body mass loading). However, the importance of body mass on J values should be noted, as in the lower limb robusticity was not always highest in the more active sample, but often in the heavier sample.

These results support the initial hypothesis, wherein individuals in the more biosocially stressed, more active Lisbon sample produced higher standardized J and MA values, lower standardized CA values, and generally have more ovoid cross-sections when compared with the less stressed, less active individuals New Mexico sample. The effect of body mass loading on the lower limbs is shown in the sections of the diaphyses closest to the knee and the hip demonstrating higher standardized J values in the New Mexico sample, while the more active Lisbon sample exhibited higher J values closer to the mid-shaft. While the upper limb, which is not influenced by loading via body mass demonstrates a pattern in the STJ values more in line with the hypothesis.

The archaeological research focusing on the impact of biosocial stress on cortical bone development in children have largely focused on the femur (Brickley et al. 2009). Van Gerven and colleagues (1985) examined an archaeological sample of children from the Nubian Medieval Christian cemetery of Kulubnarti and found that cortical area percentage of the tibial mid-shaft was affected by what they assumed to be dietary

stress, but external components of bending strength were not heavily affected as measured by cortical area. Rewekant and colleagues (1995) found similar results in their study of tibial and femoral cortical thickness among medieval Polish children, where there was significant resorption found at the endosteal surface of both bones in the stressed populations. The results found in the tibia in this study align with their results, and suggest that in addition to the femur, it is important to examine the tibia and bones of the upper limb when examining the impact of biosocial stress or various environmental stressors in archaeological populations. The upper limb bones especially, may be more suited as it is not loaded by body weight, and several studies of the lower limb have noted that the femur is more sensitive to loading histories than the tibia (Holt 2003; Stock 2006; Osipov et al. 2020).

When attempting to examine how socioeconomic status (SES) might impact cortical patterning the initial hypothesis was that high SES individuals within the Lisbon sample will demonstrate higher CA values, lower MA values, lower J values, and less ovoid cross-sections with low SES Lisbon individuals exhibiting lower CA values, higher MA and J values, and more ovoid cross-sections. The high SES New Mexico children were expected to demonstrate higher CA values, lower MA values, higher J values, and more ovoid cross-sections, while the low SES New Mexico individuals would exhibit lower CA values, higher MA values, lower J values, and less ovoid cross-sections. The reasoning behind these hypotheses was that within the Lisbon sample higher SES children would be more stressed but less active than low SES children, whereas in the New Mexico sample high SES children would be less biosocially stressed and more physically active than the low SES children. The results do not meet these hypotheses, either along the lines of the two samples or within each of the bones. Examining the patterns present within each sample may lead to some more clarity around the suitability of using socioeconomic status to examine cortical patterning.

The standardized CA values of the femora demonstrate an inverse pattern to the hypothesis in the Lisbon and some sections of the New Mexico samples. The STCA mean values were generally higher in the lower SES groups for the Lisbon sample (Table 4.8). There was only one statistically significant difference between the STCA means, within the Lisbon sample at 25% of the humerus (Table 4.12). The raw values, however, demonstrate a pattern more consistent with the initial hypothesis, in that the high SES group demonstrates higher average CA values than the low SES group,

though few were significant (Table 4.13). The standardized MA did not produce any significant differences and did not generally form any patterns between low and high SES groups within each sample (Tables 4.9, and 4.12). The raw MA values demonstrate a clearer pattern, wherein in general the high SES groups from both the Lisbon and New Mexico sample had larger average MA values, though none were significant (Table 4.14).

Among the standardized J residuals there is a slight pattern when comparing the low and high SES status groups within each of the samples, wherein the low SES groups have higher mean STJ residuals (Table 4.10). There was only one significant difference between the SES groups. At the 25% humerus section the low SES mean STJ residuals are significantly higher than the higher SES group (Table 4.12). The raw mean J values in the New Mexico sample were mostly higher in the high SES groups in the femur and the tibia (Table 4.15). There was only one instance of a significant difference in I_{max}/I_{min} indices, and it was among the Lisbon sample at 75% of the humerus (Table 4.11). Otherwise, the indices did not demonstrate a clear pattern, though mean values were more often higher among the low SES group for the Lisbon sample. There were no significant differences between low and high SES groups for I_{max}/I_{min} indices among the New Mexico sample, though most means were higher for the high SES group.

The raw values demonstrate that the high SES groups had cortices that were generally larger than those of the low SES groups in both samples. The mean STCA residuals were highest in the low SES group for both samples. Physical activity appears to be influencing the development of cortical cross-sectional area, wherein generally active, lower SES children in the Lisbon sample may be producing comparable STCA levels to their higher SES counterparts because of their high levels of physical activity. Whereas the children in New Mexico are generally more sedentary, but better nourished than the Lisbon children, which may help explain why the higher status children have higher STCA values in some cases. Many of these children may have access to extracurricular activities that would make them more active than their lower SES counterparts or potentially using different muscle groups. The higher measurements of STJ in the Lisbon sample could be the result of different physical activity levels between low and high SES groups. As many of the low SES children would have been working either inside or outside the home and could have been much more active than their

higher SES counterparts. The inverse appears to be happening in the New Mexico sample where higher SES individuals may have a more active lifestyle afforded to them by affluence than those in the lower SES groups who may not have the same opportunities. Either way, both groups of children from the modern sample would likely be enrolled in day long school programs, wherein they would experience a similarly sedentary lifestyle.

These results indicate that the relationship between SES and cortical bone patterning is complex, and dependent on the social and temporal context in which the individuals in the samples lived. The lower SES children living in 20th century Lisbon were subject to much different stressors than the lower SES children living in 21st century New Mexico. In general, it appears that the non-standardized CSG measures of the high SES children were larger than those of the low SES children for both samples. When examining the size-standardized CSG values, the low SES Lisbon children were much more active, as they likely needed to work to help support their families but were relatively more malnourished and had a higher disease load than the low SES New Mexico children, who would have lived more sedentary lives (van Mil et al. 1999; Tylavsky et al. 2019). The results of the SES analysis must also be brought into question by the general lack of homogeneity of variance and heteroscedasticity of variance among the two groups within each sample. This is likely the result of lower sample size. Any attempts to conduct similar research in the future should ensure access to a larger group of individuals with documented histories.

When compared with other studies that have examined the impact of socioeconomic status on cortical bone development, the results from this research are somewhat comparable. Mays and colleagues (2009) found there to be significantly thinner standardized cortical thickness in the lower SES group than in the higher group but found that standardized bone width did not differ significantly between the groups. Their results were standardized for age, but not for body mass, so no direct comparison Can be made with the STCA values from this study. Additionally, the study by Mays and colleagues was only able to include 5 individuals in the high SES group and only looked at one section of the femur. Newman and Gowland (2017) were able to examine around 400 individuals from low, medium, and high SES groups from 19th century London but found there to be no significant differences between the groups in an analysis of covariance that included age as a cofactor. Their observations were also limited to the

femoral mid-shaft and did not standardize the cortical thickness or bone width values for size. What Newman and Gowland (2017) did find, was that all their sample groups fell significantly below modern appositional growth standards, a pattern that fits with the observed differences between the raw CA values from the historic Lisbon sample and the modern New Mexico sample.

The results produced in the current study demonstrate a pattern that can be compared to both previous works. Newman and Gowland (2017) did not find significant differences in cortices, and in this study the majority of cross-sections had no significant difference in standardized CSG values between the SES groups. But Mays and colleagues (2009) did find a significant difference in cortical thickness, and here, raw CA values at 25% of the femur, and 25% and 75% of the tibia did demonstrate significantly higher values in the high SES group. Therefore, it can be concluded that it is important to examine different sections and bones when attempting to look for changes related to socioeconomic status and cortical development. The previous studies do not examine any bones from the upper limb, which have proven in this research to be integral for teasing out the differences biosocial stress and mechanical loading can have on bone development.

Additionally, the previous studies are limited by the nature of many historical and archaeological samples in the demographic information available on the children in their samples. Individuals in these studies are sorted into socioeconomic groups based solely on their burial location. The benefits of working with the documented skeletal collections used in this research is that each individual received a SES grouping based on factors that would have affected them while they were alive, rather than ones that are related only to their death.

When attempting to draw any broader implications on cortical growth and development from this study it is important to understand the limitations of certain aspects of the samples. As the samples are cross-sectional, as opposed to longitudinal, in nature, these are not indicative of individualized developmental and ontological growth trajectories. There are, however, potential benefits from this type of sample, in that there is a variation represented in the sample that would be missing from examining the same few individuals over the course of their growth. The Lisbon and New Mexico samples are

comprised of a more diverse subset of the population, both biologically and socially, when compared with longitudinal samples that have fewer individuals.

A statistical limitation of the research can be found in the sample size, specifically with regards to the grouping for low and high SES individuals. The number of individuals that could be grouped into either low or high SES was relatively small, especially when only examining individuals within each sample. The limitations were largely practical, for the Lisbon sample it was based on available specimen and for the New Mexico samples it was based on zip codes. Additionally, several individuals from the Lisbon and New Mexico samples had to be excluded from this categorization, as demographic data (either freguesias or census data) were not available. The interpretation of the SES results to populations or samples outside this context should keep this in mind. Measurement error in calculating the CSG variables was relatively low (see chapter 1), but it is possible that some differences between individuals at the same bone section could result from user error on the part of the researcher. Specifically, the alignment and measurement of the long bones in Dragonfly is a manual process, and thus is subject to human error.

Variation in cortical patterning can also be attributed to more individualistic conditions, rather than sample trends in biosocial and mechanical stress. Cause of death or experiences with chronic disease or disability are also factors that can influence the development of the cortices. For example, many children in the Lisbon sample died of tuberculosis, which can impact bone formation as the disease progresses. Though there were much fewer instances of chronic or acute disease as the cause of death in the New Mexico samples, there were a few individuals with known chronic illnesses, such as sickle cell anemia. Such factors can have hidden impacts on bone development, but they also demonstrate a larger pattern within the samples, wherein most causes of death were related to illness in the Lisbon sample, while most of the causes of death in the New Mexico sample were accidental. With this trend in mind, individuals with causes of death related to illness that could impact bone development were not excluded from the samples, as this would act as an accurate representation of morbidity and mortality risks in specific historical contexts. Additionally, if all individuals in the Lisbon collection were excluded based on cause of death related to disease the sample size would have been drastically reduced.

4.5.4. Conclusion

The findings presented in this paper suggest a complex relationship is present in the development of the cortices of long bones, one that is specific to environmental context and the bone being examined. The first hypothesis; that more stressed, but active individuals of the Lisbon sample had smaller cortical areas and larger medullary cavities, as well as higher torsional rigidity and more ovoid cross-sections, in the long bone cross-sections than the relatively less stressed, and less active individuals in the New Mexico sample, is generally supported in sections of the humerus. More complex dynamics were present in the lower limb. Here the cortical area and medullary cavity area were larger in the New Mexico sample, but there were instances where J values were equal or higher within individuals in the Lisbon sample. It is likely then that the biomechanical constraints surrounding loading the lower limb mean that a certain level of robusticity must be maintained, particularly in the sections of the diaphysis near the hip and the knee. The positive secular trend of body mass among children living in urbanized countries should be considered with regards to this phenomenon. The results suggest it is important to look at multiple bones and multiple sections of the bones, as it cannot be assumed that robusticity will be uniform along the entire diaphysis.

In terms of socioeconomic status and its effect on cortical bone development, the results appear to be less clear. The unstandardized CSG values appear to support the initial hypothesis of lower SES individuals having smaller cortical areas and larger medullary cavities, but there does not appear to be a pattern in the standardized residuals. These results are in line with what has been found in previous studies of cortex development and socioeconomic status. More research is needed with documented collections to understand the impact of socioeconomic status on appositional cortical growth. Socioeconomic status encompasses stress from multiple sources, including malnutrition, disease, access to medical care, and opportunities for physical activity. Pulling apart the intricacies of how these factors impact cortical development should be the focus of future research. The mechanical loading of the limb will likely continue to impact bone shape, but if nutritional requirements are not met, increased medullary cavity size and a thinner cortex will ultimately negatively impact the amount of cortical bone present. Bone rigidity and strength will then likely be impacted, but habitual loading, either through high body mass or physical activity level, can offset some of the loss of torsional rigidity. What is clear is that using modern cortical growth

standards to examine archaeological populations is problematic, and the differences in nutrition, physical activity, disease, and body mass should be taken into account when comparing modern child cortical development with those from children of the past. Children from past populations are unlikely to be both undernourished and sedentary (at least in comparison to modern children), meaning that while medullary cavity size may be large, it is possible that their bones may exhibit relatively high bone strength and robust shapes, especially when compared with those of modern children.

