Assessment of the associations of age with the classic skeletal sex indicators

Submitted by Sarah-Kelly Houston

A dissertation presented in publication format, submitted to the Department of Anatomy, School of Medicine, Faculty of Health Sciences, University of Pretoria, in fulfilment of the requirements for the degree MSc in Anatomy

13 February 2024

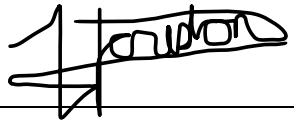
Supervisor: Ms. L Liebenberg

Co-supervisor: Prof. D Brits

Co-supervisor: Prof. J Myburgh

DECLARATION

I, Sarah-Kelly Houston, declare that the thesis (presented in publication format), which I hereby submit for the degree Master of Science (Anatomy) to the University of Pretoria, is my own work and has not previously been submitted by me for a degree at another university.



Sarah-Kelly Houston

2024/02/13

Date

ACKNOWLEDGEMENTS

This endeavour would not have been possible without everyone who supported, encouraged and believed in me. I am incredibly grateful for my supervision team, Ms. Leandi Liebenberg, Prof. Desiré Brits, and Prof. Jolandie Myburgh, thank you for all of the edits, advice, support, and coffee. Thank you for igniting my passion in this incredibly interesting field and helping me to always push myself to be the best I can be.

I would like to extend gratitude to the School of Anatomical Sciences at the University of the Witwatersrand and the curators of the Pretoria Bone Collection and the Raymond A Dart Collection of Modern Human Skeletons, Ms Gabi Krüger and Dr Brendan Billings for assisting me and granting me access to the collections and skeletal remains required for this study. Acknowledgement is also necessary to all those who donated their bodies to the furthering of research and teaching. Furthermore, I would like to acknowledge the University of Pretoria's Postgraduate Master's Research Bursary for providing funding for my tuition.

A special thanks to my fellow study friends: Stacey, Neil-John, and Callum. Our hour-long chats, coffees, and off-campus lunches where we laughed, cried, and complained really motivated me to finish my dissertation. To Alex, my parents, my family, my best friends, and housemates, thank you for standing by my side, having unwavering faith in me, and always having arms open to hug and a shoulder to cry on over the last two years. And finally, to all the fur babies in my life thanks for always being there for a cuddle or a game of fetch.

ABSTRACT

Forensic anthropologists are often required to establish a biological profile from skeletal remains, which includes the estimation of biological sex. The Walker and Phenice traits are morphoscopic methods frequently utilised to estimate sex from unknown skeletal remains. The methods employ five traits located on the skull and three on the os coxae and are considered the classic morphological skeletal sex indicators. Previous literature has stated that the growth and deterioration related to aging that occurs throughout an individual's life may affect the classic morphological skeletal sex indicators and therefore the estimation of sex. More specifically, females are reported to become more robust with older age, while younger males appear more gracile. This overlap in sexual dimorphism leads to greater misclassification and decreased accuracy in medicolegal casework. The aim of this study was to assess if age needs to be considered when attempting to classify sex using the classic morphological sex indicators.

The sex indicators were scored on a sample of 453 skulls and 429 os coxae of modern South Africans. The scores were then compared among ten different age cohorts. Separate age-specific classification models were created to assess if age affects the positive predictive performance of the sample. Significant differences were identified for the nuchal crest between age cohorts younger and older than forty years of age with the sample pooled. None of the other cranial traits demonstrated any significant differences. Regarding the classification models, accuracies increased for the younger group when separated from the older group, but the opposite was shown to be true for the older group. The os coxae did not show significant differences between the ten age cohorts. However, when separated into three large cohorts (younger, middle-aged, older) differences in accuracy was apparent between the cohorts. The results indicate that preselection of age prior to sex estimation is not necessary in a South African population. Further research is required to identify additional explanations for misclassification with sex estimation, as age does not appear to be a major influential factor for the classic morphoscopic sex indicators.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 3:

Chapter 4:

Chapter 5:

Appendix B:

Appendix D:

Appendix E:

Appendix F:

LIST OF FIGURES

Chapter 4:

Appendix A:

Appendix C:

Appendix D:

Appendix F:

LIST OF ABBREVIATIONS

Acronym	Meaning
BF	Black (South African) Female
BM	Black (South African) Male
CF	Coloured (South African) Female
CM	Coloured (South African) Male
DFA	Discriminant Function Analysis
FD3	FORDISC 3.1
GH	Growth Hormone
Gla	Glabella
GMM	Geometric Morphometrics
GSN	Greater sciatic notch
LR	Logistic Regression
Ma	Mastoid process
MDCT	Multi-detector Computed Tomography
Me	Mental eminence
MIR	Medial aspect of the ischiopubic ramus
Nu	Nuchal crest
Or	Supra-orbital margin
PCC	Percent Correctly Classified
RFM	Random Forest Model
SA	South African
SES	Socioeconomic Status
SPC	Sub-pubic contour
WF	White (South African) Female

WM	White (South African) Male
VA	Ventral arc

CHAPTER 1: INTRODUCTION

Forensic anthropologists are required to estimate the biological profile from unknown skeletal remains in the interest of finding a connection to national missing persons records. The parameters of the biological profile include estimates of sex, population affinity, age-at-death, and stature. Each parameter of the biological profile is important as it contributes to the identification of unknown skeletal remains and are estimated using numerous techniques and methods.

Sex estimation is considered one of the most important aspects of the biological profile, as accurate sex estimation has been shown to decrease the list of possible matches to the missing persons list by 50% (Krüger *et al*., 2015). Additionally, the other parameters of the biological profile are reliant on the correct estimation and preselection of sex in pursuance of high accuracy age and stature estimation. A myriad of sex estimation methods exists, with morphoscopic methods arguably being the most popular. Currently, two major morphoscopic methods are employed to estimate sex using the skull and os coxae, which are considered the classic morphological sex indicators (Walker, 2008; Klales *et al*., 2012).

The majority of anthropological methods for sex estimation originated in other countries across the globe and were then validated on local skeletal collections to ascertain whether the methods were viable for use on the South African population. Where necessary, the methods were altered to be population-specific, and different classification formulae were created in favour of increased accuracy. In particular, the Walker (2008) method, which applies traits from the skull, was validated on the South African population and recalibrated by Krüger and colleagues (2015) to increase accuracy when applied on South Africans. The recalibration by Krüger and colleagues (2015) requires the estimation and preselection of population affinity from the unknown skeletal remains so that the relevant standards can be applied for the estimation of sex. This is important as the various population groups in South Africa display varying levels of sexual dimorphism. Sexual dimorphism is typically expressed as size and/or shape differences between males and females. Similarly, the Phenice (1969) traits found on the pubic bone of the os coxa were first codified and paired with simple line drawings and statistics by Klales and colleagues (2012). The updated methodology and definitions were then validated on numerous diverse global populations and recalibrated by Kenyhercz and colleagues (2017). South African individuals were included in the global sample which resulted in increased accuracies when applying the recalibrated formulae in South Africa. For simplification, the

skull and pubic traits will hereafter be referred to as the Walker (2008) traits and Phenice (1969) traits, respectively.

Numerous factors have been noted to influence the expression of sexual dimorphism, and thus the accuracy of sex estimates (Cabo *et al.*, 2012). Some examples include the population affinity of the individual, hormonal influences during the individual's life, pregnancy (and number of pregnancies), robust muscle attachments due to physical activity, as well as the age of the individual (Albert *et al*., 2007).

In terms of the aging process, previous studies have observed changes in the craniofacial complex that occur throughout adulthood, such as an increase in the anterior facial height, increased mandibular length, and horizontal expansion of the craniofacial skeleton (Walker, 1995; Albert *et al*., 2007; Krüger *et al*., 2015). Moreover, Walker (1995) reported that older females tend to display more masculine traits as compared to younger females. This was supported by Krüger and colleagues (2015), who found that older South African females tended to misclassify more frequently, leading to lower classification accuracies. Walker (1995) also noted that younger males tend to display more feminine traits in comparison to older males. Similarly, age-related changes have been observed in the pelvis, with Waltenberger and colleagues (2022) stating that the shape of the pelvis changes throughout an adult's lifetime but that the changes are expressed more in females than in males. This was supported by Auerbach *et al.* (2017) who reported that younger females display narrower pelvic inlets than older females. Despite documented changes, the current standards for morphoscopic sex estimation do not consider the potential implications of age on the positive predictive performance of the methods.

The age changes are a matter of concern given the demographics of skeletal collections as well as the increasing age-at-death of the living population. The increasing age-at death in the living population may allow for continued, unaccounted for changes to the skeleton. The above methods were validated on the various available bone collections that exist in South Africa (such as the Pretoria Bone Collection, Raymond A. Dart Collection of Modern Human Skeletons, and the Kirsten Skeletal Collection), all of which largely consist of older individuals who were unclaimed or donated their bodies to the collections (L'Abbé *et al*., 2005; Dayal *et al.*, 2009; Alblas *et al.*, 2018). Although the collections are ultimately made up of mostly older individuals, this is particularly true for white South Africans, for whom the mean age-at-death is much greater compared to the mean age-at-death of the black South Africans (L'Abbé *et al*., 2005). The discrepancy in the mean age-at-death between groups is likely due to the differences

in socioeconomic status (SES) and, therefore, life-expectancy as well as different cultural beliefs of the different groups that make up the South African population (L'Abbé *et al*., 2021). Subsequently, the mean age-at-death of the individuals in the skeletal collections may differ from the mean age-at-death of individuals seen in forensic casework, more specifically older individuals in the collections versus younger individuals in casework. It is imperative to further explore potential skeletal differences that may exist as a result of age and assess how these differences influence our ability to estimate sex.

While previous studies have noted significant skeletal differences between varying age cohorts (Rogers, 1982; Walker, 1995; van Wyk and van Wyk, 2004; Albert *et al*., 2007; Ross and Williams, 2010; Gapert *et al.*, 2013; Small *et al.*, 2016; Waltenberger *et al*., 2021), few studies have looked specifically at how these differences will influence morphoscopic methods. Therefore, the current study focused on the application of scores of the Walker (2008) and Phenice (1969) traits to South African individuals across various adult age cohorts to ascertain if and to what degree age affects the classification accuracy of the sex estimate. In order to achieve this, the objectives of the current study included testing the inter- and intra-observer repeatability of the traits to test if the traits were scored consistently by the principal investigator; the comparison of the frequency distribution of each of the five Walker (2008) traits found on the skull and the three Phenice (1969) traits found on the os coxa among the ten age cohorts to test for significant differences attributable to age; and the analysis and comparison of the obtained accuracies when the existing South African formulae for the Walker (2008) and Phenice (1969) traits were used as well as when additional statistical analyses were implemented

CHAPTER 2: LITERATURE REVIEW

2.1 Sex estimation and the classic sex indicators

Sex estimation quantifies any sexual dimorphism evident on the skeleton; this is typically expressed as size and/or shape differences between males and females. Sexual dimorphism is present on the skeleton due to hormonal differences (oestrogen versus testosterone), functional reproductive differences (i.e. females being able to give birth), as well as differences in muscle robusticity (Cox *et al.*, 2000). The goal of sex estimation from skeletal remains is to assess whether an individual was a biological male or female. While a spectrum of genders exists (Tallman *et al.*, 2021; Flaherty *et al.*, 2023) and an individual may personally identify as either one of the sexes or even neither, skeletal morphology will express the phenotype determined by genetic material (Krüger *et al.*, 2015). The sex of the individual is defined as anatomical or chromosomal and is determined by the sex chromosomes (XX or XY) prior to birth. Gender on the other hand is considered a social construct that is expressed by the individual and is not always aligned with biological sex and can generally not be estimated from the skeleton. Unfortunately, forensic anthropologists are still not able to determine the sex of an individual with 100% certainty and rely on the metric and morphoscopic evaluation of the skeleton to produce a sex estimate. The individual can be estimated to be male or female with certain probabilities, but the possibility of an error can never be completely excluded.

South African biological and forensic anthropologists currently use three main methods to facilitate the estimation of sex from skeletal remains, which includes one metric method and two morphoscopic methods. Metric methods measure the distances between various landmarks as well as maximum and minimum diameters of bones, and generally achieve greater levels of repeatability when compared to morphoscopic methods as there is minimal subjectivity involved (Dirkmaat *et al.*, 2012). Morphoscopic methods involve visual assessment of the degree of expression as well as the absence or presence of certain traits which is assigned a numerical value. Many forensic practitioners prefer non-metric or morphoscopic techniques as they do not require any equipment and are relatively quick and straight-forward. Although morphoscopic methods have been critiqued for its reliance on observer experience, the methodology has been made more repeatable and accurate through the introduction of line drawings and statistics. For both morphoscopic methods commonly used for sex estimation the traits are scored on an ordinal scale from 1 to 5, with 1 depicting more feminine traits and 5

depicting more masculine traits. (Phenice, 1969; Walker, 2008; Klales *et al*., 2012; Krüger *et al.*, 2015; Kenyhercz *et al.*, 2017).

The pelvis is known to be the most sexually dimorphic skeletal element due to its direct involvement in human reproduction (Phenice, 1969; Stewart, 1979; Krogman and İşcan, 1986; Spradley and Jantz, 2011; Betti, 2014). Despite this, metric assessment of the pelvis is difficult, tedious and unfavoured for sex estimation as the definitions are confusing and non-specific (Steyn *et al*., 2012). The cranium is the most commonly retrieved element in forensic cases and is often the only skeletal element available for analysis and as such is still frequently used for sex estimation even though long bones have shown higher sex estimation accuracies (Spradley and Jantz, 2011; Krüger *et al.*, 2015). The current study focused on the assessment of the two morphoscopic methods most frequently used for sex estimation in South Africa which utilise the Walker (2008) traits on the skull and the Phenice (1969) traits on the os coxa.

The Walker (2008) traits include the nuchal crest, mastoid process, supra-orbital margin, and glabella on the cranium, as well as the mental eminence on the mandible. The method was originally proposed and developed in North America by Acsádi and Nemeskéri (1970) and entailed the morphological analysis of the above traits and scoring them from hyperfeminine (score of -2) to hypermasculine (score of 2), with a score of zero demonstrating intermediate sexual features or androgyny. The initial method was subsequently improved by Walker (2008) by adding statistics and simple line drawings to improve repeatability. Walker (2008) also altered the scoring system to score the traits from 1 to 5. When the Walker (2008) method was applied to a South African sample, low accuracies were noted, indicating differences in cranial morphology between South Africans and North Americans. South Africans demonstrated lower levels of sexual dimorphism and tend to produce more intermediate scores when using the original formulae (Krüger *et al.,* 2015). In 2015, Krüger and colleagues recalibrated the regression formulae to account for the variation that is present in the South African population. Different combinations of the Walker (2008) traits were noted to provide higher accuracy levels for each population group and therefore the combinations with the highest accuracy levels for each population group were incorporated into the logistic regression formulae (Krüger *et al.,* 2015). The Krüger *et al*. (2015) method displays cross validated accuracies ranging from 73% to 93% depending on the population affinity and the combination of variables employed. As a result, South African forensic anthropologists are now able to estimate the sex of unknown skeletal remains with high accuracy using the skull with adapted formulae (Krüger *et al.,* 2015). The recalibration of the Walker (2008) method for the South African population was imperative

as the degree of sexual dimorphism also varies depending on the population and population groups, necessitating the development and validation of population and population affinity-specific standards (Krüger *et al.,* 2015; Caple and Stephan, 2017).

The Phenice (1969) traits include the medial aspect of the ischiopubic ramus, the subpubic concavity, and the ventral arc located on the pubic bone of the os coxa. Although first described by Phenice (1969), Klales and colleagues (2012) were responsible for the codification and creation of line drawings (Figure A2 in APPENDIX A**: SCORING SHEETS)** to create a comparable method to the one published by Walker (2008). The modified methodology utilises the three traits in order to estimate sex with greater repeatability and published accuracy rates, which was lacking with Phenice's description of the traits. In 2017, a global recalibration was conducted to modify the Klales *et al.* (2012) standards, making it possible to successfully apply the method to numerous different populations, including South Africans (Kenyhercz *et al*., 2017). A global recalibration entails the validation of methodology, which was previously based on a single population, on numerous diverse geographic populations. Following the validation, the original equations are recalibrated to improve low resultant accuracies identified in validation studies on other populations. With the prior knowledge of population affinity, the recalibrated equations can then be applied to pooled diverse population groups with high accuracy levels (Kenyhercz *et al*., 2017). The Kenyhercz *et al*. (2017) method demonstrated a cross validated accuracy of 95.9%.

2.2 Factors affecting sex estimation

Numerous factors can influence the robusticity and thickness of bones and the expression of the classic sex indicators. As a result, sex estimation and the accuracy with which sex can be estimated may be affected. Influential factors include population affinity, hormonal influences, physical activity, pathological conditions/diseases and age.

2.2.1 Population affinity

Differences are known to exist between males and females, but the extent of these differences is dependent on the population the individual originates from (Krüger *et al.*, 2015). While the general trend of sexual dimorphism is that males tend to be more robust than females and females tend to be more gracile than males; this does not exclude the fact that males from one population can be equally or even more gracile than females from another population (Dirkmaat *et al*., 2012). For example, if we take into consideration the comparison between the

Korean and Māori (New Zealand) populations. Korean males and females tend to be smaller and more gracile than Māori males and females, who tend to be larger and more robust. Sex estimation of Korean individuals using Māori standards would most likely result in the misclassification of Korean males as females due to their gracile features (which are more similar to Māori females' features) when compared to Māori males. Similarly, sex estimation of Māori individuals using Korean standards would most likely result in the misclassification of Māori females as males due to their larger and more robust features (L'Abbé *et al*., 2013). Additionally, research by Krüger *et al.* (2017) found that South Africans tend to be less sexually dimorphic than other populations particularly when comparing the skull.

Variation in sexual dimorphism is present not only among global populations, but also among different South African population groups. For example, white South Africans have been shown to be more sexually dimorphic than black South Africans when analysing the cranial traits (Krüger *et al.,* 2015). When exploring the Walker (2008) traits, Krüger and colleagues (2015) noted that white South African males more frequently tend to produce scores of 5, black South African females tend to produce scores of 1, and both black South African males and white South African females tend to produce more intermediate scores. The study also found that greater accuracies were obtained for white South Africans when the glabella and nuchal crest were used in conjunction and for black South Africans when the glabella, mastoid process and supra-orbital margin were used in conjunction (Krüger *et al.,* 2015). The study indicated the overlap of males and females between different populations, and how different combinations of traits work better to classify sex for the different populations (Krüger *et al.,* 2015). Thus, certain sex estimation methods, such as using the Walker (2008) traits in South African individuals as well as metric methods, such as metric methods using FD3, rely on the population affinity to be preselected to increase the accuracy of the sex estimate (Krüger *et al.,* 2015).

2.2.2 Hormonal influences and the implications of physical activity

One of the main drivers of sexual dimorphism in human adults is differences in hormone production (i.e., testosterone and oestrogen) during puberty. The male sex hormone, testosterone, promotes the activity and longevity of osteoblasts (cells that deposit bone), and promotes periosteal bone apposition. Both testosterone and oestrogen work collectively in males to promote periosteal bone apposition and cortical bone growth during puberty allowing men to build wider and stronger bones with higher resistance to bending (Callewaert *et al.,*

2010). This allows males to have a greater periosteal diameter than females from mid-puberty (Veldhuis *et al.*, 2005; Callewaert *et al.,* 2010). The female sex hormone, oestrogen, has a multitude of effects on bone growth depending on its concentration: low concentrations of oestrogen promote the secretion of growth hormone which stimulates chondrocyte growth in the proliferation zone of the growth plate. High concentrations of oestrogen, however, stimulates osteoclast invasion of the growth plate and induces apoptosis of the chondrocytes which collectively halts growth (Juul, 2001). Stimulation of endosteal and trabecular bone formation, promotion of epiphyseal maturation and an increase in tensile bone strength are included in the function of oestrogen (Veldhuis *et al.*, 2005). Inevitably oestrogen stimulates endocortical apposition but limits bone expansion in females (Callewaert *et al.,* 2010). The sex hormones are known to fluctuate with aging, as is evident with menopause, thus the fluctuation of hormone levels with increasing age will have an influence on bone structure and skeletal trait expression (Veldhuis *et al.*, 2005; Callewaert *et al.,* 2010; Krüger *et al.,* 2015).

Similarly, physical activity can have implications on the skeletal muscles and the muscle attachment sites. More robust muscle attachments may result from overworking of the muscles and pressure on the muscle tendons in individuals who do a lot of physical labour in their day to day lives and/or individuals who over exercise and apply a lot of pressure on their muscles, tendons, and the insertion points (Cabo *et al.*, 2012). Hypertrophied muscles result in the production of additional bone in the form of spurs, eminences or bony thickenings which may influence the sex estimate. Thus, due to their increased physical activity levels males tend to exhibit more rugose muscle attachments (Greene, 2022).

2.2.3 Pathological conditions/diseases

There are numerous conditions/diseases that alter bone density and robusticity that should be considered. These include, but are not limited to, diseases such as osteoporosis and -petrosis, endocrine disorders such as hyperparathyroidism, hyperthyroidism, acromegaly and pituitary gigantism and metabolic disorders such as osteomalacia and scurvy that alter bone density, shape and size (Buikstra, 2019). Although most conditions are not frequently encountered, one should be knowledgeable about its potential occurrence.

2.2.4 Age-at-death and the effects on the classic sex indictors

The fundamental principle that allows the estimation of age from skeletal remains is the growth and subsequent deterioration of the human skeleton throughout an individual's life (Garvin *et*

al., 2012). The early years (from infancy to adolescence) involves the constant growth of the skeletal system based on a relatively predictable pattern which simplifies the estimation of age in younger individuals. For example, certain elements are present or absent and/or fused or unfused at certain stages of growth. Later in life, the skeletal system begins to deteriorate due to the movement and pressure on the skeletal elements over time as well as changes in hormone levels and/or the decline in nutrients and osteoblast production. Age estimation in adults is based on the deterioration of the skeletal elements due to these factors (Garvin *et al*., 2012).

The individual cranial skeletal elements grow and change at different rates during the postnatal developmental years, and it is likely that subtle morphological changes may continue to occur throughout adulthood in response to further growth and environmental factors (Ross and Williams, 2010). Certain age-related changes have been exhibited on the skull and these changes are present in numerous different forms. The changes may be observed as tooth loss leading to alveolar bone resorption, rougher granulation at muscle attachment sites, and weakening of the masticatory muscles leading to vault structure changes, as well as thickening or thinning of the cranial bones (Rogers, 1982).

Walker (1995) suggested that males younger than 30 years of age tend to have more feminine cranial features, such as sharper supra-orbital margins. In the same study, postmenopausal women were observed to appear more masculine in the same cranial regions indicating that there is continual change in the cranium throughout adult life (Walker, 1995). Additionally, Albert *et al*. (2007) noted that the greatest morphological modifications in the head, neck and face occur around 50 years of age in both males and females. Gapert and colleagues (2013) studied the effects of age on the metric sexual dimorphism of the foramen magnum on The Saint Bride's collection in London. Their findings indicated that the foramen magnum is sexually dimorphic, and that the accuracy of sex estimation is not affected by age when the sample was distributed into two groups (<50 and >50 years of age). However, an increase in sexual dimorphism was clearly visible in the >50 years group (Gapert *et al.*, 2013).

Ross and Williams (2010) conducted a geometric morphometric evaluation of 24 individuals ranging from 14 to 40 years of age and noted that significant shape differences were only present between the individuals aged 20 to 25 years. Significant size differences, however, were found between all the younger age groups and the 25-year-old age group (Ross and Williams, 2010). Furthermore, they observed significant differences between the 25-to-40-year age group and the younger age groups indicating that an expansion in cranial base, facial breadth and facial width occurs with increasing age in adults (Ross and Williams, 2010). The cause for

these changes is predicted to be multifactorial following an increase in the size of the paranasal sinuses and eruption of the third molars and/or alveolar remodelling (Ross and Williams, 2010). Additionally, the findings also suggested that individuals reach cranio-facial maturity from 14 years of age and that adult standards and methods can be applied on adolescents from 14 years and older (Ross and Williams, 2010).

Studies have also found dramatic changes in the shape and size of the facial region with age and subsequent edentulism (Small *et al.*, 2016) which may affect not only metric methods of sex estimation but also morphoscopic methods. In a survey of South African individuals, van Wyk and van Wyk (2004) noted that 12.6% of individuals between the ages of 35 and 44 years of age were completely edentulous (van Wyk and van Wyk, 2004). Coloured South Africans were identified to exhibit the largest percentage of edentulism (51.6%), and it was shown to be more common in female individuals than male individuals across population groups (van Wyk and van Wyk, 2004).

Alterations in the number of teeth present in the mouth as well as tooth loss directly affects the shape of the face (Ross and Williams, 2010). Small and colleagues (2016) noted that following tooth loss and edentulism, the maxillary and mandibular alveolar bone was significantly affected, and that this loss of dentition resulted in an impact to the upper facial height and the shape of the palate. The results further indicated that edentulism and tooth loss may influence important measurements and morphoscopic traits used for sex and population affinity estimation (Small *et al.*, 2016). Their study on white South Africans showed that edentulous individuals displayed inferior and anterior flexion of the basicranium resulting in inferior movement of the external occipital protuberance (nuchal crest) and anterior and inferior movement of the mastoid processes. Edentate individuals were also identified to display a marginally larger basicranium (Small *et al.*, 2016). Additionally, tooth loss will inevitably result in the loss (or loss of prominence) of the mental eminence due to mandibular alveolar bone resorption.

All of the above factors demonstrate how the cranial elements continue to change in shape and size throughout adulthood, possibly affecting surrounding structures. While these factors have been widely discussed, the potential effect on the skeletal elements utilised in the estimation of sex from the skull, more specifically Walker (2008) traits has not been established.

As a central element of the human skeleton, the shape and size of the pelvis is directly affected by biomechanical forces that act on the acetabula from the legs and on the sacroiliac joint from the spine (Waltenberger *et al*., 2021). Several studies have found that despite the cessation of overall growth, remodelling occurs, and certain pelvic dimensions increase into adulthood (Tague, 1994; Berger *et al*., 2011; Huseynov *et al*., 2016; Mitteroecker and Fischer, 2016; Sharma *et al*., 2016; Auerbach *et al*., 2018).

Pelvic shape is known to change throughout an individual's life, with greater changes reported for females than males. The changes in shape are estimated to be due to endocrine factors (the induction of hormones) as well as mechanical factors, such as parturition, which explains the greater magnitude of change seen in females (Waltenberger *et al*., 2021). Waltenberger and colleagues (2021) reported that females displayed three phases of change, namely: (1) during the parturition period, (2) pre-menopause and post-parturition, and (3) post-menopause, whereas males were identified to have only two phases of change, with the split occurring around 45 years of age. The same study also identified that the first phase of change seen in females was mainly due to childbirth and the degree of change increased proportionally to in the number of times a woman has given birth (Waltenberger *et al*., 2021). Thus, the increased weight and pressure of the uterus and foetus resting on the pelvic girdle has a direct effect on the underlying bone. Additionally, the increased oestrogen during pregnancy and the premenopausal phase likely results in oestrogen-induced bone remodelling which explains the widening of the pelvic inlet during the first phase. Pregnancy and parturition not only alter the shape and size of the pelvis but may also cause damage to the pelvis, especially the pubic symphysis. Although not necessarily a common occurrence, Waltenberger and colleagues (2021) found pelvic fractures were still observed to increase in females until 40 years of age and were identified to correlate to the process of birth. The obstetric dilemma is well known in humans as a difficult birthing process which consists of long and painful labours (Wells *et al*., 2012). The hormones produced during late pregnancy allow certain pelvic ligaments to relax during the birthing process. However, damage to or tearing of these ligaments has been reported leading to the subsequent deterioration of the pubic symphysis and ventral arc (Wells *et al*., 2012; Waltenberger *et al*., 2021).

The postmenopausal phase of females was shown to parallel the secondary phases seen in males. In males the two phases of change are relatively similar and mostly differ in magnitude of change. In the first phase, which occurs from 20 to 40 years of age, the pelvic inlet becomes more heart-shaped, the greater sciatic notch becomes narrower and the subpubic angle and iliac

height decreases (Waltenberger *et al*., 2021). The second phase occurs from 45 years of age and mirrors the first phase, except the anterior-posterior diameter as well as the coxal height decreases to a greater degree (Waltenberger *et al*., 2021). The alteration of shape and size in these later phases, seen in both males and females, are correlated to mechanical factors, such as continued strain, ultimately resulting in bone loss. This loss can be amplified even further in individuals with vitamin and mineral deficiencies (Waltenberger *et al*., 2021). While dimensional changes in the pelvis have been widely discussed in the literature, no existing published research that has explored these changes in relation to morphoscopic methods or the Phenice (1969) traits and how it may affect sex estimation accuracies has been identified.

CHAPTER 3: MATERIALS AND METHODS

3.1 Sample

The sample consisted of 712 individuals (453 skulls and 439 left os coxae), inclusive of black, white, and coloured South African males and females. The specimens were obtained from the Pretoria Bone Collection (University of Pretoria) as well as the Raymond A Dart Collection of Modern Human Skeletons, commonly known as the Dart Collection (University of the Witwatersrand). Ethical approval was obtained (Reference number 588/2022) from the University of Pretoria Ethics Committee and the University of the Witwatersrand Ethics Committee (APPENDIX C**: ETHICAL APPROVAL**; Figure C1).

3.2 Methodology

The age for the study sample ranged from 14 to 108 years of age. The lower end of the age range was selected based on findings of previous studies. Corron and colleagues (2021) studied the maturation scores of immature subadult os coxae and identified that at around 13-14 years of age subadults have a maturity score of around 15 or higher. A maturation score of 15 involves the fusion of particular epiphyses located on the os coxa as well as partial or complete fusion of the ischial tuberosity and, at minimum, the presence of the anterior-superior iliac spine if unfused (Corron *et al.*, 2021). These results indicate that pelvic adult sex estimation methodologies can be used on subadults around 13-14 years of age. Ross and Williams (2012) also reported that the cranium reaches adult maturity around 14 years of age. Furthermore, Klales and Burns (2017) stated that sex could accurately be estimated in subadults in the early adolescent cohort (12.6-15.5 years) using adult standards. No upper age limit was selected. Any individuals with visible pathology, ante-mortem trauma, and/or post-mortem damage that affected the accurate scoring of traits were excluded. Individuals with edentulism (loss of six or less teeth in one quadrant) were included in the sample as this is a contributing factor to age-related changes and a large number of individuals in the sample (and in the South African population) demonstrate ante-mortem tooth loss and edentulism. The sample was subdivided into ten age cohorts for comparative purposes to test whether significant differences existed in the traits between the cohorts (Table 3.1). The current study made use of convenience sampling (i.e., the inclusion of individuals into the sample was non-probabilistic and was rather based on accessibility and availability) and therefore the number of individuals per cohort as well as the demographic distribution within each cohort varies as a result of availability of specimens that met the inclusion criteria. The resultant sample distribution of the study contains an uneven

distribution of individuals throughout the age cohorts which is due to the constitution of the skeletal collections and the exclusion and inclusion criteria of the study.

