AN OSTEOMETRIC EVALUATION OF AGE AND SEX DIFFERENCES IN THE LONG BONES OF SOUTH AFRICAN CHILDREN FROM THE WESTERN CAPE

Kyra Elizabeth Stull

Submitted in fulfilment of the requirements for the degree
Doctor of Philosophy
PhD (Anatomy)

in the Faculty of Health Science
Department of Anatomy
University of Pretoria

2013

Supervisors
Professor Ericka N. L'Abbé
Dr. Stephen D. Ousley

Declaration of Original Work

I, Kyra Elizabeth Stull, hereby declare that this thesis entitled,

"An Osteometric Evaluation of Age and Sex Differences in the Long Bones of South African Children from the Western Cape"

Which I herewith submit to the University of Pretoria for the Degree of Doctor of Philosophy in Anatomy, is my own original work and has never been submitted for any academic award to any other tertiary institution for any degree.

_____ _____

KE Stull **Date**

DEDICATED TO
Glenn Eanes and
Glenn, Debra, and Natalie Stull

in loving memory of
Grams
(Roxann Stull, 1932 – 2011)

Acknowledgements

This dissertation was completed with the support and encouragement of many people. My advisors offered direction and supplied advice throughout the entire process and ultimately kept the research on 'the leading (or bleeding) edge'. Prof. L'Abbé believed in me enough to accept me as a doctoral student and continually fought for me when I needed it the most. I could not be more thankful for the opportunities and experiences you provided. Dr. Ousley was the first person to require me to find the original article and be a critic of published material. You taught me so much and ultimately shaped me into a better anthropologist. Thank you.

I have been extremely lucky to find so many people who were kind enough to assist in different aspects of the project. Dr. Trevor Hastie and Stephen Milborrow offered invaluable statistical advice and responded to numerous enquiries regarding FDA and MARS. Prof Rheeder stood behind my protocol and always offered statistical advice. Stef Steiner accommodated numerous data collection trips and provided information regarding the Lodox Statscan. Dr. Thomas Blake at Red Cross War Memorial Children's Hospital and Dr. Lornaj Martin at the Forensic Pathology Service, Salt River permitted use of the images. The medical records staff, and Leah Muller in particular, at Red Cross was of great assistance and made the data collection process entertaining. My aunts, Terri Brennan and Kirsta Stull, and the PEO AQ Chapter of New Jersey demonstrated their constant support by nominating me for the PEO Scholar Award, which I was lucky enough to receive. Lydelle Joubert welcomed me into her Cape Town home and shared the many charms of the city. I am also grateful to Elzabe Brink who provided additional accommodation in Cape Town.

The L'Abbé-Loots family welcomed us to South Africa as if we were family and provided memories we will forever cherish. You will forever hold a special place in our hearts. Marius, I will never cut a tomato without thinking of you and Maia, I will miss our Afrikaans lessons, dress up, market math, and every other moment of your childhood. Jolandie Myburgh not only assisted with data collection, but also offered friendship, friends, and an escape from reality with weekly dinner and movie dates. Thank you for letting me in, teaching me Afrikaans, and sharing your bank account. Desiré Brits was willing to share her time to be my inter-observer error.

While trying to figure out the statistical analyses, Heather Garvin became my roomie once again and shared her home with us. When I was overwhelmed, Natalie Uhl offered words

of encouragement and support. Michael Kenyhercz never lacked intellectually stimulating discussions and a shared love for R. If only future collaborations could happen in Xai-Xai or on a porch in Kruger. Katie Frazee, the beginning of this adventure began with you. Meredith Tise, Erin Chapman and Alexandra Klales have been great friends and offered homes, encouragement and hilarious memories. Dr. Kate Spradley has been a mentor, a colleague and a friend. Thank you for believing in me and offering guidance, advice, and great opportunities. Joe and Erin Andrew helped us survive for the past two and a half years. Thank you for sharing your home, your food, and your friendship; we wouldn't have been able to do it without you. My in-laws, Debi and George Eanes, supported me even though I moved their son far away for months at a time and worked during family visits.

Marlynne Micalizzi was always interested; although it seems so simple, it means so much. During the final stretch, she and Terry Fortune provided much assistance. Poppy (Glenn O. Stull) taught me the three principles to be successful in life, reminded me that if I played with the turkeys I wouldn't soar with the eagles, and above all, he and Grams provided much love and encouragement. From the first I can remember (which may be fairly late in life compared to most, but I am certain this has been since I was born), my sister, Natalie Fay, has been my biggest supporter. You have given me such joy in life and I could never imagine this world without you. Thank you for being you. Mommy and Daddy (Debra and Glenn Stull), the words 'thank you' are not enough. You have encouraged me, loved me and supported me as I chose the path in life I wanted. Our weekly video dates and your constant words of inspiration were undoubtedly the sustenance I needed. I could have never asked for a better family to stand by me and I could not have completed everything I have in my life without your words of wisdom, guidance, and constant love.

Glenn Eanes, you have loved me during the toughest times, offered words of encouragement when I needed them the most, assisted me when I needed help interpreting R code (aka 'having a bad R day'), pulling files, or editing application materials, reminded me to 'get fired up', and cooked dinner, did laundry and paid the bills while I sat in front of the computer. I will never forget how much you gave up so I could have the experience I wanted. You are my best friend and marrying you was the best decision I ever made, even if it is like climbing Mount Everest without oxygen.

Summary

An Osteometric Evaluation of Age and Sex Differences in the Long Bones of South African Children from the Western Cape
KE Stull [1]

Supervisors: Ericka N. L'Abbé[1], Stephen D. Ousley[2]
[1] Department of Anatomy: Section of Physical Anthropology, University of Pretoria, Pretoria, South Africa
[2] Department of Applied Forensic Sciences, Mercyhurst University, Erie, PA USA

Degree: PhD (Anatomy)

The main goal of a forensic anthropological analysis of unidentified human remains is to establish an accurate biological profile. The largest obstacle in the creation or validation of techniques specific for subadults is the lack of large, modern samples. Techniques created for subadults were mainly derived from antiquated North American or European samples and thus inapplicable to a modern South African population as the techniques lack diversity and ignore the secular trends in modern children. This research provides accurate and reliable methods to estimate age and sex of South African subadults aged birth to 12 years from long bone lengths and breadths, as no appropriate techniques exist.

Standard postcraniometric variables (n = 18) were collected from six long bones on 1380 (males = 804, females = 506) Lodox Statscan-generated radiographic images housed at the Forensic Pathology Service, Salt River and the Red Cross War Memorial Children's Hospital in Cape Town, South Africa. Measurement definitions were derived from and/or follow studies in fetal and subadult osteology and longitudinal growth studies. Radiographic images were generated between 2007 and 2012, thus the majority of children (70%) were born after 2000 and thus reflect the modern population.

Because basis splines and multivariate adaptive regression splines (MARS) are nonparametric the 95% prediction intervals associated with each age at death model were calculated with cross-validation. Numerous classification methods were employed namely linear, quadratic, and flexible discriminant analysis, logistic regression, naïve Bayes, and random forests to identify the method that consistently yielded the lowest error rates. Because some of the multivariate subsets demonstrated small sample sizes, the classification accuracies were

bootstrapped to validate results. Both univariate and multivariate models were employed in the age and sex estimation analyses.

Standard errors for the age estimation models were smaller in most of the multivariate models with the exception of the univariate humerus, femur, and tibia diaphyseal lengths. Univariate models provide narrower age estimates at the younger ages but the multivariate models provide narrower age estimates at the older ages. Diaphyseal lengths did not demonstrate any significant sex differences at any age, but diaphyseal breadths demonstrated significant sex differences throughout the majority of the ages. Classification methods utilizing multivariate subsets achieved the highest accuracies, which offer practical applicability in forensic anthropology (81% to 90%). Whereas logistic regression yielded the highest classification accuracies for univariate models, FDA yielded the highest classification accuracies for multivariate models. This study is the first to successfully estimate subadult age and sex using an extensive number of measurements, univariate and multivariate models, and robust statistical analyses. The success of the current study is directly related to the large, modern sample size, which ultimately captured a wider range of human variation than previously collected for subadult diaphyseal dimensions.

Key Words: subadult, Multivariate Adaptive Regression Splines (MARS), basis splines, flexible discriminant analysis, logistic regression, growth, sex estimation, age estimation, diaphysis, anthropology

Table of Contents

Tables

Figures

CHAPTER 1: INTRODUCTION

"Techniques are not ends in themselves, they are only as good as the answers they provide. The answers are only as good as the questions, and the questions are only as good as the insights generated from observations made using the technique" Seeman et al. (1996)

The main goal of a forensic anthropological analysis of unidentified human remains is to establish an accurate biological profile consisting of estimations of sex, age, ancestry and stature as well as a bone trauma analysis. Anthropologists turn towards skeletal collections as the primary source of data when designing and conducting research. However, and in contrast to techniques specific to adults, the largest obstacle in the creation or validation of techniques specific for subadults is the lack of large, modern samples (Weaver, 1988; Franklin, 2010). Only a few documented skeletal collections of subadults exist and available skeletal material is either antiquated or too limited to accurately reflect the range of human variation. The sample directly affects the validity of techniques. A historic sample may result in a technique that is applicable to bioarchaeology but inappropriate in forensic anthropology. For example, if the data originates from the 18th and 19th centuries, the created techniques do not reflect the modern population and ignores secular trends noted in children around the world (Meredith, 1976; Malina, 2004; Hawley et al., 2009; Anholts, 2013).

Subadult age and sex estimation techniques are mostly derived from European or North American populations, rendering them ineffective when applied to South Africans as well as other populations in the world. While children achieve the same general milestones for growth and development, variation in growth is inherent as populations differ in genetic composition and environmental influences (Steyn and İşcan, 1997; Bogin, 1999; Wilson et al., 2011). The largest documented skeletal collection in South Africa, the Raymond A. Dart Collection, contains only 72 individuals younger than twelve years of age. Many of the skeletons do not have all of their associated elements and/or are not considered modern with the dates of birth ranging from 1827 to 1980 (Dayal et al., 2009).

Age estimation is often the primary and only contribution to a biological profile for subadults because many anthropologists believe quantifiable sexually dimorphic differences are absent. Three of the most commonly used methods to estimate subadult age are dental development/eruption; long bone lengths; and the progression of epiphyseal fusion. Because

dental development is under stronger genetic control and is less affected by adverse environmental conditions, dental development is preferred over osseous development to estimate age (Lewis and Garn, 1960; Cardoso, 2007 a; b). In the absence of dentition, age assessment is most frequently derived from the comparison of long bone lengths to classic longitudinal growth studies. The adoption of longitudinal growth studies is likely a compensation for the paucity of subadult skeletons available for research (Stull et al., 2013). Growth studies use the length of the diaphysis to evaluate the growth of a specific child at a certain age and not to evaluate the age of the child with a certain bone length. An anthropologist cannot derive a 95% prediction interval for skeletal elements utilizing the results of published growth studies. However, supplying an error rate with estimations is in partial fulfillment of *Daubert* criteria.

Sex estimation is an essential component to the biological profile and an accurate estimation greatly increases the chance of identification. However, anthropologists regularly state that diagnostic sex differences are not fully established until the completion of adolescence (Scheuer and Black, 2000; Rissech et al., 2008, 2013; López-Costas et al., 2012; Moore, 2013). Yet, the focus of most sex estimation studies are on skeletal elements that are recognized to only be sexually dimorphic after puberty such as the pelvis and mandible (Reynolds, 1945, 1947; Boucher, 1955, 1957; Edward E. Hunt and Gleiser, 1955; Bailit and Hunt, 1964; Gindhart, 1973; Sundick, 1977; Black, 1978; Weaver, 1980; Schutkowski, 1987, 1993; De Vito and Saunders, 1990; Holcomb and Konigsberg, 1995; Molleson et al., 1998; Loth and Henneberg, 2001). However, no medical or anthropological literature indicates that the skeletal elements routinely evaluated (i.e. mandible and pelvis) will express sexual dimorphism prior to puberty. In contrast, endocrinological, biomechanical and medical literature suggest differences exist in the diaphyses prior to puberty (Malina and Johnston, 1967; Frost and Schönau, 2000; Cabo et al., 2012). Until recently, diaphyseal dimensions of children have been relatively ignored as a potential indicator of sexual dimorphism.

Growth studies investigated sex differences but focused only on long bone lengths and not diaphyseal breadths (Maresh, 1970; Gindhart, 1973; Smith and Buschang, 2004). Recent articles that investigated diaphyseal breadths did not discover any significant differences, but the authors adjusted for small sample sizes by collapsing age intervals (i.e. birth to five years and six to ten years). Collapsing ages likely masks any sex-specific trends, especially when size differences are associated with age differences. Two recent studies with relatively large sample

sizes demonstrated the potential of diaphyses in sex estimation with either high rates of correct classification or statistically significant sex differences in skeletal dimensions (Clark et al., 2007; Stull and Godde, 2013).

The lack of accurate and reliable methods available to estimate age and sex in the subadult biological profile is an impediment in forensic anthropology. Thus, the goal of this research is to provide accurate and reliable methods for estimating age and sex of South African subadults from long bone lengths and breadths. Standard postcraniometric measurements were collected from the six long bones (humerus, ulna, radius, femur, tibia, and fibula) on 1380 radiographic images (Lodox Systems Pty (Ltd), Sandton, South Africa), housed in one morgue (Salt River) and one hospital database (Red Cross War Memorial Children's Hospital). Radiographic data from hospitals and morgues provides the necessary means to amass a large modern sample in order to appropriately examine age and sex differences within the subadult skeleton. Integration of radiographic images of children, multiple linear measurements and robust statistical analyses is a novel approach to a longstanding problem within the field of forensic anthropology. Furthermore, the derived age and sex estimation techniques offer country-specific methods for South Africa.

CHAPTER 2: LITERATURE REVIEW

"Dynamic control of growth is endowed by age- and gender-dependent interactions among key genetic, environmental, dietary, socioeconomic, developmental, behavioral, nutritional, metabolic, biochemical, and hormonal factors" Veldhuis et al. (2005:114).

AGE

With the cessation of growth, estimation of age in adult skeletons is based on degenerative patterns appreciated in particular bony interfaces such as the pubic symphysis or auricular surface. In contrast to adult age estimation, subadult age estimation is based on morphologic and metric evaluation of indicators as children grow and develop toward adult size. Because of the predictable changes that occur during growth and development, age estimation methods in subadults generally produce much smaller age ranges than those of adults and consequently are considered more accurate. While the absolute error is smaller in subadult age estimation compared to adult age estimation (because the age range is smaller), the relative error may be comparable (Nawrocki, 2010). Relative error is associated with the range of variation relative to the estimated age. For example, a three year age range for a subadult with a point age estimate of 8 years is narrower than a 20 year age range for an adult with a point age estimate of 50 years; however, three years in relation to the developmental period of subadults may be comparatively as large as 20 years for the degenerative period of adults.

Age estimation is considered to be the most critical and powerful tool in establishing a presumptive identification of an unknown child because few techniques are associated with the estimation of subadult sex and ancestry and even fewer produce consistent results (Scheuer and Black, 2000; Smith, 2007; Saunders, 2008; Franklin, 2010). Chronological age is the amount of time that has passed since an individual was born whereas biological age refers to the physiological state of an individual's biological progression to maturity. Anthropologists estimate chronological age from biological age because a positive correlation exists. However, error in the estimate is introduced because the biological age is highly dependent on the adaptation of each individual to biomechanical stresses and environmental and genetic influences, all of which further introduce a source of intra-population variation (see *Inherent Sources of Variation*, page 23*)*. As a person ages, the accumulation of the effects of extrinsic

factors increase and the precision of age estimates decrease (Jones et al., 1991; Scheuer, 2002; Nawrocki, 2010; Garvin et al., 2012).

When presented with unknown skeletal remains, numerous indicators exist to estimate chronological age, namely dental formation and eruption, appearance and fusion of ossification centers, and long bone lengths. Dental development is less susceptible to environmental stressors than skeletal growth and thus more closely reflects chronological age than dental eruption, long bone lengths or epiphyseal fusion (Lewis and Garn, 1960; Moorrees et al., 1963; Demirjian et al., 1973; Sundick, 1977; Cardoso, 2007 a; b). However, in situations where an individual may be of low socio-economic status (SES), dental formation is even shown to be delayed when compared to chronological age (Lewis and Garn, 1960; Demirjian et al., 1985; Cardoso, 2005). When dental structures are not available for examination, metric analysis of the long bones is often used to estimate age (Sundick, 1978; Hoffman, 1979; Hoppa, 1992; Pfau and Sciulli, 1994; Ubelaker, 1999; Cardoso, 2007 b; Conceição and Cardoso, 2011). However, to date, few studies appropriately address subadult age estimation.

From a survey of practicing forensic anthropologists based in the United States, the longitudinal growth study of Maresh (1970) was the most frequently reported study used to estimate age at death from skeletal remains when dentition was not available. The continued application of Maresh is due to the remarkable results of the Child Research Council (University of Colorado), a longitudinal growth study that commenced in 1935 and continued through the 1960's (Maresh and Deming, 1939; Maresh, 1943, 1955; Hoffman, 1979). The sample contained 254 affluent children of European ancestry born in the early 20th century (Maresh and Deming, 1939; Maresh, 1970; Gindhart, 1973; Hoffman, 1979). Diaphyseal length measurements of the six long bones were collected at 2 months, 4 months and 6 months and then biannually from 1 year through 18 years of age with the aim to establish normal long bone lengths for specific ages (Maresh, 1955, 1970). Data were collected through radiographic images with the acknowledgment of approximately 2% – 3% magnification (Maresh, 1970). Maresh provided mean diaphyseal lengths, with and without the inclusion of the epiphysis (dependent on age), and in some publications the 10th, 50th and 90th percentiles per age. Other methods acknowledged in the survey were adopted from evaluations of growth in historic or prehistoric populations (i.e. Stewart, 1954; Johnston, 1962; Sundick, 1972; Merchant and Ubelaker, 1977). Though neither longitudinal growth studies nor evaluations of historic/prehistoric growth were designed to

predict age-at-death, anthropologists regularly apply them in forensic settings.

Longitudinal growth studies follow the same group of children for a specific duration (i.e. for several years) and take the same set of measurements from each child at regularly scheduled observations (usually 6 months or 1 year). In a cross-sectional study – analogous to data obtained from prehistoric populations – the data consists of each child being measured once and consequently all children are different in each cohort. Longitudinal and cross-sectional data have distinct designs and produce different results. Because cross-sectional data include data from each individual only one time, the data provide population means, variabilities, and ultimately patterns, for each measurement in the sample (Eveleth and Tanner, 1990; Hoppa, 1992; Lampl and Johnston, 1996; Hauspie and Roelants, 2012).

In longitudinal studies, data heaping is a common problem because measurements are collected at the same time for each individual each year. For example, the average age for 5 year olds in the Maresh sample is 5.25 years because data were recorded near their birthday and six months later, while the average age in a cross-sectional study is 5.5 years. Because cross-sectional studies offer a random sample of the population, more variation is evident for each age and less auto-correlation is evident of measurements, an outcome of repeated measurements on the same individuals, which inherently violates the assumption of independent errors (Bock and du Toit, 2004). The variability in diaphyseal measurements is far greater among children randomly sampled from a population rather than a sample of the same children over time (Ousley, 2013). Overall, a cross-sectional study design is better suited for age estimation and a longitudinal study is better designed for evaluating growth rates (Eveleth and Tanner, 1990; Ousley, 2013). Thus, when results from longitudinal studies are inappropriately applied to subadult skeletal material, a bias termed misapplication exists. Furthermore, utilizing results from growth studies usually implies the samples are antiquated and specific to one population group, thus additional biases – termed temporality and demographic composition – are introduced in the age estimation (Stull et al., 2013 a).

BIASES IN AGE ESTIMATION

Misapplication is the term for errors associated with the different aims of research and their subsequent designs, either evaluating growth or estimating age. Longitudinal growth studies generally evaluate the central tendency or normal variation in diaphyseal lengths for known chronological ages so as to understand a pattern in growth. However, forensic anthropologists need to know normal variation in chronological age using known diaphyseal lengths so as to estimate age from skeletal remains. Essentially, most studies present variation in long bones within an age cohort but do not present the potential range of variation in the age assessment (Hoppa, 1992). Although an emphasis within this literature review is placed on the inapplicability of longitudinal growth studies, several forensic and biological anthropology studies approach age estimation in the same manner (Hoffman, 1979; Pfau and Sciulli, 1994; AlQahtani et al., 2010).

The requirement to present error rates or to quantify/qualify prediction estimates is in line with best practices in forensic anthropology (Christensen, 2004; Dirkmaat et al., 2008). As would be expected from a research design evaluating normal growth, the Maresh results were published with mean diaphyseal lengths per age with associated percentiles. While an 80% prediction interval can be estimated from the 10^{th} and 90^{th} percentiles, if they are supplied, forensic anthropologists need to report a 95% prediction interval. The explicit prediction interval is important following the repercussions of the *Daubert* decision which emphasizes the use of error rates (Daubert vs. Merrell Dow Pharmaceuticals, 1993). Longitudinal data are unsuitable for application in forensic anthropology because the point estimation based on interpolation from age-specific means or percentiles of long bone lengths is unknown and probably large (Himes et al., 1977; Scheuer et al., 1980; Stull et al., 2013 a).

The Rostock Manifesto is a theoretical approach named at a workshop entitled ‘‘Mathematical Modeling for Paleodemography: Coming to Consensus’’ held at the Max Planck Institute in June 1999. The approach was developed in an attempt to discuss and provide guidelines for appropriate biostatistical models for age at death. Although the goals of the meeting specified to adult age estimation, the theoretical approaches are also applicable to subadults (Hoppa and Vaupel, 2002). From this meeting, anthropologists recognized that when they conduct forensic or paleodemographic anthropological analyses the interest is in the

probability that the remains belong to an individual of age *a* with a known diaphyseal length *l*, Pr*(a|l)* (Hoppa and Vaupel, 2002). The Pr(*l|a)* is the research design that is followed during most longitudinal growth studies or when there is a known reference population. The probability statements are not equal, thus anthropologists should not utilize the percentiles provided in growth studies to estimate age at death, as the research designs do not have interchangeable goals and outcomes.

Numerous large longitudinal growth studies commenced in the 1930's. The disparity between the population when the study commenced and the modern-day population forms the *temporality* bias. The accumulated data from the longitudinal studies are antiquated and may not be appropriate to compare to modern samples. The developmental trends identified in the growth studies that commenced in the 1900's reflect those of a population that experienced different health standards, nutrition, and environmental factors compared to a 21st century population. Use of references based upon historical samples overlooks secular changes, more specifically the well-documented secular increase in stature that has been observed between modern populations and earlier generations (Meredith, 1976; Steckel, 1994, 2008; Nadler, 1998; Meadows Jantz and Jantz, 1999; Malina, 2004; Heuzé and Cardoso, 2008; Hawley et al., 2009; Anholts, 2013).

The *demographic composition* bias is due to the tendency of researchers conducting growth studies to focus on one discrete population group. Estimation of age in one population that is derived from a different group ignores knowledge of differential environments (i.e. SES) or genetic differences among groups which causes differences in developmental and maturational trends and subsequently generates significant error in age estimations (see *Population Differences*, page 24) (Lampl and Johnston, 1996; Ontell et al., 1996; Bogin, 1999; Schmeling et al., 2000, 2006; Crowder and Austin, 2005). Systematic error is a similar concept to demographic bias that is defined as individuals from a shared environment will exhibit relatively similar patterns when compared to the reference population (Lampl and Johnston, 1996). For example, studies involving living populations routinely demonstrate that stature is highly correlated with nutrition, specifically short stature is common in populations with nutritional deficiencies (Eveleth and Tanner, 1990; Larsen, 1995; Bogin, 1999). Because stature and diaphyseal lengths are strongly correlated, the accuracy of the final age estimate is dependent on the unknown belonging to the same population with similar environmental conditions. Catch-up growth, or an increase in percentile position caused by a height velocity

that is beyond normal limits for age and/or maturity, can make up for subadult deficiencies when the cause of the growth suppression is removed (Prader et al., 1989; Boersma and Wit, 1997). However, a population that suffers from adverse environments throughout the entire growth process may not experience this phenomenon.

Most growth studies evaluate males and females separately, which is a further limitation associated with the *demographic composition bias* because of the lack of sex estimation techniques available prior to adolescence (Johnston, 1962; Ruff, 2003). Therefore, when anthropologists attempt to adapt the results of growth studies, they are inherently introducing error into the estimates. The additional error is because of either the combination of diaphyseal lengths per age and sex or the estimation of age following an estimation of sex. Because anthropologists cannot currently estimate SES, sex, or ancestry in subadult skeletal remains, the accuracy of the final age estimate decreases if the technique was not derived from the same sample to which it is applied (Stull et al., 2013 a).

The accuracy and reliability of age estimation techniques are central critiques of paleodemography and should also be of concern in forensic anthropology (Hoppa and Vaupel, 2002). Descriptive statistics should be used to assess whether the child was developing normally if a method was based on a specific reference population with the purpose to assess the Pr($l|a$) (Tanner, 1986). The biases associated with the adoption of longitudinal growth studies to estimate age at death in forensic anthropology warrants investigation of appropriate statistical techniques and the inclusion of samples reflective of the entire population rather than a specific subset.

SUBADULT AGE ESTIMATION TECHNIQUES

As noted above, studies frequently cited in forensic anthropological analyses are in actuality not designed to estimate age at death of unidentified skeletal remains. Maresh (1970) is the most commonly cited study conducted in North America though the results are based on the Child Research Council longitudinal growth study conducted in Denver, Colorado during the years of 1935 and 1967. The results were presented as normal variations (percentiles) of fat, muscle, and bone lengths in healthy children of one year age intervals from birth to 18 years in numerous publications and throughout multiple decades (Maresh and Deming, 1939; Maresh,

1943, 1955, 1961, 1966). Maresh (1955) acknowledged the group of children that comprised the Child Research Council sample should not be considered representative of all children in the United States.

An additional study that has historically been mentioned in the literature is Merchant and Ubelaker (1977), which is based on the long bone lengths of the protohistoric Arikara. As age at death was unknown for the Arikara, dental formation stages were chosen for all of the available individual teeth and the ages of each individual were averaged for a final age estimation (Merchant and Ubelaker, 1977). Similar to other limitations and sources of error as mentioned in the previous sections, compounded error in the estimates is introduced from the reliance on dental formation to estimate age and then the use of the age estimation to evaluate diaphyseal lengths. The results were used to provide a range of diaphyseal lengths per year age interval. Similar to Maresh (1970), the purpose was not to estimate age in a modern population but rather to evaluate diaphyseal growth in a protohistoric population.

Recently, Rissech et al. (2008), Rissech et al. (2013) and López-Costas et al. (2012) investigated age differences in the humerus, femur, and tibia, respectively. Age distributions of the sample included individuals from birth to young adults (17 – 25 years). The three manuscripts provided univariate and bivariate regression formulae to estimate age from diaphyseal lengths independent of sex (Rissech et al., 2008, 2013; López-Costas et al., 2012). The methodology follows the assumption that the evaluated skeletal remains will show similar growth patterns to other Western European populations. However, their samples are historic, dating from the late eighteenth to late twentieth century, and most likely do not reflect a modern population. Rissech et al., (2008) demonstrated that the diaphyseal length of the femur explained 93% of the variability in age. Analyses were conducted on a pooled male and female sample because no significant differences in diaphyseal lengths were noted between the sexes. Rissech et al. (2013) evaluated age differences in the diaphyseal length and proximal and distal breadths of the humerus. Similar to the femur, both sexes were pooled as no statistical differences were found in the measurements. Univariate models for all three variables were either first or second-degree polynomials with R-squared values of 83% - 86% (Rissech et al., 2013). The authors suggested that the humerus was the preferred element to estimate age-at-death due to the later fusion of the distal epiphysis compared to the other epiphyses. López-Costas et al. (2012)

evaluated the diaphyseal length, distal breadth and proximal breadth of the tibia. Univariate first-degree polynomial models for the tibia explained between 87% and 89% of the variability in age.

Numerous problems exist with the models. First, the presented techniques may be suitable for anthropologists evaluating historical specimens, but not for practicing forensic anthropologists who work with modern samples. Second, the research designs followed a proper intention to estimate age but several errors were introduced. A tendency exists for researchers to provide regression formulae with the associated standard error (SE) and correlations. Although a statistician could use the results (SE and correlation coefficient) to compute a prediction interval for an individual prediction, most anthropologists are not able to and thus the method does not fulfill *Daubert* criteria. Further, the presentation of a SE implies that the variation in diaphyseal dimensions is consistent through the ages. However, as previously stated, variation is known to increase as age increases. An additional limitation of these studies is the small number of individuals per age interval, which resulted in collapsing the data into five year age intervals. The authors consider this collapse as an acceptable form of dealing with small sample sizes but the collapse is biologically unsuitable for growth data as this process ignores the differential growth rates and attainments of each biological phase, which varies depending on environment, population, etc. (Cowgill et al., 2012).

APPLICATION OF LONGITUDINAL DATA TO ESTIMATE AGE

In an attempt to evaluate the accuracy of the above-mentioned studies on a modern population, Stull et al. (2013 a) obtained diaphyseal length measurements from the humerus, radius, femur, and tibia from a modern South African population to estimate age based on results from Maresh (1970) and Merchant and Ubelaker (1977). Each measurement was compared to the two references in order to obtain an age range and a point estimate. The age range was used to determine percent correct, which gauged the accuracy of each ‘method’, and the point estimate was utilized to evaluate bias, which conveyed whether the estimated age consistently over- or underestimated chronological age.

Stull et al. (2013 a) demonstrated that chronological age fell within the estimated age interval only 21.5% of the time following the Maresh standard and 33% of the time following the

Merchant and Ubelaker standard (Tables 2.1 and 2.2). The four long bones in the sample showed consistent trends in bias for both standards; age was overestimated for the first three years and then age was consistently underestimated until 12 years. Student's t-tests, with Bonferroni correction, showed statistically significant differences ($p < 0.001$) between the South African and Maresh North American sample for ages 1, 2, 7 and 11 years for the ulna, radius, humerus, and femur. Without Bonferroni correction, all long bones except for the radius demonstrated statistically significant differences by age 7 ($p < 0.05$).

Comparison of mean long bone lengths provided an evaluation of subadult growth between the samples. The South African sample had larger mean long bone lengths than the Maresh North American sample at birth. From 4 to 7 years, South Africans begin to show smaller dimensions than their North American counterparts and at 12 years of age the South Africans are smaller than the North Americans. The pattern was particularly apparent in the lower limbs where differences were as large as 30mm (Stull et al., 2013 a). Ubelaker (1999) noted, when comparing Indian Knoll – a prehistoric Native American group – to the Arikara, that their growth trajectories were similar until approximately 7 or 8 years of age. At this time, the Indian Knoll population began to lag in growth while the Arikara continued at the same rate. Similar results were found between the Maresh North American sample and modern South African sample. Differential growth rates will result in differential adult statures – either because of genetics or environmental situations. On average, adult South Africans are shorter than North Americans and thus are also expected to have shorter limbs (Steyn and Smith, 2007). Based on the results of the comparisons between modern South Africans and North Americans and between two protohistoric Native American groups, disparities in adult size may commence around 7 years of age. Besides the fact that the two studies (Maresh and Ubelaker) are not applicable methods to estimate age, population differences are apparent between the two samples.

Considering that age estimation is proclaimed to be the most powerful tool for forensic anthropologists when conducting skeletal evaluations on subadults, a review of the literature suggests that age estimation techniques are as unreliable as sex or ancestry estimation. Forensic anthropologists, irrespective of the temporal inconsistencies or inappropriateness of the data, routinely cite longitudinal growth studies when conducting age estimations. Recent publications have approached age estimation appropriately, however, the available cross-sectional data remain antiquated and unsuitable for comparison to modern children.

Table 2.1 – Percent correct for the modern South African data using Maresh (1970).

	N	Percent Correct
Humerus	101	37%
Radius	104	17%
Femur	100	18%
Tibia	85	14%
Total		**21.50%**

Table 2.2 – Percent Correct using the modern South African data using Ubelaker (1999).

	N	Percent Correct
Humerus	93	54%
Radius	86	31%
Femur	76	32%
Tibia	89	21%
Total		**33%**

SEX

Sex estimation of subadult skeletal remains is consistently recognized as inaccurate. However, the bulk of published methods focus on shape and morphological differences of sexually dimorphic elements in adults, such as the ilium and mandible, though minimal literature is available to suggest that differences exist in subadult skeletal structures. In contrast, a plethora of medical, clinical, and biomechanical literature documents skeletal differences between males and females throughout growth that should hypothetically create different skeletal structures for males and females and includes hormonal differences, anthropometric and physiological differences and sex-specific biomechanical adaptations.

HORMONAL INFLUENCES

Hormones affect the developing skeleton in a systemic nature that directly affects human body size, proportions and composition (Cabo et al., 2012). Following sex differentiation in the fetus, discrete sex characteristics are evident, quantifiable and continue through childhood. Sex and hormonal differences begin in the human embryo six to eight weeks after conception when the Sry gene of the Y chromosome commences testicular differentiation (Tanner, 1989; Knickmeyer and Baron-Cohen, 2006). Testosterone secretion is high in the male fetus between weeks 10 and 20 because of further development in the Leydig cells (Riggs et al., 2002; Saunders, 2008). In contrast, and despite low levels of estrogen being present in the fetal system, the ovaries in a female fetus are inactive (Smail et al., 1981; Grumbach et al., 2003). At week twelve, the process of sex differentiation is largely complete. However, the critically sensitive period, when environmental influences can modify tissue development, may last up to twenty-four weeks *in utero* (Knickmeyer and Baron-Cohen, 2006). The time period of eight through 24 weeks *in utero* is considered the most important in terms of sexual differentiation (Knickmeyer and Baron-Cohen, 2006).

Within the first day after birth, male testosterone levels rapidly decrease only to increase again within a week (Quigley, 2002; Riggs et al., 2002; Fechner, 2003; Knickmeyer and Baron-Cohen, 2006). Testosterone levels remain high for the first year of life with a peak between the first and third month of life with median levels equivalent to the levels associated with adolescence (Quigley, 2002; Riggs et al., 2002; Fechner, 2003; Aksglaede et al., 2006; Knickmeyer and Baron-Cohen, 2006; Saunders, 2008). Between four to six months, the testosterone levels decline and around nine months of age, testosterone levels are equivalent to typical prepubertal levels (Quigley, 2002). This phase is referred to as the neonatal surge (Riggs et al., 2002; Knickmeyer and Baron-Cohen, 2006). Although the function of the neonatal surge is not fully understood in humans, this surge is most likely related to the preparation for future development, growth and reproduction. For example, interference with the neonatal surge can later disrupt testicular function in puberty and result in reduced muscle mass, shorter final stature, and smaller bone size (Mann et al., 1989; Arfai et al., 2002; Quigley, 2002; Veldhuis et al., 2005). Because sex steroids catalyze changes during puberty, one could presume that an equivalent level of hormones as seen in the neonatal hormonal surge may be responsible for

increases in fat and musculoskeletal development during infancy and adolescence (Bogin, 1999; Arfai et al., 2002).

Female infants also demonstrate an immediate postnatal activation of the hypothalamic-pituitary-gonadal axis (i.e. endocrine system), but the dynamics are more complex and hormonal levels heterogeneous (Forest et al., 1973; Quigley, 2002; Fechner, 2003). While Luteinizing hormone (LH) is the dominant hormone in males, follicle-stimulating hormone (FSH) is the predominant hormone in females. During the first four months of life, a rapid increase in ovarian follicular maturation, which is controlled by FSH, occurs and results in an increased level of estradiol production between two and four months and continues through 24 months (Quigley, 2002). The predominant FSH in the first two years of life projects future development for females. Prepubertal males and females have low levels of sex steroids but the earlier rise of FSH than LH is again documented during the onset of puberty (Quigley, 2002; Aksglaede et al., 2006).

Sex steroids have specific functions at the organ, tissue, and cellular levels within the skeletal system. Both testosterone and estrogen affect osteoblasts and bone formation. Yet, a disparity in the distribution and accumulation of androgen and estrogen receptors causes differential tissue sensitivity (Tanner, 1989); therefore, the effects of the sex steroids on osteoblasts differ in males and females (Riggs et al., 2002). While estrogen impacts endosteal and trabecular bone growth as well as bone turnover, testosterone opposes it with periosteal apposition (Riggs et al., 2002; Veldhuis et al., 2005; Högler et al., 2008; Cabo et al., 2012). Additional responsibilities of estrogen include regulating bone resorption, promoting epiphyseal fusion, and increasing tensile bone strength. Larger bone sizes have been documented in prepubertal males compared to prepubertal females and attributed to a more rapid periosteal apposition relative to interstitial growth in boys than girls prior to puberty (Clark et al., 2007). Besides periosteal apposition – the main cause for physiological sex differences – testosterone indirectly affects the skeleton by inhibiting fat accumulation and inducing muscle development (Cabo et al., 2012).

Concentrations of estrogen and testosterone gradually increase from mid-childhood until puberty when the larger magnitude of effect is observable (Garnett et al., 2004).

ANTHROPOMETRIC AND PHYSIOLOGICAL SEX DIFFERENCES

Quantitative anthropometric and physiological differences between males and females originate in the prenatal period and continue throughout growth and development. Numerous studies documented lower average birth weights, lengths, and head circumferences in female neonates compared to male neonates (Garn and Keating, 1980; Largo et al., 1980; Tanner, 1989; Thomas et al., 2000; Malina et al., 2004; Veldhuis et al., 2005; Clark et al., 2007). Furthermore, sex differences consistently increase as age increases (Thomas et al., 2000). Divergence of male and female weight occurs at approximately 24 weeks *in utero* and statistically significant differences are apparent at birth (Veldhuis et al., 2005); however, negligible differences are noted in the weight of males and females throughout childhood (Tanner, 1989; Clark et al., 2007). Specific to South Africa, a large sample of black, white, coloured, and Indian children demonstrated significant differences between the sexes in mean heights and weights from birth through 2 years of age (Cameron et al., 1998). This trend continues through growth with males being consistently heavier and taller than females until approximately 9 years of age, with the onset of puberty in females (Högler et al., 2008). In total, sex differences noted in adult linear body dimensions are stated to emerge during adolescence (Hauspie and Roelants, 2012).

Besides anthropometric differences, sex-related disparities in body composition appear as early as the 15th week *in utero* and continue throughout puberty. One of the many roles of testosterone is to inhibit fat accumulation, thus males demonstrate greater muscle mass and more fat-free mass than females (Garnett et al., 2004; Veldhuis et al., 2005). The exact age that sex differences in the relative composition of fat-free mass are noted is unknown. However, researchers have reported differences as early as 3 years of age and statistically significant differences prior to the onset of puberty (Poissonnet et al., 1984; Prader et al., 1989; Arfai et al., 2002; Garnett et al., 2004; Malina et al., 2004; Clark et al., 2007). The sex differences in body composition, a result of differences in sex steroids, is probably responsible for the sex differences in muscle mass, which is apparent both in relative and absolute terms, in adolescence (Malina, 1974, 2004). The differences between muscle and bone mass are much greater between the sexes than the average weight differences (Tanner et al., 1981; Cabo et al., 2012).

Sex differences in maturity are also apparent within the fetal period and continue throughout development. For example, female fetuses are skeletally more advanced than male

fetuses, ranging from one and a half to three weeks in the third trimester to six weeks at birth (Tanner, 1989; Veldhuis et al., 2005). The trend continues until puberty, when females are approximately 2 years more advanced than males (Tanner, 1989; Hauspie and Roelants, 2012). Sex differences have also been apparent in growth velocity after the 36th week *in utero* (Thomas et al., 2000; Veldhuis et al., 2005). Growth trajectories, along with shape, size, and other sex differences, exist for prepubertal males and females (Tanner, 1989; Humphrey, 1998).

SEX-SPECIFIC BIOMECHANICAL DIFFERENCES DURING GROWTH

The mechanostat is the combination of all mechanical and non-mechanical (i.e. hormones) factors that allow for a healthy load-bearing bone to satisfy intended functions and demonstrate bone's adaptive capabilities (Frost, 1988). The mechanostat theory postulates that during growth an increase in muscle forces is positively correlated to dimensions, size, and strength of bone. Reduced muscle development (i.e. an unloaded bone) is negatively correlated to bone dimensions, size, and strength; essentially, requirements associated with the non-skeletal tissues controls the growth of the skeletal tissues (Moss, 1973; Frost, 1988; Schoenau et al., 2002; Schoenau and Fricke, 2008). The intrinsic relationship between muscle and bone is reflected in a linear relationship between the cross sectional areas of muscle and bone (Schoenau et al., 2002). This is in accordance with biomechanical models that state when a bone has force exerted upon it, the bone will respond with an increase in cross-sectional area. Thus in an indirect osteogenic effect, larger muscles exert higher forces on bones which results in larger bones (Vicente Rodríguez, 2006). Since the compressive forces are associated with muscle volume, the increase in transverse diameter is also apparent in non-weight bearing bones such as the upper limb. The over-proportionate increase in cortical thickness, due to the effects of testosterone, also contributes to overall thicker bones in males (Arden and Spector, 1997; Riggs et al., 2002).

Muscle contractions, rather than body weight, govern the postnatal structural adaptation to loading in the skeleton such that an increase in muscle strength allows for greater force exerted on muscle origin sites which ultimately stimulates growth (Weaver, 1980 b; Schutkowski, 1993; Schönau, 1998; Daly et al., 2004; Vicente Rodríguez, 2006; Högler et al.,

2008). For example, growth spurts in humeral cortices corresponds to a contemporaneous peak in muscle mass (Tanner et al., 1981). Therefore, the greater muscle mass of males directly affects skeletal growth due to increased compressive forces exerted on the diaphyses (Schönau, 1998; Högler et al., 2008; Cabo et al., 2012), specifically an increase in cortical width from differential distributions of sex steroids (Riggs et al., 2002). The theory is based on the knowledge that skeletal muscles develop prior to bone mass. Therefore, as muscles increase in size and strength – which occurs continuously throughout childhood – the skeletal structures adapt to the increased load with added mass, size, and strength (Schönau, 1998; Daly et al., 2004). However, the greater muscle mass of males was shown to account for only 12% to 16% of the variance in bone size, which indicates that factors besides muscle size also contribute to the bones response to loading (Daly et al., 2004). Previous research demonstrates quantitative sex differences in muscle and cortical thicknesses of humerii and tibiae of subadult males and females such that males have larger dimensions throughout the ages of 6 to 16 years (Johnston and Malina, 1966; Malina and Johnston, 1967; Rogol et al., 2000; Arfai et al., 2002; Wells, 2007). Similar findings were noted in the second metacarpal (Smithgall et al., 1966). Thus, the change in bone geometry throughout growth is consistent with a sex-specific biomechanical adaptation.

SUBADULT SEX ESTIMATION TECHNIQUES WITHIN ANTHROPOLOGY

A multitude of subadult sex estimation studies present with low classification accuracies and therefore, researchers have suggested that sexual dimorphism may not present with the same magnitude as noted in adults. However, the most popular approach for subadult sex estimation techniques has been to evaluate differences in the pelvis and the mandible, presumably because the elements are sexually dimorphic in adults. Less emphasis has been placed on osteometric analyses. Rather than assuming the absence of sexual dimorphism, the research design and sampling of previous studies needs to be examined.

MORPHOLOGICAL APPROACH

In 1876, Fehling was the first to address sex differences in the shape of the fetal and neonate pelvis and many anthropological researchers have followed this line of enquiry (Reynolds, 1947; Boucher, 1955, 1957; Weaver, 1980 a; Schutkowski, 1993; Holcomb and Konigsberg, 1995; Molleson et al., 1998; Loth and Henneberg, 2001; Sutter, 2003; Franklin and Cardini, 2007; Franklin et al., 2007; Vlak et al., 2008; Wilson et al., 2011). Although sex differences are recognized in fetal and neonate skeletal structures with resultant moderate classification accuracies, an inconsistency in the observed classification accuracies is noted when the techniques are tested on separate samples. Using the greater sciatic notch and lengths of the ilium and femur, Fazekas and Kósa (1978) achieved 80% correct classification but using the same material Schutkowski (1987) achieved only 70% correct classification.

Schutkowski (1993) more recently identified sex differences in the pelvis on a sample that ranged from birth to 11 years. The pelvic features, inclusive of the angle and depth of the greater sciatic notch, arch criterion, and curvature of the iliac crest, demonstrated sexually dimorphic differences with classification accuracies between 70 and 95% (Schutkowski, 1993). Numerous authors explored the Schutkowski (1993) methodology on various population groups. Although Sutter (2003) obtained moderately high accuracies (79 – 81%) on a prehistoric Chilean population, Vlak et al. (2008) yielded low correct classifications on a Portuguese sample and was unable to confidently identify sex differences, especially in age groups younger than 11 years. Comparable accuracy levels were identified only in the oldest age interval from 11 to 15 years.

An attempt to quantify morphological features has also been made in an effort to increase objectivity and control for increases in size as age increases (Holcomb and Konigsberg, 1995; Wilson et al., 2011). Although sexual dimorphism in the fetal pelvis was noted, specifically in the shape of the sciatic notch, levels of dimorphism were not comparable to those in adults. Furthermore, the practicality in forensic application was limited to a large overlap between males and females (Holcomb and Konigsberg, 1995). When studies included older individuals, acceptable classification accuracies were obtained only for age intervals from 11 to 14.99 years (Wilson et al., 2011). Sexually dimorphic differences for individuals between birth and 5 years were also identified in mandibular morphology, namely the protrusion of the chin, shape of the dental arcade, and gonial eversion (Schutkowski, 1993). Loth and Henneberg (2001) presented a

new methodology based on observations of the mandibular body shape and symphyseal base on a sample of 62 individuals, aged from birth to 19 years. Classification accuracies in the original article were considered acceptable (81%); however, similar to the studies focused on pelvic morphology, tests conducted on separate samples resulted in low accuracies that ranged from 58 to 64% (Scheuer, 2002; Galdames et al., 2008). Using geometric morphometrics, Franklin et al. (2007) were unable to repeat the distinct male and female shape differences that Loth and Henneberg (2001) identified on the mandible.

One common thread in all the above-mentioned research studies is that the original researchers seem to achieve higher accuracies than subsequent researchers, an effect which may be largely attributed to population differences in growth. The second similarity throughout all of the literature is that investigated areas are regions on the skeleton known to change during the pubertal hormonal surge. Females prepare for reproductive capabilities during this pubertal surge and, as one might expect, this is the primary reason for the pelvis to be the most sexually dimorphic element in the adult skeleton. Cardoso and Saunders (2008: 28) note that:

> *"because the innominate shows such a late developmental pattern and late attainment of adult size, sexually dimorphic features may not be readily recognizable before puberty. Sexual dimorphism in the greater sciatic notch-auricular surface area, and particularly the composite arch, is probably an expression of that developmental trajectory."*

The skeletal elements that exhibit sexual dimorphism in adult skeletons do not generally attain adult levels until adolescence. Furthermore, no medical literature is available to assume morphological differences in the pre-pubertal skeleton would exist. In contrast, a multitude of literature, from numerous fields, demonstrates sexually dimorphic differences in prepubertal individuals that may be evident through metric analysis of the long bones.

OSTEOMETRIC APPROACH

Sex differences in subadult skeletal remains, evaluated through osteometric analyses, have not been commonly cited in the anthropological literature in comparison to morphological analyses. While some researchers have previously evaluated bone breadths, the dimensions were

not evaluated in terms of sex estimation but rather in an attempt to understand sexually dimorphic differences in growth rate and duration or purely for understanding normal variation in growth (Maresh, 1961, 1966; Malina and Johnston, 1967). Malina and Johnston (1967) did note that males prior to the onset of adolescence displayed greater breadths in the tibia and humerii than females of the equivalent age. Similarly, Humphrey (1998) identified sex differences in long bone diameters between birth and adolescence while sex differences in long bone lengths were not observed until adolescence. Over 60% of the sexual dimorphism noted in the long bone diameters could be attributed to differential growth rates, thus "about half of the variation in sexual dimorphism is related to the age of attainment of 90% adult size" (Humphrey, 1998: 64-66). Essentially, earlier growing parts of the skeleton tend to display less sexual dimorphism than later growing elements (Schultz, 1962).

Choi and Trotter's (1970) employed a metric approach with the intention to estimate sex from the weight and length of black and white North American long bones of 114 fetuses aged 16 to 44 weeks. Statistically significant sex differences were revealed in ratios of lengths and weights of the diaphyses. The provided discriminant functions for the humerus, radius, femur and tibia correctly classified the sex of 72% of the fetal remains within their study (Choi and Trotter, 1970). Following the hypothesis of greater muscle mass in males and in conjunction with the neonatal hormonal surge, Stull and Godde (2013) utilized length and breadth measurements of the humerus and femur of infants (n=85) from birth to one year with discriminant function analysis. Discriminant function analyses elucidated sexually dimorphic differences between the sexes with correct classification ranging from 89 – 97% using the femur midshaft and humerus distal breadth. The study provided a successful model for future analyses of subadult long bones.

Rissech et al. (2013), Rissech et al. (2008) and López-Costas et al. (2012) evaluated sex differences in long bone lengths and breadths from the same samples which the authors created age estimation formulae. Females were larger than males from birth to four years, from five years onwards males were larger than females for the diaphyseal lengths of the humerus, femur, and tibia as well as the proximal and distal breadths of the humerus and proximal breadth of the tibia. The distal breadth of the tibia was the only measurement where males were larger for the entire period of birth to 25 years (López-Costas et al., 2012); however the differences were not significant.

Longitudinal growth studies also explored and identified sex differences in diaphyseal lengths. A subset of individuals of the Fels longitudinal study sample yielded remarkable sex differences for the radius; radius lengths were significantly greater in males from 1 month to 10 years but between the ages of 10.5 years and 13 years boys and girls were comparable (Gindhart, 1973). Due to the earlier adolescent growth spurt of females, this sexually dimorphic pattern is expected. The outcomes of the Fels Longitudinal Study corroborate findings from the Child Research Council sample which noted that males presented with larger radius and ulna diaphyseal lengths from infancy to 11 years and then again from 13 to 18 years (Maresh, 1970; Smith and Buschang, 2004). Males consistently presented with larger diaphyseal lengths than females until adolescence, after which the trend reversed because females experienced a younger age for the onset of adolescence. Male diaphyseal lengths were once again larger following the end of adolescence and the completion of the male growth spurt (Maresh, 1970). The femur, tibia, and humerus did not demonstrate any statistically significant differences in prepubertal periods, but as would be expected, males were always larger following puberty (Maresh, 1970). The tibia is the only element that exhibits a discrepancy between the two sets of longitudinal growth studies. While no differences in sexual dimorphism were noted in the Child Research Council sample, male tibia were always larger than female tibia in the Fels sample (Gindhart, 1973). Although not significant, Smith and Buschang (2004) demonstrated that the tibia of the females from the Child Research Council sample began to diverge from the tibia of males around 7 years of age. A reanalysis on a subset of the Child Research Council data revealed significant sex differences in long bone lengths at 16 years but no differences at 10 years (Smith and Buschang, 2005).

From the presented literature, a reasonable assumption is that size and shape differences can be quantified in long bones prior to puberty. Medical, clinical, and anthropological literature demonstrate differences exist in body dimensions both *in utero* and through growth as well as differences in physiology which yield distinct skeletal structures due to differential distribution of estrogen and androgen receptors. Furthermore, anthropological literature has identified that sexually dimorphic differences are apparent in the long bones of subadults. Although most of the presented studies have not been validated with separate samples, the original results of osteometric analyses are more consistent than those for morphological analyses. And more so, they are yielding significant differences in the younger ages, while morphological analyses

demonstrate significant differences in the older ages. Sex estimation from diaphyseal dimensions is supported by anthropological, physiological and biomechanical literature, which has previously identified differences in the skeletal structures of prepubertal males and females. Thus, the notion that diaphyseal dimensions should express sexual dimorphism has as strong theoretical basis; a principle that adheres to acceptable measurement theory models (Houle et al., 2011).

INHERENT SOURCES OF VARIATION

Populations vary in shape, size, maturation rate, and growth trajectories because of genetic differences as well as influences from the environment such as nutrition and lifestyle (Steyn and İşcan, 1997; Ulijaszek, 2001; Lewis et al., 2002; Sun et al., 2002; Veldhuis et al., 2005). Numerous studies demonstrate the necessity for population specificity through a decrease in accuracy when the original technique is applied to other population groups (Calcagno, 1981; Steyn and İşcan, 1997; Lewis et al., 2002; Spradley et al., 2008; Wilson et al., 2008). Intrinsic and extrinsic factors are fundamentally interrelated and separating the effects of each factor is complex and practically impossible in humans. However, genetic influences can be evaluated by assessing children from different populations but with similar environmental influences or environmental influences can be investigated by assessing children from within a population but with varying environmental influences (Eveleth, 1978; Stinson, 1985).

Further complications are because of differing magnitudes of either genetic or environment influences throughout growth. In childhood, environmental factors and socio-economic status (SES) are shown to have a significantly greater effect on growth and development than genetics (Garn et al., 1976; Garn and Bailey, 1978; Molteno et al., 1991). For example, adopted child-parent pairs have high correlations when compared to biological-child parent pairs in terms of stature, weight, and fat-fold thickness up to adolescence. If body dimensions were purely under genetic control then no correlation, or at least a weak correlation, would be expected between these variables and the non-biological parents (Garn et al., 1976). Therefore, growth from birth to age 7 is often described as a consequence of developmental plasticity in which environmental influences can directly affect growth potential with irreversible

modifications (Bogin and Loucky, 1997). Modern human population variation in stature is believed to be a result of nutrition during infancy and early childhood, hence the plastic ability of the skeletal structures to adapt to the environment (Eveleth and Tanner, 1990). Because the insulin dependent growth of infancy is replaced with a growth-hormone-regulated growth in childhood (and older), nutritional deficiencies in childhood and puberty slow maturation but do not affect final stature (Karlberg, 1989; Kuzawa and Bragg, 2012),

POPULATION DIFFERENCES

General size differences, specifically stature, are frequently and easily assessed in different populations and some disparities can be directly attributed to ancestry though stature is also commonly recognized as a sensitive indicator of the quality of life (Eveleth and Tanner, 1990; Steckel, 1994, 2008; Bogin et al., 2002; Ha et al., 2003). On average, modern adult South Africans are shorter than North Americans (Steyn and Smith, 2007). Stature differences between the two populations are apparent at twelve years of age when American children are already taller than South African children (Preston and Chertkow, 1986). Walker and Walker (1977) demonstrated that 20th century black South African children are significantly shorter in height and lighter in weight than the reference standards of London, Harvard-Iowa, and contemporary white children in Johannesburg. Although a strong genetic predisposition exists for stature, the degree to which stature is controlled by genetics is unknown as poor adverse conditions can prevent attainment of predisposed adult height (Hoppa, 1992). An example of the effects of population differences on age estimation can be taken from Hoffman (1979), who noted differences in diaphyseal lengths ranging from 40 mm to 60 mm between Eskimos (Stewart, 1976) and white children from the Child Research Council. Use of the Stewart (1976) growth charts could potentially overestimate age as much as 2 to 3 years if applied to a mid 20th century white North American sample (Hoffman, 1979). Eskimos are recognized as exhibiting short stature likely because of genetic variation and/or adaptation to the environment (Heller et al., 1967). Thus, the application of a growth chart derived from an Eskimo population and applied to a population with a different mean stature is to result in errors in the age estimation.

Body proportions differ between populations with disparities recognizable during childhood (Hamill et al., 1973; Eveleth, 1978; Eveleth and Tanner, 1990; Bogin, 1999). Cowgill et al. (2012) noted a strong hereditary component to intralimb indices during growth and previous studies demonstrated that individuals of the same stature from different ethnic groups, with extrinsic factors controlled, displayed proportional differences (Eveleth, 1978). Hamill et al. (1973) found that American white children have longer trunks while American black children have longer legs. Furthermore, many black groups (males and females) from African countries exhibit relatively longer legs compared to white groups from European countries (Eveleth, 1978). Bogin et al. (2002) evaluated leg lengths of Africans, Asians, Australians, and Europeans and noted that Africans have the longest legs and Asians the shortest legs. Differential proportions ultimately affect age estimates based on a single diaphyseal length.

Black populations, irrespective of continent, are considered to be genetically predisposed to mature earlier and grow more rapidly than white populations; yet, poor environmental conditions can delay growth and development (Garn and Bailey, 1978; Singer and Kimura, 1981; Cameron et al., 1993; Bogin, 1999). Differences at birth are present in North American black and white infants but by 1 or 2 years of age trends reverse and American black children present with similar or greater heights and weights than American white children (Eveleth, 1978). In the United States, 8 year old non-Hispanic black girls demonstrate significant differences in maturation when compared to non-Hispanic white and Mexican American girls; the same trend was true for non-Hispanic black boys (Sun et al., 2002). In terms of epiphyseal fusion, Crowder and Austin (2005) identified significant differences in the age of complete fusion of the distal tibia and fibula such that black and Mexican-American males complete fusion by 13 years while white American males did not complete fusion until 15 years. The above-mentioned examples show techniques derived from one population are not always applicable to other populations.