Chapter 5. Discussion and Conclusion

This section offers a brief discussion of the findings of the preceding research as well as some concluding thoughts. The results of the papers included in the dissertation will be summarized, followed by a discussion of the trends present across the papers, and how the results fit into existing literature on the topic. The conclusions that can be drawn from this research, in connection with preceding literature, regarding the impact of body mass, physical activity, and biosocial stress on cortical development in long bones are then given. This discussion will also attempt to discuss the ways in which historical and environmental context should be considered when examining the impact of the previous factors on cortical growth and development.

All three studies rely on cross-sections isolated from CT scans, from which cross-sectional geometry (CSG) measurements could be taken. The first and third paper utilise both the historical sample comprised on individuals from the Luís Lopes Documented Skeletal Collection from Lisbon's Museum of Natural History and Science (Lisbon sample) and the modern sample comprised of individuals from the New Mexico Descendent Image Database CT collection (New Mexico sample). The second paper relies solely on individuals from the New Mexico sample, as they had known weights at death, whereas weights were not recorded for the Lisbon sample and would have been estimated. Each sample consisted of individuals ranging in age from just after birth to 19 years of age. Where both the Lisbon and New Mexico samples were used, individuals were matched for age and sex.

The first paper examined intraobserver error between three rounds of cortical area (CA) measurement from dry and wet bone samples using both manual and algorithmic segmentation methods. The paper's goals were: 1) understand how comparable CSG values taken from individuals in the Lisbon and New Mexico samples were, thereby acting as a foundational basis for the results of the next two papers; and 2) determine the comparability of the algorithmic segmentation methods with the manual segmentation of bone from non-bone in the CT images. The results suggest that there are more instances of significant differences in intraobserver error among the wet bone samples than the dry bone samples, but more instances of significant differences between each segmenting method among each observer round in the dry bone sample.

However, all coefficients of reliability were over 0.95, indicating that the error rates fall within the acceptable range (Ulijaszek and Lourie 1994). Based on these results, the error rates are comparable enough to compare the two samples in the third paper. The results also demonstrate uniformity across reliability and technical measurement of error for both automated methods and the manual methods of segmentation. So, while the manual method should be taken as the gold standard, as it allows for the researcher to make a visual assessment of the limits of the cortical bone within a cross-section, the results of the study suggest that the use of algorithms are comparable to manual segmentation.

The second paper developed weight estimation formulae based on a sample of 21st century children from the New Mexico sample. The formulae used several common epiphyseal and metaphyseal breadth and torsional rigidity (J) measurements to estimate weight, as well as some previously unexamined sections of the femur and tibia. The goal of the second paper was to create a new set of weight estimation formulae for children suitable for use in a contemporary forensic context. Weight estimation formulae were created for a sample of modern children ages 1 to 19 years old of identified age, height, weight, and sex from the Albuquerque, New Mexico area. Formulae developed using J values generally had the lowest prediction error and demonstrated an increased utility for prediction of weight in obese individuals (BMI above the 95th percentile) than the bone breadth formulae. The bone breadth results were generally less accurate and suffered from a bias likely introduced from de-transforming these data but were still most suited to the youngest ago group (ages birth to 5 years old). All formulae generally underestimated weight in younger and lighter individuals, and overestimated weight in older and heavier individuals. Measurements from the femur still produced the formulae with the least amount of error, but formulae from the tibia produced comparable error rates, suggesting that in instances where the femur is not available the tibia is an equally suitable substitute for estimating weight.

The third paper examined the relationship between biosocial stress, mechanical loading, and cortical bone distribution in the femur, tibia, and humerus of the Lisbon and New Mexico samples. The goals of the third paper were: 1) determine if there is a difference in cortical bone distribution between individuals in the Lisbon sample and the New Mexico sample; 2) investigate the relationship within each sample between socioeconomic status (SES) and cortical bone distribution. It was expected that

individuals in the Lisbon sample, coming from a relatively more biosocially stressed, and more physically active population, would exhibit lower total cortical area values, higher medullary cavity area, higher torsional rigidity values, and more ovoid-shaped cross-sections. Conversely, individuals in the less active, less biosocially stressed New Mexico sample would demonstrate higher total cortical area values, lower medullary cavity area values, lower torsional rigidity values, and less ovoid cross-sections. In terms of SES within the samples, it was predicted that within the Lisbon samples high SES individuals would be less stressed and less active than low SES individuals, which would result in less robust, higher cortical area cross-sections with smaller medullary cavities. For the New Mexico sample it was predicted that high SES individuals would be less stressed and more active than low SES individuals, thus giving the higher SES individuals more robust, higher cortical area cross-sections, with smaller medullary cavities. The results indicate that in the upper limb, individuals in the Lisbon sample generally had proportionally larger medullary cavities and thinner cortices, but proportionally more rigid and more ovoid bones than the New Mexico sample (though the overall size of the cross-sections were much larger in the New Mexico sample). In the lower limb, the heavier weights for age observed in the New Mexico sample meant that some sections of the femur and tibia were more rigid and ovoid than in the Lisbon sample. This is likely due to the biomechanical size constraints surrounding loading the lower limb. In terms of socioeconomic status and its effect on cortical bone development, the results appear to be less clear. In general, high SES individuals in both samples have more total cortical area than low SES individuals, but there does not appear to be a pattern when examining the proportional results.

When viewed in tandem the results demonstrate the importance of examining the individuals in each sample within their specific environmental conditions. The social and historical context in which each child lived directly impacts the appositional development and remodeling of their cortex. Factors such as body mass, physical activity, and biosocial stress will directly impact the CSG measurements that are taken from their remains. The next three sections will explore the role of body mass, physical activity, as well as malnutrition and disease prevalence within this study's results and preceding literature.

5.1. The impacts of body mass on cortical development during ontogeny

As mentioned previously, one of the aspects that makes up an individual's loading history is the effect of body mass. The samples examined in this research are from two populations that would have demonstrated very different trajectories in weight-for-age among children. The Lisbon sample is from a period during the early 20th century, when BMIs in Portugal were much lower than those of children living in 21st century America. Children in the New Mexico sample are from a 21st century American population, where the increase in overall weight-for-age, as well as the prevalence of childhood obesity has been well documented (Sun et al. 2012; Tylavsky et al. 2019). Additionally, as per the known BMIs for the New Mexico sample, most individuals in the sample fall in the overweight (85th to 95th percentile) and obese (above 95th percentile) categories. Z-scores based on World Health Organizational averages (WHO 2006) were 1.83 for weight and 2.00 for BMI, meaning that on average individuals within the New Mexico sample were 1.83 and 2 standard deviations above the global means, respectively. When z-scores for J at the femoral mid-shaft were produced using Ruff's (2021) calculations from the Denver Growth Cohort, the mean values from the New Mexico sample were all at least above one (except for the below 5th percentile BMI subsample). Both the z-scores indicating body size and the z-scores indicating cortical appositional growth demonstrate that children in the New Mexico sample are heavier and have more torsionally rigid femoral mid shafts than an American sample from the mid 20th century. This likely reflects the increase in body mass present among the New Mexico children. While there are no known weights for the children of the Lisbon sample, when examining long bone growth patterns Cardoso (2005) noted that children among the Luís Lopes collection were small for their age, and likely would have suffered from malnutrition in the form of undernutrition, making it unlikely they would have been overweight, let alone obese. In all likelihood, children in the New Mexico sample were heavier than those in the Lisbon sample.

How these differences in body mass have influenced CSG measurements is difficult to pull apart from the effects of physical activity and biosocial stress. What can be observed in the measurements is that cross-sections from the New Mexico sample were larger-for-age than in the Lisbon sample, meaning they had larger cortices and were by in large more rigid (at least in the lower limb). When standardized for size

differences the New Mexico sample still had thicker cortices, but the Lisbon sample had larger medullary cavities. The affect of body mass was most evident when comparing the relative torsional strength and shape of the upper and lower limbs in each sample. Unsurprisingly, when standardized for size differences among the humeri, the more physically active Lisbon sample had higher torsional rigidity and more ovoid-shaped cross-sections. However, the values for diaphyseal sections of the femur and the tibia were mixed. Some diaphyseal sections had higher torsional rigidity values and more ovoid cross-sections among the Lisbon sample, and some sections had higher values among the New Mexico sample.

Previous research has demonstrated that, biomechanically, lower limb CSG are impacted by loading history due to both physical activity and body mass, while the upper limbs loading history is more heavily impacted by physical activity (Ruff 2000; Pomeroy 2018). Standardized rigidity values are larger in the lower limb sections for the Lisbon sample at 45.5%, and 75% of the femur, and 25% and 50% of the tibia, the remaining sections are larger among the New Mexico sample. It appears that the sections near the hip and the knee are more rigid for the heavier New Mexico sample, while the sections close to the midsections of the diaphysis are more rigid in the Lisbon sample. It is possible to interpret this as diaphyseal bone strength being less responsive to loading via body mass, while areas nearer to the joints are more influenced by increased body mass. Comparison between rigidity values at different portions of the diaphysis could therefore be useful in examining individuals with unknown weight and unknown patterns of habitual physical activity. With regard to shape, most of the lower limb sections are more ovoid among the Lisbon sample, significantly so at 75% and 80% of the femur. The sections that are more ovoid among the New Mexico sample are at 50% and, significantly, at 75% the tibia. It may therefore be more difficult to differentiate the influences of loading via physical activity versus body mass on lower limb diaphyseal shape.

Ruff and colleagues (1991) found no relationship between articulator breadth and body mass in the femur but did find a strong relationship between mid shaft cortical area and second moment of area (Iy) and body mass. The sample used in their study, the Denver Growth Cohort, did not contain any obese children. Interestingly, they also found a slight relationship between neck shaft diameter and body mass. They attribute this to the fact that articular dimensions have a higher proportion of trabecular bone, and a

lower proportion of cortical bone, meaning they would be less prone to remodeling in response to bone loading. This contrasts with the diaphyseal mid-shaft, which has a higher cortical bone to trabecular bone ratio. The femoral neck is at a midpoint in terms of the amount of trabecular bone versus cortical bone, and thus is slightly influenced by loading history. Therefore, the larger STJ values at 80%, and 25% of the femur and 75% of the tibia among the heavier New Mexico sample may be a by-product of their larger cross-sections, as these sections have higher trabecular bone levels than those sections closer to the midsection, and thus would be less impacted by physical activity level. Saers and colleagues (2016) observed trabecular bone architecture in three adult populations with different mobility patterns and found that while bone volume and anisotropy were plastic in response to physical activity, some parameters experienced morphometric constraints based on joint function. Ducher and colleagues (2009) examined trabecular bone density and cortical area among obese women and found that while both appeared to be responsive to obesity level, trabecular volume was more predictive of fat mass, while cortical area was a better predictor of lean mass. Overweight and obese children tend to have higher ratios of fat mass to muscle mass (Sukumer et al. 2011). This patterning in addition to the results of the current research could indicate that the higher levels of trabecular bone in the proximal and distal sections of the diaphysis are more responsive to mechanical factors regulated by body mass, while midshaft cortical sections are more plastic in response to physical activity.