Table 3.1 Sex distribution and population affinity distribution of the sample.

Age Cohort	Females	Males	BSA	WSA	CSA	*N*
<20	5	4	6	1	2	9
20-29	38	40	70	3	5	78
30-39	40	50	86	1	3	90
40-49	47	55	88	12	2	102
50-59	40	59	79	19	1	99
60-69	48	57	70	34	1	105
70-79	50	59	55	54	0	109
80-89	45	51	25	71	0	96
90-99	9	13	5	17	0	22
>99	0	2	1	0	1	2
TOTAL	**322**	**390**	**485**	**212**	**15**	**712**

Only the left os coxa and left side of the cranium was analysed and scored. This is also in line with common practice in forensic anthropology case analyses (Langley *et al.*, 2016). Previous studies have shown that differences may exist between the left and right os coxa and left and right sides of the cranium (Cole *et al.*, 2020; Houston, 2021). However, it was beyond the scope of the current study to assess and establish whether asymmetry between the left and right sides may influence the accuracy of the sex estimate when the individuals are divided into age cohorts.

The five Walker (2008) traits on the skull and three Phenice (1969) traits on the os coxa were visually assessed and scored for each of the available specimens in the sample that met the inclusion criteria (Table 3.2). Each trait was scored using the definitions and line drawings as by Walker (2008) and Klales *et al.* (2012). The line drawings and definitions can be seen in APPENDIX A**: SCORING SHEETS** and APPENDIX **B: TRAIT DEFINITIONS**, Figure A1 and Figure **A2** and Table B1 respectively. The allotted score for each trait was entered into the data collection sheet and edentulism, if present, was noted for future reference.

Table 3.2 List of sex estimation traits, the abbreviations and scoring range.

	Traits	Abbreviation	Scoring*
Walker (2008) Traits (skull)			
1	Glabella	Gla	1-5
2	Mastoid Process	Ma	1-5
3	Mental Eminence	Me	1-5
4	Nuchal Crest	Nu	1-5
5	Supra-orbital Margin	Or	1-5
Phenice (1969) Traits (os coxa)			
1	Sub-pubic Contour	SPC	1-5
2	Medial Aspect of the Ischiopubic Ramus	MIR	1-5
3	Ventral Arc	VA	1-5

*A score of 1 demonstrates a feminine expression, and a score of 5 a masculine expression.

3.3 Statistical analyses

Inter- and intra-observer error was tested to assess the repeatability of the traits. The Walker (2008) and Phenice (1969) traits were scored on 10 randomly selected individuals from the sample by the principal investigator on separate occasions and a second observer. A series of analyses was completed in order to test the inter-and intra-observer error. Additionally, it was important to quantify whether differences existed due to score differences between observations or due to age-related changes. The repeatability of the traits was assessed with Cohen's Kappa. The Landis and Koch scale was then used to better describe the level of agreement between observers (Table 3.3).

Table 3.3 Landis and Koch scale (Landis and Koch, 1977).

Kappa statistic	Strength of Agreement
0.00	Agreement equivalent to chance
0.01-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-0.99	Near perfect agreement
1.00	Perfect agreement

Given the uneven distribution of the sample, a Kruskal-Wallis test was conducted in combination with a *post-hoc* Dunn's test to establish if the scores assigned for each trait differed among the age cohorts. The Kruskal-Wallis test is a non-parametric test used to compare ordinal data (trait scores) for two or more categorical groups (age cohorts). The Dunn's test is an extension that is applied to further interpret the results and was calculated with a Bonferroni correction. The Kruskal-Wallis test is an omnibus technique that is only capable of indicating if there are any significant differences when comparing multiple groups, while the Dunn's test indicates more specifically which groups differ from one another. The Bonferroni correction is used in conjunction with the Dunn's test to avoid the incremental risk of a type 1 error when conducting a *post-hoc* analysis for multiple comparisons.

Preprocessing of the data as well as normality testing was completed prior to statistical analyses. Thereafter, logistic regression (LR) and discriminant function analysis (DFA) formulae were used to estimate the sex for each individual and the percent correct was calculated (e.g., if the individual was recorded as female and the formula correctly classified the individual as a female). LR and DFA were selected as it corresponds to the Walker (2008) and Klales *et al.* (2012) studies which is important for comparative analyses with these studies. This was first done for the entire sample with all of the age cohorts pooled together in order to get the accuracies with which the sample is classifying irrespective of age. Next, the results of the Kruskal Wallis test were used to modify the formulae for both the Walker (2008) and Phenice (1969) traits. More specifically, if significant differences were noted between cohorts, the cohorts that differed were separated to create reference groups based on age to see if the accuracy increased or decreased when estimating sex for the different age groups and how this accuracy compares to that obtained with the pooled sample. Finally, Random Forest models

(RFMs) were also implemented to further test the accuracy, since RFM has been recommended as a more suitable machine learning method that is not restricted by statistical assumptions (Klales, 2021). An interactive program called MorphoPASSE (Klales, 2020) which utilises RFMs recently emerged as a new, easily accessible means to estimate sex. With the introduction of MorphoPASSE (Klales, 2020), it is important that both the LR and DFA as well as the RFMs are tested in order to stay up to date with the current methodology and standards. Additionally, it is important to note that variation in statistical tests may give different results.

CHAPTER 4: THE EFFECTS OF AGE ON THE ACCURACY OF SEX ESTIMATION USING THE SKULL

4.1 Introduction

The purpose of this chapter is to evaluate the effects of age on the estimation of sex based on the classic cranial and mandibular sex indicators. Furthermore, this chapter identifies the presence of significant differences between the age cohorts and then compares the accuracy percentages in a pooled and divided sample.

4.2 Manuscript to be submitted

The impact of age-related changes in the skull on sex estimation

S. HOUSTON, D. BRITS, J. MYBURGH, L. LIEBENBERG

Manuscript to be submitted for publication to the International Journal of Legal Medicine.

Please note this chapter has been formatted based on the submission requirements of the International Journal of Legal Medicine.

Abstract

The sexually dimorphic traits on the skull described by Walker [3] are frequently employed in skeletal analyses for sex estimation. Previous research has highlighted various changes in the craniofacial complex associated with advancing age, as cranial remodelling persists into adulthood. Age has been recognised as a potential factor contributing to misclassification when using the Walker [3] traits. This study aimed to evaluate the positive predictive performance of these traits among individuals subdivided into different age cohorts. The traits were scored on a sample of 453 South African skulls. Only the nuchal crest exhibited statistically significant results, indicating differences between individuals younger than 40 years and those older than 40 years. Classification models showed a slight increase in accuracy for younger individuals when age-specific standards were applied, while accuracy for older individuals decreased slightly. However, the difference in accuracy with age-specific standards was negligible compared to accuracies obtained with all ages pooled. The results suggest that age did not have a substantial impact on the traits, and prior knowledge of age did not significantly influence the accuracy of the method.

KEYWORDS: Forensic anthropology, Cranium, Morphoscopic, Sexual dimorphism, Walker method, South Africa

INTRODUCTION

Sexual dimorphism, referring to the size and shape differences between males and females, has been observed throughout the human skeleton [1,2]. Various osteometric and morphoscopic techniques exist to quantify sexual dimorphism. Among these methods the Walker [3] method is considered one of the most preferred techniques to estimate sex when the skull is available for analyses. The Walker [3] method involves the examination of five morphoscopic traits on the cranium and mandible, name the glabella, supra-orbital margins, nuchal crest, mastoid processes, and the mental eminence. The methodology has been validated through numerous studies worldwide, with several authors noting population-specific variations among the traits [4,5,8]. Notably, the South African population exhibited differences compared to the North American reference sample used in the development of the method. In addition to population affinity, age has been recognised as a significant factor influencing sexual dimorphism, potentially impacting the accuracy of sex estimated [7,8].

Throughout adulthood factors such as decreased hormone levels, mechanical stress, remodelling, and a decline in nutrient availability and osteoblast production can contribute to changes in the skeleton [9]. In particular, changes in the craniofacial skeleton have been extensively documented in the literature [8, 10-15]. Age-related changes on the skull typically manifest in the form of tooth loss leading to alveolar bone resorption, rougher granulation at muscle attachment sites, structural alterations in the vault due to weakening of the masticatory muscles, as well as differential thickening or thinning of the cranial bones [11]. Albert and colleagues [8] described craniofacial aging continuing until the third decade of life; however, bony alterations may persist further into the fifth and sixth decades. Horizontal enlargement was reported between the third and eighth decades of life, with most craniofacial dimensions increasing in small increments. Moreover, anterior facial height, as well as mandibular length, was shown to also increase with age [4,7,8]. While differences are frequently observed when comparing the crania of young and old adults, some differences have also been noted among younger groups. More specifically, Ross and Williams [10] found significant shape differences among individuals aged between 20 to 25 years. Significant size differences were also identified around the 25-year mark [10]. These changes have been attributed to an increase in the size of the paranasal sinuses, eruption of the third molars and/or alveolar remodelling [10].

Indeed, the presence or absence of dentition can play a significant role in the morphology of the face [10]. Albert and colleagues [8] describe a dramatic increase in the lower anterior facial height with the continued eruption of the lower incisors, with the greatest increase in lower

anterior facial height occurring in the early twenties. Similarly, Similarly, Small *et al*. [16] found that tooth loss and edentulism affect the maxillary and mandibular alveolar bone, leading to changes in upper facial height and the shape of the palate. Mandibular alveolar resorption further contributes to the loss or reduction in prominence of the mental eminence. However, edentulism has far-reaching implications that can affect more than just the facial skeleton. A sample of edentulous South Africans exhibited inferior and anterior flexion of the basicranium, resulting in inferior movement of the external occipital protuberance (nuchal crest) and anterior and inferior movement of the mastoid processes when compared to their counterparts that did not suffer tooth loss [16].

The majority of previous studies have assessed age-related changes using geometric morphometric techniques or standard cranial measurements. But few studies have discussed the implications of aging on morphoscopic techniques applied to the skull [17]. Walker [18] indicated that cranial features associated with sex may differ throughout adult life. Males tend to see a shift from gracile to more robust during puberty; however, males younger than 30 years of age have been noted to exhibit more feminine cranial features [17,18]. The opposite is suggested to occur with females, where post-menopausal females tend to exhibit more masculine features [18]. It has actually been suggested that sex estimation using the cranium not be attempted in females older than 55 years of age, as misclassification is likely to occur [17,19]. More recent studies have indicated that a weak correlation exists, but that age did not have a large impact on the Walker [3] traits [9,15,17]. However, further research is required.

The South African-specific standards created by Kruger *et al.* [4] made use of the Pretoria Bone Collection (PBC). As is the case with numerous skeletal collections across the globe, the individuals that are accessioned often tend be of more advanced ages [20,21,22,23]. Notably, the sample used by Kruger et al. [4] demonstrate a disparity between the groups, with the mean age for the white South Africans being nearly 20 years older than the black South Africans. Thus, should age be an issue it could lead to greater misclassifications for one group if the same standards are applied regardless of age. Furthermore, the age distribution of individuals in skeletal collections may not align with that of individuals encountered in forensic casework, where younger adults are more prevalent. The present study aimed to explore the application of Walker [3] trait scores among South African individuals across different adult age cohorts with the objective of ascertaining if, and to what degree age-at-death is associated with the classification accuracy of sex estimates. By exploring the relationship between age-at-death

and sex estimation accuracies, this research seeks to enhance our understanding of the effects of age on the reliability of forensic anthropological methods.

MATERIALS AND METHODS

Sample

The sample comprised of the skulls of 453 black, white, and coloured South African males and females (Table 4.1) from the Pretoria Bone Collection (University of Pretoria) and the Raymond A. Dart Collection of Modern Human Skeletons (University of the Witwatersrand) in South Africa. Ethical approval was obtained to conduct this research (Reference number 588/2022).

The cranium has previously been reported to reach its adult size around 14 years of age [10]; thus, a lower age limit of 14 years was selected to maximise the sample size. No upper age limit was selected. The overall age range for the individuals included in the sample was between 14 and 108. The sample was then subdivided into ten age cohorts at 10-year increments (Table 4.1). As this study made use of convenience sampling, the number of individuals per cohort, as well as the demographic distribution within each cohort varies. As is often the case with cohort studies, the youngest and oldest cohorts consist of very few individuals; thus, sample size should be taken into consideration when interpreting the results. Individuals with visible pathology, ante-mortem trauma, and/or post-mortem damage that affected the accurate scoring of the traits were excluded. Individuals with partial or complete edentulism were included in the sample as a large number of individuals in skeletal collections (and in the South African population) demonstrate edentulism to some degree. The frequency of edentulism (partial and complete) for each age cohort of the sample is represented graphically in Figure 4.1. The frequency distribution shows that the percentage of individuals with edentulism is shown to increase with age, with the greatest percentage of individuals with edentulism concentrated in the older cohorts. Furthermore, partial edentulism is shown to be more concentrated in the middle-aged cohorts (40-69 years). This supports the inclusion of edentulism as a variable of aging.

Table 4.1 Total number of individuals in each age cohort and the sex and population affinity distribution of the sample.

	Sex		Population affinity			Total
Age Cohort	**Females**	**Males**	**BSA**	**WSA**	**CSA**	***n***
<20	5	4	6	1	2	9
20-29	30	30	53	2	5	60
30-39	30	30	57	0	3	60
40-49	30	31	54	6	1	61
50-59	30	31	50	10	1	61
60-69	30	30	40	19	1	60
70-79	30	30	36	24	0	60
80-89	30	30	17	43	0	60
90-99	9	11	4	16	0	20
>99	0	2	1	0	1	2
TOTAL	**224**	**229**	**318**	**121**	**14**	**453**

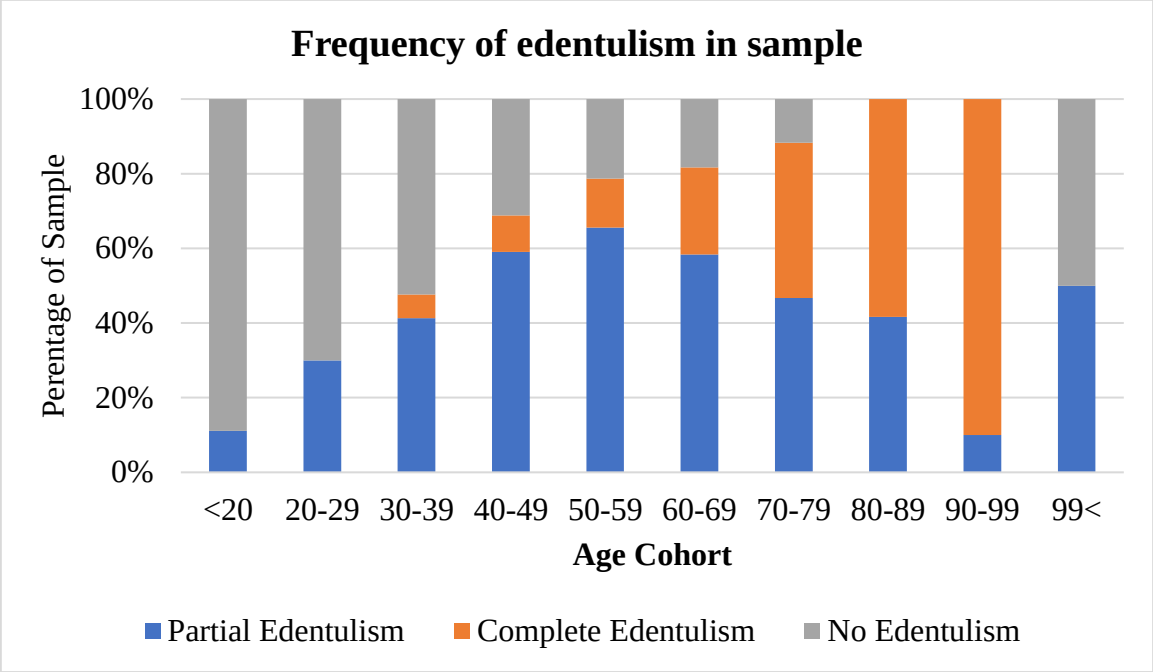


Figure 4.1 Frequency of partial and complete edentulism of the sample

Method

The five Walker [3] traits were visually assessed and scored for each skull in the sample (Table 4.2). Each trait was scored on an ordinal scale ranging between 1 and 5 using the definitions and line drawings as described in the methodology of the original paper [3]. In the case of bilateral traits, only the left side was scored.

Table 4.2 List of the Walker [3] traits on the skull and their abbreviations.

	Traits	Abbreviation	Scoring*
1	Glabella	Gla	1-5
2	Mastoid Process	Ma	1-5
3	Mental Eminence	Me	1-5
4	Nuchal Crest	Nu	1-5
5	Supra-orbital Margin	Or	1-5

*a score of 1 demonstrates a feminine expression, and a score of 5 a masculine expression.

Statistical analyses

All statistical analysis were completed with R (version 4.0.5) and the RStudio environment (version v1.4.1106) [24].

Observer agreement was tested on 10 skulls randomly selected from the sample. The sample included a combination of black and white South African males and females to encompass a range of trait expressions. The first author (SKH) underwent a training period to become proficient with the method and rescored the skulls two weeks apart to gauge the intra-observer repeatability. For the inter-observer agreement, an additional observer (LL) with more than 10 years of experience with the method also scored the skulls. Both observers engaged in multiple discussion sessions to resolve any discrepancies in scoring before the commencement of data collection; all of the data collection was performed by the first author (SKH). A quadratic weighted Cohen's Kappa was calculated for each trait using the *irr* package in R [25]. The calculated values were then interpreted using the descriptions by Landis and Koch [26] to show the strength of agreement between the scores.

Frequency distributions were calculated for each of the traits to analyse the distribution of scores within the sample. A Kruskal Wallis test was used to compare the scores for each trait among the different cohort. A *post hoc* Dunn's test was then calculated with a Bonferroni correction to further explore the results obtained with the Kruskal Wallis test. The Kruskal Wallis test is a non-parametric test suitable for ordinal data and is not subject to assumptions such as normality or homogeneity of variance. Results obtained with Kruskal Wallis is considered significant when $p<0.05$. The Dunn's test is required to determine which groups in

a multiple comparison demonstrates significant differences. The Bonferroni correction counteracts the effects of multiple comparisons and prevents increased probability of Type 1 errors occurring [27]. Multiple iterations were conducted: first the entire sample (sexes and populations) was pooled together; next the sample was separated according to sex to assess if there are different sex-specific trends among the age cohorts; finally, the sample was separated according to population affinity to identify population-specific trends among the age cohorts. Since the sample was extremely limited, the coloured South Africans were removed for the population analysis and only the black and white South Africans were compared.

Classification models were then created to establish the accuracy of the traits when the ages are pooled together, and when the ages are divided according to the Kruskal Wallis results (i.e., age-specific standards). To achieve this, multiple classification methods were employed. Firstly, ordinal logistic regression (LR) was selected to test the sex estimation accuracy of the model as this is the method used in the original paper [3] as well as the South African-specific standards [4]. The LR indicates the probability of correctly classifying sex based on the different expressions of the traits.

Secondly, random forest models (RFMs) were also implemented to further test the accuracy. RFM was selected as an additional method as it is used in MorphoPASSE, which is a computer program whereby trait scores from the skull and os coxa can be entered in statistical analyses to produce sex estimates with associated probabilities for unknown individuals [28]. The RFMs were set up so that 75% of the sample was used to train the model and the remaining 25% was used as a hold-out sample to test the model, providing an independent validation of the model performance. The *randomForest* package in R was used to conduct the RFM classifications [29]. In instances of missing values, the mode of the trait for the sex and population group that the skull belongs to was used as a data imputation technique. Data imputation was only conducted when less than 10% of the values were missing. Additionally, sex bias was calculated for the LR and RFMs by subtracting the female accuracy from the male accuracy. Sex bias indicates whether a certain sex group is classifying more accurately than the other and is useful for identifying trends in the scores that may lead to misclassification.

RESULTS

Inter- and intra-observer agreement

Inter- and intra-observer agreement was tested to confirm the repeatability and accuracy of the study to ensure that observer differences could be excluded as a variable and all significant differences identified could be attributed to age-related changes. Table 4.3 presents the Kappa values. The intra-observer agreement ranged from "No agreement" (0.00) to "Perfect agreement" (1.00), with the mental eminence and nuchal crest performing the worst and best, respectively. When comparing the results for the inter-observer agreement, the Kappa values were slightly lower, ranging from "Agreement equivalent to chance" (0.00) to "Near perfect" (0.87). Once again, the mental eminence performed the worst, and the nuchal crest performed the best.

Table 4.3 Quadratic-weighted Cohen's Kappa and values for the inter- and intra-observer agreement with the associated description following Landis and Koch [26].

	Intra-observer		Inter-observer	
Trait	**Kappa**	**Description**	**Kappa**	**Description**
Gla	0.83	Near perfect	0.84	Near perfect
Ma	0.76	Substantial	0.77	Substantial
Or	0.68	Substantial	0.46	Moderate
Nu	1.00	Perfect	0.87	Near perfect
Me	-0.13	None	0.00	Equivalent to chance

Exploratory analyses: Pooled sample

Table 4.4 indicates the frequencies for each trait score per age cohort with the sexes and population groups pooled. The frequency distribution shows that the greatest frequency of scores is concentrated around the intermediate scores (scores 2 to 4). The older cohorts demonstrated a greater frequency for a score of 5 (2.64% vs 0.95%), whereas the younger cohorts demonstrated a greater frequency for a score of 1 (6.51% vs 2.25%). The older and younger cohorts were divided based on the 40-year sectioning point identified in the Kruskal Wallis tests (Table 4.5). The frequency percentage of a score was calculated by adding all the percentages up and dividing by the number of age cohorts multiplied by the number of traits. For example, the frequency percentage of a score of 5 for the younger age cohort was calculated as follows:

Frequency of score 5 for younger age cohort = [(Gla <20 score 5 %) + (Ma <20 score 5 %) + (Or <20 score 5 %) + (Nu <20 score 5 %) + (Me <20 score 5 %) + (Gla 20-29 score 5 %) + (Ma 20-29 score 5 %) + (Or 20-29 score 5 %) + (Nu 20-29 score 5 %) + (Me 20-29 score 5 %) + (Gla 30-39 score 5 %) + (Ma 30-39 score 5 %) + (Or 30-39 score 5 %) + (Nu 30-39 score 5 %) + (Me 30-39 score 5 %)] ÷ (3 x 5)

Frequency of score 5 for younger age cohort = [0.00 + 0.00 + 0.00 + 0.00 + 5.00 + 1.67 + 0.00 + 3.39 + 0.00 + 0.00 + 0.00 + 0.00 + 0.00 + 1.96 + 2.22] ÷ (3 x 5)

Frequency of score 5 for younger age cohort = 0.95%

Table 4.4 Trait frequencies for each age cohort in the pooled sample.

	Age Cohort																	
	<20		20-29		30-39		40-49		50-59		60-69		70-79		80-89		90-99		99<	
Trait Scores	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Gla	(n = 9)		(n = 60)		(n = 60)		(n = 61)		(n = 61)		(n = 58)		(n = 60)		(n = 60)		(n = 20)		(n = 2)	
1	1	11.11	6	10.00	5	8.33	8	13.11	3	4.92	0	0.00	2	3.33	0	0.00	1	5.00	0	0.00
2	6	66.67	26	43.33	26	43.33	24	39.34	28	45.90	27	46.55	28	46.67	24	40.00	7	35.00	0	0.00
3	2	22.22	18	30.00	21	35.00	14	22.95	20	32.79	20	34.48	24	40.00	19	31.67	3	15.00	1	50.00
4	0	0.00	10	16.67	8	13.33	12	19.67	10	16.39	10	17.24	6	10.00	14	23.33	9	45.00	1	50.00
5	0	0.00	0	0.00	0	0.00	3	4.92	0	0.00	1	1.72	0	0.00	3	5.00	0	0.00	0	0.00
Ma	(n = 9)		(n = 60)		(n = 60)		(n = 61)		(n = 61)		(n = 60)		(n = 60)		(n = 59)		(n = 20)		(n = 2)	
1	0	0.00	0	0.00	1	1.67	2	3.28	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
2	3	33.33	26	43.33	30	50.00	17	27.87	26	42.62	20	33.33	22	36.67	13	22.03	6	30.00	0	0.00
3	3	33.33	21	35.00	17	28.33	28	45.90	21	34.43	19	31.67	26	43.33	31	52.54	10	50.00	1	50.00
4	3	33.33	10	16.67	11	18.33	12	19.67	8	13.11	17	28.33	9	15.00	13	22.03	4	20.00	1	50.00
5	0	0.00	3	5.00	1	1.67	2	3.28	6	4.92	4	6.67	3	5.00	2	3.39	0	0.00	0	0.00
Or	(n = 9)		(n = 59)		(n = 60)		(n = 60)		(n = 61)		(n = 59)		(n = 58)		(n = 59)		(n = 20)		(n = 2)	
1	0	0.00	6	10.17	5	8.33	4	6.67	3	4.92	4	6.78	3	5.17	0	0.00	0	0.00	0	0.00
2	2	22.22	17	28.81	19	31.67	11	18.33	16	26.23	17	28.81	14	24.14	12	20.34	3	15.00	0	0.00
3	4	44.44	21	35.59	21	35.00	30	50.00	27	44.26	25	42.37	26	44.83	29	49.15	13	65.00	1	50.00
4	3	33.33	13	22.03	15	25.00	13	21.67	14	22.95	10	16.95	14	24.14	18	30.51	4	20.00	1	50.00
5	0	0.00	2	3.39	0	0.00	2	3.33	1	1.64	3	5.08	1	1.72	0	0.00	0	0.00	0	0.00

Table 4.4 (continued)

	Age Cohort																	
	<20		**20-29**		**30-39**		**40-49**		**50-59**		**60-69**		**70-79**		**80-89**		**90-99**		**99<**	
Trait Scores	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Nu	(n = 9)		(n = 60)		(n = 60)		(n = 61)		(n = 61)		(n = 59)		(n = 60)		(n = 60)		(n = 20)		(n = 2)	
1	2	22.22	6	10.00	7	11.67	1	1.64	1	1.64	0	0.00	1	1.67	2	3.33	0	0.00	0	0.00
2	5	55.56	30	50.00	26	43.33	28	45.90	20	32.79	17	28.81	17	28.33	17	28.33	7	35.00	0	0.00
3	2	22.22	18	30.00	20	33.33	20	32.79	20	32.79	25	42.37	29	48.33	18	30.00	6	30.00	1	50.00
4	0	0.00	6	10.00	7	11.67	10	16.39	17	27.87	16	27.12	10	16.67	15	25.00	6	30.00	1	50.00
5	0	0.00	0	0.00	0	0.00	2	3.28	3	4.92	1	1.69	3	5.00	8	13.33	1	5.00	0	0.00
Me	(n = 9)		(n = 51)		(n = 45)		(n = 42)		(n = 41)		(n = 34)		(n = 24)		(n = 21)		(n = 2)		(n = 2)	
1	0	0.00	1	1.96	1	2.22	1	2.38	1	2.44	1	2.94	0	0.00	2	9.52	0	0.00	0	0.00
2	1	11.11	19	37.25	19	42.22	10	23.81	8	19.51	6	17.65	7	29.17	4	19.05	1	50.00	1	50.00
3	3	33.33	23	45.10	18	40.00	15	35.71	16	39.02	9	26.47	6	25.00	6	28.57	1	50.00	1	50.00
4	3	33.33	7	13.73	6	13.33	15	35.71	15	36.59	17	50.00	11	45.83	8	38.10	0	0.00	0	0.00
5	0	0.00	1	1.96	1	2.22	1	2.38	1	2.44	1	2.94	0	0.00	1	4.76	0	0.00	0	0.00

The results of the Kruskal Wallis test showed that significant differences existed only for the nuchal crest between some of the cohort groups (Table 4.5). More specifically, the Dunn's test showed that the younger cohorts (especially the 20–29-year-olds) differed the most from the older cohorts. The results indicate that significant differences exist between those younger than 40 years of age and those older than 40 years of age; thus a 40-year sectioning point was implemented and analysed for comparisons of accuracy between older and younger cohorts. Table 4.5 shows only the cohorts where significant differences were present, for a more detailed table representing all the age cohorts refer to APPENDIX E**: EXTENDED TABLES** (Table E1).

Table 4.5 Significant differences present in the nuchal crest of different age cohorts for the pooled sample.

Cohort comparison	P-value
<20 | 80-89	**<0.05**
20-29 | 50-59	**<0.05**
20-29 | 60-69	**<0.01**
20-29 | 70-79	**<0.05**
20-29 | 80-89	**<0.01**
30-39 | 60-69	**<0.05**
30-39 | 80-89	**<0.01**

Exploratory analyses: Sex differences

Table 4.6 indicates the frequencies for each trait score for each age cohort for the females in the sample. The frequency distribution shows that the greatest frequency (92.38%) of scores is still concentrated around the intermediate scores (scores 2 to 4) for all of the traits. The older cohorts demonstrated a slightly greater frequency for a score of 4 or 5 (13.12% versus 11.20%), whereas the younger cohorts demonstrated a greater frequency for a score of 1 or 2 (59.67% versus 46.08%). Females younger than 40 years of age are more likely to be scored a 2 for the glabella, mastoid process, nuchal crest, and the mental eminence, whereas females older than 60 years of age are more likely to be scored a 3 for the mastoid process, supra-orbital margin, nuchal crest, and the mental eminence.

Table 4.6 Trait frequencies for each age cohort for females.