ENVIRONMENTAL INFLUENCES

The environment comprises a wide variety of factors including the natural environment (high altitude and average temperatures), anthropogenics (pollutants, metals, radiation), and, the focus for the remaining section, social and/or political processes (advantaged or disadvantaged

position in society). The position in society directly affects the caloric intake, nutrition and access to adequate healthcare, amongst other variables (Schell et al., 2012). Adverse environmental conditions generate disparities between groups and the distinct combination of environmental influences produces a group of children that are relatively similar to one another and unique to other children. Disparities among groups based on environmental differences can cause systematic error in age at death estimates as growth trajectories and proportions are affected (Eveleth and Tanner, 1990; Lampl and Johnston, 1996). Furthermore, nutritional deficiencies delay the onset of pubertal growth and result in a protracted growth velocity that is slower and distributed across a longer duration (Kuzawa and Bragg, 2012). A consistently documented trend is higher SES children are heavier and taller than lower SES children (Hoppa, 1992). Specifically in South Africa, lower SES coloured children exhibit, on average, 5% shorter statures and 20% to 25% lighter weights than coloured children of higher SES in Cape Town (Henneberg and Louw, 1998).

Most radiographic data for subadults stems from systematic, non-repeatable, cross-sectional, or mixed-longitudinal radiological growth studies comprised of white and middle-class children (Maresh, 1970; Scheuer, 2002). Because the standards are derived from healthy populations, the estimated ages based on these samples may be significantly underestimating the true chronological age of an individual living in an adverse environment. Lampl and Johnston (1996) evaluated differences between skeletal age and chronological age when the sample was not middle-class and white. Skeletal and dental age was scored on a sample of Mexican children, aged birth to seven that lived under environmental stress inclusive of high infectious disease rates and moderate malnutrition. In terms of skeletal age, only 13% of the children were estimated to be the same chronological age and 22% of the sample demonstrated delayed skeletal age with differences up to 4 years between chronological age and the estimated skeletal age. Although dental age is consistently identified as being under more genetic control than skeletal age, results of the estimated dental age and true chronological age reveal an error of 3 years for 94% of the sample.

METHODOLOGICAL ERRORS

The most common error in methodology is related to sampling. Specifically for subadults, skeletal collections generally do not contain enough individuals per age, which results in an overall small sample size that is unevenly distributed across the ages (Stull et al., 2013 b). Furthermore, large datasets quickly become too small to properly test hypotheses when subset into sex and age. For example, Franklin et al. (2007) investigated sex differences in the mandible of 96 individuals from 1 to 17 years of age. Although the sample size may appear acceptable, if evenly distributed, approximately 5 individuals are in each yearly subset, independent of sex. Because of the small sample sizes, the majority of research regarding subadult age or sex estimation tends to collapse a number of ages as a means to create larger subsets. Without an appropriately large number of individuals per each investigated age, distinct differences cannot be thoroughly and accurately identified. Considering the rapid growth during the first few, years of life, each year should be investigated separately, or in small age-at-death ranges that reflect subdivisions in growth (Cowgill et al., 2012; Stull and Godde, 2013). Large, more evenly distributed datasets that exhibit substantial numbers of individuals of both sexes and all population groups for each age interval are needed for *a priori* hypothesis testing (Nawrocki, 2010).

The use of skeletal collections rather than living children to devise age at death estimations introduces error associated with the osteological paradox. Subadults in a skeletal collection are individuals that failed to survive, potentially displaying differential growth rates. The variability in the growth of skeletal collections may or may not reflect growth processes shared by their contemporaries who endured (Johnston, 1962; Wood et al., 1992; Saunders and Hoppa, 1993; Ruff, 2003). Thus, skeletal samples are biased samples of all people alive at a given age in a population (Saunders and Hoppa, 1993). However, some authors argue that the majority of subadult skeletons in collections are more likely the result of acute conditions that would not have affected bone growth (Lovejoy et al., 1990).

Errors associated with methodology and sampling bias are considered greater than those associated with biological mortality bias or the osteological paradox. With regard to age estimations specifically, the inability to accurately sex unidentified skeletal remains automatically introduces error into the estimate due to a larger range of variation across all ages

(Saunders and Hoppa, 1993). The lack of comparative data sources greatly contributes to the unreliability and inaccuracies associated with subadult age and sex estimation (Scheuer, 2002), such that it is recognized that known-age skeletal samples, which can be used to develop models or validate models, need to be identified (Hoppa and Vaupel, 2002).

DATA SOURCES

The largest obstacle for the forensic anthropological community is the shortage of known modern subadult skeletal material (Scheuer and Black, 2000). Substantial collections of modern subadult skeletons with known demographics are virtually unknown (Feldesman, 1992; Shapiro and Richtsmeier, 1997; Scheuer and Black, 2000; Rogers, 2009; Franklin, 2010). Most museums are not actively collecting skeletal material, and even in collections that are actively growing, donations by parents of their deceased children are rare (Stull and Godde, 2013). Subadult skeletal material available in skeletal collections is either of archaeological origin or is extremely limited in terms of demographic variability (mostly for individuals per chronological year), and results in a situation where statistical analyses cannot be applied (Rogers, 2009).

Generating radiographic images are a standard practice for hospitals and morgues across the world, and vast amounts of data have been collected within these institutions. In the absence of modern subadult skeletal collections, forensic anthropologists can turn towards radiographic data sources in an effort to bridge this gap. Conventional radiography inherently produces images with distortion which are unsuitable to use for metric analyses, as the exact magnification of each measurement plane may not be known (Hoffman, 1979; Schroeder et al., 1997). Distortion, or the misrepresentation in size or shape of an object when a radiographic image is generated, is dependent on the source to image distance (SID), object to image distance (OID) as well as the location of the object in relationship to the center of the beam (Figure 2.1) (Bontrager, 2001). The degree of distortion increases proportionally to the distance between the object and the image (OID) (Schroeder et al., 1997). Thus, a measurement obtained from the anterior pelvis, if the individual is in the supine position, will have more error compared to a measurement obtained from the posterior pelvis. Furthermore, images generated with a cone beam (traditional radiography) contain distortion diverging from the center (Stull et al., 2013 b).

The greater the divergence of the X-ray beam, due to a shorter SID, results in greater distortion; in contrast, less divergence of the X-ray beam, due to a greater SID, results in less distortion (Stull et al., 2013 b). If the error (distortion) associated with the radiograph is unknown, due to a lack of knowledge concerning the SID and OID, then a measurement obtained from a radiograph may not reflect the true pelvic measurements (Schroeder et al., 1997). Maresh (1955, 1970) displayed a calculated distortion error ranging from 1% to 3% when the distance to the collimator was 7 ½ feet with the limb in direct contact with the cassette. However, Green et al. (1946) suggests distortion can be approximately 4% to 6% when the SID is 6 feet. Percentage of distortion error is positively correlated to the length of the long bone such that the longer the bone the larger the error (Feldesman, 1992). On the other hand, some authors feel the effect of distortion may be so small that it does not drastically affect the final 95% prediction interval (Hoffman, 1979; Smith, 2007).

The Lodox Statscan (Lodox Systems Pty (Ltd), Sandton, South Africa) is a fast acquisition, low-dose, full body radiographic device that was originally designed for the South African diamond mining industry. In 1999, the Lodox Statscan was the first used in health sciences (Beningfield et al., 2003). Although the slot-scanning technology is renowned for the lower radiation, the technology offers images void of distortion because the X-ray source is projected through a collimated fan-beam onto a detector, which moves in synchrony over the patient (Douglas et al., 2010). Therefore, Lodox Statscan-generated radiographs can be used to evaluate the subadult skeleton with metric analyses.

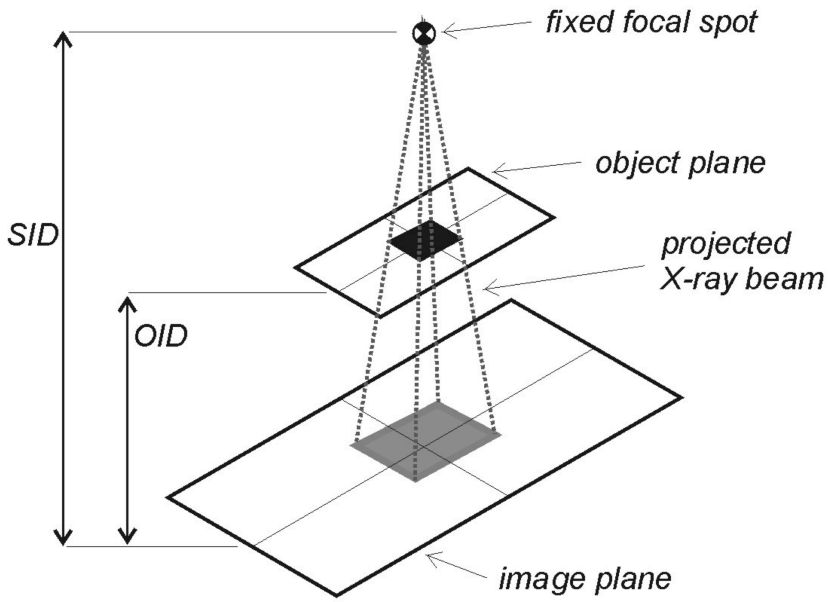


Figure 2.1 – Illustration depicting the differences in SID and OID in conventional (cone beam) radiography (adapted from Stull et al., 2013 b).

CHAPTER 3: MATERIALS

The cross-sectional sample was collected retrospectively from two institutions in the Western Cape Province of South Africa and included a total of 1380 children between birth and 12 years old. To circumvent any possible sampling issues associated with cross-sectional data such as a different number of individuals per each age interval, an equivalent number of individuals for each age were included. The sampling allowed for a large, evenly distributed sample with substantial numbers of individuals, even when separated by sex, for all years from birth to twelve. The date of birth and either date of examination or date of death was utilized to determine the chronological age at time of imaging. Age categories were based on calendar years; for example, individuals in the age category of 2 year-olds had chronological ages between 2.00 years and 2.99 years. Images comprising the sample were produced between 2007 and 2012. Random sampling of a population representative of the entire population ultimately affects sample validity in the statistical analyses and subsequent interpretations.

Demographic and biological data were collected from the medical files from each institution. As each institution has its own operating procedures dependent on its mission and scope, different variables were available at each location. For example, Salt River Forensic Pathology Laboratory provided biological data such as ancestry but the Red Cross War Memorial Children's Hospital provided variables one could use to estimate ancestry (see *Ancestry*, page 35). Variables that were consistently collected at both institutions included sex, age, height, weight, and either cause and manner of death (Salt River) or reason for diagnostic imaging (Red Cross Hospital).

The upper age limit of the current study was 12 years because of the potential for epiphyseal fusion; the diaphyseal length cannot be obtained if one of the epiphyses has fused. Although no population-specific technique exists for epiphyseal fusion in South Africa, previously researched populations generally do not display fusion in the lower or upper limbs prior to the age of 13 years (McKern and Stewart, 1957; Scheuer and Black, 2000; Crowder and Austin, 2005; Cardoso, 2008 a; b; Schaefer, 2008). An age range of birth to twelve includes individuals who do not exhibit epiphyseal fusion of the long bones, thus the use of diaphyseal dimensions is an appropriate method to estimate age.

Long bone length and breadth measurements were acquired for each individual from

Lodox Statscan (Lodox Systems Pty (Ltd), Sandton, South Africa) radiographic images using a custom imaging and viewing software, DVS or Diagnostic Viewing Station (www.lodox.com). The scanner was installed at Red Cross War Memorial Children's Hospital in 2004 and the Salt River Forensic Pathology Laboratory in 2007.

INSTITUTIONS

The combination of the two data sources was ideal for the research as it ensured that all ages would have sufficient numbers of individuals for statistical analyses. Salt River Forensic Pathology Laboratory (Salt River) and Red Cross War Memorial Children's Hospital (Red Cross) provided access to their radiographic images as well as demographic data. The Salt River sample was comprised of deceased children who were generally less than 6 years of age; the Salt River portion totaled 16% (n = 165) of the entire sample size (Table 3.1). Generally in South Africa, forensic pathology services tend to have a larger number of individuals in the younger age categories due to high infant and child mortality rates (WHO, 2003; Sartorius et al., 2011; United Nations, 2013). In the Western Cape in 2009, neonatal deaths accounted for 35% of the total deaths under five years, which commonly included cause of deaths attributed to prematurity (14%), birth asphyxia and severe infections (6%) and congenital abnormalities (5%) (Groenewald et al., 2011). The least three frequent causes of death under five years of age in the Western Cape were injuries (7%), congenital (4%) and malnutrition (3%) (Groenewald et al., 2011). Although 48% of the Salt River sample had a natural manner of death, a similarly high percentage (38%) of individuals had an accidental manner of death. Of the Salt River sample, only 7% were noted as homicides and 0.07% as undetermined.

The Red Cross War Memorial Children's Hospital is a public hospital that serves children from birth to twelve years from all nine provinces in South Africa and is considered one of the only dedicated child health institutions in the country. The Red Cross sample constitutes 84% (n = 1145) of the total sample with evenly distributed ages throughout the majority of years; the smallest number of individuals is between birth and 2 years of age (Table 3.1). Most children in the sample required diagnostic imaging prior to treatment of accidental injuries. Cases of burns or sexual abuse most likely would not be imaged, or present in the Red Cross sample,

because diagnostic imaging would not be required for treatment (personal communication, Dr. Thomas Blake). Corroborative data stems from a review of the use of the Lodox Statscan in the Trauma unit at Red Cross, which demonstrates that motor-vehicle accidents (MVA's) were the most common mode of injury, accounting for upwards of 74% of the cases imaged in the first two years of use (Douglas et al., 2010). Although all children in the Red Cross sample were living when the images were generated, a possibility exists that the child died of incurred injuries during the course of treatment.

Although all manners of death were collected, some individuals were removed in order to reduce the sampling errors associated with mortality bias (Saunders and Hoppa, 1993). Circumstances for removal included individuals with a manner of death listed as undetermined and less than 6 months of age, a cause of death listed as sudden infant death syndrome (SIDS), and all children who were less than 30 days of age (including stillborns). Mortality reports for 2009 indicate that a fairly high percentage of neonatal deaths are attributed to prematurity (Groenewald et al., 2011); however, premature children are not always noted in the medical files. Specific to individuals with a manner of death as undetermined, the processes affecting the growth are unknown and thus, the children cannot be adequately compared to their contemporaries. Essentially, the children removed from the sample represented a small subset of the population of failure to thrive children, which would most likely have exhibited a different growth rate compared to their contemporaries and if included may skew the lower ages of the total sample. Most of the individuals who remained in the sample had a manner of death as either accident or homicide and would most likely exhibit growth rates similar to the children comprising the Red Cross sample. The sample size decreased from 1380 individuals to 1330 individuals and following the removal of outliers (n=20), the sample size was further reduced to 1310 individuals.

SEX

In an attempt to combat sampling errors and unbalanced sample sizes in cross-sectional data, the minimum number of males and females was equivalent for all ages ($n > 30$) except in the less than 1 year and 12 year age groups. The final sample includes 506 females and 804

males (Table 3.2). The Red Cross sample highly influenced the total sample because of the larger number of individuals in comparison to the Salt River sample. A review of cases following 2 years of the Statscan's use in the Trauma Unit at Red Cross demonstrated that 64% of the patients who qualified for Statscan imaging were male (Douglas et al., 2010). Thus, the sample in the current study is representative of sampling from the hospital. More males than females also comprised the Salt River sample. This parallels deaths by sex and age noted in the Western Cape Mortality Report (Groenewald et al., 2011), thus the current sample also reflects normal sampling distribution from the Forensic Pathology Service.

Table 3.1 – The age distribution of the RXH sample provides the larger contribution to the overall sample, as well as more individuals older than 2 years of age. In contrast, SR sample has fewer individuals overall but more comprise the younger ages (less than 4 years of age).

Age	**Red Cross**			**Salt River**		
(years)	*Females*	*Males*	*Pooled*	*Females*	*Males*	*Pooled*
0	6	9	15	23	22	45
1	22	30	52	10	13	23
2	27	68	95	8	17	25
3	40	64	104	12	10	22
4	39	53	92	11	5	16
5	33	66	99	4	5	9
6	39	68	107	1	2	3
7	42	68	110	1	4	5
8	46	60	106	NA	3	3
9	41	57	98	NA	4	4
10	34	64	98	1	2	3
11	43	62	105	NA	3	3
12	21	43	64	2	2	4
Total	433	712	**1145**	73	92	**165**

Table 3.2 – Sample size and age distribution separated by sex.

Age (years)	Sex		
	Females	*Males*	*Pooled*
<1	29	31	60
1	32	43	75
2	35	85	120
3	52	74	126
4	50	58	108
5	37	71	108
6	40	70	110
7	43	72	115
8	46	63	109
9	41	61	102
10	35	66	101
11	43	65	108
12	23	45	68
Total	**506**	**804**	**1310**

ANCESTRY

South African law forcefully divided the country on social race from 1948, when the National Party marginally won the general election and subsequently introduced apartheid throughout South Africa, until 1991 (Cameron, 2003; Viljoen and Sekhampu, 2013). Apartheid forced segregation that resulted in major disparities between population groups and devastatingly affected people with regard to education, housing, and psychology. Perhaps the greatest legacy of Apartheid was the Population Registration Act of 1950, which forcefully defined South Africans into four groups: white, native/bantu/black, coloured and asian. Many South Africans, and bureaucratic systems, use these terms now as self-identifiers, not in the sense of a legal

definition but in terms of redress for past wrongs (Christopher, 2002; Patterson et al., 2010). In the South African census of 1996, 99.1% of the population self-classified themselves according to four socially-identified population groups namely white, black, coloured, Indian/Asian (Statistics South Africa, 1999).

In 2011, black South Africans were the largest population group constituting 79% of the population, whereas white and coloured South Africans each represented 8.9% of the population. Asians and groups classified as 'other' total only 3% of the population (Statistics South Africa, 2012). Disparate population percentages are seen between the Western Cape Province and South Africa as a whole. In the City of Cape Town in 2011, coloureds were the largest population with a frequency of 48.8%, blacks followed at 32.9% and whites were 15.7% (Statistics South Africa, 2012). Indian or Asian South Africans represent the smallest frequency (<3%) for population groups both in 2007 and 2011 in Cape Town (Western Cape Provincial Treasury, 2011; Statistics South Africa, 2012).

A total of 703 (54%) individuals from the entire sample either had a recorded ancestry or had enough demographic variables available to estimate ancestry (Table 3.3). Black South Africans were the most prevalent group, totaling 60% of the sample. Coloureds were the next largest group, accounting for 39% of the sample and whites were the least prevalent group with only seven individuals. Due to sampling in the Western Cape Province, the number of coloured individuals in the dataset is expected to be larger than black South Africans. However, the lower number of coloureds estimated in the dataset might be a skewed result because of the variables, specifically language, used to estimate ancestry (see below).

When subdivided into population and sex the numbers are in accordance with the above-mentioned population and sex descriptions. Black males have the largest numbers (36%) followed by coloured males (26%), black females (24%), and coloured females (13%) (Table 3.3). Although white males and white females represent less than 0.01% of the entire sample, they were not removed from the final sample or statistical analyses. The 0.01% reflects the number of individuals with either known or estimated ancestry, not the definite number of whites included in the sample. More individuals are likely to self-identify as white, but unfortunately social race is difficult to estimate based on the cultural identifiers described below, and thus white South Africans resulted in the smallest sample sizes among the three population groups.

Table 3.3 – Age distribution separated by population and sex. There are considerably more black males than the three additional population-sex groups. Coloured females demonstrate the smallest sample size within the study, besides white males and white females, which were not included because of their small numbers.

Age (years)	Black			Coloured		
	Female	*Male*	*Pooled*	*Female*	*Male*	*Pooled*
<1	12	13	25	14	11	25
1	17	19	36	5	10	15
2	16	34	50	6	17	23
3	20	29	49	7	17	24
4	24	26	50	8	10	18
5	9	22	31	4	13	17
6	12	22	34	11	15	26
7	11	21	32	10	16	26
8	18	15	33	4	19	23
9	8	18	26	4	14	18
10	9	17	26	6	16	22
11	9	9	18	4	12	16
12	5	9	14	6	13	19
Total	170	254	**424**	89	183	**272**

COLOURED SOUTH AFRICANS

The social term coloured was created in 1808 to describe population diversity among the slaves and African-born free persons of the Cape colonies (Patterson et al., 2010). Historical situations established the coloured[1] group as a genetically distinct population with major intra- and inter-continental genetic contributions from Europe, Africa and Indonesia (Inwood and Masakure, 2009; Tishkoff et al., 2009; Patterson et al., 2010; Quintana-Murci et al., 2010; Petersen et al., 2013). An admixed population group is defined as being formed from two or more parental populations that are genetically distinct from one another (Patterson et al., 2010). Specifically, one study using autosomal DNA noted that the coloured population exhibited 48.4% non-African (European) contribution, 17.1% Asian contribution, and 28.5% Khoe-San contribution, and the smallest contribution (15.5%) from African non-Khoe-San (Petersen et al., 2013). Studies evaluating mtDNA have shown that coloured females display a much stronger Khoe-San than European contribution while coloured males display a larger Eurasian than Khoe-San contribution in the Y-chromosome (Quintana-Murci et al., 2010). High rates of admixture are most likely facilitated from lenient opinions regarding racial ideologies and more strict views concerning social fault lines based on religion prior to the 19th century (Inwood and Masakure, 2009).

The Khoikhoi and Bushmen (San) people are sometimes pooled and referred to as Khoe-San (other spellings include Khoisan and Khoesan) because both groups are indigenous to South Africa and have linguistic similarities (Khoisan linguistic family) (Patterson et al., 2010; Schlebusch, 2010; Petersen et al., 2013). The hunter-gatherer society of the San is recognized as the oldest group to inhabit southern Africa. However, the pastoralist society of the Khoikhoi arrived prior to the Bantu expansion (*ca.* 5000 to 4000 BP) and are thus also considered indigenous to sub-Saharan Africa (de Filippo et al., 2012). Mitochondrial DNA (mtDNA), Y chromosome, and autosomal chromosome diversity studies demonstrate that within Africa, the Khoikhoi and San populations cluster together while being the furthest from other African populations (Jakobsson et al., 2008; Li et al., 2008; Tishkoff et al., 2009). The genetic

[1] Currently, the term coloured is the most widely recognized population-specific identifier within the South African community and thus is used herein (Christopher, 2002; Adhikari, 2005; Patterson et al., 2010).

contribution of the Khoikhoi appears greater than the San contribution in the coloured population (Morris, 1997).

Because population history has a great influence on the coloured genotype, the geographic location – and thus different population history – affects the percentage of group contributions in coloured South Africans. Subsequently, results vary greatly even within different neighborhoods of Cape Town. While the larger sample of coloureds demonstrates upwards of 28% and 24% Indian and Indonesian contributions, respectively, District Six coloureds – a neighborhood once considered the center of the coloured population in Cape Town due to the Dutch-East Indian Company slave trade – demonstrated 63% and 50% Indian and Indonesian contributions, respectively. The large genetic variation in atDNA (autosomal markers inclusive of SNPs) of coloureds is evidence of differential proportions of parental populations for each individual (Patterson et al., 2010; Petersen et al., 2013).

BLACK SOUTH AFRICANS

Archaeological, linguistic, and genetic data indicate significant migration events in Africa over the past several thousand years shaped the pattern of variation of black South Africans (Tishkoff and Williams, 2002; Tishkoff et al., 2009). One of the most significant was the migration of the agricultural Bantu-speakers, hypothesized to originate in Cameroon and to travel into Southern Africa *ca.* 5,000 to 3,000 BP (Tishkoff and Williams, 2002; Ribot, 2003; Berniell-Lee et al., 2009; de Filippo et al., 2012; Petersen et al., 2013). The population movement led to the spread of Bantu culture, language, and genes across Southern Africa (Diamond and Bellwood, 2003). Populations of varying genetic composition were the result of the Bantu-speakers expansion occurring along multiple routes and at different times.

Modern black South Africans are generally subdivided on factors associated with ethnicity, such as kinship, religion, language, and shared territory (Treiman, 2007). Specifically in a recent study of South African adolescents (aged 14), black children assigned more importance to language as a domain of self-identification than sex or age (Norris et al., 2008). The Bantu-speaking languages belong to Niger-Kordofanian, the largest linguistic phylum in Africa (Ribot, 2004). Nine of the 11 official South African languages are considered Bantu-

speaking languages and are associated with tribal groups. The two largest black ethnic groups within South Africa are Zulu and Xhosa (Treiman, 2007). Gene flow is presumed to have occurred between historical Bantu-speakers and indigenous groups (i.e. Khoe-San) because of the presence of click sounds in the isiXhosa language (a Bantu language group) as well as the genetic contributions of Khoe-San (mtDNA and Y chromosome DNA) identified in Bantu-speaking groups (Petersen et al., 2013).

ESTIMATION OF ANCESTRY

Although the Salt River Forensic Pathology Laboratory provided ancestry information in the medical files, the Red Cross War Memorial Children's Hospital did not. However, three variables could be collected from the medical files at the Red Cross War Memorial Children's Hospital that could be utilized to estimate ancestry, namely home language, residential address, and religion. These three variables can accurately subdivide each population group (Christopher, 2002; Treiman, 2007).

Although the legislation regarding the Group Areas Act (1950) was repealed in 1994, residential areas, both farming and urban, are currently segregated following the designs of the previous government. Consequently, residence information offers a mechanism to estimate ancestry due to the high probability of group membership based on location (Christopher, 2002; Viljoen and Sekhampu, 2013). Language is a strong cultural marker that has been demonstrated to significantly correlate with several genetic markers (Mateos, 2007). While coloureds are mainly an Afrikaans-speaking group of Christian denomination, two smaller groups of coloureds self-identify as Muslim (7%) and/or English-speakers (14%) (Treiman, 2007). As for black South Africans, the two largest lingual groups are Zulu, which includes 30% of the black population, and Xhosa, which includes 22% of the black population.

All three variables used to estimate ancestry were optionally self-reported on the personal information sheet completed upon admittance. At least two of the three variables were used to estimate ancestry of the Red Cross sample. Unfortunately, when English was reported as a home language, classification into one of the three population groups was impossible as most individuals in South Africa speak English, thus it is less of a cultural identifier. For example, a

minimum of 14% of coloureds consider English a home language. Therefore, the proportion of coloureds and whites is greatly lowered in the known or estimated ancestry dataset. Ancestry estimation of the Red Cross sample provides a thorough description of the sample utilized for age and sex estimations as well as presents potential sampling errors or biases. The estimated ancestry was not used for statistical analyses.

SOCIOECONOMIC STATUS (SES)

Because the statistical analyses did not include factors such as SES, the data was not separated on SES. However, overall SES of the sample is important to acknowledge, and is similar to identifying possible sampling errors or possible biases associated with ancestry. The Lodox Statscan is located in the Trauma Unit at the Red Cross War Memorial Children's Hospital, mainly serving lower SES groups within Cape Town. Children transferred from the other nine provinces for competent pediatric care most likely require different modes of imaging, such as magnetic resonance imaging (MRI) or computed tomography (CT), and rarely require diagnostic imaging from the trauma ward. As noted earlier, the most common mode of injury requiring Lodox Statscan imaging were motor-vehicle accidents, accounting for approximately 74% of the all the cases in the trauma unit (Douglas et al., 2010).

As for Salt River, a morgue population of children is generally agreed to represent a lower SES (Swart et al., 2012). Significantly different child mortality (death under five years of age) rates exist between higher and lower SES children (WHO, 2003). For example, in the same country a child from a privileged household has a 22% probability of death prior to 5 years of age, a child from a poor household has a 33% probability of death prior to 5 years of age (WHO, 2003). Furthermore, poor children in Africa are twice as likely to die than poor children in the Americas despite utilizing the identical definition and circumstances of poverty. Sex differences are also seen in child mortality rates, and normally males have a higher child mortality rate than females (WHO, 2003).

Socio-economic status is based on a multitude of factors with some being education, access to healthcare, and income. Specifically within South Africa, SES differences within populations are relatively small compared to between-population SES differences (Treiman et

al., 1996). Based on the 1996 South African census, whites earned upwards of 4 times the income of coloureds and more than five times the income of blacks (Statistics South Africa, 1999). Furthermore, it was estimated that nearly 40% of black South Africans are unemployed and live in poverty (Barbarin and Richter, 2001; Treiman, 2007; Viljoen and Sekhampu, 2013). Non-white South Africans that were employed constituted 81% of the manual labor force in 1996 (Treiman et al., 1996) while most (~90%) of white South Africans were working in non-manual labor jobs. The least advantaged white South Africans were considered substantially better off than either coloured or black South Africans (Treiman, 2007). Though Muslim and English-speaking coloureds are considered equivalent to white South Africans, the majority of the coloured population is considered to have comparable education levels to that of black South Africans, which is much lower than most white South Africans (Treiman, 2007). Even though South Africa has had a dramatic economic, social, and political transition since 1994, apartheid introduced dynamics that continue to perpetuate inequality in the 21st century (Sekhampu, 2013).

Although improvements have been made in basic services in some areas (i.e. housing, water, electricity), squalor still exists in most townships and many black and coloured South Africans have substandard access to education and healthcare (Sekhampu, 2013; Viljoen and Sekhampu, 2013). A township refers to an area of land, normally located on the periphery of cities, that was reserved for blacks and coloureds (separately) from the late 19th century through to the end of apartheid (Sekhampu, 2013). Currently, most townships persist with poverty and associated problems thereof. Although government housing has been supplied in some townships, corrugated metal shacks are not uncommon. And even in governmental houses, running water may be available but indoor flush toilets are rare (25%) (Barbarin and Khomo, 1997). A study in 2003 documented that 67% of households in townships live below the poverty lines (Slabbert, 2003). Furthermore, approximately 46% of households in one township that are considered poor have an income less than 50% of the poverty line (Sekhampu, 2004). The inferior quality of the urban housing standards, or townships, persists in blatant contrast to middle and upper class suburban South Africa (Viljoen and Sekhampu, 2013).

Within the entire sample for the current study, 63% (n = 875) of the individuals had known residential addresses. Within this subset, 20% of the sample resides in the township of Khayelitsha, 14% resides in the township of Mitchells Plain, and approximately 17% resides in the townships of Guguletu, Nyanga, and Gatesville. The stated five townships represent the five

largest residencies for the sample. The majority of the remaining sample resides in additional townships such as Athlone, Atlantis, Milnerton, Eerste River, Grassy Park, Hanover Park, Maitland, Kuils River, Langa, Strand, Somerset West, and Vredenberg. Only 9% (n =125) of the entire sample had specific housing information, such as living in an informal (shack) or formal (house) resident. Of the subset with documented housing, the sample was approximately equally divided with 51% living in formal residences and 48% living in informal residences. The assumption, based on the demographic data, is that the majority of the sample in the current study is considered of low SES. While it is not ideal to create age and sex estimation techniques from an inherently biased sample, this sample does represent the South African population. Thus, creation of country specific methods based on these samples will represent normal variation in age for the majority of South Africans.

CHAPTER 4: METHODOLOGY

LODOX STATSCAN

The Lodox Statscan has an X-ray tube mounted to one end of a C-arm and an image receptor mounted to the other end, thus the SID is a fixed distance of 130 cm (Figure 4.1) (Stull et al., 2013 a). The OID, which is the distance from the table to the image receptor on the C-arm, is approximately 6 cm at its lowest position. The fan beam, or linear slit scanner, is collimated to 1 mm or less and projects energy onto a detector which is located on the opposite side of the C-arm as it moves across the patient at one of the three speed options (35 mm/s, 70 mm/s, 140 mm/s)(Maree et al., 2007; Douglas et al., 2010; Whiley et al., 2012). The fan-beam design offers minimal scattered radiation that results in a higher resolution image and less radiation dose to the patient (Douglas et al., 2010). The linearly moving narrow fan beam design of the Lodox Statscan is different to the cone beam design of conventional X-ray machines, evident by the claim to produce minimal distortion generated in the y-axis (scan direction or long axis) (Figure 4.2).

In order to validate the claim of minimal distortion produced in the y-axis, Stull et al. (2013) imaged numerous bones, multiple times and in multiple circumstances. In an effort to simulate OID – as this is a large determinant of distortion – dense foam was used under each dry skeletal element. To correspond to actual patient positions, the dry skeletal elements were also placed at an angle. The dry bone measurements and the measurements obtained from the radiographic measurements were compared in terms of percent difference in an attempt to account for size differences in length and breadth measurements. Because minor errors are expected to occur within and between observers, percent agreement was utilized at the +/- 1 mm and +/- 2 mm levels. Additionally, Bland-Altman plots were used to depict agreement levels between the measurements as the statistic illustrates the spread of differences in measurements rather than just the relationship between the measurements such as a correlation coefficient (Bland and Altman, 1986).

Percent agreement in the +/- 1 mm range was fairly high at 85% and the percent agreement in the +/- 2 mm range was 97%. The percent differences ranged between 2% and 11% for y-axis and x-axis measurements, respectively. However, the average percent difference was only 0.5% larger for y-axis measurements and 4% larger for x-axis measurements (Stull et al., 2013 a). These results are comparable to error rates noted in prospective growth studies where the settings are controlled as well as intra- and inter-observer error rates on dry bone (Green et al., 1946; Maresh, 1955; Gindhart, 1973; Hoffman, 1979). Lodox Statscan images offer a potential radiographic source to image subadult skeletal structures that generates minimal distortion.

Figure 4.1 – Lodox Statscan radiographic machine.

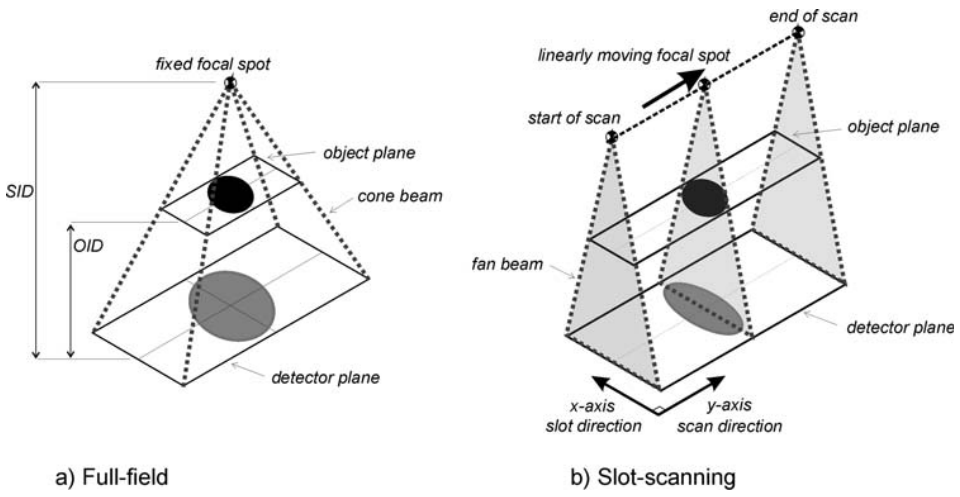


Figure 4.2 – Depiction of distortion in a) full-field conventional radiography with a cone beam geometry and b) Lodox Statscan fan beam geometry (taken from Stull et al., 2013 a).

MEASUREMENTS

Long bone length and breadth measurements were acquired for each individual in the sample from Lodox Statscan (Lodox Systems Pty (Ltd), Sandton, South Africa) radiographic images using a custom imaging and viewing software, DVS (www.lodox.com). Screen calibration was conducted upon installation of the software to ensure accuracy of measurements. Eighteen measurements were attempted on each Statscan-generated image of each individual, but because of variability in the placement of each patient, individuals could have a different number of measurements.

Measurement definitions were derived from and/or follow those from studies in fetal and subadult osteology (Fazekas and Kósa, 1978) and longitudinal growth studies (Maresh, 1970). Although most of the definition sources stem from dry bones, the definitions associated with the longitudinal growth studies were designed specifically for radiographic images (Maresh, 1955). If the measurements were not previously associated with fetal or subadult remains and/or defined in the literature, measurements specific to adult skeletons were modified from those presented in Standards (Buikstra and Ubelaker, 1994; Moore-Jansen et al., 1994). Measurements were only obtained when the bone was in proper anatomical position. Table 4.1 provides the measurement

and associated abbreviation that will be used herein. Figure 4.3 provides examples of a few measurements obtained in radiographic images when the individual is in anatomical position and Appendix I contains the measurement definitions and images of each measurement. In order to adjust for the increase in size with the increase in age, the breadth measures of each element were divided by the length measurement of the same element in order to create a ratio. The process created 12 new variables that were included in the statistical analyses stated below. Their definitions and abbreviations are noted in Table 4.2.

Because not every individual was placed perfectly for imaging, different combinations of measurements are associated with each individual. Femoral breadths and tibia breadths in the younger individuals (< 1 year) were difficult to obtain because of natural leg positions. Humeral, radial, and ulna lengths in the older individuals were difficult to obtain because of active epiphyseal fusion. Measurements were obtained from the side of the body that would yield the least distortion, which was generally measurements in the scan direction, or y-axis (Stull et al., 2013 a). If both left and right-sided elements were in proper placement than measurements were collected from the left- sided elements.

Table 4.1 – The measurements and associated abbreviations.

Humerus diaphyseal length	HDL	Radius midshaft breadth	RMSB
Humerus proximal breadth	HPB	Femur diaphyseal breadth	FDL
Humerus distal breadth	HDB	Femur distal breadth	FDB
Humerus midshaft breadth	HMSB	Femur midshaft breadth	FMSB
Ulna diaphyseal length	UDL	Tibia diaphyseal length	TDL
Ulna midshaft breadth	UMSB	Tibia proximal breadth	TPB
Radius diaphyseal length	RDL	Tibia distal breadth	TDB
Radius proximal breadth	RPB	Tibia midshaft breadth	TMSB
Radius distal breadth	RDB	Fibula diaphyseal length	FBDL

Table 4.2 – Ratios, and their associated abbreviations, included in the classification models derived for sex estimation.

	Ratio	Abbreviation
Humerus	HPB/HDL	HDPB
	HDB/HDL	HDDB
	HMSB/HDL	HDMS
Ulna	UMSB/UDL	UDMS
Radius	RPB/RDL	RDPB
	RDB/RDL	RDDB
	RMSB/RDL	RDMS
Femur	FDB/FDL	FDDB
	FMSB/FDL	FDMS
Tibia	TPB/TDL	TDPB
	TDB/TDL	TDDB
	TMSB/TDL	TDMS

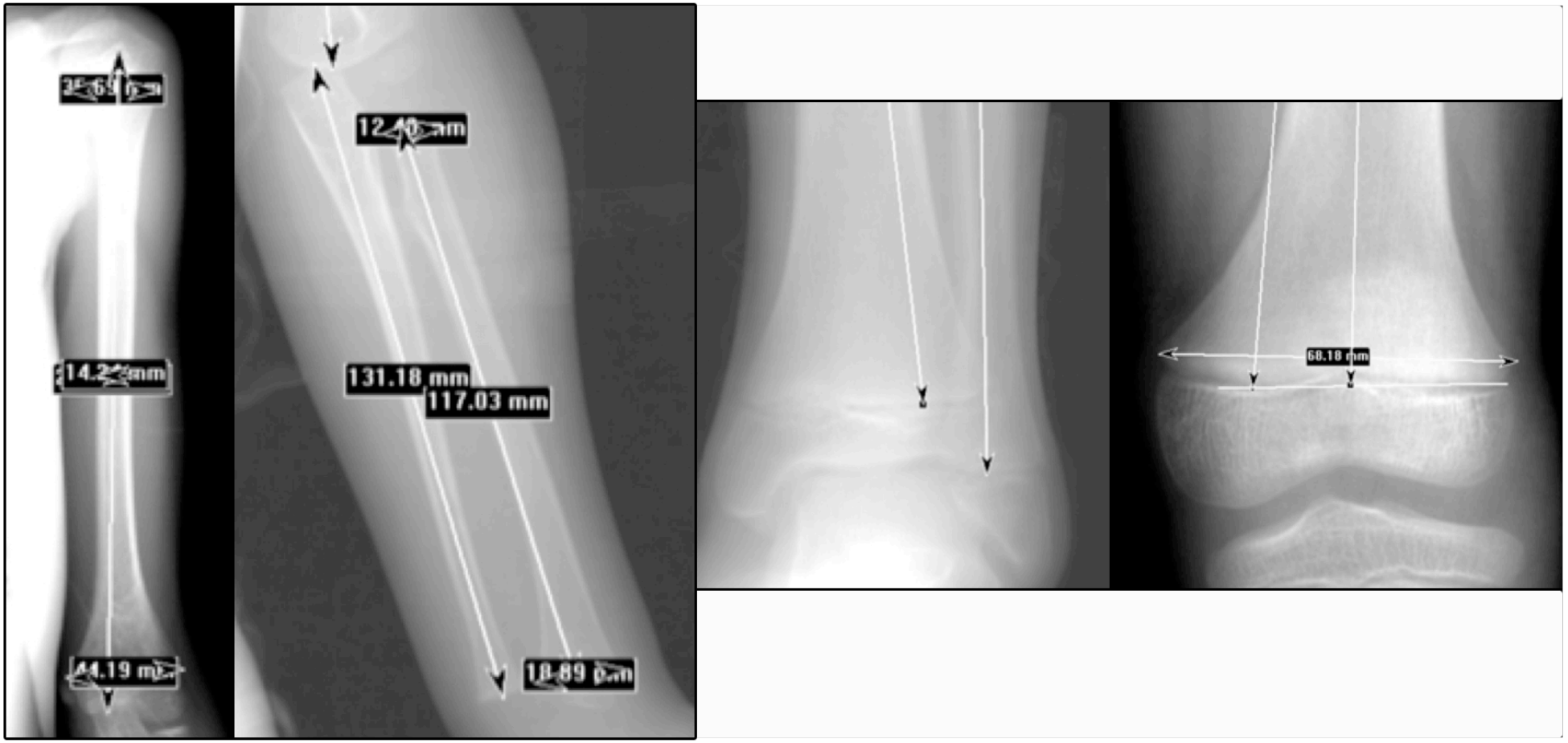


Figure 4.3. Example of measurements obtained on Lodox Statscan-generated images when the individual is in anatomical position. Measurement explanations from left to right: 1st image – HDL, HPB, HMSB, HDB; 2nd image – UDL, RDL, RPB, RDB; 3rd image – the distal extensions of TDL and FBDL; 4th image – FDB and the distal extensions of FDL.

INTER- AND INTRA-OBSERVER ERROR

Precision and reliability are directly attributed to consistency of repeated measurements on the same object and are used to assess data quality (Bland and Altman, 1986; Bailey and Bynes, 1990; Ulijaszek and Kerr, 1999). If considerable amount of variation exists in repeated measurements of the same subject, or poor repeatability, the statistical analyses associated with the measurement may be invalid or compromised, especially those which rely on regression, correlation and covariance (Bailey and Bynes, 1990; Goto and Mascie-Taylor, 2007). Therefore, the fewer and smaller measurement errors increase the probability of detecting other significant relationships (WHO Multicentre Growth Reference Study Group, 2006; Harris and Smith, 2009). Many variables can affect measurement error; however, in the current study, measurement error was mainly due to the side of the body the measurement was obtained, observer experience, and clarity of measurement definitions.

Fifteen individuals were randomly selected from the entire sample to estimate the intra- and inter-observer error, which demonstrates differences between repeated measurements by one observer and differences between single measurements taken by two observers, respectively. The definitions stated above were provided to the second observer, thus a potential of 18 measurements could be obtained from each individual, resulting in a possible comparison of 270 measurements. Because of variability in the placement of each individual, the number was reduced to 152 measurements for the intra-observer error and 149 measurements for the inter-observer error. Because both observers had the option of using left or right side elements, some of the variability in the measurement error may be due to asymmetry rather than measurement error. The intra- and inter-observer errors were assessed with technical error of measurement, relative technical error of measurement, and Bland-Altman plots.

The comparison of measurements must focus on the agreement between two variables and not the strength of the relationship between two variables, as one would assume a high correlation of the two variables that measured the same quantity (Bland and Altman, 1986). Additionally, the relationship between variables is extraneous to the degree of agreement as data that have poor agreement can produce high correlations (Bland and Altman, 1986; Rothwell, 2000). For example, one person may always disagree by 4 mm; the correlation is still high but the agreement is poor. One of the most commonly employed measures of imprecision is the

technical error of measurement (TEM), which is a measure of error variability and represents the square root of the sum of squared differences divided by twice the sample size, or simply the standard deviation between repeated measures (Dahlberg, 1940; Mueller and Martorell, 1988; Knapp, 1992; WHO Multicentre Growth Reference Study Group, 2006; Goto and Mascie-Taylor, 2007; Geeta et al., 2009). The equation for TEM is:

$$\text{TEM} = \sqrt{\frac{(\Sigma D^2)}{2N}}$$

where *D* is the difference between the measurements and *N* is the number of individuals measured (Ulijaszek and Kerr, 1999). A cluster analysis based on the inter-relationships among the absolute values of 11 estimators of measurement precision identified TEM to be most similar to mean absolute difference, percent agreement, and the average deviation (Utermohle et al.,1983). Utermohle et al. (1983) suggested the groups in the cluster with TEM were estimator of precision least affected by measurement size as all four statistics demonstrated the smallest correlations to mean measurement length. The ability to be unaffected by size is an extremely important characteristic for the current research considering the large differences between diaphyseal midshaft measurements and diaphyseal length measurements.

TEM is positively related to the measurement and carries the units of the measured object (WHO Multicentre Growth Reference Study Group, 2006). Because
large mean values have a large TEM and small mean values have a smaller TEM "comparative imprecision of different measurements cannot be assessed" (Ulijaszek and Kerr, 1999; page 167). To overcome this, TEM can be converted to relative TEM (%TEM), which is the error expressed as a percentage of the mean of the measurement analyzed (Perini et al., 2005). The converted percentage has no units and allows for direct comparisons of all measurement sizes (Ulijaszek and Kerr, 1999). Unfortunately, acceptable levels of TEM and %TEM are difficult to determine as it may be group, population, or even age dependent in anthropometric studies; therefore, the values for TEM vary across measurement type and intra- and inter-observer error

$$\%\ \text{TEM} = \left(\frac{TEM}{mean}\right) * 100$$

(Ulijaszek and Kerr, 1999). Overall, the lower the %TEM obtained, the better the precision of measurements (Geeta et al., 2009). Some authors note an acceptable level of intra-observer %TEM for skilled observers to be between 1 – 5% and 1.5 – 7.5% for beginner observers. Inter-observer %TEM is higher and ranges from 1 – 7.5% for skilled observers and 2 – 10% for beginner observers (Perini et al., 2005).

A Bland-Altman plot can be employed to visualize the amount of agreement between or within observers (Bland and Altman, 1986). The plot reveals the overall trends in the agreement of two datasets and identifies any systematic bias' and outliers by plotting the means of the repeated measures along the x-axis and the differences between the corresponding measurement pairs on y-axis (Rothwell, 2000; Geeta et al., 2009; Harris and Smith, 2009). The limit of agreement, both positive and negative, is the reference interval that is based on the mean and standard deviation. Thus the upper and lower limits of agreements are different for each sample (Bland and Altman, 1986; Geeta et al., 2009). Additionally, the limits provide insight into the amount of random variation that is present. If the observers tend to agree, the plot should show a random scatter of differences around a mean of zero; if observers tend to disagree, the scatter will extend beyond the limits of agreement.

DATA PREPARATION

All analyses were performed in R (R Core Team, 2013), an open source software program. Besides acting as an environment to perform statistical analyses, R is a programming language that is similar to S, a preferred language within statistics (R Core Team, 2013). Because the statistical analyses conducted in R are through packages, R is extremely flexible and is not limited to predetermined analyses. The packages employed for each statistical analysis will be cited accordingly.

OUTLIER DETECTION

Prior to any statistical analyses, the data were explored for outliers. Checking the data for outliers is an essential step in research as all outliers require identification prior to model creation or analysis (Ben-Gal, 2005). Outliers can be defined as a "unique combination of characteristics identifiable as distinctly different from other observations" (Hair et al., 2007; 73). Essentially, a univariate outlier is an extreme value that is exceptionally high or low in relation to the other values in the same sample. Univariate techniques are greatly influenced by outliers such that the mean and variance may mask true outliers (Jackson and Chen, 2004). In a high dimensional dataset, the multidimensional position of each observation needs to be objectively measured and compared to the overall configuration of the entire dataset. Thus, observations that were not recognized as an outlier in the univariate approaches may be recognized as an outlier in the bivariate or multivariate approach. Generally, a multivariate approach is considered superior to univariate and bivariate techniques (Jolliffe, 2002; Hair et al., 2010). Outliers have a practical effect such that they strongly influence analyses and a substantive effect such that they may reflect the outer edges of a normal population. Although not every outlier is automatically problematic, outliers require evaluation in order to investigate their influence and impact on the dataset (Hair et al., 2007). Furthermore, outliers are not always representative of unique observations but can be the result of human error.

Outlier detection was conducted with univariate, bivariate, and multivariate techniques. Visual assessment was conducted through scatterplot matrices, boxplots and basic scatterplots. Boxplots are summary statistics that use the interquartile range as a measure of the spread of the data (Wickham and Stryjewski, 2011). If an observation falls outside of one and a half times the length of the interquartile range, the point is considered an outlier. Scatterplots and scatterplot matrices display potential outliers as isolated observations. Scatterplot matrices are valuable in large datasets as it permits multiple scatterplots to be viewed at one time.

Multivariate analyses included the use of R functions mvoutlier, outlierTest and princomp, all functions located in the mvoutlier (Filzmoser and Gschwandtner, 2013), car (Fox and Weisberg, 2011), and stats (R Core Team, 2013) packages, respectively. mvoutlier (Filzmoser and Gschwandtner, 2013) interprets multivariate outliers from Euclidean space and produces an adjusted quantile plot, indicating observations with above average values for most of

the univariate variables as possible outliers. The outlierTest function was applied to linear models created for each element (age regressed on measurement). The p-value reported the largest absolute studentized residual – the measure resulting from the division of a residual by an estimate of its standard deviation – in the sample, using the *t* distribution (Fox and Weisberg, 2011). Principal component analysis (PCA) was also used to identify atypical observations and possible outliers utilizing the princomp function (Chen et al., 2009; R Core Team, 2013). PCA is a data reduction technique that linearly transforms the data into a new set of the same number of variables that are uncorrelated to one another all while retaining the total variance in the correlation matrix or the variance-covariance matrix (VCVM). By plotting the principal component (PC) scores derived from a correlation matrix or VCVM with a biplot, one can identify observations that have relatively atypical variable combinations. Furthermore, the multivariate distances to the sample centroid can be compared to evaluate the observations with the largest distances.

The DVS imaging software retained all measurements obtained on the Lodox Statscan-generated images; therefore, all observations that were noted during the outlier detection methods were checked for typographical errors. The majority of observations noted as possible outliers were typographical errors that were corrected, allowing for a large retention rate. A total of 20 individuals were removed from the sample. The removed individuals were consistently marked as outliers through univariate and multivariate techniques with no associated pathological reason for being exceptionally large or small for age. Tabachnick and Fidell (2007) noted that the removal of outliers presents few consequences in large samples. Following the removal of the outliers, the data was again processed with the multivariate techniques to ensure all true outliers were removed.

EXPLORING RELATIONSHIPS BETWEEN SEX, AGE AND POSTCRANIOMETRIC VARIABLES

The data required graphical examination in order to evaluate the type of relationship the diaphyseal dimensions had with age and sex. Scatterplots of age regressed on a measurement and separated by sex were plotted with loess (local regression) smoothing lines to visualize

relationships. Loess uses locally weighted polynomial regression to fit a smooth curve through data without making assumptions about the form of the relationship. Based on the graphs, the relationship between age and postcraniometric variables was obviously nonlinear, reflective of a biological growth curve. The relationship of the data provides insight to the most appropriate analyses, specifically flexible models for age at death estimation that can compensate for the nonlinearity.

Correlations, inclusive of both sexes, population groups, and all ages were conducted on all measurements. Correlations provide a summary coefficient of the extent of the linear relationship that exists between variables (Kachigan, 1991). The results of the correlation matrix demonstrate the direction and strength of the relationship between variables and consequently illustrates the potential for predictive information (Kachigan, 1991). Because correlations only represent the linear relationship between two variables, nonlinear relationships will not be reflected in the Pearson correlation coefficient and some relationships within the dataset may be stronger than reflected in the coefficients (Hair et al., 2010).

A Holm's adjustment was incorporated into the p-values for the correlations because an increase in the number of statistical tests directly increases the likelihood of a Type I error, or rejection of the null hypothesis when in fact it is true. Holm (1979) offers one of the most powerful sequential tests that is universally valid, distribution free and is designed to increase the power of the tests while managing the inflation of alpha (Holm, 1979; Levin, 1996; Abdi, 2010). The test first obtains all of the p-values and then employs a Bonferroni correction to the observation with the smallest probability; if the test is significant the Holm's procedure continues until the first observation is not significant (Abdi, 2010). A Holm's correction was also incorporated into the two-tailed Student's t-test that was conducted on the mean male and female values of each measurement for each individual measurement-age dataset (i.e. the proximal breadth of the humerus for all individuals between birth and 12 years) to test for sex differences at each age. Assumptions with the t-test include normality; if the sample violates the assumption of normality then an equivalent, non-parametric Mann-Whitney test should be utilized.

Analysis of variance (ANOVA) is a univariate hypothesis test used to analyze differences of group means for more than two groups. Multivariate analysis of variance (MANOVA) is a multivariate extension of an ANOVA that compares multiple means at one time by combining the dependent variables into a linear composite, or new variable, with each single variable

contributing to the overall relationship based on the strength of the individual correlation (Weinfurt, 1995; Tabachnick and Fidell, 2007). Significant MANOVA results indicate that the groups (i.e. sex and age) differ in respect to the new composite variable. MANOVAs do not work well if the predictor variables suffer from multicollinearity, which is demonstrated by correlation coefficients greater than 0.9 (Tabachnick and Fidell, 2007). Consequently, ANOVAs were also conducted. Both ANOVA and MANOVA were conducted to identify the effects of age, sex, and the interaction of age and sex on the measurements.

Numerous assumptions are associated with both tests. Specifically for the ANOVA, the assumptions include linearity, normality, and equal variances of the predictor variable. The same assumptions are true for the MANOVA except the analog to equal variances is homogeneity of the variance-covariance matrices (VCVMs) of the multiple predictor variables (Weinfurt, 1995; Tabachnick and Fidell, 2007). Because Quantile-Quantile plots (QQ plots) are used to evaluate shapes of distributions, the test was used as a visual tool to assess normality and demonstrated the variables were somewhat normally distributed. Scatterplot matrices (Fox and Weisberg, 2011) were used to visualize the relationships among the predictor variables. A Levene's test, a test for homogeneity of variances in *k* groups based on the ANOVA statistic applied to absolute deviations of observations from the corresponding group mean, was conducted on each univariate measure prior to the ANOVA (Fox and Weisberg, 2011). The assumption was made that the variance-covariance matrices were equal for males and females.

While the ANOVAs were conducted on the univariate predictor variables, the sample was subdivided into bone subsets and four additional subsets based on element type and measurement type for the MANOVAs in an effort to retain sample sizes. The bone and measurement type and element type subsets include: humerus (n=497), radius (n=281), ulna (n=400), femur (n=778), tibia (n=521), upper limbs measurements (n=96), lower limb measurements (n=458), proximal element measurements (n=316), and distal element measurements (n=100)). All of the subsets were used throughout the results for the multivariate analyses. This process was followed so superfluous variables would be removed that would most likely not contribute to final models while making an effort to retain adequate sample sizes.

DATA TRANSFORMATION: PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) was conducted on each subset in order to extract the total variance from the dataset and cope with multicollinearity, all while reducing dimensionality. PCA uses orthogonal transformations to convert the raw data into uncorrelated, linear combinations of variables that maximizes the variance from the residual correlations (Everitt and Dunn, 2001; Jolliffe, 2002; Tabachnick and Fidell, 2007; Hastie et al., 2009; Wright and London, 2009; Rousseeuw and Hubert, 2011). The new variables, or principal components (PC), are derived in decreasing order of proportional variance. For example, the first principal component includes the combination of variables that expresses the largest variance in the dataset while the last principal component includes a combination of variables that expresses the smallest variance in the dataset. If the correlations among all variables are positive, then the first principal component (PC1) is interpreted as a measure of size, illustrated by all either positive or negative coefficients – the sign is arbitrary as the variance is unchanged (Jungers et al., 1995; Klingenberg, 1996; Jolliffe, 2002). In contrast, PC2 – PC*x* present with positive and negative coefficients, which are recognized to explain variation resulting from differences in shape while controlling for size (Jungers et al., 1995; Jolliffe, 2002). Thus, the removal of PC1 from analyses is recognized as an approach that successfully adjusts for size, which is evident among individuals of different ages in the sample. PCA was performed using the variance-covariance matrix with the princomp function in the stats package (R Core Team, 2013). The principal component (PC) scores were used in both age and sex estimation analyses described below.

The aims of the current study were to provide sex and age estimation models for South African subadults, thus two different types of analyses needed to be applied – classification and regression. Overall, the analyses were conducted independently (i.e. age estimation analyses were independent of sex estimation analyses); however, some variables from the age estimation were used in the classification models for sex estimation (i.e. y-axis residuals). The complexity of the analyses progressively increased for the classification and regression models from univariate models, multivariate bone models and subsets to the all-measurement model. Again the purpose of the subsets was to preserve large sample sizes. The variables chosen for each multivariate model were based on the significant relationships with sex or age in the

MANOVA/ANOVA and Student's t-tests results as well as the results from the univariate and bone and subset models.