The results from the second paper also demonstrate that breadth measurements are more strongly correlated with weight in younger individuals, so it is likely that these parameters may be more impacted by loading history via body mass early on in life, while the metaphyses and epiphyses are still forming. But once the longitudinal growth of the bone slows, the trabecular bone becomes less plastic, and changes in response to loading are more influential on cortical bone. This could also demonstrate that J is a better parameter for estimating body mass in adults that limb breadth measurements. The New Mexico children generally possess larger cross-sections and are heavier, thus it is possible that their bone diameters towards the hip and the knee were impacted earlier in development, and thus are larger. The sample composition also demonstrates that a large proportion of children are obese early in life, during the formation periods of these skeletal elements (Degnin et al. 2010; Berendsen and Olsen 2015). While sections

with more cortical bone towards the mid-shaft respond more to changes in loading history based on physical activity as individuals age.

Cowgill (2018) examines the two main studies that have produced body mass estimation formulae by Ruff (2007) and Robbins and colleagues (2010) and found that when used to estimate mass in several different spatially and temporally distinct samples of children both sets for formulae tend to produce higher estimates when using the CSG value than the breadth measurements. Cowgill explains these results via the original reference sample (the Denver Growth Cohort) being from an urbanized, affluent population. When examining the utility of these formulae on children from the NMDID, Spake and colleagues (2021) found that both the breadth and J formulae underestimated weight, but that the J formulae were slightly less likely to do so. These results could demonstrate a similar trend to the result of the current study, wherein lighter, but more active children from the mid 20th century Denver Cohort have smaller bone sections and articular breadth, but larger mid-shaft J values, whereas children from the New Mexico sample, having been heavier when their epiphyses and metaphyses were forming demonstrate larger breadth measurements, while J at the midshaft is still large due to loading from body mass, but not being impacted by increased levels of physical activity. In essence, the Lisbon children were small from the beginning, their midshaft diaphyseal J values being influenced by their high level of physical activity and their distal and proximal J values being relatively smaller due to lower loading via body mass at a young age. Conversely, the New Mexico children were large from the beginning, with their midshaft diaphyseal J values being small in comparison due to lack of physical activity, but their distal and proximal J values reflecting loading via a larger body mass from a young age.

The results of the current studies, along with the results presented by Spake and colleagues (2021) and Yim and colleagues (2020), reinforce the idea that body mass is better reflected in unstandardized J values than in breadth values from the long bones, particularly in older children. Both Spake and colleagues (2021) and Yim and colleagues found that weight estimation formulae derived from torsional rigidity parameters underestimated body mass less in modern populations that formulae based on breadth parameters. In younger children breadth measurements may be more appropriate, as these elements are still forming in response to loading via body mass. Body mass estimates still appear to be best predicted by the femoral midshaft in juveniles and

adolescent, but the proximal tibia appears to provide similar levels of prediction error, and this should be considered if the femoral midshaft is unavailable The results also demonstrate how J values are not consistent across populations of children, and how the J values produced are influenced by multiple factors, only one of which is body mass. Therefore, it is important to diversify the reference samples used to produce these formulae if the relationships between cortical parameters and body weight in past and contemporary children is to be properly examined. In future research if samples size was increased, it would be useful to examine cross-sectional values pre- and post- puberty and in sex-based samples. This would help investigate how changes in body mass during puberty affect CSG values, as well as how the different growth trajectories of males and female post puberty affect these values.

Additionally, the results from this research suggest that cortical torsional strength at the midshafts of the lower limb bones may be more responsive to physical activity, whereas the sections of the diaphysis closer to the joints may be more reflective of body mass (or at least overall body size). Much of the research that has been conducted on the cortical bone geometry of children focuses on the mid-shaft of the femur, tibia, and/or humerus (Osipov et al. 2016; Harrington and Osipov 2018; Osipov et al. 2020). In the future consideration should be paid to other sections of the diaphysis. Future research should also attempt to investigate the impact of lean and fat mass on deposition of cortical bone via loading in obese children as a means of investigating the relationship between loading via physical activity and body mass. Examining the differences in cortical distribution in the diaphyses of the lower and upper limb bones may aid in pulling apart the influences of body mass and physical activity on cortical deposition.

5.2. The impact of physical activity on cortical development during ontogeny

Having established that body mass is one factor affecting loading history, it is necessary to examine the impacts of physical activity on CSG. As mentioned in the previous section all the unstandardized CSG values were higher among individual in the New Mexico sample, whereas among the size-standardized values, cortical area was higher in the New Mexico sample, but standardized torsional rigidity values were higher in the Lisbon sample at 45.5%, and 75% of the femur, 25% and 50% of the tibia, and 25%, 35%, 41% and 50% of the humerus. The Lisbon sample also generally had more

ovoid shaped cross-section, except for the sections at 50% and 75% of the tibia, and 41% of the humerus.

The individuals in the Lisbon samples are relatively more robust because of the differences in physical activity that are present between the samples. The individuals from the Lisbon sample would have been much more active than those in the New Mexico sample, due to both the increase in sedentary lifestyles that has occurred over the course of the 20th century in industrialized populations (Carlson et al. 2015), and the tendency for children in early 20th century Portugal to be engaged in both paid and unpaid labour (Valente 1986; Campinho 1995). For children living in early 20th century Lisbon, walking would have been their main source of transportation, and the terrain in and around central Lisbon is very steep (a factor which has been suggested to influence torsional robusticity (Harrington 2010)). Additionally, many of the children in the Lisbon sample were likely working as paid domestic servants and apprentices, activities that would have increased their daily physical activity level. Even those who were not engaged in paid labour would have been expected to do a large amount of active work in the home (Cardoso 2005). Children in the United States have been especially affected by the trend towards sedentary lifestyles and decreased physical activity (van Mil et al. 1999; Tylavsky et al. 2019). Children in the New Mexico sample, by comparison, would have been less likely to be in the workforce and would have been enrolled in a compulsory public school system ending in grade 12, something that did not exist in Portugal until 1986 (Margarida Marques et al. 2007), further increasing their tendency towards sedentary lifestyles. The labour that the children in the Lisbon sample were undertaking, in conjunction with poor nutrition and high disease load, likely explains the relatively small size of the sample cross-sections when compared with those taken from the New Mexico sample, as they are known to be stunted for their age (Cardoso 2005).

The standardized torsional rigidity values appear to reflect this difference in physical activity between the children in each sample. The rigidity values at the 45.5% and 75% sections of the femur and the 25% and 75% of the tibia were proportionally higher among the Lisbon sample. In order to examine the effect of physical activity with less consideration for body mass, the upper limb must be examined. In the humeri rigidity values were proportionally larger among the individuals from the Lisbon sample in all areas except 75% of the humeral diaphysis. This indicates that children in the Lisbon sample had higher torsional robusticity values in their humeri than the children in

the New Mexico sample. While body size is still an important mechanical factor that affects the robusticity of the cortex in the upper limb (see the general raw size difference in the J values between the samples in Table 4.7), the upper limb is much more influenced by physical activity in the form of muscular loading than the lower limb (Ruff 2003; Warner et al. 2006). These results are reinforced by the cross-sectional shapes being more ovoid in all sections of the humerus, with the exception of the 41% midshaft section. Thus, both the robusticity data and the demographic data point to the children in the Lisbon sample being more active than those in the New Mexico sample.

The Influence of physical activity and mechanical loading on bone development in children has been examined in terms of bone formation. There are highly correlated relationships between bone mineralization in the long bones, in the form of bone mineral density and concentration, and habitual physical activity among children (Gunnes and Lehmann 1996; Vincent-Rodriguez 2007). In their examination of how mechanical and metabolic properties impact cortical bone development Eleazer and Jankaukas (2016) found that mechanical bone properties tend to maintain a certain degree of bone strength, even in the face of bone loss due to biosocial factors like malnutrition and disease. There appears to be a similar pattern among the children of the Lisbon sample, wherein torsional robusticity and ovoid shape are preserved despite the overall smaller size of the cortex. Future research should examine the break down of age-dependent locomotion and mechanical loading among different age groups, particularly among the Lisbon sample where children began to become more involved in paid and unpaid labour around puberty.

While there has been relatively little biomechanical analysis of cortical bone deposition in children, the previous research that has been done on mechanical loading of the long bones via activity among prehistoric children has produced similar result to those in this study (Ruff 2003a; Cowgill and Hager 2007; Cowgill 2010; Cowgill 2014; Harrington and Osipov 2018). When examining Cis-Baikal hunter gather groups from the Early and Late Neolithic, Osipov and colleagues found a relative decrease in cortical area and increasingly round cross-sections among children from the less mobile Late Neolithic group (2020). The pattern of the more physically active sample exhibiting higher torsional rigidity values in the humerus than the less active sample is also replicated. Interestingly, Osipov and colleagues (2020) found a decrease in midshaft torsional strength among the lower limb of the more stressed, but more mobile Early

Neolithic group, an opposite finding to what is present in this research. This difference is attributed to the greater body mass estimates for the Late Neolithic group being larger than those of the Early Neolithic group. Within the current research torsional rigidity was less impacted by activity at the proximal and distal diaphysis, and more by activity in the midshaft in the lower limb. This finding highlights the importance of examining multiple sections of the diaphysis when attempting to pull apart the affects of loading via body mass and physical activity on cortical bone distribution. When compared with the previous biomechanical research conducted on ontogenetic cortical development, the current research is unique, both in terms of the individual level data available for the sample, but also in terms of how cross-sectional parameters were standardized for body size. Body mass estimates were not used, but rather a breadth measurement often highly correlated with body size (femoral head breadth) was used to create the standardized residuals. This eliminates any potential error associated with estimating body mass from being introduced when the residuals were created. It is recommended for future use among researchers looking to standardize cortical dimensions when working with individuals without a known body mass. The mechanical loading factors of body mass and physical activity have a much more localized affect on cortical bone distribution (even at different sections of the same diaphysis), whereas the degree of biosocial stress has a much more systematic influence on cortical bone deposition throughout the body. The influence of biosocial stress on the cortical development of individuals in both samples is the subject of the next section.

5.3. The impact of biosocial stress on cortical development during ontogeny

The presence of biosocial stress in the growth environment causes an energetic trade-off between cortical bone development and other areas of growth which results in a negative impact on cortical bone deposition (Mays et al. 2009; Newman and Gowland 2017). The degree of biosocial stress present likely had a large impact on the differences in cortical development observed between the two samples. Differentiating the effects of malnutrition and disease on these samples is difficult, as it is likely that they do not act independently on bone development, but work together and in tangent with other stress factors that can disrupt cortical bone deposition. This is why biosocial stress has been used as the larger grouping for these types of stressors.

As noted previously, the general size of the unstandardized cross-sections is larger among the children in the New Mexico sample, meaning that children in the New Mexico sample had larger cortical cross-sections, more cortical bone, and larger medullary cavities than the children in the Lisbon sample (significantly so in terms of cortical area for all but one of the cross-sections). Previous research has established that internal endosteal surfaces are more likely to be impacted by malnutrition and disease, while l measures of robusticity and shape are more likely to be influenced by mechanical loading history (Garn et al., 1969; Rewekant et al. 1995; Eleazer and Jankaukas 2016). Therefore, with the understanding that there is a general size difference in cortex between samples, it becomes necessary to examine the size-standardized values for cortical area and medullary cavity area when attempting to understand how biosocial stress impacted these samples.

When examining the size-standardized cortical areas, the individuals in the New Mexico sample still have more cortical bone than those from among the Lisbon sample (the majority of which were significantly larger). Conversely, the size-standardized medullary cavity areas were higher among individuals from the Lisbon sample, with only one section of the humerus being higher among the New Mexico sample. Few differences in standardized medullary cavity size were significant, but this opposing pattern demonstrated that the ratio of cortical area to medullary cavity area in the size-standardized sections is higher in the New Mexico sample and lower in the Lisbon sample. Children in the Lisbon sample appear to have larger medullary cavities relative to the total amount of cortical bone, while children in the New Mexico sample have smaller medullary cavity sizes relative to the total amount of cortical bone present in the cross-section. Garn and colleagues (1964; 1969) found similarly significant reductions in midshaft cortical bone area when examining Guatemalan children who were experiencing protein-caloric malnutrition. They also found that medullary shaft areas had increased when compared with national averages (Garn 1965; Garn 1969). Subsequent research on nutrient deficiencies has reaffirmed reduction in total cortical area and increased medullary size in response to undernutrition across different populations over different time periods (Hummert 1983; Van Gerven et al. 1985; Brickley and Ives 2008; Mays et al. 2000), a finding that aligns with this research. Van Gerven and colleagues (1985) also determined that robusticity and shape parameters were more likely to be

preserved, which aligns with the higher relative torsional rigidity and ovoid shape observed in the stressed but active Lisbon sample.