	Age Cohort																	
	<20		**20-29**		**30-39**		**40-49**		**50-59**		**60-69**		**70-79**		**80-89**		**90-99**		**99<**	
Trait Scores	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Gla	(n = 5)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n =30)		(n = 30)		(n = 30)		(n = 9)		(n = 0)	
1	1	20.00	6	20.00	4	13.33	7	23.33	3	10.00	0	0.00	2	6.67	0	0.00	1	11.11	0	0.00
2	3	60.00	18	60.00	16	53.33	16	53.33	20	66.67	19	63.33	13	43.33	15	50.00	5	55.56	0	0.00
3	1	20.00	5	16.67	8	26.67	6	20.00	6	20.00	8	26.67	15	50.00	11	36.67	2	22.22	0	0.00
4	0	0.00	1	3.33	2	6.67	1	3.33	1	3.33	3	10.00	0	0.00	4	13.33	1	11.11	0	0.00
5	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Ma	(n = 5)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 29)		(n = 9)		(n = 0)	
1	0	0.00	0	0.00	0	0.00	2	6.67	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
2	2	40.00	19	63.33	21	70.00	15	50.00	20	66.67	12	40.00	13	43.33	8	27.59	4	44.44	0	0.00
3	1	20.00	8	26.67	4	13.33	10	33.33	10	33.33	11	36.67	15	50.00	15	51.72	5	55.56	0	0.00
4	2	40.00	2	6.67	5	16.67	3	10.00	0	0.00	5	16.67	2	6.67	6	20.69	0	0.00	0	0.00
5	0	0.00	1	3.33	0	0.00	0	0.00	0	0.00	2	6.67	0	0.00	0	0.00	0	0.00	0	0.00
Or	(n = 5)		(n = 29)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 29)		(n = 30)		(n = 9)		(n = 0)	
1	0	0.00	6	20.69	5	16.67	3	10.00	2	6.67	2	6.67	2	6.90	0	0.00	0	0.00	0	0.00
2	1	20.00	10	34.48	11	36.67	8	26.67	8	26.67	10	33.33	7	24.14	7	23.33	2	22.22	0	0.00
3	3	60.00	11	37.93	12	40.00	18	60.00	19	63.33	15	50.00	15	51.72	15	50.00	5	55.56	0	0.00
4	1	20.00	2	6.90	2	6.67	1	3.33	1	3.33	2	6.67	5	17.24	8	26.67	2	22.22	0	0.00
5	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	3.33	0	0.00	0	0.00	0	0.00	0	0.00

Table 4.6 (continued)

	Age Cohort																	
	<20		**20-29**		**30-39**		**40-49**		**50-59**		**60-69**		**70-79**		**80-89**		**90-99**		**99<**	
Trait Scores	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Nu	(n =5)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 29)		(n = 30)		(n = 30)		(n = 9)		(n = 0)	
1	2	40.00	5	16.67	5	16.67	1	3.33	1	3.33	0	0.00	1	3.33	2	6.67	0	0.00	0	0.00
2	2	40.00	18	60.00	16	53.33	16	53.33	13	43.33	7	24.14	8	26.67	11	36.67	5	55.56	0	0.00
3	1	20.00	7	23.33	9	30.00	10	33.33	12	40.00	14	48.28	15	50.00	10	33.33	3	33.33	0	0.00
4	0	0.00	0	0.00	0	0.00	3	10.00	2	6.67	7	24.14	3	10.00	6	20.00	1	11.11	0	0.00
5	0	0.00	0	0.00	0	0.00	0	0.00	2	6.67	1	3.45	3	10.00	1	3.33	0	0.00	0	0.00
Me	(n = 4)		(n = 28)		(n = 24)		(n = 18)		(n = 17)		(n = 18)		(n = 8)		(n = 8)		(n = 1)		(n = 0)	
1	0	0.00	1	3.57	1	4.17	0	0.00	1	5.88	1	5.56	0	0.00	1	12.50	0	0.00	0	0.00
2	1	25.00	16	57.14	12	50.00	5	27.78	6	35.29	5	27.78	3	37.50	2	25.00	1	100.00	0	0.00
3	1	25.00	10	35.71	10	41.67	10	55.56	7	41.18	4	22.22	3	37.50	5	62.50	0	0.00	0	0.00
4	2	50.00	1	3.57	1	4.17	3	16.67	3	17.65	7	38.89	2	25.00	0	0.00	0	0.00	0	0.00
5	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	5.56	0	0.00	0	0.00	0	0.00	0	0.00

Table 4.7 indicates the frequencies for each trait score for each age cohort for the males in the sample. The frequency distribution again shows that the greatest frequency (94.45%) of scores is concentrated around the intermediate scores (scores 2 to 4). The younger cohorts demonstrated a greater frequency for a score of 1 or 2 (33.31%) compared to the older cohorts (20.36%). Whereas the older cohorts demonstrated a greater frequency for a score of 4 or 5 (40.83%) than the younger cohorts (26.93%).

Table 4.7 Frequencies for each trait score in each age cohort for males.

	Age Cohort																	
	<20		**20-29**		**30-39**		**40-49**		**50-59**		**60-69**		**70-79**		**80-89**		**90-99**		**99<**	
Trait Scores	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Gla	(n = 4)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n =28)		(n = 30)		(n = 30)		(n = 11)		(n = 2)	
1	0	0.00	0	0.00	1	3.33	1	3.33	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
2	3	75.00	8	26.67	10	33.33	7	23.33	7	23.33	8	28.57	15	50.00	9	30.00	2	18.18	0	0.00
3	1	25.00	13	43.33	13	43.33	8	26.67	14	46.67	12	42.86	9	30.00	8	26.67	1	9.09	1	50.00
4	0	0.00	9	30.00	6	20.00	11	36.67	9	30.00	7	25.00	6	20.00	10	33.33	8	72.73	1	50.00
5	0	0.00	0	0.00	0	0.00	3	10.00	0	0.00	1	3.57	0	0.00	3	10.00	0	0.00	0	0.00
Ma	(n = 4)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 11)		(n = 2)	
1	0	0.00	0	0.00	1	3.33	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
2	1	25.00	7	23.33	9	30.00	2	6.67	5	16.67	8	26.67	9	30.00	5	16.67	2	18.18	0	0.00
3	2	50.00	13	43.33	13	43.33	17	56.67	11	36.67	8	26.67	11	36.67	16	53.33	5	45.45	1	50.00
4	1	25.00	8	26.67	6	20.00	9	30.00	8	26.67	12	40.00	7	23.33	7	23.33	4	36.36	1	50.00
5	0	0.00	2	6.67	1	3.33	2	6.67	6	20.00	2	6.67	3	10.00	2	6.67	0	0.00	0	0.00
Or	(n = 4)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 29)		(n = 29)		(n = 29)		(n = 11)		(n = 2)	
1	0	0.00	0	0.00	0	0.00	1	3.33	1	3.33	2	6.90	1	3.45	0	0.00	0	0.00	0	0.00
2	1	25.00	7	23.33	8	26.67	3	10.00	7	23.33	7	24.14	7	24.14	5	17.24	1	9.09	0	0.00
3	1	25.00	10	33.33	9	30.00	12	40.00	8	26.67	10	34.48	11	37.93	14	48.28	8	72.73	1	50.00
4	2	50.00	11	36.67	13	43.33	12	40.00	13	43.33	8	27.59	9	31.03	10	34.48	2	18.18	1	50.00
5	0	0.00	2	6.67	0	0.00	2	6.67	1	3.33	2	6.90	1	3.45	0	0.00	0	0.00	0	0.00

Table 4.7 (continued)

	Age Cohort																	
	<20		20-29		30-39		40-49		50-59		60-69		70-79		80-89		90-99		99<	
Trait Scores	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Nu	(n =4)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 30)		(n = 11)		(n = 2)	
1	0	0.00	1	3.33	2	6.67	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
2	3	75.00	12	40.00	10	33.33	11	36.67	7	23.33	10	33.33	9	30.00	6	20.00	2	18.18	0	0.00
3	1	25.00	11	36.67	11	36.67	10	33.33	7	23.33	11	36.67	14	46.67	8	26.67	3	27.27	1	50.00
4	0	0.00	6	20.00	7	23.33	7	23.33	15	50.00	9	30.00	7	23.33	9	30.00	5	45.45	1	50.00
5	0	0.00	0	0.00	0	0.00	2	6.67	1	3.33	0	0.00	0	0.00	7	23.33	1	9.09	0	0.00
Me	(n = 3)		(n = 23)		(n = 21)		(n = 23)		(n = 24)		(n = 16)		(n = 16)		(n = 13)		(n = 1)		(n = 2)	
1	0	0.00	0	0.00	0	0.00	1	4.35	0	0.00	0	0.00	0	0.00	1	7.69	0	0.00	0	0.00
2	0	0.00	3	13.04	7	33.33	4	17.39	2	8.33	1	6.25	4	25.00	2	15.38	0	0.00	1	50.00
3	2	66.67	13	56.52	8	38.10	5	21.74	9	37.50	5	31.25	3	18.75	1	7.69	1	50.00	1	50.00
4	1	33.33	6	26.09	5	23.81	12	52.17	12	50.00	10	62.50	9	56.25	8	61.54	0	0.00	0	0.00
5	0	0.00	1	4.35	1	4.76	1	4.35	1	4.17	0	0.00	0	0.00	1	7.69	0	0.00	0	0.00

The results of the Kruskal Wallis test showed that significant differences exist between some of the cohort groups for the nuchal crest and the supra-orbital margin when comparing females. The Dunn's test highlighted that the younger cohorts (especially the 20-29- and 30–39-year-olds) differed the most from the older cohorts (Table 4.8). The results of the Kruskal Wallis test on cohort differences between males for each trait showed that no significant differences were present between any of the cohort groups for any of the traits.

Table 4.8 Significant differences present in females of different age cohorts using the Kruskal Wallis test with a *post hoc* Dunn's test.

Trait	Cohort comparison	P-value
Or	20-29 | 80-89	**<0.05**
Nu	20-29 | 60-69	**<0.01**
	30-39 | 60-69	**<0.01**
	20-29 | 70-79	**<0.01**
	30-39 | 70-79	**<0.05**

Exploratory analyses: Population differences

When the sample was tested for population differences, significant differences were identified between black and white South Africans for the nuchal crest and the mental eminence (Table E3, APPENDIX E**: EXTENDED TABLES**).

Classification models

Four different classification iterations were conducted to assess if age impacted the classification of sex for both the LR and the RFM; this included (1) the pooled sample (with all of the variables included and all of the age cohorts combined); (2) the younger sample (including only the individuals younger than 40 years of age); (3) the older sample (including only the individuals older than 40 years of age); and (4) excluding the mental eminence (including all the of the age cohorts combined and all of the variables but excluding the mental eminence). The 40-year sectioning point was selected based on the results of the Kruskal-Wallis test that indicated significant differences from 40 years onwards. The mental eminence was excluded due to the results of the Kappa tests which indicated a low level of repeatability as well as the extremely small sample size and thus lack of data on the trait. Also, the high rate of edentulism in the sample has an effect on the normal size and shape of the mental eminence.

The pooled sample (1)

The accuracy of the pooled sample LR function (75.00%) and the RFM (87.61%) are beyond what would be expected by random chance. The correct classification for the LR function was 73.39% for females and 76.43% for males (Table 4.9). The RFM testing model, which was the 25% of the data that was not utilised to build the model, gave a correct classification accuracy

of 89.29% for females and 85.97% for males (Table 4.9). Thereafter, sex bias was calculated to indicate whether a certain sex group was classifying more accurately than the other. A large sex bias indicates that the accuracy of one of the sex groups is much lower than the other and is dramatically bringing down the overall accuracy of the model. The sex bias for the pooled sample LR was 3.04% which minor.

The LR function showed that the two traits that were not significant were the mastoid process and nuchal crest which means that they did not have high variable importance in the model. The RFM variable importance scale illustrated that the mental eminence showed the greatest level of importance, followed by the glabella, the supra-orbital margin, the mastoid process, and then lastly the nuchal crest (Figure 4.2 (a)).

When referring to Table 4.9, it is evident throughout the classification iterations that the RFMs are outperforming the LR functions.

Table 4.9 Detailed accuracies of each of the iteration types for the LR and RFM.

		Iteration				
			Age-specific			
Model		**Pooled Sample**	**Younger Sample**	**Older Sample**	**Total****	**Without Mental Eminence**
LR	Combined	75.00	78.43	72.22	75.33	70.65
	Females	73.39	80.00	68.18	74.09	75.00
	Males	76.43	76.60	75.00	75.80	66.37
	Sex bias*	3.04	-3.40	6.82	1.71	-8.63
RFMs	Combined	87.61	87.50	78.75	83.13	71.68
	Females	89.29	81.25	74.36	77.81	75.00
	Males	85.97	93.75	82.93	88.34	68.42
	Sex bias*	-3.32	12.50	8.57	10.53	-6.58

*Sex bias = Male percent correct – female percent correct x 100

**Total = (accuracy of the younger sample + accuracy of the older sample) ÷ 2

The age-specific samples (2 and 3)

The younger sample RFM (87.50%) had a greater accuracy than the younger sample LR function (78.43%); however, both are beyond what would be expected by random chance (Table 4.9). The younger sample RFM model indicated a higher accuracy percentage when compared to the pooled sample LR model but not to the pooled sample RFM. The younger sample LR function achieved a correct classification percentage of 80.00% for females and 76.60% for males (Table 4.9), while the RFM testing model gave a correct classification accuracy of 81.25% for females and 93.75% for males (Table 4.9). For the younger sample LR function the glabella, supra-orbital margin and mental eminence were shown to be statistically significant and have a high variable importance; however, for the younger sample RFM the variable importance scale illustrated that the mental eminence showed the greatest level of importance, followed by the supra-orbital margin, the mastoid process, the nuchal crest, and then lastly the glabella (Figure 4.3 (a)).

The older sample models returned lower accuracies than the above with the LR function achieving 72.22% and the RFM achieving 78.75%. Both accuracies are still beyond what would be expected from random chance. Despite the separation into cohorts, the older sample RFM model demonstrated a lower accuracy percentage when compared to the pooled sample RFM, however, the accuracy is greater than the pooled sample LR accuracy. Furthermore, the older sample RFM demonstrated a much lower accuracy than the younger sample RFM. A correct classification percentage of 68.18% was obtained for females and 75.00% for males from the LR function, whereas the RFM testing model gave a correct classification accuracy of 74.36% for females and 82.93% for males (Table 4.9). Once again, the LR function indicated that the glabella, left orbit and mental eminence were statistically significant. The older sample RFM model variable importance scale illustrated that the mental eminence showed the greatest level of importance, followed by the supra-orbital margin, the mastoid process, the glabella, and then lastly the nuchal crest (Figure 4.3 (b)).

The sample without the mental eminence (4)

The accuracy of the LR function, excluding the mental eminence was 70.65% which is beyond what would be expected by random chance. From the function a correct classification percentage of 75.00% was obtained for females and of 66.37% for males (Table 4.9). The glabella, mastoid process and supra-orbital margin were shown to be statistically significant and therefore have high variable importance for the LR function. Alternatively, the RFM

excluding the mental eminence achieved a 71.68% classification accuracy. The testing model gave a correct classification accuracy of 75.00% for females and 68.42% for males. The female accuracy is shown to stay the same between the LR model and the RFM, however the male accuracy increases with the RFM. The model's the variable importance scale displayed the glabella with the greatest level of importance, followed by the supra-orbital margin, the mastoid process, and then lastly the nuchal crest (Figure 4.2 (b)).

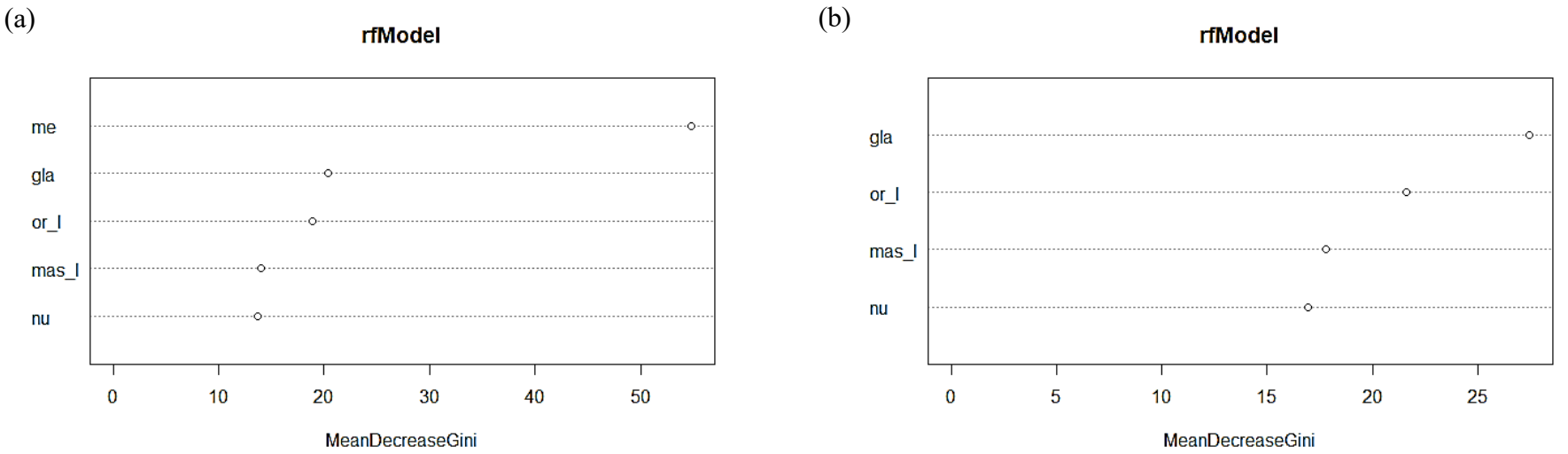


Figure 4.2 Visualisation and comparison of the variable importance in (a) the pooled sample RFM and (b) the without mental eminence RFM

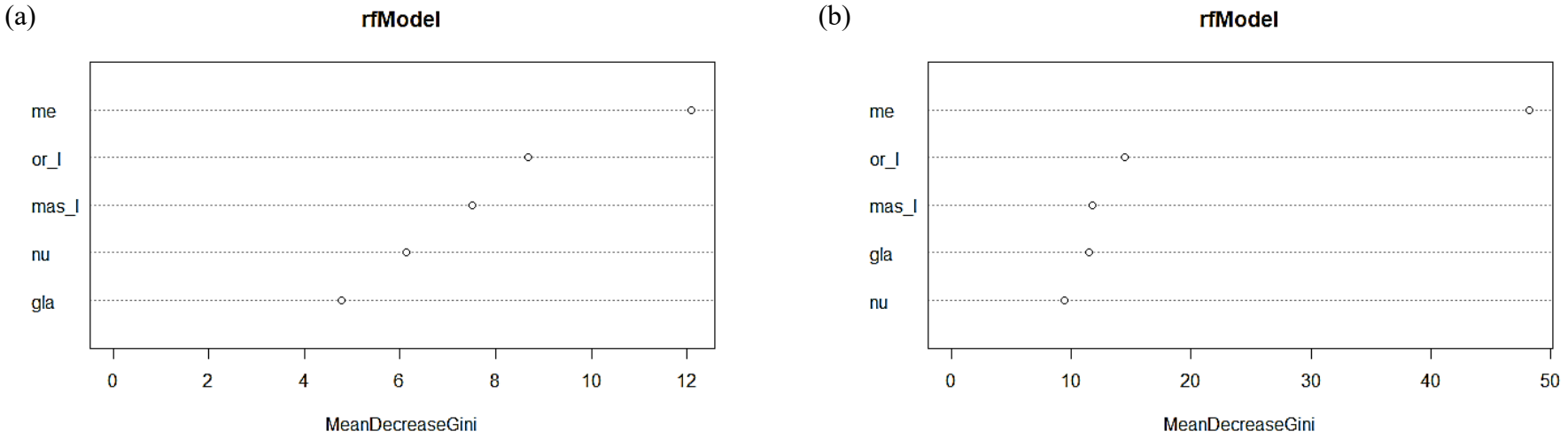


Figure 4.3 Visualisation and comparison of the variable importance in (a) the younger sample RFM and (b) the older sample RFM

DISCUSSION & CONCLUSION

Previous research has ascribed misclassification in sex estimation using the Walker [2] traits as the result of changes that occur with advancing age [8,10-15]. The current study explored potential trends in the trait scores that may be attributable to age, and assessed how age can affect the positive predictive performance of the method when estimating sex.

Overall, very few statistically significant differences were observed in the trait scores among the age cohorts. More specifically, the nuchal crest was the only trait to demonstrate any significant differences in the pooled sample, where individuals younger than forty years of age were noted to differ from individuals older than forty years of age. The nuchal crest is a major attachment site for the nuchal ligament and the superior portion of the trapezius muscle. The nuchal ligament assists in the attachment of the superior portion of the trapezius muscle to the cranium, but in most individuals the trapezius muscle also attaches to the occipital bone as an extension of itself [30]. De La Paz and colleagues [30] analysed the muscle structures and attachment sites of the head and neck for sex, population, and age-related differences [30]. The study focused on the anatomical relation of the muscles and subsequent attachment sites to the morphology of the nuchal crest [30]. Typically, nuchal crests that are considered to be more robust (i.e., that would be given a score of 5) include both a prominent, often hooked nuchal crest paired with rough nuchal lines. In their results, De La Paz and colleagues [30] did not find significant sex or age-related differences in the muscle attachment site itself; however, the authors did identify variation in the types of attachments. The results demonstrated the variable attachment of the nuchal ligament and trapezius muscle as a greater percentage of individuals exhibit no proximal cranial attachment by the superior portion of the trapezius muscle. Thus, the appearance of the nuchal crest as employed by forensic anthropologists may be influenced by the type of muscle attachment so that the nuchal area appears more robust. However, this apparent robusticity is not necessarily the result of sexual dimorphism and does not appear to be the result of aging. When the sample is divided so that males and females are assessed separately the results revealed that the males did not demonstrate any significant differences for the nuchal crest, and all significant differences were only observed with the females in the sample. More specifically, younger females (between 20-39 years) differed from the older females (between 60-79 years), while females that fall within the middle-aged cohort do not differ significantly from either the younger or older cohorts. Thus, the significant differences noted for the nuchal crest in the pooled sample was most likely due to significant differences between the cohorts of female individuals, and thereby emphasises the importance of exploring

the sexes separately to more effectively identify trends in craniofacial changes as each sex appears to be undergoing changes differently. Changes in the nuchal crest are assumed to be the result of continued and increasing strain of the nuchal ligament and trapezius muscle throughout an individual's lifespan. Additionally, any stress experienced by the vertebral column as the result of morphological changes, such as vertebral wedging that may occur with osteoporosis or general vertebral body compression that occurs with age, would also have an effect on the nuchal ligament as it is a superior and posterior extension of the supraspinous ligament [31,32]. However, it is unclear why the females in the sample presented with changes in the nuchal region while the males did not. Sex differences in musculoskeletal markers are typically attributed to differences in activity patterns or sexual division of labour, with males frequently engaging in harder physical labour resulting in increased hypertrophy of muscles and muscle attachment sites [33, 34]. By extension, males are expected to demonstrate greater differences at muscle attachment sites as these changes tend to be accumulative and have been correlated with age [33,34]. Weiss [33] noted that both sex and age play a role in the robusticity of muscle markers of the humerus, where females demonstrated slighter entheseal changes compared to their male counterparts. Yet the results from the current study did not follow this trend with the nuchal crest. It should be acknowledged that the majority of studies that have explored musculoskeletal stress markers and age have looked at postcranial skeletal elements, and changes on muscle markers of the cranium are incompletely described in the literature. However, further factors that should be considered to better understand age-related changes to the nuchal crest (in addition to sexual dimorphism) include body size, muscle size, muscle attachment size, and population affinity [30,33,34]. Weiss [33] also argues that aggregate muscle marker analyses (i.e., from more than one muscle attachment site) be conducted to gain a better understanding of the effects that the above factors have on the muscle attachments sites of individuals.

In addition to the nuchal crest, the females also demonstrated some significant differences (between the 20-year-old and 80-year-old cohorts) for the super-orbital margin when assessing the sexes separately. Once again, the males did not present with any differences, nor were the differences observed with the pooled sample. Many other studies have analysed age changes to the orbit, however these studies mainly focused on metric changes. Kahn and Shaw [35] noted significant changes to the orbital width and area with age in both males and females and suggested that dramatic changes occur to the entire bony orbit throughout an individual's life. Özer and colleagues [36] found slight correlations between changes in the width and height of

the orbital aperture and increasing age in females but not in males. Additionally, Ugradar and Lambros [37] identified a relationship between increasing orbital volume and increasing age in females. However, as the current study only identifies significant differences between two of the cohorts there is a distinct possibility that it is an artefact from the sample. Therefore, it is imperative that further studies be conducted to better explore sex differences in the aging of the orbit, especially with regards to morphoscopic methods. Another aspect that should be considered is secular trends. While it was not within the scope of the study to explore Secular trends, it should be considered that secular changes in the cranium may affect the traits. Goode [38] studied individuals from the Hamann-Todd and William M. Bass Skeletal Collections and identified the presence of some secular trends in the cranial morphoscopic traits in North America males and females from 1849–1960. Jantz and Jantz [39] and Grine *et al*. [40] identified metric secular trends that affect the cranial morphology in 19th to 20th century North Americans and South Africans respectively. Majority of the literature focuses on metric changes and there has been little to no focus on whether (and to what extent) the morphoscopic traits are affected. Further research should be conducted to determine if secular trends are still present in the morphoscopic traits and whether such trends are occurring in the South African population.

The results of the current study produced only a few significant differences between the age cohorts, however, this does not negate the existence of age-related differences. It is likely that morphoscopic methods may be unable to quantify the small differences caused by age-related changes between the cohorts. For example, it is unlikely that a female's score would change so drastically (such as from a score of 1 to a score of 3) in one lifespan. The shift from a score of a very gracile 1 to a score of an intermediate 3 would be radical and would require a large amount stress for the bone to alter so significantly. As was noted in previous literature [8,10-15] that was carried out on different populations, significant differences were present between the age cohorts, but they were slight and did not have a considerable effect on the sex estimate (Combined accuracy of 71.90%; <50 cohort accuracy of 69.10%; and >50 cohort accuracy of 81.30%) [3,15]. Thus, to account for the more subtle changes, metric methods and geometric morphometrics may be better able to quantify the more nuanced changes to the skeleton that occur from age.

Significant differences were also identified between the population groups, specifically between black and white South Africans, for the nuchal crest and mental eminence. However, the results differ from that of Krüger and colleagues [4] where significant differences between

all traits, with the exception of the supra-orbital margin and the mental eminence for males, were observed. The results of the Krüger *et al*. [4] study demonstrated that the sexual dimorphism of all the cranial traits is population-specific and population-specific formulae should be applied. The current study only observed significant population specificity for the mental eminence and nuchal crest between black and white South Africans. The difference in the results between the studies is unexpected as they were completed on similar samples. The study completed by Krüger *et al*. [4] also did not include coloured SAs in their sample and therefore additional assessment may be necessary to establish whether significant differences exist between coloured SAs and the other SA population groups. Furthermore, the presence of significant differences in the nuchal crest for both the age-cohorts and the population groups may possibly indicate that it may be more attributed to population influences rather than age-related changes. In such instances, it is important to always corroborate results with multiple skeletal elements and methodology. For this reason, multivariate analysis, and software such as MorphoPASSE [28] are advantageous since multiple skeletal elements and methods are run simultaneously.

Despite limited significant differences, the sample was subdivided into two broad groups (younger versus older) to see if there would be any difference in classification accuracy that could possibly justify prior knowledge of age to estimate sex. The results from the classification models indicated a slight increase in accuracy in the younger sample, which showed that the younger age cohorts classify more accurately when separated from the older age cohorts. The increase in accuracy in the younger sample is largely due to a substantial increase in the accuracy of the younger females, with the younger males only showing a slight improvement. The removal of the older, more masculine-presenting, females increased the accuracy of the sex estimate. In contrast, the older sample produced a decrease in accuracy when compared to the pooled model. The decrease in accuracy can potentially be attributed to the greater sex bias (6.82%) seen in the older group with a large decrease in accuracy for the females and only a slight decrease in accuracy for males. These results are consistent with the literature that states that females become more robust with age [18] resulting in a decrease in their sexual dimorphism where their traits appear more masculine. However, the changes are slight and given the amount of overlap that naturally occurs with the traits, the accuracy is not greatly affected when age-specific standards are used. Furthermore, as seen in the variable importance scale for the younger and older sample RFMs, the younger sample lists the nuchal crest as more important than the glabella and the older sample lists the glabella as more important than the

nuchal crest. The alteration in the variable importance between the younger and older sample is most likely attributed to the continued and increasing strain of the nuchal ligament and trapezius muscle during a females lifespan. Krogman and İscan [19] stated that, due to the effects of age on the skeleton, sex should only be estimated from the cranium for individuals between the ages of twenty and fifty-five years due to the increasing robusticity of females in older age cohorts. However, the classification accuracy observed with the older cohort classification model contests this, as the accuracy was not substantially lower the either the combined or younger models. Thus, sex estimation can be performed with high accuracy regardless of age. This supports other studies, such as the one completed by Garvin and colleagues in 2014 [9] or the study done by Lesciotto and Doershuk in 2018 [15], which stated that even though they found significant differences between age and the Walker [2] traits the correlation was weak and did not influence the accuracy of the sex estimate or the traits enough to validate the creation of new standards for medicolegal contexts.

The mental eminence posed numerous problems throughout the current study. Firstly, the mental eminence demonstrated the lowest level of agreement between the observers. The mental eminence is known by forensic anthropologists as a trait that commonly does not provide high levels of repeatability [41]. Studies done by Krüger and colleagues [4], Braun and colleagues [42] and Klales [17] further confirm the lack of repeatability of the menton (mental eminence) and the MorphoPASSE manual [28] recommends completely excluding the mental eminence due to its poor repeatability. Secondly, the alveolar resorption that occurs with ante-mortem tooth loss greatly affected the scoring of the mental eminence, and thus limited the sample. Previous literature [8,10,16] has noted the influence of edentulism of facial remodelling and although not proportionally linked to age, there was a large percentage of individuals in the sample with edentulism, particularly in the older age cohorts. Since there were so many individuals that could not be scored, data imputation was used to supplement the sample in order to run the RFM; however, it should be acknowledged that the magnitude of data imputation that was performed essentially nullifies any potential variation in the sample regarding the mental eminence and thus the model that includes the mental eminence is not statistically valid or reliable. The high variable importance that the RFM assigned to the mental eminence is likely a consequence of the mental eminence introducing a large amount of statistical noise into the model. The accuracy of the model is shown to decrease with the exclusion of the mental eminence, but it also produces a more accurate depiction of the true accuracy of the model with more realistic results due to the mental eminence's unreliability.

In conclusion, the findings of the study indicate that while there are some statistically significant differences in the sex-specific traits on the cranium and that the positive predictive performance of the models may be affected it may not be necessary to preselect age prior to the estimation of sex. The presence of significant differences may not necessarily just be attributed to age-related changes since population specific differences, edentulism, as well as the quasi-continuous nature of the morphoscopic methods may have had an impact on the results. The older sample experienced a decrease in accuracy while the accuracy increases when the younger individuals are isolated. However, this increase in accuracy for the younger sample wasn't extreme enough to warrant the pre-selection and isolation of the younger age cohorts. Therefore, the preselection of age before the estimation of sex using the Walker [3] method is not currently necessary in a medicolegal context.