AGE ESTIMATION

Based on the exploratory analyses, linear regression models were not flexible enough to fit the nonlinear relationships evident between diaphyseal dimensions and age. Consequently, basis splines and multivariate adaptive regression splines (MARS) were chosen as the preferred methods for the age at death estimation models. The use of the basis spline and MARS models present a novel approach within the field of anthropology to estimate age at death. The models created in this dissertation use age as the response variable and the measurement(s) as the predictor variable(s). Regressing age on measurement allows for the prediction of age rather than diaphyseal length and subsequently, 95% prediction intervals can be derived, which are part of the requirements for best practices in the field. Males and females were pooled for all age at death models, possibly resulting in increased variation per age though this process eliminates compounding errors in the estimates.

Numerous variations of basis splines models were analyzed for each univariate model and similarly, numerous variations of MARS models were also analyzed. The best two fits – based on standard errors, R-squared, Akaike information criterion (AIC) and ANOVA – were compared to one another to determine the most appropriate method for each specific measurement (see *Model Selection, page 65*). The multivariate bone and element type and measurement type subsets as well as the all-measurement subset were only subjected to MARS as the model is recognized to work well with variable interactions. Additionally, the PC scores for each of the subsets were also incorporated into MARS models. Following each model derivation, appropriate prediction intervals were created through resampling because both basis splines and MARS are nonparametric models (see *Resampling: Prediction Intervals,* page 67).

BASIS SPLINES

Additive models involve piecing together different curves to map out the relationship between the predictor and response variable. Different sub-regions of *x* may exhibit different relationships between the two variables. A polynomial function is applied to each sub-region that exhibits a different slope and is connected 'piecewise' to the other polynomial functions to ultimately create a smooth fit, or spline (Figure 4.4). The basis spline is considered a generalization of the Bézier curve, a spline that consists of no interior knots – the location where two piecewise polynomials join – and is an extension of the linear model into an additive model (Wright and London, 2009; Racine, 2012). However, the basis spline is more adaptable than the Bézier curve because the basis spline fit is constructed piecewise from a different polynomial function in each contiguous interval of *X*, which presents as one parametric curve (Hastie and Tibshirani, 1990; Eilers and Marx, 1996; Wood, 2006; Racine, 2012). In contrast to linear models where the *X* variable(s) are multiplied by a scalar (*B*), the scalar is replaced by a function in additive models (Wright and London, 2009). By attaching two regression lines with different slopes at one point, a linear regression fits nonlinear data. Piecewise-polynomials are considered superior to polynomial transformations – which have been recently presented in age estimation techniques – because polynomial transformations tend to be erratic at the boundaries when the coefficients are adjusted (Hastie et al., 2009).

The type of function, or order (*m*), specified is associated with the type and complexity of the spline and corresponds to different types of regression (Wright and London, 2009). For example, a 0-degree polynomial is analogous to the constant (β0), a second-degree polynomial is a quadratic regression, and a 3rd-degree polynomial is considered a cubic regression (Wright and London, 2009). The order (*m*) and number of knots (*N*) defines a basis spline. Fundamentally, two knots are always at the endpoints making the total number of knots $N + 2$ and the transition between each polynomial is continuous. The remaining knots are located in a specified location (non-uniform) or equidistant (uniform) for the best fit. This process is fundamentally different from smoothing splines which use the observations themselves to smooth the fit (Racine, 2012). The quantiles knot sequence has knots located at the quantiles of the empirical distribution of the predictor variable; the number of observations per interval are equal but the segments may not be an equal length (Racine, 2012). If not specified in the bs function of the splines package (R Core

Team, 2013), the knot, in a one-knot model, is placed at the median of the predictor variable and in a two-knot model, at the first and second tertiaries. The number of knots is directly related to the fit of the data such that underfitting is the result of too few knots while overfitting is the result of too many knots (Figure 4.5) (Eilers and Marx, 1996).

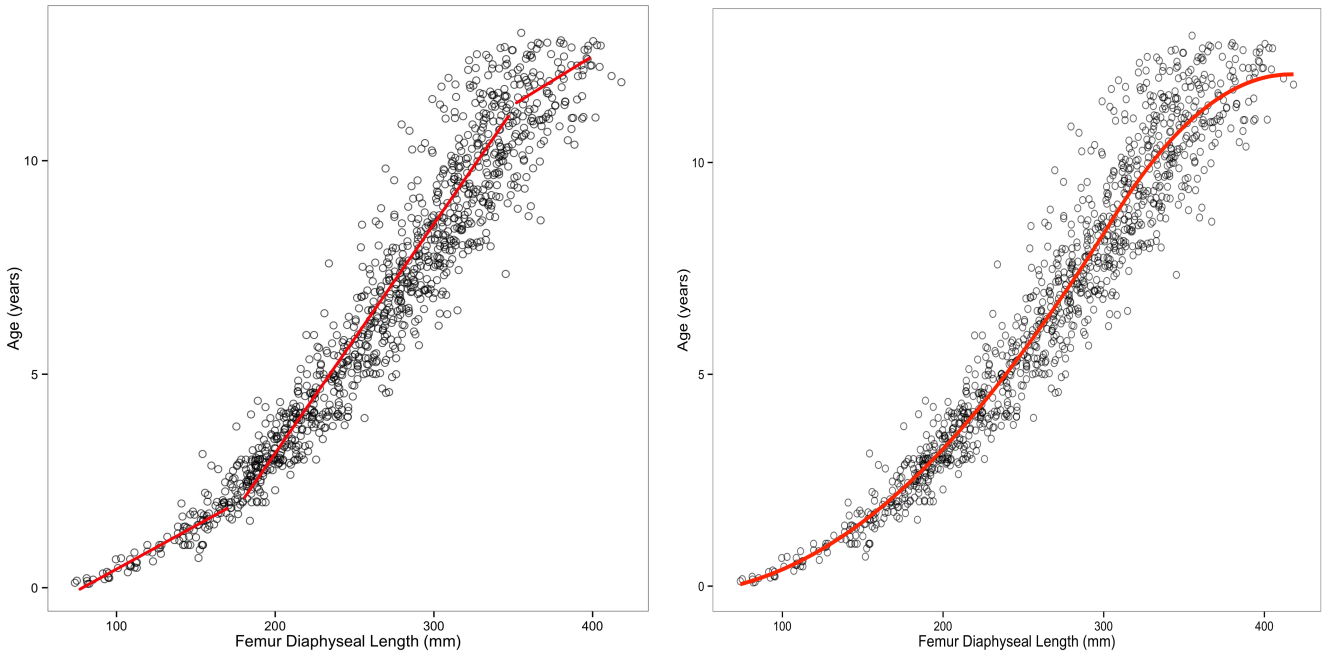


Figure 4.4 – The left figure demonstrates different slopes for three sub-regions of x. The right figure demonstrates a quadratic spline with three knots that 'connects' the three regression lines observed in the left figure.

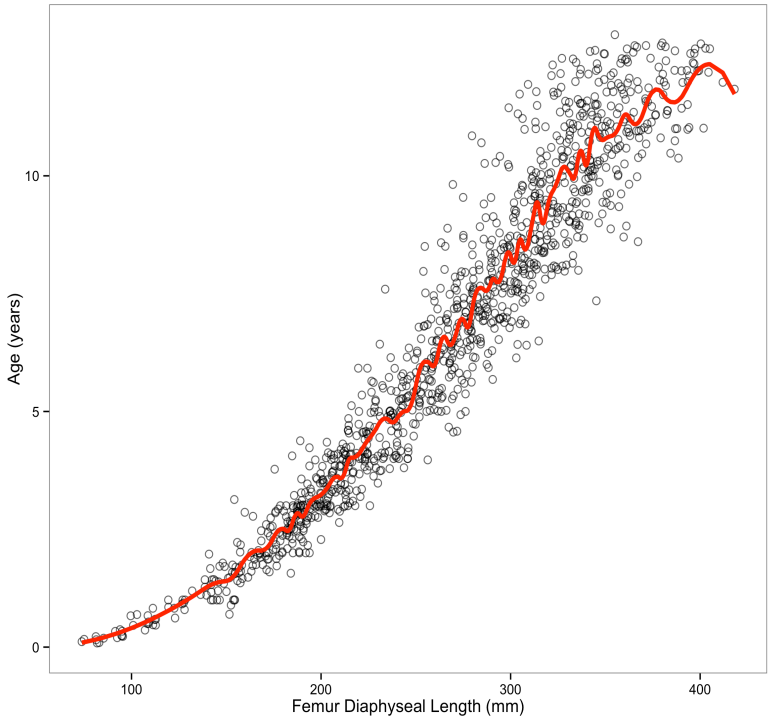


Figure 4.5 – An example of an overfit basis spline model with 60 knots.

Any spline function can be expressed as a linear combination of basis splines. In the basis spline algorithm, *t* denotes the knot sequence, *m* denotes the order, and *x* is the variable.

$$S_{m,t}(x) = \sum_i \alpha_i B_{i,m,t}(x)$$

The foundation of the basis spline is the basis function

$$N_i^m(x) = \frac{x - t_i}{t_{i+m} - t_i} N_i^{m-1}(x) + \frac{t_{i+m} - x}{t_{i+m} - t_{i+1}} N_{i+1}^{m-1}(x)$$

$$N_i^1(x) = \begin{cases} 1 & if\ t_i \le x < t_{i+1} \\ 0 & otherwise \end{cases}$$

where $N_i^m(x)$ is the *i*th basis spline function of order m
and t_i is a non-decreasing set of real numbers known as the knot sequence
and *x* is the parameter variable. The feature of basis functions is the knot sequence and is defined as $m + 1$ knots or *m* intervals. The number of basis functions in each model is dependent on one subtracted from the degree (m = 1, 2, or 3) and then subtracted by the total number of knots (e.g. m – 1 – N). Once the basis functions have been identified, the models are considered linear and fitting of the model can proceed as in linear regression (Hastie et al., 2009). The basis splines models were conducted using the bs function in the splines package (R Core Team, 2013).

MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS)

Multivariate adaptive regression splines (MARS) is a flexible, nonlinear, nonparametric regression modeling technique that fits general nonlinear multivariate functions through subsets of the variables that makes no assumptions about the relationships of the variables (Friedman, 1991; De Veaux and Ungar, 1994; Butte et al., 2010). The model is viewed as a generalization of recursive partitioning regression or additive models (Friedman, 1991). Specifically, MARS amalgamates linear regression, truncated basis functions, and binary recursive partitioning to approximate the underlying function and model relationships (Muñoz and Felicísimo, 2004). MARS requires large datasets to build a suitable model while it is still competitive in low dimension situations (i.e. fewer predictor variables); however, the analysis is considered ideal in situations where there are between two and 20 predictor variables (moderate to high dimension) and between 50 and 1000 observations (moderate sample sizes) (Friedman, 1991; Sekulic and Kowalski, 1992; De Veaux and Ungar, 1994). Additionally, because MARS is an adaptive technique the final model is considered more accurate compared to a technique that has a fixed set of basis functions (Barron, 1994).

MARS is considered to be sensitive to outliers and high leverage points in predictor space (Friedman, 1991; De Veaux and Ungar, 1994). While an outlier is an observation that does not conform to the overall variation demonstrated by a large positive or negative residual, a high leverage point presents with an unusual combination of predictor variables presenting as an

outlier in predictor space (Kabacoff, 2011). Overall, MARS strength and flexibility can be attributed to its heavy use of the response values to construct the basis functions. The bias of model estimates is usually greatly reduced, simultaneously increasing the variance as additional parameters are adjusted to better fit the data (Friedman, 1991).

Achieving the appropriate *f(X),* or basis function, is the ultimate goal of MARS. This basis function is estimated by subdividing *X* into regions and obtaining estimates of *f(X)* for each region (Friedman, 1991; Sekulic and Kowalski, 1992). MARS uses expansions in piecewise linear basis functions of the form of x > *t,* then (*x* – *t*) and if x < *t,* then (*t* – *x*), whichever permits the term to result in a positive value (Sekulic and Kowalski, 1992; Hastie et al., 2009). Each piecewise linear function is separated by

$$(x-t)_+ = \begin{cases} x-t & if\ x > t, otherwise \\ 0, & \end{cases}$$

and

$$(t-x)_+ = \begin{cases} t-x & if\ x < t, otherwise \\ 0, & \end{cases}$$

the knot, *t*, which forms a reflected pair. Data are separated into two parts for each predictor variable. The knot represents a change in slope or the transition from one polynomial to the next (Friedman, 1991; Muñoz and Felicísimo, 2004; Butte et al., 2010). This process, termed a hinge function or term, is conducted for each predictor variable (*x)* and every possible value of *t*. Thus, all points will be positive for all points to the left or right of the knot, respectively. For example, if the hinge function chosen for the model is (h_mxl – 222), the hinge function will only be used if the humerus diaphyseal length is larger than 222 mm; if the measurement is smaller than 222 mm, the hinge function will be disregarded in the equation.

There are 2*Np* basis functions if all predictor variables are distinct and each basis function takes the form of a constant (i.e. the intercept), a hinge function, or a combination of hinge functions. In contrast to recursive partitioning, basis functions overlap. Where recursive partitioning follows a two-at-a-time backward deletion strategy, MARS employs a one-at-a-time deletion – along with subset selection – through a forward and backward iterative selection to adaptively construct a set of basis functions (Friedman, 1991, 1993). With the inclusion of an

additional variable, the hinge function can occur dependent of the previous knot, only on one side of the previous knot, or independent of the previous knot. The coefficients for each function are then estimated by minimizing the residual sum of squares (Hastie et al., 2009). The knot and variable pair which provides the best fit, based on least squares, is retained in the model (Friedman, 1991; Sekulic and Kowalski, 1992; De Veaux and Ungar, 1994). The forward pass deliberately over-fits the training data and continues until one of the following is met: 1) reached the maximum number of terms for the model; 2) the R-squared changes less than 0.001; 3) the R-squared is greater than or equal to 0.999; 4) generalized R-squared value (explained below) is less than -10; or 5) no new term increases the R-squared value (Friedman, 1993; Muñoz and Felicísimo, 2004; Milborrow, 2013).

The backward pruning pass – employed to remove excess basis functions that no longer contribute to the accuracy of the fit – utilizes the generalized criterion value (GCV) with a goal to produce the most generalizable approximation (Friedman, 1991, 1993). The GCV, an approximation of the prediction error ascertained by a leave-one-out cross-validation, is based on the residual sum-of-squares (RSS) (Milborrow, 2013). Following the pruning pass, the MARS model determines the fitted values, residuals and coefficients by regressing y on the basis matrix following an ordinary least squares regression. The MARS models were performed using the earth package in R (Milborrow, 2013).

A fundamental concept of MARS is that in different areas of multivariate space, variables will have differing degrees of importance to the response surface (Sekulic and Kowalski, 1992). The term ‘adaptive’ in the technique name refers to the capability of the algorithm to choose the dominant variables for each sub-region in multivariate space rather than just interactions with the predictor variable (Sekulic and Kowalski, 1992; De Veaux and Ungar, 1994; Muñoz and Felicísimo, 2004). However, MARS is considered to be vulnerable to multicollinearity (Friedman, 1991; De Veaux and Ungar, 1994; Muñoz and Felicísimo, 2004; Milborrow, 2013). During the forward pruning pass, MARS may arbitrarily (if variables have the same RSS or GCV criterion) choose one variable over another because it is difficult to isolate the contributions and interactions of two highly correlated variables (Friedman, 1991). Since application of PCA creates new uncorrelated variables, the multicollinearity problems associated with MARS are alleviated; generally when data are nonlinear and there are strong inter-variable correlations, MARS with PCA tends to outperform MARS without PCA (De Veaux and Ungar,

1994). MARS conducts automatic variable selection based on the effect of the variable on the response variable, which is averaged for the entire population. MARS models base variable selection on three criteria: the number of model subsets that include the variable; variables that cause for a decreased residual sum of squares; and the GCV criterion, if the GCV increases, the variable has less importance (Milborrow, 2013).

In terms of model selection and goodness of fit, numerous outputs are necessary to consider. earth builds ten-fold cross-validated models on the data. Cross-validation partitions the data into 10 equal subsets, repeatedly builds a model on all but one subset (90%), and the left-out subset is used to measure model performance (10%) (Efron and Tibshirani, 1993; Hastie et al., 2009; Milborrow, 2013). An averaged out-of-fold R-squared (cross-validated R-squared [cv.rsq]) is obtained from the left-out subset and is an estimate of the model performance on independent data (Milborrow, 2013). The GRSq, or generalized R-squared, is based on the raw GCV and is a model assessment or generalization of model performance. Adding terms (or hinge functions) generally always increases the R-squared, but the GRSq may actually lower in regard to the predictive powers and ultimately the generalization to independent data (Milborrow, 2013). Once the MARS model is created it is transformed into a linear model and compared to the best spline model (see below).

MODEL SELECTION: STANDARD ERROR, R-SQUARED, AIC, ANOVA

Age in years was the response variable and bone measurements were the predictor variables for each of the models created, which ranged from simple linear regression with no transformation and without the inclusion of splines to more complex models such as quadratic functions with 3 knots or a MARS model with a cube root transformation of age. First the models created from the same method were compared, such as all the spline models or all of the MARS models, by standard error (SE), *F*-value and RSS in an ANOVA, R-squared and AIC (Wright and London, 2009). Criteria to choose the best MARS model included the SE, cv.rsq and GCV as well as model complexity. Residuals of each model were evaluated to look for any outliers and ensure homoscedasticity. After the chosen MARS model was transformed into a linear regression, the model was compared to the chosen spline model through SE and R-squared

and AIC when possible. Because the models were created with different transformations to age, not all models could be directly compared with AIC, thus the SE and R-squared were the primary values for comparison.

The priority in model selection is seeking the model that yields the best balance between model fit and complexity. Model selection based on the Akaike Information Criterion (AIC) represents the information-theoretic selection approach based on a publication by Kullback and Leibler (K-L) (1951) (Burnham and Anderson, 2004). The K-L chooses a model that loses the least information relative to the additional models being compared. AIC combines estimation and model selection with an unbiased estimator of K-L information and ranks the models from best to worst, based on the lowest and highest AIC, respectively (Burnham and Anderson, 2004; Hastie et al., 2009; Wright and London, 2009; Burnham et al., 2011). Whether one model is significantly better than another model – when comparing with ANOVA – is based on the *F* statistic, which is a test statistic that evaluates the variance of group means divided by the mean of within group variances based on the RSS. AIC is considered superior to sole dependency on the RSS and *F*-values in an ANOVA since the RSS will always decrease with the addition of new variables, no matter how overfit the model is. Essentially, AIC considers the complexity of the model and the goodness of fit of the model.

In regression, the standard error is the ordinary least squares estimate of the standard deviation of the underlying errors and is considered an accuracy check of the model (Efron and Tibshirani, 1993). R-squared is the proportion of the total variability explained by the model. Whereas the SE can be compared across datasets, the R-squared cannot because it does not follow a distribution; however the R-squared can be compared when used on the same dataset (Hawkins, 2004). Generally, the more variability explained by the R-squared, the better the model since R-squared is the square of the correlation between the model's predicted values and the actual values. The square of the correlation ranges from 0 to 1 and the correlation can range from -1 to 1. The greater the magnitude of the correlation between the predicted values and the actual values, the greater the R-squared, regardless of whether the correlation is positive or negative. Kvålseth (1985) devised eight criteria for a good R-squared statistic: 1) a measure of goodness of fit and reasonable interpretation, 2) a dimensionless number, independent of the variables in the model, 3) a number that ranges from 0 to 1, with 1 representing a perfect fit, 4) general enough to be applicable to any type of model, 5) applicable to all model fitting

techniques, 6) its values for different models fitted to the same dataset are comparable, 7) harmonious with other statistics derived from measures of fit (i.e. standard error of prediction), and 8) positive and negative residuals have equivalent weights.

RESAMPLING: PREDICTION INTERVALS

As previously stated in *Chapter 2,* useable age at death estimates require corresponding prediction intervals. The prediction interval considers the variability in the conditional distribution and the error in the conditional mean and takes into account the models capability to predict a future observation; a confidence interval only takes into account the conditional mean of the data and acts as a measure of the accuracy for the parameters of the model (Efron and Tibshirani, 1993). Essentially, the interval estimated for a future prediction, based on the variability in the observed data, is the prediction interval. Because the basis splines are semiparametric and MARS models are nonparametric, applying a parametric prediction interval is invalid as parametric intervals are constructed following assumptions based on a normal distribution, an assumption to which semi- or nonparametric models do not adhere (Lei and Wasserman, 2012). Thus, the prediction intervals associated with basis splines and MARS model need to be calculated indirectly through resampling methods, particularly cross-validation.

Cross-validation provides a realistic estimation of prediction error because a test sample is incorporated that is different from the training sample. To circumvent the problem of collecting another dataset, cross-validation utilizes a portion of the data to fit the model and a portion of the data to test the model, ultimately estimating the average error (Efron and Tibshirani, 1993). Specifically, k-fold cross-validation separates the data into K equal sized parts, for the k^{th} parth the model will be fit to the other $K - 1$ parts and the prediction error of the fitted model will be calculated when predicting the k^{th} part (Efron and Tibshirani, 1993; Kohavi, 1995; Hastie et al., 2009). This is conducted for as many subsets chosen, which in the current study is 10, and then the K estimates of prediction error are combined. Utilizing $K = 10$ allows for a lower variance, but a possible higher bias; however, this is alleviated if the training sample has approximately 100 observations (Efron and Tibshirani, 1993). In the current study, each training set will be a different size because the dataset specific to each variable is different; the largest

univariate samples (i.e. HDL, FDL) will have training sets with more than 100 observations though all of the multivariate subsets may not. Although cross-validation may overestimate the true prediction interval, a tenfold sample is generally recommended as the best compromise for bias and variance (Kohavi, 1995; Hastie et al., 2009). Specific to cross-validation in the MARS models, a cross-validated model is constructed and then predictions are made on the out-of-fold data. The out-of-fold data refers to the 10% holdout sample for each model. This process was repeated 100 times and the 95% predictions intervals will be based on the spread of the predictions.

The primary difference between a nonparametric cross-validated prediction interval and a classic parametric prediction interval is how the variance of the prediction error is estimated.

$$V(x) + v$$

A cross-validation was conducted on each basis spline and MARS model to obtain the variance of the predicted values *(V(x))*; essentially, how much the predicted values – for each observation – vary over 100 iterations of the cross-validated models.

$$Var\big(f(x)\big)\ at\ x$$

However this does not include the noise variance, or *v*, which can be estimated from the RSS of the original basis spline or MARS model divided by the number of observations. The equation for RSS is

$$\sum_{i=1}^{N}(y_1 - \hat{y})^2$$

The cross-validated variance of the prediction error is then used in the prediction interval equation

$$f(x) +/-2\sqrt{V(x) + v}$$

Ultimately, the prediction interval is similar to a standard parametric prediction interval with the difference being how *V(x)* was estimated. In the current study, both 100 and 1000 iterations were conducted; the values did not differ greatly so 100 iterations was chosen.

An age at death table was created for each univariate model. The tables provide a cross-validated 95% prediction interval associated with the cross-validated fitted value for the range of each diaphyseal measurement. For example, if the range of TDL was from 115 mm to 368 mm, the chart would include the range of the measurement – in 1 mm increments – from 115 to 368 mm and each diaphyseal length would have an associated fitted value and 95% upper and lower prediction intervals. An age at death chart was also created for the all-measurement model. However, the data presented in the age at death chart of the all-measurement model are the variable combinations for the individuals in the current study, as the possible number of variable combinations is too large to account for in age at death tables. Thus, the all-measurement age at death table is merely used as an example of the potential success of multivariate models employed for age at death estimation of subadult skeletal remains.

SEX ESTIMATION

Sex estimation required a multitude of classification models to expose the analysis best suited for the data. The classification models consisted of linear, quadratic, and flexible discriminant analysis, logistic regression, naïve Bayes, and random forests. The models utilized the original variables, ratios, PC scores derived from the variance-covariance matrix (VCVM), as well as the y-axis residuals from each of the chosen age estimation basis splines or MARS models. The goal of the inclusion of the PC scores, ratios and residuals was to retain the variation but remove the increase in size as age increases (Jungers et al., 1995).

MEASURING SEXUAL DIMORPHISM

Sympercent (sp) differences were calculated to quantify sexually dimorphic differences per measurement and age in the dataset. Sympercent differences removes size by removing the unit of measure and rather provides a log of a ratio while still retaining the relationship between males and females (Cole, 2000). The sympercent approach is in contrast to the typical way that dimorphism is reported, which is by percent differences. Both sympercent and percent differences account for the drastic differences in size of diaphyseal lengths and breadths. For example, the midshaft measurements may have a range of only 10 mm whereas a diaphyseal length measurement may have a range of 300 mm. Similarly, a 5 mm difference on the midshaft is drastically different from a 5 mm difference on a diaphyseal length and thus the magnitude of each is different. However, percent differences present with signs as well as magnitudes dependent on what variable is chosen as the numerator and denominator (Cole, 2000; Wells, 2012). For example, a group of females may be 7.7% shorter than a group of males, but the males are 8.4% taller than the same group of females. Therefore, percent differences between variables are asymmetrical and not additive (Cole, 2000).

Logarithms convert a ratio to a difference, which corresponds to fractional differences on the original scale of a log-transformed scale, thus solving the problems of additivity and asymmetry. Multiply the logarithmic ratio by 100 and the natural log differences are equivalent to symmetric percent differences (Cole, 2000).

$$\left(100(log_e x_2)\right) - \left(100(log_e x_1)\right) = s\%$$

Because there is no denominator, the sp difference allows for males and females to be symmetrically larger or smaller than one another, which results in a value that is easier to interpret. The sp difference is unique among fractional percentages because it is additive. Furthermore, there is only one way to calculate sp differences, while there are many ways to calculate fractional differences (Cole, 2000).

CLASSIFICATION MODELS

A multitude of different classification methods were employed in an attempt to identify the methods that consistently yielded the highest probability of a correct conclusion. Generally the goal of classification methods is to predict the response variable in future analyses. The process followed the same pattern for all elements, ranging from the most simple univariate situation to a multivariate bone model. Following all univariate and multivariate bone analyses, the classification models were applied to a subset of variables (all-measurement model) that demonstrated statistical significance in the MANOVA, ANOVA, and Student's t-tests results as well as the subsets (i.e. upper limb, proximal elements, etc.). All analyses mentioned within this section apply to the original measurements and most analyses also include the ratios, residuals, and PC scores.

Discriminant analysis (DA) is a method for identifying relationships between qualitative dependent variables and quantitative independent variables with a goal to predict group membership from a set of predictors (Kachigan, 1991; Tabachnick and Fidell, 2007). Essentially, DA classifies the quantitative independent variables (i.e. postcraniometrics) into a qualitative label (i.e. sex). Group membership of the unknown is subsequently based on the discriminant function score and its relative proximity – based on Mahalanobis distance – to the centroids of each reference group, or males and females (Hastie et al., 1993). Mahalanobis distance is a measure of similarity of an unknown to a known reference group based on the covariance among variables and the group's measurement means. The reference group to which the unknown is most similar is the group into which it is classified.

Different types of DA are available for certain data structures; the most popular of these are linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). Fundamental differences between LDA and QDA are apparent on the hyperplane that separates the classes. LDA uses a linear boundary between the classes while the QDA uses a quadratic boundary (Michie et al., 1994; Hastie et al., 2009). Multivariate normality is required for both LDA and QDA and was analyzed through QQ plots (Hastie et al., 1993; Tabachnick and Fidell, 2007). While LDA requires common VCVMs of the different groups in the sample ($p > 0.05$), the QDA is employed if the VCVMs significantly differ ($p < 0.05$) (Hastie et al., 2009).

A major consideration when employing a DA is the sample size in comparison to the number of variables utilized in the model as the estimation of means, variances, and covariances all depend on sample size (Ousley and Jantz, 2012). A suggestion that will be acknowledged during the statistical analyses is that the minimum sample size among groups must be at least three times the number of measurements (3*m*) (Huberty, 1994); a standard that provides estimates less subjective to sampling variation and that achieve more successful classifications (Ousley and Jantz, 2012). DA also has the capability to apply stepwise selection of variables, an approach that removes superfluous variables that should not help classification. The result is a smaller number of predictor variables that discriminates between the groups as well as all of the predictor variables.

Both LDA and QDA provide classification accuracies for each function, indicative of the validity of the result. Within this research, all DA models assumed equal prior probabilities. Therefore, the probability of being classified as a male or female is 50% whereas a proportional prior probability is based on the sample distributions. For the purpose of this study, classification accuracies that were 50% greater than chance, or 75%, were considered practical for use in forensic anthropological analyses. The qda and lda functions are in the MASS package (Venables and Ripley, 2002) and the stepclass function, used for stepwise variable selection, is in the klaR package (Roever et al., 2013).

In addition to LDA and QDA, flexible discriminant analysis (FDA) substitutes a flexible, nonparametric fit in the place of the linear regression in LDA and achieves a more flexible classifier that is motivated by generalized optimal scoring (Hastie et al., 2009; Milborrow, 2013). Essentially, rather than a linear hyperplane as used in LDA, FDA employs a flexible (nonlinear) hyperplane to separate the classes (Hastie et al., 1993, 2009; Milborrow, 2013). FDA is similar to MARS models (see *MARS*, page 62) in that it uses adaptive additive spline regressions and then performs a linear regression in the enlarged space (Hastie et al., 2009). Essentially, the MARS algorithm – which has capability to capture variable interactions in a hierarchical manner – was applied with an adaption to a multiple response variable such that the residual sum of squares and GCV criterion is summed over the number of response variables (Hastie et al., 1993). Generally, the flexible modeling better separates classes, especially those greater than two, and is considered superior to LDA and QDA, demonstrated by consistently higher overall correct classifications (Friedman, 1991; Milborrow, 2013). FDA was conducted using the fda function in

the mda package, which was based on the original S code written by Hastie and Tibshirani (Leisch et al., 2011).

Logistic regression is used for binary predictions and is useful when describing relationships between independent variables and a binary dependent variable without requiring a normal distribution, linear relationships among variables, or equal variances among groups (Kachigan, 1991; Motulsky, 1995; Tabachnick and Fidell, 2007; Hastie et al., 2009). Logistic regression is related to and answers the same questions as DA, however logistic regression is sometimes preferred as it is considered more flexible and has fewer assumptions. Logistic regression is considered especially useful when the predictor variables display a nonlinear relationship (Tabachnick and Fidell, 2007). Similar to most multivariate analyses, logistic regression is extremely sensitive to multicollinearity and the effects are identified as large standard errors in the fitted model (Tabachnick and Fidell, 2007). Furthermore, multicollinearity may erroneously skew the results such that the predictor variables that demonstrate significance when initially evaluated (i.e. t-tests, ANOVA) may demonstrate no significance in a logistic regression.

Rather than a least squares estimate of the parameters, logistic regression adopts a maximum likelihood approach to estimate coefficients through an iterative process (Hastie et al., 2009). Essentially, the goal is to find the best weights for predictors that maximizes the correct prediction (Hair et al., 2010); the parameters are adjusted until the likelihood of the data does not change significantly (Tabachnick and Fidell, 2007). The likelihood function measures the probability of observing the particular set of predictor variables that occur in the sample. The maximum likelihood estimate involves finding the coefficients that results in the log of the likelihood function as large as possible. A measure of model fit using chi-squared is conducted by taking -2 times the log of the likelihood value. Logistic regression directly produces probabilities of group membership, unlike DA.

To understand the fit, the difference between the observed and fitted values needs to be examined. Whereas this would be achieved by the R-squared and the RSS in an ordinary least squares approach, the model generalization for maximum likelihood models is a chi-squared of the residual deviance or a coefficient of determination. The former can be achieved by conducting an ANOVA on the logistic regression model, specifying for the chi-squared test.

Small variation in the residuals of the fitted model along with no systematic tendency is an indicator of the models goodness-of-fit (Hosmer et al., 1997).

McFadden (1974) proposed one of the most common methods cited to calculate a coefficient of determination, or pseudo R-squared, for logistic regression. The formula proposed is analogous to the RSS in ordinary least squares regression, which corresponds to a proportional reduction in error. The McFadden R-squared can be computed as the difference between the initial and model -2 log likelihood statistics, divided by the initial -2 log-likelihood statistic (McFadden, 1974). Following an empirical test of multiple R-squared analgoues, Menard (2000) noted that the McFadden R-squared satisfied all of Kvålseth (1985) criteria for a good R-squared. While evaluating the fit, the sample size should always be taken into consideration; for example, a larger sample size effects the significance of the model, such that the more likely statistical significance will be noted even when it is of no practical importance (Tabachnick and Fidell, 2007). Ideally, the size of each outcome group should be at least 10 times the number of estimated model coefficients; this condition was met in all logistic regression analyses (Hair et al., 2010). The glm package was used for logistic regression models (R Core Team, 2013).

Naïve Bayes is a classifier function that assigns an outcome according to an example following Bayes Theorem, which is based on conditional independence. Unlike the previous methodologies, naïve Bayes reflects a causal relationship (Markov and Russell, 2007). Conditional independence states that all attributes are independent given the class variable, such that X is conditionally independent of Y, only if the probability distribution of X is independent of Y given Z. Essentially, the resulting posterior probability is proportional to the product of the prior and conditional probabilities; the specified class is based on which has the highest posterior probability (Friedman et al., 1997; Bennett, 2000; Malovini et al., 2012). Naïve Bayes has been shown to be asymptotically optimal, or that the model can reach the highest accuracy if provided with a large training set that is consistent and reflective of the true population prior probabilities. Even though naïve Bayes inherently violates the basic assumption of independence, the method is considered a robust and simple classifier that is routinely recognized to predict equally well as other classification methods (Friedman et al., 1997; Zhang, 2004, 2005; Zhang and Su, 2004; Hastie et al., 2009).

The last type of classification model that was employed was random forests, which consists of a number of un-pruned trees (~500) that are independent while all retaining the same

distribution (Breiman, 1999, 2001; Cutler et al., 2007). Un-pruned trees refer to each tree built at its maximal size, which ultimately reduces bias. Each tree is constructed with a bootstrap sample from the training set, termed bagging, an acronym for 'bootstrap aggregating' (Breiman, 1996 a). Bagging is considered a superior process as it enhances the accuracy as well as provides continuous estimates of error, strength, and correlation. A randomized set of predictor variables is utilized concurrently to bagging to find the best split for each node into the children nodes. Consequently, approximately one-third of the variables (i.e., the square root of the number of variables) are excluded from the bootstrap sample and not used in the construction of the k^{th} tree (Cutler and Zhao, 2001; Cutler et al., 2007). This process is based on out-of-bag estimates (OOB), or estimates of generalization error noted above. Once the specified number of trees is created, the forest chooses the class that has the most votes. Because OOB estimates are not used in the fitting of the trees, they represent a cross-validated accuracy estimate that eliminates the need for an independent test sample (Breiman, 1996 b, 2001; Cutler et al., 2007). Outcomes for random forests include a reduction in variation, improvement of the performance of procedures, and unbiased OOB estimates; the process is generally believed to outperform naïve Bayesian models (Breiman, 1996 a, 1999, 2001; Cutler et al., 2007). All random forest analyses were conducted with the randomForest function within the randomForest package (Liaw and Wiener, 2002).

ASSESSING CLASSIFICATION ACCURACY

A bootstrap was conducted on each of the multivariate subsets for each classification method. Specifically, the model for each subset was subject to 1000 iterations. The classification error of the bootstrapped results was averaged and thus provided a more realistic classification error. The resampling was conducted on both the raw measurements and the PC scores using the VCVM. The first PC was removed for all bootstrap analyses to account for the size differences that are inherent with an increase in age. Stepwise selection was conducted prior to resampling for the LDA model, however the process is integral to both the logistic regression and FDA models. Thus, not every model will consider the same variables as significantly contributing to sex estimation.

The bootstrap is a resampling method for the model and statistics within a model and is the most suitable resampling method for smaller sample sizes, which is especially apparent in some of the multivariate subsets. The process consists of generating *B* bootstrap samples, usually 100 or 1000, estimating the model of each *B* sample, and then applying the fitted model to the original samples (Efron, 1979; Efron and Tibshirani, 1993; Michie et al., 1994; Hastie et al., 2009). An important note is that the *B* bootstrapped datasets are the same sample size as the raw data and are sampled with replacement from the training data. Thus, the bootstrapped samples are not identical to the original sample and some observations may be omitted entirely. Approximately 63% of the *B* sample will be unique observations from the original sample, the remaining 37% of the *B* sample will be duplicates of the original sample (Michie et al., 1994; Kohavi, 1995). Bootstrapped estimates do tend to produce optimistic classification accuracies based on the amount the average residual squared error underestimates the true prediction error (Efron and Tibshirani, 1993; Hastie et al., 2009); however, bootstrapped samples are considered comparable to cross-validation results.

MODEL ASSESSMENT

The ideal way to achieve the goals of appropriate model selection as well as model assessment is to test the generated models on an independent sample, after using the original data to train and test the models (Hastie et al., 2009). Essentially, the holdout sample provides a more realistic generalization performance of the chosen models, as the sample was not used as the training sample for model creation. The holdout sample for the current study was an additional sample of Lodox Statscan-generated images of 30 South African children aged between birth and 12 years collected from the Red Cross War Memorial Children's Hospital. Essentially, the holdout sample was to estimate the overall generalization performance of the appropriate age and sex estimation models. No individuals from this sample were included in the original dataset. The sample stems from the same population base and exhibits the same demographics and distributions.

CHAPTER 5: RESULTS

Descriptive statistics, including the sample size, mean and standard deviation of each measurement by age and sex, are located in Tables 5.1 – 5.6. A different sample size was obtained for each measurement because of traumatic injuries and/or inaccurate placement, amongst other factors (Table 5.7). Femoral and tibia breadths were difficult to obtain in the younger individuals (< 1 year) because of natural leg positions, while the humeral, radial, and ulna lengths were difficult to obtain in the older individuals because of active epiphyseal fusion. The small sample sizes of the midshaft breadths of the radius and ulna (n=441 and n=406, respectively) resulted from children not being placed in anatomical position.

In an effort to retain sample sizes and elucidate possible patterns among the diaphyseal dimensions and age and sex, the sample was partitioned into four element type and measurement type subsets, in addition to the bone subsets (i.e. humerus, radius, ulna, femur, tibia, and fibula). The element and measurement type subsets include the upper limbs measurements (n=96), lower limb measurements (n=458), proximal element measurements (n=316), and distal element measurements (n=100). An all-measurement subset was also utilized in the multivariate models for both age and sex estimation. However, the variables in the all-measurement model were different for both age and sex estimation analyses, thus the variables included for each analysis will be described in the associated sections.

MEASUREMENT RELIABILITY

Technical error of measurement (TEM) and relative TEM (%TEM) were conducted for each measurement to evaluate intra- and inter-observer error. From the original dataset, 15 individuals were measured and re-measured by a second observer and the author, respectively. If both left and right sides were acceptable, the left was used; however, both the observer and the author had the option of choosing either left or right-sided elements because a measurement could only be obtained from an element in the proper position. For each of the 15 individuals, as with the original dataset, each measurement obtained a different sample size. The intra-observer

error for all measurements is lower than the inter-observer error for all measurements. The mean intra-observer TEM and %TEM are small at 0.45 mm (0.07 mm to 0.91 mm) and 0.22% (0.02% to 0.83%), respectively (Table 5.8). The mean inter-observer TEM and %TEM were similarly low at 0.76 mm and 0.40% (Table 5.8), but as would be expected the ranges were slightly wider for both the TEM (0.02 mm to 1.7 mm) and %TEM (0.02% to 2.3%). The measurements with the highest inter-observer %TEM were midshaft breadths of the ulna and radius whereas the measurements with the highest intra-observer %TEM were the ulna midshaft breadth and radius proximal breadth.

Based on the Bland-Altman plot, neither the intra- or inter-observer error show systematic bias. Most measurement differences are within 2 mm on the intra-observer error Bland-Altman plot (Figure 5.1). Generally, the inter-observer has the majority of measurements within upper and lower agreements of 2 mm, though the spread of differences is larger than the intra-observer error spread of differences (Figure 5.2).

Table 5.1 – Summary statistics separated by age and sex for the measurements associated with the humerus.

Females

Age (years)	HDL			HPB			HDB			HMSB		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	32	89.73	16.24	23	18.69	3.46	17	23.16	3.30	22	8.59	2.17
1	26	120.53	10.92	13	25.62	2.37	14	30.31	2.34	20	12.09	1.92
2	38	138.33	10.47	24	25.89	2.45	24	32.71	2.24	32	12.45	1.01
3	39	152.04	12.22	27	27.67	1.76	27	35.51	2.55	36	13.28	1.08
4	38	164.39	11.51	33	28.85	2.38	23	36.36	3.07	36	13.57	1.20
5	24	178.43	11.50	19	30.56	1.81	15	37.42	3.64	23	14.10	1.42
6	34	192.60	10.54	21	33.02	1.94	22	40.87	2.48	30	14.70	1.49
7	36	202.62	10.99	31	32.43	2.69	27	40.40	3.73	32	14.66	1.59
8	36	217.11	16.46	28	34.05	2.89	21	42.29	3.34	33	15.09	1.39
9	31	227.91	14.37	20	35.44	2.25	18	44.15	3.97	28	15.49	1.77
10	27	238.64	16.98	20	36.00	2.49	19	48.14	5.79	28	16.37	2.03
11	23	239.14	18.99	30	37.53	2.50	17	47.95	5.38	26	16.69	1.42
12	8	255.18	14.96	15	40.42	3.68	9	52.20	2.12	15	18.08	2.25

Males

Age (years)	HDL			HPB			HDB			HMSB		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	31	85.72	14.39	25	18.38	3.66	25	24.33	3.55	24	8.61	1.86
1	44	121.85	9.65	24	26.02	2.67	32	31.66	2.35	36	12.43	1.54
2	77	137.12	8.74	47	26.98	2.18	51	33.84	2.37	68	12.82	1.20
3	53	150.80	11.22	43	28.87	2.18	31	35.62	3.17	42	13.59	1.39
4	50	163.83	12.91	42	30.39	2.37	34	37.34	2.97	48	13.97	1.60
5	52	181.29	11.97	33	32.79	2.27	36	39.87	2.76	48	14.76	1.56
6	60	194.05	13.30	48	33.72	2.39	37	40.89	2.65	51	15.07	1.44
7	55	201.79	12.57	46	34.41	2.82	34	42.25	3.33	54	15.60	1.51
8	58	216.21	14.29	50	36.67	3.05	41	44.02	3.44	53	16.46	1.56
9	55	225.57	13.27	38	37.65	2.67	35	46.15	2.96	45	16.59	1.58
10	57	233.15	15.91	44	38.99	2.68	40	48.03	4.41	51	17.09	1.55
11	50	242.26	17.90	41	39.14	2.67	44	48.96	5.30	50	17.30	1.75
12	37	250.37	18.28	29	40.48	4.18	20	49.45	3.35	34	17.61	2.13

Table 5.2 – Summary statistics separated by age and sex for the four measurements associated with the radius.

Females

Age (years)	RDL			RPB			RDB			RMSB		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	29	72.10	11.69	21	8.22	1.52	21	12.48	1.53	16	5.41	1.17
1	25	94.03	7.89	10	10.88	0.83	12	17.16	1.62	9	8.54	0.77
2	38	106.74	8.64	15	11.09	0.67	17	17.77	1.46	9	8.09	0.45
3	43	118.25	10.23	20	11.79	0.93	18	18.42	1.43	17	8.59	0.99
4	41	124.75	8.60	21	12.56	0.90	14	19.29	1.57	15	9.38	0.83
5	32	137.15	8.74	10	13.23	0.98	17	20.47	1.30	8	10.04	1.18
6	34	147.38	9.70	8	14.39	1.30	17	21.91	1.28	8	10.75	0.55
7	38	152.48	8.59	18	14.28	0.74	21	22.12	1.53	18	10.88	0.95
8	42	163.88	13.53	20	15.18	1.27	23	22.59	1.57	20	11.02	1.09
9	33	172.62	11.37	12	15.23	1.17	17	24.23	1.80	8	10.82	1.12
10	35	181.49	15.29	13	16.41	0.86	18	24.88	2.68	12	12.65	1.86
11	33	184.98	16.38	18	17.21	1.27	20	25.56	1.90	17	12.14	1.40
12	16	203.79	16.44	10	19.33	2.28	7	27.42	2.00	9	13.52	2.00

Males

Age (years)	RDL			RPB			RDB			RMSB		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	32	69.15	11.52	17	8.37	1.28	17	12.96	1.83	14	5.87	1.07
1	43	96.46	7.03	22	10.90	1.05	28	17.10	1.63	19	7.86	1.06
2	78	107.49	6.40	33	11.40	0.74	34	18.47	1.46	28	8.42	0.69
3	63	117.30	8.18	29	12.10	0.84	32	19.14	1.47	28	8.82	0.93
4	55	126.89	8.93	22	13.33	1.13	24	20.64	1.55	18	9.44	0.88
5	63	137.78	9.27	21	13.94	1.08	18	21.32	1.39	19	9.81	0.83
6	69	148.73	10.89	31	14.65	0.93	30	22.17	1.37	24	10.24	0.88
7	67	154.51	10.90	29	14.84	1.23	33	23.30	1.96	25	11.21	1.11
8	59	164.22	11.98	35	16.22	1.47	31	24.61	2.28	27	11.86	1.32
9	56	172.34	9.05	25	16.57	1.50	30	25.38	2.13	20	12.16	1.57
10	63	178.85	14.34	26	17.41	1.21	25	25.94	1.77	22	12.94	1.11
11	55	187.02	15.36	22	17.35	1.21	27	26.42	2.13	18	12.46	1.23
12	41	194.56	13.78	15	18.90	1.72	19	27.37	2.19	13	12.91	1.21

Table 5.3 – Summary statistics separated by age and sex for the two measurements associated with the ulna.

Age (years)	Females					
	UDL			UMSB		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	27	81.77	13.42	16	5.69	1.29
1	25	107.32	8.88	9	8.15	1.02
2	37	120.49	9.27	9	7.69	0.60
3	41	133.04	11.25	16	8.01	0.82
4	40	139.51	8.84	15	8.73	0.83
5	29	151.80	9.75	7	9.24	1.12
6	32	163.84	10.50	7	9.79	0.66
7	34	167.65	8.03	17	9.75	0.63
8	32	181.65	14.41	16	10.12	0.98
9	25	189.15	12.54	7	10.41	1.33
10	16	197.21	17.65	6	11.35	1.23
11	14	193.51	14.07	10	10.74	0.91
12	6	208.38	9.85	3	13.01	3.02

Age (years)	Males					
	UDL			UMSB		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	31	77.56	12.71	14	5.63	1.10
1	42	107.59	6.91	19	7.87	1.22
2	77	119.60	6.77	27	8.02	0.79
3	62	130.17	9.16	28	8.54	0.73
4	50	140.46	9.53	18	9.12	0.93
5	59	152.19	9.90	20	9.68	1.18
6	66	163.40	10.99	24	10.09	1.28
7	61	169.17	10.72	25	10.63	1.03
8	58	179.26	13.04	28	11.21	1.22
9	49	188.17	9.46	20	11.44	1.73
10	55	193.58	13.75	21	11.86	1.09
11	43	201.05	15.30	17	11.81	1.45
12	22	205.86	13.74	7	11.38	0.80

Table 5.4 – Summary statistics separated by age and sex for the three measurements associated with the femur.

	Females								
Age (years)	**FDL**			**FDB**			**FMSB**		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	27	115.33	23.16	3	11.99	1.27	0	NA	NA
1	28	158.17	14.65	14	13.32	1.06	10	43.28	5.05
2	35	190.39	15.17	30	14.64	1.40	26	47.30	3.47
3	40	214.13	18.08	36	15.86	1.71	31	51.01	3.75
4	39	229.32	16.19	35	16.43	1.80	35	53.63	4.73
5	31	253.04	15.94	29	17.45	1.17	27	55.08	2.57
6	36	273.89	15.67	35	18.43	1.63	34	58.33	2.91
7	39	286.87	15.13	36	18.47	1.44	33	58.75	3.26
8	41	310.80	24.22	41	19.89	1.83	40	62.60	4.27
9	35	325.09	22.16	35	20.49	1.87	31	65.00	4.57
10	33	344.41	23.85	31	21.41	1.94	29	67.55	5.10
11	33	351.37	27.28	33	22.21	2.43	29	69.73	5.10
12	14	374.10	16.99	16	23.29	2.84	15	69.95	5.79

	Males								
Age (years)	**FDL**			**FDB**			**FMSB**		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	18	109.80	24.00	2	12.04	0.21	1	41.85	NA
1	33	160.83	15.06	14	13.96	1.35	9	46.67	4.32
2	69	185.49	11.84	41	15.05	1.60	25	50.42	3.79
3	62	208.09	15.44	52	15.83	1.36	46	53.23	3.23
4	54	229.23	18.80	43	16.82	1.64	43	56.21	3.69
5	56	252.36	16.37	49	17.38	1.36	44	58.51	4.00
6	61	274.36	17.46	53	18.72	1.40	48	61.60	4.19
7	60	289.81	19.29	56	19.43	2.10	47	63.06	4.22
8	61	305.95	19.90	60	20.79	2.12	49	65.80	4.16
9	56	323.61	17.56	54	21.61	1.86	44	68.02	4.13
10	59	338.02	23.87	55	21.88	1.79	42	70.48	5.06
11	59	352.79	24.30	55	23.16	2.51	44	72.53	5.51
12	41	364.53	26.08	37	23.54	2.73	32	73.49	6.59

Table 5.5 – Summary statistics separated by age and sex for the four measurements associated with the tibia.

Females

Age (years)	TDL			TPB			TDB			TMSB		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	20	97.38	19.34	0	NA	NA	1	24.52	NA	0	NA	NA
1	26	130.46	12.96	10	35.86	3.96	9	25.53	3.26	9	12.85	1.58
2	37	154.67	13.46	22	39.98	2.52	21	26.78	1.50	19	13.46	1.37
3	35	176.55	16.39	26	43.32	3.13	23	29.37	2.56	20	14.51	1.47
4	40	187.59	14.84	30	46.25	3.20	25	30.52	2.50	21	15.27	1.24
5	30	207.15	13.97	21	46.82	2.20	21	31.26	1.50	20	15.67	1.28
6	36	225.19	17.13	30	50.39	2.41	27	34.51	2.11	25	18.02	1.73
7	37	236.53	13.83	31	50.45	3.41	28	34.98	2.77	26	17.23	1.71
8	42	258.80	25.58	38	53.58	3.90	32	37.23	3.39	30	18.74	2.16
9	37	269.88	21.96	26	55.66	4.48	22	39.01	3.90	25	18.66	1.79
10	31	284.57	22.10	25	58.59	4.51	18	40.93	3.49	16	20.73	2.06
11	37	293.52	26.66	32	60.87	4.07	22	42.01	2.98	24	20.78	1.66
12	14	318.57	20.96	15	61.82	4.59	8	43.68	4.04	9	20.93	1.52

Males

Age (years)	TDL			TPB			TDB			TMSB		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	14	91.05	19.08	1	35.43	NA	1	24.17	NA	1	11.63	NA
1	27	134.40	12.33	10	37.90	3.97	8	25.95	3.01	7	13.01	1.07
2	62	153.04	10.53	32	41.81	3.39	28	27.91	2.03	29	14.42	1.46
3	55	172.39	14.40	36	44.57	2.62	38	29.42	1.47	33	14.84	1.03
4	52	187.28	13.64	38	48.37	2.63	33	31.26	2.30	29	15.50	1.10
5	52	204.30	15.04	41	50.01	3.25	35	32.72	2.23	34	16.44	1.56
6	59	223.52	16.09	45	51.89	3.58	41	34.82	2.49	34	17.47	1.66
7	59	236.95	18.85	47	53.98	3.58	43	36.38	2.83	37	18.26	1.71
8	58	252.22	20.71	46	56.90	3.66	46	38.77	3.18	34	19.32	1.81
9	49	266.45	14.74	39	58.77	4.55	26	40.03	2.53	21	20.66	1.47
10	54	281.50	19.79	41	60.50	4.46	31	41.99	3.24	25	20.95	2.10
11	53	291.58	24.18	39	61.74	4.38	31	42.60	3.37	23	21.33	1.81
12	35	303.60	26.37	29	63.71	6.50	21	44.51	4.75	15	22.78	3.01

Table 5.6 – Summary statistics separated by age and sex for the fibula diaphyseal length.

Age (years)	Females - FBDL			Males - FBDL		
	n	*Mean*	*St. Dev.*	*n*	*Mean*	*St. Dev.*
<1	27	81.77	13.42	16	5.69	1.29
1	25	107.32	8.88	9	8.15	1.02
2	37	120.49	9.27	9	7.69	0.60
3	41	133.04	11.25	16	8.01	0.82
4	40	139.51	8.84	15	8.73	0.83
5	29	151.80	9.75	7	9.24	1.12
6	32	163.84	10.50	7	9.79	0.66
7	34	167.65	8.03	17	9.75	0.63
8	32	181.65	14.41	16	10.12	0.98
9	25	189.15	12.54	7	10.41	1.33
10	16	197.21	17.65	6	11.35	1.23
11	14	193.51	14.07	10	10.74	0.91
12	6	208.38	9.85	3	13.01	3.02

Table 5.7 – Sample sizes for each measurement.

Measurement	n
HDL	1071
HPB	814
HDB	713
HMSB	965
RDL	1183
RPB	523
RDB	570
RMSB	441
UDL	1033
UMSB	406
FDL	1120
FDB	814
FMSB	945
TDL	1051
TPB	750
TDB	639
TMSB	566
FBDL	1031

Table 5.8 – TEM and %TEM for inter-observer error and intra-observer error.

	Inter-observer Error		Intra-observer Error	
	TEM	*%TEM*	*TEM*	*%TEM*
HDL	1.68	0.08	0.66	0.04
HPB	0.80	0.24	0.63	0.22
HDB	0.68	0.58	0.24	0.16
HMSB	0.30	0.22	0.29	0.20
UMXL	1.18	0.10	0.91	0.07
UMSB	0.66	1.92	0.30	0.84
RDL	1.15	0.06	0.35	0.02
RPB	0.23	0.49	0.26	0.75
RDB	0.02	0.04	0.25	0.44
RMSB	0.79	2.27	0.19	0.52
FDL	1.58	0.06	0.91	0.03
FDB	0.62	0.13	0.65	0.19
FMSB	0.15	0.09	0.16	0.12
TDL	0.87	0.06	0.70	0.04
TPB	0.40	0.16	0.29	0.09
TDB	0.81	0.49	0.19	0.21
TMSB	0.26	0.56	0.07	0.16
FBDL	0.48	0.02	0.71	0.03
Min	0.02	0.02	0.07	0.02
Max	1.70	2.27	0.91	0.84
Mean	0.76	0.40	0.45	0.22

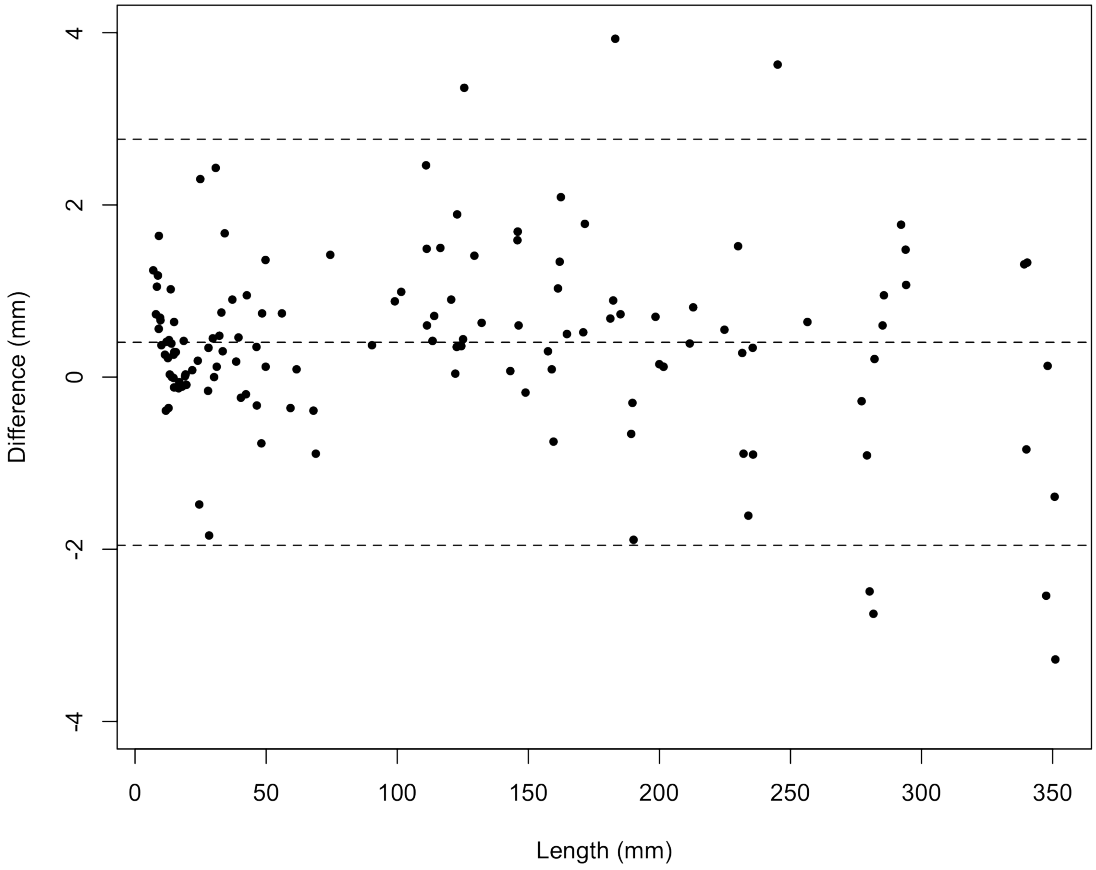


Figure 5.1 – Bland-Altman plot illustrating the inter-observer agreement between the same measurements obtained during two separate observations of 15 individuals. The dashed lines indicate the upper and lower agreement levels and are based on the standard deviation.