When examining the types of biosocial stress that would have been present in each sample there are two main factors that would likely have the most impact: nutrition and disease. Nutritional status among the children in the Lisbon sample would have been relatively poor when compared with that of the children in the New Mexico sample. Along with their large number of children in the sample involved in wage labour, malnutrition in the form of undernutrition and poor living conditions were prevalent in Lisbon during their lives, enough so to cause stunting among the children in the Luís Lopes collection (Cardoso 2005; Cardoso and Garcia 2009). The likelihood that individuals in the Lisbon sample would have experienced nutritional deficiencies is high. In contrast, individuals in the New Mexico sample demonstrated a tendency to fall close to or above the 95th percentile for BMI (though there were several individuals who fell below the 5th percentile for BMI). Additionally, previous research on the children of the NMDID collection found that they were not stunted (Spake 2020). This, in addition to the high BMI values found amongst the New Mexico sample, indicates that the individuals in the sample were not likely to suffer from protein-caloric deficiencies and likely experienced caloric excess. Whether they were malnourished in the form of access to adequate micronutrients (i.e., over-nourished) in their diets is unknown. The effects of over-nutritional on cortical and medullary dimensions in children is also unknown. When comparing the weights and BMI z-scores for the New Mexico sample, the sample means were 1.83 and 2.00, respectively. Individuals that fell above the 95th percentile (almost half the sample) had a mean weight z-score of 3.51 and BMI z-score of 3.92. These scores put them at a significantly higher value than the World Health Organization averages, as z-scores were calculated based on WHO weight-for-age and BMI-for-age averages (WHO 2006).

Disease load and infection are another aspect of biosocial stress that can influence cortical resorption and medullary cavity size. Chronic disease and long-term exposure to pathogens have been demonstrated to impact bone growth in the form of stunting and lower bone mineral density (Gowland 2015; Maratova et al. 2017). Infectious disease rates were much higher in Lisbon during the period when the children in the sample were alive (Cardoso 2005). This was likely the result of overpopulated and unsanitary living conditions in urban environments as more and more people began to

migrate from rural communities (Cardoso and Garcia 2009). The demographic information provided for each individual on cause of death can give some context to the disease environment present in each population. Among the children in the Lisbon sample acute and chronic tuberculosis were the most common manners of death, with many other children dying from other forms of infectious disease, such as upper respiratory disease and pneumonia. Accidental death was less common. Conversely, in the New Mexico sample manner of death was mostly associated with automotive accidents (and occasionally homicides). Death due to disease was rare, and in the cases where it did happen the diseases were not infectious. An attempt was made to have relatively even distribution of cause of death between the samples, but there simply were not enough disease related deaths in the New Mexico sample or accidental deaths in the Lisbon sample for this to occur without drastically reducing the sample size. Modern medical care and the use of antibiotic treatment can be seen in the different manners of death common to each sample. The individuals in the Lisbon sample, thus, would have had a higher disease load relative to children in the New Mexico sample.

The different causes of death potentially indicate a mortality bias among the Lisbon sample, as many of the children present died of chronic diseases, such as tuberculosis and upper respiratory diseases, which could have drastically stunted their growth (Wood et al. 1992). In contrast, there has not been a strong mortality bias found among the New Mexico sample when comparing those who died of accidental and natural causes of death (Spake 2020). It is therefore important to keep in mind the idea that the proportionally smaller cortical area and larger medullary cavity size found among the Lisbon sample could be the result of the sample being composed largely of children who died of disease exposure. While this is likely a factor, the contextual information demonstrates that living conditions were poor for the entire childhood population of Lisbon at the time. Thus, it is possible there were many children who survived their exposure to this biosocial stress and demonstrated similar cortical distribution patterns into adult life or experienced catch up growth later (Wood et al. 1992; Spake 2020). Additionally, though there may be mortality bias towards stunted appositional growth among the non-survivors of the Lisbon sample, Saunders and Hoppa (1993) have noted that the mortality bias error introduced by using non-survivors as samples when studying linear growth is minimal, though there is some evidence of mortality bias in linear growth among contemporary subadult populations (Spake et al. 2022). Future research should

investigate the relationship between survivorship in childhood and cortical distribution in adults.

Socioeconomic status has the potential to create a more nuanced understanding within each sample, rather than just modern versus historic, or light versus heavy. Within this sample the unstandardized cortical area and medullary cavity values suggest a larger general cross-sectional size among the high socioeconomic status groups from both samples, but the differences were generally not significant. This may indicate that in both samples individuals in higher SES groups have fewer nutritional deficiencies and lower disease loads, but further research is needed with larger sample sizes. The overall impact of socioeconomic status on either mechanical loading or biosocial stress is yet to be fully understood. Previous studies have suggested a relationship between reduced cortical thickness in lower SES groups when compared with their higher SES counterparts (Mays et al. 2009; Beauchamp 2017), a finding that matches the overall smaller size of cortical area observed in the low SES groups of both samples. The nature of these samples have been archaeological, and individual-level documentation has been limited. Thus, there have been no opportunities to tease out the different factors affecting bone deposition that would amount to SE stress. Clearly biosocial factors like nutrition and disease would be variable between low and high SES groups, but the impact of different activity levels or body mass has been largely taken for granted. This research represents the first attempt to investigate the different factors that make up SES stress. Future research should attempt to examine in further detail the differences in mechanical loading present within different SES groups. Additionally, future research could examine the impact of membership to racialized and/or marginalized groups to determine if cortical bone deposition may be responsive to the biosocial stress these individuals can experience.

Based on the standardized medullary cavity area values and the biosocial stress levels present in each sample it appears that the children in the Lisbon sample were experiencing higher rates of endosteal resorption than those in the New Mexico sample. It has been hypothesized that medullary loss in response to energetic deficiencies may result from a need for bone minerals to be released and/or the need to expand the amount of medullary tissue present in the bone (Garn 1969; Eleazer and Jankauskas 2016). When examining how caloric restriction impacts growing mice, Devlin and colleagues (2010) found that subadult mice that were placed on calorie restricted diets

demonstrated significantly lower instances of cortical bone formation and significantly higher levels of bone endosteal resorption in the femur. This is likely due to increased adipogenesis occurring in the medullary cavity, a process that can be brought on in response to inadequate caloric intake (Piotrowska and Tarnowski 2021). It is possible that a similar type of resorption is occurring during development in the Lisbon sample, a pattern that is absent from the children in the New Mexico sample.

5.4. Conclusions and Future Steps

The results of this study reinforce the importance of examining the contextual nature of skeletal growth and development. The children in the two samples examined in this research are separated in time by less than a century, yet the patterns of cortical bone development present in their femora, tibiae, and humeri demonstrate a very different pattern of growth which reflects the different periods' social, cultural, political, and economic contexts. The goal of this research was to examine and untangle some of the factors that influence cortical bone deposition during growth and development, from examining the methods researchers use to quantify differences in cortical bone distribution to the impacts of the growth environment on cortical bone deposition. This was undertaken using a unique set of samples consisting of children from documented skeletal collections from Lisbon, Portugal and New Mexico, USA. This dissertation presents several novel conclusions regarding the measurement of cortical bone distribution and the impact of the growth environment on cortical bone deposition among children.

The first paper examined intraobserver error among cortical area measurements taken from CT scans of dry bone and wet bone samples, with the additional goal of determining if cortical areas derived from algorithmic methods were comparable to those found using the manual segmentation method. The results suggest that there are more instances of significant differences in intra-observer error among the wet bone samples than the dry bone samples, but fewer instances of significant differences between each method among each observer round in the wet bone sample. The results of the study also suggest that the use of automated algorithms is comparable to manual methods of segmentation. While the error rates between rounds of measurements taken from dry and wet bone do not differ hugely when taken from comparable low resolution CT scans, further research using the same bone samples, both in dry and wet, is needed to confirm

whether differences exist between cortical area measurements taken from the same specimen.

The second paper developed weight estimation formulae based on sections of the femur and tibia using a documented weight and age reference sample of 70 children from the New Mexico sample many of whom were overweight and obese. Weight estimates created using the formulae developed in the study were then compared with known weight to examine error. Formulae developed using J values generally had the lowest MSE and MAR values and demonstrated an increased utility for individuals with BMI falling above the 95th percentile than the breadth formulae. The formulae produced using the breadth measurements were not as precise, and demonstrated a bias in mean residuals, likely resulting from the heteroscedasticity between the sexes and age categories combined with the process of de-transforming the data. These formulae generally underestimated weight in younger and lighter individuals, and overestimated weight in older and heavier individuals. In terms of the new cross-sections examined within this study, the tibia performed similarly to the femoral midsection when it came to acting as a weight estimation parameter. While the femoral mid-shaft remained generally the most precise weight estimation parameter, tibial cross-section formulae could be used to estimate weight if the femoral midpoint were not available. Additionally, formulae using several sections of the femur demonstrated comparable MSE, MAR, and MR values to those produced by femoral mid-shaft formulae, and thus could be used if the mid-shaft was not available or was damaged. Based on these results it is suggested that the J formulae are more useful when it comes to accurately estimating weight in children ages birth to 5 years. These formulae are uniquely suited for estimating weight among forensic populations of modern children in industrialized nations, many of whom are overweight or obese. Future research should attempt to examine the differential effects of lean muscle and fat mass on torsional rigidity among children, so as to better understand the impact of obesity on bone strength.

The third paper examined the relationships between biosocial stress, mechanical loading, and the amount and distribution of cortical bone. The study used cross-sectional bone data from known sex and age children from the Luís Lopes Documental Skeletal Collection (n=61) and the New Mexico Decedent Image Database (n=45). The study addressed two main research questions: 1) is there a difference in cortical distribution between individuals in the early 20th century Portuguese sample and the early 21st

century New Mexico sample; and 2) is there a relationship within each sample between socioeconomic status (SES) and cortical bone distribution. This research confirms the influence of biosocial stress on increased medullary cavity size and decreased the thickness of the cortex. It is likely that the increased biosocial stress present among the individuals the Lisbon sample caused increased bone resorption in the endosteal aspect of the bone and smaller unstandardized cross-sectional areas, while the high level of mechanical loading preserved a relatively high degree of bone robusticity, especially in the upper limb. In addition to having generally larger cross-sections, the New Mexico sample appears to demonstrate more cortical area and smaller medullary cavities among most of the bone cross-sections. This is likely the result of the relatively low levels of biosocial stress present in the New Mexico population. The robusticity of the humeri in the New Mexico sample are smaller than those in the Lisbon sample, likely resulting from their decreased physical activity. The increased BMI typical of those in the New Mexico sample appeared to result in increased robusticity in some sections of the lower limb close to the hip and knee, while other areas closer to the mid-shaft appeared to be more influenced by physical activity. Thus, though physical activity has decreased in the modern sample, overall bone loading may not have been as drastically reduced due to the impact of heavier body mass among New Mexico children. Future studies attempting to tease out the effects of body weight and physical activity on bending rigidity and strength should not only examine the mid-shafts of the femur, but also multiple sections of the lower limb and of the upper limb. Future research on the topic should also look to improve an understanding of the mechanisms by which something like SES can impact mechanical loading of the bone.

Overall, this research demonstrates that skeletal development is population (and perhaps subpopulation) specific, and that biosocial and mechanical experiences during growth and development are not universal among children. Individual long bones, or sections of long bones can contain dramatically different cortical distributions depending on body mass, physical activity level, nutritional status, and disease load. Thus, caution should be taken when attempting to use generalized methodologies to learn about children in both the past and the present. Children are not simply small adults, nor is every child or group of children alike. They deserve research and methods that are developed for their unique experiences of growth and development.

References

Agarwal SC, Grynpas, MD. 2009. Measuring and interpreting age-related loss of vertebral bone mineral density in a medieval population. American Journal of Physical Anthropology, 139(2): 244–252.