DECLARATIONS

The authors have no competing interests to declare that are relevant to the content of this article.

ETHICS APPROVAL

All procedures performed in the study involving human remains were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Ethics Committee (Reference number 588/2022) of the University of Pretoria and the University of the Witwatersrand.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Department of Anatomy at the University of Pretoria and the School of Anatomical Sciences at the University of the Witwatersrand. Additionally, the authors acknowledge the curators of the skeletal collections Ms. GC Krüger and Dr. B. Billings for permission and access to the Pretoria Bone Collection and the Dart Collection respectively. Acknowledgement must also go to those individuals who have donated their remains for research and teaching.

REFERENCES

1. Krüger, G.C., L'Abbé, E.N. and Stull, K.E. 2017. Sex estimation from the long bones of modern South Africans. International Journal of Legal Medicine. 131(1):275-285. https://doi.org/10.1007/s00414-016-1488-z
2. Spradley, M.K., Jantz, R.L. 2011. Sex estimation in forensic anthropology: Skull versus postcranial elements. Journal of Forensic Sciences. 56(2):289-296. https://doi.org/10.1111/j.1556-4029.2010.01635.x
3. Walker, P.L. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. American Journal of Biological Anthropology. 136(1):39-50. https://doi.org/10.1002/ajpa.20776
4. Krüger, G.C., L'Abbé, E.N., Stull, K.E., Kenyhercz, M.W. 2015. Sexual dimorphism in cranial morphology among modern South Africans. International Journal of Legal Medicine. 129(4):869-875. https://doi.org/10.1007/s00414-014-1111-0
5. Lewis, C.J. and Garvin, H.M. 2016. Reliability of the walker cranial nonmetric method and implications for sex estimation. Journal of Forensic Sciences. 61(3):743-751. https://doi.org/10.1111/1556-4029.13013
6. Jilala, W., Ng'walali, P., Russa, D. and Bushozi, P. 2021. Sexing contemporary Tanzanian skeletonized remains using skull morphology: A test of the walker sex assessment method. Forensic Science International: Reports. 3:100195. https://doi.org/10.1016/j.fsir.2021.100195
7. Cabo, L.L., Brewster, C.P., Azpiazu, J.L. 2012. Sexual dimorphism: Interpreting sex markers. In: Dirkmaat, D.C. ed. A companion to forensic anthropology. Blackwell Publishing Ltd. p. 248-286. https://doi.org/10.1002/9781118255377.ch13
8. Albert, A.M., Ricanek, K., Patterson, E. 2007. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. Forensic Science International. 172(1):1-9. https://doi.org/10.1016/j.forsciint.2007.03.015
9. Garvin, H.M., Passalacqua, N.V., Uhl, N.M., Gipson, D.R., Overbury, R.S, Cabo, L.L. 2012. Developments in forensic anthropology: Age-at-death estimation. In: Dirkmaat, D.C. ed. A companion to forensic anthropology. Blackwell Publishing Ltd. p. 202-219. https://doi.org/10.1002/9781118255377.ch10

10. Ross, A.H., Williams, S.E. 2010. Craniofacial Growth, Maturation, and Change: Teens to Midadulthood. The Journal of Craniofacial Surgery. 21(2):458-461. https:/doi.org/10.1097/SCS.0b013e3181cfea34
11. Rogers, S.L. 1982. The aging skeleton: Aspects of human bone involution. Springfield, Illinois: Thomas C.C.
12. Ross, A.H., Jantz, R.L. McCormick, W.F. 1998. Cranial thickness in American females and males. Journal of Forensic Sciences. 43(2):67-72. https://doi.org/10.1520/JFS16218J
13. Lynnerup, N. 2001. Cranial thickness in relation to age, sex and general body build in a Danish forensic sample. Forensic Science International. 117(1-2):45-51. https://doi.org/10.1016/S0379-0738(00)00447-3
14. Gapert, R., Black, S., Last, J. 2013. Test of age-related variation in the craniometry of the adult human foramen magnum region: Implications for sex determination methods. Forensic Science, Medicine, and Pathology. 9(4):478-488. https://doi.org/10.1007/s12024-013-9437-3
15. Lesciotto, K.M., Doershuk, L.J. 2018. Effect of age on nonmetric cranial traits for sex estimation in adults. Forensic Anthropology. 1(3):150-159. https://doi.org/10.5744/fa.2018.0016
16. Small, C., Brits, D., Hemingway, J. 2016. Assessing the effects of tooth loss in adult crania using geometric morphometrics. International Journal of Legal Medicine. 130(1):233-243. https://doi.org/10.1007/s00414-015-1174-6
17. Klales, A.R. 2021. Current state of sex estimation in forensic anthropology. Forensic Anthropology. 4(2):118-127. https://doi.org/10.5744/fa.2020.3033
18. Walker, P.L. 1995. Problems of preservation and sexism in sexing: Some lessons from historical collections for palaeodemographers. In: Saunders, S.R. and Herring, Grave reflections: portraying the past through cemetery studies. Toronto: Canadian Scholars' Press. p. 31-47.
19. Krogman, W.M. and İscan, M.Y. 2013. The human skeleton in forensic medicine [Internet]. Springfield, Illinois, U.S.A.: Thomas, C.C. [cited 2023 Jun 2]. pp.143-193. Available from: https://search-ebscohost-com.uplib.idm.oclc.org/login.aspx?direct=true&db=nlebk&AN=646615&site=ehost-live&scope=site.

20. L'Abbé, E., Loots, M., Meiring, J. 2005. The Pretoria Bone Collection: A modern South African skeletal sample. Homo. 56(2):197-205. https://doi.org/10.1016/j.jchb.2004.10.004

21. Dayal, M.R., Kegley, A.D., Štrkalj, G., Bidmos, M.A., Kuykendall, K.L. 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 Biological Anthropology. 140(2):324-335. https://doi.org/10.1002/ajpa.21072

22. Alblas, A., Greyling, L.M., Geldenhuys, E. 2018. Composition of the Kirsten Skeletal Collection at Stellenbosch University. South African Journal of Science. 114(1-2):1-6. https://doi.org/10.17159/sajs.2018/20170198

23. L'Abbé, E.N., Krüger, G.C., Theye, C.E., Hagg, A.C., Sapo, O. 2021. The Pretoria Bone Collection: A 21st century skeletal collection in South Africa. Forensic Sciences. 1(3):220-227. https://doi.org/10.3390/forensicsci1030020

24. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Internet]. Available from: https://www.R-project.org/

25. Gamer, M., Lemon, J., Fellows, I., Singh, P. 2019. *irr*: Various coefficients of interrater reliability and agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr

26. Landis, J.R., Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics. 33(1):159-174. https://doi.org/10.2307/2529310

27. Ali, Z., and Bhaskar, S.B. 2016. Basic statistical tools in research and data analysis. Indian Journal of Anaesthesia. 60(9):662. https://doi.org/10.4103%2F0019-5049.190623

28. Klales, A.R. 2020. MorphoPASSE: Morphological pelvis and skull sex estimation program. In: Sex estimation of the human skeleton. United States: Elsevier Academic Press. p. 271-278. https://doi.org/10.1016/B978-0-12-815767-1.00016-X

29. Liaw, A. and Wiener, M. 2002. Classification and regression by *randomForest*. R news. 2(3):18-22. https://cran.r-project.org/doc/Rnews/

30. De La Paz, J.S., Buckley, H.R., Halcrow, S.E., Techataweewan, N., Woodley, S.J. 2023. Architecture of head and neck soft tissues and associated entheses: An exploration of sexual dimorphism in, and population differences between, New Zealand and Thai individuals. Journal of Anatomy. 243(1):1-18. https://doi.org/10.1111/joa.13853

31. Mercer, S.R. and Bogduk, N. 2003. Clinical anatomy of ligamentum nuchae. Clinical Anatomy: The Official Journal of the American Association of Clinical Anatomists and the British Association of Clinical Anatomists. 16(6):484-493. https://doi.org/10.1002/ca.10121

32. Grados, F., Fardellone, P., Benammar, M., Muller, C., Roux, C. and Sebert, J. 1999. Influence of age and sex on vertebral shape indices assessed by radiographic morphometry. Osteoporosis International. 10:450-455. https://doi.org/10.1007/s001980050253

33. Weiss, E. 2003. Understanding muscle markers: Aggregation and construct validity. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists. 121(3):230-240. https://doi.org/10.1002/ajpa.10226

34. Weiss, E. 2010. Cranial muscle markers: A preliminary examination of size, sex, and age effects. Homo. 61(1):48-58. https://doi.org/10.1016/j.jchb.2009.11.001

35. Kahn, D.M. and Shaw Jr, R.B. 2008. Aging of the bony orbit: A three-dimensional computed tomographic study. Aesthetic Surgery Journal. 28(3):258-264. https://doi.org/10.1016/j.asj.2008.02.007

36. Özer, C.M., Öz, I.I., Serifoglu, I., Büyükuysal, M.Ç. and Barut, Ç. 2016. Evaluation of eyeball and orbit in relation to gender and age. Journal of Craniofacial Surgery. 27(8):e793-e800. https://doi.org/10.1097/SCS.0000000000003133

37. Ugradar, S. and Lambros, V. 2019. Orbital volume increases with age: A computed tomography–based volumetric study. Annals of Plastic Surgery. 83(6):693-696. https://doi.org/10.1097/SAP.0000000000001929

38. Godde, K. 2015. Secular trends in cranial morphological traits: A socioeconomic perspective of change and sexual dimorphism in North Americans 1849–1960. Annals of Human Biology. 42(3):255-261. https://doi.org/10.3109/03014460.2014.941399

39. Jantz, R.L. and Meadows Jantz, L. 2000. Secular change in craniofacial morphology. American Journal of Human Biology: The Official Journal of the Human Biology Association. 12(3):327-338. https://doi.org/10.1002/(SICI)1520-6300(200005/06)12:3%3C327::AID-AJHB3%3E3.0.CO;2-1

40. Grine, F.E., Lee, C., Mongle, C.S., Billings, B.K., Wallace, I.J. and Mngomezulu, V. 2020. Secular trends in cranial size and shape among black South Africans over the

late 19th and 20th centuries. Annals of Human Biology. 47(5):446-456. https://doi.org/10.1080/03014460.2020.1783361

41. Lewis, C.J. and Garvin, H.M. 2016. Reliability of the walker cranial nonmetric method and implications for sex estimation. Journal of Forensic Sciences. 61(3):743-751. https://doi.org/10.1111/1556-4029.13013

42. Braun, S., Ridel, A.F., L'Abbé, E.N., Oettlé, A.C. 2023. Analysis of the hard-tissue menton shape variation in adult South Africans using cone-beam computed tomography (CBCT) scans. Forensic Imaging. 32:1-7. https://doi.org/10.1016/j.fri.2023.200532

CHAPTER 5: ASSOCIATION BETWEEN AGE AND THE ACCURACY OF SEX ESTIMATION USING THE PUBIC BONE

4.1 Introduction

The purpose of this chapter is to assess the implications of age on accurate scoring of the Phenice (1969) traits in a South African sample. Furthermore, this chapter identifies the lack of significant differences that exist in the traits between the age cohorts. However, previous research utilising similar methodology and alternate aspects of the os coxa noted significant differences between age cohorts. Thus, based on previous research as well as the average age of menopause in South Africa, age cohorts that symbolise 'younger, middle-aged and older adult' were utilised to compare the accuracy percentage in a pooled and divided sample.

4.2 Manuscript to be submitted

Assessing the implications of age on accurately scoring the Phenice (1969) traits in a South African sample

S. HOUSTON, J. MYBURGH, D. BRITS, L. LIEBENBERG

Manuscript to be submitted for publication to the Forensic Science International Journal.

Abstract

Forensic anthropologists consider the pelvis to be the most accurate skeletal element for sex estimation in medicolegal casework. Three morphoscopic traits on the pubic bone are frequently utilised to assess sexual dimorphism. Previous literature has indicated that the growth and deterioration related to aging that occurs throughout an individual's life may affect the classic skeletal sex indicators used for sex estimation. The study thus aimed to assess the association of age-related changes with the positive predictive performance of the classic skeletal sex indicators found on the pelvis.

The os coxae of 429 modern black, white, and coloured South Africans between the ages of 16 and 108 were scored using the pubic traits. The inter- and intra-observer agreement ranged from moderate to near perfect. Few significant differences were identified between age cohorts, with only males and black South Africans demonstrating some variation. Slight increases were noted in the classification accuracy when the sample was separated into three age cohorts (younger, middle-aged, and older adults). However, the increased accuracies were accompanied by a greater sex bias observed with the older age cohorts, mainly due to a decrease in accuracy among females. Any changes attributed to age were minor and the age-related variation did not greatly influence the presentation of the traits. While the preselection of age prior to the estimation of sex using the pubic traits is not necessary in a South African population the effects of age on the classic skeletal indicators may be population specific.

KEYWORDS: Forensic anthropology, Pubic bone, Pelvis, Morphoscopic, Sexual dimorphism

INTRODUCTION

Forensic anthropologists consider the pelvis to be the skeletal element that provides the greatest accuracies for sex estimation due to the direct role that it plays in sexual reproduction and parturition (Klales *et al.*, 2012). Phenice (1969) identified three traits on the pubic bone that could be used to differentiate between the sexes, namely the medial aspect of the ischiopubic ramus, the subpubic concavity, and the ventral arc. The traits were initially described as being present or absent in one sex or the other; however, greater variation was later recognised to be present in the traits. Thereafter, Klales and colleagues (2012) utilised the Phenice (1969) traits in a new modified method. The modified methodology involved the codification of the traits into an ordinal scale with five trait states. The method, which was paired with comparative line drawings, made use of robust classification statistics to allow for a repeatable and scientifically valid method that takes into account a wider range of variation (Klales *et al.*, 2012). Furthermore, the modification allowed the new methodology to satisfy the Daubert criteria (Dirkmaat *et al.*, 2008). Subsequent validation studies have been performed using the Klales *et al.* (2012) methodology on different populations (Gómez-Valdés *et al.*, 2017; Kenyhercz *et al.*, 2017; Lesciotto and Doershuk, 2018; Klales, 2020). In 2017, a global recalibration of the method was published that enables the application of the method to diverse populations across the globe while achieving high accuracies (Kenyhercz *et al.*, 2017). The recalibration was conducted since the application of the Klales *et al.* (2012) method on other populations produced accuracies that were much lower than the original study done using a North American sample. One of the most recent developments in sex estimation is the program MorphoPASSE, which recently emerged as a multivariate technique to estimate sex using the scores of the Walker (2008) and the Phenice (1969) traits in conjunction. MorphoPASSE utilises Random Forest Models (RFMs) and a database of numerous diverse population groups to estimate sex of unknown individuals with high accuracies (Klales, 2020). Additional benefits of MorphoPASSE include that it is easily accessible, user-friendly, freely available, is able to incorporate missing data, and it uses the existing Walker (2008) and Klales *et al.* (2012) methods with updated definitions and alternative statistical techniques.

Numerous factors have been reported to influence sexual dimorphism of the skeleton and, consequently, the accuracy with which sex can be estimated (Cabo *et al.*, 2012). Examples of these factors include population affinity, hormonal influences during an individual's lifetime, pregnancy (including the number of pregnancies), robust muscle attachments resulting from physical activity, and the age of the individual (Albert *et al.*, 2007; Waltenberger *et al.*, 2021).

The current study specifically focusses on the association between age and the accuracy of sex estimation.

Due to age-associated extrinsic and intrinsic factors, the human skeleton experiences growth and deterioration throughout an individual's life (Garvin *et al.*, 2012). Forensic anthropologists estimate age from the skeleton through the analysis of these age-related changes and the link between the timeline within which they occur. A number of the methods commonly employed for age estimation involve the pelvis, suggesting that age may play a role in the morphology of the pelvis.

As the central element of the human skeleton, the shape and size of the pelvis are directly influenced by biomechanical forces acting on the acetabula from the legs and on the sacroiliac joint from the spine (Waltenberger *et al*., 2021). Several studies have confirmed that despite the cessation of overall growth, remodelling occurs in certain pelvic dimensions throughout adulthood making the pelvis a target bone for age estimation methods (Tague, 1994; Berger *et al*., 2011; Huseynov *et al*., 2016; Mitteroecker and Fischer, 2016; Sharma *et al*., 2016; Auerbach *et al*., 2018). Two landmarks used for the estimation of age are located on the pelvis, namely the pubic symphysis and the auricular surface. Meindl and colleagues (1985) suggest that the pubic symphysis is the most reliable skeletal indicator for age estimation (Garvin *et al.*, 2012). Brooks and Suchey (1990) refined the original Todd (1920) and Acsádi-Nemeskéri (1970) methods of age estimation from the pubic symphysis into two (male and female) six phase systems that encompasses more variation identified in modern humans. The Suchey-Brooks (1990) method utilises robust statistics with clear descriptions, images, and casts of each phase to provide age-ranges that are repeatable and accurate (Franklin, 2010). Their method involves separate methodology and related casts for each of the sexes as the age-related changes that occur differ between males and females. The Phenice (1969) method also uses traits that are located on the pubic symphysis for sex estimation. Therefore, any degenerative changes to the pelvis which are associated with age may possibly also affect sex estimates. Thus, it is important to consider whether these sex-specific age-related changes may have an effect on the accuracy of the sex estimate.

The shape of the pelvis, particularly the shape and size of the pelvic inlet, undergoes changes throughout an individual's life with more pronounced changes occurring in females compared to males (Waltenberger *et al.*, 2021). These changes are influenced by endocrine factors (hormonal induction) and mechanical factors (e.g., childbirth) which explains the greater magnitude of change that is observed in females (Waltenberger *et al.*, 2021). Waltenberger and

colleagues (2021) reported that females displayed three phases of change: (1) during the parturition period; (2) pre-menopause and post-parturition; and (3) post-menopause. Males were reported to only display two phases of change, with the split occurring around 45 years of age. The first phase in females is primarily attributed to childbirth, with the degree of change increasing with the number of childbirths. The increased weight and pressure of the uterus and foetus on the pelvic girdle directly affects the underlying bone, namely the os coxae and the sacrum. Additionally, increased oestrogen production during pregnancy and the premenopausal phase likely leads to oestrogen-induced bone remodelling, resulting in the widening of the pelvic inlet during the first phase. Thus, the pelvic widening is likely to change due to the obstetric needs during a female's lifetime in order to ease cephalo-pelvic disproportion (Waltenberger *et al*., 2021). Although not a common occurrence, pregnancy and childbirth can also result in damage to the pelvis, particularly the pubic symphysis. Waltenberger and colleagues (2021) found that the recurrence of pelvic fractures in females continues to increase until the age of 40, which is correlated to the birthing process. These pelvic fractures can also affect the shape, size, and functionality of the pelvis. The obstetric dilemma, which is characterized by difficult and prolonged labours, is well-known in human females (Wells *et al.*, 2012). The body naturally produces hormones during the late stages of pregnancy to relax certain pelvic ligaments to facilitate childbirth. However, these ligaments can still become damaged or torn resulting in damage to the pubic symphysis and ventral arc (Wells *et al.*, 2012; Waltenberger *et al.*, 2021). The two phases of change in males are relatively similar but differ in the extent of change. The first phase, which occurs between 20 and 40 years of age, involves the pelvis becoming more heart-shaped, with narrowing of the greater sciatic notch, and a decrease in the subpubic angle and iliac height (Waltenberger *et al.*, 2021). The second phase in males, which parallels the postmenopausal phase observed in females, begins around 45 years of age, and mirrors the first phase but with a greater reduction in the anterior-posterior diameter and coxal height (Waltenberger *et al.*, 2021). The change in size and shape that is observed in the later phases of both males and females (postmenopausal phase) are attributed to mechanical factors, such as the continued strain (by muscles on the pelvis as well as forces caused by bipedal locomotion) throughout an individual's life, which ultimately results in bone loss. The bone loss observed can be further exacerbated in individuals who suffer from vitamin and mineral deficiencies (Waltenberger *et al*., 2021).

Few studies have attempted to assess age-related changes using non-metric or morphoscopic methods. DesMarais and colleagues (2023) assessed the effect of age on sex estimation of the

greater sciatic notch (GSN) in an Australian population using Multidetector Computed Tomography (MDCT) scans. The sample age ranged from 18 to 96 years of age; based on the average age of menopause in Australian females. The sample was divided into three age cohorts (18-49 years, 50-69 years, and 70+ years). When comparing the sex estimation accuracy of the 18-49 years age group and the 70+ years age group, the estimation accuracy was shown to decrease in older females but increase in older males. The results of the study demonstrate the effect of age on the morphology of the GSN (DesMarais *et al.*, 2023). Despite the depth of information available on skeletal changes that may occur with age, it is unclear if any substantial changes occur to the Phenice (1969) traits and the potential impact thereof on the positive predictive performance of these methods.

While previous studies have identified significant differences among different age cohorts (Tague, 1994; Albert *et al*., 2007; Berger *et al*., 2011; Huseynov *et al*., 2016; Mitteroecker and Fischer, 2016; Sharma *et al*., 2016; Auerbach *et al*., 2018; Waltenberger *et al*., 2021; DesMarais *et al.*, 2023), limited research has specifically examined the influence of these differences on the Phenice (1969) traits. Therefore, the current study aims to evaluate the possible association between age-at-death and the classic sex indicators on the pubic bone and, if present, the extent to which age-at-death affects the classification accuracy of the sex estimate in the South African population.

MATERIALS AND METHODS

Sample

The sample consisted of the left os coxae of 429 black (BSA), white (WSA) and coloured (CSA) male and female South Africans (Table 5.1) from the Pretoria Bone Collection (University of Pretoria) and the Raymond A Dart Collection of Modern Human Skeletons (University of the Witwatersrand) in South Africa. The individuals in the sample ranged from 16 to 108 years of age. The lower limit was selected as previous studies have demonstrated that subadults in early adolescence (from 13 years and older) could be accurately sexed using adult standards of the pelvis (Klales and Burns, 2017; Corron *et al.*, 2021). No upper age limit was selected in order to maximise the sample size. Any individuals with visible pathology, ante-mortem trauma, and/or post-mortem damage that affected the accurate scoring of traits were excluded. The sample was subdivided into ten age cohorts in increments of 10 years (Table 5.1) Since the current study made use of convenience sampling, the number of individuals per cohort as well as the demographic distribution within each cohort varied depending on the

availability of specimens that met the inclusion criteria. Notably, the sample for the youngest and oldest cohorts, as well as the coloured South Africans, are limited and thus the effect of sample size on statistical results should be taken into consideration. Ethical approval to conduct this research was obtained (Reference number 588/2022).

Table 5.1 Sample distribution.

	Sex		Population affinity			Total
Age Cohort	**Females**	**Males**	**BSA**	**WSA**	**CSA**	***n***
<20	3	0	1	1	1	3
20-29	30	30	52	3	5	60
30-39	30	29	58	0	1	59
40-49	29	29	50	7	1	58
50-59	30	29	43	15	1	59
60-69	28	31	36	23	0	59
70-79	29	30	22	37	0	59
80-89	29	29	10	48	0	58
90-99	4	9	3	10	0	13
>99	0	1	1	0	0	1
TOTAL	**212**	**217**	**276**	**150**	**9**	**429**

Method

The three Phenice (1969) traits on the os coxae were visually assessed and scored for each of the 429 individuals in the sample (Table 5.2). Each trait was scored on an ordinal scale ranging between 1 and 5 using the definitions and line drawings as described by Klales *et al.* (2012). In order for consistency throughout global studies, the use of the left side for estimation of the biological profile is standardised (Langley *et al.*, 2016); thus, the current study scored and analysed only the left os coxa unless it was not available in which case it was excluded.

Table 5.2 List of the Phenice (1969) traits on the os coxae and their associated abbreviations.

	Traits	Abbreviation	Scoring*
1	Sub-pubic contour	SPC	1-5
2	Medial aspect of the ischiopubic ramus	MIR	1-5
3	Ventral arc	VA	1-5

*A score of 1 demonstrates feminine expression, and a score of 5 masculine expression.

Statistical analyses

All statistical analysis were run using R (version 4.0.5) and the RStudio environment (version v1.4.1106).

Observer agreement testing involved the examination of 10 randomly selected os coxa from the sample, which comprised of black and white South African males and females, to ensure a range of trait expressions. To ensure proficiency with the method, the first author (SKH) underwent a training period to master the methodology and reevaluated the os coxa two weeks later to assess intra-observer repeatability. To ascertain inter-observer agreement, an additional observer (LL) with over a decade of experience in the methodology scored the os coxa. Prior to data collection, both observers participated in multiple discussion sessions to reconcile any scoring discrepancies. Thereafter, all data collection was carried out exclusively by the first author (SKH). Quadratic-weighted Cohen's Kappa values for each trait were calculated using the *irr* package in R (Gamer *et al.*, 2019), and the strength of agreement between the scores was interpreted based on the descriptions by Landis and Koch (1977).

Frequency distributions were computed to assess the distribution of scores for each trait within the sample. Thereafter, Kruskal-Wallis tests were employed to compare trait scores across the different age cohorts. *Post hoc* analysis using Dunn's test with a Bonferroni correction was then conducted to further elucidate the findings derived from the Kruskal-Wallis test. The Kruskal-Wallis test is a non-parametric method that is not subject to assumptions such as normality or homogeneity of variance which makes it well-suited for ordinal data. Statistical significance in Kruskal-Wallis results was determined at a threshold of $p<0.05$. The Dunn's test allowed for the identification of which cohort comparisons contained significant differences in the multiple comparisons. Application of the Bonferroni correction aimed to alleviate the

ramifications of multiple comparisons, mitigating the likelihood of Type I errors (Ali and Bhaskar, 2016). Several analytical iterations were conducted: initially the entire sample (which encompassed both sexes and all population groups) was pooled together; subsequently, the sample was divided by sex to assess the presence of sex-specific trends within age cohorts; finally, the sample was partitioned based on population affinity to discern population-specific trends across age cohorts. Due to the constrained sample size, the population analysis excluded CSAs, and only the black and white South Africans were compared.

Subsequent to obtaining the Kruskal-Wallis results, preprocessing the data and assuring the data was normally distributed, classification models were developed to assess trait accuracy in two scenarios. The scenarios consisted of when the cohorts were pooled together and when the cohorts were divided based on significant differences present (age-specific standards). This involved the application of multiple classification methods. Initially discriminant function analyses (DFA) were employed to evaluate the sex estimation accuracy, aligning with the methodology utilised in the original study (Klales *et al.*, 2012) and the global recalibration (Kenyhercz *et al.,* 2017). DFA provides probabilities for correctly classifying sex based on the different trait expressions.

Subsequently, random forest models (RFMs) were incorporated as a secondary measure to enhance the accuracy assessment. The choice of RFM as an additional method was motivated by its integration into MorphoPASSE, a computer program enabling the input of trait scores from the skull and os coxa into statistical analyses. This process generates sex estimates alongside associated probabilities for unknown individuals (Klales, 2020). The RFMs were configured to use 75% of the dataset for model training and the remaining 25% for independent validation, assessing model performance. The *randomForest* package in R facilitated the RFM classifications (Liaw and Wiener, 2002). In cases of missing values, data imputation involved utilising the mode of the trait within the corresponding sex and population group, restricted to instances where less than 10% of values were missing. Moreover, sex bias calculations for DFA and RFMs, derived from the subtraction of the male accuracy from female accuracy. The sex bias served as a valuable tool for indicting sex classification trends and identifying potential patterns in scores that might lead to misclassification.

RESULTS

Inter- and intra-observer agreement

Inter- and intra-observer agreement was tested to confirm the repeatability of the traits. The Kappa values for intra-observer error ranged from “Substantial” to “Near perfect agreement”, with the VA (0.80) performing the best and the SPC (0.67) performing worst (Table 5.3). Similarly, when comparing the results for the inter-observer agreement, the Kappa values ranged from “Substantial” to “Near perfect agreement” (Table 5.3) with the MIR (0.83) performing the best and the VA (0.71) performing the worst.

Table 5.3 Quadratic-weighted Cohen’s Kappa values for the inter- and intra-observer agreement with the associated description following Landis and Koch (1977).

	Intra-observer		Inter-observer	
Trait	**Kappa**	**Description**	**Kappa**	**Description**
VA	0.80	Near perfect	0.71	Substantial
SPC	0.67	Substantial	0.75	Substantial
MIR	0.75	Substantial	0.83	Near perfect

Descriptive Statistics

Table 5.4 displays the frequency distribution of the score for each trait in the age cohorts. The frequency distribution shows that the greatest frequency (88.74%) of scores is concentrated in the middle of the score range (scores 2-4). Concentration of the scores in the centre range with limited extremes (11.26%) is not out of the ordinary. The frequency percentage of a score was calculated by summing up the percentages and dividing by the number of age cohorts multiplied by the number of traits.

For example, the frequency percentage of a score of 1 or 5 for the sample was calculated as follows:

Frequency of score 1/5 for sample = [[(VA <20 score 1 %) + (MIR <20 score 1 %) + (SPC <20 score 1 %) + (VA 20-29 score 1 %) + (MIR 20-29 score 1 %) + (SPC 20-29 score 1 %) + (VA 30-39 score 1 %) + (MIR 30-39 score 1 %) + (SPC 30-39 score 1 %) + (VA 40-49 score 1 %) + (MIR 40-49 score 1 %) + (SPC 40-49 score 1 %) + (VA 50-59 score 1 %) + (MIR 50-59 score 1 %) + (SPC 50-59 score 1 %) + (VA 60-69 score 1 %) + (MIR 60-69 score 1 %) + (SPC 60-69 score 1 %) + (VA 70-79 score 1 %) + (MIR 70-79 score 1 %) + (SPC 70-79 score 1 %) + (VA 80-89 score 1 %) + (MIR 80-89 score 1 %) + (SPC 80-89 score 1 %) + (VA 90-99 score 1 %) + (MIR 90-99 score 1 %) + (SPC 90-99 score 1 %) + (VA 99< score 1 %) + (MIR 99< score 1 %) + (SPC 99< score 1 %) + (VA <20 score 5 %) + (MIR <20 score 5 %) + (SPC <20 score 5 %) + (VA 20-29 score 5 %) + (MIR 20-29 score 5 %) + (SPC 20-29 score 5 %) + (VA 30-39 score 5 %) + (MIR 30-39 score 5 %) + (SPC 30-39 score 5 %) + (VA 40-49 score 5 %) + (MIR 40-49 score 5 %) + (SPC 40-49 score 5 %) + (VA 50-59 score 5 %) + (MIR 50-59 score 5 %) + (SPC 50-59 score 5 %) + (VA 60-69 score 5 %) + (MIR 60-69 score 5 %) + (SPC 60-69 score 5 %) + (VA 70-79 score 5 %) + (MIR 70-79 score 5 %) + (SPC 70-79 score 5 %) + (VA 80-89 score 5 %) + (MIR 80-89 score 5 %) +

(SPC 80-89 score 5 %) + (VA 90-99 score 5 %) + (MIR 90-99 score 5 %) + (SPC 90-99 score 5 %) + (VA 99< score 5 %) + (MIR 99< score 5 %) + (SPC 99< score 5 %)] ÷ (10 x 3)

Frequency of score 1/5 for sample = [0.00 + 0.00 + 0.00 + 3.33 + 3.33 + 5.00 + 5.26 + 5.08 + 5.17 + 13.79 + 1.72 + 6.90 + 10.17 + 6.78 + 13.56 + 10.34 + 10.17 + 11.86 + 11.86 + 5.08 + 15.25 + 12.07 + 6.90 + 15.52 + 12.07 + 6.90 + 15.52 + 7.69 + 0.00 + 15.38 + 0.00 + 0.00 + 0.00 + 1.67 + 3.33 + 8.33 + 12.28 + 5.08 + 5.17 + 3.45 + 6.90 + 1.72 + 5.08 + 3.39 + 5.08 + 6.90 + 8.47 + 10.17 + 10.17 8.47 + 6.78 + 5.17 + 1.72 + 6.90 + 0.00 + 0.00 + 7.69] ÷ (10 x 3)

Frequency of score 1/5 for sample = 11.26%

Table 5.4 Trait frequencies for each age cohort in the pooled sample. Refer to Table 5.2 for trait abbreviations.