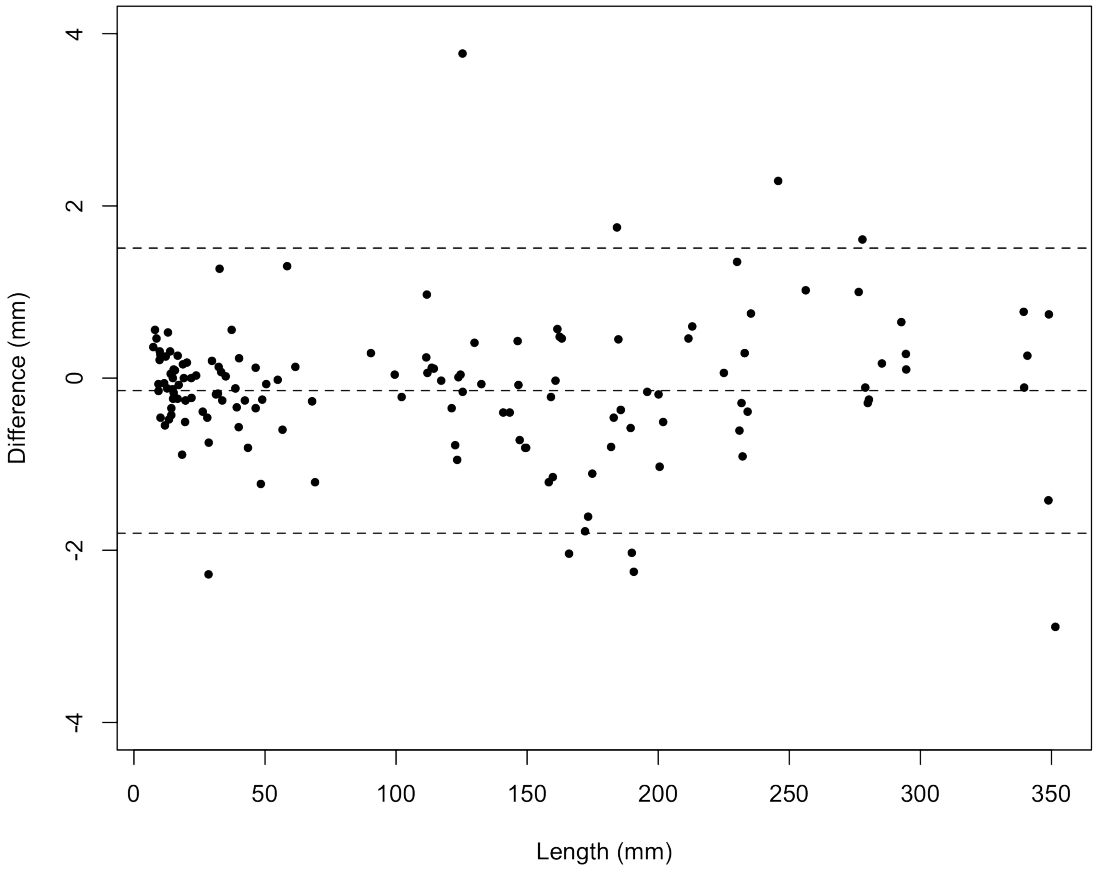


Figure 5.2 – Bland-Altman plot illustrating the intra-observer agreement between the same measurements obtained during two separate observations of 15 individuals. The dashed lines indicate the upper and lower agreement levels and are based on the standard deviation.

STUDENT'S T-TEST, CORRELATIONS, ANOVAS AND MANOVAS

Numerous exploratory analyses were conducted to provide insight into the relationships between the diaphyseal dimensions and age and sex. Specifically, the analyses included Student's t-tests, correlations, ANOVAs and MANOVAs. Student's t-tests were conducted to evaluate statistically significant differences between the sexes for each measurement and each age interval. The Holm's adjustment was included to account for multiple tests and the likelihood of increased type 1 errors, or the conclusion that a difference exists though it does not. Mean differences between male and females per measurement for each age and whether the difference was significant are noted in Tables 5.9 – 5.13. Raw measurements of the humerus and femur demonstrated the greatest number of statistically significant differences between the sexes when compared to the other four skeletal elements. The greatest number of original measurement-age datasets, or the combination of a specific measurement and age, with significant differences between the sexes were HPB, FDB, and TPB. Ratios demonstrated more significant differences between sexes than the original measurements, specifically the ratios that included the most sexually dimorphic original variables (i.e. FDDB, TDPB and HDPB). The disparities in the male and female ratios were apparent at one year of age and persisted through 12 years, however the differences were smaller at the extreme ages. Not all ratios demonstrated statistical significance; no age in either UMXMS or RMXMS subsets demonstrated significant differences between the sexes. Diaphyseal length measurements did not demonstrate statistically significant sex differences for any bone-age datasets and midshaft breadths displayed the fewest statistically significant sex differences among the breadth measurements for all bones and ages.

Table 5.9 – Differences (mm) of the midshaft measurements. Female measurements were subtracted from male measurements, thus a negative number indicates the mean female dimension is larger than the mean male dimension.

Age (years)	HMSB	RMSB	UMSB	FMSB	TMSB
<1	0.02	0.46	-0.06	0.05	NA
1	0.34	-0.68	-0.28	0.64	0.16
2	0.37	0.33	0.33	0.41	0.96
3	0.31	0.23	0.53	-0.03	0.33
4	0.4	0.06	0.39	0.39	0.23
5	0.66	-0.23	0.44	-0.07	0.77
6	0.37	-0.51	0.3	0.29	-0.55
7	0.94	0.33	**0.88***	0.96	1.03
8	**1.37***	0.84	**1.09***	0.9	0.58
9	1.1	1.33	1.03	**1.12***	2
10	0.72	0.29	0.51	0.47	0.22
11	0.61	0.32	1.08	0.95	0.55
12	-0.47	-0.61	-1.63	0.25	1.85

Table 5.10 – Differences (mm) of distal breadth measurements. Female measurements were subtracted from male measurements, thus a negative number indicates the mean female dimension is larger than the mean male dimension.

Age (years)	HDB	RDB	FDB	TDB
<1	-4.01	-2.95	NA	-6.33
1	1.32	2.43	3.39	3.94
2	-1.21	0.75	**3.12***	-1.63
3	-1.24	-0.95	2.22	-4.16
4	-0.56	**2.14***	2.58	-0.31
5	2.86	0.63	**3.43***	**-2.85**
6	1.45	1.35	**3.27***	-1.67
7	-0.83	2.03	**4.31***	0.42
8	-0.9	**0.34***	**3.2***	-6.58
9	-2.34	-0.28	3.02	-3.43
10	-5.49	-2.64	2.93	-3.07
11	3.12	2.04	2.8	-1.94
12	**-4.81**	-9.23	3.54	-14.97

Table 5.11 – Differences (mm) between maximum length measurements. Female measurements were subtracted from male measurements, thus a negative number indicates the mean female dimension is larger than the mean male dimension.

Age (years)	HDL	RDL	UDL	FDL	TDL	FBDL
<1	-4.01	-2.95	-4.21	-5.53	-6.33	-1.11
1	1.32	2.43	0.27	2.66	3.94	4.4
2	-1.21	0.75	-0.89	-4.9	-1.63	-2.84
3	-1.24	-0.95	-2.87	-6.03	-4.16	-4.77
4	-0.56	2.14	0.96	-0.1	-0.31	-0.09
5	2.86	0.63	0.39	-0.68	-2.85	-2.02
6	1.45	1.35	-0.44	0.47	-1.67	-1.69
7	-0.83	2.03	1.52	2.94	0.42	2.44
8	-0.9	0.34	-2.39	-4.85	-6.58	-3.8
9	-2.34	-0.28	-0.98	-1.48	-3.43	-1.79
10	-5.49	-2.64	-3.63	-6.39	-3.07	-6.3
11	3.12	2.04	7.54	1.42	-1.94	1.13
12	-4.81	-9.23	-2.52	-9.57	-14.97	-11.93

Table 5.12 – Differences (mm) of proximal breadth measurements. Female measurements were subtracted from male measurements, thus a negative number indicates the mean female dimension is larger than the mean male dimension.

Age (years)	HPB	RPB	TPB
<1	-0.31	0.15	NA
1	0.4	0.02	2.04
2	1.09	0.31	**1.83***
3	**1.2***	0.31	1.25
4	1.54	0.77	**2.12***
5	**2.23***	0.71	3.19
6	0.7	0.26	1.5
7	**1.98***	0.56	**3.53***
8	**2.62***	1.04	**3.32***
9	**2.21***	**1.34***	**3.11***
10	**2.99***	**1***	1.91
11	1.61	0.14	0.87
12	0.06	-0.43	1.89

Table 5.13 – Differences (mm) of the ratio measures. Female measurements were subtracted from male measurements, thus a negative number indicates the mean female dimension is larger than the mean male dimension. Ratios that did not demonstrate significant differences were excluded from the table.

Age (years)	HDPB	HDDB	HDMS	RDPB	RDDB	FDDB	FDMS	TDPB	TDDB	TDMS
<1	0.01	0.02	0	0.01	0.01	NA	0	NA	0	NA
1	0	0.01	0	0	0	0.01	0	0	0	0
2	0.01	**0.01***	0	0	0	**0.02***	0	**0.01***	**0.01***	0
3	0	0.01	0	0	0.01	**0.01***	0.01	0.01	0	0
4	**0.01***	0.01	0.01	0.01	**0.01***	**0.01***	0	**0.01***	0.01	0
5	0.01	0.01	0	0	0.01	**0.01***	0	**0.01***	**0.01***	0
6	0	0	0	0	0	**0.01***	0	**0.01***	0.01	0
7	**0.01***	0.01	0.01	0.01	0.01	**0.02***	0.01	**0.01***	0	0.01
8	**0.01***	0.02	**0.01***	**0.01***	**0.01***	**0.01***	**0.01***	**0.02***	0	0.01
9	**0.02***	**0.01**	0	0.01	0.01	**0.01***	**0.01***	**0.01***	0	**0.01***
10	**0.02***	0.01	0	**0.01***	0	**0.01***	0	**0.01***	0.01	0
11	0	0.01	0	0	0	**0***	**0.01***	0	0.01	0
12	0.01	0	0	0.01	0.01	**0.02***	0	0.02	0.01	0

Correlations, inclusive of sexes, population groups and ages, were conducted on the original measurements (Table 5.14). The results included a Holm's adjustment to significance figures to account for multiple comparisons among variables. The correlation coefficients were greater than 0.71 in all of the relationships between measurement and age, indicative of a strong, positive relationship. The correlation coefficients between all diaphyseal lengths and age were greater than 0.92, while RPB, RDB, FMSB, TDB, and TPB demonstrated correlation coefficients greater than 0.85. The remaining correlation coefficients were less than 0.85. All inter-variable correlation coefficients were greater than 0.73. The high correlations indicate multicollinearity, a factor recognized to compromise statistical analyses when bivariate relationships are greater than 0.70, and especially when greater than 0.90 (Neter et al., 1985; Tabachnick and Fidell, 2007). Figure 5.3 provides a graphical display of the inter-variable correlations and correlations between age and measurements with shape and color. The narrower eclipses indicate stronger relationships; a straight line – as on the diagonal – represents a perfect correlation ($r = 1.0$). The color scale is located at the bottom of the triangle with red indicating a negative correlation and blue indicating a positive relationship. The ellipse was removed if the p-value was not significant ($p > 0.05$).

Scatterplots of measurements regressed on age was indicative of a nonlinear relationship, similar to a biological growth curve. This was expected considering the diaphyseal dimensions, particularly lengths, exhibit different growth velocities during the growth process. In contrast, linear relationships were apparent between the postcraniometric variables. An example of the patterns can be observed in Figure 5.4.

ANOVAs were conducted to assess the relationship between each measurement and age, sex, and the interaction of age and sex, whereas a MANOVA was used to test mean differences among the multivariate subsets. Multicollinearity is recognized to diminish the power of MANOVA, which is also why ANOVAs were employed in the analysis. Quantile-quantile (QQ) plots showed that the variables generally follow multivariate normality; however, a few outliers departed from normality in the tail. The Levene's test, which was applied to all univariate measures to test for equal variance, demonstrated non-significant p-values (> 0.05) for all variables, and thus an indication of equal variances. Because all of the measurement and element type subsets potentially violate normality in the MANOVA, a Pillai's trace test was utilized as it

is recognized as the most robust generalization of the univariate *F* statistic when MANOVA assumptions are not met (Olson, 1976).

Interpretation of the ANOVA results indicated statistical significance ($p < 0.05$) for the diaphyseal length measurements of the six long bones for all three factors – age, sex, and the interaction of age and sex. Results specific to the breadth measurements indicated statistical significance ($p < 0.05$) for age and sex but not the interaction of age and sex (Appendix II). MANOVA results indicated significant relationships with both sex and age ($p < 0.05$) for all bones and multivariate subsets (Table 5.15). However, only two subsets demonstrated significant relationships with the interaction of age and sex and included the femur and distal element subset ($p = 0.01$ and 0.05, respectively).

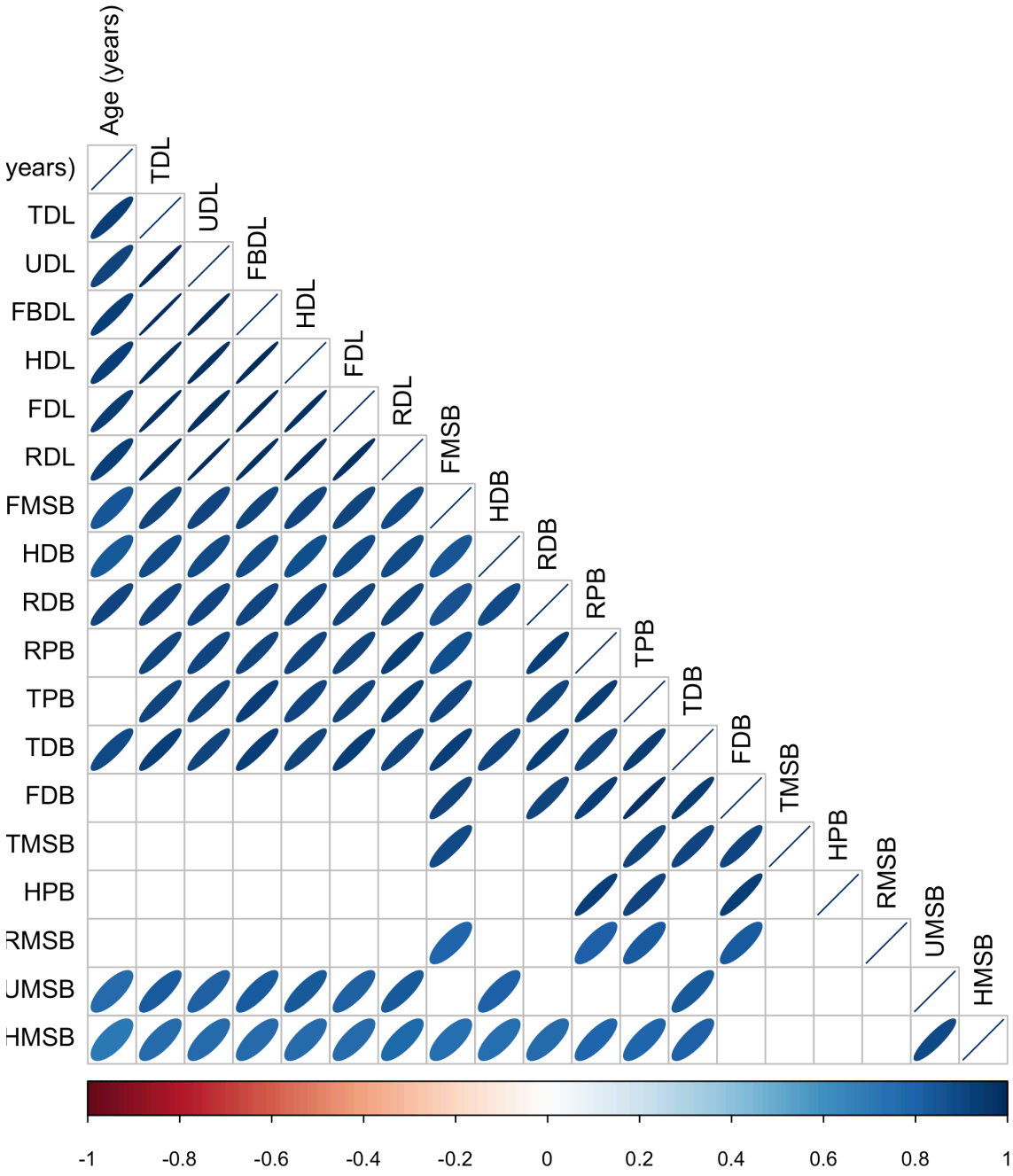


Figure 5.3 – Visualization for the inter-variable correlations and the correlations between the measurements and age. The ellipses were removed if the p-value was not significant ($p > 0.05$).

Table 5.14 – Correlation matrix with a Holm's adjustment. All measurements are statistically significant ($p < 0.05$).

	Age	HDL	HPB	HDB	HMSB	RDL	RPB	RDB	RMSB	UDL	UMSB	FDL	FDB	FMSB	TDL	TPB	TDB	TMSB	FBDL
Age	1	0.924	0.798	0.834	0.709	0.921	0.861	0.903	0.837	0.917	0.76	0.944	0.844	0.849	0.925	0.858	0.888	0.817	0.93
HDL	0.924	1	0.879	0.878	0.76	0.983	0.907	0.907	0.836	0.983	0.836	0.988	0.882	0.905	0.989	0.902	0.913	0.853	0.99
HPB	0.798	0.879	1	0.834	0.854	0.898	0.932	0.884	0.816	0.88	0.871	0.883	0.925	0.849	0.894	0.909	0.915	0.84	0.891
HDB	0.834	0.878	0.834	1	0.753	0.886	0.844	0.882	0.837	0.885	0.805	0.88	0.844	0.857	0.888	0.864	0.901	0.813	0.887
HMSB	0.709	0.76	0.854	0.753	1	0.775	0.796	0.766	0.862	0.762	0.894	0.768	0.785	0.731	0.767	0.785	0.805	0.809	0.766
RDL	0.921	0.983	0.898	0.886	0.775	1	0.92	0.914	0.844	0.995	0.835	0.98	0.906	0.898	0.987	0.926	0.917	0.859	0.989
RPB	0.861	0.907	0.932	0.844	0.796	0.92	1	0.937	0.801	0.903	0.831	0.916	0.939	0.872	0.915	0.941	0.915	0.873	0.915
RDB	0.903	0.907	0.884	0.882	0.766	0.914	0.937	1	0.873	0.901	0.837	0.919	0.91	0.86	0.905	0.912	0.933	0.86	0.904
RMSB	0.837	0.836	0.816	0.837	0.862	0.844	0.801	0.873	1	0.832	0.863	0.852	0.823	0.791	0.825	0.827	0.853	0.872	0.834
UDL	0.917	0.983	0.88	0.885	0.762	0.995	0.903	0.901	0.832	1	0.818	0.98	0.894	0.904	0.99	0.919	0.918	0.854	0.991
UMSB	0.76	0.836	0.871	0.805	0.894	0.835	0.831	0.837	0.863	0.818	1	0.815	0.848	0.821	0.822	0.84	0.839	0.82	0.827
FDL	0.944	0.988	0.883	0.88	0.768	0.98	0.916	0.919	0.852	0.98	0.815	1	0.893	0.915	0.989	0.919	0.925	0.87	0.991
FDB	0.844	0.882	0.925	0.844	0.785	0.906	0.939	0.91	0.823	0.894	0.848	0.893	1	0.912	0.891	0.974	0.948	0.903	0.893
FMSB	0.849	0.905	0.849	0.857	0.731	0.898	0.872	0.86	0.791	0.904	0.821	0.915	0.912	1	0.915	0.908	0.93	0.888	0.919
TDL	0.925	0.989	0.894	0.888	0.767	0.987	0.915	0.905	0.825	0.99	0.822	0.989	0.891	0.915	1	0.919	0.922	0.857	0.997
TPB	0.858	0.902	0.909	0.864	0.785	0.926	0.941	0.912	0.827	0.919	0.84	0.919	0.974	0.908	0.919	1	0.94	0.918	0.92
TDB	0.888	0.913	0.915	0.901	0.805	0.917	0.915	0.933	0.853	0.918	0.839	0.925	0.948	0.93	0.922	0.94	1	0.909	0.92
TMSB	0.817	0.853	0.84	0.813	0.809	0.859	0.873	0.86	0.872	0.854	0.82	0.87	0.903	0.888	0.857	0.918	0.909	1	0.863
FBDL	0.93	0.99	0.891	0.887	0.766	0.989	0.915	0.904	0.834	0.991	0.827	0.991	0.893	0.919	0.997	0.92	0.92	0.863	1

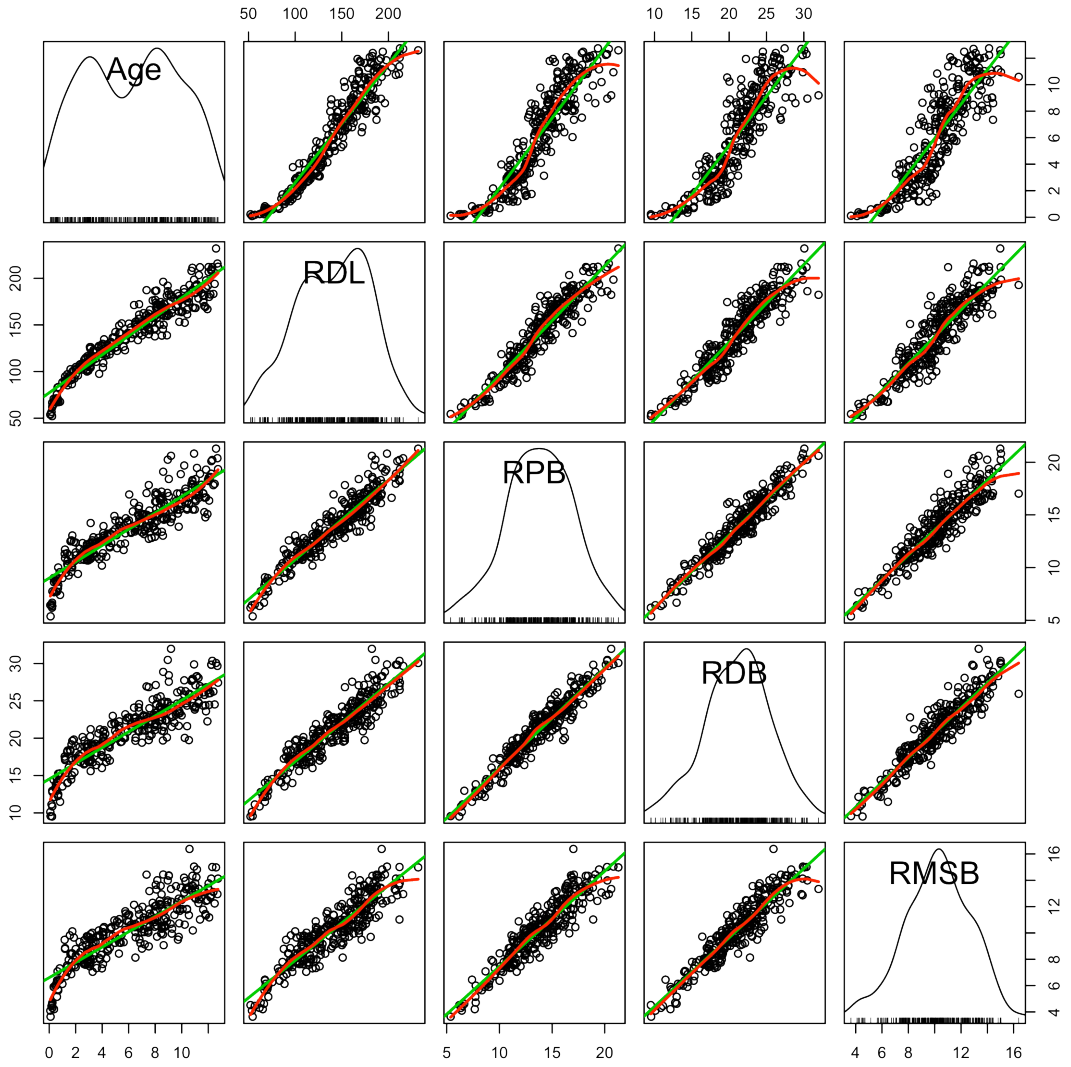


Figure 5.4 – Scatterplot matrix of age and the four measurements obtained from the radius. The green line is the linear regression line and the red line is the loess line.

Table 5.15 – MANOVA conducted on the different subsets of measurements. Bold indicates significance.

	Sex		**Age**		**Sex*Age**	
	Pillai	*Pr(>F)*	*Pillai*	*Pr(>F)*	*Pillai*	*Pr(>F)*
Humerus	0.15	**<0.001*** **	0.908	**<0.001*** **	0.007	0.45
Radius	0.123	**<0.001*** **	0.912	**<0.001*** **	0.011	0.57
Ulna	0.159	**<0.001*** **	0.899	**<0.001*** **	0.011	0.12
Femur	0.240	**<0.001*** **	0.894	**<0.001*** **	0.015	**0.01** **
Tibia	0.160	**<0.001*** **	0.880	**<0.001*** **	0.011	0.2
Upper Limbs	0.916	**<0.001*** **	0.324	**<0.001*** **	0.043	0.954
Lower Limbs	0.903	**<0.001*** **	0.236	**<0.001*** **	0.016	0.4
Proximal Elements	0.895	**<0.001*** **	0.279	**<0.001*** **	0.021	0.48
Distal Elements	0.880	**<0.001*** **	0.325	**<0.001*** **	0.195	**0.05***

***p < 0.01; **p<0.001; ***p<0.0001**

AGE ESTIMATION

In an effort to mitigate heteroscedasticity, especially observed in the residuals, the response variable – age – required a square root (sqrt) or cube root (cbrt) transformation (Figure 5.5). Because the response variable (age) required a transformation, the standard errors also reflected the transformation; however, the SE's presented herein are simply in years.

The single predictor variables, the multivariate bone, measurement and element type subsets, and all-measurement subset were used in a multitude of methods to identify the best fit

for the data. Measurements were subject to a linear model with and without a transformation of age; linear models with different basis spline combinations with and without the transformation of age; and a MARS model with and without a transformation of age. In order to compare the MARS models to the basis spline models, the MARS models were converted into linear models. The model that presented with the highest R-squared, lowest standard error (SE), and lowest Akaike Information Criterion (AIC) was chosen as the most suitable for each particular predictor variable(s). The simpler of the two models was chosen if the models were similar. For each measurement(s), 95% prediction intervals were created through cross-validation.

UNIVARIATE AGE AT DEATH MODELS

The univariate models obtained R-squared values ranging from 0.66 to 0.95 and standard errors (SE) from 0.97 years to 2.18 years (Table 5.16). The femur diaphyseal length displayed the highest R-squared value of all models while the lowest R-squared value was the humerus midshaft breadth. The six diaphyseal lengths presented with the best models as demonstrated by the highest R-squared values (0.93 - 0.95) and the smallest SE's. Four of the five midshaft breadths demonstrated the models with the smallest R-squared values (0.65 – 0.75) and widest SE's. The proximal and distal breadths cluster between these two limits. Cube root transformations of age were conducted for the majority of univariate models as this alleviated the heteroscedasticity that was present in the y-axis residuals, although square root transformations of age and basis splines were also employed. Tables 5.17 and 5.18 are exemplars for MARS and spline models.

Although the cross-validated 95% prediction intervals present the same size intervals from birth to twelve when a transformation is applied to age, in actuality the intervals are dynamic and present with narrow intervals for the younger ages and wider intervals for the older ages (Figure 5.6). The summary tables of each fit and the associated age at death tables are listed in Appendices III – VIII for each univariate measure. If the lower 95% prediction interval included negative ages (younger than birth) the associated diaphyseal measurements were removed.

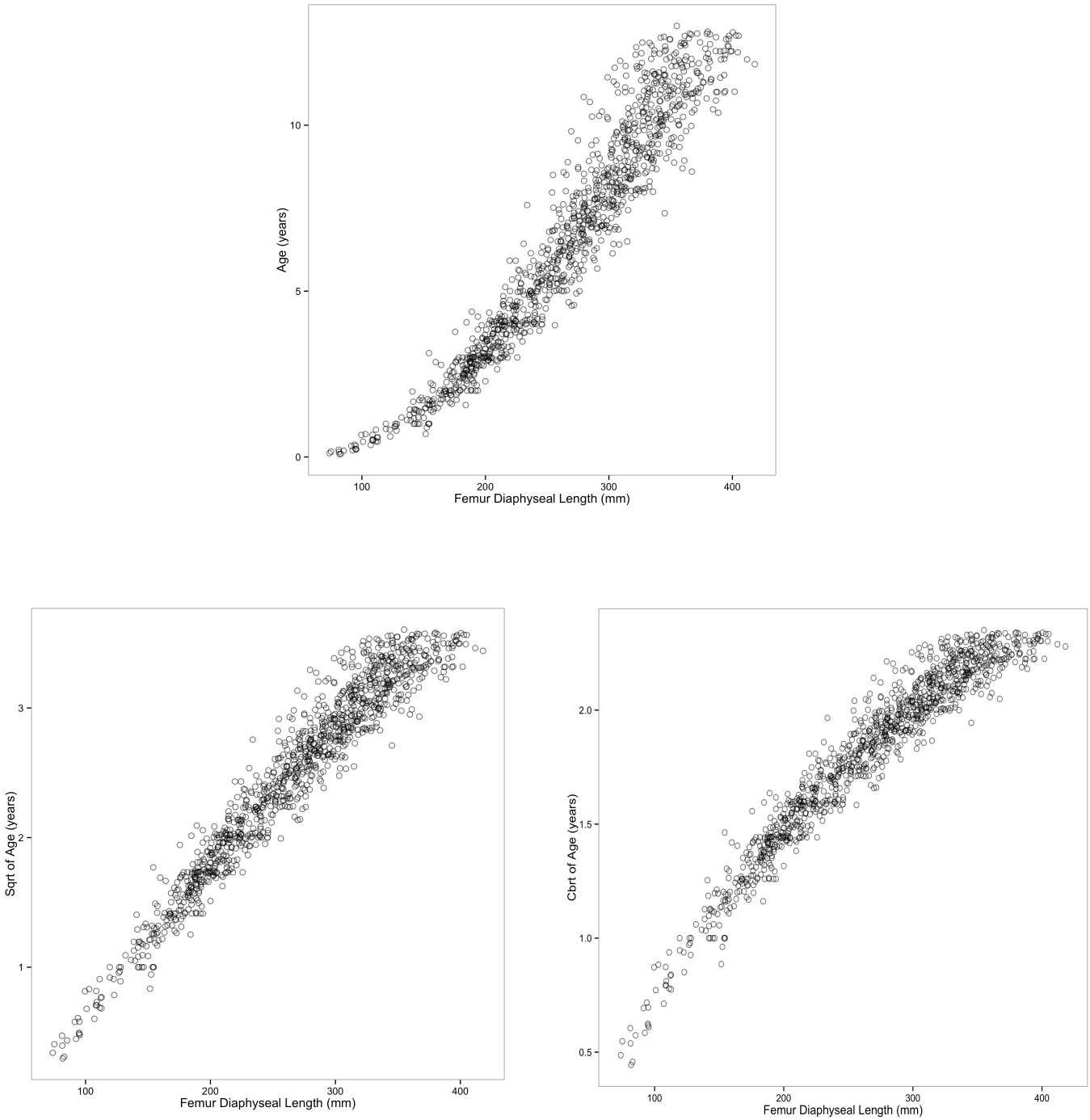


Figure 5.5 – The relationship between age and femur diaphyseal length with and without transformations of age. Age regressed on femur diaphyseal length without a transformation of age (top center); with a square root transformation of age (bottom left); and cube root transformation of age (bottom right).

Table 5.16 – Summary table of univariate models including the R-squared, model type and cross-validated percent correct. Variables are listed in decreasing order of SE. Spline models include the order and number of knots + 2. Cbrt (cube root) and sqrt (square root) MARS models refer to the type of transformation applied to the response variable (age).

	Model	SE	Adj R2	cv R2
FDL	cbrt MARS	0.9	0.95	0.95
TDL	sqrt MARS	0.95	0.94	0.94
HDL	cbrt MARS	0.97	0.95	0.95
FBDL	sqrt MARS	0.99	0.93	0.93
UDL	cbrt MARS	1.01	0.93	0.93
RDL	cbrt MARS	1.02	0.94	0.94
TDB	spline (2,4)	1.35	0.8	-
RPB	sqrt MARS	1.38	0.87	0.86
TPB	sqrt MARS	1.39	0.81	0.79
RDB	cbrt MARS	1.47	0.86	0.86
FDB	sqrt MARS	1.51	0.77	0.76
HDB	cbrt MARS	1.55	0.85	0.84
TMSB	cbrt MARS	1.59	0.72	0.71
HPB	spline (2,3)	1.6	0.83	-
FMSB	cbrt MARS	1.62	0.75	0.74
RMSB	cbrt MARS	1.76	0.81	0.79
UMSB	cbrt MARS	1.76	0.8	0.78
HMSB	sqrt MARS	2.18	0.66	0.64

Table 5.17 – MARS model for femur diaphyseal length. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cube root of age
(Intercept)	2.234***
h(FDL - 282.32)	0.006***
h(282.32 - FDL)	-0.008***
h(FDL - 167.91)	-0.002***
h(FDL - 342.03)	-0.002***
Observations	1117
cv R2	0.95
Adjusted R2	0.95
Residual Std. Error	0.90
F Statistic	5678.00***
Note:	***p < 0.01; **p<0.001; ***p<0.0001***

Table 5.18 – Spline model for the humerus proximal breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Sqrt of age
(Intercept)	0.278*
bs(HDB, degree = 2, df = 3)	0.867***
bs(HDB, degree = 2, df = 3)	3.467***
bs(HDB, degree = 2, df = 3)	3.040***
Observations	813
Adjusted R2	0.83
Residual Std. Error	1.6
F Statistic	1284.00***
Note:	***p < 0.01; **p<0.001; ***p<0.0001***

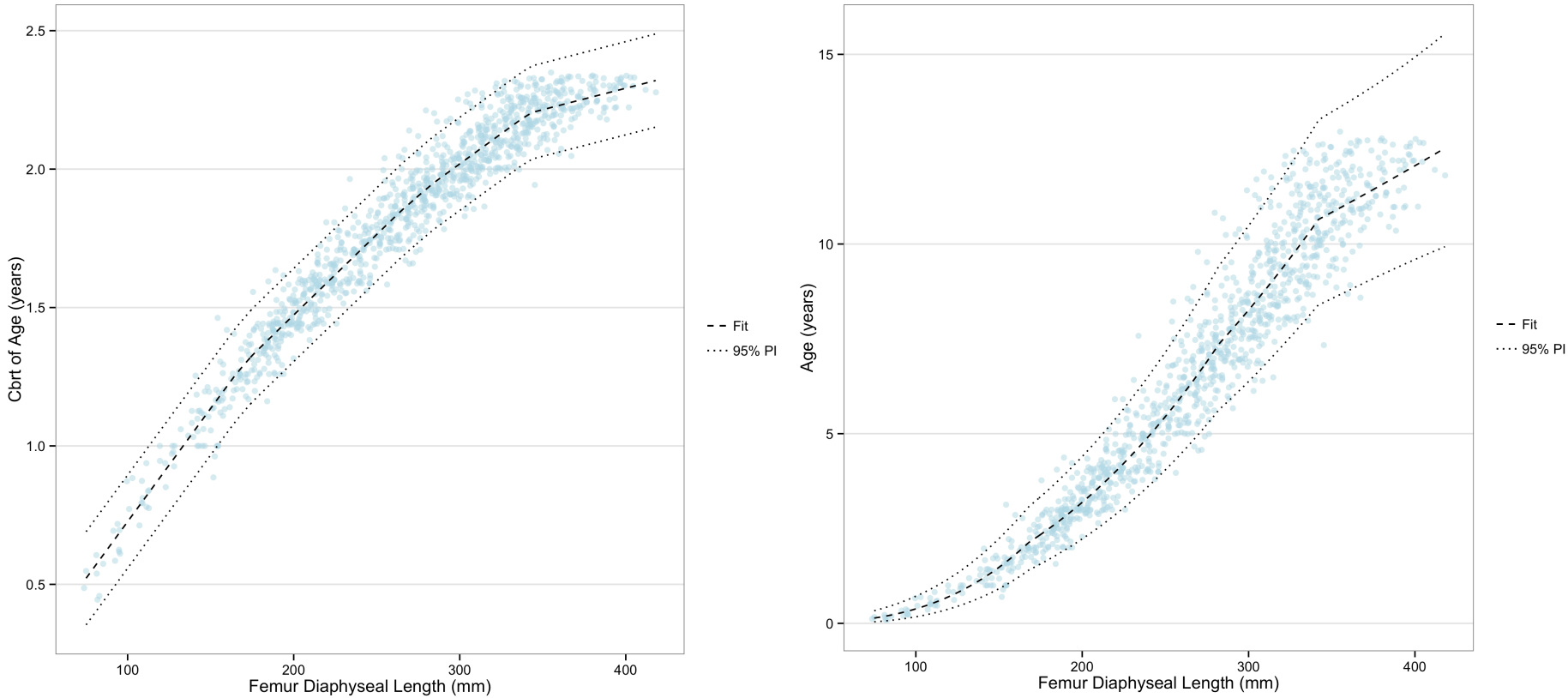


Figure 5.6 – The cross-validated 95% prediction intervals when age is regressed on femur diaphyseal length. The figure on the left displays the prediction intervals with the cube root transformation of age and the figure on the right displays how the prediction interval adjusts as age increases.

MULTIVARIATE MODELS

Since multicollinearity is known to significantly affect MARS models, the multivariate models were subjected to PCA using the variance-covariance matrix to remove the high inter-variable correlations. Multivariate models were created with both original variables and the PC scores. The model type, variables utilized in model creation, SE and R-squared for each multivariate model is located in Table 5.19. For the humerus, ulna, and femur models – both with the original variables and the PC scores – only one variable (HDL, UDL and FDL, respectively) contributed to model creation. Thus, the inclusion of breadth measurements did not increase predictive abilities; rather the univariate model utilizing the diaphyseal length of the humerus, ulna and femur yielded the most accurate estimation.

All multivariate models presented with SE's smaller than 1 year, with the exception of the upper limb subset using the PC scores that presented with a SE of 1.04 years. The all-measurement subset using the PC scores and the distal element subset, using the raw variables and the PC scores, presented with the smallest SE's of all the multivariate models (0.77, 0.78 and 0.79 years, respectfully). The distal element subset was also the only model that did not demonstrate a better fit with a transformation of age. The diaphyseal lengths or PC1 were retained and recognized as the most influential predictor variable in each model. Overall, the trend was for the SE to decrease when utilizing the PC scores, indicative of a higher accuracy, which is a likely consequence of uncorrelated diaphyseal dimensions (e.g. removal of multicollinearity).

The PCs identified as important variables exhibited contributions from the same measurements that were identified in the model using the original measurements. For example, the raw measurements that were chosen for the multivariate tibia model were TDL and TPB. The PCs identified as influential for the same model were PC1 and PC2. Loadings display which variables contributed to each PC, as this is the weight each standardized raw measurement is multiplied by to obtain the component score. The results of the PCA revealed that TDL accounted for nearly all of the variation in PC1; this is expected as PC1 usually accounts for the largest variance and is recognized as expressing size variation (Jolliffe, 2002; Berner, 2011). PC2 had substantial contributions from TPB as well as smaller contributions from TDB and TMSB.

All measurements were used to create the full model (all-measurement model), but the sample size was drastically reduced, thus some measurements were excluded to increase the sample size. Variables removed from the all-measurement model were UMSB, RMSB, RPB, and RDB. MARS selected six of the 14 measurements, which explained 93% of the variation in age with a standard error of 0.80 years (Table 5.20). FDL, HDB, TPB, RDL, TDL, and UDL were the stepwise selected models. The 14 measurements utilized in the all-measurement model were subject to PCA using the VCVM. MARS selected six of the 14 predictors, which explained 93% of the variation in age with a standard error of 0.78 years (Table 5.21). PC1, PC3, PC2, PC8, PC5 and PC10, were stepwise selected and appear in the order of variable importance. PC1-3 and PC5 have large contributions from diaphyseal lengths (FDL, TDL, and HDL) whereas PC8 and PC10 have large contributions from TDB, HPB and TMSB (Table 5.22). The cross-validated 95% prediction intervals for both all-measurement models ranged between 1 and

5 years for the youngest and oldest ages, respectively. The multivariate models serve as an example of the potential success of the application of MARS to subadult skeletal remains when estimating age at death. Appendix IX contains the age at death chart for the all-measurement model using the original variables. Because it is impossible to account for all variable combinations, the age at death chart only includes variables from the current dataset.

Table 5.19 – Standard error (SE), adjusted R-squared (adj R2) and cross-validated R-squared (cv R2) for the multivariate models.

	Model	Variables	SE	Adj R2	cv R2
Radius	sqrt MARS	RDL, RMSB	0.98	0.95	0.94
	cbrt MARS	PC1, PC2, PC4	0.97	0.96	0.95
Tibia	cbrt MARS	TDL, TPB	0.98	0.91	0.90
	cbrt MARS	PC1, PC2	0.97	0.91	0.90
Upper	cbrt MARS	UDL, RMSB, RDB, HMSB	0.99	0.95	0.87
	cbrt MARS	PC1, PC9	1.04	0.95	0.85
Lower	cbrt MARS	FDL, TDB	0.86	0.92	0.91
	cbrt MARS	PC1, PC2, PC5, PC6	0.85	0.92	0.91
Proximal	cbrt MARS	FDL, HDB, HPB, HDL	0.87	0.92	0.90
	cbrt MARS	PC1, PC3	0.88	0.91	0.90
Distal	cbrt MARS	RMSB, TPB, TDB, FBDL, UDL, RDB, UMSB, TDL	0.78	0.93	0.82
	cbrt MARS	PC1, PC4, PC6, PC7, PC9, PC10	0.79	0.93	0.77
All-measurement	cbrt MARS	FDL, HDB, TPB, RDL, UDL, TDL	0.80	0.93	0.84
	MARS	PC1, PC2, PC3, PC5, PC8, PC10	0.77	0.92	0.83

Table 5.20 – MARS all-measurement model explaining 93% of the variation in age. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cube root of age
(Intercept)	1.593***
h(FDL - 252.55)	0.0037***
h(252.55 - FDL)	-0.0087***
h(HDB - 40.24)	0.0074*
h(40.24 - HDB)	0.0195***
h(42.66 - TPB)	-0.0302***
h(RDL - 124.85)	0.0080***
h(220.29 - TDL)	0.0052**
h(UDL - 167.78)	-0.0090***
Observations	157
cv R2	0.85
Adjusted R2	0.93
Residual Std. Error	0.80
F Statistic	253.1***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table 5.21 – MARS all-measurement model (with PCA) explaining 92% of the variation in age.

	Predictor Variable
	Age
(Intercept)	3.037***
h(PC1 - 125.525)	-0.01461**
h(125.525 - PC1)	0.02727***
h(-8.26472 - PC3)	0.2296***
h(-5.62152 - PC2)	-0.1498***
h(PC8 - 2.56825)	0.7755**
h(2.56825 - PC8)	0.08215*
h(PC10 - 1.888)	-0.9047**
h(PC5 - -2.43291)	0.05599**
Observations	157
cv R2	0.82
Adjusted R2	0.92
Residual Std. Error	0.77
F Statistic	227.1***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table 5.22 – The eigenvectors and proportion of variance for each of the principal component scores using the variance-covariance matrix for the all-measurement model. The PC scores chosen in model creation were PC1, PC2, PC3, PC8, PC5 and PC10; presented in order of most to least important.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
TDL	-0.460	-0.158	0.517	-0.110	-0.231	0.052	-0.614	-0.192	0.063	-0.028	-0.028	0.084	0.019	-0.033
TPB	-0.067	-0.087	-0.210	-0.457	-0.033	-0.117	-0.043	0.236	0.215	0.142	-0.748	0.046	0.077	0.172
TMSB	-0.025	-0.023	-0.058	-0.157	-0.001	0.036	0.076	-0.003	-0.108	0.479	0.041	0.558	-0.198	-0.608
TDB	-0.051	-0.028	-0.097	-0.302	-0.057	0.035	-0.094	-0.042	-0.293	0.543	0.176	-0.677	0.074	-0.073
UDL	-0.274	-0.472	-0.293	0.187	0.314	0.104	-0.202	0.343	-0.505	-0.218	-0.067	0.018	-0.013	-0.079
FDL	-0.540	0.700	-0.129	-0.062	0.439	0.028	-0.027	-0.030	-0.032	-0.028	0.001	0.018	0.012	0.018
FDB	-0.072	-0.091	-0.255	-0.561	-0.076	-0.281	-0.069	0.212	0.227	-0.353	0.533	0.055	0.033	-0.094
FMSB	-0.025	-0.029	-0.045	-0.115	-0.044	0.054	-0.036	0.042	-0.190	0.274	0.235	0.307	-0.438	0.724
RDL	-0.262	-0.412	-0.277	0.216	0.319	-0.194	0.056	-0.442	0.480	0.235	0.083	-0.045	-0.026	0.066
HMSB	-0.016	-0.058	-0.081	-0.114	-0.040	0.097	0.109	-0.256	-0.244	0.075	0.097	0.332	0.812	0.211
HDB	-0.047	-0.094	-0.141	-0.209	-0.021	0.901	0.114	-0.097	0.240	-0.139	0.027	-0.056	-0.093	-0.030
HDL	-0.354	0.136	-0.447	0.356	-0.718	-0.007	0.033	0.095	0.044	0.042	0.002	-0.005	0.015	-0.019
HPB	-0.042	-0.027	-0.140	-0.247	-0.146	-0.157	0.182	-0.645	-0.398	-0.353	-0.215	-0.039	-0.296	-0.065
FBDL	-0.458	-0.206	0.430	-0.076	-0.058	-0.059	0.705	0.214	-0.031	0.002	0.023	-0.076	-0.003	0.028
Proportion of Variance	0.985	0.004	0.003	0.003	0.002	0.0005	0.0004	0.0004	0.0003	0.0002	0.0001	0.0001	<0.0001	<0.0001

BIAS

Plots of the bias and chronological age were constructed to determine if the point estimates of the models tended to under- or overestimate age. The bias is the difference between the chronological age and the point estimate for each individual (i.e. residual). Although the mean bias was normally distributed for each model and always approximating zero, and the majority of individuals were aged within one year of their true chronological age, an increase in error with an increase in age was apparent. Whereas the range of bias error was +/- 3 years in the univariate models, the range of bias error was only +/- 2 years in the multivariate models (Table 5.23). All univariate models demonstrated an increase in bias error as age increased (i.e. heteroscedastic). Most univariate models were unbiased until approximately 10 years of age. Following 10 years of age, the degree of underestimation increased as age increased (Figure 5.7). Overall, the bias consistently paralleled zero for the multivariate models but the heteroscedasticity persisted (Figure 5.8). Multivariate subsets had smaller sample sizes, which affect the range of error associated with bias. The original dataset was used to verify the cross-validated 95% prediction intervals associated with each univariate and multivariate model. The accuracy of the chronological age falling within the prediction intervals ranged from 94% to 100%. Thus, even though the models lose precision and may be slightly biased, 95% – or more – of the observed values fell within the prediction intervals.

Loess lines by sex were plotted to further investigate if trends in bias differed between the sexes. Males and females followed similar trends in bias for all models. However, larger disparities were apparent in the breadth measurements than the length measurements (Figure 5.9). The trend for the diaphyseal breadths was for the females to be underestimated whereas the trend for the diaphyseal lengths was for the males to be slightly underestimated. The different patterns demonstrate the tendency for males to have larger breadths than females and the tendency for females to have larger lengths than males. In contrast to the univariate models, no sex difference was noted in the bias error of the multivariate models, which suggests the multivariate models negated the effects of sex.

When the bias of the femur diaphyseal length models was assessed with the holdout sample (n=28), the error range was reduced to a range of +/- 1 year and a mean of 0.14 years.

Because none of the individuals in the holdout sample had all of the measurements required to test the all-measurement model, the tibia multivariate model was utilized as this model only required two variables and consequently had one of the larger samples (n=14). The mean bias and range in the residuals for the multivariate tibia model was 0.09 years and -2.4 years to 1.76 years, respectively.

Table 5.23 – Bias is listed in decreasing order based on the range.

Model	Bias of Point Estimates	
	Mean	*Range*
All-measurement	-0.02	-2.2 to 1.9
All-measurement (PCA)	0.00	-2.2 to 2.1
Distal (PCA)	-0.03	-2.4 to 1.4
Upper (PCA)	-0.05	-2.7 to 2.18
Distal	-0.03	-2.7 to 1.8
Upper	-0.05	-2.9 to 2.0
Lower	-0.04	-2.9 to 2.1
Radius	-0.03	-2.9 to 3.1
Lower (PCA)	-0.04	-3.0 to 2.0
Proximal (PCA)	0.00	-3.1 to 2.1
Proximal	-0.05	-3.2 to 2.1
Radius (PCA	-0.05	-3.2 to 3.0
TDL	-0.03	-3.5 to 2.6
FDL	-0.05	-3.6 to 3.3
HDL	-0.05	-3.7 to 3.0
UDL	-0.06	-3.7 to 2.8
FBDL	-0.03	-3.9 to 3.5
RDL	-0.06	-4.2 to 3.4
Tibia	-0.05	-4.4 to 2.4
RPB	-0.08	-4.3 to 4.6
Tibia (PCA)	-0.05	-4.4 to 2.4
TDB	-0.06	-4.6 to 3.8
TPB	-0.06	-4.6 to 3.9
TMSB	-0.14	-4.8 to 4.1
FDB	-0.08	-5.4 to 4.8
RDB	-0.13	-5.6 to 4.5
HPB	-0.10	-5.8 to 4.6
RMSB	-0.17	-5.8 to 3.6
FMSB	-0.14	-5.9 to 5.0
HDB	-0.13	-6.0 to 4.0
UMSB	-0.18	-6.2 to 4.2
HMSB	-0.19	-7.0 to 6.0

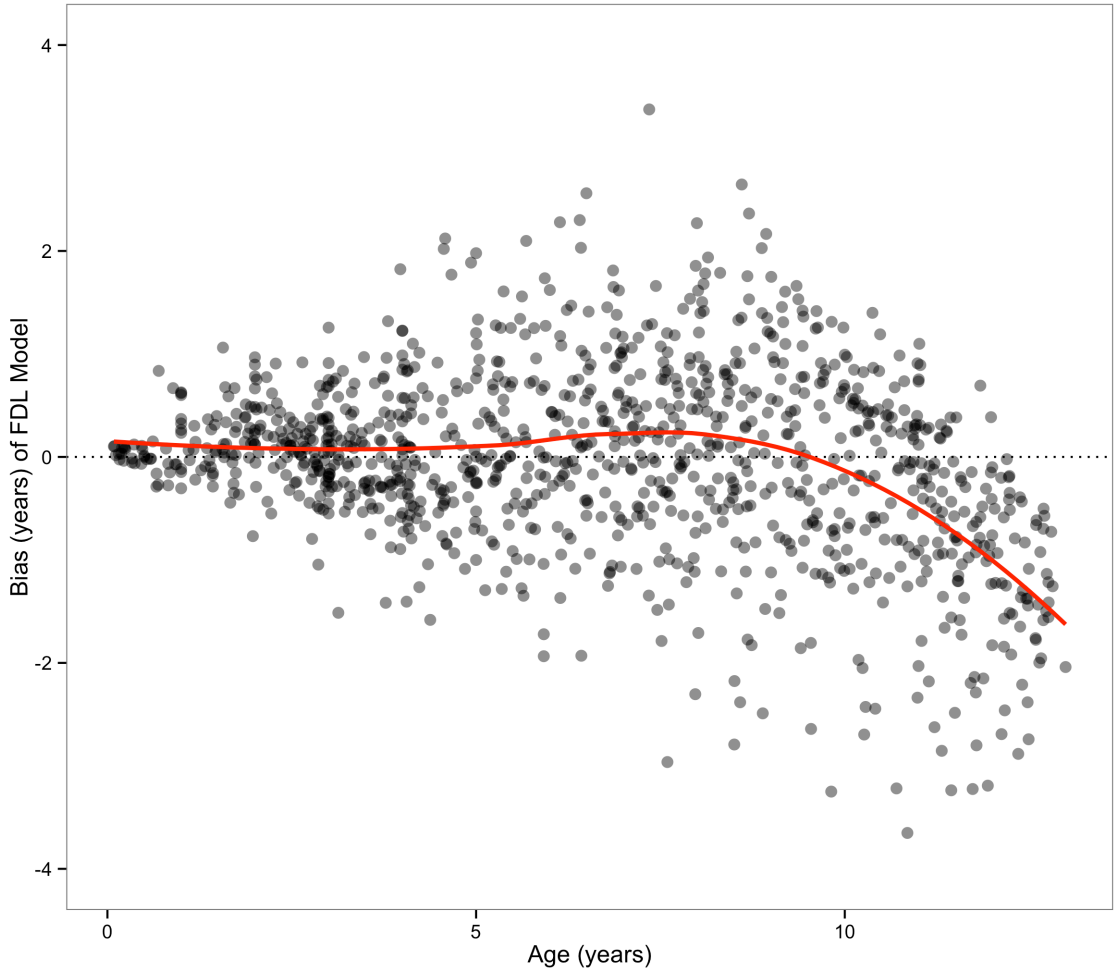


Figure 5.7 – The difference between the predicted values and true chronological age (residuals) for the femur diaphyseal length model plotted against age with a loess line depicting the local estimated bias.

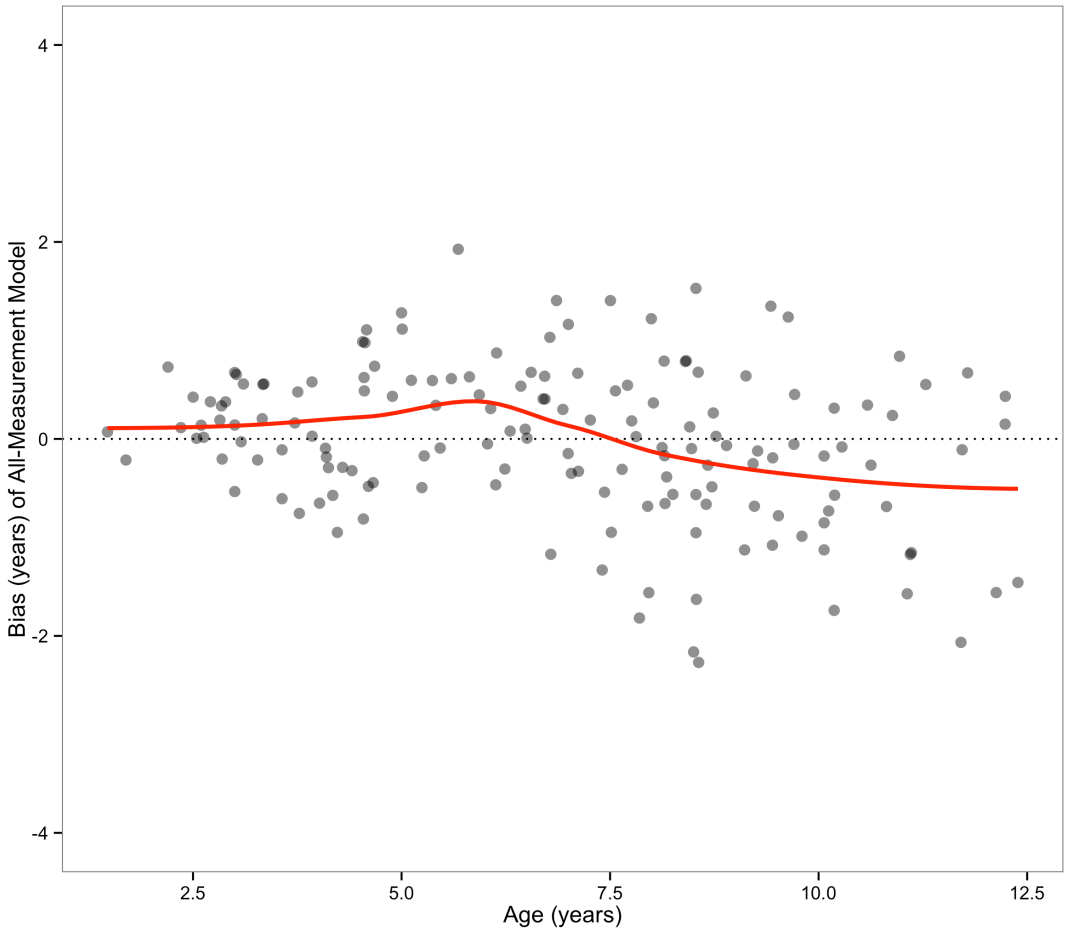


Figure 5.8 – The difference between the predicted values and true chronological age (residuals) for the all-measurement model plotted against age with a loess line depicting the local estimated bias.