Agarwal SC. 2016. Bone morphologies and histories: Life course approaches in bioarchaeology. American Journal of Physical Anthropology 159: S130–S149.

AlQahtani SJ, Hector MP, Liversidge HM. 2010. Brief communication: the London atlas of human tooth development and eruption. American Journal of Physical Anthropology. 142:481–490

Auerbach, BM, Ruff CB. 2004. Human body mass estimation: A comparison of "morphometric" and "mechanical" methods. American Journal of Physical Anthropology 125(4): 331–342.

Bass WM. 1979. Developments in the identification of human skeletal material (1968–1978). American Journal of Physical Anthropology. 51(4):555-62.

Bass SL. 2000. The prepubertal years: a uniquely opportune stage of growth when the skeleton is most responsive to exercise? Sports Medicine 30(2): 73–78.

Beck TJ, Petit MA, Wu G, LeBoff, MS, Cauley JA, Chen Z. 2009. Does obesity really make the femur stronger? BMD, geometry, and fracture incidence in the women's health initiative-observational study. Journal of Bone and Mineral Research. 24(8):1369-79.

Berendsen, AD, Olsen, BR. 2015. Bone development. Bone. 80:14–18. https://doi.org/10.1016/j.bone.2015.04.035

Bonjour JP, Ammann P, Chevalley T, Rizzoli R. 2001. Protein intake and bone growth. Canadian Journal of Applied Physiology. 26:153-166.

Bouxsein ML, Myburgh KH, van der Meulen MCH, Lindenberger E, Marcus R. 1994. Age-related differences in cross-sectional geometry of the forearm bones in healthy women. Calcified Tissue International 54(2): 113–118.

Brenton BP, Paine RR. 2007. Revaluating the health and nutritional status of maize-dependent populations: evidence for the impact of pellagra on human skeletons from South Africa. Ecology of Food and Nutrition 46:345–360.

Brickley M, Ives R. 2008. The bioarchaeology of metabolic bone disease. San Diego: Elsevier.

Bogin B. 1988. Patterns of human growth. Cambridge University Press: Cambridge.

Bogin B. 1999. Evolutionary perspective on human growth. Annual Review of Anthropology 28:109–153.

Bogin B, Varela Silva MI, Rios L. 2007. Life History Trade-Offs in Human Growth: Adaptation or Pathology? American Journal of Human Biology.19; 631–642. https://doi.org/10.1002/ajhb

Bourne R. 2010. ImageJ. Fundamentals of Digital Imaging in Medicine. Springer Science & Business Media

Campinho A. 1995. Regime jurı´dico do contrato de trabalho de menores. Histo´ria, legislac¸a˜o, anotac¸o˜es e convenc¸o˜es interna- cionais. Braga: Editora Correio do Minho.

Cardoso HFV. 2005. Patterns of Growth and Development of the Human Skeleton and Dentition in Relation to Environmental Quality. PhD thesis. McMaster University.

Cardoso HFV. 2006. Brief communication: The collection of identified human skeletons housed at the Bocage Museum (National Museum of Natural History), Lisbon, Portugal. American Journal of Physical Anthropology 129(2): 173–176.

Cardoso HFV. 2007. Environmental effects on skeletal versus dental development: using a documented subadult skeletal sample to test a basic assumption in human osteological research. American Journal of Physical Anthropology 132(2):223-33.

Cardoso HFV, Garcia, S. 2009. The not-so-dark ages: Ecology for human growth in medieval and early twentieth century Portugal as inferred from skeletal growth profiles. American Journal of Physical Anthropology 138(2): 136–147.

Cardoso HFV, Abrantes J, Humphrey LT. 2014. Age estimation of immature human skeletal remains from the diaphyseal length of the long bones in the postnatal period. International Journal of Legal Medicine, 128(5): 809–824.

Cardoso HFV, Meyers J, Liversidge, HM. 2019. A Reappraisal of Developing Deciduous Tooth Length as an Estimate of Age in Human Immature Skeletal Remains. Journal of Forensic Sciences 64(2): 385–- 392.

Carter DR, Beaupré GS. 2001. Skeletal function and form. Cambridge: Cambridge University Press.

Clarke B. 2008. Normal Bone Anatomy and Physiology. Clinical Journal of the American Society of Nephrology, 3: 131–139. https://doi.org/10.2215/CJN.04151206

Cole TJ, Freeman, JV Preece, MA. 1998. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. Statistics in Medicine 17(4): 407-429.

Conceição EL, Cardoso HF. 2011. Environmental effects on skeletal versus dental development II: further testing of a basic assumption in human osteological research. American Journal of Physical Anthropology 144(3):463-70.

Corron L, Marchal F, Condemi S, Chaumoitre K, Adalian P. 2017. Evaluating the consistency, repeatability, and reproducibility of osteometric data on dry bone surfaces, scanned dry bone surfaces, and scanned bone surfaces obtained from living individuals. Bulletins et mémoires de la Socié'é ''anthropologie de Paris (1):33-53.

Cowgill LW, Hager, LD. 2007. Variation in the development of postcranial robusticity: An example from Catalhoyuk, Turkey. International Journal of Osteoarchaeology 17(3): 235–252.

Cowgill LW, Trinkaus E, Zeder MA. 2007. Shanidar 10: a Middle Paleolithic immature distal lower limb from Shanidar Cave, Iraqi Kurdistan. Journal of Human Evolution 53: 213-223.

Cowgill LW. 2010. The ontogeny of Holocene and Late Pleistocene human postcranial strength. American Journal of Physical Anthropology 141(1): 16–37.

Cowgill LW. 2014. Femoral diaphyseal shape and mobility: an ontogenetic perspective. Reconstructing mobility: Environmental, behavioral, and morphological determinants. 193-208.

Cowgill LW. 2018. Juvenile body mass estimation: A methodological evaluation. Journal of Human Evolution 115: 78–84.

Cunningham C, Scheuer L, Black S. 2016. Developmental juvenile osteology. Academic press.

Davies TG, Stock JT. 2014. The influence of relative body breadth on the diaphyseal morphology of the human lower limb. American Journal of Human Biology. 26(6), 822–835.

Degnin CR, Laederich MB, Horton WA. 2010. FGFs in endochondral skeletal development. Journal of Cellular Biochemistry, 110(5): 1046–1057. https://doi.org/10.1002/jcb.22629

Devlin MJ. 2011. Estrogen, exercise, and the skeleton. Evolutionary Anthropology 20(2): 54–61.

Doube M, Kłosowski MM, Arganda-Carreras I, Cordelières FP, Dougherty RP, Jackson JS, Schmid B, Hutchinson JR, Shefelbine SJ. 2010. BoneJ: free and extensible bone image analysis in ImageJ. Bone 47(6):1076-9.

Elliott J. 2022 Radiographic Technique for Archaeological Human Dry Bones: a scoping review, Internet Archaeology 59.

Eleazer CD, Jankauskas R. 2016. Mechanical and Metabolic Interactions in Cortical Bone Development. American Journal of Physical Anthropology 333(2015): 317–333.

Garn S, Rohmann C, Behar M, Viteri F, Guzman M. 1964. Compact bone deficiency in protein calorie malnutrition. Science (145):1444–1445

Garn, SM, Guzmán, MA, Wagner, B. 1969. Subperiosteal gain and endosteal loss in protein-calorie malnutrition. American Journal of Physical Anthropology 30(1): 153–155.

Gooderham E, Marinho L, Spake L, Fisk S, Prates C, Sousa S, Oliveira C, Santos AL, Cardoso HF. 2020. Severe skeletal lesions, osteopenia and growth deficit in a child with pulmonary tuberculosis (mi^{d-}20th century, Portugal). International Journal of Paleopathology1(30):47-56.

Goodman AH, Armelagos GJ. 1988. Childhood Stress and Decreased Longevity in a Prehistoric Population. American Anthropologist 90(4): 936–944.

Gosman JH, Hubbell ZR, Shaw CN, Ryan TM. 2013. Development of cortical bone geometry in the human femoral and tibial diaphysis. The Anatomical Record. 296(5):774-87.

Goulart P, Bedi AS. 2017. The evolution of child labor in Portugal, 1850–2001. Social Science History 41(2): 227–254.

Gowland RL. 2015. Entangled Lives: Implications of the Developmental Origins of Health and Disease Hypothesis for Bioarchaeology and the Life Course. American Journal of Physical Anthropology (July): 530–540.

Green WT, Wyatt OM, Anderson M. 1946. Orthoroentgenography as a Method of Measuring the Bones of the Lower Extremities. Journal of Bone and Joint Surgery 28: 60-65.

Guardado Moreira MJ, de Castro Henriques F. 2016. Demographic and health changes in Portugal (1900–2013). Hygiea Internationalis. 12(1): 9–39.

Halfon N, Larson K, Son J, Lu M, Bethell C. 2017. Income Inequality and the Differential Effect of Adverse Childhood Experiences in US Children. Academic Pediatrics. 17(7): S70–S78.

Harrington L. 2010. Ontogeny of postcranial robusticity among Holocene hunter-gatherers of southernmost Africa. PhD Thesis. University of Toronto.

Harrington L, Osipov B. 2018. The Developing Forager: Reconstructing Childhood Activity Patterns from Long Bone Cross-sectional Geometry. In S. Crawford, D. M. Hadley, & G. Shephard (Eds.), The Oxford Handbook of the Archaeology of Childhood 1–23.

Hind K, Gannon L, Whatley E, Cooke, C, Truscott J. 2012. Bone cross-sectional geometry in male runners, gymnasts, swimmers and non-athletic controls: a hip-structural analysis study. European Journal of Applied Physiology 112(2): 535-541.

Holick, MF, Nieves JW. 2015. Nutrition and Bone Health. Springer Scie148usinessness Media.

Holt BM. 2003. Mobility in Upper Paleolithic and Mesolithic Europe: evidence from the lower limb. American Journal of Physical Anthropology 122: 200–215.

Hummert JR. 1983. Cortical bone growth and dietary stress among subadults from Nubia's Batn El Hajar. American Journal of Physical Anthropology 62:167–176.

Humphrey LT. 1998. Growth patterns in the modern human skeleton. American Journal of Physical Anthropology 105:57–72.

Huss-Ashmore R. 1981. Bone growth and remodeling as a measure of nutritional stress. Biocultural Adaptation: Comprehensive Approaches to Skeletal Analysis, Martin DL, Bumsted MP (eds). Research Reports (20): 84-95.

Hoyert DL, Kochanek KD, Murphy SL. 1999. Deaths: final data for 1997. Hyattsville, Maryland: US Department of Health and Human Services, CDC, National Center for Health Statistics. National Vital Statistics Report 47(20).

Johnson ML, Veldhuis JD, Lampl M. 1996. Is growth saltatory? The usefulness and limitations of frequency distributions in analyzing pulsatile data. Endocrinology 137(12):5197-204. DOI: 10.1210/endo.137.12.8940335.

Karr LP, Outram AK. 2012. Tracking changes in bone fracture morphology over time: environment, taphonomy, and the archaeological record. Journal of Archaeological Science 39(2):555-9.

Kuczmarski RJ, Ogden CL, Guo SS, Grummer-Strawn LM, Flegal MS, Mei Z, Wei R, Curtin LR, Roche AF, Johnson CL. 2002. 2000 CDC Growth Charts for the United States: methods and development. Vital Health Statistics 11(246):1–190.

Kurki HK, Holland S, MacKinnon M, Cowgill L, Osipov B, Harrington L. Appositional long bone growth: Implications for measuring cross-sectional geometry. American Journal of Biological Anthropology. 2022 Oct;179(2):291-306.

Kuzawa CW. 2005. Fetal origins of developmental plasticity: are fetal cues reliable predictors of future nutritional environments? American Journal of Human Biology. 17(1):5-21.

Komlos J, Breitfelder A, Sunder M. 2009. The transition to post-industrial BMI values among US children. American Journal of Human Biology 21:151–60.

Konigsberg LW, Hens SM, Jantz LEEM, Jungers WL. 1998. Stature Estimation and Calibration: Bayesian and Maximum Likelihood Perspectives in Physical Anthropology 41: 65–92.