	Age Cohort																	
	<20		**20-29**		**30-39**		**40-49**		**50-59**		**60-69**		**70-79**		**80-89**		**90-99**		**99<**	
Trait Scores	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
VA	(n = 3)		(n = 60)		(n = 57)		(n = 58)		(n = 59)		(n =58)		(n = 59)		(n = 58)		(n = 13)		(n = 1)	
1	0	0,00	2	3,33	3	5,26	8	13,79	6	10,17	6	10,34	7	11,86	7	12,07	1	7,69	0	0,00
2	1	33,33	20	33,33	12	21,05	12	20,69	20	33,90	12	20,69	12	20,34	18	31,03	3	23,08	0	0,00
3	2	66,67	18	30,00	19	33,33	15	25,86	10	16,95	15	25,86	9	15,25	7	12,07	3	23,08	0	0,00
4	0	0,00	19	31,67	16	28,07	21	36,21	20	33,90	21	36,21	25	42,37	23	39,66	6	46,15	1	100,00
5	0	0,00	1	1,67	7	12,28	2	3,45	3	5,08	4	6,90	6	10,17	3	5,17	0	0,00	0	0,00
MIR	(n = 3)		(n = 60)		(n = 59)		(n = 58)		(n = 59)		(n = 59)		(n = 59)		(n = 58)		(n = 13)		(n = 1)	
1	0	0,00	2	3,33	3	5,08	1	1,72	4	6,78	6	10,17	3	5,08	4	6,90	0	0,00	0	0,00
2	0	0,00	11	18,33	13	22,03	14	24,14	17	28,81	10	16,95	11	18,64	13	22,41	3	23,08	0	0,00
3	2	66,67	19	31,67	20	33,90	25	43,10	21	35,59	20	33,90	19	32,20	23	39,66	3	23,08	0	0,00
4	1	33,33	26	43,33	20	33,90	14	24,14	15	25,42	18	30,51	21	35,59	17	29,31	7	53,85	1	100,00
5	0	0,00	2	3,33	3	5,08	4	6,90	2	3,39	5	8,47	5	8,47	1	1,72	0	0,00	0	0,00
SPC	(n = 3)		(n = 59)		(n = 58)		(n = 58)		(n = 59)		(n = 59)		(n = 59)		(n = 58)		(n = 13)		(n = 1)	
1	0	0,00	3	5,00	3	5,17	4	6,90	8	13,56	7	11,86	9	15,25	9	15,52	2	15,38	0	0,00
2	2	66,67	16	26,67	15	25,86	17	29,31	18	30,51	14	23,73	13	22,03	19	32,76	2	15,38	0	0,00
3	1	33,33	20	33,33	15	25,86	16	27,59	11	18,64	14	23,73	10	16,95	6	10,34	2	15,38	0	0,00
4	0	0,00	15	25,00	22	37,93	20	34,48	19	32,20	18	30,51	23	38,98	20	34,48	6	46,15	1	100,00
5	0	0,00	5	8,33	3	5,17	1	1,72	3	5,08	6	10,17	4	6,78	4	6,90	1	7,69	0	0,00

Inferential Statistics

The results of the Kruskal Wallis test on cohort differences showed that no significant differences were present between any of the cohort groups for any of the traits. The results also did not reveal any significant cohort differences between females for each of the traits. However, when the males were compared, the results of the Kruskal Wallis test showed that significant differences exist between the 20-29 and 70-79 cohorts for the left ventral arc. A detailed table representing all the age cohorts is available in APPENDIX E**: EXTENDED TABLES** (Table E2).

When the sample was analysed for population differences, significant differences were noted between black and white South Africans for all of the traits (Table 5.5). Analysis of the trait scores revealed a greater frequency of lower scores (2 and 3) for the WSAs whereas the BSAs displayed a greater frequency of higher scores (3 and 4). Unfortunately, as the CSAs were not a valid sample due to small sample size, they were not included in population comparisons. Refer to APPENDIX F**: ADDITIONAL TESTS** for additional studies performed on the population differences.

Table 5.5 Significant differences present between the Phenice (1969) traits of the different population affinities using the Kruskal Wallis test with a post hoc Dunn's test. Bold indicates significant differences.

Trait	Population comparison	P-value
VA	BSA | WSA	**<0.01**
SPC	BSA | WSA	**<0.05**
MIR	BSA | WSA	**<0.01**

Classification models were created using DFA to analyse the positive predictive performance of the sample and various models. As no significant differences were identified for the pooled results the individuals were not divided into variable cohort groups based on those results. The sample was divided into three larger groups that roughly correlates with younger adults (14-50 years of age), middle-aged adults 50-69 years of age), and older adults (70+ years of age) based on the results of previous literature (DesMarais *et al.*, 2023). Four different models were created to assess if age influenced the accuracy with which sex can be classified; more specifically the same tests were run using (1) the pooled sample; (2) the

younger sample; (3) the middle-aged sample; and (4) the older sample. RFMs were additionally utilised to further analyse the classification accuracy for the sample.

The accuracy of the **(1) Pooled Sample** DFA was 90.33% with a correct classification of 89.37% for females and 91.24% for males (Table 5.6). When stepwise variable selection was applied the accuracy decreased slightly to 90.07%. The pooled RFM yielded an accuracy of 88.99 while the testing model (which was the 25% of the data that was not utilised to build the model) achieved a correct classification rate of 83.33% for the females and 94.55% for the males (Table 5.6). For the pooled sample DFA the variable importance indicated that the VA displayed the highest variable importance, followed by the SPC and then lastly the MIR which is consistent throughout the models. However, throughout the RFMs it was indicated that the SPC had the highest variable importance followed by the VA and then the MIR (Table 5.7; Figure F1-Figure F4 in APPENDIX F**: ADDITIONAL TESTS**).

Table 5.6 Summary of accuracies for each of the models using DFA and RFM.

		Model				
			Age-specific			
Function		**Pooled Sample**	**Younger Sample**	**Middle-aged Sample**	**Older Sample**	**Total****
DFA	Combined	90.33	90.34	91.45	92.37	91.39
	Females	89.37	89.77	85.96	87.10	87.61
	Males	91.24	90.91	96.67	97.10	94.89
	Sex bias*	1.87	1.14	10.71	10.00	7.28
RFMs	Combined	88.99	91.11	90.00	87.88	89.66
	Females	83.33	86.96	80.00	87.50	84.82
	Males	94.55	95.46	100.00	88.24	94.57
	Sex bias*	11.22	8.50	20.00	0.74	9.75

*Sex bias = Male percent correct – females percent correct x 100

**Total = (accuracy of the younger sample + accuracy of the middle-aged sample + accuracy of the older sample) ÷ 2

The **(2) Younger Sample** DFA yielded an accuracy of 90.33% while the younger sample RFM yielded an accuracy of 91.11%. From the DFA function a correct classification percentage of 89.37% was obtained for females and of 91.24% for males (Table 5.6). Stepwise variable selection was applied, and the accuracy decreased slightly to 89.80%. The RFM testing model gave a correct classification accuracy of 81.25% for females and 93.75% for males (Table 5.6). Both the younger sample DFA and RFM show higher accuracies when compared to both the pooled sample DFA and the pooled sample RFM. The younger sample RFM is slightly higher than the younger sample DFA (91.11% versus 90.33%) however there is a much larger sex bias (8.50% versus 1.14%)

The accuracy of the **(3) Middle-aged Sample** DFA was 91.45%. When stepwise variable selection was applied, MIR was removed from the model and the accuracy increased slightly to 93.18%. An accuracy of 85.96% was obtained for females and of 96.67% for males for the DFA function (Table 5.6). The RFM testing model gave a correct classification accuracy of 80.00% for females and 100.00% for males with a total accuracy of 90.00% (Table 5.6). The middle-aged DFA indicated a higher accuracy when compared to the pooled sample DFA, furthermore, the middle-aged sample RFM showed a higher accuracy percentage when compared to the pooled sample RFM but not to the pooled sample DFA.

The **(4) Older Sample** iteration achieved a 92.37% accuracy and obtained an 87.10% accuracy for females and 97.10% for males (Table 5.6). With stepwise variable selection, the accuracy decreased slightly to 92.31%. The RFM for the older sample yielded an accuracy of 87.88% and the testing model gave a correct classification accuracy of 87.50% for females and 88.24% for males (Table 5.6). The older sample DFA also showed a higher accuracy when compared to the pooled sample DFA, but the older sample RFM showed a lower accuracy percentage when compared to both the pooled sample DFA and the pooled sample RFM.

Table 5.7 Variable importance for the different RFMs.

	Model			
	Pooled Sample	**Younger Sample**	**Middle-aged Sample**	**Older Sample**
Phenice (1969) Trait	Variable Importance			
VA	44.29	15.43	11.74	14.68
MIR	28.49	13.02	8.28	8.68
SPC	46.62	16.69	16.71	16.75

DISCUSSION

The primary aim of the study was to investigate the associations of age-related changes with the classic skeletal sex indicators located on the pubic bone and the subsequent effect on the accuracy of morphoscopic sex estimation methods used by forensic anthropologists. Previous research (Tague, 1994; Albert *et al*., 2007; Berger *et al*., 2011; Huseynov *et al*., 2016; Mitteroecker and Fischer, 2016; Sharma *et al*., 2016; Auerbach *et al*., 2018; Waltenberger *et al*., 2021; DesMarais *et al.*, 2023) has acknowledged this relationship, but a comprehensive examination of the potential implications has not yet been undertaken.

Numerous studies have discussed the age-related changes that occur to the pelvis throughout an individual's lifetime. For example, Lovell (1989) noted that the accuracy of the sex estimate decreased with increasing age and Rogers and Saunders (1994) observed that age may influence the sex estimation accuracy. Studies have observed differences between the various age cohorts in the shape and size of the pelvic inlet (Waltenberger *et al.*, 2021), in the pelvic dimensions (Tague, 1994; Berger *et al*., 2011; Huseynov *et al*., 2016; Mitteroecker and Fischer, 2016; Sharma *et al*., 2016; Auerbach *et al*., 2018) and when using Geometric Morphometrics (GMM; DesMarais *et al.*, 2023).

DesMarais and colleagues (2023) utilised the same methodology as the current study on the greater sciatic notch (GSN) in an Australian population. DesMarais and colleagues (2023) observed that the mean score for each sex increased with age; however, only the increase in scores in females was statistically significant. The resultant logistic regression (LR) accuracy decreased for females when comparing the youngest and oldest cohorts (99% and 91% respectively) but increased for males when comparing the above cohorts (79.4% and 87.3%

respectively); however, the study demonstrated a relatively large sex bias. The study agreed with Walker (2005) who observed that the morphology pattern of the greater sciatic notch changes with age and due to this, older females are more likely to be misclassified than males. Therefore, DesMarais and colleagues (2023) results showed that age affects GSN morphology, particularly in older, postmenopausal females as on average the GSN becomes narrower with increasing age and thus they recommend that when conducting sex estimation using the GSN morphology in forensic casework that age is considered. This demonstrates that age-related differences can be quantified using a quasi-continuous scoring system; however, the current study was not able to identify any significant differences when using the Phenice (1969) traits. This suggests that the pubic bone does not experience substantial age-related changes like the ilium. However, other factors, such as observer error, population variation and/or the lack of malleability of the pubic bone should also be considered. Despite the lack of significant differences observed among the cohorts, the current study employed the same age divisions as DesMarais and colleagues (2023) to explore whether any differences in classification accuracy could be detected. The average age of menopause in South African females is around 49.5 years of age (Benjamin, 1960; Frere, 1971; Walker *et al.*, 1984) and therefore the cohort groupings were kept the same for ease and comparability between the two studies. Regardless of the lack of significant differences between the cohorts, the division resulted in slight improvements in the accuracies of the models when compared to the pooled sample. However, with the exception of the younger sample, an increased sex bias accompanied the accuracy increases which can be attributed to an improvement in the male accuracies along with a reduction in the female accuracies. The improvement in males and reduction in female accuracies can be potentially related to the increasing robusticity of the skeleton with increasing age. Separating the age cohorts in this manner, allows the younger and more gracile males to not be confused with the older and more robust females thereby dramatically increasing the older male predictive performance. The pooled sample of the current study achieved a sex estimation classification accuracy of 90.57%. The aforementioned classification accuracy obtained is slightly lower but still in line with the original classification accuracy (95%) obtained by Klales *et al.* (2012) on the Hamman-Todd Osteological Collection. The current study also achieved similar accuracies to Kelley (1979), McBride *et al.* (2001) and Ubelaker and Volk (2002) while achieving greater accuracies than Lovell (1989) and MacLaughlin and Bruce (1990). The DFA model of the current study also achieved high accuracies for both males and females with a low sex bias which is in agreement with Klales *et al.* (2012) as well. A low sex bias is a good indicator as it demonstrates that problems or variations in one of the

sexes does not lower the overall accuracy percentage of the model. For example, if a large percentage of females were misclassifying as males, the female positive predictive performance would be much lower than the males and ultimately the accuracy of the entire model would be affected.

RFMs were also implemented to classify the individuals and test the accuracy percentages as an auxiliary analysis. For the majority of the samples, the testing model achieved a lower accuracy than the DFA model as well as a much greater sex bias. The above may be due to a number of factors, such as the data imputation method chosen, the sample size, and the variable importance of the model. The variable importance of the model stated that the subpubic contour and the ventral arc have the greatest variable importance and the medial aspect of the ischiopubic ramus has the lowest variable importance with one of the models even opting to exclude the trait completely. Rogers and Saunders (1994) and Johnstone-Belford and colleagues (2018) reported that the MIR displayed the lowest classification accuracy and was the most variable between the sexes, therefore, low variable importance for medial aspect of the ischiopubic ramus is not unexpected.

Sutherland and Suchey (1987) proposed that some features, such as the ventral arc for example, do not become prominent until an individual reaches the third decade of life while Kelley (1979) and Tague (1989) proposed that other features such as the preauricular sulcus, subpubic contour, and dorsal pitting diminish in their characteristic appearance with increasing age. The accuracies achieved by Rogers and Saunders (1994) in the three age cohorts did not show any significant differences, despite the 25-to-44-year cohort showing a perceived pattern of increased accuracy. Despite the in-depth research into the precision and accuracy of seventeen sexually dimorphic traits, Rogers and Saunders (1994) found no indication of an effect of age but related the lack of significant differences to potentially be due to the small sample size.

A study to assess the accuracy of using multi-detector computed tomography (MDCT) scans of the pelvis to estimate sex using the Phenice (1969) traits was completed by Johnstone-Belford and colleagues (2018). The study also included the analysis of age on the sex estimation accuracy. For the ventral arc the study identified no noticeable age-bias in the assignment of the scores in the male sample. The female sample showed much lower accuracies which the authors attributed to potentially be due to the poor performance of the younger females with only 25% of the females aged 18-20 years being correctly classified as female. With the exclusion of the 18–20-year-old females the overall classification accuracy of the trait was

shown to increase substantially from 79.28% to 85.86% (Johnstone-Belford *et al.*, 2018). Johnstone-Belford and colleagues (2018) also identified that the subpubic concavity was the most discriminatory trait for the Western Australian female population and those individuals that were misclassified were all females aged 25 years and younger. Furthermore, a more concave contour was also frequently identified in males which was shown to decrease the overall accuracy, particularly in males aged 31 to 40 years. The study did not identify any specific age-bias in the morphoscopic analysis of the medial aspect of the ischiopubic ramus for the Western Australian sample and differing classification accuracies between males and females was attributed to the lack of detail of the MDCT scans when attempting to analyse the trait (Johnstone-Belford *et al.*, 2018). Johnstone-Belford and colleagues (2018) also identified the ventral arc as the Phenice (1969) trait that is the most accurately indicative of sex in the Western Australian sample. However, this is only true for individuals aged older than 25 years of age as it was noted that some females younger than 25 years lacked a ventral arc due to incomplete development and maturation (Johnstone-Belford *et al.*, 2018). Huseynov *et al.* (2016) identified that continued growth was present in the obstetric dimensions of the pelvis in females between 20 and 40 years of age which were especially apparent in the ischiopubic region. Selliah and colleagues (2020) completed a study on sex estimation using metric and morphoscopic techniques on older individuals with a mean age of 75 years from the CAL Milano Cemetery Skeletal Collection. The outcomes of their study confirm that the effect of age on sex estimation methods is not substantial and that the existing methods can continue to be applied on middle to late adulthood individuals (Selliah *et al.*, 2020).

The differing results of the current study to the previous literature (Rogers and Saunders, 1994; DesMarais *et al.*, 2023) may be due to the focus of the pelvic elements used. The current study focused on the Phenice (1969) traits found in the ischiopubic segment whereas other studies explored the effects of age on multiple elements of the os coxa found in both the ischiopubic segment and the sacroiliac segment. According to the systems approach, two basic functional systems are apparent in the os coxa, namely the ischiopubic segment and the sacroiliac segment. The ischiopubic segment (which is made up of the pubis and the ischium) directly reflects the sexually dimorphic function of the pelvis, namely reproduction and parturition which are under hormone control and influence (Novotný, 1986). The sacroiliac segment (which is made up of the sacrum and the ilium) is more influenced by hominisation and the adaption to bipedal locomotion (Novotný, 1986). Therefore, as the sacroiliac segment is influenced more by mechanical and local factors, it is evident that continued forces

throughout an individual's life would have an effect on its presentation. This explains why age would have an effect on the more plastic sacroiliac segment but less so on the ischiopubic segment. For example, skeletal elements such as the greater sciatic notch (GSN), as studied by DesMarais and colleagues (2023), would show more plasticity and be influenced by continuous mechanical factors and ultimately age.

There are a number of additional factors that may cause alterations to the pelvis; these factors include but are not limited to population-specificity / latitude-based differences, socioeconomic status (SES) of the reference sample, and the lower level of plasticity of the pelvis when compared to other aspects of the os coxa. Auerbach and colleagues (2018) indicated that females exhibit a change in pelvic size and shape with increasing age. Additionally, their results identified that majority of the pelvis dimensions are greater in higher-latitude groups when compared to mid-latitude groups, indicating that it is necessary to consider population when studying and making estimates from the pelvis (Auerbach *et al.*, 2018). The current study's results of the Kruskal Wallis tests showed significant differences exist between black and white South Africans for all of the Phenice (1969) traits. Analysis of the scores revealed that WSAs displayed lower scores meaning their trait presentation is more gracile whereas BSA displayed greater scores meaning their trait presentation is more robust. Population affinity is not ordinarily pre-selected and separated prior to sex estimation using the Phenice (1969) traits on the South African population. Ordinarily the Kenyhercz *et al.* (2017) global recalibration is utilised, but the results of the current study show there is a potential need to pre-select population affinity prior to using the Phenice (1969) traits for sex estimation in a South African context. Garvin & Klales (2020) found that the global calibrations tend to be biased towards the mean when the groups contain varying levels of sexual dimorphism.

CONCLUSION

The current study did not identify any significant differences between the cohorts. The lack of significant differences identified could be due to variability in scoring due to observer error or simply due to the nature of the quasi-continuous scoring system which may mask any potential age-related changes that are present in the traits. Although no significant differences between the age cohorts were identified in the current study, the accuracies of the sex estimate did change slightly when the groups were divided into three larger cohorts.

Age-related differences may be present (as seen in other literature using metric methods and GMM) but may only be present as slight variations which are not discernible during analysis

when using morphoscopic techniques. Additionally, these age-related variations may be more frequently noted in the sacroiliac complex than the pubis. Therefore, the pre-selection of age prior to the scoring of the sex-specific indicators does not appear to be necessary following the results of the current study despite slight increases in accuracies when separating younger individuals from older individuals. Forensic practitioners need to make use of a multitude of methods to confirm and support their conclusions, especially when analysing younger individuals as they tend to misclassify more frequently (American Academy of Forensic Sciences Standards Board, 2019). Furthermore, in the case of a very young individual or where limited skeletal material is available (i.e., only one method can be used for estimations) it is important for forensic practitioners to disclose in the report that multiple analyses could not be conducted. Furthermore, it was noted when conducting the study, how the differing phase of the pubic symphysis may result in confusion and mis-scoring of the Phenice (1969) traits. Therefore, future studies should be conducted to find whether correlations exist between the phase of the pubic symphysis scored during age-at-death estimation and the Phenice (1969) traits scores which are estimated during sex estimation in order to more conclusively determine whether age-at-death has an effect on the sex estimate.

Lastly, the degree to which age may affect the classic skeletal sex indicators may be population specific with some significant differences observed between age cohorts, differences that are present but do not require the pre-selection of age and creation of new standards, however this needs to be explored further. South African population-specific differences present in the Phenice (1969) traits still need to be considered as the current study found significant differences between the population groups.

DECLARATIONS OF INTEREST: None.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the Department of Anatomy at the University of Pretoria and the Department of Anatomical Sciences at the University of the Witwatersrand, as well as Ms. GC Krüger and Dr B. Billings for permission and access to the Pretoria Bone Collection and the Dart Collection respectively. Additionally, acknowledgement must be made to those who donated their bodies for the furthering of research and teaching.

REFERENCES

Acsádi, G.R., Nemeskéri, J.N. 1970. History of human life span and mortality, Budapest, Akadémiai Kiadó. 346. https://www.jstor.org/stable/2740808

Albert, A.M., Ricanek, K., Patterson, E. 2007. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. Forensic Science International. 172(1):1-9. https://doi.org/10.1016/j.forsciint.2007.03.015

Ali, Z., and Bhaskar, S.B. 2016. Basic statistical tools in research and data analysis. Indian Journal of Anaesthesia. 60(9):662. https://doi.org/10.4103%2F0019-5049.190623

American Academy of Forensic Sciences Standards Board. 2019. Standard 090: Standard for sex estimation in forensic anthropology. AAFS Standards Board, Colorado Springs, CO. 1:1-4.

Auerbach, B.M., King, K.A., Campbell, R.M., Campbell, M.L., Sylvester, A.D. 2018. Variation in obstetric dimensions of the human bony pelvis in relation to age-at-death and latitude. American Journal of Biological Anthropology. 167(3):628-643. https://doi.org/10.1002/ajpa.23690

Benjamin, F. 1960. The age of the menarche and of the menopause in white South African women and certain factors influencing these times. South African Medical Journal. 34(16):316-320.

Berger, A.A., May, R., Renner, J.B., Viradia, N., Dahners, L.E. 2011. Surprising evidence of pelvic growth (widening) after skeletal maturity. Journal of Orthopaedic Research. 29(11):1719-1723. https://doi.org/10.1002/jor.21469

Brooks, S., and Suchey, J.M. 1990. Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. Human Evolution. 5:227-238. https://doi.org/10.1007/BF02437238

Cabo, L.L., Brewster, C.P., Azpiazu, J.L. 2012. Sexual dimorphism: Interpreting sex markers. In: Dirkmaat, D.C. ed. A companion to forensic anthropology. Blackwell Publishing Ltd. p. 248-286. https://doi.org/10.1002/9781118255377.ch13

Corron, L.K., Santos, F., Adalian, P., Chaumoitre, K., Guyomarc'h, P., Marchal, F., Brůžek, J. 2021. How low can we go? A skeletal maturity threshold for probabilistic visual sex estimation from immature human os coxae. Forensic Science International. 325:1-15. https://doi.org/10.1016/j.forsciint.2021.110854

DesMarais, A., Obertova, Z. and Franklin, D. 2023. The influence of age on greater sciatic notch morphology: Testing the Walker method in an Australian population. International Journal of Legal Medicine. 1-9. https://doi.org/10.1007/s00414-023-02988-1

Dirkmaat, D.C., Cabo, L.L., Ousley, S.D. and Symes, S.A. 2008. New perspectives in forensic anthropology. American Journal of Physical Anthropology. 137(S47):33-52. https://doi.org/10.1002/ajpa.20948

Franklin, D. 2010. Forensic age estimation in human skeletal remains: Current concepts and future directions. Legal Medicine. 12(1):1-7. https://doi.org/10.1016/j.legalmed.2009.09.001

Frere, G. 1971. Mean age at menopause and menarche in South Africa. The South African Journal of Medical Sciences. 36(1):21-24.

Gamer, M., Lemon, J., Fellows, I., Singh, P. 2019. *irr*: Various coefficients of interrater reliability and agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr

Garvin, H.M. and Klales, A.R. 2020. Adult skeletal sex estimation and global standardization. In: Parra R.C., Zapico S.C., and Ubelaker D.H. ed. Forensic science and humanitarian action: Interacting with the dead and the living. United States of America: John Wiley & Sons. p.199-209. https://doi.org/10.1002/9781119482062.ch14

Garvin, H.M., Passalacqua, N.V., Uhl, N.M., Gipson, D.R., Overbury, R.S, Cabo, L.L. 2012. Developments in forensic anthropology: Age-at-death estimation. In: Dirkmaat, D.C. ed. A companion to forensic anthropology. Blackwell Publishing Ltd. p. 202-219. https://doi.org/10.1002/9781118255377.ch10

Gómez-Valdés, J.A., Menéndez Garmendia, A., García-Barzola, L., Sánchez-Mejorada, G., Karam, C., Baraybar, J.P. 2017. Recalibration of the Klales *et al.* (2012) method of sexing the human innominate for Mexican populations. American Journal of Physical Anthropology. 162(3):600-604. https://doi.org/10.1002/ajpa.23157

Huseynov, A., Zollikofer, C.P., Coudyzer, W., Gascho, D., Kellenberger, C., Hinzpeter, R. 2016. Developmental evidence for obstetric adaptation of the human female pelvis. Proceedings of the National Academy of Sciences. 113(19):5227-5232. https://doi.org/10.1073/pnas.1517085113

Johnstone-Belford, E.C. Age and sex estimation based on morphoscopic observations in clinical pelvic CT scans. [Master's dissertation]. The University of Western Australia. 2016

Kelley, M.A. 1979. Parturition and pelvic changes. American Journal of Biological Anthropology. 51(4):541-546. https://doi.org/10.1002/ajpa.1330510405

Kenyhercz, M.W., Klales, A.R., Stull, K.E., McCormick, K.A., Cole, S.J. 2017. Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales *et al.* method. Forensic Science International. 277(1):259e1-259e8. https://doi.org/10.1016/j.forsciint.2017.05.001

Klales, A.R. 2020. MorphoPASSE: Morphological pelvis and skull sex estimation program. In: Sex estimation of the human skeleton. United States of America: Elsevier Academic Press. p. 271-278. https://doi.org/10.1016/B978-0-12-815767-1.00016-X

Klales, A.R. and Burns, T.L. 2017. Adapting and applying the Phenice (1969) adult morphological sex estimation technique to subadults. Journal of Forensic Sciences. 62(3):747-752. https://doi.org/10.1111/1556-4029.13332

Klales, A.R., Ousley, S.D., Vollner, J.M. 2012. A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. American Journal of Biological Anthropology. 149(1):104-114. https://doi.org/10.1002/ajpa.22102

Landis, J.R., Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics. 33(1):159-174. https://doi.org/10.2307/2529310

Langley, N.R., Jantz, L.M., Ousley, S.D., Jantz, R.L. and Milner, G. 2016. Data collection procedures for forensic skeletal material 2.0. University of Tennessee and Lincoln Memorial University.

Lesciotto, K.M. and Doershuk, L.J. 2018. Accuracy and reliability of the Klales *et al.* (2012) morphoscopic pelvic sexing method. Journal of Forensic Sciences. 63(1):214-220. https://doi.org/10.1111/1556-4029.13501

Liaw, A. and Wiener, M. 2002. Classification and regression by randomForest. R news. 2(3):18-22. https://cran.r-project.org/doc/Rnews/

Lovell, N.C. 1989. Test of Phenice's technique for determining sex from the os pubis. American Journal of Biological Anthropology. 79(1):117-120. https://doi.org/10.1002/ajpa.1330790112

MacLaughlin, S.M. and Bruce, M.F. 1990. The accuracy of sex identification in European skeletal remains using the Phenice characters. Journal of Forensic Sciences. 35(6):1384-1392. https://doi.org/10.1520/JFS12974J

Mcbride, D.G., Dietz, M.J., Vennemeyer, M.T., Benfer, R., Furbee, N.L. and Meadors, S. 2001. Bootstrap methods for sex determination from the os coxae using the ID3 algorithm. Journal of Forensic Sciences. 46(3):427-431. https://doi.org/10.1520/JFS14992J

Meindl, R.S., Lovejoy, C.O., Mensforth, R.P. and Walker, R.A. 1985. A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. American Journal of Physical Anthropology. 68(1):29-45. https://doi.org/10.1002/ajpa.1330680104

Mitteroecker, P., Fischer, B. 2016. Adult pelvic shape change is an evolutionary side effect. Proceedings of the National Academy of Sciences. 113(26):E3596. https://doi.org/10.1073/pnas.1607066113

Novotný, V. 1986. Sex determination of the pelvic bone: A systems approach. Anthropologie (1962-). 24(2):197-206. https://www.jstor.org/stable/26294933

Phenice, T.W. 1969. A newly developed visual method of sexing the os pubis. American Journal of Biological Anthropology. 30(2):297-301. https://doi.org/10.1002/ajpa.1330300214

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Internet]. Available from: https://www.R-project.org/.