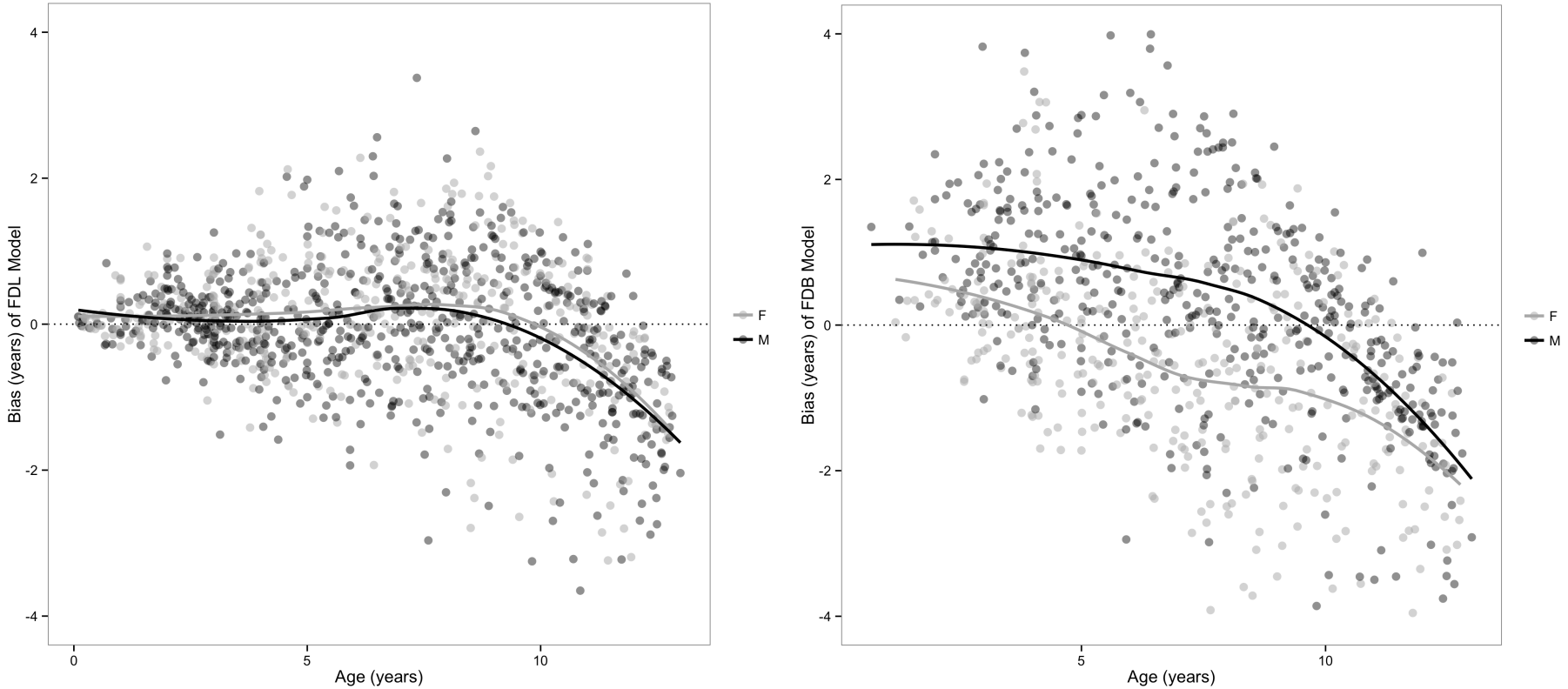


Figure 5.9 – The two figures depict the difference between the predicted values and true chronological age (residuals) for the femur diaphyseal length (left) and femur distal breadth (right) plotted against age with a loess lines depicting the local estimated bias of males and females.

SEXUAL DIMORPHISM

Sex differences were shown in sympercents (sp) in order to provide objective size differences in showing proportional differences between males and females. For an easier comparison, measurements were grouped according to type of measurement (i.e. proximal breadths, distal breadths, lengths, and midshaft breadths) rather than skeletal element. As seen in Figure 5.10 and Tables 5.24 – 5.27, males are larger than females in the midshaft breadths, distal breadths, and proximal breadths throughout most of the growth period. On average from birth to nine years, males have 4.2 sp larger proximal breadths, 3.4 sp larger distal breadths, and 3.0 sp larger midshaft breadths than females. Ten through 12 year olds demonstrate a smaller disparity between males and females for the same variables. Specifically, males have 2.7 sp larger proximal breadths, 1.9 sp larger distal breadths, and 1.7 sp larger midshaft breadths than females. However, different tendencies are noted between the upper limb breadths and the lower limb breadths at 12 years of age. Sympercent differences in upper limb breadths decrease between males and females as age increases whereas males are consistently larger than females in the lower limb breadths.

Diaphyseal length measurements follow a different pattern than diaphyseal breadth measurements (Table 5.27). Females demonstrate larger diaphyseal lengths at birth and again around 12 years; specifically, in the first year of life, females displayed diaphyseal lengths that were approximately 5 sp larger than males. For ten to 12 year olds, females displayed diaphyseal lengths 1.3 sp larger than males. However, from one to nine years of age females had, on average, only a 0.6 sp advantage. Mean male and female ages were similar for each age interval, thus for most of childhood, negligible sex differences are noted in diaphyseal lengths.

SEX CLASSIFICATION

Multiple classification models were employed in an attempt to identify the model that consistently yielded the highest correct sex classification. The models included quadratic, linear and flexible discriminant analysis (QDA, LDA and FDA, respectively), logistic regression (Log Reg), naïve Bayes (NB), and random forests (RF). The classification models were conducted on the raw measurements, ratios, y-axis residuals of the basis spline or MARS age estimation models (i.e. size-adjusted variables), and PC scores using the VCVM. Each measurement and ratio was subject to univariate classification models whereas the multivariate subsets, which included the multivariate bone subsets, multivariate subsets (i.e. measurement type and element type subsets) and the all-measurement subset, were subject to multivariate classification models. Stepwise variable selection for the QDA and LDA models was used for the multivariate bone subsets. The FDA automatically conducts variable selection and the logistic regression identifies which variables significantly contributes to model creation, thus the reported classification accuracies may or may not include all the variables that were included in the classification models.

Table 5.24 – Sympercent differences for the proximal breadths. Female dimensions were subtracted from male dimensions, thus negative sympercents indicate females are larger than males.

Age (years)	**Sympercent Differences**		
	Humerus	*Radius*	*Tibia*
<1	-1.67	1.81	NA
1	1.55	0.18	5.53
2	4.12	2.76	4.48
3	4.25	2.6	2.84
4	5.2	5.95	4.48
5	7.04	5.23	6.59
6	2.1	1.79	2.93
7	5.93	3.85	6.76
8	7.41	6.63	6.01
9	6.05	8.43	5.44
10	7.98	5.92	3.21
11	4.2	0.81	1.42
12	0.15	-2.25	3.01

Table 5.25 – Sympercent differences for the midshaft breadths. Female dimensions were subtracted from male dimensions, thus negative sympercents indicate females are larger than males.

Age (years)	Sympercent Differences				
	Humerus	*Ulna*	*Radius*	*Femur*	*Tibia*
<1	0.23	-1.06	8.16	0.42	NA
1	2.77	-3.50	-8.30	4.69	1.24
2	2.93	4.20	4.00	2.76	6.89
3	2.31	6.41	2.64	-0.19	2.25
4	2.91	4.37	0.64	2.35	1.49
5	4.57	4.65	-2.32	-0.40	4.80
6	2.49	3.02	-4.86	1.56	-3.10
7	6.21	8.64	2.99	5.07	5.81
8	8.69	10.23	7.35	4.43	3.05
9	6.86	9.43	11.59	5.32	10.18
10	4.30	4.40	2.27	2.17	1.06
11	3.59	9.59	2.60	4.19	2.61
12	-2.63	-13.39	-4.62	1.07	8.47

Table 5.26 – Sympercent differences for the distal breadths. Female dimensions were subtracted from male dimensions, thus negative sympercents indicate females are larger than males.

Age (years)	Sympercent Differences			
	Humerus	*Radius*	*Femur*	*Tibia*
<1	4.93	3.77	NA	-1.44
1	4.36	-0.35	7.54	1.63
2	3.4	3.86	6.39	4.13
3	0.31	3.83	4.26	0.17
4	2.66	6.76	4.7	2.4
5	6.34	4.07	6.04	4.56
6	0.05	1.18	5.45	0.89
7	4.48	5.2	7.08	3.92
8	4.01	8.56	4.99	4.05
9	4.43	4.64	4.54	2.58
10	-0.23	4.17	4.25	2.56
11	2.08	3.31	3.94	1.39
12	-5.41	-0.18	4.94	1.88

Table 5.27 – Sympercent differences for the diaphyseal lengths. Female dimensions were subtracted from male dimensions, thus negative sympercents indicate females are larger than males.

Age (years)	Sympercent Differences					
	Humerus	*Ulna*	*Radius*	*Femur*	*Tibia*	*Fibula*
<1	-4.57	-5.29	-4.18	-4.91	-6.72	-1.17
1	1.09	0.25	2.55	1.67	2.98	3.37
2	-0.88	-0.74	0.7	-2.61	-1.06	-1.87
3	-0.82	-2.18	-0.81	-2.86	-2.38	-2.78
4	-0.34	0.69	1.7	-0.04	-0.17	-0.05
5	1.59	0.26	0.46	-0.27	-1.39	-0.98
6	0.75	-0.27	0.91	0.17	-0.74	-0.76
7	-0.41	0.9	1.32	1.02	0.18	1.03
8	-0.42	-1.32	0.21	-1.57	-2.58	-1.49
9	-1.03	-0.52	-0.16	-0.46	-1.28	-0.68
10	-2.33	-1.86	-1.47	-1.87	-1.08	-2.25
11	1.3	3.82	1.1	0.4	-0.66	0.39
12	-1.9	-1.22	-4.63	-2.59	-4.81	-3.89

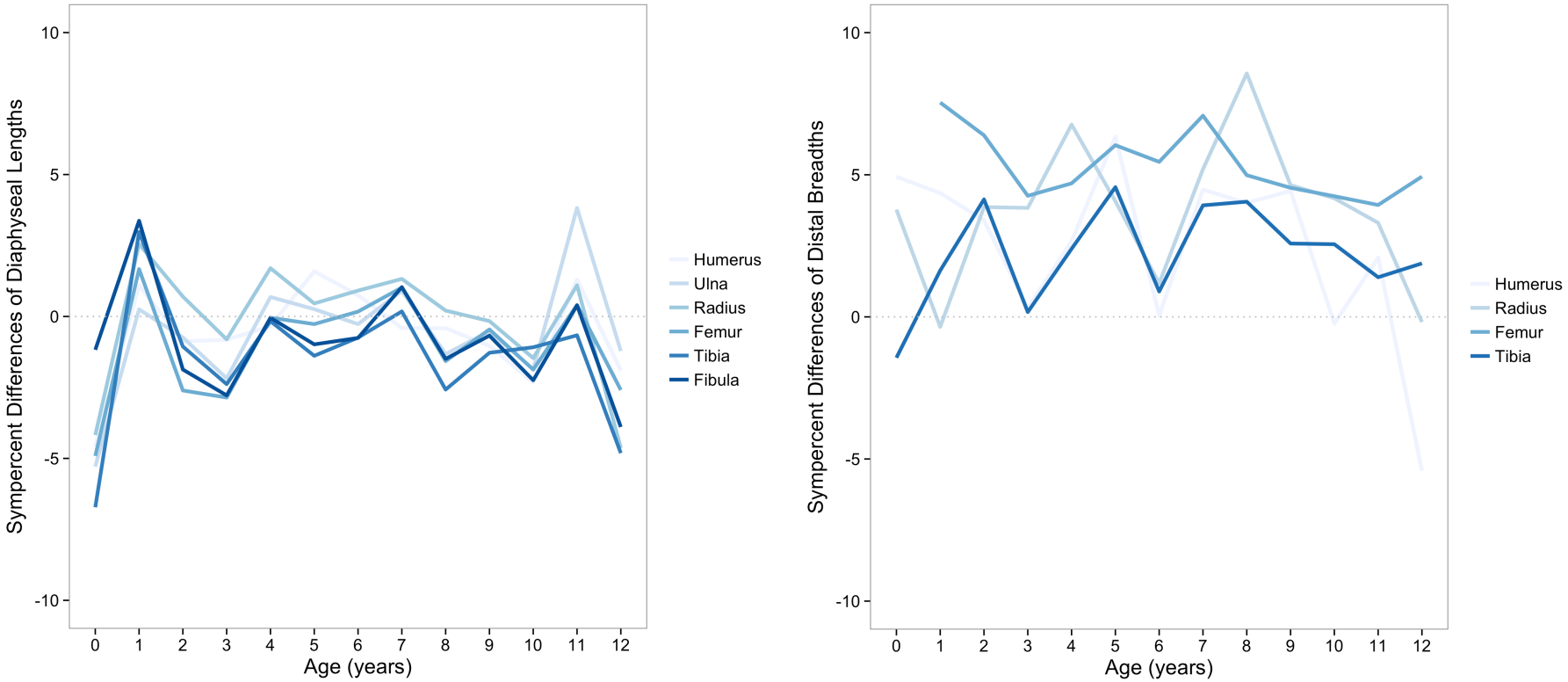


Figure 5.10 – The figure on the left displays the sympercent differences of male and female diaphyseal length measurements and the figure on the right displays the sympercent differences of male and female distal breadth measurements. Female dimensions were subtracted from male dimensions, thus negative sympercents indicate females are larger than males.

UNIVARIATE AND MULTIVARIATE BONE MODELS

Classification accuracies for the univariate and multivariate bone models for each element are listed in Tables 5.28 – 5.33. Classification accuracies ranged from 47% to 74%. A comparison of all elements demonstrated the humerus (71%) and the femur (74%) achieved the highest number of correct classifications. In the univariate models, ratios generally yielded slightly higher accuracies than original variables and the breadths always demonstrated higher accuracies than lengths. Multivariate bone subsets and stepwise selected models yielded significantly better classifications than the univariate models.

Evaluation of variable importance in the multivariate bone models demonstrated that breadth measurements were the most important, but the diaphyseal lengths were always retained in the model. The breadth measurements that contributed the most to each of the specific multivariate bone models were HPB, RDB, FDB and TPB. An evaluation of all of the models revealed that logistic regression achieved the overall highest classification rates. Though the univariate QDA and LDA models usually achieved a lower number of correct classifications, stepwise selection for the multivariate bone models was usually comparable to the logistic regression as well as the FDA. The NB and RF rarely achieved similar accuracies.

Table 5.28 – Classification accuracies for the humerus. The multivariate bone model only includes original variables (no ratios). Model abbreviations: QDA = quadratic discriminant analysis; LDA = linear discriminant analysis; FDA = flexible discriminant analysis; Log Reg = logistic regression; NB= naïve Bayesian; RF = random forest. Refer to Table 4.1 and 4.2 for measurement abbreviations.

	n	Classification Accuracy					
		QDA	*LDA*	*FDA*	*Log Reg*	*NB*	*RFM*
HDL	1069	54%	54%	54%	63%	-	-
HPB	814	57%	58%	55%	62%	-	-
HDB	710	56%	55%	50%	65%	-	-
HMSB	963	57%	58%	55%	62%	-	-
HDPB	773	58%	54%	64%	65%	-	-
HDDB	667	47%	51%	45%	66%	-	-
HDMS	930	49%	49%	60%	64%	-	-
Multivariate bone model	497	66%	66%	69%	**71%**	57%	65%
Stepwise selected measurements	497	67%	67%	-	-	-	-

Table 5.29 – Classification accuracies for the ulna. The accuracy in bold is the best for the element. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Classification Accuracy					
		QDA	*LDA*	*FDA*	*Log Reg*	*NB*	*RFM*
UDL	1182	56%	54%	49%	65%	-	-
UMSB	438	57%	56%	54%	66%	-	-
UDMS	440	60%	54%	58%	**67%**	-	-
Multivariate bone model	281	58%	57%	58%	**67%**	66%	57%
Stepwise selected measurements	281	56%	57%	-	-	-	-

Table 5.30 – Classification accuracies for the radius. The accuracy in bold is the best for the element. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Classification Accuracy					
		QDA	*LDA*	*FDA*	*Log Reg*	*NB*	*RFM*
RDL	1182	54%	52%	-	63%	-	-
RPB	523	57%	56%	56%	63%	-	-
RDB	565	55%	54%	55%	62%	-	-
RMSB	438	53%	50%	64%	63%	-	-
RDPB	511	58%	55%	58%	63%	-	-
RDDB	547	58%	56%	61%	64%	-	-
RDMS	440	61%	61%	59%	63%	-	-
Multivariate bone model	281	63%	66%	64%	**67%**	57%	57%
Stepwise selected measurements	281	66%	66%	-	-	-	-

Table 5.31 – Classification accuracies for the femur. The accuracy in bold is the best for the element. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Classification Accuracy					
		QDA	*LDA*	*FDA*	*Log Reg*	*NB*	*RFM*
FDL	1069	53%	53%	61%	62%	-	-
FDB	710	60%	60%	59%	60%	-	-
FMSB	963	53%	55%	60%	61%	-	-
FDDB	667	56%	56%	64%	64%	-	-
FDMS	930	56%	52%	60%	61%	-	-
Multivariate bone model	497	70%	72%	73%	**74%**	59%	68%
Stepwise selected measurements	497	71%	72%	-	-	-	-

Table 5.32 – Classification accuracies for the tibia. The accuracy in bold is the best for the element. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Classification Accuracy					
		QDA	*LDA*	*FDA*	*Log Reg*	*NB*	*RFM*
TDL	1069	53%	52%	60%	60%	-	-
TPB	814	56%	57%	56%	59%	-	-
TDB	710	53%	54%	54%	60%	-	-
TMSB	963	50%	52%	57%	57%	-	-
TDPB	773	58%	58%	63%	64%	-	-
TDDB	667	56%	55%	63%	63%	-	-
TDMS	930	58%	56%	63%	63%	-	-
Multivariate bone model	497	64%	67%	67%	**69%**	56%	65%
Stepwise selected measurements	497	68%	68%		-	-	-

Table 5.33 – Classification accuracies for the fibula. The accuracy in bold is the best for the element. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Classification Accuracy					
		QDA	*LDA*	*FDA*	*Log Reg*	*NB*	*RFM*
FBDL	1182	54%	53%	-	**60%**	**60%**	52%

RESIDUALS

The residuals obtained from the spline or MARS age estimation models were used in the classification models as size-independent measures (Jungers et al., 1995; Berner, 2011). All residuals demonstrated a normal distribution and a mean of zero. Univariate classification models inclusive of QDA, LDA, FDA, and logistic regression were used.

Classification accuracies utilizing the residuals of the age estimation models ranged from 52% to 66% (Tables 5.34 – 5.39). The results were similar to the univariate models using the original measures and ratios. The highest correct classifications for each bone were achieved by residuals of HDB, UMSB, RDB and RPB, TPB; thus, residuals of breadths performed better than lengths in sex estimation. The classification method that yielded the overall highest number of correct classifications was logistic regression.

Table 5.34 – Classification accuracies using the residuals of the humeral MARS and spline models. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Model*	Classification Accuracy			
			QDA	*LDA*	*FDA*	*Log Reg*
HDL	1069	cbrt MARS	58%	51%	60%	63%
HPB	814	BS (2,3)	61%	59%	61%	63%
HDB	710	cbrt MARS	58%	58%	59%	**65%**
HMSB	963	sqrt MARS	55%	55%	56%	63%

***If the model is BS, then the parentheses include the degree and df for the model**

Table 5.35 – Classification accuracies using the residuals of the ulna MARS models. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Model	Classification Accuracy			
			QDA	*LDA*	*FDA*	*Log Reg*
UDL	1031	cbrt MARS	56%	52%	60%	65%
UMSB	403	cbrt MARS	63%	55%	60%	**66%**

Table 5.36 – Classification accuracies using the residuals of the tibia MARS models. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Model	Classification Accuracy			
			QDA	*LDA*	*FDA*	*Log Reg*
RDL	1069	cbrt MARS	54%	52%	53%	63%
RPB	814	sqrt MARS	59%	58%	59%	**64%**
RDB	710	cbrt MARS	58%	57%	61%	**64%**
RMSB	963	cbrt MARS	61%	53%	55%	63%

Table 5.37 – Classification accuracies using the residuals of the femur MARS models. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Model	Classification Accuracy			
			QDA	*LDA*	*FDA*	*Log Reg*
FDL	1069	cbrt MARS	54%	53%	53%	61%
FDB	710	sqrt MARS	61%	61%	62%	**64%**
FMSB	963	cbrt MARS	56%	53%	57%	61%

Table 5.38 – Classification accuracies using the residuals of the radius MARS models. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Model	Classification Accuracy			
			QDA	*LDA*	*FDA*	*Log Reg*
TDL	1047	sqrt MARS	54%	53%	53%	60%
TPB	747	sqrt MARS	60%	60%	60%	**63%**
TDB	637	cbrt MARS	57%	53%	53%	60%
TMSB	564	sqrt MARS	58%	57%	58%	59%

Table 5.39 – Classification accuracies using the residuals of the fibula MARS model. Refer to 4.1 and 4.2 for measurement abbreviations and Table 5.26 for model abbreviations.

	n	Model	Classification Accuracy			
			QDA	*LDA*	*FDA*	*Log Reg*
FBDL	1030	sqrt MARS	54%	53%	53%	**60%**

PRINCIPAL COMPONENT SCORES AND CLASSIFICATION

The principal component scores derived from a variance-covariance matrix for each multivariate bone subset were utilized in the classification models (QDA, LDA, FDA, Log Reg, NB, RF) to remove the effects of multicollinearity. Three different analyses were run for each element. The first run was conducted on all of the PC scores, the second run included only the stepwise selected variables from the QDA and LDA (Table 5.40), and the final run was conducted on the PC scores for each element with the exception of PC1. As previously stated, PC1 is recognized to reflect size variation within the dataset (Jolicoeur, 1963); removing the substantial effects of size in sex estimation permits the shape variables (i.e. breadth measurements), captured in PC2 through PC*x*, to be used in model creation.

The classification accuracy of all models for the humerus, femur, and tibia ranged from 58% (naïve Bayes results for the tibia) to 74% (FDA and logistic regression results for the

femur) (Table 5.41). Similar to the multivariate bone models using the original measurements, the femur and humerus achieved the highest correct classifications of 74% and 72%, respectively. The radius and ulna had the lowest number of correct classifications (52% to 68%) of all five elements. Overall, the removal of PC1 did not drastically increase correct classifications. Logistic regression, FDA and NB consistently yielded the highest number of correct classifications of all models. RF had similarly bleak results to QDA and LDA.

Table 5.40 – Stepwise selected principal components for the QDA and LDA models.

	QDA Stepwise	**LDA Stepwise**
Humerus	PC1 – PC4	PC2 and PC3
Radius	PC2 and PC3	PC2 – PC4
Ulna	PC1	PC2
Femur	PC2	PC2 and PC3
Tibia	PC2-4	PC1-2

Table 5.41 – Classification accuracies for the principal components of all long bone datasets. The highest correct classification for each combination of variables is noted in bold. Refer to Table 5.26 for model abbreviations.

	n	Variables	Classification Accuracy					
			QDA	*LDA*	*FDA*	*Log Reg*	*NB*	*RFM*
Humerus	497	PC1-PC4	66%	66%	70%	71%	**72%**	68%
		PC2-PC4	66%	66%	**70%**	**70%**	66%	69%
		Stepwise	66%	68%	-	-	-	-
Radius	281	PC1-PC4	63%	66%	65%	67%	**69%**	56%
		PC2-PC4	62%	65%	65%	**68%**	61%	58%
		Stepwise	65%	65%	-	-	-	-
Ulna	400	PC1-PC2	53%	57%	62%	**67%**	**67%**	56%
		PC2	52%	58%	58%	**67%**	**67%**	57%
		Stepwise	57%	58%	-	-	-	-
Femur	778	PC1-PC3	70%	72%	73%	**74%**	72%	66%
		PC2-PC3	70%	71%	**74%**	72%	60%	64%
		Stepwise	71%	72%	-	-	-	-
Tibia	521	PC1-PC4	64%	67%	69%	**70%**	58%	63%
		PC2-PC4	67%	67%	**70%**	69%	58%	63%
		Stepwise	67%	68%	-	-	-	-

MULTIVARIATE CLASSIFICATION MODELS

The multivariate subsets were utilized to preserve a suitable sample size while including a large number of predictor variables. Based on the success of ratios in the univariate classification models, a ratios-only dataset was also created. The variables utilized in the five subsets are noted in Table 5.40. Similar to the previous analyses, and because multicollinearity is such a problem for the classification methods, the multivariate subsets were subject to PCA using a variance-covariance matrix and the PC scores were also utilized in model creation. Because some of the subset sample sizes were small, the classification accuracies of each model were bootstrapped in order to validate the results. Bootstrapped classification accuracies were conducted for each of the subsets for the LDA, FDA, and logistic regression classification models as these models either consistently yielded the highest correct classifications (FDA, logistic regression) or the methodology is commonly employed in anthropology (LDA).

Correct classifications – excluding the results of the ratios subset – ranged from 74% to 89% utilizing the original variables and 72% to 90% utilizing the PC scores (Tables 5.43 and 5.44). Specifically for the original variables, the all-measurement subset and upper limbs subset achieved the highest accuracies (89% and 81%, respectively) and the ratios subset achieved the lowest accuracy (66%). The bootstrapped estimates provided a standard error for the estimated classification accuracies, which resulted in 95% confidence intervals. For example, the all-measurement model bootstrapped 95% confidence intervals for accuracy were 73% – 85% for the LDA, 86% - 92% for the FDA and 84% – 94% for the logistic regression. Results demonstrate that FDA and logistic regression yielded higher correct classifications than the stepwise LDA. An evaluation of variables that either had significant relationships with sex or were recognized as important variables in the model creation for the multivariate subset were similar for the logistic regression and FDA. Specifically for the all-measurement subset, the logistic regression identified FDB and TDB as making significant contributions to sex estimation while the FDA model retained diaphyseal lengths with breadth measurements, such as HPB, HDL, FDB, FDL, TDB, TPB and RPB (listed in order from most important to least important).

The classification accuracy slightly increased when using the PC scores in the multivariate subsets but some subsets slightly decreased, such as the proximal element subsets. The first PC was not included in any model creation, thus the effects of size were removed.

Similar to the original variable models, the highest accuracies were observed in the all-measurement (90%) and upper limbs (83%) subsets, the ratios presented with the lowest number of correct classifications and the FDA and logistic regression yielded the overall highest number of correct classifications. The standard error for each of the estimated classification accuracies was similarly low as compared to the models utilizing the original variables. Although the best performing model was the FDA for the multivariate subsets, which is in contrast to the univariate approaches, the logistic regression had similar results to FDA for most of the multivariate subsets.

PC2, 3, 5 and 6 contributed the most to the logistic regression and FDA models. The only difference was that the FDA model also included PC4. While PC2 had the largest contributions from diaphyseal lengths, namely the femur and humerus, PC3, PC4, PC5, and PC6 had contributions from a combination of diaphyseal breadths, namely FDB, TPB, HDP, HPB and TDB (Table 5.45). All of the variables utilized in the FDA and logistic regression models utilizing the PC scores were generally the same variables that were considered important variables in the multivariate bone models. As the PCA of the variance-covariance matrix retains all of the variation present in the original dataset even though the PC scores are uncorrelated and transformed, the above-mentioned outcomes are expected (Jolliffe, 2002).

Table 5.42 – Measurements included in each multivariate subset.

	n	Variables
All-measurement	63	FDL, FDB, FMSB, HDL, HPB, HDB, HMSB, RPB, RDB, TPB, TDB*
Ratios	394	HDPB, TDPB, FDDB
Upper Limbs	96	HDL, HPB, HDB, RDL, RPB, RDB, RMSB, UDL, UMSB
Lower Limbs	458	FDL, FDB, FMSB, TDL, TPB, TDB, TMSB, FBDL
Proximal Elements	316	HDL, HPB, HDB, HMSB, FDL, FDB, FMSB
Distal Elements	100	RDL, RPB, RDB, RMSB, UDL, UMSB, TDL, TPB, TDB, TMSB, FBDL

***Some variables were excluded due to the effects on sample size**

Table 5.43 – Bootstrapped classification accuracies for the generated multivariate subsets utilizing original measurements.

	n	LDA		FDA		*Log Reg*	
		Accuracy	*95% CI*	*Accuracy*	*95% CI*	*Accuracy*	*95% CI*
All-measurement	63	79%	73% – 85%	**89%**	86% – 92%	**89%**	84% – 94%
Ratios	394	65%	63% – 67%	65%	63% – 67%	**66%**	64% – 68%
Upper Limbs	96	74%	69% – 79%	**81%**	77% – 85%	74%	69% – 79%
Lower Limbs	458	70%	68% – 72%	**72%**	70% – 74%	**72%**	70% – 74%
Proximal Elements	316	73%	70% – 74%	**78%**	76% – 80%	77%	75% – 79%
Distal Elements	100	73%	67% – 79%	67%	62% – 72%	**76%**	71% – 81%

Table 5.44 – Bootstrapped classification accuracies for the generated multivariate subsets using the PC scores, with the exception of PC1.

	n	LDA		FDA		*Log Reg*	
		Accuracy	*95% CI*	*Accuracy*	*95% CI*	*Accuracy*	*95% CI*
All-measurement	63	70%	64% – 76%	**90%**	88% – 92%	87%	82% – 92%
Ratios	394	57%	55% – 59%	64%	61% – 67%	**63%**	61% – 65%
Upper Limbs	96	73%	68% – 78%	**83%**	80% – 86%	76%	71% – 81%
Lower Limbs	458	69%	67% – 71%	**72%**	70% – 74%	71%	69% – 73%
Proximal Elements	316	75%	73 – 77%	75%	73% – 77%	**76%**	74% – 78%
Distal Elements	100	70%	65% – 75%	**81%**	77% – 85%	77%	72% – 82%

Table 5.45 – The eigenvectors and proportion of variance for each of the principal component scores using a variance-covariance matrix for the all-measurement subset.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
FDL	0.816	0.544	0.185	0.034	0.042	0.006	-0.019	-0.009	0.001	0.007	-0.012
FDB	0.110	-0.315	0.437	-0.241	-0.163	0.124	-0.673	0.216	0.143	0.270	-0.070
FMSB	0.038	-0.077	0.097	-0.015	-0.180	-0.266	-0.253	0.011	-0.446	-0.737	-0.271
HDL	0.539	-0.606	-0.583	-0.016	-0.045	0.021	0.018	-0.001	0.007	0.008	0.002
HPB	0.061	-0.224	0.246	-0.222	0.751	0.227	0.063	-0.273	0.193	-0.295	-0.132
HDB	0.074	-0.279	0.343	0.871	-0.051	0.032	0.062	-0.102	0.137	-0.026	-0.051
HMSB	0.021	-0.094	0.092	0.055	0.238	-0.011	-0.068	-0.312	-0.788	0.443	0.054
RPB	0.032	-0.072	0.086	0.069	0.199	0.080	-0.037	0.439	-0.166	-0.237	0.811
RDB	0.042	-0.111	0.134	-0.009	0.251	-0.242	0.357	0.711	-0.132	0.176	-0.406
TPB	0.104	-0.219	0.396	-0.303	-0.459	0.351	0.568	-0.092	-0.135	-0.067	0.067
TDB	0.079	-0.174	0.239	-0.180	-0.016	-0.820	0.136	-0.253	0.201	0.087	0.266
Proportion of Variance	0.987	0.005	0.005	0.001	0.0006	0.0004	0.0003	0.0002	0.0002	0.0002	0.0001

CHAPTER 6: DISCUSSION

MEASUREMENT ERROR AND RELIABILITY

Utilization of radiographic images as a data source mitigates two major problems surrounding the creation of techniques for the subadult biological profile. First, the lack of large samples of modern children in skeletal collections from which to create or validate techniques; and second, the problems associated with mortality bias in cemetery populations. High agreement between measurements obtained from radiographic images of dry bones and the dry bones themselves (Stull et al., 2013) and the low intra- and inter-observer error in the measurements obtained in radiographic images indicate that Lodox Statscan-generated images are a valid data source to obtain metric variables from the subadult skeleton. Thus, large, modern samples are acquirable from which to evaluate variation in subadults.

An evaluation of published literature demonstrates the intra- and inter-observer TEM and %TEM obtained in the current study are comparable even though the specifics for each study vary (i.e. anthropometrics, craniometrics) (Utermohle et al., 1983; Ulijaszek and Kerr, 1999; Cardoso, 2005; WHO Multicentre Growth Reference Study Group, 2006; Sicotte et al., 2010). The largest measurement errors in the current study were associated with the smallest measurements, specifically, UMSB (inter-observer %TEM = 1.92; inter-observer %TEM = 1.84) and RMSB (inter-observer %TEM = 2.27). Comparisons of measurement error obtained from skeletal data are limited. One exception is Cardoso (2005), however, the only available comparative measurement is femur diaphyseal length. Cardoso (2005) reported an intra-observer TEM of 0.40 mm, which is smaller than the intra- and inter-observer error observed in the current study (0.91 mm and 1.58mm, respectively).

The Bland-Altman plots indicate no systematic bias; larger measurements did not incur larger errors. Furthermore, most measurements were within the upper and lower agreement levels, which were approximately +/- 2 mms. Error between the two observers can be attributed to observer experience as well as the option of choosing either left or right-sided elements for each of the 15 individuals. The results indicate that the repeatability of the measures is high.

The measurements applied in this study are similar to the measurements associated with standard data collection for adults, thus the variables are not drastically different from normal standards, which may attribute to the high consistency in measurements (Fazekas and Kósa, 1978; Buikstra and Ubelaker, 1994). Consequently, measurements can be reliably obtained from Lodox Statscan-generated images and the results, including the measurement definitions, are applicable. The reliability of the measurements indicates that sources other than skeletal collections can act as a data source for anthropologists interested in evaluating subadult or adult skeletal structures. Additionally, the definitions presented herein can be applied to skeletal remains – because the measurements are not specific to radiographic images – which would permit direct comparison between samples.

AGE ESTIMATION

Diaphyseal dimensions were historically acknowledged as having the potential to be good indicators of subadult age as seen in longitudinal studies that predict diaphyseal lengths from age and the few anthropological publications that attempted to predict age from diaphyseal lengths (i.e. Maresh, 1970; Gindhart, 1973; Hoffman, 1979; Rissech et al., 2008, 2013; López-Costas et al., 2012). However, previously published studies are inadequate for two reasons: 1) inappropriate statistics and 2) unsuitable samples. Black and coloured South African children aged six to ten years have significantly increased in height and weight between 1962 and 2013 (Hawley et al., 2009; Anholts, 2013). In the current study over 70% of the sample was born after 2000. Of the sample that was born before 2000, most were post 1996 though the earliest date of birth was 1992. Thus, the data should reflect the variation within the current South African population. Most published age estimation research is derived from historic samples, which ignores secular changes and is an inadequate reflection of the population.

The current study is also the first to provide suitable statistical analyses to estimate age at death. MARS and basis spline models allow for the creation of 95% prediction intervals, which offer an appropriate technique to estimate age in subadults. Overall, MARS models are recognized as presenting with a good bias-variance trade-off. MARS models are flexible enough to model nonlinearity and variable interactions (low bias) while the basis functions prevent too

much flexibility (low variance) (Milborrow, 2013). Diaphyseal dimensions exhibit different velocities through growth, hence the square or cube root transformation of age in all age estimation models and the need for a flexible fit within the model. The results of this study are in contrast to recently published literature that show linear relationships between age and the diaphyseal lengths of the tibia, humerus, and femur, humerus proximal breadth (transverse diameter breadth), humerus distal (epicondylar) breadth, and tibia proximal breadth (Rissech et al., 2008, 2013; López-Costas et al., 2012). The appearance of linearity is most likely a product of the small sample sizes. Thus, the current study offers the capability to generalize to the entire population and employs statistical methods that satisfy *Daubert* criteria.

PREDICTION INTERVALS, BIAS, STANDARD ERRORS AND R-SQUARED

Dynamic prediction intervals were calculated for each model that compensated for smaller variation in diaphyseal dimensions at younger ages and larger variation in diaphyseal dimensions in older ages. Using FDL as an example, the 95% prediction interval for the smallest diaphyseal length (75 mm) is approximately a three month interval (0.04 years to 0.33 years). When the femoral diaphysis is 348 mm the associated 95% prediction interval is approximately five years (8.5 years to 13.4 years). The lower prediction intervals of all models were adjusted to only include the age range of the collected sample. For example, the size of FDL ranged from 73 mm to 418 mm and the first two measurements (i.e. 73 mm and 74 mm) had lower 95% prediction intervals as negative ages. Femoral diaphyses greater than 398 mm in length have fitted values of 12 years and upper 95% prediction intervals of 15 years of age. Because the current sample did not amass data on 13 through 15 year olds, prediction intervals that extend beyond 12 years of age (inclusive) should be substantiated with additional age indicators/techniques such as epiphyseal fusion or dental formation. If long bone lengths are the only age indicator available, then one must acknowledge that the ages included in the estimate were not incorporated in the original dataset.

In the published literature, the standard error is frequently supplied and offers a probable error for future prediction. However, a prediction interval offers a value in which future observations should fall and is ultimately more informative than the standard error and is

required in most court systems (Stine, 1985). Because the models were semi- or nonparametric, the prediction intervals were obtained indirectly through resampling. The resampling resulted in cross-validated prediction intervals and fitted values, offering an explicit interval with upper and lower bounds. A holdout sample was used as a means to gauge model accuracy and generalization performance. Using femur diaphyseal length on a discrete sample of 28 individuals resulted in 100% of the individuals having a chronological age within the 95% prediction intervals. Results of the validation test confirm the applicability of the age at death models.

The majority of age estimation models had decreased precision as age increased though the models retained high accuracy because the 95% prediction intervals accounted for error in the point estimates. Overall, bias, or precision of the point estimate, is generally of less importance than accuracy because anthropologists should only use point estimates to measure model performance and not as an age estimate. Mean bias for the created models were effectively zero, but bias error increased as age increased. For the diaphyseal length models, the trend for underestimating age does not seem to be evident until approximately 10 years of age; females entering their adolescent growth spurt would normally explain the underestimation. However, a sex-specific trend was not noted in the bias error for diaphyseal lengths which indicate other factors are likely responsible. For example, South African black, white, and coloured children aged between 6 and 10 years demonstrate statistically significant differences in height. Thus, ancestry should be further explored as an influential factor in age estimation. The bias for diaphyseal breadths showed a larger discrepancy between the sexes as well as the underestimation of age for females even though the overall trend for both sexes was the same. The pattern in bias for diaphyseal breadths is likely due to the higher number of significant differences between males and females in breadth than in length measurements.

Sex-specific MARS femur diaphyseal length and distal breadth models were created to evaluate sex differences. Both male and female FDL models demonstrated similar R-squared and SEs and both had three hinge functions. However, the three hinges were required approximately 1 to 2 years earlier (~20 mm) in females than males. Essentially, the female trajectory changes at an earlier age as females present with larger diaphyses at younger ages (Figure 6.1). In contrast, to the sex-specific diaphyseal length models, the hinge functions in the sex-specific FDB models were both at 68 mm and approximately 10 years of age. Although the bias error in the original

breadth models demonstrated a larger disparity between the sexes than the diaphyseal length models, the trajectory of the FDB is similar between males and females.

The R-squared values and SEs in the current study are larger and smaller, respectively, when compared to all available published models (Table 6.1). The one exception is the lower SE of the tibia distal breadth model presented by López-Costas et al. (2012), however, this is likely an artifact of a small sample size. Although few studies acknowledge the value of diaphyseal breadth measurements in age estimation, the majority of the univariate models in the current study – including breadths – resulted with SEs that are comparable or smaller than previously published work of Rissech et al. (2008, 2013) and López-Costas et al. (2012). For example, in the current study, only one measurement – the humerus midshaft breadth – resulted in a SE as wide as 2 years while every other univariate model achieved a SE between 0.90 and 1.8 years. In contrast, the published literature provided SEs that ranged from 1.19 years to 2.32 years; SEs that are on average much wider than this study. R-squared values of the current study are also higher than recently published R-squared values obtained from linear regression models (i.e. Rissech et al., 2008, 2013; López-Costas et al., 2012) (Table 6.1). However, it is acknowledged that the R-squared value is inadequate to compare different studies to one another because it does not follow a distribution (Hawkins, 2004).

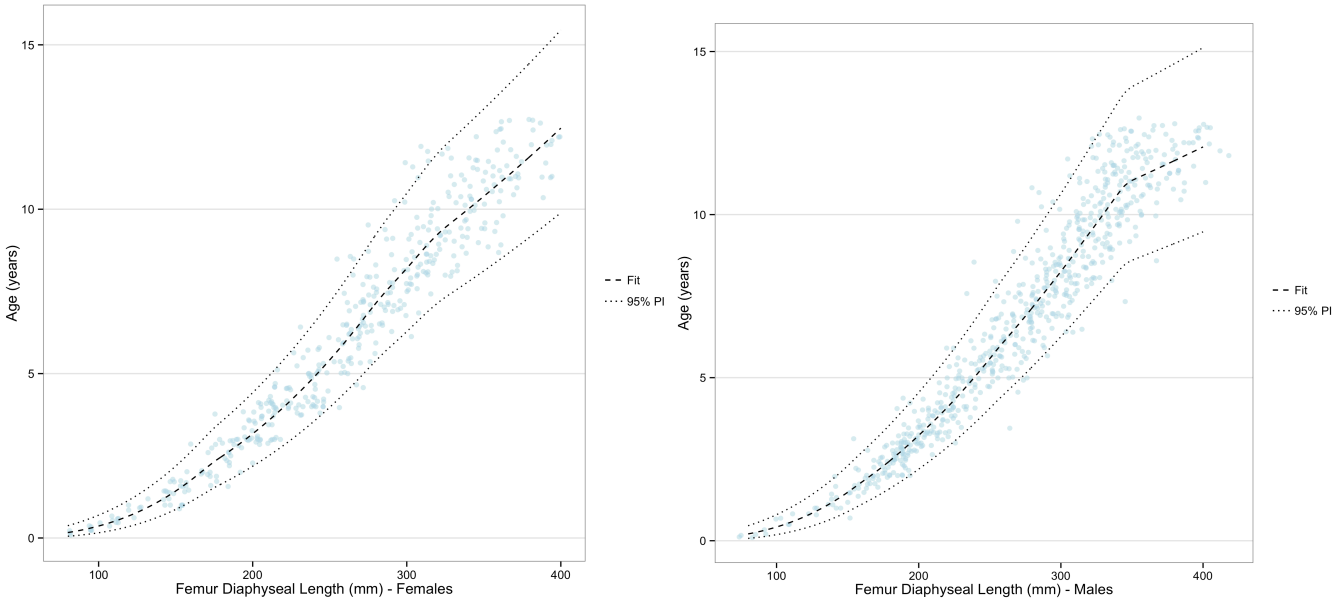


Figure 6.1 – Sex-specific femur diaphyseal length models.

Though the SE provides a means to compare models, the lack of associated prediction intervals in the published studies does not meet *Daubert* criteria. The published studies (i.e. López-Costas et al., 2012; Rissech et al., 2013) only provide a SE with their formulae and the authors state the SE is not a reliable indicator until 8 years of age as the value does not adjust to the increased variance in the measurement as age increases. However, even when the age is greater than 8 years, the SE is still inflexible and unable to adjust to the increase in variance that coincides an increase in age. Besides being fixed and inapplicable at the lower ages, the application of SE in forensic estimates still fails to fulfill *Daubert*, as the statistic is a measure of the error of the mean and not the population.

Table 6.1 – Comparison of the SE and adjusted R-squared values obtained in the current study compared to recently published SE and R-squared values of the same measurements. All models were with sexes combined.

Measure	South African Sample		Literature	
	SE	*Adj R2*	*SE**	*Adj R2**
HDL	0.97	0.95	1.39	0.87
HPB	1.60	0.83	2.32	0.84
HDB	1.55	0.85	2.15	0.83
FDL	0.90	0.95	-	0.93
TDL	0.95	0.94	1.66	0.87
TDB	1.35	0.80	1.19	0.88

***Humerus variables (Rissech et al., 2013); femur variables (Rissech et al., 2008); tibia variables (López-Costas et al., 2012)**

MODEL SELECTION: AGE ESTIMATION

In the current study standard errors were smaller in most of the multivariate models than the univariate models with the exception of HDL, FDL and TDL (Table 6.2). Evaluation of the cross-validated prediction intervals associated with each multivariate model reveal that the intervals are flexible but the models are not applicable to younger ages (~ < 2 years). The inapplicability to younger ages is a result of sampling, as sample sizes were small for the breadth measurements in the younger age categories. In contrast, prediction intervals associated with univariate diaphyseal length models allow for a more narrow prediction interval at the younger ages. Essentially, univariate models provide narrower estimates at the younger ages but the multivariate models provide narrower estimates at the older ages. If the available data includes the femur, ulna, or humerus diaphyseal length, the breadths do not increase the likelihood of a narrower prediction interval. In application of a multivariate model, the PCA models will yield more accurate age estimations than the original variables because MARS cannot handle multicollinearity.

The preference of a multivariate model at the older ages is demonstrated using the femur distal breadth model that yielded age ranges as wide as nine years in the older groups. In contrast, the multivariate subsets offer age ranges of five or six years for comparable age groups. Because a diaphyseal length was always chosen in the multivariate models, a model was created excluding diaphyseal lengths (breadths-only model). The SE was 1.22 years with an associated cv Rsq of 0.78 and the bias was +/- 3 years; the results are comparable or better than most univariate breadth models (Table 6.2). Evaluation of the cross-validated prediction intervals revealed age ranges between 2 and 7 years, which are slightly larger than some of the univariate breadth models, but not the majority of them. Generally, the multivariate breadths-only model followed the same tendencies as the other multivariate models, which is to have wider intervals at younger ages but narrower intervals at the older ages than the univariate models. Furthermore, the bias associated with the multivariate breadths-only model did not exhibit the same discrepancies between males and females as the univariate breadth models. Thus, if diaphyseal lengths are not available, the application of a multivariate model is encouraged prior to a univariate breadth model.

The difference between the applications of a univariate versus multivariate approach is related to the relationship each measurement has with age and how this changes with growth. For example, if data are separated into age intervals that generally follow the subdivisions of growth then the correlations between age and diaphyseal length change (Table 6.3). Strong correlations are noted between the lengths and age in the two youngest age groups, which indicates the inclusion of multiple diaphyseal lengths will not improve the model. The correlations gradually weaken as age increases, which suggest a multivariate model is more appropriate for older children. The causes for a weaker relationship with age are due to a multitude of factors all of which result in adults of different sizes and proportions. Furthermore, the majority of breadth measurements only have a range of 15 mm; the range is small and thus the variation is high, especially in the older ages. The inclusion of more variables results in a reduction in the size of variation and thus produces a narrower age estimate.

For all models, the widest prediction intervals are noted in the period of adolescence, which is the transition from childhood to adulthood. During this period children increase in size, however, the timing – or tempo of growth – is highly variable (Hauspie and Roelants, 2012). The inclusion of both males and females within the estimates further increases variation, as individuals – within each sex and between the sexes – present with different ages for the onset of puberty. Thus, large differences in diaphyseal dimensions are apparent in children of the same age and subsequently the prediction intervals compensate for this variability. Although the provided prediction intervals appear wider than most anthropologists typically provide for subadults, it is a proper prediction interval that compensates for 95% of the population obtaining an accurate age estimation.

Table 6.2 – Aggregation of all of the models, standard errors, and adjusted R-squared values, in order from smallest to largest standard error.

	Model	SE	Adj R2
Multivariate (PCA)	MARS	0.77	0.92
Distal	cbrt MARS	0.78	0.93
Distal (PCA)	cbrt MARS	0.79	0.93
Multivariate	cbrt MARS	0.8	0.93
Lower (PCA)	cbrt MARS	0.85	0.92
Lower	cbrt MARS	0.86	0.92
Proximal	cbrt MARS	0.87	0.92
Proximal (PCA)	cbrt MARS	0.88	0.91
FDL	cbrt MARS	0.9	0.95
TDL	sqrt MARS	0.95	0.94
HDL	cbrt MARS	0.97	0.95
Radius (PCA)	cbrt MARS	0.97	0.96
Tibia (PCA)	cbrt MARS	0.97	0.91
Radius	sqrt MARS	0.98	0.95
Tibia	cbrt MARS	0.98	0.91
FBDL	sqrt MARS	0.99	0.93
Upper	cbrt MARS	0.99	0.95
UDL	cbrt MARS	1.01	0.93
RDL	cbrt MARS	1.02	0.94
Upper (PCA)	cbrt MARS	1.04	0.95
TDB	spline (2,4)	1.35	0.8
RPB	sqrt MARS	1.38	0.87
TPB	sqrt MARS	1.39	0.81
RDB	cbrt MARS	1.47	0.86
FDB	sqrt MARS	1.51	0.77
HDB	cbrt MARS	1.55	0.85
TMSB	cbrt MARS	1.59	0.72
HPB	spline (2,3)	1.6	0.83
FMSB	cbrt MARS	1.62	0.75
RMSB	cbrt MARS	1.76	0.81

Table 6.3 – Pearson correlation coefficients of diaphyseal lengths and age that demonstrate how the relationship between the two weakens as age increases.

	Age (years)		
	<4 (n=199)	*4 – 8 (n=332)*	*9 – 12 (n=165)*
HDL	0.881	0.843	0.377
RDL	0.867	0.809	0.363
UDL	0.862	0.807	0.369
FDL	0.913	0.853	0.435
TDL	0.897	0.82	0.417
FBDL	0.898	0.826	0.409

SEXUAL DIMORPHISM IN DIAPHYSEAL DIMENSIONS

The current study is the first to discover sex differences in absolute diaphyseal breadth measurements from infancy through adolescence. Although breadth measurements are recognized to perform well as discriminators between adult males and females (Pearson, 1915; France, 1998; Spradley and Jantz, 2011; Tise et al., 2013), attempts to apply them to subadult postcrania is rarely noted in the literature. In this study, the most marked sex differences were in diaphyseal breadths between the ages of two and ten years. The absence of sex differences among children within the first two years of life may be the consequence of sampling, as positioning a young individual (~ < 2 years) into proper anatomical position is difficult and often not performed. Thus, the majority of the breadth measurements on the lower limbs were not possible to collect in the younger samples. For example, FDB was one measurement that exhibited the most sexually dimorphic differences throughout many of the sampled ages; however, in the less than 1 year age interval, only five individuals had an associated

measurement. Following 10 years of age, sexually dimorphic differences are less apparent as females and males present with differing growth rates, which subsequently affect size and shape related differences. Sex differences – especially of the upper limb – decrease as some individuals, especially girls, enter adolescence. Adolescence is marked by a sudden and rapid increase in growth rate, thus if some children present the transition to adolescence earlier, they will likely exhibit larger sizes than those who have not yet transitioned into adolescence (Bogin, 1997). The greatest factor in early and late onset ages is attributed to sex differences (Cameron, 2012; Hauspie and Roelants, 2012); but, as previously mentioned, population differences may also be an influential factor and should be considered in future analyses.

The humerus and femur consistently resulted in the largest number of mean significant differences by age compared to the other four elements. With regard to measurements, the HPB, TPB and FDB showed the highest number of statistically significant mean differences, though the ratios including them were more sexually dimorphic than the original variables. The results of this study contradict the few studies that have examined diaphyseal breadths for sex differences. Particularly, Rissech et al. (2013) did not identify statistically significant differences in the humeral proximal breadth or humeral distal breadth and López-Costas et al. (2012) did not identify any statistically significant differences in the tibia distal breadth. The differences in the results are likely attributed to the five-year age intervals the above-mentioned authors had grouped their samples which likely obscured sexually dimorphic patterns. Further, the population differences in the studies may also be a reason for differential results.

In the current study, midshaft measurements rarely showed significant differences between the sexes. The lack of sex differences in midshaft measurements supports previous research that did not find any significant differences in midshaft cross-sectional area and cortical area of prepubertal boys and girls matched for age, height, and weight (Gilsanz et al., 1997; Schoenau et al., 2002; Nieves et al., 2005).

In contrast to results based on longitudinal growth studies in the United States (Maresh, 1970; Gindhart, 1973; Smith and Buschang, 2004), no sexually dimorphic differences were found in any of the absolute diaphyseal lengths in South Africans from birth to twelve years in the current study. Multiple researchers observed the tendency of North American males to have absolutely longer radii and tibiae than females (Maresh, 1970; Gindhart, 1973; Smith and Buschang, 2004; Smith, 2007), but this tendency was not evident in the South African sample.

As expected from the Student's t-tests results, the sympercent differences of the diaphyseal lengths are minimal (~1 sp) between the sexes from approximately two years until approximately 10 years of age. Similar results were demonstrated by Clark et al. (2007) who identified a 1% difference in humeral diaphyseal length in a large modern sample of 9 year old English children – with females being larger than males. Earlier maturation and attainment of an earlier peak height velocity in females is likely responsible for the increased disparity in diaphyseal lengths (mean 4 sp) in the current sample around 12 years of age, with larger differences in the lower limbs than upper limbs. Previous research has recognized that the growth spurt in the lower limbs of females is concurrent to peak height velocity while the growth spurt in the upper limbs of females occurs approximately one year later (Smith and Buschang, 2005). Smith and Buschang (2005) also noted the earlier maturation of the girls produces longer diaphyseal lengths with differing magnitudes for each bone. As in the current study, the sex differences in the Child Research Council sample were greater in the lower limb elements than in the upper limbs (Smith and Buschang, 2005).

Furthermore, the lack of significant differences in lengths suggests a coinciding lack of differences in stature. A Student's t-test revealed no sex differences in stature or weight by age in the current South African sample. The pattern of reduced sexual dimorphism in the diaphyseal lengths follows a similar pattern to body weight and stature; more than half of the dimorphism that is apparent in adults is supposed to result from the slower development of the male compared to the female in adolescence (Willner and Martin 1985, Harrison et al 1988, Humphrey 1998). Because body size was accounted for in the sample, as no differences were found in height or weight at any ages, the differences in diaphyseal breadths can be related to factors independent of body size, such as muscle mass and sex-specific tissue responses.

MODEL EVALUATION AND APPLICABILITY: SEX ESTIMATION

The current research was able to elucidate sex differences in the skeletal dimensions of males and females because of the large sample size and multivariate approach. Utilizing FDA or logistic regression with a large number of predictor variables, an anthropologist can accurately

estimate sex between birth and 12 years with up to 90% accuracy. The classification methods utilizing multivariate subsets achieved higher accuracies than the univariate and multivariate bone models. The univariate and multivariate bone models obtained classification accuracies that were comparably as low as previously published metric and morphological sex estimation techniques (Choi and Trotter, 1970; Weaver, 1980; Schutkowski, 1987; Holcomb and Konigsberg, 1995; Scheuer, 2002; Galdames et al., 2008). Higher accuracies for multivariate subsets when compared to univariate and multivariate bone models are expected for several reasons. The multivariate subset includes the greatest number of variables and as previously stated, when estimating a dichotomous biological variable a large amount of overlap is usually present. Generally the use of more measurements maximizes differences among groups, especially in discriminant analyses, and thus in turn increases correct classification (Ousley and Jantz, 2012). The hypothesis tests, particularly the t-tests, suggested that the best discriminators would be the ratios. However, the three variables included in the ratios subset compared to the eleven in the all-measurement subset may be the explanation for the lower correct classifications.

Of all the classification techniques employed, logistic regression and FDA consistently yielded the highest number of correct classifications. Logistic regression was superior to FDA when employed on univariate models but FDA was comparable or superior to logistic regression when employed on multivariate models. Both methods are recognized as more flexible than QDA and LDA and thus tend to outperform other classification methods when nonlinear relationships are evident. FDA and logistic regression were expected to, and did, outperform the other classification methods in the current sample. Because both logistic regression and FDA perform poorly if multicollinearity is present, the ability of FDA and logistic regression would likely improve if PCA were applied (De Veaux and Ungar, 1994). Even though the classifications were only slightly increased with the use of PCA, the models including the PC scores are nonetheless superior to the utilization of the original variables because the transformations associated with PCA removes multicollinearity amongst diaphyseal measurements. Thus, the PC scores allow for a more reliable model that permits practical use within the field of anthropology.