Lampl M, Veldhuis JD, Johnson ML. 1992. Saltation and Stasis: A Model of Human Growth. Science 258(October): 801–804.

Lannon F. 1987. Privilege, persecution and prophecy. The catholic Church in Spain, 1875-1975, Oxford, Clarendon Press.

Lasker GW. 1969. Human Biological Adaptability: The ecological approach in physical anthropology. Science. 166(3912):1480-6.

Lieberman DE, Devlin MJ, Pearson OM. 2001. Articular area responses to mechanical loading: effects of exercise, age, and skeletal location. American Journal of Physical Anthropology 116: 266–277.

Lovejoy C, Owen, Burstein AH, Heiple KG. 1976. The biomechanical analysis of bone strength: a method and its application to platycnemia. American Journal of Physical Anthropology 44(3): 489-505.

Lucy D. 2005. Introduction to statistics for forensic scientists. Chichester, West Sussex, England: Wiley.

Macintosh AA, Davies TG, Ryan TM, Shaw CN, Stock JT. 2013. Periosteal versus true cross-sectional geometry: A comparison along humeral, femoral, and tibial diaphyses. American Journal of Physical Anthropology. 150(3): 442–452.

Macintosh AA., Pinhasi R, Stock JT. 2014. Divergence in male and female manipulative behaviors with the intensification of metallurgy in Central Europe. PLoS One 9(11): 112-116.

MacKelvie K, Khan K, McKay H, Sanborn C. 2002. Is there a critical period for bone response to weight-bearing exercise in children and adolescents? A systematic review. British Journal of Sports Medicine 36(4): 250–257.

Mackie EJ, Ahmed YA, Tatarczuch L, Chen K, Mirams M. 2008. Endochondral ossification: How cartilage is converted into bone in the developing skeleton. The International Journal of Biochemistry and Cell Biology, 40: 46–62. https://doi.org/10.1016/j.biocel.2007.06.009

Maratova K, Hradsky O, Matyskova J, Copova I, Soucek O, Sumnik Z, Bronsky J. 2017. Musculoskeletal system in children and adolescents with inflammatory bowel disease: normal muscle force, decreased trabecular bone mineral density and low prevalence of vertebral fractures. European Journal of Pediatrics 176(10) 1355–1363.

Maresh MM. 1943. Growth of major long bones in healthy children: a preliminary report on successive roentgenograms of the extremities from early infancy to twelve years of age. American Journal of Diseases of Children 66(3): 227–257.

Maresh MM. 1955. Linear growth of long bones of extremities from infancy through adolescence; continuing studies. American Journal of Diseases of Children, 89(6): 725–742.

Maresh MM. 1970 Measurements from roentgenograms. In (RW McCammon, Ed.) Human Growth and Development. Springfield, IL: Charles C.

Margarida Marques M, Valente Rosa MJ, Martins JL. 2007. School and diversity in a weak state: The Portuguese case. Journal of Ethnic and Migration Studies 33(7): 1145–1168.

Marphatia AA, Cole TJ, Grijalva-Eternod C, Wells JCK. 2016. Associations of gender inequality with child malnutrition and mortality across 96 countries. Global Health, Epidemiology and Genomics 1: 1-8.

Mathur P, Pillai R. 2019. Overnutrition: Current scenario & combat strategies. Indian Journal of Medical Research 149(6):695-705. DOI: 10.4103/ijmr.IJMR_1703_18.

May S, Ives R, Brickley M. 2009. The effects of socioeconomic status on endochondral and appositional bone growth, and acquisition of cortical bone in children from 19th century Birmingham, England. American Journal of Physical Anthropology. 140(3): 410–416.

McCammon RW. 1970. Human growth and development. Springfield, IL: Charles C. Thomas. Merchant

Meckel, RA. 1990. Save the babies: American public health reform and the prevention of infant mortality, 1850-1929. Baltimore, Maryland: The Johns Hopkins University Press.

Moreira MV. 1950. Problemas da habitac¸a˜o (ensaios sociais). Lisboa: Minerva.

Murray AA, Erlandson MC. 2021. Tibial cortical and trabecular variables together can pinpoint the timing of impact loading relative to menarche in premenopausal females. American Journal of Human Biology 1–19.

Must A, Dallal GE, Dietz WH. 1991. Reference data for obesity: 85th and 95th percentiles of body mass index (wt/ht2) and triceps skinfold thickness. The American Journal of Clinical Nutrition 53(4): 839-846.

Newman SL, Gowland RL. 2017. Dedicated Followers of Fashion? Bioarchaeological Perspectives on Socio-Economic Status, Inequality, and Health in Urban Children from the Industrial Revolution (18th–19th C), England. International Journal of Osteoarchaeology. 27(2): 217–229.

Nyman JS, Roy A, Shen X, Acuna RL, Tyler JH, Wang X. 2006. The influence of water removal on the strength and toughness of cortical bone. Journal of Biomechanics 39(5):931-8.

Ogden CL, Carroll MD, Flegal KM. 2008. High body mass index for age among US children and adolescents, 2003-2006. Journal of the American Medical Association 299(20): 2401–2405.

O'Neill MC, Ruff CB. 2004. Estimating human long bone cross-sectional geometric properties: A comparison of noninvasive methods. Journal of Human Evolution 47(4): 221–235.

Osipov B, Temple D, Cowgill L, Harrington L, Bazaliiskii VI, Weber AW. 2016. Evidence for genetic and behavioral adaptations in the ontogeny of prehistoric hunter-gatherer limb robusticity. Quaternary International 405: 134–146.

Osipov B. 2018. The Ontogeny of Postcranial Robusticity and Shape in Middle Holocene Cis-Baikal Hunter-Gatherer Populations. PhD Thesis. University of Alberta.

Osipov B, Harrington L, Temple D, Bazaliiskii VI, Weber AW. 2020. Chronological and regional variation in developmental stress and behavior of Early and Late Neolithic Cis-Baikal hunter-gatherer juveniles: Insights from diaphyseal cross-sectional geometry. Archaeological Research in Asia 24 (100231).

Otsu N. 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9(1): 62-66.

Ottow C, Schulz R, Pfeiffer H, Heindel W, Schmeling A, Vieth V. 2017. Forensic age estimation by magnetic resonance imaging of the knee: the definite relevance in bony fusion of the distal femoral- and the proximal tibial epiphyses using closest-to-bone T1 TSE sequence. Forensic Medicine 27: 5041–5048.

Piotrowska K, Tarnowski M. 2021. Bone Marrow Adipocytes — Role in Physiology and Various Nutritional Conditions in Human and Animal Models. Nutrients 13(1412): 1–14.

Pinhasi R, Shaw P, White B, Ogden AR. 2006. Morbidity, rickets and long-bone growth in post-medieval Britain—a cross-population analysis. Annual Review of Human Biology 33:372–389.

Pokines JT. 2016. Taphonomic Alterations to Terrestrial Surface-Deposited Human Osseous Remains in a New England Environment. Journal of Forensic Identification 66(1).

Pokines JT, Appel N, Pollock C, Eck CJ, Maki AG, Joseph AS, Cadwell L, Young CD. 2017. Anatomical Taphonomy at the Source: Alterations to a Sample of 84 Teaching Skulls at a Medical School. Journal of Forensic Identification 67(4).

Pomeroy E, Stock JT, Stanojevic S, Miranda JJ, Cole TJ, Wells JCK. 2012. Trade-Offs in Relative Limb Length among Peruvian Children: Extending the Thrifty Phenotype Hypothesis to Limb Proportions. PloS ONE 7(12).

Pomeroy E, Macintosh A, Wells JCK, Cole TJ, Stock JT. 2018. Relationship between body mass, lean mass, fat mass, and limb bone cross-sectional geometry: Implications for estimating body mass and physique from the skeleton. American Journal of Physical Anthropology. 166(1): 56–69.

Prentice A. 2001. The relative contribution of diet and genotype to bone development. Proceedings of the Nutritional Society 60: 45–52.

Raichlen DA, Gordon AD, Foster AD, Webber JT, Sukhdeo SM, Scott RS, … Ryan TM. 2015. An ontogenetic framework linking locomotion and trabecular bone architecture with applications for reconstructing hominin life history. Journal of Human Evolution (81):1–12.

Rajamannan NM. 2018. Osteocardiology: Cardiac Bone Formation. In Osteocardiology (pp. 45–54). https://doi.org/10.1007/978-3-319-64994-8

Rasband WS. 2018. ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA.

Reeves NM. 2014. Augmenting functional adaptation: Does obesity have a systemic effect on bone strength properties in humans? PhD Thesis. University of Tennesse.

Rewekant A, Jerszynska B. 1995. Patterns of cortical bone growth in children: an example from medieval populations. Anthropologie 1:79-82.

Ridler TW, Calvard S. 1978. Picture thresholding using an iterative selection method. IEEE Transactions on Systems, Man, and Cybernetics 8(8): 630-632.

Robbins G, Sciulli PW, Blatt SH. 2010. Estimating body mass in subadult human skeletons. American Journal of Physical Anthropology. 143:146–150.

Robbins Schug G., Gupta S, Cowgill LW, Sciulli PW, Blatt SH. 2013. Panel regression formulas for estimating stature and body mass from immature human skeletons: A statistical approach without reference to specific age estimates. Journal of Archaeological Science 40(7): 3076–3086.

Robling AG, Castillo AB, Turner CH. 2006. Biomechanical and Molecular Regulation of Bone Remodeling. Annual Review of Biomedical Engineering 8: 455–498.

Rogol AD, Clark PA, Roemmich JN. 2000. Growth and pubertal development in children and adolescents: effects of diet and physical activity 1 – 4. The American Journal for Clinical Nutrition. 72: 521S-528S.

Ruff CB, Hayes WC. 1983. Cross-sectional geometry of Pecos Pueblo femora and tibiae—A biomechanical investigation: II. Sex, age, and side differences. American Journal of Physical Anthropology 60(3): 383–400.

Ruff CB, Trinkaus E, Walker A, Spencer Larsen C. 1993. Post- cranial robusticity in Homo I: temporal trends and bio- mechanical interpretation. American Journal of Physical Anthropology 91(53):21–53.

Ruff CB. 1994. Morphological adaptation to climate in modern and fossil hominids. Yearbook of Physical Anthropology 37:65–107.

Ruff CB, Walker A, Trinkaus, E. 1994. Postcranial robusticity in Homo. III: ontogeny. American Journal of Physical Anthropology 93: 35–54.

Ruff CB. 2000. Body size, body shape, and long bone strength in modern humans. Journal of Human Evolution 38:269–290.

Ruff CB. 2003a. Growth in bone strength, body size, and muscle size in a juvenile longitudinal sample. Bone. 33(3): 317–329.

Ruff CB, 2003b. Ontogenetic adaptation to bipedalism: age changes in femoral to humeral length and strength proportions in humans, with a comparison to baboons Journal of Human Evolution 45(4): 317-349.

Ruff CB, Holt B, Trinkaus E. 2006. Who's afraid of the big bad Wolff?: "Wolff's law" and bone functional adaptation. American Journal of Physical Anthropology 129(4): 484–498.

Ruff, CB. 2007. Body size prediction from juvenile skeletal remains. Skeletal Remains. American Journal of Physical Anthropology 716: 698–716. https://doi.org/10.1002/ajpa

Ruff CB, Garofalo E, Holmes MA. 2013. Interpreting skeletal growth in the past from a functional and physiological perspective. American Journal of Physical Anthropology 150(1): 29–37.

Ruff CB. 2021. Cortical thickness data from Denver Growth Study (Unpublished). Attained through personal correspondence.

Saers JP, Demars LJ, Stephens NB, Jashashvili T, Carlson KJ, Gordon AD, Ryan TM Stock JT. 2021. Automated resolution independent method for comparing in vivo and dry trabecular bone. American Journal of Physical Anthropology 174: 822–831.

Sasaki N, Enyo A. 1995. Viscoelastic properties of bone as a function of water content. Journal of Biomechanics 28, 809–815

Saunders SR, Hoppa RD. 1993. Growth deficit in survivors and non-survivors: biological mortality bias in subadult skeletal samples. American Journal of Physical Anthropology. 36(17):127-51.