Rogers, T. and Saunders, S. 1994. Accuracy of sex determination using morphological traits of the human pelvis. Journal of Forensic Sciences. 39(4):1047-1056. http://dx.doi.org/10.1016%2F1353-1131(95)90091-8

Selliah, P., Martino, F., Cummaudo, M., Indra, L., Biehler-Gomez, L., Campobasso, C.P., Cattaneo, C. 2020. Sex estimation of skeletons in middle and late adulthood: Reliability of pelvic morphological traits and long bone metrics on an Italian skeletal collection. International Journal of Legal Medicine. 134:1683-1690. https://doi.org/10.1007/s00414-020-02292-2

Sharma, K., Gupta, P., Shandilya, S. 2016. Age related changes in pelvis size among adolescent and adult females with reference to parturition from Naraingarh, Haryana (India). Homo. 67(4):273-293. https://doi.org/10.1016/j.jchb.2016.04.002

Sutherland, L. D. and Suchey, J. M. Use of the ventral arc in sex determination of the os pubis. Paper presented at the Thirty-ninth Annual Meeting of the American Academy of Forensic Sciences; 1987; San Diego, California.

Tague, R.G. 1989. Variation in pelvic size between males and females. American Journal of Biological Anthropology. 80(1):59-71. https://doi.org/10.1002/ajpa.1330800108

Tague, R.G. 1994. Maternal mortality or prolonged growth: Age at death and pelvic size in three prehistoric Amerindian populations. American Journal of Biological Anthropology. 95(1):27-40. https://doi.org/10.1002/ajpa.1330950103

Todd T.W., 1920. Age changes in the pubic bone. I. The male white pubis. American Journal of Physical Anthropology, 3:285–334. https://doi.org/10.1002/ajpa.1330030301

Ubelaker, D.H. and Volk, C.G. 2002. A test of the Phenice method for the estimation of sex. Journal of Forensic Sciences. 47(1):19-24. https://doi.org/10.1520/JFS15200J

Walker, P.L. 2005. Greater sciatic notch morphology: sex, age, and population differences. American Journal of Biological Anthropology. 127(4):385–391. https://doi.org/10.1002/ajpa.10422

Walker, P.L. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. American Journal of Biological Anthropology. 136(1):39-50. https://doi.org/10.1002/ajpa.20776

Walker, A.R., Walker, B.F., Ncongwane, J. and Tshabalala, E.N. 1984. Age of menopause in black women in South Africa. British Journal of Obstetrics and Gynaecology. 91(8):797-801. https://doi.org/10.1111/j.1471-0528.1984.tb04853.x

Waltenberger, L., Rebay-Salisbury, K., Mitteroecker, P. 2021. Age dependent changes in pelvic shape during adulthood. Journal of Biological and Clinical Anthropology. 79(2):143-156. https://doi.org/10.1127/anthranz/2021/1463

Wells, J.C., Desilva, J.M., Stock, J.T. 2012. The obstetric dilemma: An ancient game of Russian roulette, or a variable dilemma sensitive to ecology? American Journal of Biological Anthropology. 149(S55):40-71. https://doi.org/10.1002/ajpa.2

CHAPTER 6: DISCUSSION AND CONCLUSION

Sex estimation methods have been tested and validated on the South African population and throughout the world. However, few studies have done a thorough assessment of how the classic skeletal sex indicators may be influenced by age and how this may affect the resultant accuracy of the morphoscopic sex estimation methods. The current study looked at two skeletal regions of the same individuals, to get a better understanding of how age may influence the skeleton and what implications this may have on future skeletal analyses.

Differences in robusticity and gracility between males and females is what allows forensic anthropologists to estimate sex from skeletal remains (Cabo *et al.*, 2012). Sex estimation methods are based on the quantification of sexual dimorphism. The different methods are ranked by forensic anthropologists from most to least accurate (i.e., pelvis, long bones, skull, other); however, the choice of method in medicolegal casework is often more reliant on the presence and preservation of the skeletal elements (Klales, 2020). Choice of methodology does not affect the estimate alone as choice of analytical and statistical techniques may also have an influence on the outcome of the sex estimate. The Phenice (1969) and the Walker (2008) traits, which use morphoscopic scoring systems, are ranked as the first and third most accurate methodologies for sex estimation, respectively. However, these methods are known to be subjective and variable as their definitions are not always clear and do not cover the wide range of variation that has been identified within scores. Implementing techniques such as geometric morphometrics (GMM), as opposed to morphoscopics, may allow for less variation in scores and higher repeatability as GMM is better able to quantify the shape and size of sexually dimorphic structures. Furthermore, as seen in the results of the current study, the variation in statistical technique (LR versus DFA versus RFM) can have an effect on the sex estimation accuracy and associated sex bias. Consequently, it is important to compare scenarios where only the independent variable differs, and all the other variables remain constant. Moreover, running additional tests to assess which statistical technique results in the highest positive predictive performance for the population group being studied is also necessary.

A multitude of factors are known to have an influence on the human skeleton. The factors that are more important for the context of this study include physical activity and locomotion, hormone activity, population, sex, and age-at-death. As population specificity and latitude have a significant effect on many skeletal structures and elements it is very important that they are taken into consideration. For example, the global recalibration by Kenyhercz and colleagues

(2017) on the Klales *et al.* (2012) method set out to improve accuracies for a multitude of populations. However, the results of the current study showed that significant differences exist between South African population-affinity groups supporting the need for not only population specific standards but also population-affinity specific standards for South Africans. The results of the current study identified significant differences between the population groups, specifically between black and white South Africans. When analysing the skull, the current study only observed significant population specificity for the mental eminence and nuchal crest between black and white South Africans. Whereas Krüger *et al*. (2015) demonstrated that the sexual dimorphism of all the cranial traits is population-specific and population-specific formulae should be applied. The difference in the results between the studies is unexpected as they were completed on similar samples. The presence of this significant difference in the nuchal crest for both the age-cohorts and the population groups may possibly indicate that it may be more attributed to population influences rather than age-related changes. In such instances, it is important to always corroborate results with multiple skeletal elements and methodology. For this reason, multivariate analysis, and software such as MorphoPASSE (2020) are advantageous since multiple skeletal elements and methods are run simultaneously.

The quantification of sexual dimorphism from the skull and pelvis both utilise morphoscopic methodology, however, the skull and pelvis are sexually dimorphic for vastly different reasons. The pelvis is known to be sexually dimorphic due to its direct involvement in reproduction and parturition (Phenice, 1969; Stewart, 1979; Krogman and İşcan, 1986; Spradley and Jantz, 2011; Betti, 2014; Krishan *et al.*, 2016), while the skull possibly displays sexually dimorphic qualities due to hormonal influences, muscles attachments and robusticity, and mate attraction (Klales, 2021; Kleisner *et al.*, 2021). Unfortunately, although highly accurate for sex estimation, the pelvis is brittle and often exposed to weathering, insect, or carnivore activity in completely skeletonised remains and thus often breaks before or during recovery (Krishan *et al.*, 2016). The skull is not only more durable than the pelvis, but it is also much more recognisable by the general public (Spradley and Jantz, 2011; Krüger *et al.*, 2015). Due to the skulls common appearance and importance in forensic casework much more in-depth studies have been conducted on how it may be affected by other aspects of the biological profile and in particular age-at-death. Although morphoscopic studies on the skull are less popular, the various measurements of the skull have been thoroughly analysed (L'Abbé *et al.*, 2013; Stull *et al.*, 2014; Krüger *et al.*, 2015). The effects of the other aspects of the biological profile on the pelvis have not been as thoroughly analysed and thus it is important that additional research is

undertaken to better understand the pelvis and how it may be affected by other aspects of the biological profile (especially in terms of morphoscopics). This is important as the pelvis is the most sexually dimorphic skeletal element, and it is imperative that the most accurate and reliable sex estimation methodology is used to successfully decrease the number of possible matches to the missing persons list.

The results of the current study identified that the skull is influenced slightly more by increasing age than the pelvis which showed little to no significant differences between the age cohorts. The lack of significant differences identified in the pelvis is surprising as the pelvis is a skeletal element that is commonly utilised in age estimation such as scoring the different phases of the pubic symphysis and the auricular surface of the sacroiliac joint. Thus, the authors suggest that further research is needed with regards to analysing and comparing the different age phases of the pelvis with the Phenice (1969) traits to ascertain whether associations exist or if age related changes to the pelvis is strictly more metrically based. The presence of more morphoscopic age-related changes on the skull compared to the pelvis can potentially be attributed to the vast range of shape and size changes that occur to the skull throughout an individual's life. The decreased effects of age identified on the pubic symphysis in the current study can be attributed to the fact that the pubic symphysis demonstrates less plasticity than other regions of the pelvis and therefore is more constrained and less likely to undergo major changes as a result of age. Additionally, the influence of increasing edentulism with increasing age on the skull cannot be ignored. Patriquin (2013) measured and compared the mechanical forces of mastication for orthognathic and prognathic individuals and the results showed that there are mechanical stress differences that, over time, may cause morphological changes to the cranium. Patriquin (2013) also indicated that bite morphology affects the morphological characteristics of the skull. Patriquin's (2013) findings can be related to the mandibular and maxillary section of the oral cavity since partial and complete edentulism has also been shown to affect the skeletal morphology in a similar way (Small *et al*., 2016; Braun *et al*., 2023). The age-related changes of the skull, paired with the significant presence of partial to complete edentulism in the sample may be the cause for the decreased positive predictive performance of the older sample group when separated from the younger sample group.

Nevertheless, additional studies are required on a number of other factors that have been identified to potentially increase sex estimation error and the associations of age with the skeletal sex indicators. As it was beyond the scope of the study, it is suggested that additional research be conducted on the effects of asymmetry on sex estimates. In medicolegal casework

the left side is favoured by forensic anthropologists (Langley *et al.*, 2016), however, when the left side is unavailable and the right side is used, it is necessary to consider whether the alteration will cause discrepancies in the estimation of sex of the unknown individual (Cole *et al.*, 2020).

Scoring systems used for morphoscopic sex estimation can be subjective (Klales, 2021) as they involve the scoring of a trait based on its degree of expression. The method is based on the comparison of a trait to line drawings with assigned values (i.e., from gracile (1) to robust (5)). The repeatability and reliability of these scoring systems is assessed by establishing whether the initial observer can obtain the same score when scored at a later date (intra-observer error) as well as whether the initial observer can obtain the same score as a second observer (inter-observer error). Overall, the inter- and intra-observer agreement for the skull ranged from no agreement to perfect agreement. However, the level of agreement was similar to Krüger and colleagues (2015) who obtained only slightly higher agreement levels using a comparable sample. Similarly, for the Phenice (1969) traits, the inter- and intra-observer agreement ranged from substantial to near perfect agreement and the agreement levels are comparable to those achieved by Klales *et al.* (2012) and Lesciotto and Doershuk (2018). It is important to consider whether the low Kappa values for the observer agreement may have a potential effect on the outcomes of the study and whether the reason only a few significant differences were identified between the cohorts was due to observer error or age-related differences. Ordinal morphological scoring techniques can be greatly influenced by subjectivity as they depend on the visual assessment of traits (Kemkes-Grottenthaler *et al.*, 2002; Steyn et al., 2004); this can make the methods unreliable as it becomes difficult to apply them to a large sample in a replicable manner (Konigsberg and Hens, 1998). The coding system employed in morphoscopic sex estimation methods is quasi-continuous which means that it resembles a continuous system, ranging from zero to five, but it fails to account for the variation present within each score. The presence of within-score variation poses challenges in defining clear boundaries for each score. For example, what appears to be a ‘small 2’ to one observer may be scored as a ‘large 1’ by another observer, this variation in scoring may result in different outcomes of the sex estimate. Klales *et al.* (2012) stated that the Phenice (1969) traits clearly show sexual dimorphism but the morphoscopic methodology fails to describe the full range of variation identified in the trait expressions. Therefore, to accommodate this variability, Klales (2021) suggested a tolerance system be allowed. The inclusion of a tolerance system may allow a substantial improvement in accuracy which assists in satisfying the Daubert criteria

(Dirkmaat *et al.*, 2008), however, this may further mask variation present between various groups (such as between age cohorts). Walker (2008) utilised a variation of the tolerance system by way of calculating the frequency distribution of the observer's assigned scores and their deviation from the modal score. Lesciotto and Doershuk (2018) tested the accuracy and reliability of the Klales *et al.* (2012) method on the Hamann-Todd Osteological Collection and found that, when a tolerance of one was applied not only did the traits become much more repeatable but the probability of a tolerance of 1 completely changing the estimated sex was only 9% (Lesciotto and Doershuk, 2018). Therefore, Lesciotto and Doershuk (2018) deemed a tolerance of 1 acceptable as there was a low chance that the sex estimate would be affected. It is imperative that the tolerance system is tested and validated so that it could potentially be utilised for the analysis of repeatability and reliability of methods. Additionally, it is important for more detailed definitions, with regards to differentiating between scores (i.e., a "small 2" versus a "large 1"), to be created to allow observers to score the Walker traits consistently and accurately. Updated definitions were provided by Klales (2020) in the MorphoPASSE manual and so the use of these definitions should be explored to see if repeatability increases.

The current study identified that a range of within-score variation exists which has the potential to alter the sex estimate. Therefore, due to the quasi-continuous nature of the scoring system, it is suggested that researchers become experienced with skeletal variation to improve scoring and understanding of the traits. Additionally, researchers should run precursory comparative analyses on the reference population and the study sample to ensure that they contain similar amounts of variation. Alternatively, potentially decreasing the number of phases from five to three may decrease the confusion as a trait then will either be classified as being more gracile, more robust, or neutral. Another important factor to consider is the influence of secular trends. Secular trends are defined as the change that occurs to the mean size or shape of a population that occurs from one generation to the next (Tobias, 1985; Godina, 2009). The current study, as well as most standard-creation and validation studies across the globe, have made use of skeletal collections consisting of remains with dates of birth dating back to the 19th Century. It is assumed by the majority of forensic anthropologists that secular trends, if present, are negligible and few studies have assessed the extent of the presence of secular trends between the skeletal collections and modern individuals (Brůžek *et al.*, 2012). However, in order to maintain accuracy and record high levels of repeatability and reliability, the presence of secular trends needs to be considered, as significant differences may develop as a result (Klales, 2021). Klales (2016) assessed the effects of secular trends on the Phenice (1969) traits

and discovered that significant differences were present for all of the traits except for the MIR in males. Accuracies for the modern collection were lower than the historic collection, however, Klales (2016) stated that this may be due to a greater mean age in the modern sample. Therefore, it is imperative that the effects of age and secular trends are analysed both individually and in conjunction to assess whether sex estimation accuracies are influenced. Furthermore, as previously mentioned, the skeletal collections utilised for the creation of standards and validation of methods worldwide are largely comprised of individuals who were of lower SES prior to death (de la Cova, 2010; L'Abbé *et al*., 2021). Lower levels of SES have been correlated to greater environmental stresses and ultimately malnutrition which are known to result in decreased body size. Additionally, as stated by the female buffering hypothesis, males are more vulnerable to the effects of environmental stresses than females. As described by this hypothesis males would experience more extreme effects of the stresses than females resulting in decreased sexual dimorphism (Nikitovic and Bogin, 2014).

Unfortunately, despite all efforts, the study faced limitations that were unavoidable. The limitations included the distribution of age and population affinity in the skeletal collections as well as the quasi-continuous nature of the scoring systems. The age distribution in the skeletal collections is limited and majority of the individuals have an age-at-death that is middle- to old-aged. Furthermore, there is a considerable lack of individuals that are below 20 years of age, subadult, or above 90 years of age. Additionally, the population affinity throughout the age cohorts is not evenly distributed and BSAs tend to fall more into the young to middle-aged cohorts whereas WSAs tend to fall into the middle-aged to old cohorts. The unequal distribution prevents the comparison of age-related changes of the sexes between population groups. CSAs are also not widely represented in the skeletal collections, and this prevents accurate analysis of the population group due to small sample sizes.

In conclusion, based on the results of the current study, age does not need to be factored in or pre-selected prior to the estimation of sex for either the skull or the os coxae. While age-related differences are present, the accuracy is not affected to the extent to require the creation of new methodology and formulae. However, additional studies are required on multiple other related aspects (asymmetry, secular trends, GMM, pelvic phase, etc.) in order to exhaust all possible influences of age on the positive predictive performance of the sex estimate.

REFERENCES

Acsádi, G.R., Nemeskéri, J.N. 1970. History of human life span and mortality, Budapest, Akadémiai Kiadó. 346. https://www.jstor.org/stable/2740808

Albert, A.M., Ricanek, K., Patterson, E. 2007. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. Forensic Science International. 172(1):1-9. https://doi.org/10.1016/j.forsciint.2007.03.015

Alblas, A., Greyling, L.M., Geldenhuys, E.-M. 2018. Composition of the Kirsten Skeletal Collection at Stellenbosch University. South African Journal of Science. 114(1-2):1-6. https://doi.org/10.17159/sajs.2018/20170198

Ali, Z., and Bhaskar, S.B. 2016. Basic statistical tools in research and data analysis. Indian Journal of Anaesthesia. 60(9):662. https://doi.org/10.4103%2F0019-5049.190623

American Academy of Forensic Sciences Standards Board. 2019. Standard 090: Standard for sex estimation in forensic anthropology. AAFS Standards Board, Colorado Springs, CO. 1:1-4.

Auerbach, B.M., King, K.A., Campbell, R.M., Campbell, M.L., Sylvester, A.D. 2018. Variation in obstetric dimensions of the human bony pelvis in relation to age-at-death and latitude. American Journal of Biological Anthropology. 167(3):628-643. https://doi.org/10.1002/ajpa.23690

Benjamin, F. 1960. The age of the menarche and of the menopause in white South African women and certain factors influencing these times. South African Medical Journal. 34(16):316-320.

Berger, A.A., May, R., Renner, J.B., Viradia, N., Dahners, L.E. 2011. Surprising evidence of pelvic growth (widening) after skeletal maturity. Journal of Orthopaedic Research. 29(11):1719-1723. https://doi.org/10.1002/jor.21469

Betti, L. 2014. Sexual dimorphism in the size and shape of the os coxae and the effects of microevolutionary processes. American Journal of Biological Anthropology. 153(2):167-177. https://doi.org/10.1002/ajpa.22410

Braun, S., Ridel, A.F., L'Abbé, E.N., Oettlé, A.C. 2023. Analysis of the hard-tissue menton shape variation in adult South Africans using cone-beam computed tomography (CBCT) scans. Forensic Imaging. 32:1-7. https://doi.org/10.1016/j.fri.2023.200532

Brooks, S., and Suchey, J.M. 1990. Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. Human Evolution. 5:227-238. https://doi.org/10.1007/BF02437238

Brůžek, J., Guyomarc'h, P., Veleminske, J., Svenkrtova, I., Chaumoitre, K. and Mestekova, S. 2012. Impact of secular trends on the reliability of sex assessment methods from the skeleton: contribution of the virtual anthropology. SLOVENSKÁ ANTROPOLÓGIA. 15(1):5-11

Buikstra, J.E. 2019. Ortner's identification of pathological conditions in human skeletal remains. 3rd edn. London, UK: Elsevier

Cabo, L.L., Brewster, C.P., Azpiazu, J.L. 2012. Sexual dimorphism: Interpreting sex markers. In: Dirkmaat, D.C. ed. A companion to forensic anthropology. Blackwell Publishing Ltd. p. 248-286. https://doi.org/10.1002/9781118255377.ch13

Callewaert, F., Sinnesael, M., Gielen, E., Boonen, S., Vanderschueren, D. 2010. Skeletal sexual dimorphism: Relative contribution of sex steroids, growth hormone-insulin-like growth factor-i (gh-igf-i) and mechanical loading. Journal of Endocrinology. 207(2):127-134.

Caple, J. and Stephan, C.N. 2017. Photo-realistic statistical skull morphotypes: new exemplars for ancestry and sex estimation in forensic anthropology. Journal of Forensic Sciences. 62(3):562-572. https://doi.org/10.1111/1556-4029.13314

Cole, S.J., Hulse, C.N. and Stull, K.E. 2020. The effects of skeletal asymmetry on accurate sex classification. In: Sex estimation of the human skeleton. San Diego, US: Elsevier Academic Press. p. 307-325. https://doi.org/10.1016/B978-0-12-815767-1.00019-5

Corron, L.K., Santos, F., Adalian, P., Chaumoitre, K., Guyomarc'h, P., Marchal, F., Brůžek, J. 2021. How low can we go? A skeletal maturity threshold for probabilistic visual sex estimation from immature human os coxae. Forensic Science International. 325:1-15. https://doi.org/10.1016/j.forsciint.2021.110854

Cox, M., Mays, S., 2000. Human osteology: In archaeology and forensic science [Internet]. London: Greenwich Medical Media. [cited 2022 March 31] Available from ResearchGate: https://www.researchgate.net/publication/235692961_Human_Osteology_in_Archaeology_and_Forensic_Science

Dayal, M.R., Kegley, A.D., Štrkalj, G., Bidmos, M.A., Kuykendall, K.L. 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 Biological Anthropology. 140(2):324-335. https://doi.org/10.1002/ajpa.21072

de la Cova, C. 2010. Cultural patterns of trauma among 19th-century-born males in cadaver collections. American Anthropologist. 112(4):589-606. https://doi.org/10.1111/j.1548-1433.2010.01278.x

de la Paz, J.S., Buckley, H.R., Halcrow, S.E., Techataweewan, N. and Woodley, S.J. 2023. Architecture of head and neck soft tissues and associated entheses: An exploration of sexual dimorphism in, and population differences between, New Zealand and Thai individuals. Journal of Anatomy. 243(1):1-18. https://doi.org/10.1111/joa.13853

DesMarais, A., Obertova, Z. and Franklin, D. 2023. The influence of age on greater sciatic notch morphology: Testing the Walker method in an Australian population. International Journal of Legal Medicine. 1-9. https://doi.org/10.1007/s00414-023-02988-1

Dirkmaat, D.C. 2012. A companion to forensic anthropology [Internet]. Malden, MA: Wiley-Blackwell. Available from Wiley Online Library: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118255377#page=38

Dirkmaat, D.C., Cabo, L.L., Ousley, S.D. and Symes, S.A. 2008. New perspectives in forensic anthropology. American Journal of Physical Anthropology. 137(S47):33-52. https://doi.org/10.1002/ajpa.20948

Flaherty, T.M., Johnson, L.J., Woollen, K.C., Lopez, D., Gaddis, K., Horsley, S.L., Byrnes, J.F. 2023. Speaking of sex: Critical reflections for forensic anthropologists. Humans. 3(4):251-270. https://doi.org/10.3390/humans3040020

Franklin, D. 2010. Forensic age estimation in human skeletal remains: Current concepts and future directions. Legal Medicine. 12(1):1-7. https://doi.org/10.1016/j.legalmed.2009.09.001

Frere, G. 1971. Mean age at menopause and menarche in South Africa. The South African Journal of Medical Sciences. 36(1):21-24.

Gapert, R., Black, S., Last, J. 2013. Test of age-related variation in the craniometry of the adult human foramen magnum region: Implications for sex determination methods. Forensic Science, Medicine, and Pathology. 9(4):478-488. https://doi.org/10.1007/s12024-013-9437-3

Gamer, M., Lemon, J., Fellows, I., Singh, P. 2019. *irr*: Various coefficients of interrater reliability and agreement. R package version 0.84.1. https://CRAN.R-project.org/package=irr

Garvin, H.M. and Klales, A.R. 2020. Adult skeletal sex estimation and global standardization. In: Parra R.C., Zapico S.C., and Ubelaker D.H. ed. Forensic science and humanitarian action: Interacting with the dead and the living. United States of America: John Wiley & Sons. p.199-209. https://doi.org/10.1002/9781119482062.ch14

Garvin, H.M., Passalacqua, N.V., Uhl, N.M., Gipson, D.R., Overbury, R.S, Cabo, L.L. 2012. Developments in forensic anthropology: Age-at-death estimation. In: Dirkmaat, D.C. ed. A companion to forensic anthropology. Blackwell Publishing Ltd. p. 202-219. https://doi.org/10.1002/9781118255377.ch10

Garvin, H.M., Sholts, S.B., Mosca, L.A. 2014. Sexual dimorphism in human cranial trait scores: Effects of population, age, and body size. American Journal of Biological Anthropology. 154(2):259-269. https://doi.org/10.1002/ajpa.22502

Godina, E.Z. 2009. The secular trend: History and prospects. Human Physiology. 35(6):770-776. https://doi.org/10.1134/S0362119709060152

Gómez-Valdés, J.A., Menéndez Garmendia, A., García-Barzola, L., Sánchez-Mejorada, G., Karam, C., Baraybar, J.P. 2017. Recalibration of the Klales *et al.* (2012) method of sexing the human innominate for Mexican populations. American Journal of Physical Anthropology. 162(3):600-604. https://doi.org/10.1002/ajpa.23157

Grados, F., Fardellone, P., Benammar, M., Muller, C., Roux, C. and Sebert, J. 1999. Influence of age and sex on vertebral shape indices assessed by radiographic morphometry. Osteoporosis International. 10:450-455. https://doi.org/10.1007/s001980050253

Greene, M.M. Musculoskeletal markers: A comparison of the influence of three biological factors on robusticity [dissertation]. Louisiana State University and Agricultural & Mechanical College; 2022. https://digitalcommons.lsu.edu/gradschool_theses/5562

Houston, S. Validating the use of the mastoid triangle for the estimation of sex and ancestry in black and white South Africans [BSc (Hons) dissertation]. Pretoria. University of Pretoria; 2021.

Huseynov, A., Zollikofer, C.P., Coudyzer, W., Gascho, D., Kellenberger, C., Hinzpeter, R. 2016. Developmental evidence for obstetric adaptation of the human female pelvis. Proceedings

of the National Academy of Sciences. 113(19):5227-5232. https://doi.org/10.1073/pnas.1517085113

Johnstone-Belford, E.C. Age and sex estimation based on morphoscopic observations in clinical pelvic CT scans. [Master's dissertation]. The University of Western Australia. 2016.

Juul, A. 2001. The effects of oestrogens on linear bone growth. Human Reproduction Update. 7(3):124-134.

Kelley, M.A. 1979. Parturition and pelvic changes. American Journal of Biological Anthropology. 51(4):541-546. https://doi.org/10.1002/ajpa.1330510405

Kemkes-Grottenthaler, A., Löbig, F., Stock, F. 2002. Mandibular ramus flexure and gonial eversion as morphologic indicators of sex. Homo. 53(2):97-111. https://doi.org/10.1078/0018-442X-00039

Kenyhercz, M.W., Klales, A.R., Stull, K.E., McCormick, K.A., Cole, S.J. 2017. Worldwide population variation in pelvic sexual dimorphism: A validation and recalibration of the Klales *et al.* method. Forensic Science International. 277(1):259e1-259e8. https://doi.org/10.1016/j.forsciint.2017.05.001

Klales, A.R. 2016. Secular change in morphological pelvic traits used for sex estimation. Journal of Forensic Sciences. 61(2):295-301. https://doi.org/10.1111/1556-4029.13008

Klales, A.R. 2020. MorphoPASSE: Morphological pelvis and skull sex estimation program. In: Sex estimation of the human skeleton. United States: Elsevier Academic Press. p. 271-278. https://doi.org/10.1016/B978-0-12-815767-1.00016-X

Klales, A.R. 2020. Practitioner preferences for sex estimation from human skeletal remains. Sex estimation of the human skeleton. Elsevier. https://doi.org/10.1016/B978-0-12-815767-1.00002-X

Klales, A.R. 2021. Current state of sex estimation in forensic anthropology. Forensic Anthropology. 4(2):118-127. https://10.5744/fa.2020.3033

Klales, A.R. and Burns, T.L. 2017. Adapting and applying the Phenice (1969) adult morphological sex estimation technique to subadults. Journal of Forensic Sciences. 62(3):747-752. https://doi.org/10.1111/1556-4029.13332

Klales, A.R., Ousley, S.D., Vollner, J.M. 2012. A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. American Journal of Biological Anthropology. 149(1):104-114. https://doi.org/10.1002/ajpa.22102

Kleisner, K., Tureček, P., Roberts, S.C., Havlíček, J., Valentova, J.V., Akoko, R.M., Leongómez, J.D., Apostol, S., Varella, M.A.C., Saribay, S.A. 2021. How and why patterns of sexual dimorphism in human faces vary across the world. Scientific Reports. 11(1):1-14. https://doi.org/10.1038/s41598-021-85402-3

Konigsberg, L.W. and Hens, S.M. 1998. Use of ordinal categorical variables in skeletal assessment of sex from the cranium. American Journal of Biological Anthropology. 107(1):97-112. https://doi.org/10.1002/(SICI)1096-8644(199809)107:1%3C97::AID-AJPA8%3E3.0.CO;2-A

Krishan, K., Chatterjee, P.M., Kanchan, T., Kaur, S., Baryah, N. and Singh, R. 2016. A review of sex estimation techniques during examination of skeletal remains in forensic anthropology casework. Forensic Science International. 261:165.e1-165.e8. http://dx.doi.org/10.1016/j.forsciint.2016.02.007

Krogman, W.M. and İşcan, M.Y. 2013. The human skeleton in forensic medicine [Internet]. Springfield, Illinois, U.S.A.: Thomas, C.C. [cited 2023 Jun 2]. pp.143-193. Available from: https://search-ebscohost-com.uplib.idm.oclc.org/login.aspx?direct=true&db=nlebk&AN=646615&site=ehost-live&scope=site

Krüger, G.C., L'Abbé, E.N., Stull, K.E., Kenyhercz, M.W. 2015. Sexual dimorphism in cranial morphology among modern South Africans. International Journal of Legal Medicine. 129(4):869-875. https://doi.org/10.1007/s00414-014-1111-0

Krüger, G.C., L'Abbé, E.N., Stull, K.E. 2017. Sex estimation from the long bones of modern South Africans. International Journal of Legal Medicine. 131(1):275-285. https://doi.org/10.1007/s00414-016-1488-z

L'Abbé, E.N, Loots, M., Meiring, J. 2005. The Pretoria Bone Collection: A modern South African skeletal sample. Homo. 56(2):197-205. https://doi.org/10.1016/j.jchb.2004.10.004

L'Abbé, E.N., Kenyhercz, M., Stull, K.E., Keough, N., Nawrocki, S. 2013. Application of FORDISC 3.0 to explore differences among crania of North American and South African

blacks and whites. Journal of Forensic Sciences. 58(6):1579-1583. https://doi.org/10.1111/1556-4029.12198

L'Abbé, E.N., Krüger, G.C., Theye, C.E., Hagg, A.C., Sapo, O. 2021. The Pretoria Bone Collection: A 21st century skeletal collection in South Africa. Forensic Sciences. 1(3):220-227. https://doi.org/10.3390/forensicsci1030020

Landis, J.R., Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics. 33(1):159-174. https://doi.org/10.2307/2529310

Langley, N.R., Jantz, L.M., Ousley, S.D., Jantz, R.L. and Milner, G. 2016. Data collection procedures for forensic skeletal material 2.0. University of Tennessee and Lincoln Memorial University.