The results are examples of the classification potential of long bone dimensions when all ages were combined. Although age and sex estimations can be conducted independent of the other, the number of correct classifications may increase if a smaller age range is included in the

sex estimation analyses. For example, although the proximal subset achieved a 78% correct with FDA with the inclusion of all ages, a smaller age range from 5 to 8 years resulted in a classification accuracy of 84%. However, a smaller age range does not automatically result with an increased accuracy. For example, an age range of 2 to 4 years also achieved a higher classification accuracy (82%) than the original model but an age interval of 9 to 12 years achieved a lower classification accuracy (76%) than the original model. The classification accuracy will be specific to each case, the measurements available and the age of the individual as the levels of sexual dimorphism vary through the ages and vary among the measurements. Thus it is not possible to guarantee the result of a higher classification rate with the inclusion of an age estimate.

SOFTWARE PROGRAM

The vast number of age at death and sex estimation tables required to represent all permutations of multivariate models would not be concise nor would it be an easily accessible resource for anthropologists. The age-at-death tables provided in Appendices III - IX are based on the analyses conducted within this dissertation. Although the univariate models are immediately applicable, the multivariate examples are specific to the combination of variables chosen and were chosen as examples of the potential of a multivariate model. Multivariate age estimation models are necessary to employ, especially when presented with remains of older individuals, and will require cross-validated 95% prediction intervals specific to each situation.

Similarly, sex estimation with high accuracy is dependent on more rather than fewer predictor variables. Furthermore, the inclusion of age estimates and many predictor variables may drastically reduce sample sizes, so resampling in the form of bootstrapping is necessary to ensure a realistic rather than an optimistic result. For both age and sex estimations, the number of possible permutations is too large for one to provide all possible combinations and thus a software program is required to account for all possible multivariate models and meet the requirement of cross-validated prediction intervals or bootstrapped classification accuracies, which will be specific to each skeletal analysis.

CHAPTER 7: CONCLUSION

The current study is the first to successfully estimate subadult age and sex using an extensive number of measurements and univariate and multivariate methods. The successful results of the current study are directly related to the largest, modern sample size ever collected for subadult diaphyseal dimensions. Subadult studies have been plagued by small sample sizes, which ultimately affects their validity. The use of Lodox Statscan-generated images have proven to be a valuable resource as there is minimal distortion noted in the radiographic images and thus the images provide a means to collect data from many subadults. The populations that comprised this sample are reflective of the current South African population, which allowed for the creation of proper country-specific techniques.

In contrast to previously published age at death studies, the current research is the first to present statistics appropriate for predicting age that include 95% prediction intervals along with point age estimates. Although both basis splines and MARS were equally explored as viable age estimation models, MARS outperformed basis splines in almost every situation. MARS makes no assumptions about the underlying functional relationship between the dependent and independent variables nor does MARS imply causality between the predictor and response variables (Butte et al., 2010). The latter is especially appropriate because age of an individual does not cause a specific diaphyseal length. Most importantly, MARS is flexible enough to model nonlinear relationships and allows for the interactive effects of the predictor variables in estimating the response variable (Hastie and Tibshirani, 1990; Muñoz and Felicísimo, 2004). Because MARS can handle a large number of predictor variables, the current study offers the first multivariate approach to subadult age estimation utilizing long bone lengths.

Each MARS model has associated 95% prediction intervals derived from cross-validation resampling, which results in the first age at death technique – utilizing long bone lengths – that fulfills *Daubert* criteria for presenting results with known error rates. Differences between the sexes were evident, but the variation between the sexes was smaller than the 95% prediction intervals associated with the age estimation models. Nonetheless, if sex were known, a sex-specific age at death estimation model would yield smaller age estimates. The application of a univariate model versus a multivariate model will be dependent on the remains being analyzed

with the suggestion to employ a univariate model for remains of younger individuals and multivariate model for remains of older individuals.

Previous sex estimation techniques were generally conducted on one element and did not capture the differences exhibited over numerous skeletal structures. In this study, univariate and multivariate bone models resulted in inapplicable classification accuracies. In contrast, the use of more variables led to increased classification accuracies that offer practical application in the field. Bootstrapped classification accuracies were evaluated because small sample sizes could lead to overly optimistic results. Breadth measurements were consistently recognized as the most sexually dimorphic measurements, particularly the knee and the shoulder, which is further supported by the biomechanical and endocrinological literature. The combination and interaction of the variables in the multivariate model achieve correct classifications that are comparable to commonly used sex estimation techniques in adults. Classification models were inclusive of all ages, but accuracies could increase if a smaller age range were provided. However, because the levels of sexual dimorphism vary through the ages and among the measurements, the best model for sex estimation will be specific to the unique combination of measurements available for each case.

Comparison of the six classification methods clearly demonstrated the better performance of the logistic regression and FDA. Both models require the fulfillment of fewer assumptions and are preferred when nonlinear relationships are apparent in the data. Though the logistic regression outperformed FDA in the univariate models, the correct classifications were either comparable or the FDA yielded higher accuracies in the multivariate models. Because FDA utilizes the MARS algorithm, the better results are not unexpected.

The current study provided only a subset of the potential variable combinations and results for age and sex estimation, as it is impossible and impractical to provide results for all the number of possible variable combinations. Thus, a computer software program will be created based on the South African subadult data that will allow for any possible variable combinations and provide the results of the different models employed with their associated error rates. Consequently, the user will be able to provide age and sex estimations specific to each anthropological analysis for South African children aged between birth and 12 years.

REFERENCES

Abdi H. 2010. Holm's Sequential Bonferroni Procedure. In: Salkind N, editor. Encyclopedia of Research Design. Thousand Oaks, California: SAGE Publications, Inc. p 574–578. Available from: http://knowledge.sagepub.com/view/researchdesign/n178.xml

Adhikari M. 2005. Contending approaches to coloured identity and the history of coloured people of South Africa. History Compass 3:1–6.

Aksglaede L, Juul A, Leffers H, Skakkebæk NE, Andersson A-M. 2006. The sensitivity of the child to sex steroids: possible impact of exogenous estrogens. Hum Reprod Update 12:341–349.

AlQahtani SJ, Hector MP, Liversidge HM. 2010. Brief Communication: The London Atlas of Human Tooth Development and Eruption. Am J Phys Anthropol 142:481–490.

Anholts A. 2013. Secular trends in the height and weight of South African children aged 6 to 10 years. Honours Thesis: University of Pretoria.

Arden NK, Spector TD. 1997. Genetic influences on muscle strength, lean body mass, and bone mineral density: a twin study. Journal of Bone and Mineral Research 12:2076–2081.

Arfai K, Pitukcheewanont PD, Goran MI, Tavare CJ, Heller L, Gilsanz V. 2002. Bone, Muscle, and Fat: Sex-related Differences in Prepubertal Children1. Radiology 224:338–344.

Bailey R, Bynes J. 1990. A New, Old Method for Assessing Measurement Error in both Univariate and Multivariate Morphometric Studies. Syst Zool 39:124–130.

Bailit H, Hunt EE. 1964. The sexing of childrens skeletons from teeth alone and its genetic implications. American Journal of Physical Anthropology 22:171–173.

Barbarin OA, Khomo N. 1997. Indicators of Economic Status and Social Capital in South African Townships What Do They Reveal About the Material and Social Conditions in Families of Poor Children? Childhood 4:193–222.

Barbarin OA, Richter L. 2001. Economic Status, Community Danger and Psychological Problems among South African Children. Childhood 8:115–133.

Barron AR. 1994. Approximation and estimation bounds for artificial neural networks. Mach Learn 14:115–133.

Beningfield S, Potgieter H, Nicol A, van As S, Bowie G, Hering E, Latti E. 2003. Report on a new type of trauma full-body digital x-ray machine. Emer Radio 10:23–29.

Bennett P. 2000. Assessing the calibration of Naive Bayes' posterior estimates. School of Computer Science: Carnegie Mellon University.

Berner D. 2011. Size correction in biology: how reliable are approaches based on (common) principal component analysis? Oecologia 166:961–971.

Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mouguiama-Daouda P, van der Veen L, Hombert J-M, Quintana-Murci L, Comas D. 2009. Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. Mol Biol Evol 26:1581–1589.

Black T. 1978. Sexual dimorphism in the tooth-crown diameters of the deciduous teeth. Am J Phys Anthropol 48:77–82.

Bland M, Altman D. 1986. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. Lancet:307–310.

Bock R, du Toit S. 2004. Parameter estimation in the context of non-linear longitudinal growth models. In: Hauspie R, Cameron N, Molinari L, editors. Methods in Human Growth Research. Cambridge University Press.

Boersma B, Wit J. 1997. Catch-up Growth. Endocrine Reviews 18:646–661.

Bogin B, Loucky J. 1997. Plasticity, political economy, and physical growth status of Guatemala Mayan children living in the United States. American Journal of Physical Anthropology 102:17–32.

Bogin B, Smith P, Orden AB, Varela Silva MI, Loucky J. 2002. Rapid change in height and body proportions of Maya American children. Am J Hum Biol 14:753–761.

Bogin B. 1997. Evolutionary Hypotheses for Human Childhood. Yrbk Phys Anthropol 40:63–89.

Bogin B. 1999. Patterns of Human Growth. Cambridge: Cambridge University Press.

Bontrager K. 2001. Textbook of Radiographic Positioning and Related Anatomy. 5th Edition. Philade: Mosby.

Boucher B. 1955. Sex differences in the foetal sciatic notch. Journal of Forensic Medicine 2:51–54.

Boucher B. 1957. Sex differences in the foetal pelvis. Am J Phys Anthropol 15:581–600.

Breiman L. 1996a. Bagging Predictors. Machine Learning 24:123–140.

Breiman L. 1996b. Out-of-bag estimation. Available from: http://ftp.stat.berkeley.edu/pub/users/breiman/OOBestimation.ps

Breiman L. 1999. Using adaptive bagging to debias regressions. Statistics Department, UCB.

Breiman L. 2001. Random Forests. Machine Learning 45:5–32.

Buikstra JE, Ubelaker DH. 1994. Standards for Data Collection from Human Skeletal Remains: Proceedings of a Seminar at the Field Museum of Natural History. Fayetteville: Arkansas Archaeological Research Series.

Burnham KP, Anderson DR, Huyvaert KP. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol 65:23–35.

Burnham KP, Anderson DR. 2004. Multimodel inference understanding AIC and BIC in model selection. Sociological methods & research 33:261–304.

Butte NF, Wong WW, Adolph AL, Puyau MR, Vohra FA, Zakeri IF. 2010. Validation of Cross-Sectional Time Series and Multivariate Adaptive Regression Splines Models for the Prediction of Energy Expenditure in Children and Adolescents Using Doubly Labeled Water. J Nutr 140:1516–1523.

Cabo L, Brewster C, Azpiazu J. 2012. Sexual Dimorphism: Interpreting Sex Markers. In: Dirkmaat D, editor. A Companian to Forensic Anthropology. Malden, MA: John WIley & Sons. p 248–286.

Calcagno J. 1981. On the applicability of sexing human skeletal material by discriminant function analysis. Journal of Human Evolution 10:189–198.

Cameron N, Grieve C, Kruger A, Leschner K. 1993. Secondary sexual development in rural and urban South African black children. Ann Hum Biol 20:583–93.

Cameron N, De Wet T, Ellison G, Bogin B. 1998. Growth in Height and Weight of South African Urban Infants From Birth to Five Years: The Birth to Ten Study. Am J Hum Biol 10:495–504.

Cameron N. 2003. Physical growth in a transitional economy: the aftermath of South African apartheid. Economics & Human Biology 1:29–42.

Cardoso H, Saunders S. 2008. Two arch criteria of the ilium for sex determination of immature skeletal remains: A test of their accuracy and an assessmend of intra- and inter- observer error. Forensic Sci Int 178:24–29.

Cardoso H. 2005. Patterns of Growth and Development of the Human Skeleton and Dentition in Relation to Environmental Quality: A biocultural analysis of a sample of 20th century Portuguese subadult documented skeletons. Dissertation: McMaster University.

Cardoso H. 2007a. Differential Sensitivity in Growth and Developement of Dental and Skeletal Tissue to Environmental Quality. Arquivos De Medicina 21:19–23.

Cardoso HF. 2008a. Epiphyseal Union at the Innominate and Lower Limb in a Modern Portuguese Skeletal Sample, and Age Estimation in Adolescent and Young Adult Male and Female Skeltons. American Journal of Physical Anthropology 135:161–170.

Cardoso HFV. 2007b. Environmental Effects on Skeletal Versus Dental Developement: Using a Documented Subadult Skeletal Sample to Test a Basic Assumption in Human Osteological Research. American Journal of Physical Anthropology 132:223–233.

Cardoso HFV. 2008b. Age Estimation of Adolescent and Young Adult Male and Female Skeletons II, Epiphyseal Union at the Upper Limb and Scapular Girdle in a Modern Potuguese Skeletal Sample. American Journal of Physical Anthropology 137:97–105.

Chen T, Martin E, Montague G. 2009. Robust probabilistic PCA with missing data and contribution analysis for outlier detection. Computational Statistics & Data Analysis 53:3706–3716.

Choi S, Trotter M. 1970. Statistical study of the multivariate structure and race-sex differences of American White and Negro fetal skeletons. Am J Phys Anthropol 33:307–312.

Christensen AM. 2004. The Impact of Daubert: Implications for Testimony and Research in Forensic Anthropology (and the Use of Frontal Sinuses in Personal Identification). Journal of Forensic Sciences 49:1–4.

Christopher A. 2002. "To Define the Indefinable": Population Classification and the Census in South Africa. Area 34:401–408.

Clark EM, Ness AR, Tobias JH. 2007. Gender differences in the ratio between humerus width and length are established prior to puberty. Osteoporos Int 18:463–470.

Cole TJ. 2000. Sympercents: symmetric percentage differences on the 100 loge scale simplify the presentation of log transformed data. Statistics in Medicine 19:3109–3125.

Conceição E l. n., Cardoso H f. v. 2011. Environmental effects on skeletal versus dental development II: Further testing of a basic assumption in human osteological research. American Journal of Physical Anthropology 144:463–470.

Cowgill LW, Eleazer CD, Auerbach BM, Temple DH, Okazaki K. 2012. Developmental variation in ecogeographic body proportions. American Journal of Physical Anthropology 148:557–570.

Crowder C, Austin D. 2005. Age Ranges of Epiphyseal Fusion in the Distal Tibia and Fibula of Contemporary Males and Females. J Forensic Sci 50:1–7.

Cutler A, Zhao G. 2001. PERT - perfect random tree ensembles. Computing Science and Statistics:497–504.

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ. 2007. Random Forests for classification in ecology. Ecology 88:2783–2792.

Dahlberg G. 1940. Statistical Methods for Medical and Biological Students. London: George Allen & Unwin, Ltd.

Daly RM, Saxon L, Turner CH, Robling AG, Bass SL. 2004. The relationship between muscle size and bone geometry during growth and in response to exercise. Bone 34:281–287.

Daubert vs. Merrell Dow Pharmaceuticals. 1993. 113 Supreme Court 2786.

Dayal M, Kegley A, Strkalj G, Bidmos M, Kuykendall K. 2009. The History and Composition of the Raymond A. Dart Collection of Human Skeletons at the University of the Witwatersrand, Johannesburg, South Africa. American Journal of Physical Anthropology 140:324–335.

Demirjian A, Buschang PH, Tanguay R, Patterson DK. 1985. Interrelationships among measures of somatic, skeletal, dental, and sexual maturity. Am J Orthod 88:433–438.

Demirjian A, Goldstein H, Tanner JM. 1973. A new system of dental age assessment. Hum Biol 45:211–227.

Diamond J, Bellwood P. 2003. Farmers and Their Languages: The First Expansions. Science 300:597–603.

Dirkmaat D, Cabo L, Symes S, Ousley S. 2008. New perspectives in forensic anthropology. American Journal of Physical Anthropology 137:33–52.

Douglas T, Pitcher R, van As A. 2010. Full-body digital radiographic imaging of the injured child. Continuing Medical Education 28:108–112.

Edward E. Hunt, Gleiser I. 1955. The estimation of age and sex of preadolescent children from bones and teeth. American Journal of Physical Anthropology 13:479–487.

Efron B, Tibshirani R. 1993. An introduction to the bootstrap. New York: Chapman & Hall.

Efron B. 1979. Bootstrap Methods: Another Look at the Jackknife. Ann Statist 7:1–26.

Eilers PH, Marx BD. 1996. Flexible smoothing with B-splines and penalties. Statistical science:89–102.

Eveleth P, Tanner J. 1990. Worldwide Variation in human growth. 2nd Edition. Cambridge: Cambridge University Press.

Eveleth P. 1978. Population Differences in Growth. Environmental and Genetic Factors. In: Falkner F, Tanner J, editors. Human Growth. Vol. 1. New York: Plenum. p 221–239.

Everitt B, Dunn G. 2001. Applied Multivariate Data Analysis. 2nd ed. London, UK: Arnold.

Fazekas S, Kósa F. 1978. Forensic Fetal Osteology. Budapest: Akademiai Kiado.

Fechner P. 2003. The biology of puberty: New developments in sex differences. In: Hayward C, editor. Gender differences at puberty. Cambridge: Cambridge University Press. p 17–28.

Feldesman M. 1992. Femur/stature ratio and estimates of stature in children. Am J Phys Anthropol 87:447–459.

De Filippo C, Bostoen K, Stoneking M, Pakendorf B. 2012. Bringing together linguistic and genetic evidence to test the Bantu expansion. Proceedings of the Royal Society B: Biological Sciences 279:3256–3263.

Filzmoser P, Gschwandtner M. 2013. mvoutlier: Multivariate outlier detection based on robust methods. Available from: http://CRAN.R-project.org/package=mvoutlier

Forest MG, Cathiard AM, Bertrand JA. 1973. Evidence of Testicular Activity in Early Infancy. JCEM 37:148–151.

Fox J, Weisberg S. 2011. An {R} Companion to Applied Regression. 2nd ed. Thousand Oads CA: Sage. Available from: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

France DL. 1998. Observation and Metric Analysis of Sex in the Skeleton. In: Reichs KJ, editor. Forensic Osteology: Advances in the indentification of human remains. Springfield: Charles C. Thomas.

Franklin D, Cardini A. 2007. Mandibular morphology as an indicator of human subadult age: interlandmark approaches. Journal of Forensic Sciences 52:1015–1019.

Franklin D, Oxnard CE, O'Higgins P, Dadour I. 2007. Sexual Dimorphism in the Subadult Mandible: Quantification Using Geometric Morphometrics*. J Forensic Sci 52:6–10.

Franklin D. 2010. Forensic age estimation in human skeletal remains: Current concepts and future directions. Legal Medicine 12:1–7.

Friedman J. 1993. Fast MARS. Stanford, CA: Department of Statistics, Stanford University.

Friedman JH. 1991. Multivariate Adaptive Regression Splines. The Annals of Statistics 19:1–67.

Friedman N, Geiger D, Goldszmidt M. 1997. Bayesian Network Classifiers. Mach Learn 29:131–163.

Frost HM, Schönau E. 2000. The "muscle-bone unit" in children and adolescents: a 2000 overview. J Pediatr Endocrinol Metab 13:571–590.

Frost HM. 1988. Vital biomechanics: proposed general concepts for skeletal adaptations to mechanical usage. Calcif Tissue Int 42:145–156.

Ben-Gal I. 2005. Outlier Detection. In: Maimon O, Rockach L, editors. Data Mining and Knowledge Discovery Handbook: A complete guide for practitioners and researchers. Kluwer Academic Publishers. p 131–146.

Galdames I, Matamala D, Smith R. 2008. Blind Test of Mandibular Morphology with Sex Indicator in Subadult Mandibles. Int J Morphol 26:845–848.

Garn S, Bailey S. 1978. Genetics of the maturational process. In: Falkner F, Tanner J, editors. Human Growth. Vol. 1. New York: Plenum. p 307–30.

Garn S, Keating M. 1980. Effect of various prenatal determinants on size and growth through seven years. Ecol Food Nutr 9:109–112.

Garn SM, Bailey SM, Cole PE. 1976. Similarities between parents and their adopted children. American Journal of Physical Anthropology 45:539–543.

Garnett SP, Högler W, Blades B, Baur LA, Peat J, Lee J, Cowell CT. 2004. Relation between hormones and body composition, including bone, in prepubertal children. Am J Clin Nutr 80:966–972.

Garvin HM, Passalacqua NV, Uhl NM, Gipson DR, Overbury RS, Cabo LL. 2012. Developments in Forensic Anthropology: Age-at-Death Estimation. In: Dirkmaat DC, editor. A Companion to Forensic Anthropology. John Wiley & Sons, Ltd. p 202–223.

Geeta A, Jamaiyah H, Safiza M, Khor G, Kee C, Ahmad A, Suzana S, Rahmah R, Faudzi A. 2009. Reliability, technical error of measurements and validity of instruments for nutritional status assessment of adults in Malaysia. Singapore Med J 50:1013–18.

Gilsanz V, Kovanlikaya A, Costin G, Roe TF, Sayre J, Kaufman F. 1997. Differential effect of gender on the sizes of the bones in the axial and appendicular skeletons. J Clin Endocrinol Metab 82:1603–1607.

Gindhart PS. 1973. Growth standards for the tibia and radius in children aged one month through eighteen years. American Journal of Physical Anthropology 39:41–48.

Goto R, Mascie-Taylor N. 2007. Precision of Measurement as a Component of Human Variation. J Physio Anthropol 26:253–256.

Green W, Wyatt G, Anderson M. 1946. Orthoroentgenography as a method of measuring the bones of the lower extremities. J Bone Joint Surg Am 28:60–65.

Groenewald P, Bradshaw D, Msemburi W, Neetheling I, Matzopolous R, Naledi T, Daniels J, Dombo M. 2011. Western Cape Mortality Profile 2009. Cape Town: South African Medical Research Council.

Grumbach M, Hughes I, Conte F. 2003. Williams Textbook of Endocrinology. (Larsen P, Kronenberg H, Melmed S, Polonsky K, editors.). Philadelphia: WB. Saunders.

Ha JH, Choi IH, Chung CY, Cho T-J, Jung ST, Lee H-S, Park S-S, Lee HY, Oh C-W, Kim IO. 2003. Distribution of lengths of the normal femur and tibia in Korean children from three to sixteen years of age. J Korean Med Sci 18:715–721.

Hair JF, Black W, Babin BJ, Anderson RE, Tatham RL. 2007. Multivariate Data Anaysis. Sixth. Upper Saddle River, NJ: Pearson Prentice Hall.

Hair JF, Black WC, Babin BJ, Anderson RE. 2010. Multivariate Data Analysis. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Hamill, Johnston, Lemshow. 1973. Body weight, stature, and sitting height: white and negro youths 12 - 17 years, United States. Washington DC; US Government Printing Office.

Harris EF, Smith RN. 2009. Accounting for measurement error: A critical but often overlooked process. Archives of Oral Biology 54, Supplement 1:S107–S117.

Hastie T, Tibshirani R, Buja A. 1993. Flexible Discriminant Analysis by Optimal Scoring. Journal of the American Statistical Association 89:1255–1270.

Hastie T, Tibshirani R, J Friedman. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer-Verlag.

Hastie T, Tibshirani R. 1990. Generalized Additive Models. London, UK: Chapman & Hall.

Hauspie R, Roelants M. 2012. Adolescent Growth. In: Cameron N, Bogin B, editors. Human Growth and Development. 2nd ed. New York: Elsevier. p 57–79.

Hawkins D. 2004. The Problem of Overfitting. J Chem Inf Comput Sci 44:1–12.

Hawley N, Rousham E, Norris S, Pettifor J, Cameron N. 2009. Secular trends in skeletal maturity in South Africa: 1962-2001. Ann Hum Biol 36:584–594.

Heller C, Scott E, Hammes L. 1967. Height, weight, and growth of alaskan eskimos. Am J Dis Child 113:338–344.

Henneberg M, Louw G. 1998. Cross-sectional Survery of Growth of Urban and Rural "Cape Coloured" Schoolchildren: Anthropometry and Functional Tests. Am J Hum Biol 10:73–85.

Heuzé Y, Cardoso HFV. 2008. Testing the quality of nonadult Bayesian dental age assessment methods to juvenile skeletal remains: The Lisbon collection children and secular trend effects. American Journal of Physical Anthropology 135:275–283.

Himes JH, Yarbrough C, Martorell R. 1977. Estimation of stature in children from radiographically determined metacarpal length. J Forensic Sci 22:452–455.

Hoffman JM. 1979. Age Estimation from Diaphyseal Lengths: Two Months to Twelve Years. J Forensic Sci 24:461–469.

Högler W, Blimkie C, Cowell C, Inglis D, Rauch F, Kemp A, Wiebe P, Duncan C, Farpour-Lambert N, Woodhead H. 2008. Sex-specific developmental changes in muscle size and bone geometry at the femoral shaft. Bone 42:982–989.

Holcomb S, Konigsberg L. 1995. Statistical study of sexual dimorphism in the human fetal sciatic notch. Am J Phys Anthropol 97:113–125.

Holm S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics:65–70.

Hoppa R, Vaupel J. 2002. The Rostock Manifesto for Paleodemography: the way from stage to age. In: Hoppa R, Vaupel J, editors. Paleodemography: age distributions from skeletal samples. Cambrige: Cambrige University Press.

Hoppa RD. 1992. Evaluating human skeletal growth: An Anglo-Saxon example. International Journal of Osteoarchaeology 2:275–288.

Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. 1997. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 16:965–980.

Houle D, Pelabon C, Wagner G, Hansen T. 2011. Measurement and Meaning in Biology. Q Rev Biol 86:3–34.

Huberty C. 1994. Applied Discriminant Analysis. New York: Wiley and Sons.

Humphrey LT. 1998. Growth patterns in the modern human skeleton. Am J Phys Anthropol 105:57–72.

Inwood K, Masakure O. 2009. The Historical Roots of Poverty and Inequality in South Africa: the Coloured population.

Jackson DA, Chen Y. 2004. Robust principal component analysis and outlier detection with ecological data. Environmetrics 15:129–139.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung H-C, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451:998–1003.

Johnston F. 1962. Growth of the long bones of infants and young children at Indian Knoll. Am J Phys Anthropol 20:249–254.

Johnston FE, Malina RM. 1966. Age changes in the composition of the upper arm in Philadelphia children. Hum Biol 38:1–21.

Jolicoeur P. 1963. The Multivariate Generalization of the Allometry Equation. Biometrics 19:497–499.

Jolliffe IT. 2002. Principal Component Analysis. 2nd ed. New York: Springer.

Jones TB, Wolfe HM, Zador IE. 1991. Biparietal diameter and femur length discrepancies: are maternal characteristics important? Ultrasound in Obstetrics and Gynecology 1:405–409.

Jungers W, Falsetti AB, Wall C. 1995. Shape, Relative Size, and Size Adjustments in Morphometrics. Yearbook of Physical Anthropology 38:137–161.

Kabacoff R. 2011. R in Action: Data Analysis and graphics with R. Shelter Island, NY: Manning Publications Co.

Kachigan S. 1991. Multivariate Statistical Analyses: A Conceptual Introduction. Second. New York: Radius Press.

Karlberg J. 1989. A Biologically-Oriented Mathematical Model (ICP) for Human Growth. Acta Pædiatrica 78:70–94.

Klingenberg C. 1996. Multivariate Allometry. In: Marcus L, Corti M, Loy A, Naylor G, Slice D, editors. Advances in Morphometrics. Vol. 284. NATO ASI Series A: Natural Sciences. New York: Plenum Press. p 23–49.

Knapp TR. 1992. Technical error of measurement: A methodological critique. American Journal of Physical Anthropology 87:235–236.

Knickmeyer R, Baron-Cohen S. 2006. Fetal testosterone and sex differences. Early Hum Dev 82:755–760.

Kohavi R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2. IJCAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. p 1137–1143. Available from: http://dl.acm.org/citation.cfm?id=1643031.1643047

Kullback S, Leibler RA. 1951. On Information and Sufficiency. The Annals of Mathematical Statistics 22:79–86.

Kuzawa CW, Bragg JM. 2012. Plasticity in Human Life History Strategy: Implications for Contemporary Human Variation and the Evolution of Genus Homo. Current Anthropology 53:S369–S382.

Kvålseth T. 1985. Cautionary note about R2. The American Statistician 39:279–285.

Lampl M, Johnston FE. 1996. Problems in the aging of skeletal juveniles: perspectives from maturation assessments of living children. Am J Phys Anthropol 101:345–355.

Largo R, Walli R, Duc G, Fanconi A, Prader A. 1980. Evaluation of perinatal growth. Presentation of combined intra- and extrauterine growth standards for weight, length and head circumference. Helvetica Paediatrica Acta 35:419–436.

Larsen CS. 1995. Biological Changes in Human Populations with Agriculture. Annual Review of Anthropology 24:185–213.

Lei J, Wasserman L. 2012. Distribution Free Prediction Bands.

Leisch F, Hornik K, Ripley BD. 2011. mda: Mixture and flexible discriminant analysis.

Levin B. 1996. On the Holm, Simes, and Hochberg multiple test procedures. Am J Public Health 86:628–629.

Lewis A, Garn S. 1960. The Relationship between Tooth Formation and Other Maturational Factors. The Angle Orthodontist 30:70–77.

Lewis C, Lavy C, Harrison W. 2002. Delay in skeletal maturity in Malawian children. Journal of Bone and Joint Surgery 84:732–734.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. 2008. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science 319:1100–1104.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. R News 2:18–22.

Lodox Systems Pty (Ltd), Sandton, South Africa. Available from: http://www.lodox.com

López-Costas O, Rissech C, Trancho G, Turbón D. 2012. Postnatal ontogenesis of the tibia. Implications for age and sex estimation. Forensic Sci Int 214:207.e1–11.

Loth S, Henneberg M. 2001. Sexually dimorphic mandibular morphology in the first few years of life. Am J Phys Anthropol 115:179–186.

Lovejoy CO, Russell KF, Harrison ML. 1990. Long bone growth velocity in the Libben population. American Journal of Human Biology 2:533–541.

Malina R, Bouchard C, Bar-Or O. 2004. Body Composition. In: Growth, Maturation, and Physical Activity. 2nd Edition. Champaign, IL: Human Kinetics.

Malina R, Johnston F. 1967. Relations between bone, muscle, and fat widths in the upper arms and claves of boys and girls studied cross-sectionally at ages 6 to 16 years. Human Biology 39:211–223.

Malina R. 2004. Secular trends in growth, maturation, and physical performance: a review. Przeglad Antropologiczny-Anthropological Review 67:3–31.

Malina RM. 1974. Adolescent Changes in Size, Build, Composition and Performance. Human Biology 46:117–131.

Malovini A, Barbarini N, Bellazzi R, De Michelis F. 2012. Hierarchical Naive Bayes for genetic association studies. BMC Bioinformatics 13:S6.

Mann D, Gould K, Collins D. 1989. Blockade of neonatal activation of the pituitary-testicular axis: Effect on peripubertal luteinizing hormone and testosterone secretion and on testicular development in male monkeys. Journal of Clinical Endocrinology and Metabolism 68:600–607.

Maree GJ, Irving BJ, Hering ER. 2007. Paediatric dose measurement in a full-body digital radiography unit. Pediatr Radiol 37:990–997.

Maresh M, Deming J. 1939. The growth of long bones in 80 infants: roentgenograms versus anthropometry. Child Development 10:91–106.

Maresh M. 1943. Growth of major long bones in healthy children. American Journal of Diseases of Children 66.

Maresh M. 1955. Linear growth of long bones of extremities from infancy through adolesence. American Journal of Diseases of Children 89:725–742.

Maresh M. 1966. Changes in tissue widths during growth: Roentgenographic measurements of bone, muscle, and fat widths from infancy through adolescence. Am J Dis Child 111:142–155.

Maresh M. 1970. Measurements from Roentgenograms. In: McCammon R, editor. Human Growth and Development. Springfield, IL: CC. Thomas. p 157–200.

Maresh MM. 1961. Bone, Muscle and Fat Measurements. Longitudinal Measurements of the Bone, Muscle and Fat Widths from Roentgenograms of the Extremities During the First Six Years of Life. Pediatrics 28:971–984.

Markov Z, Russell I. 2007. Probabilistic reasoning with naïve bayes and Bayesian networks. Available from: http://www.cs.ccsu.edu/~markov/ccsu_courses/probabilisticreasoning.pdf

Mateos P. 2007. A review of name-based ethnicity classification methods and their potential in population studies. Population, Space and Place 13:243–263.

McFadden D. 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, editor. Frontiers in Econometrics. New York: Academic Press.

McKern T, Stewart T. 1957. Skeletal age changes in young American males: analysed from the standpoint of age identification. Natick, MA: Headquarters Quartermaster Research and Development Command.

Meadows Jantz L, Jantz RL. 1999. Secular change in long bone length and proportion in the United States,1800-1970. American Journal of Physical Anthropology 110:57–67.

Menard S. 2000. Coefficients of Determination for Multiple Logistic Regression Analysis. The American Statistician 54:17–24.

Merchant V, Ubelaker D. 1977. Skeletal growth of the protohistoric Arikara. Am J Phys Anthropol 46:61–72.

Meredith H. 1976. Findings from Asia, Australia, Europe, and North America on Secular Change in Mean Height of Children, Youths, and Young Adults. Am J Phys Anthropol 44:315–326.

Michie D, Spiegelhalter DJ, Taylor CC, Campbell J eds. 1994. Machine learning, neural and statistical classification. Upper Saddle River, NJ, USA: Ellis Horwood.

Milborrow S. 2013. earth: Multivariate Adaptive Regression Spline Models. Available from: http://CRAN.R-project.org/package=earth

Molleson T, Cruse K, Mays S. 1998. Some sexually dimorphic features of the human juvenile skull and their values in sex determination in immature remains. J Archaeol Sci 25:719–728.

Molteno CD, Hollingshead J, Moodie AD, Bradshaw D, Willoughby W, Bowie MD, Smallman LA. 1991. Growth of preschool coloured children in Cape Town. S Afr Med J 79:670–676.

Moore MK. 2013. Sex Estimation and Assessment. In: DiGangi EA, Moore MK, editors. Research Methods in Human Skeletal Biology. 1st ed. Oxford: Academic Press. p 91–112.

Moore-Jansen PH, Ousley SD, Jantz RL. 1994. Data Collection Procedures for Forensic Skeletal Material. Knoxville: Department of Anthropology, The University of Tennessee.

Moorrees C, Fanning E, Hunt E. 1963. Age Variation of Formation Stages for Ten Permanent Teeth. J Dent Res 42:1490–1502.

Morris AG. 1997. The Griqua and the Khoikhoi: biology, ethnicity, and the construction of identity. Kronos:106–118.

Moss M. 1973. A functional cranial analysis of primate craniofacial growth. In: Zingeser M, editor. Fourth International Congress of Primatologists. Vol. 3. Basel: Karger. p 191–208.

Motulsky H. 1995. Intuitive Biostatistics. First. New York: Oxford University Press.

Mueller W, Martorell R. 1988. Reliability and Accuracy of Measurement. In: Lohman T, Roche A, Martorell R, editors. Anthropometric Standardisation Reference Manual. Champaign, IL: Human Kinetics Books. p 83–86.

Muñoz J, Felicísimo ÁM. 2004. Comparison of statistical methods commonly used in predictive modelling. J Veg Sci 15:285–292.

Nadler G. 1998. Earlier dental maturation: fact or fiction? Angle Orthodontist 68:535–538.

Nawrocki stephen P. 2010. The Nature anf Sources of Error in the Estimation of Age at Death from the Skeleton. In: Krista E. Latham, Finnegan M, editors. Age Estimation of the Human Skeleton. Springfield: Charles C. Thomas.

Neter J, Wasserman W, Kutner MH. 1985. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs. 2nd ed. Homewood, Illinois: Richard D. Irwin, Inc.

Nieves J, Formica C, Ruffing J, Zion M, Garrett P, Lindsay R, Cosman F. 2005. Males have larger skeletal size and bone mass than females, despite comparable body size. J Bone Mineral Res 20:529–535.

Norris SA, Roeser RW, Richter LM, Lewin N, Ginsburg C, Fleetwood SA, Taole E, van der Wolf K. 2008. South African-ness Among Adolescents: The Emergence of a Collective Identity within the Birth to Twenty Cohort Study. J Early Adolesc 28:51–69.

Olson CL. 1976. On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin 83:579–586.

Ontell F, Ivanovic M, Ablin D, Barlow TW. 1996. Bone Age in Children of Diverse Ethnicity. American Journal of Roentgenology 167:1395–1398.

Ousley S, Jantz R. 2012. Fordisc 3 and Statistical Mehods for Estimating Sex and Ancestry. In: Dirkmaat D, editor. A Companian to Forensic Anthropology. Blackwell Publishing Ltd. p 311–329.

Ousley S. 2013. Final Report: A Radiographic Database for Estimating Biological Parameters in Modern Subadults. Erie, PA: Mercyhurst University.

Patterson N, Petersen DC, van der Ross RE, Sudoyo H, Glashoff RH, Marzuki S, Reich D, Hayes VM. 2010. Genetic structure of a unique admixed population: implications for medical research. Hum Mol Genet 19:411–419.

Pearson K. 1915. On the Problem of Sexing Osteometric Material. Biometrika 10:479–487.

Perini T, de Oliveira G, dos Santos Ornellas J, de Oliveira F. 2005. Technical Error of Measurement in Anthropometry. Rev Bras Med Esporte 11.

Petersen DC, Libiger O, Tindall EA, Hardie R-A, Hannick LI, Glashoff RH, Mukerji M, Fernandez P, Haacke W, Schork NJ, Hayes VM, Indian Genome Variation Consortium. 2013. Complex Patterns of Genomic Admixture within Southern Africa. PLoS Genet 9:e1003309.

Pfau R, Sciulli P. 1994. A Method for Establishing the Age of Subadults. Journal of Forensic Sciences 39:165–176.

Poissonnet CM, Burdi AR, Garn SM. 1984. The chronology of adipose tissue appearance and distribution in the human fetus. Early Hum Dev 10:1–11.

Prader A, Largo RH, Molinari L, Issler C. 1989. Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. Helv Paediatr Acta Suppl 52:1–125.

Preston C, Chertkow S. 1986. Aspects of growth in a group of South African negro children. In: Singer R, Lundy J, editors. Variation, Culture and Evolution in African Populations. Johannesburg: Witwatersrand University Press.

Quigley CA. 2002. The Postnatal Gonadotropin and Sex Steroid Surge—Insights from the Androgen Insensitivity Syndrome. JCEM 87:24–28.

Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden PD, Hoal EG, Behar DM. 2010. Strong Maternal Khoisan Contribution to the South African Coloured Population: A Case of Gender-Biased Admixture. Am J Hum Genet 86:611–620.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org/

Racine J. 2012. A Primer on Regression Splines.

Reynolds E. 1945. The Bony Pelvic Girdle in Early Infancy. Am J Phys Anthropol 3:321–354.

Reynolds E. 1947. The bony pelvis in prepubertal childhood. Am J Phys Anthropol 3:165–200.

Ribot I. 2003. Craniometrical analysis of Central and East Africans in relation to history: a case study based on unique collections of known ethnic affiliation. Anthropologica et Praehistorica 114:25–50.

Ribot I. 2004. Differentiation of Modern Sub-Saharan African Populations: Craniometric interpretations in relation to geography and history. Bulletins et Mémoires de la Société d'Anthropologie de Paris 16:143–165.

Riggs BL, Khosla S, Melton LJ 3rd. 2002. Sex steroids and the construction and conservation of the adult skeleton. Endocr Rev 23:279–302.

Rissech C, López-Costas O, Turbón D. 2013. Humeral development from neonatal period to skeletal maturity—application in age and sex assessment. International Journal of Legal Medicine 127:201–212.

Rissech C, Schaefer M, Malgosa A. 2008. Development of the femur--Implications for age and sex determination. Forensic Science International 180:1–9.

Roever C, Raabe N, Leubke K, Ligges U, Szepannek G, Zentgraf M. 2013. klaR: Classification and visualization.

Rogers T. 2009. Sex determination of adolescent skeletons using the distal humerus. American Journal of Physical Anthropology 140:143–148.

Rogol AD, Clark PA, Roemmich JN. 2000. Growth and pubertal development in children and adolescents: effects of diet and physical activity. Am J Clin Nutr 72:521s–528s.

Rothwell P. 2000. Analysis of Agreement Between Measurements of Continuous Variables: General Principles and Lessons from Studies of Imaging of Carotid Stenosis. J Neurol 415:825–834.

Rousseeuw PJ, Hubert M. 2011. Robust statistics for outlier detection. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1:73–79.

Ruff C. 2003. 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 45:317–349.

Sartorius BK, Sartorius K, Chirwa TF, Fonn S. 2011. Infant mortality in South Africa - distribution, associations and policy implications, 2007: an ecological spatial analysis. Int J Health Geogr 10:61.

Saunders S. 2008. Juvenile skeletons and growth-related studies. In: Katzenberg M, Saunders S, editors. Biological Anthropology of the Human Skeleton. 2nd ed. Hoboken: John WIley & Sons. p 117–147.

Saunders SR, Hoppa RD. 1993. Growth deficit in survivors and non-survivors: Biological mortality bias in subadult skeletal samples. American Journal of Physical Anthropology 36:127–151.

Schaefer MC. 2008. A summary of epiphyseal union timings in Bosnian males. International Journal of Osteoarchaeology 18:536–545.

Schell L, Knutson K, Bailey S. 2012. Environmental Effects on Growth. In: Cameron N, Bogin B, editors. Human Growth and Development. 2nd ed. New York: Elsevier. p 245–286.

Scheuer J, Musgrave J, Evans S. 1980. The estimation of late fetal and perinatal age form limb bone length by linear and logarithmic regression. Ann Hum Biol 7:257–265.

Scheuer L, Black S. 2000. Developmental Juvenile Osteology. New York: Elsevier Academic Press.

Scheuer L. 2002a. Application of osteology to forensic medicine. Clin Anat 15:297–312.

Scheuer L. 2002b. A Blind Test of Mandibular Morphology for Sexing Mandibles in the First Few Years of Life. Am J Phys Anthropol 119:189–191.

Schlebusch C. 2010. Issues raised by use of ethnic-group names in genome study. Nature 464:487–487.

Schmeling A, Reisinger W, Geserick G, Olze A. 2006. Age estimation of unaccompanied minors Part I. General Considerations. Forensic Science International 159S:S61–S64.

Schmeling A, Reisinger W, Loreck D, Vendure K, Markus W, Geserick G. 2000. Effects of ethnicity on skeletal maturation: consequences for forensic age estimations. International Journal of Legal Medicine 113:253–258.

Schoenau E, Fricke O. 2008. Mechanical influences on bone development in children. Eur J Endocrinol 159 Suppl 1:S27–31.

Schoenau E, Neu C, Beck B, Manz F, Rauch F. 2002. Bone Mineral Content per Muscle Cross-Sectional Area as an Index of the Functional Muscle-Bone Unit. J Bone Mineral Res 17:1095–1101.

Schönau E. 1998. The development of the skeletal system in children and the influence of muscular strength. Horm Res 49:27–31.

Schroeder C, Schmidtke S, Bidez M. 1997. Measuring the human pelvis: a comparison of direct and radiographic techniques using a modern United States-based sample. Am J Phys Anthropol 103:471–9.

Schultz AH. 1962. Metric age changes and sex differences in primate skulls. Zeitschrift für Morphologie und Anthropologie 52:239–255.

Schutkowski H. 1987. Sex determination of fetal and neonatal skeletons by means of discriminant analysis. Int J Anthropol 2:347–352.

Schutkowski H. 1993. Sex determination of infant and juvenile skeletons: I. morphognostic features. Am J Phys Anthropol 90:199–205.

Seeman E, Hopper JL, Young NR, Formica C, Goss P, Tsalamandris C. 1996. Do genetic factors explain associations between muscle strength, lean mass, and bone density? A twin study. Am J Physiol 270:E320–327.

Sekhampu TJ. 2004. An in-depth micro-economic analysis of the poor in the Bophelong community with special reference to the activities that they use to sustain themselves.

Sekhampu TJ. 2013. Determinants of Poverty in a South African Township. J Soc Sci 34:145–153.

Sekulic S, Kowalski BR. 1992. MARS: A tutorial. Journal of Chemometrics 6:199–216.

Shapiro D, Richtsmeier J. 1997. Brief communication: A sample of pediatric skulls available for study. Am J Phys Anthropol 103:415–416.

Sicotte M, Ledoux M, Zunzunegui M-V, Aboubacrine SA, Nguyen V-K. 2010. Reliability of anthropometric measures in a longitudinal cohort of patients initiating ART in West Africa. BMC Medical Research Methodology 10:102.

Singer R, Kimura K. 1981. Body height, weight, and skeletal maturation in Hottentot (Khoikhoi) children. Am J Phys Anthropol 54:401–413.

Slabbert T. 2003. Bophelong: A socio-economic and environmental analysis. Vanderbijlpark: Vaal Research Group (VRG).

Smail P, Reyes F, Winter J, Faiman C. 1981. The fetal hormonal environment and its effect on the morphogenesis of the genital system. In: Kogan S, Hafez E, editors. Pediatric Andrology. Boston: Martinus Nijhoff. p 9–19.

Smith S, Buschang P. 2004. Variation in Longitudinal Diaphyseal Long Bone Growth in Children Three to Ten Years of Age. American Journal of Human Biology 16:648–657.

Smith S. 2007. Stature Estimation of 3-10 year-old Children from Long Bone Lengths. J Forensic Sci 52:538–46.

Smith SL, Buschang PH. 2005. Longitudinal models of long bone growth during adolescence. Am J Hum Biol 17:731–745.

Smithgall E, Johnston F, Malina R, Galbraith M. 1966. Developmental Changes in Compact Bone Relationships in the Second Metacarpal. Human Biology 38:141–151.

Spradley M, Jantz R. 2011. Sex estimation in forensic anthropology: Skull versus postcranial elements. Journal of Forensic Sciences 56:289–296.

Spradley MK, Jantz RL, Robinson A, Peccerelli F. 2008. Demographic change and forensic identification: Problems in metric identification of Hispanic skeletons. Journal of Forensic Sciences [Internet] 53. Available from: internal-pdf://jfo_614-4233172992/jfo_614.pdf

Statistics South Africa. 1999. The people of South Africa: population census 1996. Pretoria.

Statistics South Africa. 2012. Census 2011: Statistical Release.

Steckel RH. 1994. Heights and health in the United States, 1710-1950. In: Komlos J, editor. Stature, living standards and economic development: essays in anthropometric history. Chicago: University of Chicago Press. p 153–170.

Steckel RH. 2008. Heights and Human Welfare: Recent Developments and New Directions. National Bureau of Economic Research. Available from: http://www.nber.org/papers/w14536

Stewart T. 1954. Evaluation of Evidence From the Skeleton. In: Gradwohl R, editor. Legal Medicine. 1st ed. St. Louis: C. V. Mosby Co. p 407–450.

Stewart T. 1976. Identification of Skeletal Structures. In: Camps FE, editor. Gradwhol's Legal Medicine. 3rd ed. Bristol. p 109–135.

Steyn M, İşcan M. 1997. Sex determination from the femur and tibia in South African Whites. Forensic Science International 90:111–119.

Steyn M, Smith J. 2007. Interpretation of ante-mortem stature estimates in South Africans. Forensic Sci Int 171:97–102.

Stine RA. 1985. Bootstrap Prediction Intervals for Regression. Journal of the American Statistical Association 80:1026.

Stinson S. 1985. Sex differences in environmental sensitivity during growth and development. American Journal of Physical Anthropology 28:123–147.

Stull K, L'Abbe E, Ousley S. 2013a. A comparison of the long bone lengths of North American and South African children as a means for appropriate age-at-death techniques. In: Washington, DC.

Stull KE, Godde K. 2013. Sex estimation of infants between birth and one year through discriminant analysis of the humerus and femur. J Forensic Sci 58:13–20.

Stull KE, L'abbé EN, Steiner S. 2013b. Measuring distortion of skeletal elements in Lodox Statscan-generated images. Clin Anat 26:780–786.

Sun S, Schubert C, Chumlea W, Roche A, Kulin H, Lee P, Himes J, Ryan A. 2002. National Estimates of the Timing of Sexual Maturation and Racial Differences Among US Children. Pediatrics 110:911–928.

Sundick R. 1972. Human skeletal growth and dental development as observed in the Indian Knoll population.

Sundick R. 1977. Age and sex determination of subadult skeletons. J Forensic Sci 22:141–144.

Sundick R. 1978. Human skeletal growth and age determination. HOMO: Journal of Comparative Human Biology 29:228–249.

Sutter R. 2003. Nonmetric Subadult Skeletal Sexing Traits: l. A Blind Test of the Accuracy of Eight Previously Proposed Methods Using Prehistoric Known-Sex Mummies from Northern Chile. J Forensic Sci 48:927–935.

Swart L, Laher H, Seedat M, Gantchev G. 2012. Urban and rural differences in child injury deaths in South Africa: A one-year review. African Safety Promotion Journal 10:28–40.

Tabachnick BG, Fidell LS. 2007. Using multivariate statistics. 5th ed. Boston: Pearson Education.

Tanner J. 1986. The use and abuse of growth standards. In: Falkner F, Tanner J, editors. Human Growth: A Comprehensive Treatise. Vol. 3: Methodology. 2nd ed. New York: Plenum Press. p 95–109.

Tanner J. 1989. Foetus into Man: Physical Growth from Conception to Maturity. Cambridge: Howard University Press.

Tanner JM, Hughes PC, Whitehouse RH. 1981. Radiographically determined widths of bone muscle and fat in the upper arm and calf from age 3-18 years. Ann Hum Biol 8:495–517.

Thomas P, Peabody J, Tunier V, Clark R. 2000. A new look at intrauterine growth and the impact of race, altitude, and gender. Pediatrics 106.

Tise ML, Spradley MK, Anderson BE. 2013. Postcranial sex estimation of individuals considered Hispanic. J Forensic Sci 58 Suppl 1:S9–14.

Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo J-M, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM. 2009. The genetic structure and history of Africans and African Americans. Science 324:1035–1044.

Tishkoff SA, Williams SM. 2002. Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet 3:611–621.

Treiman DJ, McKeever M, Fodor E. 1996. Racial differences in occupational status and income in South Africa, 1980 and 1991. Demography 33:111–132.

Treiman DJ. 2007. The Legacy of Apartheid: Racial Inequalities in the New South Africa. In: Heath AF, Cheung SY, editors. Unequal chances: ethnic minorities in Western labour markets. Oxford: Published for the British Academy by Oxford University Press.

Ubelaker DH. 1999. Human Skeletal Remains: Excavation, Analysis, Interpretation. Washington: Taraxacum.

Ulijaszek S, Kerr D. 1999. Anthropometric Measurement Error and the Assessment of Nutritional Status. Br J Nutr 82:165–177.

Ulijaszek S. 2001. Secular trends in growth: the narrowing of ethnic differences in stature. Nutrition Bulletin 26:43–51.

United Nations. 2013. World Population Prospects: The 2012 Revision. Department of Economic and Social Affairs.

Utermohle CJ, Zegura SL, Heathcote GM. 1983. Multiple observers, humidity, and choice of precision statistics: factors influencing craniometric data quality. Am J Phys Anthropol 61:85–95.

De Veaux RDD, Ungar LH. 1994. Multicollinearity: A tale of two nonparametric regressions. In: Oldford RW, editor. Selecting Models from Data. Lecture Notes in Statistics. Springer New York. p 393–402. Available from: http://link.springer.com/chapter/10.1007/978-1-4612-2660-4_40

Veldhuis J, Roemmich J, Richmond E, Rogol A, Lovejoy J, Sheffield-Moore M, Mauras N, Bowers C. 2005. Endocrine control of body composition in infancy, childhood, and puberty. Endocr Rev 26:114–146.

Venables W, Ripley B. 2002. Mordern Applied Statistics with S. 4th ed. New York: Springer.

Vicente Rodríguez G. 2006. How does exercise affect bone development during growth? SPORTS MEDICINE REVIEWS OF APPLIED MEDICINE AND SCIENCE IN SPORT AND EXERCISE 36:561.

Viljoen D, Sekhampu TJ. 2013. The Impact of Apartheidon Urban Poverty in South Africa: What we can Learn from History. Mediterranean Journal of Social Sciences 4:729.

De Vito C, Saunders S. 1990. A discriminant function analysis of deciduous teeth to determine sex. Journal of Forensic Sciences 35:845–848.

Vlak D, Roksandic M, Schillaci M. 2008. Greater sciatic notch as a sex indicator in juveniles. Am J Phys Anthropol 137:309–315.

Walker A, Walker B. 1977. Studies on increases in growth rate of South African Black schoolchildren and their significance to Health. S Afr Med J 51:707–712.

Weaver D. 1980a. Sex differences in the ilia of a known sex and age sample of fetal and infant skeletons. Am J Phys Anthropol 52:191–195.

Weaver D. 1988. Forensic Aspects of Fetal and Neonatal Skeletons. In: Reichs K, editor. Forensic Osteology: advances in the identification of human remains. 2nd ed. Springfield: Charles C. Thomas. p 187–203.

Weaver DS. 1980b. Sex differences in the Ilia of a known sex and age sample of fetal and infant Skeletons. American Journal of Physical Anthropology 52:191–195.

Weinfurt KP. 1995. Multivariate analysis of variance. In: Grimm LG, Yarnold PR, editors. Reading and understanding multivariate statistics. Washington, DC: American Psychological Association. p 245–276.

Wells JCK. 2007. Sexual dimorphism of body composition. Best Practice & Research Clinical Endocrinology & Metabolism 21:415–430.

Wells JCK. 2012. Sexual dimorphism in body composition across human populations: Associations with climate and proxies for short- and long-term energy supply. American Journal of Human Biology 24:411–419.

Western Cape Provincial Treasury. 2011. Regional Development Profile City of Cape Town. Cape Town, South Africa.

Whiley SP, Mantokoudis G, Ott D, Zimmerman H, Exadaktylos AK. 2012. A Review of Full-Body Radiography in Nontraumatic Emergency Medicine. Emergency Medicine

International [Internet] 2012. Available from: http://www.hindawi.com/journals/emi/2012/108129/abs/

WHO Multicentre Growth Reference Study Group. 2006. Reliability of anthropometric measurements in the WHO Multicentre Growth Reference Study. Acta Paediatr Suppl 450:38–46.

WHO. 2003. The world health report 2003 - shaping the future. World Health Organization. Available from: http://www.who.int/whr/2003/en/

Wickham H, Stryjewski L. 2011. 40 years of boxplots.

Wilson L, Cardoso H, Humphrey L. 2011. On the reliability of a geometric morphometric approach to sex determination: A blind test of six criteria of the juvenile ilium. Forensic Science International 206:35–42.

Wilson LA, MacLeod N, Humphrey LT. 2008. Morphometric Criteria for Sexing Juvenile Human Skeletons Using the Ilium. Journal of Forensic Sciences 53:269–278.

Wood J, Milner G, Harpending H, Weiss K. 1992. The Osteological Paradox: Problems of Inferring Prehistoric Health from Skeletal Samples. Current Anthropology 33:343–370.

Wood SN. 2006. Generalized Additive Models: An Introduction with R. CRC Press.

Wright D, London K. 2009. Modern Regression Techniques Using R: A Practical Guide for Students and Researchers. London: SAGE Publications, Inc.

Zhang H, Su J. 2004. Naive Bayesian Classifiers for Ranking. In: Boulicaut J-F, Esposito F, Giannotti F, Pedreschi D, editors. Machine Learning: ECML 2004. Lecture Notes in Computer Science. Springer Berlin Heidelberg. p 501–512. Available from: http://link.springer.com/chapter/10.1007/978-3-540-30115-8_46

Zhang H. 2004. The Optimality of Naive Bayes. In: Vol. 1. . p 3.

Zhang H. 2005. Exploring conditions for the optimality of naive Bayes. International Journal of Pattern Recognition and Artificial Intelligence 19:183–198.

APPENDIX I - MEASUREMENT DEFINITIONS

Humerus diaphyseal length (HDL) – The maximum distance between the most proximal edge of the diaphysis to the most distal edge of the diaphysis (modified from Fazekas and Kósa, 1978).
Comment: The most distal portion is generally the medial portion.

Humerus proximal breadth (HPB) – The distance between the most medial and lateral edges of the proximal diaphysis, perpendicular to the long axis of the bone, when the element is in anatomical position.
Comment: This is not a maximum breadth.

Humerus distal breadth (HDB) – The distance between the most medial and lateral points on the distal diaphysis, perpendicular to the long axis of the bone, when the element is in anatomical position (Fazekas and Kósa, 1978).

Humerus midshaft breadth (HMSB) – The distance between the most medial and lateral edges at midshaft, perpendicular to the long axis of the bone, when the bone is in anatomical position (Fazekas and Kósa, 1978).
Comment: Determine midshaft when obtaining diaphyseal length. Note, this is not a minimum or maximum.

Ulna diaphyseal length (UDL) – The maximum distance between the most proximal edge of the diaphysis to the most distal edge of the diaphysis (Fazekas and Kósa, 1978).

Ulna midshaft breadth (UMSB) – The distance between the most medial and lateral edges at midshaft, perpendicular to the long axis of the bone, when the bone is in anatomical position (Fazekas and Kósa, 1978).
Comment: Determine midshaft when obtaining diaphyseal length. Note, this is not a minimum or maximum.

Radius diaphyseal length (RDL) – The maximum distance between the most proximal edge of the diaphysis to the most distal edge of the diaphysis (Fazekas and Kósa, 1978).

Radius proximal breadth (RPB) – The distance between the most medial and lateral edges of the proximal diaphysis, perpendicular to the long axis of the bone, when the bone is in anatomical position (modified from Urcid, 1992).