Schell LM, Gallo MV, Ravenscroft J. 2009. Environmental influences on human growth and development: historical review and case study of contemporary influences. Annals of Human Biology. 36(5):459-77.

Schell LM, Rousham EK. 2022. Environmental effects on growth. In Human growth and development (pp. 261-315). Academic Press.

Sciulli PW, Blatt SH. 2008. Evaluation of juvenile stature and body mass prediction. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists 136(4): 387-393.

Shaw CN, Stock JT. 2009. Habitual Throwing and Swimming Correspond With Upper Limb Diaphyseal Strength and Shape in Modern Human Athletes. American Journal of Physical Anthropology 140: 160–172.

Shaw CN, Stock JT. 2011. The influence of body proportions on femoral and tibial midshaft shape in hunter-gatherers. American Journal of Physical Anthropology. 144(1): 22–29. https://doi.org/10.1002/ajpa.21363

Shaw CN, Ryan, TM. 2012. Does skeletal anatomy reflect adaptation to locomotor patterns? Cortical and trabecular architecture in human and nonhuman anthropoids. American Journal of Physical Anthropology 147(2): 187–200.

Shaw CN, Hofmann CL, Petraglia MD, Stock JT, Gottschall JS. 2012. Neandertal humeri may reflect adaptation to scraping tasks, but not spear thrusting. PloS ONE 7(7): 1–8.

Scheuer L, MacLaughlin-Black S. 1994. Age estimation from the pars basilaris of the fetal and juvenile occipital bone. International Journal of Osteoarchaeology.4(4):377-80.

Sims NA, Gooi JH. 2008. Seminars in Cell & Developmental Biology Bone remodeling : Multiple cellular interactions required for coupling of bone formation and resorption. Seminars in Cell & Developmental Biology. 19: 444–451. https://doi.org/10.1016/j.semcdb.2008.07.016

Sinclair D. Human growth after birth. London: Oxford University Press, 1978.

Sládek V, Berner M, Sailer R, Stock JT, Macintosh A. 2015. Lower limb biomechanics and habitual mobility among mid-Holocene populations of the Cis-Baikal. Quaternary International 130(3): 1–10.

Smith DM, Nance WE, Kang KW, Christian JC, Johnson CC. 1973. Genetic factors in determining bone mass. Journal of Clinical Investigation 52: 2800– 2808.

Smith RJ. 1993. Logarithmic transformation bias in allometry. American Journal of Physical Anthropology 90:215–228.

Spake, L. 2020. Using anthropometrics and dental formation stages of contemporary children to investigate the impact of biological mortality bias on interpretations of past population health.

Spake L, Cardoso HF. 2019. Indirect evidence for biological mortality bias in growth from two temporo-spatially distant samples of children. Anthropologischer Anzeiger. 8;76(5):379-90.

Spake L, Meyers J, Blau S, Cardoso HFV, Lottering N. 2020. A simple and software-independent protocol for the measurement of post-cranial bones in anthropological contexts using thin-slab maximum intensity projection. Forensic Imaging 20(200354): 1–5.

Spake L, Meyers J, Cardoso HFV. 2021. Juvenile Body Mass Estimation from the Femur Using Postmortem Computed Tomography Data. Human Biology 93(20: 125-137.

Spake L, Hoppa RD, Blau S, Cardoso HF. 2022. Biological mortality bias in diaphyseal growth of contemporary children: Implications for paleoauxology. American Journal of Biological Anthropology. 178(1):89-107.

Stock J, Pfeiffer S. 2001. Linking structural variability in long bone diaphysis to habitual behaviors: Forager from the South African Later Stone Age and the Andaman Islands. American Journal of Physical Anthropology 115(4): 337–348.

Stock JT. 2006. Hunter-gatherer postcranial robusticity relative to patterns of mobility, climatic adaptation, and selection for tissue economy. Am. J. Phys. Anthropol. 131: 194–204.

Stock JT, Shaw CN. 2007. Which Measures of Diaphyseal Robusticity Are Robust? A Comparison of External Methods of Quantifying the Strength of Long Bone Diaphyses to Cross-Sectional Geometric Properties. American Journal of Physical Anthropology 134(4): 412–423.

Sumner DR, Mockbee B, Morse K, Cram T, Pitt M. 1985. Computed tomography and automated image analysis of prehistoric femora. American Journal of Physical Anthropology 68: 225–232.

Sun SS, Deng X, Sabo R, Carrico R, Schubert CM, Wan W, Sabo C. 2012. Secular trends in body composition for children and young adults: the Fels Longitudinal Study. American Journal of Human Biology: The Official Journal of the Human Biology Council 24(4): 506–514.

Swan KR, Ives R, Wilson LAB, Humphrey LT. 2020. Ontogenetic changes in femoral cross-sectional geometry during childhood locomotor development. American Journal of Physical Anthropology, (May), 1–16.

Temple DH, Goodman AH. 2014. Bioarcheology has a “health” problem: Conceptualizing “stress” and “health” in bioarcheological research. American Journal of Physical Anthropology 155(2): 186–191.

Temple DH. 2019. Bioarchaeological evidence for adaptive plasticity and constraint: Exploring life-history trade-offs in the human past. Evolutionary Anthropology 28(1): 34–46.

Trent M, Dooley DG, Dougé J. 2019. The impact of racism on child and adolescent health. Pediatrics. 144(2); 1-16.

Trinkaus E, Ruff, CB, Esteves F, Coelho JMS, Silva M, Mendova M. 2002. The lower limb remains. In: Zilh~ao J, Trinkaus E. (Eds.), Portrait of the Artist as a Child. Intituto Portugues de Arqueologia, Lisbon: 435-465.

Troiano RP, Flegal KM, Kuczmarski RJ, Campbell SM, Johnson CL. 1995. Overweight prevalence and trends for children and adolescents: the National Health and Nutrition Examination Surveys, 1963 to 1991. Archives of Pediatrics & Adolescent Medicine 149(10): 1085-1091.

Trowbridge FL, Marks JS, Lopez de Romana G, Madrid S, Boutton TW, Klein PD. 1987. Body composition of Peruvian children with short stature and high weight-for-height. II Implications for the interpretation for weight-for-height as an indicator of nutritional status, The American Journal of Clinical Nutrition 46(3): 411–418.

Tylavsky FA, Ferrara A, Catellier DJ, Oken E, Li X, Law A, Dabelea D, Rundle A, Gilbert-Diamond D, Hivert MF, Breton CV, Cassidy-Bushrow AE, Mueller NT, Hunt KJ, Arteaga SS, Lombo T, Mahabir S, Ruden D, Sauder K, Hedderson MM, Zhu Y, Polk S, Mihalopoulos NL, Vos M, Pyles L, Roary M, Aschner J, Karagas MR, Trasande L. 2019. Understanding childhood obesity in the US: the NIH environmental influences on child health outcomes (ECHO) program. International Journal of Obesity.

Ulijaszek SJ, Kerr DA. 1999. Review article Anthropometric measurement error and the assessment of nutritional status. British Journal of Nutrition 44(82):165–177.

Ulijaszek SJ, Lourie JA. 1994. Intra- and inter-observer error in anthropometric measurement. In Anthropometry: the Individual and the Population, pp. 30–55, SJ Ulijaszek and CGN Mascie- Taylor, editors. Cambridge: Cambridge University Press

Valente A. 1986. A problema´tica do trabalho de menores em Portugal. Lisboa: Textos de Apoio.

Van der Sluis IM, de Muinck Keizer-Schrama SMPF. 2001. Osteoporosis in Childhood: Bone Density of Children in Health and Disease. Journal of Pediatric Endocrinology and Metabolism 4(7): 817-832.

Van Gerven DP, Hummert JR, Burr DB. 1985. Cortical bone maintenance and geometry of the tibia in prehistoric children from Nubia's Batn el Hajar. American Journal of Physical Anthropology 66:275– 28.

Walker CS, Yapuncich GS, Sridhar S, Cameron N, Churchill SE. 2018. Evaluating morphometric body mass prediction equations with a juvenile human test sample: accuracy and applicability to small-bodied hominins. Journal of Human Evolution 115: 65-77.

Warner SE, Shea JE, Miller SC, Shaw JM. 2006. Adaptations in cortical and trabecular bone in response to mechanical loading with and without weight bearing. Calcified Tissue International 79(6): 395–403.

Watts R. 2015. The long-term impact of developmental stress. Evidence from Later Medieval and Post-Medieval London (AD1117–1853). American Journal of Physical Anthropology 158(4):569-580.

Weaver CM, Wastney M, Spence LA. 2015. Quantitative clinical nutrition approaches to the study of calcium and bone metabolism. Nutrition and bone health. 361-377.

Wood JW, Milner GR, Harpending HC, Weiss KM. 1992. The Osteological Paradox: problems of inferring prehistoric health from skeletal samples. Current Anthropology 33:343-58.

World Health Organization. 2006. WHO child growth standards: length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development

Yim AD, Koningsberg LW, Hwa HL. 2020. Subadult body mass estimation from skeletal remains: Validation for femoral cross-section methods in a contemporary Taiwanese population. In Program of the Annual Meeting of the American Academy of Forensic Sciences, February 2020, Anaheim, CA, USA. Colorado Springs, CO: AAFS, 116.

Appendix A. All Segmentation Algorithms available in ImageJ

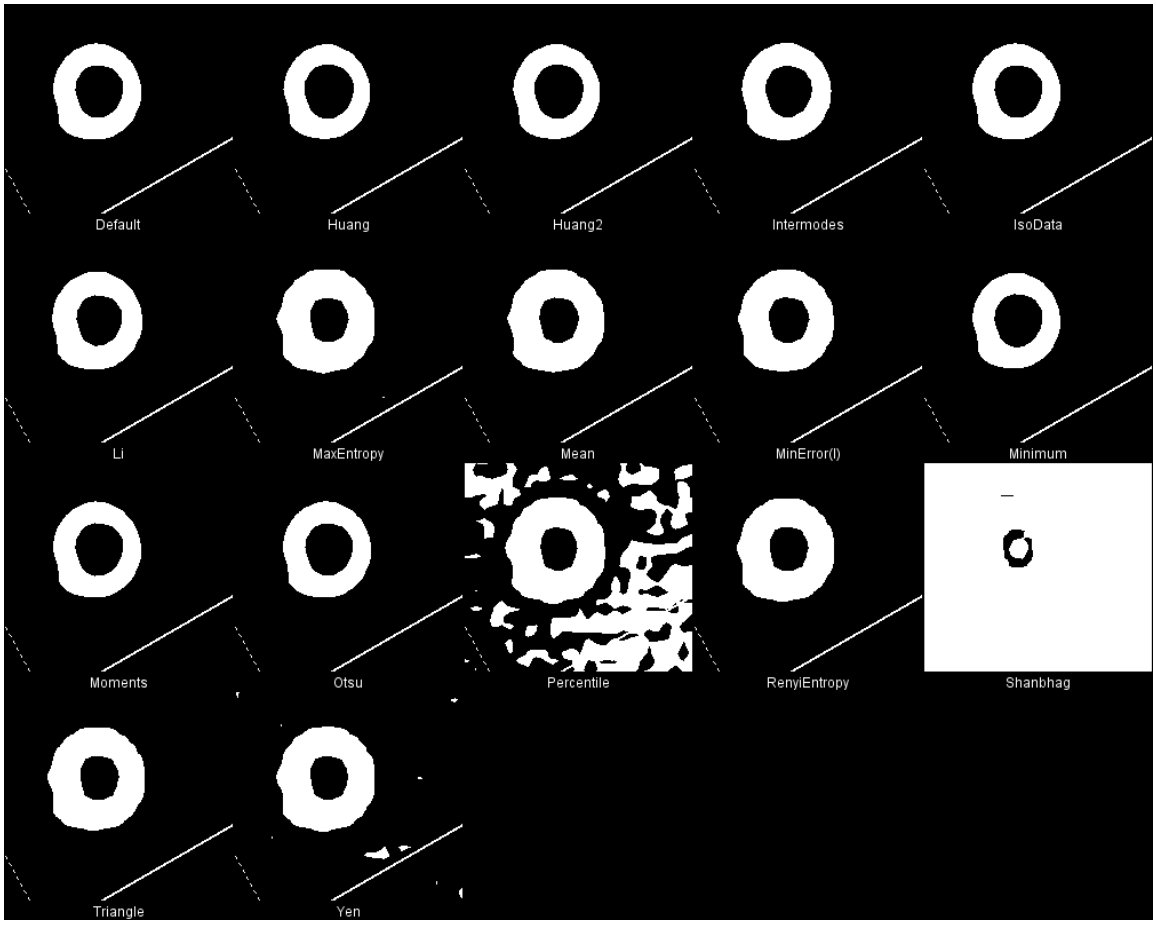


Figure A.1 Assembly of the masks produced from each of the possible thresholding algorithms derived from specimen Case000001 from the wet bone sample (New Mexico).Figure A1. Assembly of the masks produced from each of the possible thresholding algorithms derived from specimen Case000001 from the wet bone sample (New Mexico).