Lesciotto, K.M. and Doershuk, L.J. 2018. Accuracy and reliability of the Klales *et al.* (2012) morphoscopic pelvic sexing method. Journal of Forensic Sciences. 63(1):214-220. https://doi.org/10.1111/1556-4029.13501

Lesciotto, K.M., Doershuk, L.J. 2018. Effect of age on nonmetric cranial traits for sex estimation in adults. Forensic Anthropology. 1(3):150-159. https://doi.org/10.5744/fa.2018.0016

Lewis, C.J. and Garvin, H.M. 2016. Reliability of the walker cranial nonmetric method and implications for sex estimation. Journal of Forensic Sciences. 61(3):743-751. https://doi.org/10.1111/1556-4029.13013

Liaw, A. and Wiener, M. 2002. Classification and regression by randomForest. R news. 2(3):18-22. https://cran.r-project.org/doc/Rnews/

Lovell, N.C. 1989. Test of Phenice's technique for determining sex from the os pubis. American Journal of Biological Anthropology. 79(1):117-120. https://doi.org/10.1002/ajpa.1330790112

MacLaughlin, S.M. and Bruce, M.F. 1990. The accuracy of sex identification in European skeletal remains using the Phenice characters. Journal of Forensic Sciences. 35(6):1384-1392. https://doi.org/10.1520/JFS12974J

Mcbride, D.G., Dietz, M.J., Vennemeyer, M.T., Benfer, R., Furbee, N.L. and Meadors, S. 2001. Bootstrap methods for sex determination from the os coxae using the ID3 algorithm. Journal of Forensic Sciences. 46(3):427-431. https://doi.org/10.1520/JFS14992J

Meindl, R.S., Lovejoy, C.O., Mensforth, R.P. and Walker, R.A. 1985. A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. American Journal of Physical Anthropology. 68(1):29-45. https://doi.org/10.1002/ajpa.1330680104

Mercer, S.R. and Bogduk, N. 2003. Clinical anatomy of ligamentum nuchae. Clinical Anatomy: The Official Journal of the American Association of Clinical Anatomists and the British Association of Clinical Anatomists. 16(6):484-493. https://doi.org/10.1002/ca.10121

Mitteroecker, P., Fischer, B. 2016. Adult pelvic shape change is an evolutionary side effect. Proceedings of the National Academy of Sciences. 113(26):E3596. https://doi.org/10.1073/pnas.1607066113

Nikitovic, D. and Bogin, B. 2014. Ontogeny of sexual size dimorphism and environmental quality in Guatemalan children. American Journal of Human Biology. 26(2):117-123. https://doi.org/10.1002/ajhb.22492

Novotný, V. 1986. Sex determination of the pelvic bone: A systems approach. Anthropologie (1962-). 24(2):197-206. https://www.jstor.org/stable/26294933

Patriquin, M. L. 2013. The relationship between masticatory stress and prognathism: a finite element and morphometric study (PhD Dissertation). University of Pretoria. http://hdl.handle.net/2263/36682

Petaros, A., Sholts, S.B., Slaus, M., Bosnar, A., Wärmländer, S.K.T.S. 2015. Evaluating sexual dimorphism in the human mastoid process: A viewpoint on the methodology. Clinical Anatomy. 28(5):593-601. https://doi.org/10.1002/ca.22545

Phenice, T.W. 1969. A newly developed visual method of sexing the os pubis. American Journal of Biological Anthropology. 30(2):297-301. https://doi.org/10.1002/ajpa.1330300214

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Internet]. Available from: https://www.R-project.org/.

Rogers, S.L. 1982. The aging skeleton: Aspects of human bone involution. Springfield, Illinois: Thomas C.C.

Rogers, T. and Saunders, S. 1994. Accuracy of sex determination using morphological traits of the human pelvis. Journal of Forensic Sciences. 39(4):1047-1056. http://dx.doi.org/10.1016%2F1353-1131(95)90091-8

Ross, A.H., Williams, S.E. 2010. Craniofacial growth, maturation, and change: teens to midadulthood. The Journal of Craniofacial Surgery. 21(2):458-461. https:/doi.org/10.1097/SCS.0b013e3181cfea34

Santos, F., Guyomarc'h, P., Rmoutilova, R., Bruzek, J. 2019. A method of sexing the human os coxae based on logistic regressions and Bruzek's nonmetric traits. American Journal of Biological Anthropology. 169(3):435-447. https://doi.org/10.1002/ajpa.23855

Selliah, P., Martino, F., Cummaudo, M., Indra, L., Biehler-Gomez, L., Campobasso, C.P., Cattaneo, C. 2020. Sex estimation of skeletons in middle and late adulthood: Reliability of pelvic morphological traits and long bone metrics on an Italian skeletal collection. International Journal of Legal Medicine. 134:1683-1690. https://doi.org/10.1007/s00414-020-02292-2

Sharma, K., Gupta, P., Shandilya, S. 2016. Age related changes in pelvis size among adolescent and adult females with reference to parturition from Naraingarh, Haryana (India). Homo. 67(4):273-293. https://doi.org/10.1016/j.jchb.2016.04.002

Small, C., Brits, D., Hemingway, J. 2016. Assessing the effects of tooth loss in adult crania using geometric morphometrics. International Journal of Legal Medicine. 130(1):233-243. https://doi.org/10.1007/s00414-015-1174-6

Spradley, M.K., Jantz, R.L. 2011. Sex estimation in forensic anthropology: Skull versus postcranial elements. Journal of Forensic Sciences. 56(2):289-296. https://doi.org/10.1111/j.1556-4029.2010.01635.x

Stewart, T.D. 1979. Essentials of forensic anthropology: Especially as developed in the United States, Charles C. Thomas Springfield, IL.

Steyn, M., Becker, P.J., L'Abbé, E.N., Scholtz, Y. and Myburgh, J. 2012. An assessment of the repeatability of pubic and ischial measurements. Forensic Science International. 214(1-3):210.e1-2010.e4. https://doi.org/10.1016/j.forsciint.2011.07.049

Steyn, M., Pretorius, E. and Hutten, L. 2004. Geometric morphometric analysis of the greater sciatic notch in South Africans. Homo. 54(3):197-206. https://doi.org/10.1078/0018-442x-00076

Stull, K.E., Kenyhercz, M.W. and L'Abbé, E.N. 2014. Ancestry estimation in South Africa using craniometrics and geometric morphometrics. Forensic Science International. 245:206.e1-206.e7. https://doi.org/10.1016/j.forsciint.2014.10.021

Sutherland, L. D. and Suchey, J. M. Use of the ventral arc in sex determination of the os pubis. Paper presented at the Thirty-ninth Annual Meeting of the American Academy of Forensic Sciences; 1987; San Diego, California.

Tague, R.G. 1989. Variation in pelvic size between males and females. American Journal of Biological Anthropology. 80(1):59-71. https://doi.org/10.1002/ajpa.1330800108

Tague, R.G. 1994. Maternal mortality or prolonged growth: Age at death and pelvic size in three prehistoric Amerindian populations. American Journal of Biological Anthropology. 95(1):27-40. https://doi.org/10.1002/ajpa.1330950103

Tallman, S.D., Kincer, C.D. and Plemons, E.D. 2021. Centering transgender individuals in forensic anthropology and expanding binary sex estimation in casework and research. Forensic Anthropology. 5(2):161-180. http://dx.doi.org/10.5744/fa.2020.0030

Tobias, P.V. 1985. The negative secular trend. Journal of Human Evolution. 14(4):347-356. https://doi.org/10.1016/S0047-2484(85)80041-5

Todd T.W., 1920. Age changes in the pubic bone. I. The male white pubis. American Journal of Physical Anthropology, 3:285–334. https://doi.org/10.1002/ajpa.1330030301

Ubelaker, D.H. and Volk, C.G. 2002. A test of the Phenice method for the estimation of sex. Journal of Forensic Sciences. 47(1):19-24. https://doi.org/10.1520/JFS15200J

van Wyk, P.J., van Wyk, C. 2004. Oral health in South Africa. International Dental Journal. 54(6):373-377. https://doi.org/10.1111/j.1875-595X.2004.tb00014.x

Veldhuis, J.D., Roemmich, J.N., Richmond, E.J., Rogol, A.D., Lovejoy, J.C., Sheffield-Moore, M., Mauras, N., Bowers, C.Y. 2005. Endocrine control of body composition in infancy, childhood, and puberty. Endocrine Reviews. 26(1):114-146. https://doi.org/10.1210/er.2003-0038

Walker, P.L. 1995. Problems of preservation and sexism in sexing: Some lessons from historical collections for palaeodemographers. In: Saunders, S.R. and Herring, Grave reflections: portraying the past through cemetery studies. Toronto: Canadian Scholars' Press. p. 31-47.

Walker, P.L. 2005. Greater sciatic notch morphology: sex, age, and population differences. American Journal of Biological Anthropology. 127(4):385–391. https://doi.org/10.1002/ajpa.10422

Walker, P.L. 2008. Sexing skulls using discriminant function analysis of visually assessed traits. American Journal of Biological Anthropology. 136(1):39-50. https://doi.org/10.1002/ajpa.20776

Walker, A.R., Walker, B.F., Ncongwane, J. and Tshabalala, E.N. 1984. Age of menopause in black women in South Africa. British Journal of Obstetrics and Gynaecology. 91(8):797-801. https://doi.org/10.1111/j.1471-0528.1984.tb04853.x

Waltenberger, L., Rebay-Salisbury, K., Mitteroecker, P. 2021. Age dependent changes in pelvic shape during adulthood. Journal of Biological and Clinical Anthropology. 79(2):143-156. https://doi.org/10.1127/anthranz/2021/1463

Wells, J.C., Desilva, J.M., Stock, J.T. 2012. The obstetric dilemma: An ancient game of Russian roulette, or a variable dilemma sensitive to ecology? American Journal of Biological Anthropology. 149(S55):40-71. https://doi.org/10.1002/ajpa.20776

White, T.D., Folkens, P.A. 2005. The human bone manual. California: Elsevier Academic Press.

APPENDIX A: SCORING SHEETS

Refer to Figure A1 and Figure A2 for the guidelines related to the scoring of the Walker and Phenice traits respectively.

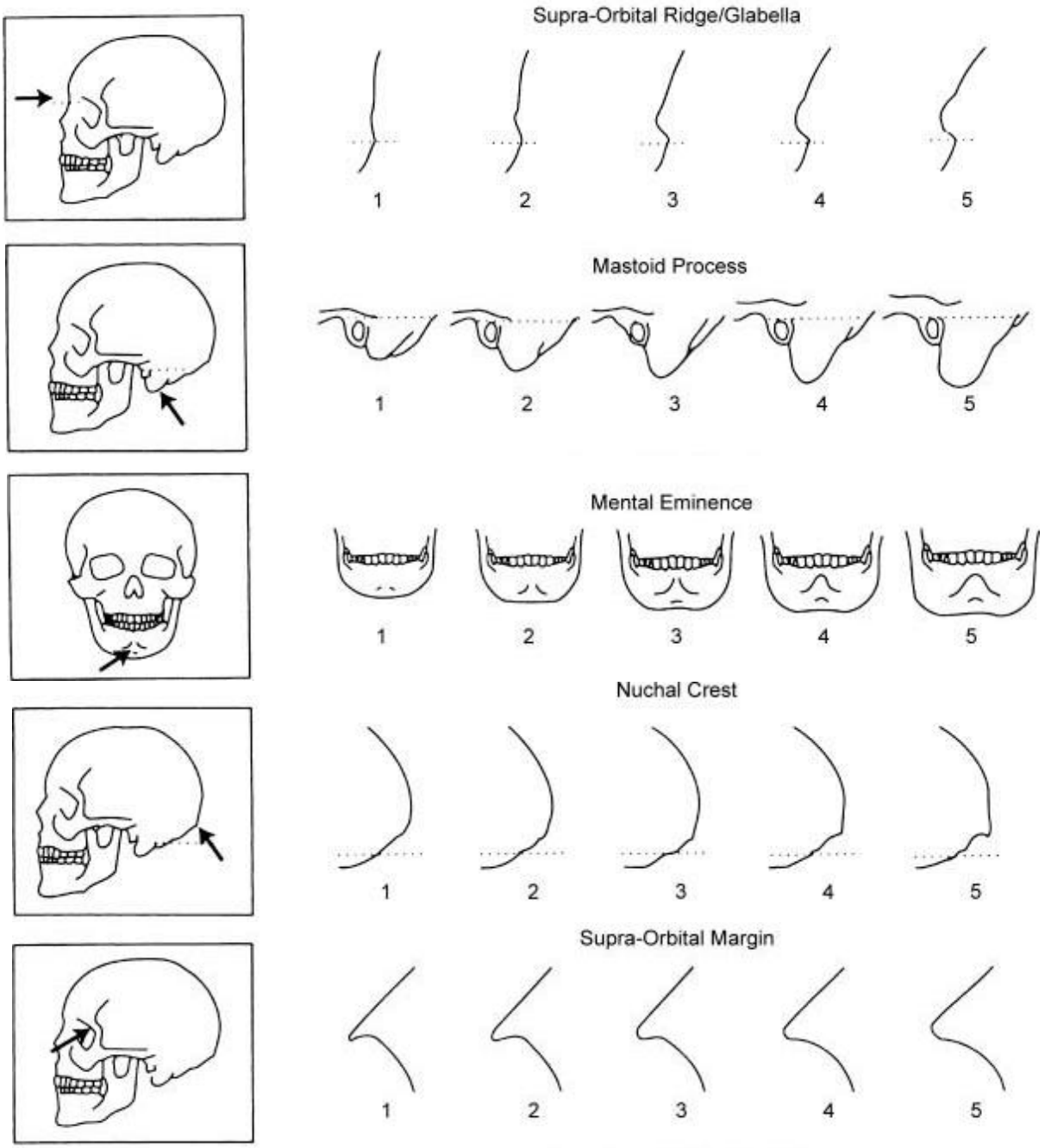


Figure A1. The Walker (2008) traits and associated line drawings (Taken from Walker, 2008:41)

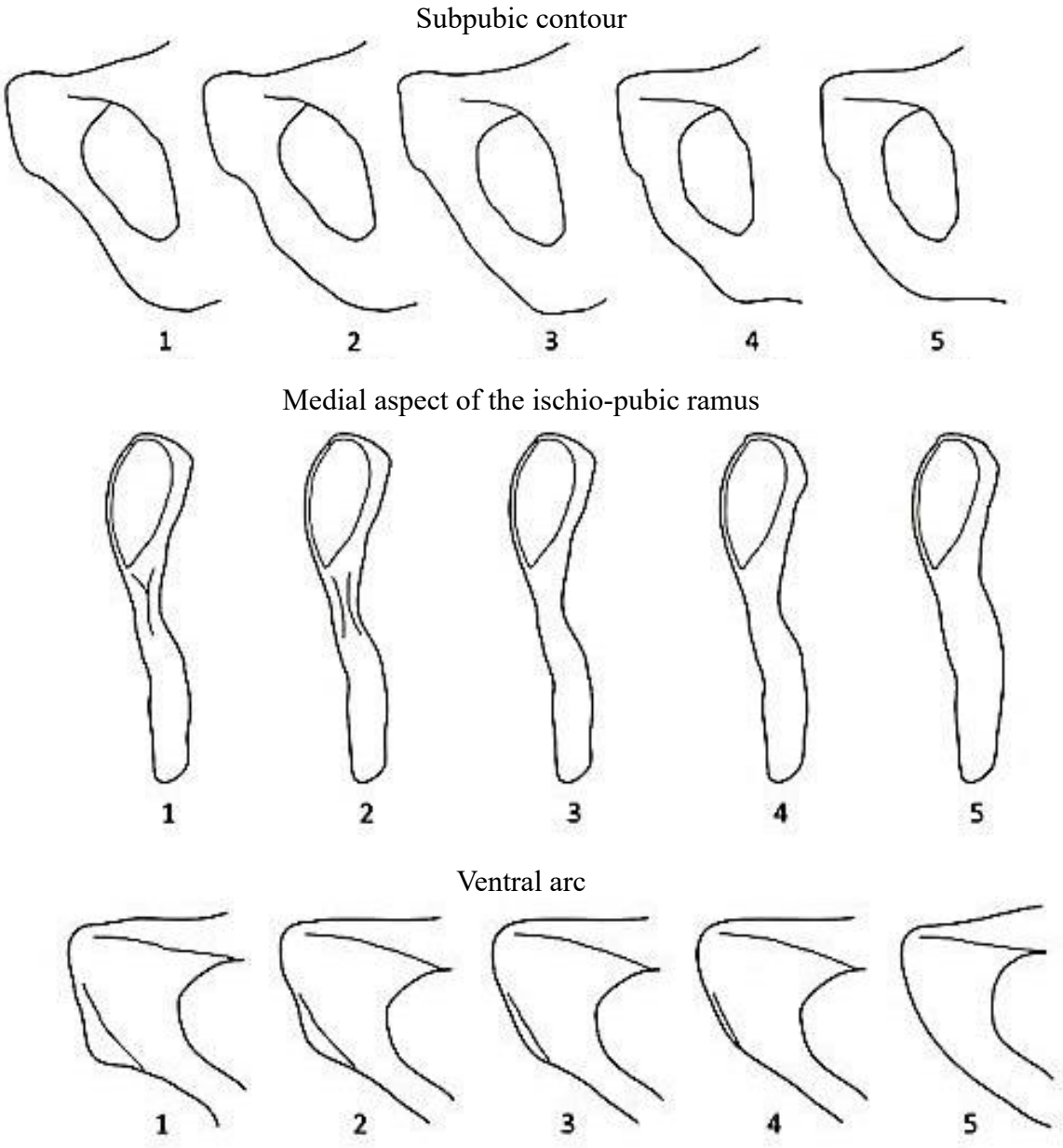


Figure A2. The Phenice (1969) traits and associated line drawings proposed by Klales *et al*. (2012) (Taken from Klales *et al.*, 2012:107).

APPENDIX B: TRAIT DEFINITIONS

For each landmark scored, the skull and os coxa was held in the required plane as shown in the scoring sheets (Figures A1 – A2).

Table B1. Table of trait definitions and descriptions.

Term	Definition	Scoring**
Walker (2008) Traits (skull)		
Glabella	A type II* landmark at the most anteriorly projecting part of the lower margin of the frontal bone in the mid-sagittal plane at the level of the supra-orbital ridge (White and Folkens, 2005; Small *et al.*, 2016). To score the glabella the skull was positioned in the Frankfort Horizontal Plane.***	1-5
Mastoid process	A projection on the cranium that extends from the temporal bone located posterior to the external auditory meatus (Petaros *et al.*, 2015). To score the mastoid process the skull was positioned in the Frankfort Horizontal Plane.	1-5
Mental eminence	A raised triangular protuberance on the medial anterior corpus of the mandible. Prominence and size are increased in male individuals (White and Folkens, 2005). To score the mental eminence the mandible was viewed from anteriorly.	1-5
Nuchal crest	The most posterior projecting roughened crest(s) of bone present on the posteromedial surface of the occipital bone (White and Folkens, 2005; Small *et al.*, 2016). To score the nuchal crest the skull was positioned in the Frankfort Horizontal Plane.	1-5
Supra-orbital margin	The superolateral margin of the orbit when observed from an inferior perspective. Appears blunt in males and sharp in females (White and Folkens, 2005). To score the supra-orbital margins the skull was viewed from inferiorly.	1-5

Table B1 continued…

Term	Definition	Scoring**
Phenice (1969) Traits (os coxa)		
Sub-pubic contour	Males: a lateral convexity or straight contour in the ischio-pubic ramus which occurs just below the lower margin of the pubic symphysis (Phenice, 1969). Females: a lateral concavity in the ischio-pubic ramus which occurs just below the lower margin of the pubic symphysis (Phenice, 1969). To score the Sub-pubic Contour the os coxa needs to be positioned with the dorsal surface towards the observer.	1-5
Medial aspect of the ischio-pubic ramus	Males: present as a broad surface found immediately below the symphyseal surface on the medial aspect of the ischio-pubic ramus (Phenice, 1969). Females: present as a sharp ridge found immediately below the symphyseal surface on the medial aspect of the ischio-pubic ramus (Phenice, 1969). To score the Medial Aspect of the Ischio-pubic Ramus the os coxa needs to be positioned with the medial surface (with the pubic symphysis) towards the observer.	1-5
Ventral arc	An elevated ridge of bone that arcs inferiorly from the pubic crest, towards the most lateral continuation of the subpubic concavity where it then merges with the medial aspect of the ischio-pubic ramus. Only present in females (Phenice, 1969). To score the Ventral Arc the os coxa needs to be positioned with the ventral surface towards the observer (acetabulum towards the observer).	1-5

*A type II landmark is defined as a point of maximum curvature.

**With 1 being more feminine and 5 being more masculine.

***A plane aligning the most inferior point of the lower margin of the orbit with the most superior point of the external auditory meatus.

APPENDIX C: ETHICAL APPROVAL

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Health Sciences

Institution: 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 18 March 2022 and Expires 18 March 2027.
- IORG #: IORG0001762 OMB No. 0990-0278 Approved for use through August 31, 2023

Faculty of Health Sciences **Research Ethics Committee**

1 February 2023

Approval Certificate
New Application

Dear Ms S Houston

Ethics Reference No.: 588/2022
Title: Assessment of the associations of age with the classic skeletal sex indicators

The **New Application** as supported by documents received between 2022-11-29 and 2023-02-01 for your research, was approved by the Faculty of Health Sciences Research Ethics Committee on 2023-02-01 as resolved by its quorate meeting.

Please note the following about your ethics approval:

- Ethics Approval is valid for 1 year and needs to be renewed annually by 2024-02-01.
- Please remember to use your protocol number (588/2022) on any documents or correspondence with the Research Ethics Committee regarding your research.
- Please note that the Research Ethics Committee may ask further questions, seek additional information, require further modification, monitor the conduct of your research, or suspend or withdraw ethics approval.

Ethics approval is subject to the following:

- The ethics approval is conditional on the research being conducted as stipulated by the details of all documents submitted to the Committee. In the event that a further 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.

We wish you the best with your research.

Yours sincerely

On behalf of the FHS REC, Dr R Sommers
MBChB, MMed (Int), MPharmMed, PhD
***Deputy Chairperson** of the Faculty of Health Sciences Research Ethics Committee, University of Pretoria*

The Faculty of Health Sciences Research Ethics Committee complies with the SA National Act 61 of 2003 as it pertains to health research and the United States Code of Federal Regulations Title 46 and 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes, Second Edition 2016 (Department of Health)

Research Ethics Committee
Room 1.69, Level 1, Tswelopele Building
University of Pretoria, Private Bag x323
Gezina 0031, South Africa
Tel: +27 (0)12 356 3084
Email: deepeka.behari@up.ac.za
www.up.ac.za

Fakulteit Gesondheidswetenskappe
Lefapha la Litsebene tsa Maphelo

Figure C1. Ethical Approval Certificate

APPENDIX D: FREQUENCY DISTRIBUTIONS

Table D1. Frequency distribution of partial and complete edentulism.

Cohort	Individuals with Partial Edentulism	Individuals with Complete Edentulism	Number of Individuals with any Edentulism	*n*	Percent with any Edentulism (%)
<20	1	0	1	9	11.11
20-29	18	0	18	60	30.00
30-39	23	4	27	60	45.00
40-49	36	6	42	61	68.85
50-59	40	8	48	61	78.69
60-69	35	14	49	60	81.67
70-79	28	25	53	60	88.33
80-89	25	35	60	60	100.00
90-99	2	18	20	20	100.00
>99	1	0	1	2	50.00
TOTAL	**209**	**110**	**319**	**453**	**70.42**

Table D2. Frequencies for each trait score in each age cohort pooled population.

Cohort	Trait	Trait Score Frequency				
		1	**2**	**3**	**4**	**5**
<20	Glabella	1	6	2	0	0
	Left Mastoid	0	3	3	3	0
	Left Orbit	0	2	4	3	0
	Nuchal Crest	2	5	2	0	0
	Mental Eminence	0	1	3	3	0
20-29	Glabella	6	26	18	10	0
	Left Mastoid	0	26	21	10	3
	Left Orbit	6	17	21	13	2
	Nuchal Crest	6	30	18	6	0
	Mental Eminence	1	19	23	7	1
30-39	Glabella	5	26	21	8	0
	Left Mastoid	1	30	17	11	1
	Left Orbit	5	19	21	15	0
	Nuchal Crest	7	26	20	7	0
	Mental Eminence	1	19	18	6	1
40-49	Glabella	8	24	14	12	3
	Left Mastoid	2	17	28	12	2
	Left Orbit	4	11	30	13	2
	Nuchal Crest	1	28	20	10	2
	Mental Eminence	1	10	15	15	1
50-59	Glabella	3	28	20	10	0
	Left Mastoid	0	26	21	8	6
	Left Orbit	3	16	27	14	1
	Nuchal Crest	1	20	20	17	3
	Mental Eminence	1	8	16	15	1
60-69	Glabella	0	27	20	10	1
	Left Mastoid	0	20	19	17	4
	Left Orbit	4	17	25	10	3
	Nuchal Crest	0	17	25	16	1
	Mental Eminence	1	6	9	17	1

Table D2. (continued)

Cohort	Trait	Trait Score Frequency				
		1	**2**	**3**	**4**	**5**
70-79	Glabella	2	28	24	6	0
	Left Mastoid	0	22	26	9	3
	Left Orbit	3	14	26	14	1
	Nuchal Crest	1	17	29	10	3
	Mental Eminence	0	7	6	11	0
80-89	Glabella	0	24	19	14	3
	Left Mastoid	0	13	31	13	2
	Left Orbit	0	12	29	18	0
	Nuchal Crest	2	17	18	15	8
	Mental Eminence	2	4	6	8	1
90-99	Glabella	1	7	3	9	0
	Left Mastoid	0	6	10	4	0
	Left Orbit	0	3	13	4	0
	Nuchal Crest	0	7	6	6	1
	Mental Eminence	0	1	1	0	0
99<	Glabella	0	0	1	1	0
	Left Mastoid	0	0	1	1	0
	Left Orbit	0	0	1	1	0
	Nuchal Crest	0	0	1	1	0
	Mental Eminence	0	1	1	0	0

Table D3. Frequencies for each trait score in each age cohort for females.

Cohort	Trait	Trait Score Frequency				
		1	**2**	**3**	**4**	**5**
<20	Glabella	1	3	1	0	0
	Left Mastoid	0	2	1	2	0
	Left Orbit	0	1	3	1	0
	Nuchal Crest	2	2	1	0	0
	Mental Eminence	0	1	1	2	0
20-29	Glabella	6	18	5	1	0
	Left Mastoid	0	19	8	2	1
	Left Orbit	6	10	11	2	0
	Nuchal Crest	5	18	7	0	0
	Mental Eminence	1	16	10	1	0
30-39	Glabella	4	16	8	2	0
	Left Mastoid	0	21	4	5	0
	Left Orbit	5	11	12	2	0
	Nuchal Crest	5	16	9	0	0
	Mental Eminence	1	12	10	1	0
40-49	Glabella	7	16	6	1	0
	Left Mastoid	2	15	10	3	0
	Left Orbit	3	8	18	1	0
	Nuchal Crest	1	16	10	3	0
	Mental Eminence	0	5	10	3	0
50-59	Glabella	3	20	6	1	0
	Left Mastoid	0	20	10	0	0
	Left Orbit	2	8	19	1	0
	Nuchal Crest	1	13	12	2	2
	Mental Eminence	1	6	7	3	0
60-69	Glabella	0	19	8	3	0
	Left Mastoid	0	12	11	5	2
	Left Orbit	2	10	15	2	1
	Nuchal Crest	0	7	14	7	11
	Mental Eminence	1	5	4	7	1

Table D3. (continued)

Cohort	Trait	Trait Score Frequency				
		1	**2**	**3**	**4**	**5**
70-79	Glabella	2	13	15	0	0
	Left Mastoid	0	13	15	2	0
	Left Orbit	2	7	15	5	0
	Nuchal Crest	1	8	15	3	3
	Mental Eminence	0	3	3	2	0
80-89	Glabella	0	15	11	4	0
	Left Mastoid	0	8	15	6	0
	Left Orbit	0	7	15	8	0
	Nuchal Crest	2	11	10	6	1
	Mental Eminence	1	2	5	0	0
90-99	Glabella	1	5	2	1	0
	Left Mastoid	0	4	5	0	0
	Left Orbit	0	2	5	2	0
	Nuchal Crest	0	5	3	1	0
	Mental Eminence	0	1	0	0	0
99<	Glabella	0	0	0	0	0
	Left Mastoid	0	0	0	0	0
	Left Orbit	0	0	0	0	0
	Nuchal Crest	0	0	0	0	0
	Mental Eminence	0	0	0	0	0

Table D4. Frequencies for each trait score in each age cohort for males.