Radius distal breadth (RDB) – The distance between the most medial and lateral edges of the distal diaphysis, perpendicular to the long axis of the bone, when the bone is in anatomical position.
Comment: The measurement is obtained from the anterior projections on the distal diaphysis.

Radius midshaft breadth (RMSB) – The distance between the most medial and lateral edges at midshaft, perpendicular to the long axis of the bone, when the bone is in anatomical position (modified from Fazekas and Kósa, 1978).
Comment: Determine midshaft when obtaining diaphyseal length. Note, this is not a minimum or maximum.

Femur diaphyseal length (FDL) – The maximum distance between the most proximal edge of the diaphysis to the most distal edge of the diaphysis (Fazekas and Kósa, 1978).
Comment: The most distal point is generally the medial projection on the metaphysis. The expression is slight in infants but becomes more pronounced as age increases.

Femur distal breadth (FDB) – The distance between the most medial and lateral edges of the distal diaphysis, perpendicular to the long axis of the bone, when the bone is in anatomical position (modified from Fazekas and Kósa, 1978).

Femur midshaft breadth (FMSB) – The distance between the most medial and lateral edges at midshaft, perpendicular to the long axis of the bone, when the bone is in anatomical position (Fazekas and Kósa, 1978).
Comment: *Determine midshaft when obtaining diaphyseal length. Note, this is not a minimum or maximum.*

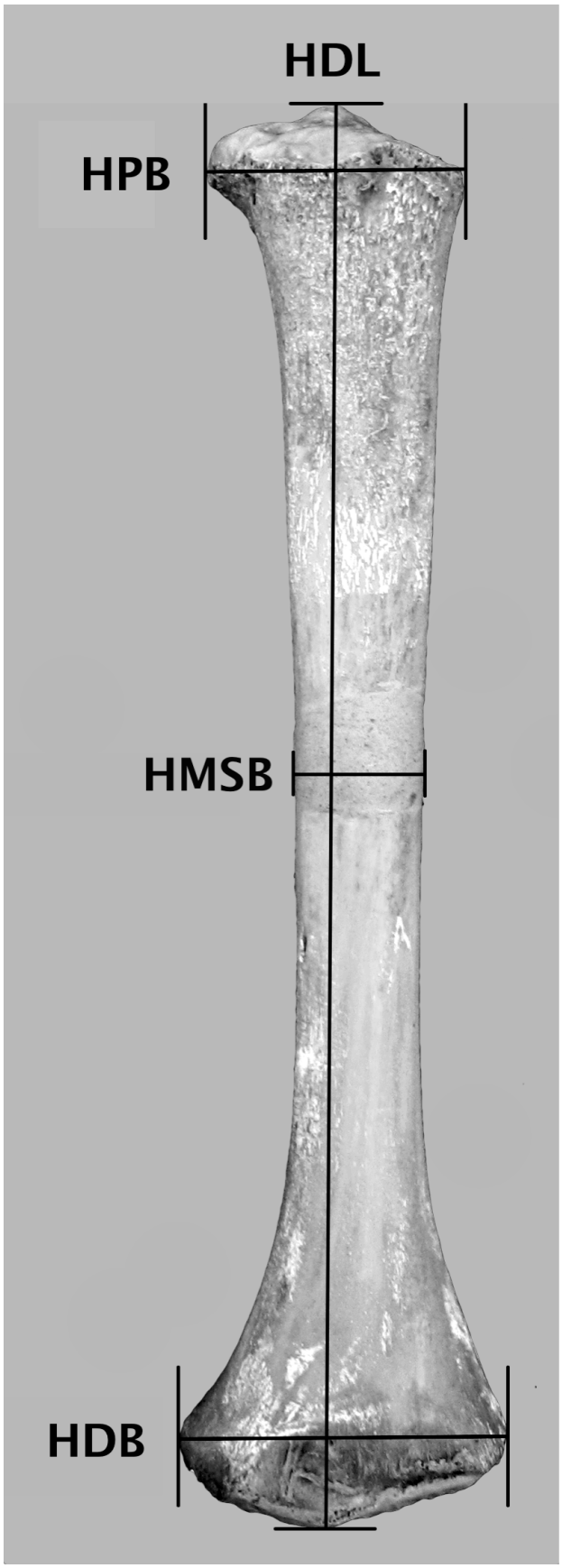
Tibia diaphyseal length (TDL) – The maximum distance between the most proximal edge of the diaphysis to the most distal edge of the diaphysis (Fazekas and Kósa, 1978).
Comment: Generally, the most proximal point is medial and the most distal point is lateral.

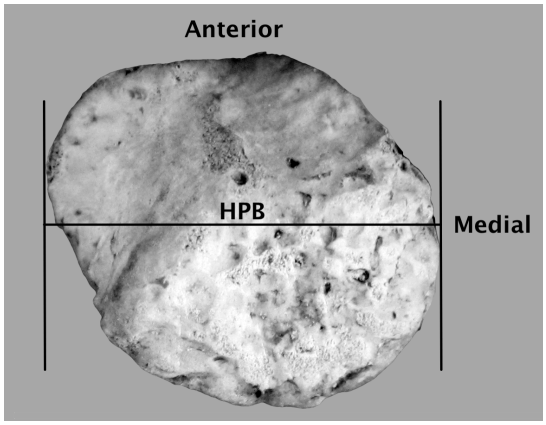
Tibia proximal breadth (TPB) – The distance between the most medial and lateral edges of the proximal diaphysis, perpendicular to the long axis of the bone, when the bone is in anatomical position (modified from Moore-Jansen et al., 1994).

Tibia distal breadth (TDB) – The distance between the most medial and lateral edges of the distal diaphysis, perpendicular to the long axis of the bone in anatomical position (modified from Moore-Jansen et al., 1994).
Comment: The lateral edge is the anterior projection of the fibular notch.

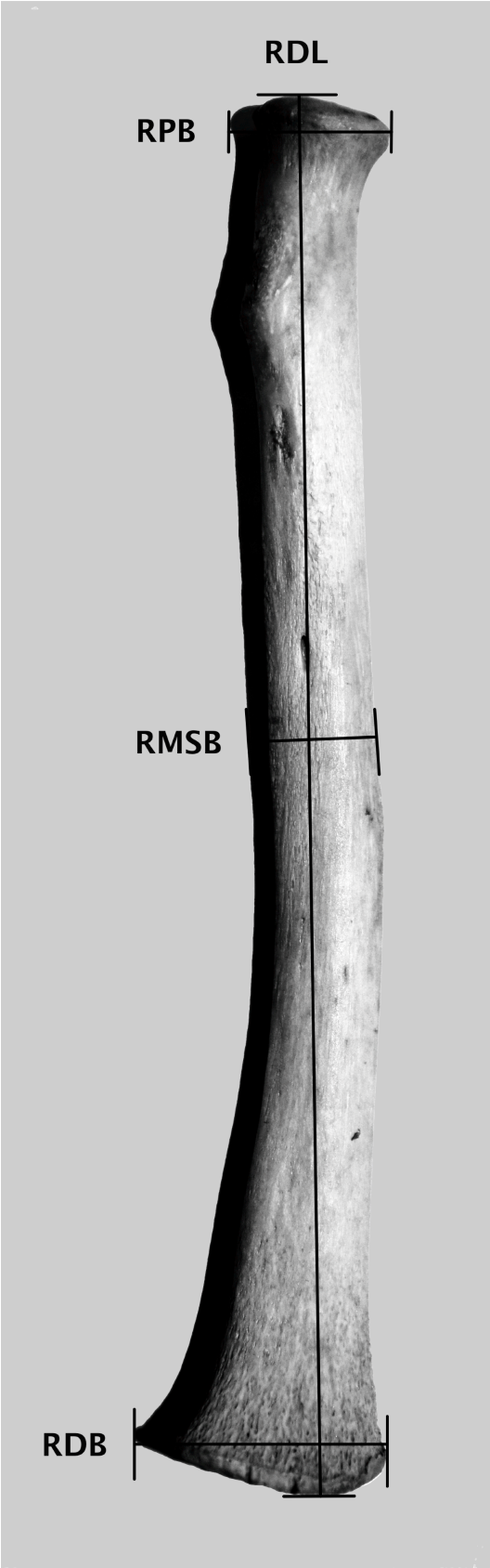
Tibia midshaft breadth (TMSB) – The distance between the most medial and lateral edges at midshaft, perpendicular to the long axis of the bone, when the bone is in anatomical position (Fazekas and Kósa, 1978).
Comment: Determine midshaft when obtaining diaphyseal length.

Fibula diaphyseal length (FBDL) – The maximum distance between the most proximal edge of the diaphysis to the most distal edge of the diaphysis (Fazekas and Kósa, 1978).

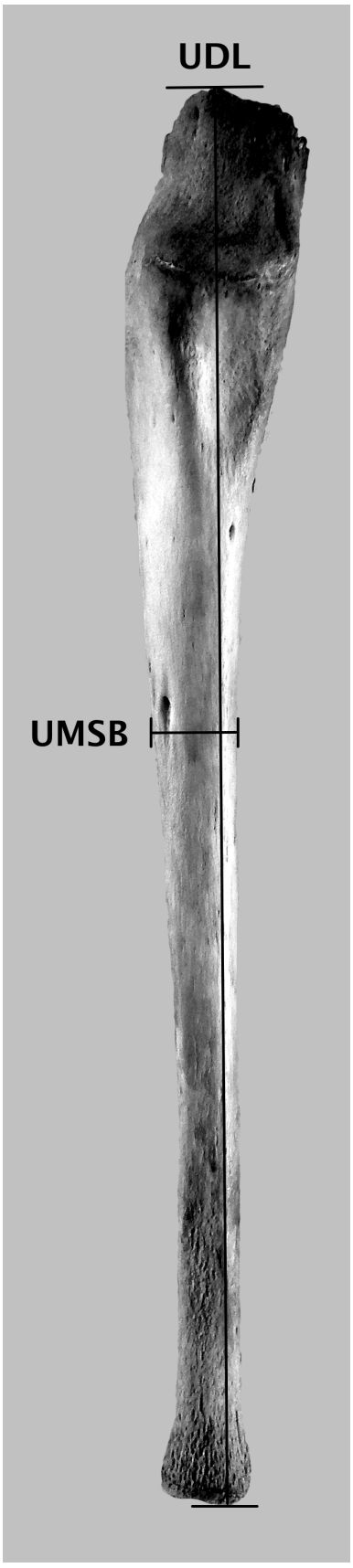


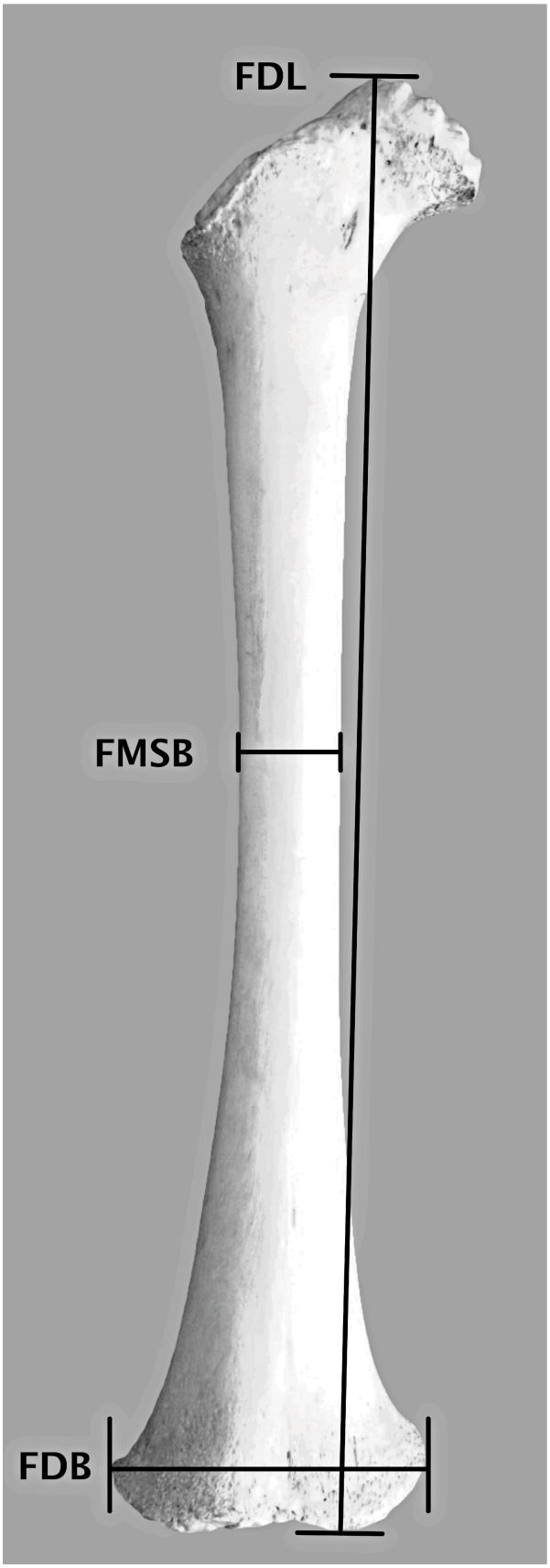


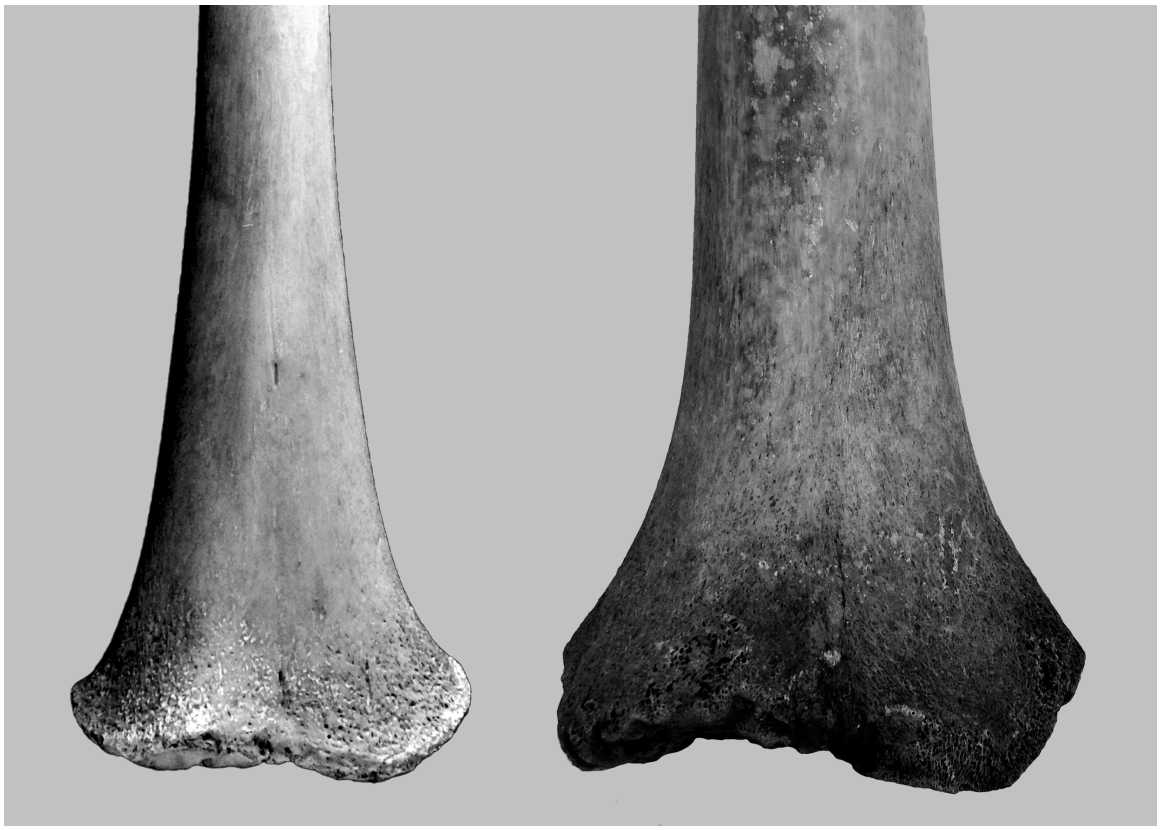
Superior view of the proximal humerus diaphysis to emphasize the measurement is not a maximum but rather a medial-lateral.



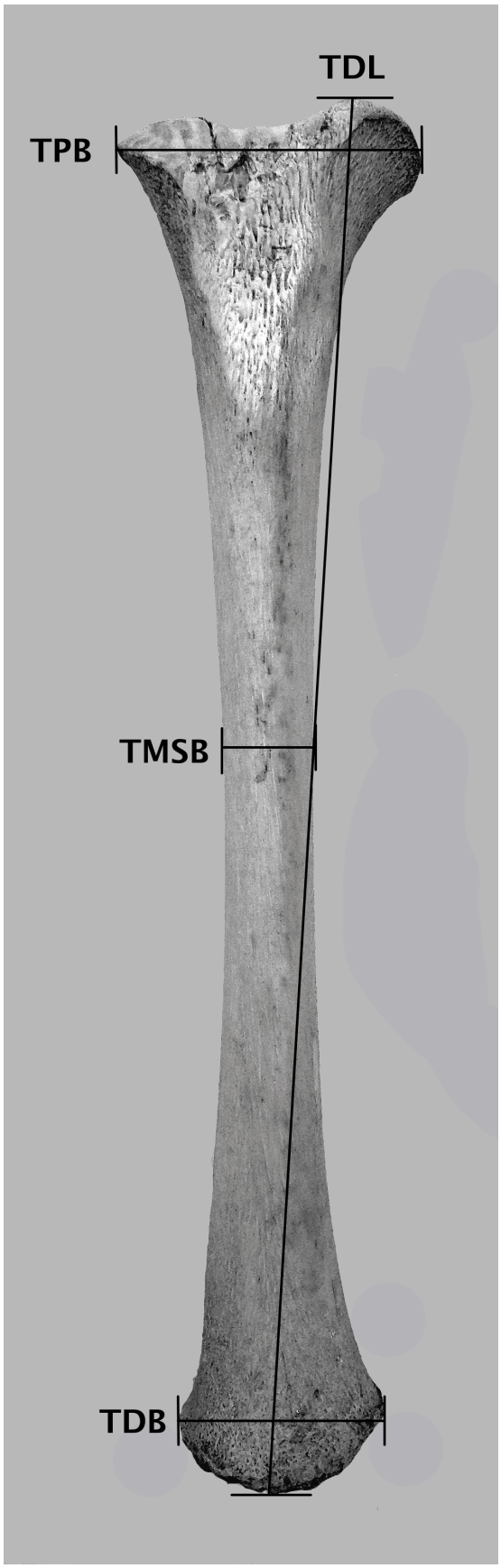
Additional image of the RDB to emphasize the measurement is obtained from the anterior projections on the distal diaphysis.

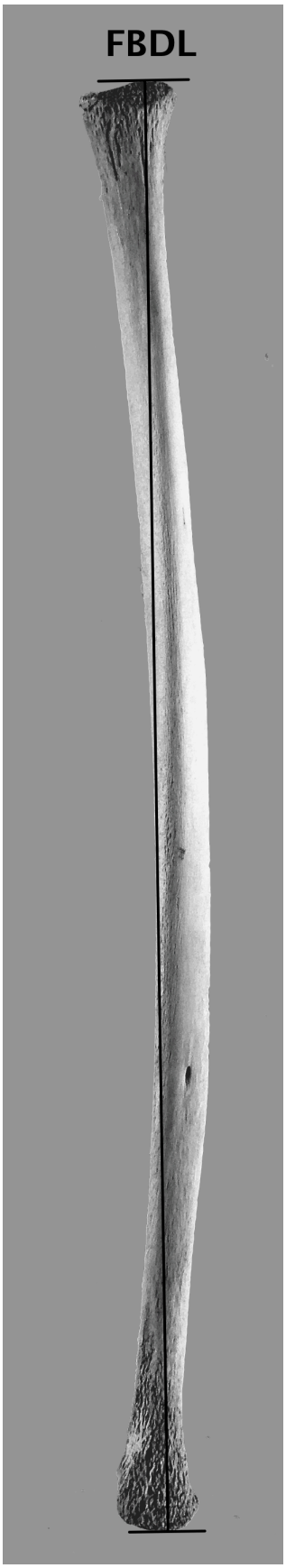






The medial projection of the distal diaphysis of the femur. The expression is slight in younger subadults (left) but is more pronounced in older subadults (right).





APPENDIX II – ANOVA

Table A2 – ANOVA results evaluating the statistical significance of age, sex, and the interaction of age and sex for each measurement. Bold indicates significance.

		Sex		**Age**		**Sex*Age**	
		F-value	*Pr(>|t|)*	*F*-value	*Pr(>|t|)*	*F*-value	*Pr(>|t|)*
Humerus	HMXL	50.90	**<0.001*** **	10271.30	**<0.001*** **	10.60	**0.001** **
	HPB	85.87	**<0.001*** **	2626.72	**<0.001*** **	1.24	0.27
	HDB	31.71	**<0.001*** **	2532.80	**<0.001*** **	2.53	0.11
	HMSB	47.30	**<0.001*** **	1333.70	**<0.001*** **	0.00	0.97
Radius	RMXL	13.78	**<0.001*** **	10136.03	**<0.001*** **	5.54	**<0.001*** **
	RPB	45.87	**<0.001*** **	2442.14	**<0.001*** **	0.03	0.87
	RDB	40.60	**<0.001*** **	2291.03	**<0.001*** **	0.96	0.33
	RMSB	7.28	**0.007** **	1188.02	**<0.001*** **	0.38	0.537
Ulna	UMXL	62.24	<0.001***	8059.45	<0.001***	4.22	0.04*
	UMSB	49.97	<0.001***	485.34	<0.001***	2.54	0.11
Femur	FMXL	41.90	<0.001***	12248.70	<0.001***	10.3	0.001***
	FDB	157.70	<0.001***	2398.56	<0.001***	0.01	0.94
	FMSB	52.58	<0.001***	2187.34	<0.001***	1.52	0.22
Tibia	TMXL	12.26	<0.001***	9685.54	<0.001***	7.49	0.006**
	TPB	69.50	<0.001***	2493.50	<0.001***	0.00	0.99
	TDB	27.93	<0.001***	2278.80	<0.001***	0.23	0.63
	TMSB	8.60	0.004**	1279.50	<0.001***	2.30	0.13
Fibula	FBMXL	22.51	<0.001***	8975.59	<0.001***	8.85	0.003**

Note: ***p < 0.05; **p<0.01; ***p<0.001**

APPENDIX III – HUMERUS: UNIVARIATE AGE ESTIMATION MODELS AND THE ASSOCIATED PREDICTION TABLES

Table A3.1 – MARS model for humerus diaphyseal length. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	2.114***
h(HDL – 191.21)	0.011***
h(191.21 – HDL)	-0.013***
h(HDL – 125.49)	-0.003***
h(HDL – 223.52)	-0.004***
Observations	1064
cv R2	0.95
Adjusted R2	0.95
Residual Std. Error	0.97
F Statistic	<0.0001***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A3.2 – The point estimate (fit) and 95% prediction interval (in years) for HDL.

HDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
60	0.10	0.02	0.28
61	0.11	0.02	0.29
62	0.12	0.03	0.31
63	0.13	0.03	0.33
64	0.14	0.03	0.35
65	0.15	0.04	0.37
66	0.16	0.04	0.39
67	0.17	0.05	0.41
68	0.18	0.05	0.43
69	0.19	0.06	0.45
70	0.20	0.06	0.47
71	0.22	0.07	0.49
72	0.23	0.08	0.52
73	0.25	0.08	0.54
74	0.26	0.09	0.57
75	0.28	0.10	0.59
76	0.29	0.11	0.62
77	0.31	0.12	0.65
78	0.33	0.13	0.68
79	0.35	0.13	0.71
80	0.36	0.14	0.74
81	0.38	0.16	0.77
82	0.40	0.17	0.80
83	0.43	0.18	0.83
84	0.45	0.19	0.87
85	0.47	0.20	0.90
86	0.49	0.22	0.94
87	0.52	0.23	0.98
88	0.54	0.25	1.01
89	0.57	0.26	1.05
90	0.59	0.28	1.09
91	0.62	0.29	1.13
92	0.65	0.31	1.17
93	0.68	0.33	1.22
94	0.71	0.35	1.26
95	0.74	0.36	1.30
96	0.77	0.38	1.35
97	0.8	0.4	1.4
98	0.83	0.42	1.44
99	0.87	0.45	1.49
100	0.90	0.47	1.54
101	0.94	0.49	1.59
102	0.97	0.52	1.65
103	1.01	0.54	1.70
104	1.05	0.57	1.75
105	1.09	0.59	1.81
106	1.13	0.62	1.87
107	1.17	0.65	1.92
108	1.21	0.68	1.98
109	1.26	0.70	2.04
110	1.30	0.73	2.10
111	1.34	0.76	2.16
112	1.39	0.80	2.22
113	1.44	0.83	2.29
114	1.48	0.86	2.35
115	1.53	0.89	2.42
116	1.58	0.93	2.48
117	1.63	0.96	2.55
118	1.68	1.00	2.62
119	1.74	1.04	2.69
120	1.79	1.08	2.76
121	1.84	1.12	2.84
122	1.90	1.16	2.91
123	1.96	1.20	2.99
124	2.02	1.24	3.06
125	2.07	1.28	3.14
126	2.13	1.32	3.21
127	2.18	1.36	3.28
128	2.23	1.40	3.35
129	2.28	1.43	3.42
130	2.33	1.47	3.48
131	2.38	1.51	3.55
132	2.43	1.55	3.61
133	2.49	1.58	3.68
134	2.54	1.62	3.75

HDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
135	2.59	1.66	3.81
136	2.64	1.70	3.88
137	2.69	1.74	3.95
138	2.75	1.78	4.02
139	2.80	1.82	4.09
140	2.86	1.86	4.16
141	2.91	1.90	4.23
142	2.97	1.95	4.30
143	3.03	1.99	4.38
144	3.09	2.03	4.45
145	3.15	2.08	4.53
146	3.20	2.12	4.60
147	3.27	2.17	4.68
148	3.33	2.21	4.76
149	3.39	2.26	4.84
150	3.45	2.31	4.92
151	3.51	2.36	5.00
152	3.58	2.41	5.08
153	3.64	2.46	5.16
154	3.71	2.51	5.24
155	3.77	2.56	5.32
156	3.84	2.61	5.41
157	3.91	2.66	5.49
158	3.98	2.72	5.58
159	4.05	2.77	5.67
160	4.12	2.83	5.76
161	4.19	2.88	5.85
162	4.26	2.94	5.94
163	4.33	2.99	6.03
164	4.41	3.05	6.12
165	4.48	3.11	6.21
166	4.56	3.17	6.30
167	4.63	3.23	6.40
168	4.71	3.29	6.49
169	4.79	3.35	6.59
170	4.87	3.41	6.69
171	4.95	3.48	6.79
172	5.03	3.54	6.89
173	5.11	3.60	6.99
174	5.19	3.67	7.09

HDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
175	5.27	3.74	7.19
176	5.36	3.80	7.29
177	5.44	3.87	7.40
178	5.53	3.94	7.50
179	5.62	4.01	7.61
180	5.70	4.08	7.71
181	5.79	4.15	7.82
182	5.88	4.22	7.93
183	5.97	4.29	8.04
184	6.06	4.36	8.15
185	6.15	4.43	8.26
186	6.24	4.51	8.37
187	6.33	4.58	8.47
188	6.42	4.65	8.58
189	6.51	4.72	8.69
190	6.60	4.80	8.80
191	6.69	4.87	8.91
192	6.77	4.94	9.01
193	6.85	5.00	9.11
194	6.93	5.07	9.21
195	7.01	5.13	9.30
196	7.09	5.20	9.40
197	7.17	5.26	9.50
198	7.25	5.33	9.60
199	7.34	5.40	9.70
200	7.42	5.46	9.80
201	7.50	5.53	9.90
202	7.59	5.60	10.00
203	7.67	5.67	10.10
204	7.76	5.74	10.20
205	7.85	5.81	10.31
206	7.93	5.88	10.41
207	8.02	5.95	10.52
208	8.11	6.03	10.62
209	8.20	6.10	10.73
210	8.29	6.17	10.84
211	8.38	6.25	10.94
212	8.47	6.32	11.05
213	8.56	6.40	11.16
214	8.65	6.47	11.27

HDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
215	8.74	6.55	11.38
216	8.84	6.63	11.49
217	8.93	6.70	11.60
218	9.03	6.78	11.72
219	9.12	6.86	11.83
220	9.22	6.94	11.95
221	9.31	7.02	12.06
222	9.41	7.10	12.18
223	9.51	7.18	12.29
224	9.60	7.26	12.40
225	9.67	7.32	12.49
226	9.74	7.37	12.57
227	9.80	7.42	12.64
228	9.85	7.47	12.70
229	9.90	7.51	12.76
230	9.95	7.55	12.81
231	10.00	7.59	12.87
232	10.05	7.63	12.93
233	10.09	7.67	12.98
234	10.14	7.71	13.04
235	10.19	7.74	13.09
236	10.23	7.78	13.15
237	10.28	7.82	13.20
238	10.32	7.86	13.26
239	10.37	7.90	13.31
240	10.42	7.94	13.37
241	10.46	7.98	13.42
242	10.51	8.02	13.48
243	10.56	8.06	13.53
244	10.61	8.10	13.59
245	10.65	8.13	13.65
246	10.70	8.17	13.70
247	10.75	8.21	13.76
248	10.80	8.25	13.81
249	10.84	8.29	13.87
250	10.89	8.33	13.93
251	10.94	8.38	13.98
252	10.99	8.42	14.04
253	11.04	8.46	14.10
254	11.09	8.50	14.16
255	11.14	8.54	14.21
256	11.18	8.58	14.27
257	11.23	8.62	14.33
258	11.28	8.66	14.39
259	11.33	8.70	14.44
260	11.38	8.74	14.50
261	11.43	8.79	14.56
262	11.48	8.83	14.62
263	11.53	8.87	14.68
264	11.58	8.91	14.74
265	11.63	8.95	14.80
266	11.68	9.00	14.86
267	11.73	9.04	14.92
268	11.78	9.08	14.97
269	11.83	9.12	15.03
270	11.89	9.17	15.09
271	11.94	9.21	15.15
272	11.99	9.25	15.21
273	12.04	9.30	15.27
274	12.09	9.34	15.34
275	12.14	9.38	15.40
276	12.20	9.43	15.46
277	12.25	9.47	15.52
278	12.30	9.52	15.58
279	12.35	9.56	15.64
280	12.40	9.60	15.70
281	12.46	9.65	15.76
282	12.51	9.69	15.82
283	12.56	9.74	15.89
284	12.62	9.78	15.95
285	12.67	9.83	16.01
286	12.72	9.87	16.07
287	12.78	9.92	16.14
288	12.83	9.96	16.20
289	12.88	10.01	16.26

Table A3.3 – MARS model for humerus distal breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	2.085***
h(HDB – 43.96)	0.017***
h(43.96 – HDB)	-0.061***
Observations	706
cv R2	0.84
Adjusted R2	0.85
Residual Std. Error	1.56
F Statistic	<0.0001***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A3.4 – The point estimate (fit) and 95% prediction interval (in years) for HDB.

HDB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
18	0.12	0	0.55
19	0.17	0.01	0.69
20	0.24	0.02	0.84
21	0.31	0.04	1.02
22	0.41	0.07	1.22
23	0.52	0.11	1.44
24	0.65	0.16	1.69
25	0.8	0.22	1.97
26	0.97	0.29	2.27
27	1.16	0.38	2.61
28	1.38	0.49	2.98
29	1.62	0.61	3.38
30	1.89	0.75	3.81
31	2.19	0.92	4.28
32	2.51	1.1	4.79
33	2.87	1.31	5.34
34	3.26	1.55	5.92
35	3.69	1.81	6.55
36	4.14	2.1	7.21
37	4.64	2.41	7.92
38	5.16	2.76	8.67
39	5.72	3.13	9.45

HDB (mm)	***Fit***	***Lower 95% PI***	***Upper 95% PI***
40	6.3	3.52	10.26
38	5.16	2.76	8.67
39	5.72	3.13	9.45
40	6.3	3.52	10.26
41	6.92	3.94	11.11
42	7.58	4.39	12.01
43	8.27	4.88	12.95
44	8.94	5.36	13.86
45	9.23	5.56	14.24
46	9.5	5.75	14.59
47	9.76	5.94	14.94
48	10.01	6.12	15.27
49	10.25	6.29	15.59
50	10.49	6.47	15.91
51	10.74	6.65	16.24
52	10.99	6.83	16.57
53	11.24	7.01	16.9
54	11.5	7.2	17.24
55	11.76	7.39	17.58
56	12.03	7.59	17.93
57	12.3	7.79	18.28
58	12.57	7.99	18.63

Table A3.5 – MARS model for humerus midshaft breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Sqrt of age
(Intercept)	2.042***
h(HMSB – 15.93)	0.041***
h(15.93 – HMSB)	-0.153***
Observations	959
cv R2	0.67
Adjusted R2	0.68
Residual Std. Error	2.19
F Statistic	<0.0001***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A3.6 – The point estimate (fit) and 95% prediction interval (in years) for HMSB.

HMSB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
9	0.91	0.00	3.43
10	1.43	0.09	4.38
11	2.18	0.34	5.64
12	3.10	0.75	7.07
13	4.19	1.32	8.66
14	5.43	2.06	10.42
15	6.85	2.96	12.35
16	8.42	4.02	14.43
17	9.23	4.58	15.49
18	9.75	4.96	16.16
19	10.29	5.34	16.85
20	10.85	5.74	17.56
21	11.41	6.16	18.28
22	12.00	6.59	19.01

APPENDIX VI – RADIUS: UNIVARIATE AGE ESTIMATION MODELS AND THE ASSOCIATED PREDICTION TABLES

Table A4.1 – MARS model for radius diaphyseal length. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	2.177***
h(RDL – 149.85)	0.013***
h(149.85 – RDL)	-0.017***
h(RDL – 97.87)	-0.005***
h(RDL – 180.07)	-0.006***
Observations	1177
cv R2	0.94
Adjusted R2	0.94
Residual Std. Error	1.024
F Statistic	<0.0001***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table A4.2 – The point estimate (fit) and 95% prediction interval (in years) for RDL.

RDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
56	0.18	0.05	0.45
57	0.20	0.06	0.48
58	0.22	0.06	0.51
59	0.24	0.07	0.54
60	0.26	0.08	0.58
61	0.28	0.09	0.62
62	0.30	0.10	0.65
63	0.32	0.12	0.69
64	0.35	0.13	0.73
65	0.37	0.14	0.78
66	0.40	0.16	0.82
67	0.43	0.17	0.87
68	0.46	0.19	0.92
69	0.49	0.20	0.96
70	0.52	0.22	1.02
71	0.56	0.24	1.07
72	0.59	0.26	1.12
73	0.63	0.29	1.18
74	0.67	0.31	1.24
75	0.71	0.33	1.30
76	0.75	0.36	1.36
77	0.79	0.38	1.42
78	0.84	0.41	1.49
79	0.89	0.44	1.56
80	0.93	0.47	1.63
81	0.98	0.50	1.70
82	1.04	0.54	1.78
83	1.09	0.57	1.85
84	1.14	0.61	1.93
85	1.20	0.65	2.01
86	1.26	0.68	2.10
87	1.32	0.72	2.18
88	1.39	0.77	2.27
89	1.45	0.81	2.36
90	1.52	0.86	2.45
91	1.59	0.90	2.55
92	1.66	0.95	2.64
93	1.73	1.00	2.74
94	1.81	1.06	2.84
95	1.88	1.11	2.95
96	1.96	1.17	3.06
97	2.04	1.22	3.16
98	2.13	1.28	3.27
99	2.21	1.34	3.38
100	2.28	1.39	3.48
101	2.34	1.44	3.56
102	2.41	1.49	3.65
103	2.48	1.54	3.74
104	2.54	1.59	3.83
105	2.61	1.64	3.92
106	2.69	1.69	4.01
107	2.76	1.74	4.11
108	2.83	1.80	4.21
109	2.91	1.85	4.30
110	2.98	1.91	4.40
111	3.06	1.97	4.50
112	3.14	2.03	4.61
113	3.22	2.09	4.71
114	3.30	2.15	4.81
115	3.39	2.21	4.92
116	3.47	2.27	5.03
117	3.56	2.34	5.14
118	3.65	2.41	5.25
119	3.73	2.47	5.37
120	3.82	2.54	5.48
121	3.92	2.61	5.60
122	4.01	2.68	5.71
123	4.10	2.75	5.83
124	4.20	2.83	5.96
125	4.30	2.90	6.08
126	4.40	2.98	6.20
127	4.50	3.06	6.33
128	4.60	3.14	6.46
129	4.70	3.22	6.59
130	4.81	3.30	6.72
131	4.92	3.38	6.85
132	5.02	3.47	6.99

RDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
133	5.13	3.55	7.13
134	5.25	3.64	7.27
135	5.36	3.73	7.41
136	5.47	3.82	7.55
137	5.59	3.91	7.69
138	5.71	4.00	7.84
139	5.83	4.10	7.99
140	5.95	4.20	8.14
141	6.07	4.29	8.29
142	6.20	4.39	8.44
143	6.32	4.49	8.60
144	6.45	4.59	8.75
145	6.58	4.70	8.91
146	6.71	4.80	9.07
147	6.85	4.91	9.24
148	6.98	5.02	9.40
149	7.11	5.12	9.55
150	7.23	5.22	9.71
151	7.33	5.30	9.83
152	7.43	5.38	9.95
153	7.53	5.46	10.06
154	7.62	5.53	10.18
155	7.72	5.61	10.30
156	7.82	5.69	10.42
157	7.92	5.77	10.54
158	8.02	5.85	10.66
159	8.12	5.94	10.78
160	8.22	6.02	10.90
161	8.32	6.10	11.03
162	8.43	6.19	11.15
163	8.53	6.27	11.28
164	8.64	6.36	11.40
165	8.74	6.44	11.53
166	8.85	6.53	11.66
167	8.96	6.62	11.79
168	9.07	6.71	11.92
169	9.18	6.80	12.05
170	9.29	6.89	12.19
171	9.40	6.98	12.32
172	9.51	7.07	12.45
173	9.62	7.16	12.59
174	9.74	7.26	12.72
175	9.85	7.35	12.86
176	9.97	7.45	13.00
177	10.08	7.54	13.14
178	10.20	7.64	13.28
179	10.32	7.74	13.42
180	10.42	7.82	13.54
181	10.50	7.89	13.64
182	10.57	7.94	13.71
183	10.62	7.99	13.78
184	10.66	8.02	13.83
185	10.70	8.05	13.88
186	10.74	8.09	13.92
187	10.78	8.12	13.97
188	10.83	8.16	14.02
189	10.87	8.19	14.07
190	10.91	8.22	14.12
191	10.95	8.26	14.17
192	10.99	8.29	14.22
193	11.03	8.33	14.26
194	11.07	8.36	14.31
195	11.11	8.39	14.36
196	11.15	8.43	14.41
197	11.19	8.46	14.46
198	11.24	8.50	14.51
199	11.28	8.53	14.56
200	11.32	8.57	14.61
201	11.36	8.60	14.66
202	11.40	8.64	14.71
203	11.45	8.67	14.76
204	11.49	8.71	14.81
205	11.53	8.74	14.86
206	11.57	8.78	14.91
207	11.61	8.81	14.96
208	11.66	8.85	15.01
209	11.70	8.88	15.06
210	11.74	8.92	15.11
211	11.78	8.95	15.16
212	11.83	8.99	15.21

RDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
213	11.87	9.02	15.26
214	11.91	9.06	15.31
215	11.96	9.10	15.36
216	12.00	9.13	15.41
217	12.04	9.17	15.46
218	12.09	9.20	15.52
219	12.13	9.24	15.57
220	12.17	9.28	15.62
221	12.22	9.31	15.67
222	12.26	9.35	15.72
223	12.31	9.39	15.77
224	12.35	9.42	15.83
225	12.39	9.46	15.88
226	12.44	9.50	15.93
227	12.48	9.54	15.98
228	12.53	9.57	16.04
229	12.57	9.61	16.09
230	12.62	9.65	16.14
231	12.66	9.69	16.19

Table A4.3 – MARS model for radius proximal breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Sqrt of age
(Intercept)	2.371***
h(RPB – 15.73)	-0.211***
h(15.73 – RPB)	0.095***
Observations	519
cv R2	0.86
Adjusted R2	0.87
Residual Std. Error	1.389
F Statistic	<0.0001***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table A4.4 – The point estimate (fit) and 95% prediction interval (in years) for RPB.

RPB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
8	0.57	0.03	1.77
9	0.94	0.15	2.40
10	1.61	0.48	3.41
11	2.46	0.98	4.60
12	3.48	1.66	5.98
13	4.79	2.59	7.66
14	6.35	3.77	9.59
15	7.85	4.95	11.43
16	9.25	6.06	13.10
17	9.86	6.56	13.83
18	10.48	7.07	14.57
19	11.08	7.57	15.27
20	11.69	8.07	15.99
21	12.32	8.60	16.72
22	12.97	9.14	17.47

Table A4.5 – MARS model for radius distal breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	2.014***
h(RDB – 24.4)	0.02**
h(24.4 – RDB)	-0.111***
Observations	562
cv R2	0.86
Adjusted R2	0.86
Residual Std. Error	1.48
F Statistic	<0.0001***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A4.6 – The point estimate (fit) and 95% prediction interval (in years) for RDB.

RDB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
9	0.08	0.00	0.41
10	0.15	0.01	0.62
11	0.27	0.04	0.90
12	0.44	0.09	1.25
13	0.66	0.17	1.68
14	0.95	0.30	2.19
15	1.31	0.47	2.81
16	1.75	0.70	3.53
17	2.28	1.00	4.36
18	2.91	1.38	5.31
19	3.65	1.83	6.40
20	4.50	2.38	7.62
21	5.48	3.03	8.99
22	6.58	3.78	10.51
23	7.83	4.65	12.20
24	9.22	5.64	14.05
25	9.98	6.19	15.05
26	10.26	6.40	15.43
27	10.55	6.61	15.80
28	10.84	6.82	16.19
29	11.14	7.04	16.58
30	11.44	7.27	16.97
31	11.75	7.50	17.38

Table A4.7 – MARS model for radius midshaft breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	1.976***
h(RMSB – 10.73)	0.099**
h(10.73 – RMSB)	-0.219***
h(RMSB – 12.71)	-0.081***
Observations	431
cv R2	0.81
Adjusted R2	0.82
Residual Std. Error	1.72
F Statistic	678***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A4.8 – The point estimate (fit) and 95% prediction interval (in years) for RMSB.

RMSB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
4	0.12	0.00	0.63
5	0.37	0.05	1.26
6	0.83	0.19	2.20
7	1.56	0.51	3.52
8	2.63	1.06	5.29
9	4.10	1.90	7.56
10	6.05	3.11	10.42
11	8.11	4.47	13.32
12	9.11	5.15	14.72
13	10.09	5.82	16.05
14	10.82	6.33	17.05
15	11.58	6.86	18.08
16	12.38	7.43	19.14

APPENDIX V – ULNA: UNIVARIATE AGE ESTIMATION MODELS AND THE ASSOCIATED PREDICTION TABLES

Table A5.1 – MARS model for ulna diaphyseal length. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	2.064***
h(UDL – 161.55)	0.012***
h(161.44 – UDL)	-0.015***
h(UDL – 195.38)	-0.006***
h(UDL – 112.52)	-0.004***
Observations	1026
cv R2	0.93
Adjusted R2	0.93
Residual Std. Error	1.01
F Statistic	<0.0001***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table A5.2 – The point estimate (fit) and 95% prediction interval (in years) for UDL.

UDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
54	0.08	0.01	0.25
55	0.08	0.01	0.27
56	0.09	0.02	0.29
57	0.10	0.02	0.31
58	0.11	0.02	0.33
59	0.12	0.03	0.35
60	0.14	0.03	0.37
61	0.15	0.03	0.40
62	0.16	0.04	0.42
63	0.18	0.04	0.45
64	0.19	0.05	0.48
65	0.21	0.06	0.50
66	0.22	0.06	0.53
67	0.24	0.07	0.56
68	0.26	0.08	0.60
69	0.28	0.09	0.63
70	0.30	0.10	0.66
71	0.32	0.11	0.70
72	0.34	0.12	0.74
73	0.36	0.13	0.77
74	0.39	0.14	0.81
75	0.41	0.16	0.85
76	0.44	0.17	0.90
77	0.46	0.18	0.94
78	0.49	0.20	0.98
79	0.52	0.22	1.03
80	0.55	0.23	1.08
81	0.58	0.25	1.13
82	0.61	0.27	1.18
83	0.65	0.29	1.23
84	0.68	0.31	1.28
85	0.72	0.33	1.34
86	0.76	0.35	1.39
87	0.80	0.38	1.45
88	0.84	0.40	1.51
89	0.88	0.43	1.57
90	0.92	0.45	1.63
91	0.96	0.48	1.70
92	1.01	0.51	1.76
93	1.06	0.54	1.83
94	1.10	0.57	1.90
95	1.15	0.60	1.97
96	1.20	0.63	2.04
97	1.26	0.67	2.12
98	1.31	0.70	2.20
99	1.37	0.74	2.27
100	1.42	0.78	2.35
101	1.48	0.82	2.44
102	1.54	0.86	2.52
103	1.61	0.90	2.61
104	1.67	0.94	2.69
105	1.73	0.99	2.78
106	1.80	1.04	2.87
107	1.87	1.08	2.97
108	1.94	1.13	3.06
109	2.01	1.18	3.16
110	2.09	1.24	3.26
111	2.16	1.29	3.36
112	2.24	1.34	3.46
113	2.30	1.39	3.55
114	2.36	1.43	3.63
115	2.43	1.48	3.71
116	2.49	1.52	3.80
117	2.56	1.57	3.89
118	2.62	1.62	3.97
119	2.69	1.67	4.06
120	2.76	1.72	4.15
121	2.83	1.77	4.25
122	2.90	1.82	4.34
123	2.97	1.87	4.44
124	3.05	1.93	4.53
125	3.12	1.98	4.63
126	3.20	2.04	4.73
127	3.28	2.10	4.83
128	3.36	2.16	4.93
129	3.44	2.22	5.03
130	3.52	2.28	5.14

UDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
131	3.60	2.34	5.24
132	3.68	2.40	5.35
133	3.77	2.47	5.46
134	3.85	2.53	5.57
135	3.94	2.60	5.68
136	4.03	2.67	5.80
137	4.12	2.73	5.91
138	4.21	2.80	6.03
139	4.31	2.87	6.15
140	4.40	2.95	6.27
141	4.50	3.02	6.39
142	4.59	3.09	6.51
143	4.69	3.17	6.64
144	4.79	3.25	6.76
145	4.89	3.33	6.89
146	5.00	3.41	7.02
147	5.10	3.49	7.15
148	5.21	3.57	7.28
149	5.31	3.65	7.41
150	5.42	3.74	7.55
151	5.53	3.82	7.69
152	5.64	3.91	7.83
153	5.76	4.00	7.97
154	5.87	4.09	8.11
155	5.99	4.18	8.25
156	6.10	4.27	8.40
157	6.22	4.37	8.54
158	6.34	4.46	8.69
159	6.47	4.56	8.84
160	6.59	4.65	8.99
161	6.71	4.75	9.14
162	6.81	4.83	9.27
163	6.91	4.91	9.40
164	7.01	4.99	9.51
165	7.11	5.07	9.63
166	7.20	5.14	9.75
167	7.30	5.22	9.86
168	7.39	5.30	9.98
169	7.49	5.38	10.10
170	7.59	5.45	10.22

UDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
171	7.69	5.53	10.34
172	7.79	5.61	10.46
173	7.89	5.69	10.58
174	7.99	5.77	10.70
175	8.09	5.85	10.83
176	8.19	5.93	10.95
177	8.29	6.02	11.07
178	8.39	6.10	11.20
179	8.50	6.18	11.32
180	8.60	6.27	11.45
181	8.70	6.35	11.58
182	8.81	6.44	11.70
183	8.92	6.53	11.83
184	9.03	6.61	11.96
185	9.13	6.70	12.09
186	9.24	6.79	12.23
187	9.35	6.88	12.36
188	9.47	6.97	12.49
189	9.58	7.06	12.63
190	9.69	7.16	12.76
191	9.81	7.25	12.90
192	9.92	7.34	13.04
193	10.03	7.43	13.17
194	10.13	7.52	13.30
195	10.23	7.60	13.41
196	10.31	7.66	13.50
197	10.38	7.72	13.59
198	10.45	7.77	13.67
199	10.51	7.82	13.74
200	10.56	7.87	13.81
201	10.61	7.91	13.87
202	10.65	7.94	13.92
203	10.68	7.97	13.95
204	10.71	8.00	13.99
205	10.75	8.02	14.03
206	10.78	8.05	14.07
207	10.81	8.07	14.10
208	10.84	8.10	14.14
209	10.87	8.12	14.18
210	10.90	8.15	14.22

UDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
211	10.94	8.18	14.25
212	10.97	8.20	14.29
213	11.00	8.23	14.33
214	11.03	8.26	14.37
215	11.06	8.28	14.41
216	11.10	8.31	14.45
217	11.13	8.34	14.48
218	11.16	8.36	14.52
219	11.19	8.39	14.56
220	11.22	8.42	14.60
221	11.26	8.44	14.64
222	11.29	8.47	14.68
223	11.32	8.50	14.71
224	11.35	8.52	14.75
225	11.39	8.55	14.79
226	11.42	8.58	14.83
227	11.45	8.60	14.87
228	11.49	8.63	14.91
229	11.52	8.66	14.95
230	11.55	8.69	14.99
231	11.58	8.71	15.03
232	11.62	8.74	15.07
233	11.65	8.77	15.11
234	11.68	8.79	15.15
235	11.72	8.82	15.18
236	11.75	8.85	15.22
237	11.78	8.88	15.26
238	11.82	8.91	15.30
239	11.85	8.93	15.34
240	11.88	8.96	15.38
241	11.92	8.99	15.42
242	11.95	9.02	15.46

Table A5.3 – MARS model for ulna midshaft breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	1.886***
h(10.32 – UMSB)	-0.214***
h(UMSB – 5.71)	0.034**
Observations	392
cv R2	0.79
Adjusted R2	0.80
Residual Std. Error	1.76
F Statistic	810***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A5.4 – The point estimate (fit) and 95% prediction interval (in years) for UMSB.

UMSB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
4	0.14	0.00	0.70
5	0.41	0.05	1.40
6	0.93	0.22	2.46
7	1.83	0.61	4.07
8	3.18	1.32	6.27
9	5.07	2.43	9.15
10	7.59	4.03	12.79
11	8.83	4.86	14.53
12	9.27	5.15	15.14
13	9.72	5.46	15.77
14	10.19	5.78	16.42

APPENDIX VI – FEMUR: UNIVARIATE AGE ESTIMATION MODELS AND THE ASSOCIATED PREDICTION TABLES

Table A6.1 – The point estimate (fit) and 95% prediction interval (in years) for FDL.

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
75	0.14	0.04	0.33
76	0.15	0.05	0.34
77	0.16	0.05	0.35
78	0.16	0.05	0.37
79	0.17	0.06	0.38
80	0.18	0.06	0.39
81	0.19	0.07	0.4
82	0.19	0.07	0.42
83	0.2	0.07	0.43
84	0.21	0.08	0.45
85	0.22	0.08	0.46
86	0.23	0.09	0.48
87	0.24	0.09	0.49
88	0.25	0.1	0.51
89	0.26	0.1	0.52
90	0.27	0.11	0.54
91	0.28	0.11	0.55
92	0.29	0.12	0.57
93	0.3	0.13	0.59
94	0.31	0.13	0.61
95	0.32	0.14	0.62
96	0.33	0.15	0.64
97	0.35	0.15	0.66
98	0.36	0.16	0.68
99	0.37	0.17	0.7
100	0.38	0.17	0.72
101	0.4	0.18	0.74
102	0.41	0.19	0.76
103	0.42	0.2	0.78
104	0.44	0.21	0.8
105	0.45	0.21	0.82
106	0.47	0.22	0.84
107	0.48	0.23	0.86
108	0.5	0.24	0.89
109	0.51	0.25	0.91
110	0.53	0.26	0.93
111	0.54	0.27	0.96
112	0.56	0.28	0.98
113	0.58	0.29	1
114	0.59	0.3	1.03
115	0.61	0.32	1.05
116	0.63	0.33	1.08
117	0.65	0.34	1.11
118	0.67	0.35	1.13
119	0.69	0.36	1.16
120	0.71	0.38	1.19
121	0.72	0.39	1.21
122	0.74	0.4	1.24
123	0.77	0.42	1.27
124	0.79	0.43	1.3
125	0.81	0.44	1.33
126	0.83	0.46	1.36
127	0.85	0.47	1.39
128	0.87	0.49	1.42
129	0.9	0.5	1.45
130	0.92	0.52	1.48
131	0.94	0.53	1.52
132	0.97	0.55	1.55
133	0.99	0.57	1.58
134	1.01	0.58	1.61
135	1.04	0.6	1.65
136	1.06	0.62	1.68
137	1.09	0.64	1.72
138	1.12	0.66	1.75
139	1.14	0.67	1.79

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
140	1.17	0.69	1.83
141	1.2	0.71	1.86
142	1.22	0.73	1.9
143	1.25	0.75	1.94
144	1.28	0.77	1.98
145	1.31	0.79	2.01
146	1.34	0.82	2.05
147	1.37	0.84	2.09
148	1.4	0.86	2.13
149	1.43	0.88	2.18
150	1.46	0.9	2.22
151	1.5	0.93	2.26
152	1.53	0.95	2.3
153	1.56	0.97	2.34
154	1.59	1	2.39
155	1.63	1.02	2.43
156	1.66	1.05	2.48
157	1.7	1.07	2.52
158	1.73	1.1	2.57
159	1.77	1.13	2.61
160	1.8	1.15	2.66
161	1.84	1.18	2.71
162	1.88	1.21	2.75
163	1.91	1.23	2.8
164	1.95	1.26	2.85
165	1.98	1.29	2.89
166	2.02	1.31	2.94
167	2.05	1.34	2.98
168	2.09	1.37	3.03
169	2.12	1.39	3.07
170	2.16	1.42	3.12
171	2.19	1.45	3.16
172	2.23	1.47	3.21
173	2.27	1.5	3.26
174	2.3	1.53	3.3
175	2.34	1.56	3.35
176	2.37	1.58	3.39
177	2.4	1.6	3.43
178	2.43	1.63	3.47
179	2.47	1.65	3.51

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
180	2.5	1.68	3.55
181	2.53	1.7	3.59
182	2.56	1.73	3.63
183	2.6	1.75	3.67
184	2.63	1.78	3.71
185	2.66	1.8	3.76
186	2.7	1.83	3.8
187	2.73	1.86	3.84
188	2.76	1.88	3.88
189	2.8	1.91	3.93
190	2.83	1.94	3.97
191	2.87	1.96	4.01
192	2.9	1.99	4.06
193	2.94	2.02	4.1
194	2.97	2.05	4.15
195	3.01	2.08	4.19
196	3.05	2.1	4.24
197	3.08	2.13	4.28
198	3.12	2.16	4.33
199	3.16	2.19	4.38
200	3.2	2.22	4.42
201	3.23	2.25	4.47
202	3.27	2.28	4.52
203	3.31	2.31	4.57
204	3.35	2.34	4.61
205	3.39	2.37	4.66
206	3.43	2.4	4.71
207	3.47	2.43	4.76
208	3.51	2.47	4.81
209	3.55	2.5	4.86
210	3.59	2.53	4.91
211	3.63	2.56	4.96
212	3.67	2.6	5.01
213	3.71	2.63	5.06
214	3.76	2.66	5.11
215	3.8	2.7	5.17
216	3.84	2.73	5.22
217	3.88	2.76	5.27
218	3.93	2.8	5.32
219	3.97	2.83	5.38

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
220	4.01	2.87	5.43
221	4.06	2.9	5.49
222	4.1	2.94	5.54
223	4.15	2.97	5.59
224	4.19	3.01	5.65
225	4.24	3.05	5.7
226	4.28	3.08	5.76
227	4.33	3.12	5.82
228	4.38	3.16	5.87
229	4.42	3.2	5.93
230	4.47	3.23	5.99
231	4.52	3.27	6.05
232	4.57	3.31	6.1
233	4.61	3.35	6.16
234	4.66	3.39	6.22
235	4.71	3.43	6.28
236	4.76	3.47	6.34
237	4.81	3.51	6.4
238	4.86	3.55	6.46
239	4.91	3.59	6.52
240	4.96	3.63	6.58
241	5.01	3.67	6.64
242	5.06	3.71	6.7
243	5.11	3.76	6.77
244	5.17	3.8	6.83
245	5.22	3.84	6.89
246	5.27	3.88	6.96
247	5.32	3.93	7.02
248	5.38	3.97	7.08
249	5.43	4.01	7.15
250	5.49	4.06	7.21
251	5.54	4.1	7.28
252	5.59	4.15	7.34
253	5.65	4.19	7.41
254	5.7	4.24	7.48
255	5.76	4.28	7.54
256	5.82	4.33	7.61
257	5.87	4.38	7.68
258	5.93	4.42	7.75
259	5.99	4.47	7.81

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
260	6.04	4.52	7.88
261	6.1	4.56	7.95
262	6.16	4.61	8.02
263	6.22	4.66	8.09
264	6.27	4.71	8.16
265	6.33	4.75	8.22
266	6.39	4.8	8.29
267	6.44	4.85	8.35
268	6.5	4.89	8.42
269	6.55	4.94	8.48
270	6.61	4.98	8.55
271	6.66	5.03	8.62
272	6.72	5.07	8.68
273	6.77	5.12	8.75
274	6.83	5.17	8.81
275	6.88	5.21	8.88
276	6.94	5.26	8.95
277	7	5.31	9.01
278	7.05	5.35	9.08
279	7.11	5.4	9.15
280	7.17	5.45	9.21
281	7.22	5.49	9.28
282	7.28	5.54	9.35
283	7.34	5.59	9.42
284	7.39	5.64	9.48
285	7.45	5.68	9.55
286	7.5	5.72	9.61
287	7.55	5.77	9.67
288	7.6	5.81	9.73
289	7.65	5.85	9.79
290	7.7	5.89	9.85
291	7.75	5.94	9.91
292	7.81	5.98	9.97
293	7.86	6.02	10.03
294	7.91	6.07	10.09
295	7.96	6.11	10.15
296	8.01	6.16	10.21
297	8.07	6.2	10.28
298	8.12	6.24	10.34
299	8.17	6.29	10.4

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
300	8.23	6.33	10.46
301	8.28	6.38	10.52
302	8.33	6.42	10.59
303	8.39	6.47	10.65
304	8.44	6.51	10.71
305	8.5	6.56	10.78
306	8.55	6.61	10.84
307	8.6	6.65	10.91
308	8.66	6.7	10.97
309	8.72	6.74	11.04
310	8.77	6.79	11.1
311	8.83	6.84	11.17
312	8.88	6.89	11.23
313	8.94	6.93	11.3
314	9	6.98	11.36
315	9.05	7.03	11.43
316	9.11	7.08	11.5
317	9.17	7.13	11.56
318	9.22	7.17	11.63
319	9.28	7.22	11.7
320	9.34	7.27	11.76
321	9.4	7.32	11.83
322	9.46	7.37	11.9
323	9.51	7.42	11.97
324	9.57	7.47	12.04
325	9.63	7.52	12.11
326	9.69	7.57	12.18
327	9.75	7.62	12.24
328	9.81	7.67	12.31
329	9.87	7.72	12.38
330	9.93	7.77	12.45
331	9.99	7.83	12.53
332	10.05	7.88	12.6
333	10.11	7.93	12.67
334	10.17	7.98	12.74
335	10.23	8.03	12.81
336	10.3	8.08	12.88
337	10.36	8.14	12.95
338	10.41	8.18	13.01
339	10.46	8.23	13.07

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
340	10.51	8.27	13.13
341	10.56	8.31	13.19
342	10.62	8.36	13.25
343	10.66	8.39	13.3
344	10.7	8.43	13.35
345	10.73	8.46	13.38
346	10.76	8.48	13.41
347	10.78	8.5	13.44
348	10.8	8.52	13.47
349	10.83	8.54	13.49
350	10.85	8.56	13.52
351	10.87	8.58	13.55
352	10.9	8.6	13.57
353	10.92	8.61	13.6
354	10.94	8.63	13.63
355	10.97	8.65	13.65
356	10.99	8.67	13.68
357	11.01	8.7	13.71
358	11.04	8.72	13.74
359	11.06	8.74	13.76
360	11.08	8.76	13.79
361	11.11	8.78	13.82
362	11.13	8.8	13.85
363	11.15	8.82	13.87
364	11.18	8.84	13.9
365	11.2	8.86	13.93
366	11.23	8.88	13.95
367	11.25	8.9	13.98
368	11.27	8.92	14.01
369	11.3	8.94	14.04
370	11.32	8.96	14.07
371	11.34	8.98	14.09
372	11.37	9	14.12
373	11.39	9.02	14.15
374	11.42	9.04	14.18
375	11.44	9.06	14.2
376	11.46	9.08	14.23
377	11.49	9.1	14.26
378	11.51	9.12	14.29
379	11.54	9.14	14.32

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
380	11.56	9.17	14.34
381	11.59	9.19	14.37
382	11.61	9.21	14.4
383	11.63	9.23	14.43
384	11.66	9.25	14.46
385	11.68	9.27	14.48
386	11.71	9.29	14.51
387	11.73	9.31	14.54
388	11.76	9.33	14.57
389	11.78	9.35	14.6
390	11.81	9.37	14.63
391	11.83	9.4	14.65
392	11.86	9.42	14.68
393	11.88	9.44	14.71
394	11.91	9.46	14.74
395	11.93	9.48	14.77
396	11.95	9.5	14.8
397	11.98	9.52	14.83
398	12	9.54	14.85
399	12.03	9.57	14.88

FDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
400	12.05	9.59	14.91
401	12.08	9.61	14.94
402	12.1	9.63	14.97
403	12.13	9.65	15
404	12.15	9.67	15.03
405	12.18	9.69	15.06
406	12.2	9.72	15.08
407	12.23	9.74	15.11
408	12.25	9.76	15.14
409	12.28	9.78	15.17
410	12.31	9.8	15.2
411	12.33	9.82	15.23
412	12.36	9.85	15.26
413	12.38	9.87	15.29
414	12.41	9.89	15.32
415	12.43	9.91	15.35
416	12.46	9.93	15.38
417	12.48	9.96	15.41
418	12.51	9.98	15.44

Table A6.2 – MARS model for femur midshaft breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	2.041***
h(FMSB – 19.89)	0.053***
h(19.89 – FMSB)	-0.106***
h(FMSB – 22.56)	-0.038**
Observations	940
cv R2	0.73
Adjusted R2	0.75
Residual Std. Error	1.62
F Statistic	937***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A6.3 – The point estimate (fit) and 95% prediction interval (in years) for FMSB.