Figure A.2 Overlayed masks of manually thresholded total cross-sectional area and medullary cavity area for Case000001.Overlayed masks of manually thresholded total cross-sectional area and medullary cavity area for Case000001.

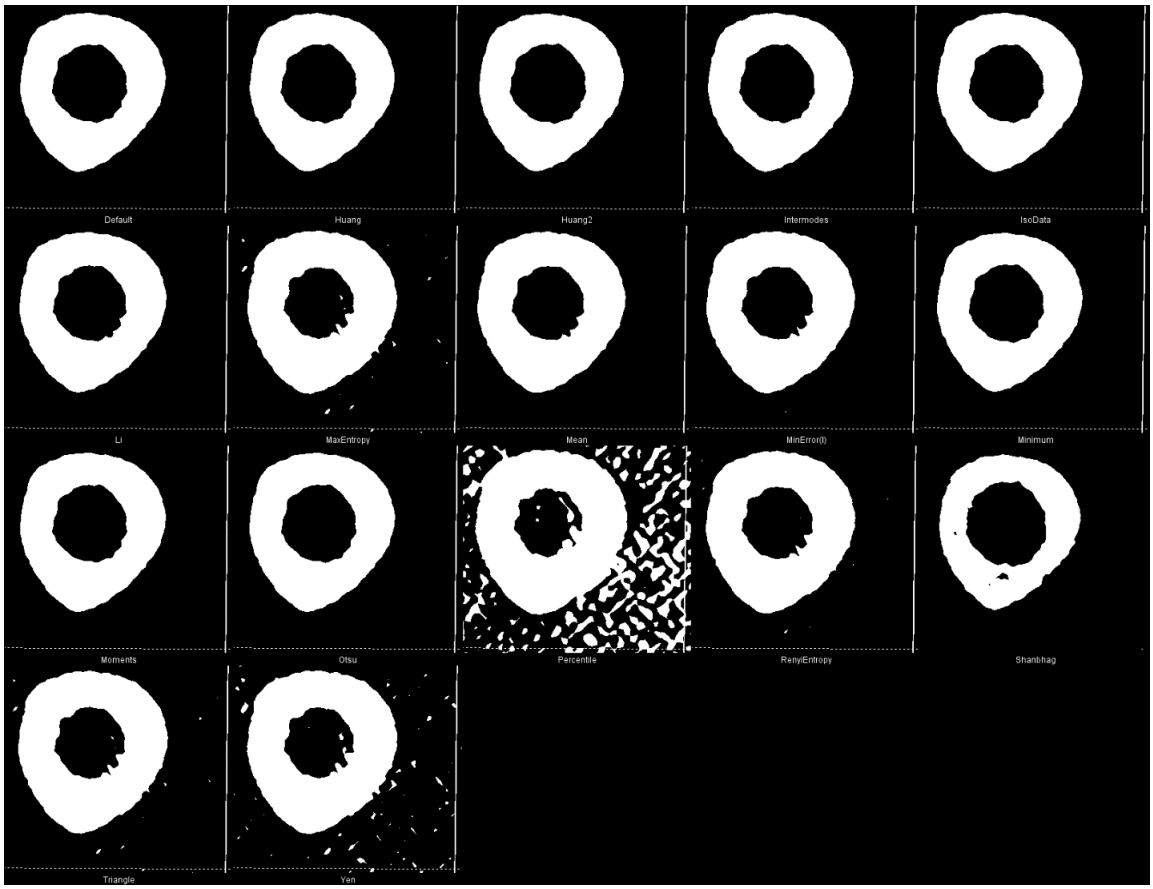


Figure A.3 Assembly of the masks produced from each of the possible thresholding algorithms derived from specimen Case000407 from the wet bone sample (New Mexico).Assembly of the masks produced from each of the possible thresholding algorithms derived from specimen Case000407 from the wet bone sample (New Mexico).

Figure A.4 Overlayed masks of manually thresholded total cross-sectional area and medullary cavity area for Case000407.Overlayed masks of manually thresholded total cross-sectional area and medullary cavity area for Case000407.

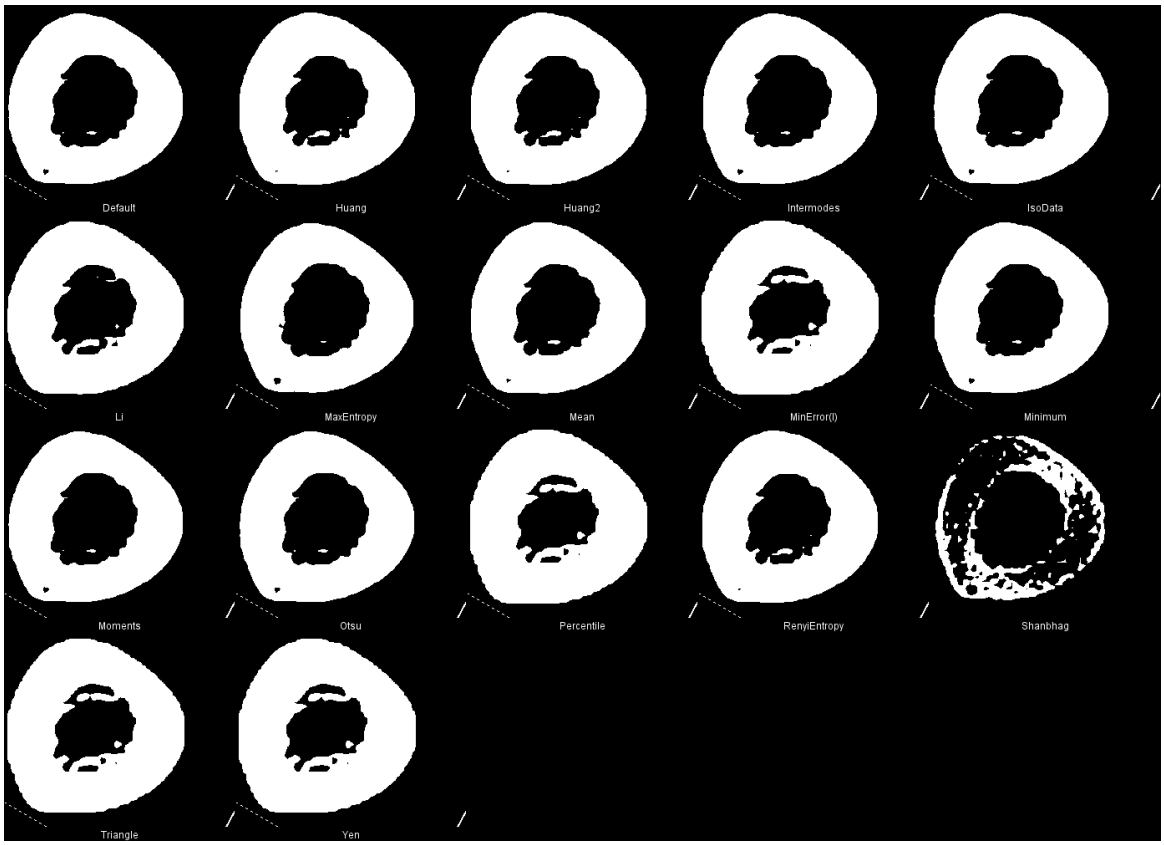


Figure A.5 Assembly of the masks produced from each of the possible thresholding algorithms derived from specimen L_139 from the dry bone sampleAssembly of the masks produced from each of the possible thresholding algorithms derived from specimen L_139 from the dry bone sample (Lisbon).

Figure A.6 Overlayed masks of manually thresholded total cross-sectional area and medullary cavity area for L_139.

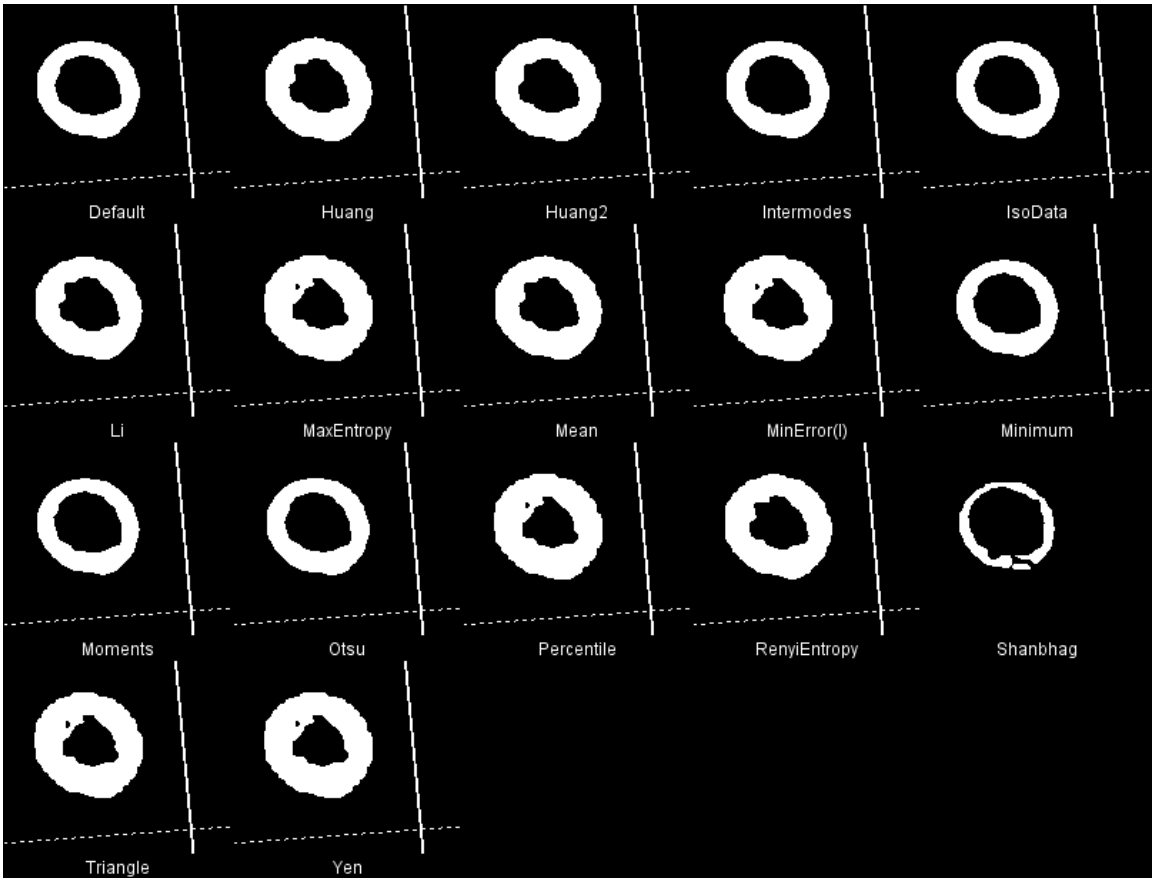


Figure A.7 Assembly of the masks produced from each of the possible thresholding algorithms derived from specimen L_561 from the dry bone sample (Lisbon).

Figure A.8 Overlayed masks of manually thresholded total cross-sectional area and medullary cavity area for L_561.