FMSB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
10	0.98	0.31	2.25
11	1.33	0.48	2.83
12	1.74	0.70	3.51
13	2.24	0.98	4.29
14	2.83	1.32	5.18
15	3.51	1.74	6.19
16	4.29	2.24	7.31
17	5.22	2.85	8.63
18	6.30	3.58	10.12
19	7.40	4.35	11.63
20	8.50	5.13	13.09
21	9.22	5.65	14.06
22	10.00	6.21	15.08
23	10.51	6.58	15.75
24	10.71	6.73	16.01
25	10.91	6.88	16.28
26	11.12	7.03	16.55
27	11.33	7.18	16.82
28	11.54	7.34	17.09
29	11.75	7.50	17.37
30	11.97	7.66	17.65

Table A6.4 – MARS model for femur distal breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Sqrt of age
(Intercept)	2.148***
h(FDB – 67.47)	0.009***
h(67.47 – FDB)	-0.035***
Observations	809
cv R2	0.76
Adjusted R2	0.77
Residual Std. Error	1.51
F Statistic	<0.0001***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table A6.5 – The point estimate (fit) and 95% prediction interval (in years) for FDB.

FDB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
36	1.24	0.28	2.88
37	1.36	0.34	3.06
38	1.49	0.40	3.26
39	1.62	0.48	3.45
40	1.76	0.55	3.65
41	1.91	0.64	3.86
42	2.06	0.72	4.08
43	2.22	0.82	4.30
44	2.38	0.92	4.52
45	2.55	1.03	4.76
46	2.72	1.14	4.99
47	2.90	1.26	5.24
48	3.09	1.38	5.49
49	3.29	1.51	5.75
50	3.55	1.69	6.09
51	3.82	1.88	6.45
52	4.11	2.08	6.82
53	4.40	2.29	7.20
54	4.71	2.52	7.59
55	5.03	2.75	7.99
56	5.36	2.99	8.40
57	5.69	3.25	8.82
58	6.04	3.51	9.26
59	6.40	3.79	9.70
60	6.77	4.07	10.15
61	7.15	4.37	10.61
62	7.54	4.67	11.09
63	7.94	4.99	11.57
64	8.35	5.31	12.06
65	8.77	5.65	12.57
66	9.20	6.00	13.08
67	9.64	6.35	13.60
68	10.07	6.70	14.11
69	10.20	6.81	14.27
70	10.32	6.91	14.41
71	10.44	7.01	14.56
72	10.56	7.11	14.70
73	10.69	7.21	14.85
74	10.81	7.31	14.99
75	10.93	7.41	15.14
76	11.06	7.52	15.28
77	11.19	7.62	15.43
78	11.31	7.73	15.58
79	11.44	7.83	15.73
80	11.57	7.94	15.88
81	11.70	8.04	16.03
82	11.83	8.15	16.18
83	11.96	8.26	16.34
84	12.09	8.37	16.49

APPENDIX VII – TIBIA: UNIVARIATE AGE ESTIMATION MODELS AND THE ASSOCIATED PREDICTION TABLES

Table A7.1 – MARS model for tibia diaphyseal length. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Sqrt of age
(Intercept)	2.271***
h(TDL – 226.13)	0.009***
h(226.13 – TDL)	-0.011***
h(TDL – 130.53)	-0.004***
h(TDL – 272.25)	-0.003***
Observations	1036
cv R2	0.94
Adjusted R2	0.94
Residual Std. Error	0.95
F Statistic	<0.0001***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table A7.2 – The point estimate (fit) and 95% prediction interval (in years) for TDL.

TDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
61	0.16	0.00	0.59
62	0.18	0.00	0.62
63	0.19	0.00	0.64
64	0.20	0.01	0.66
65	0.21	0.01	0.68
66	0.22	0.01	0.70
67	0.24	0.01	0.73
68	0.25	0.02	0.75
69	0.26	0.02	0.77
70	0.28	0.03	0.80
71	0.29	0.03	0.82
72	0.31	0.03	0.84
73	0.32	0.04	0.87
74	0.34	0.05	0.89
75	0.35	0.05	0.92
76	0.37	0.06	0.95
77	0.38	0.06	0.97
78	0.40	0.07	1.00
79	0.42	0.08	1.03
80	0.44	0.09	1.05
81	0.45	0.09	1.08
82	0.47	0.10	1.11
83	0.49	0.11	1.14
84	0.51	0.12	1.17
85	0.53	0.13	1.20
86	0.55	0.14	1.23
87	0.57	0.15	1.26
88	0.59	0.16	1.29
89	0.61	0.17	1.32
90	0.63	0.18	1.35
91	0.65	0.20	1.38
92	0.68	0.21	1.41
93	0.70	0.22	1.44
94	0.72	0.23	1.48
95	0.74	0.25	1.51
96	0.77	0.26	1.54
97	0.79	0.27	1.58
98	0.81	0.29	1.61
99	0.84	0.30	1.64
100	0.86	0.32	1.68
101	0.89	0.33	1.71
102	0.91	0.35	1.75
103	0.94	0.36	1.78
104	0.97	0.38	1.82
105	0.99	0.40	1.86
106	1.02	0.42	1.89
107	1.05	0.43	1.93
108	1.08	0.45	1.97
109	1.10	0.47	2.01
110	1.13	0.49	2.05
111	1.16	0.51	2.08
112	1.19	0.53	2.12
113	1.22	0.55	2.16
114	1.25	0.57	2.20
115	1.28	0.59	2.24
116	1.31	0.61	2.28
117	1.34	0.63	2.32
118	1.37	0.65	2.36
119	1.40	0.67	2.41
120	1.44	0.69	2.45
121	1.47	0.72	2.49
122	1.50	0.74	2.53
123	1.54	0.76	2.58
124	1.57	0.79	2.62
125	1.60	0.81	2.66
126	1.64	0.83	2.71
127	1.67	0.86	2.75
128	1.71	0.88	2.80
129	1.74	0.91	2.84
130	1.78	0.94	2.89
131	1.81	0.96	2.93
132	1.85	0.99	2.98
133	1.89	1.02	3.03
134	1.92	1.04	3.07
135	1.96	1.07	3.12
136	2.00	1.10	3.17
137	2.04	1.13	3.22

TDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
138	2.08	1.16	3.26
139	2.12	1.18	3.31
140	2.15	1.21	3.36
141	2.19	1.24	3.41
142	2.23	1.27	3.46
143	2.27	1.30	3.51
144	2.32	1.34	3.56
145	2.36	1.37	3.61
146	2.40	1.40	3.67
147	2.44	1.43	3.72
148	2.48	1.46	3.77
149	2.52	1.50	3.82
150	2.57	1.53	3.87
151	2.61	1.56	3.93
152	2.65	1.60	3.98
153	2.70	1.63	4.04
154	2.74	1.66	4.09
155	2.79	1.70	4.14
156	2.83	1.73	4.20
157	2.88	1.77	4.25
158	2.92	1.81	4.31
159	2.97	1.84	4.37
160	3.02	1.88	4.42
161	3.06	1.92	4.48
162	3.11	1.95	4.54
163	3.16	1.99	4.59
164	3.21	2.03	4.65
165	3.25	2.07	4.71
166	3.30	2.11	4.77
167	3.35	2.15	4.83
168	3.40	2.19	4.89
169	3.45	2.23	4.95
170	3.50	2.27	5.01
171	3.55	2.31	5.07
172	3.60	2.35	5.13
173	3.65	2.39	5.19
174	3.71	2.43	5.25
175	3.76	2.47	5.31
176	3.81	2.52	5.37
177	3.86	2.56	5.44

TDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
178	3.92	2.60	5.50
179	3.97	2.65	5.56
180	4.02	2.69	5.63
181	4.08	2.73	5.69
182	4.13	2.78	5.75
183	4.19	2.82	5.82
184	4.24	2.87	5.88
185	4.30	2.91	5.95
186	4.35	2.96	6.01
187	4.41	3.01	6.08
188	4.47	3.05	6.15
189	4.52	3.10	6.21
190	4.58	3.15	6.28
191	4.64	3.19	6.35
192	4.69	3.24	6.41
193	4.75	3.29	6.48
194	4.81	3.34	6.55
195	4.87	3.39	6.62
196	4.93	3.44	6.69
197	4.99	3.49	6.76
198	5.05	3.54	6.83
199	5.11	3.59	6.90
200	5.17	3.64	6.97
201	5.23	3.69	7.04
202	5.29	3.74	7.11
203	5.35	3.79	7.18
204	5.42	3.85	7.25
205	5.48	3.90	7.33
206	5.54	3.95	7.40
207	5.60	4.00	7.47
208	5.67	4.06	7.54
209	5.73	4.11	7.62
210	5.80	4.17	7.69
211	5.86	4.22	7.77
212	5.92	4.28	7.84
213	5.99	4.33	7.92
214	6.06	4.39	7.99
215	6.12	4.44	8.07
216	6.19	4.50	8.14
217	6.25	4.56	8.22

TDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
218	6.32	4.61	8.30
219	6.39	4.67	8.37
220	6.46	4.73	8.45
221	6.53	4.79	8.53
222	6.59	4.85	8.61
223	6.66	4.91	8.69
224	6.73	4.97	8.77
225	6.80	5.03	8.84
226	6.87	5.09	8.92
227	6.94	5.15	9.00
228	7.01	5.21	9.08
229	7.08	5.27	9.17
230	7.15	5.33	9.25
231	7.23	5.39	9.33
232	7.30	5.45	9.41
233	7.37	5.52	9.49
234	7.44	5.58	9.57
235	7.52	5.64	9.66
236	7.59	5.71	9.74
237	7.66	5.77	9.82
238	7.74	5.84	9.91
239	7.81	5.90	9.99
240	7.89	5.97	10.08
241	7.96	6.03	10.16
242	8.04	6.10	10.25
243	8.11	6.16	10.33
244	8.19	6.23	10.42
245	8.27	6.30	10.51
246	8.34	6.36	10.59
247	8.42	6.43	10.68
248	8.50	6.50	10.77
249	8.58	6.57	10.86
250	8.66	6.64	10.94
251	8.74	6.71	11.03
252	8.81	6.78	11.12
253	8.89	6.85	11.21
254	8.97	6.92	11.30
255	9.05	6.99	11.39
256	9.14	7.06	11.48
257	9.22	7.13	11.57

TDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
258	9.30	7.20	11.66
259	9.38	7.27	11.76
260	9.46	7.34	11.85
261	9.54	7.42	11.94
262	9.63	7.49	12.03
263	9.70	7.55	12.11
264	9.75	7.60	12.17
265	9.79	7.63	12.22
266	9.83	7.67	12.26
267	9.87	7.70	12.30
268	9.91	7.74	12.34
269	9.94	7.76	12.38
270	9.97	7.79	12.41
271	10.00	7.82	12.45
272	10.03	7.85	12.48
273	10.06	7.87	12.52
274	10.09	7.90	12.55
275	10.12	7.93	12.59
276	10.15	7.96	12.62
277	10.19	7.98	12.66
278	10.22	8.01	12.69
279	10.25	8.04	12.73
280	10.28	8.07	12.76
281	10.31	8.09	12.80
282	10.34	8.12	12.83
283	10.37	8.15	12.87
284	10.41	8.18	12.90
285	10.44	8.21	12.94
286	10.47	8.23	12.97
287	10.50	8.26	13.01
288	10.53	8.29	13.04
289	10.56	8.32	13.08
290	10.60	8.35	13.11
291	10.63	8.37	13.15
292	10.66	8.40	13.18
293	10.69	8.43	13.22
294	10.72	8.46	13.25
295	10.75	8.49	13.29
296	10.79	8.52	13.33
297	10.82	8.54	13.36

TDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
298	10.85	8.57	13.40
299	10.88	8.60	13.43
300	10.92	8.63	13.47
301	10.95	8.66	13.50
302	10.98	8.69	13.54
303	11.01	8.72	13.58
304	11.05	8.75	13.61
305	11.08	8.77	13.65
306	11.11	8.80	13.68
307	11.14	8.83	13.72
308	11.18	8.86	13.76
309	11.21	8.89	13.79
310	11.24	8.92	13.83
311	11.27	8.95	13.87
312	11.31	8.98	13.90
313	11.34	9.01	13.94
314	11.37	9.04	13.98
315	11.41	9.07	14.01
316	11.44	9.10	14.05
317	11.47	9.13	14.08
318	11.50	9.16	14.12
319	11.54	9.18	14.16
320	11.57	9.21	14.20
321	11.60	9.24	14.23
322	11.64	9.27	14.27
323	11.67	9.30	14.31
324	11.70	9.33	14.34
325	11.74	9.36	14.38
326	11.77	9.39	14.42
327	11.80	9.42	14.45
328	11.84	9.45	14.49
329	11.87	9.48	14.53
330	11.91	9.51	14.57
331	11.94	9.54	14.60
332	11.97	9.57	14.64
333	12.01	9.60	14.68
334	12.04	9.63	14.71
335	12.07	9.66	14.75
336	12.11	9.69	14.79
337	12.14	9.73	14.83
338	12.18	9.76	14.87
339	12.21	9.79	14.90
340	12.24	9.82	14.94
341	12.28	9.85	14.98
342	12.31	9.88	15.02
343	12.35	9.91	15.05
344	12.38	9.94	15.09
345	12.42	9.97	15.13
346	12.45	10.00	15.17
347	12.49	10.03	15.21
348	12.52	10.06	15.24
349	12.55	10.09	15.28
350	12.59	10.13	15.32

Table A7.3 – MARS model for tibia proximal breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	2.038***
h(TPB – 226.13)	0.049***
h(226.13 – TPB)	-0.034***
h(TPB – 130.53)	-0.023***
h(TPB – 272.25)	0.021***
h(TPB – 226.13)	-0.039***
Observations	735
cv R2	0.80
Adjusted R2	0.81
Residual Std. Error	1.4
F Statistic	643***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table A7.4 – The point estimate (fit) and 95% prediction interval (in years) for TPB.

TPB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
31	1.22	0.31	2.76
32	1.36	0.38	2.97
33	1.51	0.46	3.18
34	1.67	0.54	3.41
35	1.83	0.64	3.64
36	2.00	0.74	3.88
37	2.18	0.85	4.12
38	2.37	0.97	4.38
39	2.56	1.09	4.64
40	2.76	1.23	4.91
41	3.01	1.39	5.24
42	3.33	1.61	5.65
43	3.66	1.85	6.09
44	4.01	2.10	6.53
45	4.38	2.36	7.00
46	4.76	2.65	7.48
47	5.15	2.94	7.97
48	5.55	3.24	8.46
49	5.93	3.54	8.94
50	6.31	3.84	9.40
51	6.70	4.14	9.87
52	7.09	4.45	10.35
53	7.50	4.77	10.84
54	7.92	5.11	11.34
55	8.35	5.45	11.86
56	8.79	5.81	12.38
57	9.24	6.18	12.91
58	9.70	6.56	13.46
59	10.17	6.95	14.01
60	10.41	7.14	14.29
61	10.53	7.25	14.44
62	10.66	7.35	14.58
63	10.78	7.45	14.73
64	10.91	7.56	14.87
65	11.04	7.66	15.02
66	11.16	7.77	15.17
67	11.29	7.88	15.32
68	11.42	7.98	15.47
69	11.55	8.09	15.62
70	11.68	8.20	15.77
71	11.81	8.31	15.92
72	11.94	8.42	16.08
73	12.07	8.53	16.23
74	12.21	8.64	16.38
75	12.34	8.76	16.54
76	12.48	8.87	16.70
77	12.61	8.98	16.85

Table A7.5 – Spline model for tibia distal breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Sqrt of age
(Intercept)	1.188***
bs(TDB, degree = 2, df = 4)	0.407**
bs(TDB, degree = 2, df = 4)	1.532***
bs(TDB, degree = 2, df = 4)	2.230***
bs(TDB, degree = 2, df = 4)	2.282***
Observations	630
Adjusted R2	0.80
Residual Std. Error	1.35
F Statistic	650***
Note:	***p < 0.01; **p<0.001; ***p<0.0001**

Table A7.6 – The point estimate (fit) and 95% prediction interval (in years) for TDB.

TDB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
21	1.36	0.33	3.09
22	1.54	0.45	3.28
23	1.75	0.59	3.53
24	2.00	0.76	3.83
25	2.27	0.94	4.18
26	2.58	1.15	4.58
27	2.93	1.39	5.03
28	3.32	1.66	5.53
29	3.75	1.98	6.09
30	4.24	2.33	6.71
31	4.78	2.74	7.38
32	5.37	3.19	8.12
33	6.00	3.68	8.88
34	6.60	4.15	9.61
35	7.17	4.61	10.29
36	7.71	5.04	10.94
37	8.20	5.44	11.52
38	8.65	5.81	12.05
39	9.06	6.15	12.54
40	9.46	6.48	13.00
41	9.83	6.79	13.44
42	10.18	7.08	13.85
43	10.50	7.35	14.22
44	10.80	7.59	14.57
45	11.07	7.82	14.88
46	11.31	8.02	15.16
47	11.52	8.19	15.40
48	11.69	8.34	15.62
49	11.84	8.45	15.80
50	11.95	8.53	15.95
51	12.03	8.58	16.07

Table A7.7 – MARS model for tibia midshaft breadth. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Cbrt of age
(Intercept)	2.061***
h(TMSB – 18.99)	0.037***
h(18.99 – TMSB)	-0.108***
Observations	554
cv R2	0.71
Adjusted R2	0.72
Residual Std. Error	1.59
F Statistic	720***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A7.8 – The point estimate (fit) and 95% prediction interval (in years) for TMSB.

TMSB (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
10	1.30	0.47	2.78
11	1.72	0.69	3.47
12	2.23	0.97	4.27
13	2.83	1.33	5.18
14	3.53	1.76	6.21
15	4.33	2.27	7.37
16	5.25	2.88	8.67
17	6.30	3.59	10.11
18	7.47	4.40	11.71
19	8.65	5.24	13.29
20	9.22	5.65	14.05
21	9.73	6.02	14.72
22	10.25	6.40	15.41
23	10.78	6.79	16.10
24	11.33	7.19	16.81
25	11.88	7.60	17.53
26	12.46	8.03	18.28

APPENDIX VIII – FIBULA: UNIVARIATE AGE ESTIMATION MODELS AND THE ASSOCIATED PREDICTION TABLES

Table A8.1 – MARS model for fibula diaphyseal length. The residual standard error (Residual Std. Error) is in years and is not affected by the transformation of age.

	Predictor Variable
	Sqrt of age
(Intercept)	2.281***
h(FBDL – 239.72)	0.007***
h(239.72 – FBDL)	-0.009***
h(FBDL – 135.89)	-0.003***
h(FBDL – 271.99)	-0.003***
Observations	1025
cv R2	0.93
Adjusted R2	0.93
Residual Std. Error	0.99
F Statistic	<0.0001***

Note: ***p < 0.01; **p<0.001; ***p<0.0001**

Table A8.2 – The point estimate (fit) and 95% prediction interval (in years) for FBDL.

FBDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
65	0.26	0.02	0.79
66	0.28	0.02	0.81
67	0.29	0.03	0.83
68	0.31	0.03	0.86
69	0.32	0.04	0.88
70	0.33	0.04	0.91
71	0.35	0.05	0.93
72	0.37	0.05	0.96
73	0.38	0.06	0.98
74	0.40	0.07	1.01
75	0.42	0.07	1.04
76	0.43	0.08	1.06
77	0.45	0.09	1.09
78	0.47	0.10	1.12
79	0.49	0.10	1.15
80	0.50	0.11	1.17
81	0.52	0.12	1.20
82	0.54	0.13	1.23
83	0.56	0.14	1.26
84	0.58	0.15	1.29
85	0.60	0.16	1.32
86	0.62	0.17	1.35
87	0.64	0.18	1.38
88	0.67	0.20	1.41
89	0.69	0.21	1.45
90	0.71	0.22	1.48
91	0.73	0.23	1.51
92	0.75	0.25	1.54
93	0.78	0.26	1.57
94	0.80	0.27	1.61
95	0.82	0.29	1.64
96	0.85	0.30	1.68
97	0.87	0.31	1.71
98	0.90	0.33	1.74
99	0.92	0.35	1.78
100	0.95	0.36	1.81
101	0.97	0.38	1.85
102	1.00	0.39	1.89
103	1.03	0.41	1.92
104	1.05	0.43	1.96
105	1.08	0.44	2.00
106	1.11	0.46	2.03
107	1.14	0.48	2.07
108	1.16	0.50	2.11
109	1.19	0.52	2.15
110	1.22	0.54	2.19
111	1.25	0.56	2.23
112	1.28	0.58	2.27
113	1.31	0.60	2.30
114	1.34	0.62	2.35
115	1.37	0.64	2.39
116	1.40	0.66	2.43
117	1.43	0.68	2.47
118	1.47	0.70	2.51
119	1.50	0.72	2.55
120	1.53	0.75	2.59
121	1.56	0.77	2.64
122	1.60	0.79	2.68
123	1.63	0.82	2.72
124	1.66	0.84	2.77
125	1.70	0.86	2.81
126	1.73	0.89	2.85
127	1.77	0.91	2.90
128	1.80	0.94	2.94
129	1.84	0.97	2.99
130	1.87	0.99	3.03
131	1.91	1.02	3.08
132	1.95	1.04	3.13
133	1.98	1.07	3.17
134	2.02	1.10	3.22
135	2.06	1.13	3.27
136	2.10	1.16	3.32
137	2.13	1.18	3.36
138	2.17	1.21	3.41
139	2.21	1.24	3.46
140	2.25	1.27	3.51
141	2.29	1.30	3.56

FBDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
142	2.33	1.33	3.61
143	2.37	1.36	3.66
144	2.41	1.39	3.71
145	2.45	1.42	3.76
146	2.50	1.46	3.81
147	2.54	1.49	3.87
148	2.58	1.52	3.92
149	2.62	1.55	3.97
150	2.66	1.59	4.02
151	2.71	1.62	4.07
152	2.75	1.65	4.13
153	2.79	1.69	4.18
154	2.84	1.72	4.24
155	2.88	1.76	4.29
156	2.93	1.79	4.35
157	2.97	1.83	4.40
158	3.02	1.86	4.46
159	3.07	1.90	4.51
160	3.11	1.93	4.57
161	3.16	1.97	4.62
162	3.21	2.01	4.68
163	3.25	2.05	4.74
164	3.30	2.08	4.80
165	3.35	2.12	4.85
166	3.40	2.16	4.91
167	3.45	2.20	4.97
168	3.49	2.24	5.03
169	3.54	2.28	5.09
170	3.59	2.32	5.15
171	3.64	2.36	5.21
172	3.69	2.40	5.27
173	3.74	2.44	5.33
174	3.80	2.48	5.39
175	3.85	2.52	5.45
176	3.90	2.56	5.51
177	3.95	2.61	5.57
178	4.00	2.65	5.64
179	4.06	2.69	5.70
180	4.11	2.74	5.76
181	4.16	2.78	5.83

FBDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
182	4.22	2.82	5.89
183	4.27	2.87	5.95
184	4.33	2.91	6.02
185	4.38	2.96	6.08
186	4.44	3.00	6.15
187	4.49	3.05	6.21
188	4.55	3.10	6.28
189	4.60	3.14	6.35
190	4.66	3.19	6.41
191	4.72	3.24	6.48
192	4.78	3.28	6.55
193	4.83	3.33	6.61
194	4.89	3.38	6.68
195	4.95	3.43	6.75
196	5.01	3.48	6.82
197	5.07	3.53	6.89
198	5.13	3.58	6.96
199	5.19	3.63	7.03
200	5.25	3.68	7.10
201	5.31	3.73	7.17
202	5.37	3.78	7.24
203	5.43	3.83	7.31
204	5.49	3.88	7.38
205	5.55	3.93	7.45
206	5.62	3.99	7.52
207	5.68	4.04	7.60
208	5.74	4.09	7.67
209	5.80	4.15	7.74
210	5.87	4.20	7.82
211	5.93	4.25	7.89
212	6.00	4.31	7.96
213	6.06	4.36	8.04
214	6.13	4.42	8.11
215	6.19	4.47	8.19
216	6.26	4.53	8.26
217	6.32	4.59	8.34
218	6.39	4.64	8.42
219	6.46	4.70	8.49
220	6.52	4.76	8.57
221	6.59	4.81	8.65

FBDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
222	6.66	4.87	8.72
223	6.73	4.93	8.80
224	6.80	4.99	8.88
225	6.86	5.05	8.96
226	6.93	5.11	9.04
227	7.00	5.17	9.12
228	7.07	5.23	9.20
229	7.14	5.29	9.28
230	7.21	5.35	9.36
231	7.28	5.41	9.44
232	7.36	5.47	9.52
233	7.43	5.53	9.60
234	7.50	5.59	9.68
235	7.57	5.66	9.77
236	7.64	5.72	9.85
237	7.72	5.78	9.93
238	7.79	5.85	10.01
239	7.86	5.91	10.10
240	7.94	5.98	10.18
241	8.01	6.04	10.27
242	8.09	6.10	10.35
243	8.16	6.17	10.43
244	8.24	6.24	10.52
245	8.31	6.30	10.61
246	8.39	6.37	10.69
247	8.47	6.43	10.78
248	8.54	6.50	10.86
249	8.62	6.57	10.95
250	8.70	6.64	11.04
251	8.78	6.70	11.13
252	8.85	6.77	11.21
253	8.93	6.84	11.30
254	9.01	6.91	11.39
255	9.09	6.98	11.48
256	9.17	7.05	11.57
257	9.25	7.12	11.66
258	9.33	7.19	11.75
259	9.41	7.26	11.84
260	9.49	7.33	11.93
261	9.57	7.40	12.02

FBDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
262	9.66	7.48	12.11
263	9.73	7.55	12.20
264	9.81	7.61	12.28
265	9.88	7.67	12.36
266	9.94	7.72	12.43
267	9.98	7.76	12.48
268	10.02	7.80	12.52
269	10.05	7.82	12.55
270	10.08	7.85	12.59
271	10.11	7.87	12.62
272	10.14	7.90	12.65
273	10.17	7.93	12.69
274	10.20	7.95	12.72
275	10.23	7.98	12.75
276	10.26	8.00	12.78
277	10.28	8.03	12.82
278	10.31	8.06	12.85
279	10.34	8.08	12.88
280	10.37	8.11	12.92
281	10.40	8.14	12.95
282	10.43	8.16	12.98
283	10.46	8.19	13.02
284	10.49	8.21	13.05
285	10.52	8.24	13.08
286	10.55	8.27	13.12
287	10.58	8.29	13.15
288	10.61	8.32	13.18
289	10.64	8.35	13.22
290	10.67	8.37	13.25
291	10.70	8.40	13.28
292	10.73	8.43	13.32
293	10.76	8.45	13.35
294	10.79	8.48	13.38
295	10.82	8.51	13.42
296	10.85	8.53	13.45
297	10.88	8.56	13.49
298	10.91	8.59	13.52
299	10.95	8.62	13.55
300	10.98	8.64	13.59
301	11.01	8.67	13.62

FBDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
302	11.04	8.70	13.66
303	11.07	8.72	13.69
304	11.10	8.75	13.72
305	11.13	8.78	13.76
306	11.16	8.81	13.79
307	11.19	8.83	13.83
308	11.22	8.86	13.86
309	11.25	8.89	13.90
310	11.28	8.92	13.93
311	11.31	8.94	13.96
312	11.35	8.97	14.00
313	11.38	9.00	14.03
314	11.41	9.03	14.07
315	11.44	9.05	14.10
316	11.47	9.08	14.14
317	11.50	9.11	14.17
318	11.53	9.14	14.21
319	11.56	9.16	14.24
320	11.59	9.19	14.28
321	11.63	9.22	14.31
322	11.66	9.25	14.35
323	11.69	9.28	14.38
324	11.72	9.30	14.42
325	11.75	9.33	14.45
326	11.78	9.36	14.49
327	11.82	9.39	14.52
328	11.85	9.42	14.56
329	11.88	9.45	14.59
330	11.91	9.47	14.63
331	11.94	9.50	14.66
332	11.97	9.53	14.70
333	12.01	9.56	14.73
334	12.04	9.59	14.77
335	12.07	9.62	14.80
336	12.10	9.65	14.84
337	12.13	9.67	14.87
338	12.17	9.70	14.91
339	12.20	9.73	14.94
340	12.23	9.76	14.98
341	12.26	9.79	15.02

FBDL (mm)	*Fit*	*Lower 95% PI*	*Upper 95% PI*
342	12.30	9.82	15.05
343	12.33	9.85	15.09
344	12.36	9.88	15.12
345	12.39	9.90	15.16
346	12.43	9.93	15.20
347	12.46	9.96	15.23
348	12.49	9.99	15.27
349	12.52	10.02	15.30
350	12.56	10.05	15.34
351	12.59	10.08	15.38
352	12.62	10.11	15.41
353	12.65	10.14	15.45
354	12.69	10.17	15.48
355	12.72	10.20	15.52
356	12.75	10.23	15.56
357	12.78	10.26	15.59
358	12.82	10.28	15.63
359	12.85	10.31	15.67
360	12.88	10.34	15.70

APPENDIX IX – ALL-MEASUREMENT MODEL: AGE ESTIMATION PREDICTION TABLE

Table A9 – The chronological age and measurements of individuals in the original sample used in the all-measurement model with the point estimate (fit) and 95% prediction interval (in years). The table is an example of prediction intervals associated with multivariate models.

Age (years)	TDL	TPB	TMSB	TDB	UDL	FDL	FDB	FMSB	RDL	HMSB	HDB	HDL	HPB	FBDL	Fit	Lower 95% PI	Upper 95% PI
1.47	127	32	13	26	106	152	47	14	95	13	26	121	27	127	2.02	1.23	3.09
1.69	131	35	12	24	102	155	41	13	90	12	33	117	26	126	1.61	0.94	2.54
2.20	145	37	13	26	112	185	47	14	99	12	30	135	26	143	2.99	1.94	4.36
2.35	156	38	13	27	120	183	47	16	108	12	32	135	27	154	2.67	1.70	3.94
2.50	150	40	13	26	125	186	46	14	110	11	33	140	24	152	3.03	1.97	4.41
2.54	147	39	11	24	114	173	47	12	99	10	30	125	22	144	2.45	1.54	3.66
2.59	150	36	13	24	115	185	40	14	101	12	29	133	21	150	2.78	1.79	4.09
2.62	152	38	13	25	117	184	46	14	108	12	32	136	25	147	2.69	1.72	3.97
2.71	154	40	15	28	119	189	50	16	106	12	33	139	27	154	3.11	2.03	4.51
2.82	159	44	14	30	126	189	50	16	112	15	36	140	28	158	3.02	1.97	4.40
2.84	152	42	15	29	116	188	53	14	106	15	35	139	31	150	3.13	2.05	4.53
2.85	141	38	11	24	109	176	44	13	97	11	30	130	22	137	2.54	1.61	3.77
2.89	170	45	14	30	127	202	55	15	116	13	37	148	28	167	3.33	2.20	4.79
3.00	158	45	14	30	120	196	54	15	110	14	38	141	28	160	3.09	2.02	4.49
3.00	146	37	13	26	106	179	46	15	93	12	31	123	24	139	2.37	1.48	3.55
3.00	174	41	14	26	135	213	49	16	119	14	35	160	25	176	4.04	2.74	5.69
3.02	154	43	15	29	118	192	49	15	105	12	34	143	27	154	3.54	2.36	5.06
3.07	155	39	13	26	124	195	46	13	109	12	33	142	26	155	3.11	2.03	4.51
3.10	166	43	12	29	126	203	48	16	110	13	35	148	26	161	3.70	2.48	5.27
3.27	159	42	15	28	127	190	48	15	114	13	35	141	26	159	2.99	1.95	4.36
3.33	173	46	15	29	124	210	55	17	111	13	38	146	28	171	3.63	2.43	5.17
3.33	187	45	14	30	138	215	55	16	123	15	33	157	31	178	4.26	2.91	5.97

Age (years)	TDL	TPB	TMSB	TDB	UDL	FDL	FDB	FMSB	RDL	HMSB	HDB	HDL	HPB	FBDL	Fit	Lower 95% PI	Upper 95% PI
3.35	178	45	14	31	139	221	56	16	125	12	39	154	27	177	4.03	2.73	5.68
3.56	166	44	14	29	119	198	53	15	103	14	34	140	28	165	3.30	2.17	4.75
3.56	157	44	14	28	121	190	53	15	110	14	37	151	31	154	2.81	1.81	4.13
3.72	180	44	16	29	140	217	53	16	126	14	37	162	28	177	3.89	2.63	5.51
3.75	189	46	17	32	140	224	54	16	126	13	35	163	27	185	4.37	3.00	6.10
3.77	139	38	13	27	111	176	45	15	98	13	28	126	27	137	2.61	1.66	3.86
3.92	175	45	15	28	133	216	55	17	120	14	33	153	28	174	4.54	3.13	6.32
3.92	174	50	15	32	135	213	59	16	123	13	36	151	28	172	4.09	2.78	5.75
4.01	167	42	13	26	133	203	50	15	120	13	35	147	29	163	3.29	2.17	4.74
4.08	181	49	14	32	143	224	58	16	126	15	39	162	28	181	4.14	2.82	5.81
4.09	176	43	14	27	139	209	49	15	122	13	33	153	25	174	3.78	2.54	5.36
4.12	164	43	13	28	129	204	50	15	116	13	35	145	24	163	3.58	2.39	5.11
4.17	189	45	16	32	140	222	54	18	124	14	38	156	28	188	3.76	2.53	5.34
4.22	165	47	16	31	123	194	56	18	110	14	34	146	29	162	3.16	2.07	4.57
4.29	180	49	17	33	139	224	58	17	126	15	39	167	29	174	4.11	2.79	5.77
4.40	185	52	15	32	137	223	60	17	124	13	36	149	31	184	4.34	2.98	6.07
4.53	193	48	14	30	147	241	56	15	133	13	37	176	28	192	5.36	3.77	7.33
4.54	176	46	15	30	133	215	54	17	121	14	38	158	31	176	3.78	2.54	5.37
4.54	200	47	16	31	154	243	55	17	138	14	40	176	28	200	5.09	3.56	7.00
4.55	177	46	14	27	138	222	51	14	125	13	32	160	28	183	4.90	3.41	6.76
4.55	213	48	17	35	152	270	57	18	134	12	40	185	30	212	6.15	4.40	8.30
4.57	215	54	18	37	151	272	66	21	134	15	40	188	33	207	6.38	4.59	8.58
4.60	176	44	14	28	135	214	50	16	120	12	35	152	27	172	4.01	2.72	5.65
4.65	186	47	15	31	142	228	57	16	125	13	37	162	30	188	4.40	3.03	6.15
4.67	190	46	16	29	142	240	55	18	132	14	37	175	30	192	5.72	4.06	7.78
4.88	189	49	15	31	142	235	57	16	128	14	35	169	31	187	5.28	3.71	7.23
4.99	208	52	17	37	158	263	63	19	142	15	41	186	34	210	6.22	4.46	8.38
5.00	218	48	17	32	154	262	51	18	147	13	40	188	31	215	6.10	4.37	8.24

Age (years)	TDL	TPB	TMSB	TDB	UDL	FDL	FDB	FMSB	RDL	HMSB	HDB	HDL	HPB	FBDL	Fit	Lower 95% PI	Upper 95% PI
5.11	206	48	17	31	153	252	56	17	137	15	39	179	30	208	5.66	4.02	7.71
5.24	186	43	16	30	133	229	53	17	120	14	34	162	29	185	4.45	3.06	6.21
5.26	192	41	13	27	147	243	51	16	132	12	37	179	28	192	4.86	3.38	6.72
5.36	223	53	17	36	164	259	62	18	146	16	44	189	35	219	5.92	4.22	8.03
5.40	202	51	17	34	157	253	61	18	135	17	39	183	37	201	5.89	4.20	7.99
5.45	207	48	15	31	145	250	56	16	130	13	37	171	30	201	5.41	3.82	7.40
5.59	226	56	19	34	158	276	67	20	142	15	43	191	36	221	6.56	4.73	8.80
5.67	243	52	18	35	180	291	62	18	165	17	43	209	32	238	8.04	5.94	10.59
5.80	221	51	15	34	159	273	59	16	146	12	39	187	32	216	6.49	4.68	8.72
5.92	207	49	18	33	154	252	55	18	141	18	37	173	32	207	6.17	4.42	8.33
6.02	224	50	18	36	162	268	63	18	144	17	42	193	33	217	6.10	4.36	8.24
6.06	210	52	18	37	159	265	59	18	144	17	41	184	33	210	6.23	4.47	8.40
6.12	230	49	19	33	154	269	58	20	139	14	41	193	32	226	5.77	4.10	7.84
6.13	218	54	18	36	163	281	61	19	149	15	41	189	33	215	7.00	5.09	9.34
6.23	198	48	17	32	152	251	56	18	139	16	40	177	36	198	5.41	3.82	7.39
6.29	216	53	18	37	162	269	63	18	146	16	41	185	35	221	6.34	4.55	8.53
6.42	226	51	16	36	161	284	62	17	147	16	42	197	36	230	7.03	5.11	9.37
6.47	236	47	16	32	159	284	58	18	144	13	41	187	30	233	7.06	5.14	9.41
6.49	230	52	17	33	169	261	62	19	153	15	39	189	34	230	6.46	4.66	8.68
6.54	234	51	18	37	163	284	61	21	148	13	39	200	32	231	7.38	5.40	9.80
6.68	235	50	15	32	165	279	61	18	152	15	41	204	34	235	7.02	5.10	9.36
6.70	230	51	17	35	166	278	61	19	155	18	42	203	36	226	6.91	5.01	9.22
6.71	220	54	19	35	168	279	62	19	152	16	41	195	36	223	7.01	5.09	9.34
6.77	254	52	18	38	177	299	62	20	159	14	43	214	34	250	8.06	5.95	10.62
6.78	208	51	17	34	155	252	60	18	139	16	42	177	31	210	5.15	3.61	7.07
6.84	250	54	21	36	170	308	60	20	153	15	42	209	32	250	8.46	6.28	11.10
6.92	234	49	17	36	173	286	58	20	156	17	40	207	32	232	7.41	5.42	9.84
6.98	232	55	19	34	172	276	63	19	155	15	42	200	34	235	6.87	4.99	9.18

Age (years)	TDL	TPB	TMSB	TDB	UDL	FDL	FDB	FMSB	RDL	HMSB	HDB	HDL	HPB	FBDL	Fit	Lower 95% PI	Upper 95% PI
6.99	257	54	22	38	179	296	62	20	165	17	45	201	33	255	8.00	5.91	10.55
7.02	229	54	18	35	158	287	62	18	143	15	40	190	33	226	7.16	5.22	9.53
7.10	253	48	17	35	177	298	56	19	161	15	42	210	30	249	7.93	5.84	10.46
7.11	242	50	16	35	169	283	55	17	149	13	40	200	30	237	6.91	5.02	9.23
7.25	213	45	15	28	159	272	53	15	146	12	36	191	27	214	6.80	4.93	9.10
7.39	223	50	16	34	158	275	60	19	143	14	40	200	33	221	6.37	4.58	8.57
7.42	231	55	20	37	167	283	68	19	149	16	40	203	35	233	7.11	5.18	9.47
7.49	245	50	18	35	178	306	58	18	165	15	46	214	33	246	8.73	6.50	11.42
7.50	207	50	18	32	154	255	57	17	140	13	36	181	29	213	6.16	4.42	8.32
7.55	248	60	20	40	175	302	69	21	157	17	43	218	39	244	8.26	6.12	10.86
7.63	227	43	14	29	161	278	51	15	146	11	36	186	27	223	6.88	5.00	9.20
7.69	236	54	18	39	173	304	65	19	158	15	40	215	32	238	8.50	6.31	11.15
7.74	244	57	18	38	171	289	65	20	157	16	44	211	34	242	7.49	5.49	9.93
7.79	225	52	17	36	168	280	63	17	153	15	37	194	32	220	7.56	5.55	10.02
7.84	222	48	16	34	154	271	59	18	144	13	41	187	29	215	5.83	4.15	7.91
7.93	219	48	18	36	160	276	59	18	145	14	36	197	33	219	7.21	5.26	9.59
7.95	231	50	17	37	167	274	59	18	146	16	41	192	33	229	6.48	4.67	8.71
7.98	290	60	20	41	206	336	68	21	186	16	45	231	41	287	9.90	7.46	12.81
8.00	271	53	19	40	184	314	60	20	167	15	41	217	34	271	8.76	6.52	11.45
8.11	257	53	16	36	193	311	61	20	172	14	39	219	34	256	8.93	6.66	11.66
8.13	265	58	21	40	178	312	65	21	163	18	45	225	39	265	8.84	6.59	11.56
8.14	235	48	17	34	168	298	61	18	152	13	42	210	32	232	7.70	5.66	10.18
8.14	229	56	18	37	165	290	64	20	150	14	43	207	34	230	7.22	5.27	9.61
8.16	254	55	18	36	181	301	65	19	161	16	44	213	36	256	8.03	5.93	10.58
8.23	240	57	21	41	172	290	67	21	155	17	45	205	38	240	7.55	5.54	10.00
8.38	265	54	18	40	183	318	66	21	170	16	44	222	39	264	9.01	6.73	11.75
8.40	259	58	19	38	178	316	67	21	167	15	42	217	36	256	9.02	6.74	11.77
8.44	259	54	18	38	181	306	63	19	168	15	43	226	37	254	8.44	6.26	11.07

Age (years)	TDL	TPB	TMSB	TDB	UDL	FDL	FDB	FMSB	RDL	HMSB	HDB	HDL	HPB	FBDL	Fit	Lower 95% PI	Upper 95% PI
8.46	264	57	19	41	181	318	65	22	161	16	42	225	36	262	8.84	6.59	11.55
8.48	204	47	17	32	157	255	59	19	139	13	38	184	28	210	5.78	4.11	7.85
8.51	247	52	18	36	178	299	63	20	162	16	42	215	32	246	7.85	5.78	10.37
8.51	247	56	18	37	173	289	65	20	156	16	44	209	35	244	7.36	5.38	9.77
8.51	264	56	19	37	185	324	67	20	171	17	36	227	39	261	10.93	8.32	14.04
8.52	231	53	18	35	161	280	59	19	147	15	39	194	33	229	7.03	5.11	9.37
8.54	283	61	20	40	198	329	73	21	179	17	48	233	37	280	9.73	7.33	12.62
8.54	175	51	16	31	144	239	56	16	128	14	35	181	31	175	5.01	3.50	6.91
8.63	245	57	17	37	182	306	64	19	162	17	45	221	39	245	8.29	6.14	10.89
8.65	254	54	19	37	174	299	63	19	163	15	42	205	31	252	8.01	5.91	10.56
8.70	253	57	18	37	181	300	65	20	165	16	45	215	36	253	8.07	5.96	10.62
8.72	254	56	19	37	179	315	64	20	162	16	47	230	34	249	8.90	6.64	11.62
8.75	261	55	19	37	188	310	62	19	175	16	45	223	36	260	8.63	6.42	11.30
8.88	255	55	19	37	188	316	65	21	171	17	46	222	37	254	8.79	6.55	11.50
9.09	227	50	18	34	172	287	58	18	160	14	39	202	35	232	7.73	5.68	10.22
9.11	270	54	20	39	194	329	64	22	181	19	46	233	41	266	9.47	7.11	12.31
9.19	291	58	20	41	199	322	66	23	180	17	48	242	37	280	9.19	6.88	11.97
9.21	264	55	18	37	188	313	61	19	174	16	41	220	35	268	8.78	6.54	11.48
9.25	274	63	23	42	204	326	75	24	185	18	48	240	36	274	9.35	7.01	12.16
9.41	299	65	22	47	217	351	78	25	198	22	54	251	41	298	11.11	8.47	14.25
9.43	260	53	15	37	178	307	61	18	164	14	41	216	34	258	8.40	6.23	11.02
9.43	280	54	17	38	195	335	64	20	178	18	42	246	36	279	9.54	7.17	12.38
9.50	265	63	22	40	189	316	74	21	170	16	47	216	37	265	8.90	6.64	11.63
9.61	281	60	23	41	196	352	66	24	180	16	48	248	38	276	10.74	8.17	13.81
9.68	289	57	21	40	201	339	65	23	184	15	44	233	34	286	9.71	7.31	12.59
9.69	268	55	19	37	188	340	64	20	176	15	42	231	35	263	9.98	7.53	12.91
9.78	255	60	22	40	184	306	70	22	166	20	50	208	37	250	8.52	6.33	11.17
10.04	278	58	19	40	191	333	68	21	173	18	50	235	40	278	9.86	7.43	12.77

Age (years)	TDL	TPB	TMSB	TDB	UDL	FDL	FDB	FMSB	RDL	HMSB	HDB	HDL	HPB	FBDL	Fit	Lower 95% PI	Upper 95% PI
10.04	258	57	20	36	180	321	67	19	165	16	44	213	37	259	9.10	6.80	11.86
10.04	260	51	17	34	176	313	60	18	165	16	41	215	37	258	8.66	6.44	11.34
10.10	276	59	21	44	198	328	72	23	178	18	50	223	39	275	9.50	7.13	12.33
10.16	280	65	24	43	195	342	73	24	176	18	52	243	43	273	10.48	7.95	13.50
10.16	244	59	21	41	177	299	69	21	162	17	47	208	38	241	8.05	5.94	10.60
10.17	279	70	24	44	197	332	82	23	185	18	41	232	42	274	9.35	7.01	12.15
10.26	265	62	21	44	188	324	74	23	179	19	47	227	42	272	9.40	7.05	12.21
10.56	294	61	24	43	211	361	69	22	193	19	48	251	39	296	11.14	8.50	14.29
10.60	285	65	22	46	211	352	70	23	189	18	53	242	38	287	10.73	8.16	13.80
10.79	298	56	21	38	206	347	68	22	190	14	44	250	36	297	9.96	7.52	12.89
10.86	278	65	20	45	202	350	75	22	187	15	51	246	37	279	10.82	8.23	13.90
10.94	292	56	20	44	183	369	69	20	167	16	45	240	39	287	11.95	9.17	15.23
11.04	293	58	19	41	202	342	69	22	185	16	41	243	38	293	9.62	7.24	12.48
11.07	277	59	19	45	180	333	68	20	167	16	39	238	41	270	9.81	7.39	12.70
11.09	262	60	22	40	185	331	69	21	170	18	47	220	37	266	9.66	7.27	12.53
11.26	302	60	21	40	203	376	70	25	187	16	44	253	38	302	11.52	8.81	14.72
11.68	259	55	19	41	182	323	65	19	170	16	44	221	36	251	9.14	6.84	11.91
11.69	309	63	23	44	201	380	69	23	181	18	44	251	39	304	11.78	9.04	15.04
11.76	293	65	23	44	198	354	73	20	189	20	53	252	41	296	11.35	8.67	14.53
12.10	300	60	20	42	211	353	70	21	192	16	50	257	39	297	10.45	7.92	13.47
12.21	343	72	23	49	229	393	82	24	210	17	48	274	46	338	12.75	9.85	16.17
12.21	329	66	22	47	224	387	75	24	208	19	50	264	38	322	12.45	9.59	15.82
12.36	290	62	20	46	203	358	74	24	183	18	50	237	40	288	10.84	8.24	13.92

The Research Ethics Committee, Faculty Health Sciences, University of Pretoria complies with ICH-GCP guidelines and has US Federal wide Assurance.

* **FWA** 00002567, Approved dd 22 May 2002 and Expires 13 Jan 2012.
* **IRB** 0000 2235 IORG0001762 Approved dd 13/04/2011 and Expires 13/04/2014.

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences Research Ethics Committee
Fakulteit Gesondheidswetenskappe Navorsingsetiekkomitee

DATE: 30/09/2011

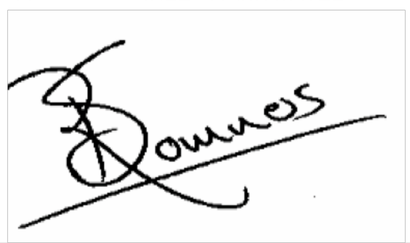
NUMBER	**191/2011**
TITLE OF THE PROTOCOL	An Osteometric Evaluation of Age and Sex Differences in the Long Bones of South African Children from the Western Cape
PRINCIPAL INVESTIGATOR	Mrs K E Stull **Dept:** Anatomy; University of Pretoria. **Cell:** 0799840465 **E-Mail:** Kstullster@gmail.com
SUB INVESTIGATOR	None
STUDY COORDINATOR	None
SUPERVISOR	Dr EN L'Abbè **E-Mail:** ericka.labbe@up.ac.za
STUDY DEGREE	PhD
SPONSOR COMPANY	None
MEETING DATE	28/09/2011

The Protocol was approved on 28/09/2011 by a properly constituted meeting of the Ethics Committee subject to the following conditions:

1. The approval is valid for a 2 year period [till the end of December 2013] , and
2. The approval is conditional on the receipt of 6 monthly written Progress Reports, and
3. The approval is conditional on the research being conducted as stipulated by the details of the documents submitted to and approved by the Committee. In the event that a need arises to change who the investigators are, the methods or any other aspect, such changes must be submitted as an Amendment for approval by the Committee.

Members of the Research Ethics Committee:

Prof M J Bester	(female)BSc (Chemistry and Biochemistry); BSc (Hons)(Biochemistry); MSc(Biochemistry); PhD (Medical Biochemistry)
Prof R Delport	(female)BA et Scien, B Curationis (Hons) (Intensive care Nursing), M Sc (Physiology), PhD (Medicine), M Ed Computer Assisted Education
Prof JA Ker	MBChB; MMed(Int); MD – Vice-Dean (ex officio)
Dr NK Likibi	MBB HM – Representing Gauteng Department of Health) MPH
Dr MP Mathebula	(female)Deputy CEO: Steve Biko Academic Hospital; MBCHB, PDM, HM
Prof A Nienaber	(female) BA(Hons)(Wits); LLB; LLM; LLD(UP); PhD; Dipl.Datametrics(UNISA) – Legal advisor
Mrs MC Nzeku	(female) BSc(NUL); MSc(Biochem)(UCL, UK) – Community representative
Prof L M Ntlhe	MbChB (Natal) FCS (SA)
Snr Sr J Phatoli	(female) BCur(Eet.A); BTec(Oncology Nursing Science) – Nursing representative
Dr R Reynders	MBChB (Prêt), FCPaed (CMSA) MRCPCH (Lon) Cert Med. Onc (CMSA)
Dr T Rossouw	(female) MBChB (cum laude); M.Phil (Applied Ethics) (cum laude), MPH (Biostatistics and Epidemiology (cum laude), D.Phil
Dr L Schoeman	(female) B.Pharm, BA(Hons)(Psych), PhD – Chairperson: Subcommittee for students' research
Mr Y Sikweyiya	MPH; SARETI Fellowship in Research Ethics; SARETI ERCTP; BSc(Health Promotion)Postgraduate Dip (Health Promotion) – Community representative
Dr R Sommers	(female) MBChB; MMed(Int); MPharmMed – **Deputy Chairperson**
Prof TJP Swart	BChD, MSc (Odont), MChD (Oral Path), PGCHE – School of Dentistry representative
Prof C W van Staden	MBChB; MMed (Psych); MD; FCPsych; FTCL; UPLM - **Chairperson**



DR R SOMMERS; MBChB; MMed(Int); MPharmMed.
Deputy Chairperson of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria
♦ Tel:012-3541330 ♦ Fax:012-3541367 / 0866515924 ♦ E-Mail: manda@med.up.ac.za
♦ Web: //www.healthethics-up.co.za ♦ H W Snyman Bld (South) Level 2-34 ♦ Private Bag x 323, Arcadia, Pta, S.A., 0007