Cohort	Trait	Trait Score Frequency				
		1	**2**	**3**	**4**	**5**
<20	Glabella	0	3	1	0	0
	Left Mastoid	0	1	2	1	0
	Left Orbit	0	1	1	2	0
	Nuchal Crest	0	3	1	0	0
	Mental Eminence	0	0	2	1	0
20-29	Glabella	0	8	13	9	0
	Left Mastoid	0	7	13	8	2
	Left Orbit	0	7	10	11	2
	Nuchal Crest	1	12	11	6	0
	Mental Eminence	0	3	13	6	1
30-39	Glabella	1	10	13	6	0
	Left Mastoid	1	9	13	6	1
	Left Orbit	0	8	9	13	0
	Nuchal Crest	2	10	11	7	0
	Mental Eminence	0	7	8	5	1
40-49	Glabella	1	7	8	11	3
	Left Mastoid	0	2	17	9	2
	Left Orbit	1	3	12	12	2
	Nuchal Crest	0	11	10	7	2
	Mental Eminence	1	4	5	12	1
50-59	Glabella	0	7	14	9	0
	Left Mastoid	0	5	11	8	6
	Left Orbit	1	7	8	13	1
	Nuchal Crest	0	7	7	15	1
	Mental Eminence	0	2	9	12	1
60-69	Glabella	0	8	12	7	1
	Left Mastoid	0	8	8	12	2
	Left Orbit	2	7	10	8	2
	Nuchal Crest	0	10	11	9	0
	Mental Eminence	0	1	5	10	0

Table D4. (continued)

Cohort	Trait	Trait Score Frequency				
		1	**2**	**3**	**4**	**5**
70-79	Glabella	0	15	9	6	0
	Left Mastoid	0	9	11	7	3
	Left Orbit	1	7	11	9	1
	Nuchal Crest	0	9	14	7	0
	Mental Eminence	0	4	3	9	0
80-89	Glabella	0	9	8	10	3
	Left Mastoid	0	5	16	7	2
	Left Orbit	0	5	14	10	0
	Nuchal Crest	0	6	8	9	7
	Mental Eminence	1	2	1	8	1
90-99	Glabella	0	2	1	8	0
	Left Mastoid	0	2	5	4	0
	Left Orbit	0	1	8	2	0
	Nuchal Crest	0	2	3	5	1
	Mental Eminence	0	0	1	0	0
99<	Glabella	0	0	1	1	0
	Left Mastoid	0	0	1	1	0
	Left Orbit	0	0	1	1	0
	Nuchal Crest	0	0	1	1	0
	Mental Eminence	0	1	1	0	0

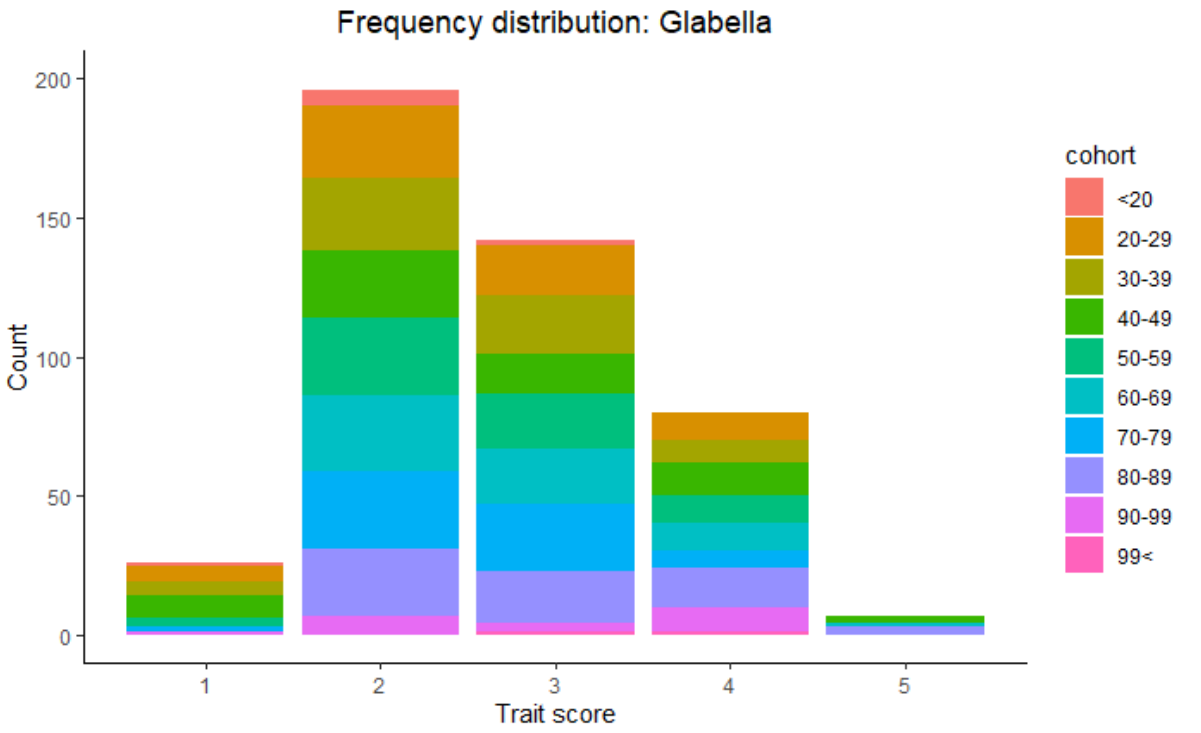


Figure D1. Illustration of the frequency distribution of the scores of the glabella for the age cohorts.

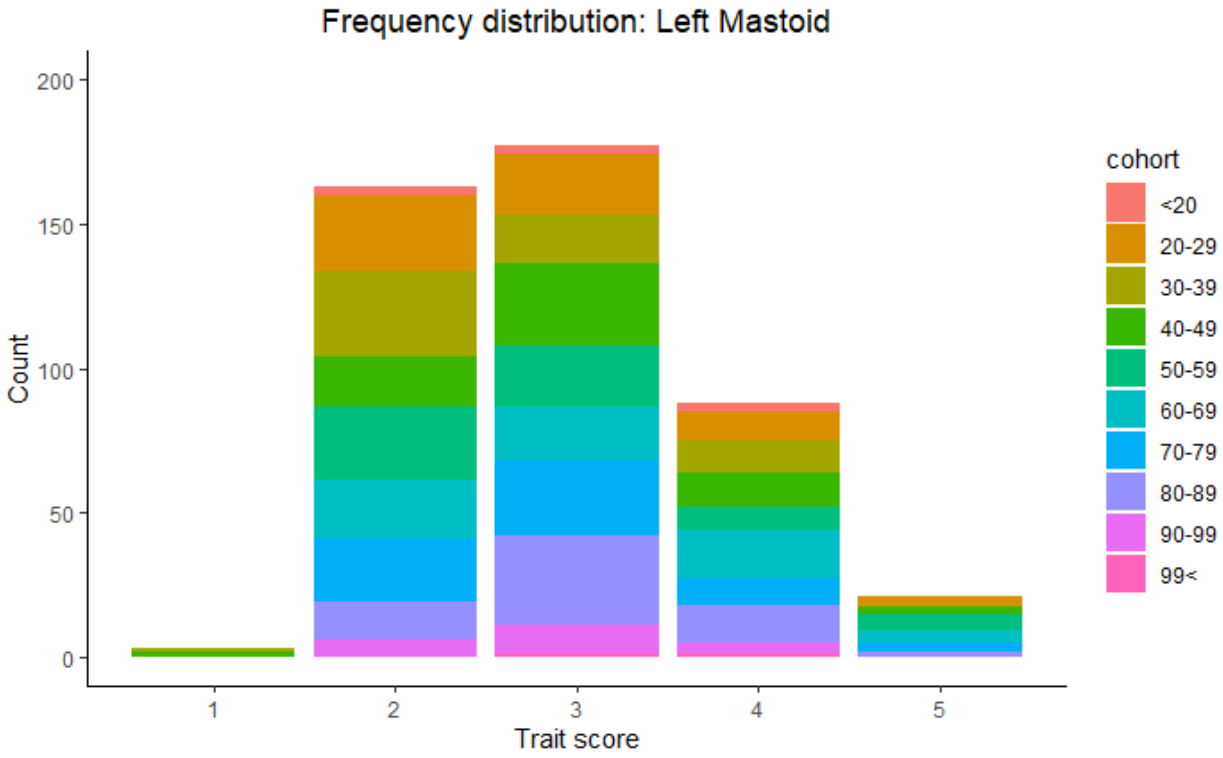


Figure D2. Illustration of the frequency distribution of the scores of the left mastoid for the age cohorts.

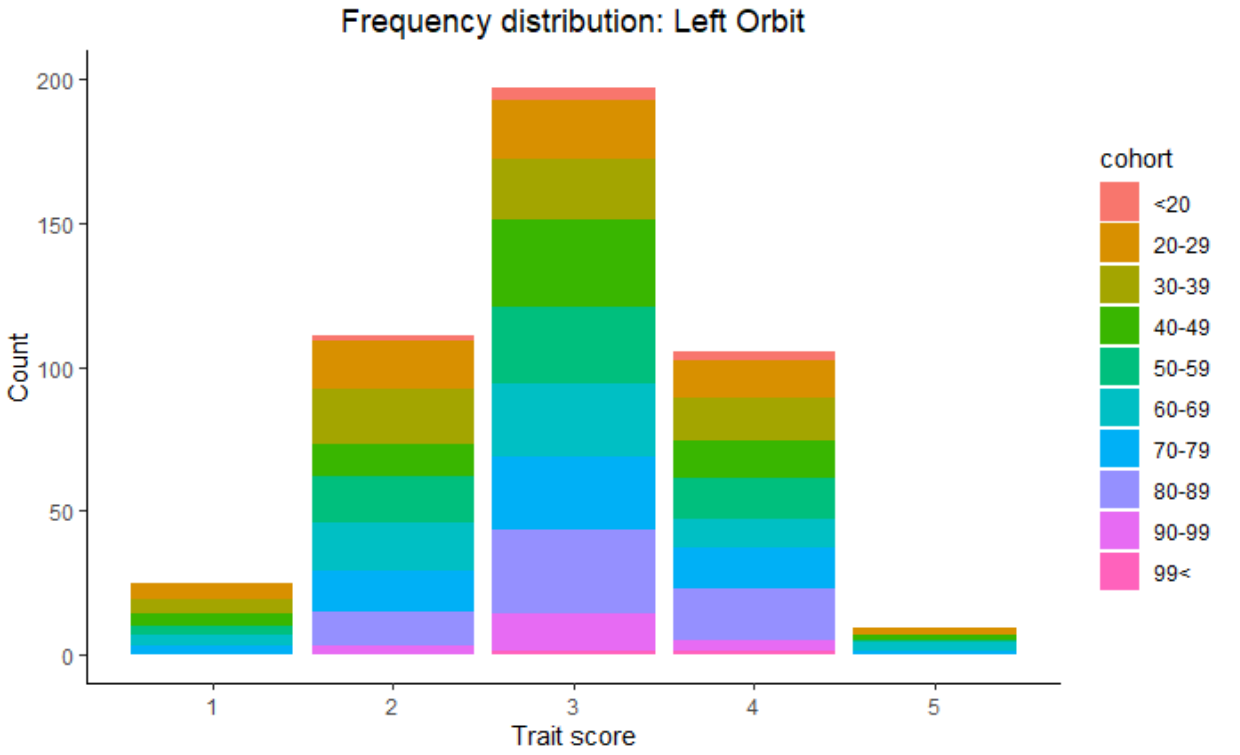


Figure D3. Illustration of the frequency distribution of the scores of the left orbit for the age cohorts.

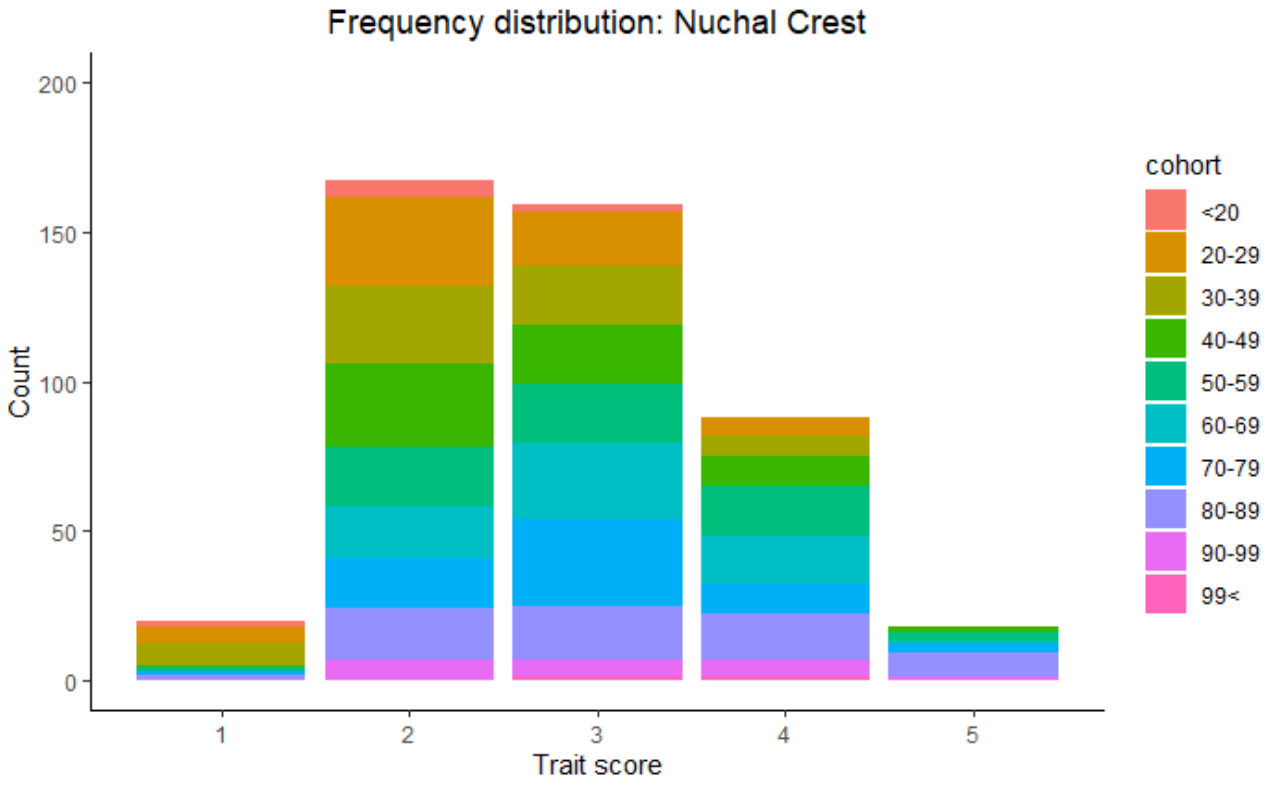


Figure D4. Illustration of the frequency distribution of the scores of the nuchal crest for the age cohorts.

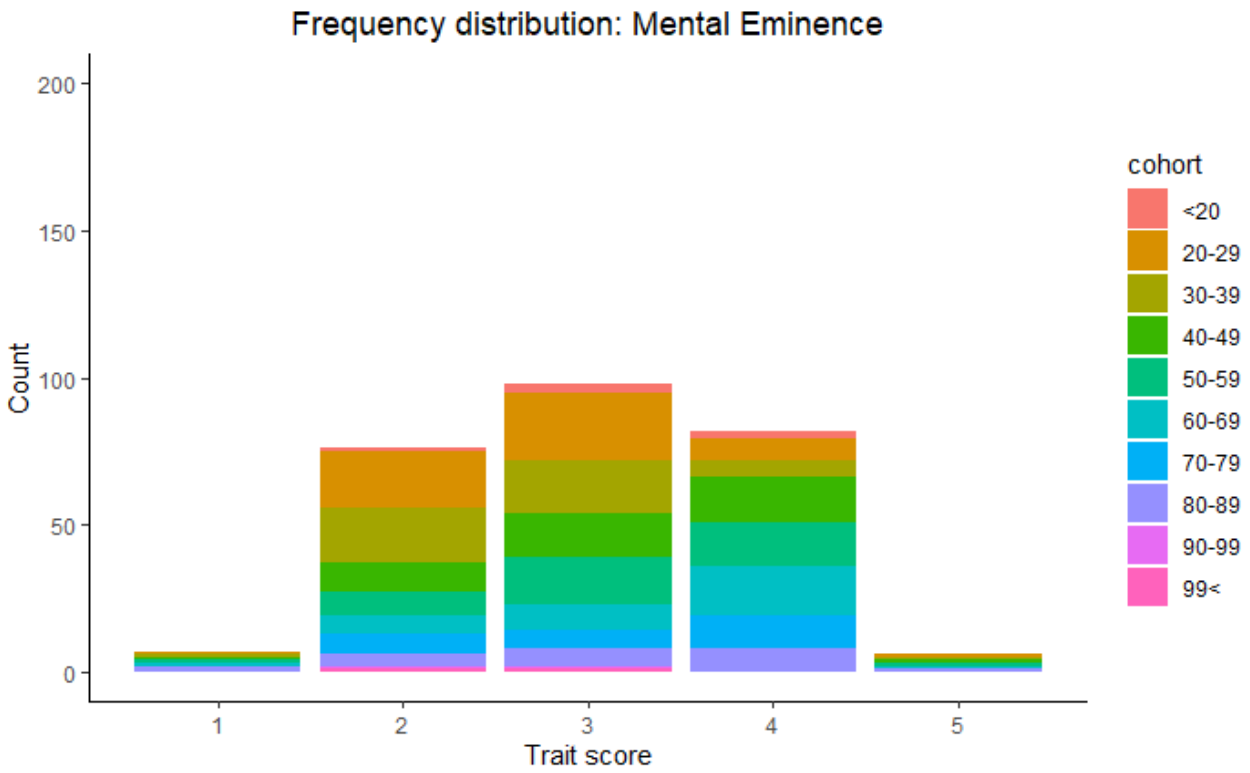


Figure D5. Illustration of the frequency distribution of the scores of the mental eminence for the age cohorts.

Table D5. Frequencies for each trait score in each age cohort.

Cohort	Trait	Trait Score Frequency				
		1	**2**	**3**	**4**	**5**
<20	Left VA	0	1	2	0	0
	Left MIR	0	0	2	1	0
	Left SPC	0	2	1	0	0
20-29	Left VA	2	20	18	19	1
	Left MIR	2	11	19	26	2
	Left SPC	3	16	20	15	5
30-39	Left VA	3	12	19	16	7
	Left MIR	3	13	20	20	3
	Left SPC	3	15	15	22	3
40-49	Left VA	8	12	15	21	2
	Left MIR	1	14	25	14	4
	Left SPC	4	17	16	20	1
50-59	Left VA	6	20	10	20	3
	Left MIR	4	17	21	15	2
	Left SPC	8	18	11	19	3
60-69	Left VA	6	12	15	21	4
	Left MIR	6	10	20	18	5
	Left SPC	7	14	14	18	6
70-79	Left VA	7	12	9	25	6
	Left MIR	3	11	19	21	5
	Left SPC	9	13	10	23	4
80-89	Left VA	7	18	7	23	3
	Left MIR	4	13	23	17	1
	Left SPC	9	19	6	20	4
90-99	Left VA	1	3	3	6	0
	Left MIR	0	3	3	7	0
	Left SPC	2	2	2	6	1
99<	Left VA	0	0	0	1	0
	Left MIR	0	0	0	1	0
	Left SPC	0	0	0	1	0

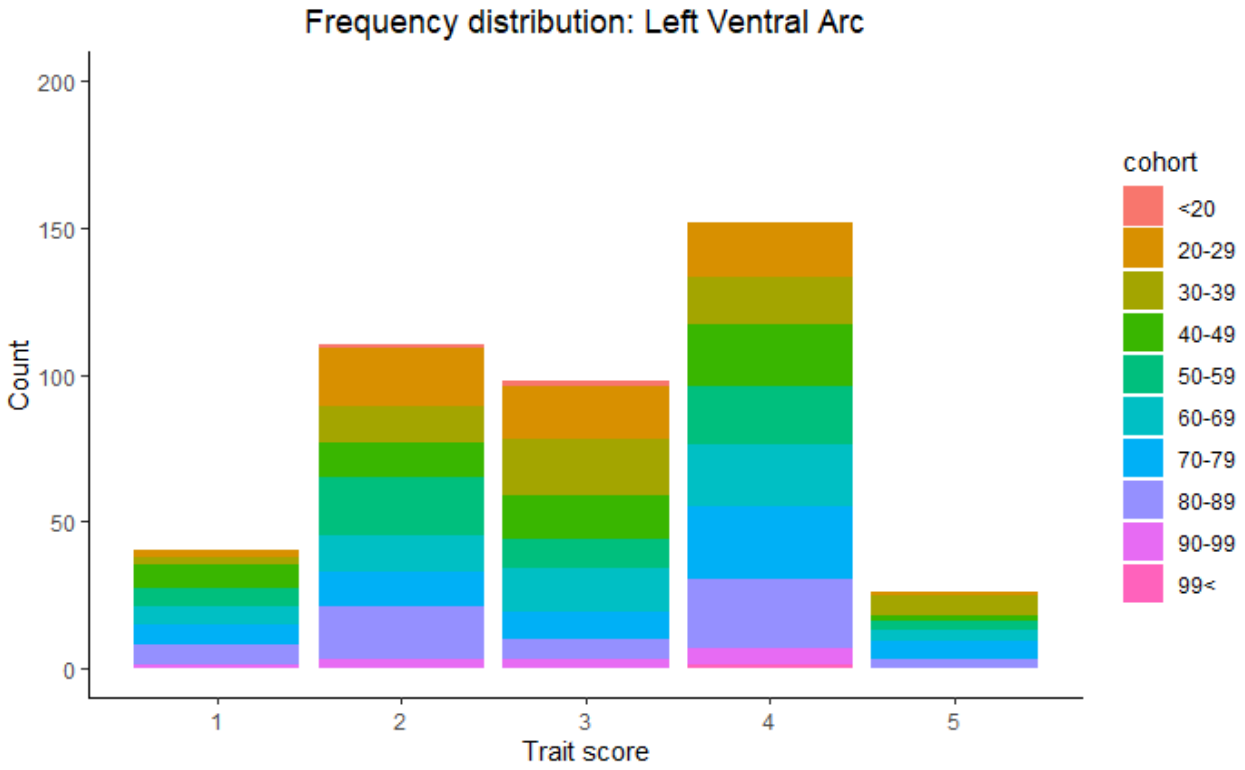


Figure D6. Illustration of the frequency distribution of the scores of the left ventral arc for the age cohorts.

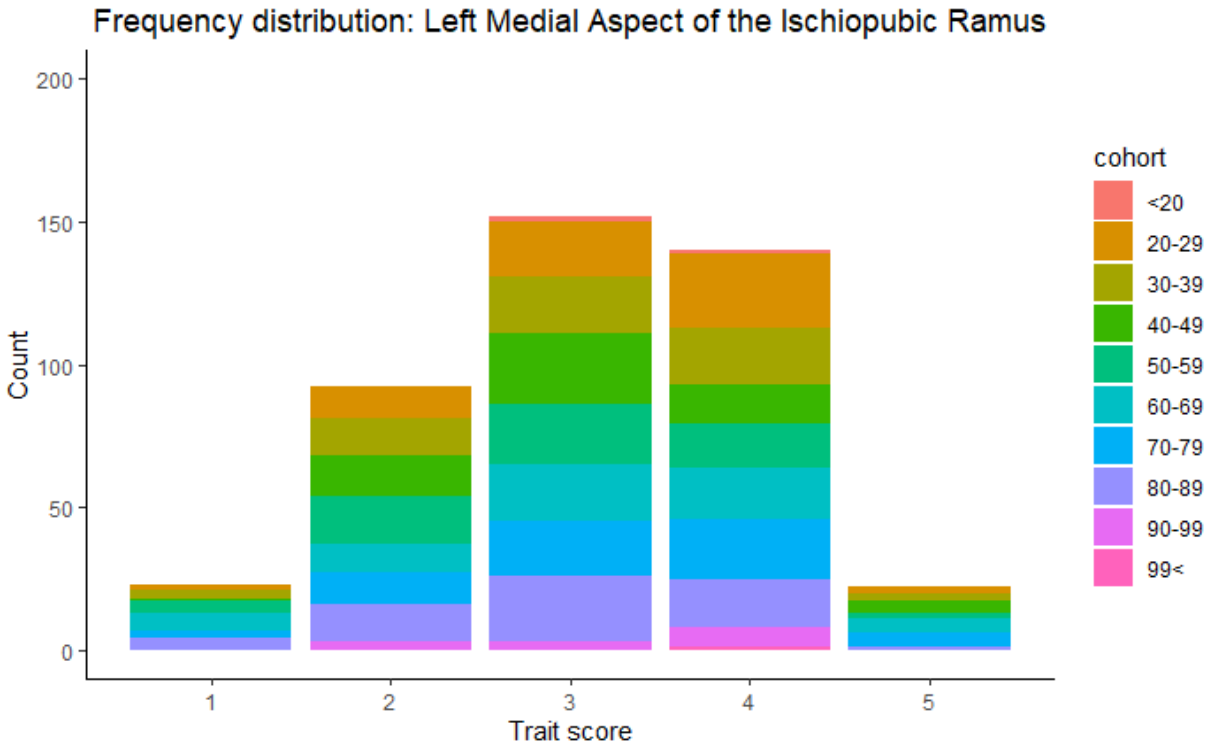


Figure D7. Illustration of the frequency distribution of the scores of the left medial aspect of the ischiopubic ramus for the age cohorts.

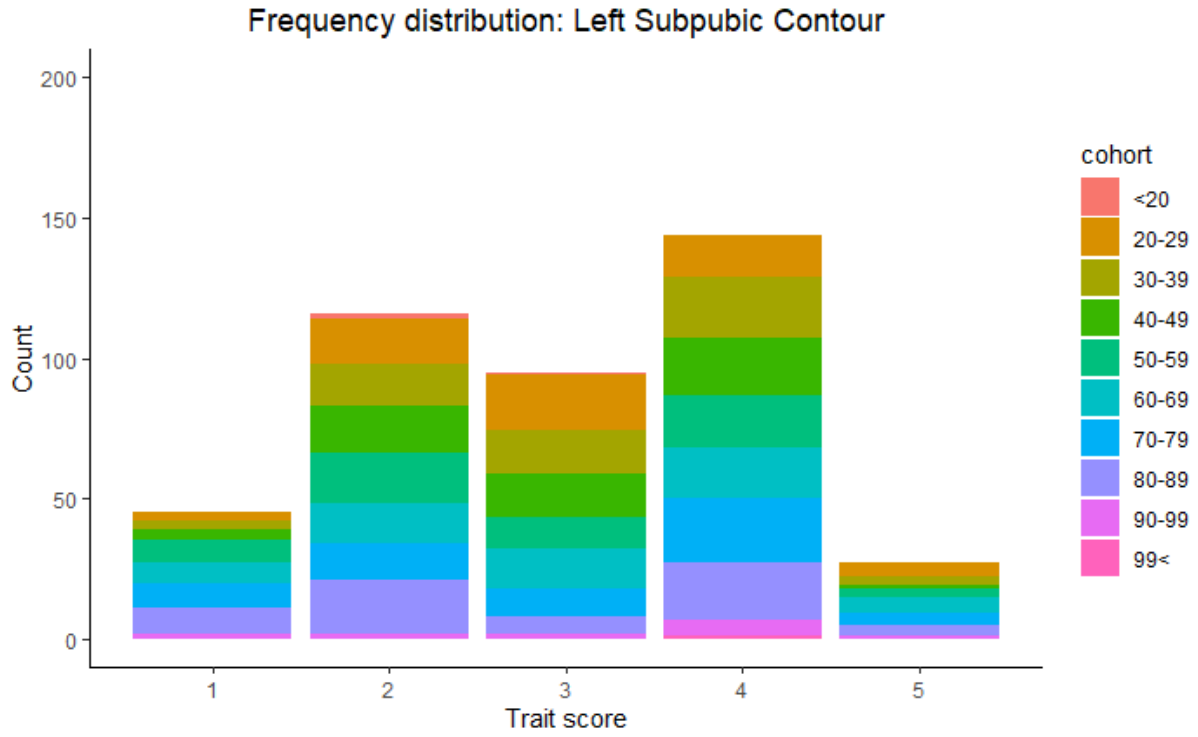


Figure D8. Illustration of the frequency distribution of the scores of the left subpubic contour for the age cohorts.

APPENDIX E: EXTENDED TABLES

Table E1. Significant differences present between the Nuchal crest of different age cohorts using the Kruskal Wallis test with a post hoc Dunn's test and a Bonferroni correction. Bold indicates significant differences.

Cohort comparison	p-value
<20 - 20-29	1.00
<20 - 30-39	1.00
20-29 - 30-39	1.00
<20 - 40-49	1.00
20-29 - 40-49	1.00
30-39 - 40-49	1.00
<20 - 50-59	0.12
20-29 - 50-59	**<0.05**
30-39 - 50-59	0.07
40-49 - 50-59	1.00
<20 - 60-69	0.09
20-29 - 60-69	**<0.01**
30-39 - 60-69	**<0.05**
40-49 - 60-69	1.00
50-59 - 60-69	1.00
20-29 - 70-79	**<0.05**
30-39 - 70-79	0.18
40-49 - 70-79	1.00
50-59 - 70-79	1.00
60-69 - 70-79	1.00
<20 - 80-89	**<0.05**
20-29 - 80-89	**<0.01**
30-39 - 80-89	**<0.01**
40-49 - 80-89	0.87
50-59 - 80-89	1.00
60-69 - 80-89	1.00
70-79 - 80-89	1.00

Table E1. (continued)

Cohort comparison	p-value
<20 - 90-99	0.27
20-29 - 90-99	0.40
30-39 - 90-99	0.86
40-49 - 90-99	1.00
50-59 - 90-99	1.00
60-69 - 90-99	1.00
70-79 - 90-99	1.00
80-89 - 90-99	1.00
<20 - 99<	1.00
20-29 - 99<	1.00
30-39 - 99<	1.00
40-49 - 99<	1.00
50-59 - 99<	1.00
60-69 - 99<	1.00
70-79 - 99<	1.00
80-89 - 99<	1.00
90-99 - 99<	1.00

Table E2. Significant differences present in males for the left VA of different age cohorts using the Kruskal Wallis test with a post hoc Dunn's test and a Bonferroni correction. Bold indicates significant differences.

Cohort comparison	p-value
20-29 - 30-39	0.75
20-29 - 40-49	1.00
30-39 - 40-49	1.00
20-29 - 50-59	1.00
30-39 - 50-59	1.00
40-49 - 50-59	1.00
20-29 - 60-69	1.00
30-39 - 60-69	1.00

Table E2. (continued)

Cohort comparison	p-value
40-49 - 60-69	1.00
50-59 - 60-69	1.00
20-29 - 70-79	**<0.05**
30-39 - 70-79	1.00
40-49 - 70-79	0.52
50-59 - 70-79	1.00
60-69 - 70-79	1.00
20-29 - 80-89	1.00
30-39 - 80-89	1.00
40-49 - 80-89	1.00
50-59 - 80-89	1.00
60-69 - 80-89	1.00
70-79 - 80-89	1.00
20-29 - 90-99	1.00
30-39 - 90-99	1.00
40-49 - 90-99	1.00
50-59 - 90-99	1.00
60-69 - 90-99	1.00
70-79 - 90-99	1.00
80-89 - 90-99	1.00
20-29 - 99<	1.00
30-39 - 99<	1.00
40-49 - 99<	1.00
50-59 - 99<	1.00
60-69 - 99<	1.00
70-79 - 99<	1.00
80-89 - 99<	1.00
90-99 - 99<	1.00

Table E3. Significant differences present between the Nuchal crest and Mental Eminence of different populations using the Kruskal Wallis test with a post hoc Dunn's test and a Bonferroni correction. Bold indicates significant differences.

	Population comparison	**P-value**
Nuchal Crest	black | white	**<0.01**
Mental Eminence	black | white	**<0.01**

APPENDIX F: ADDITIONAL TESTS

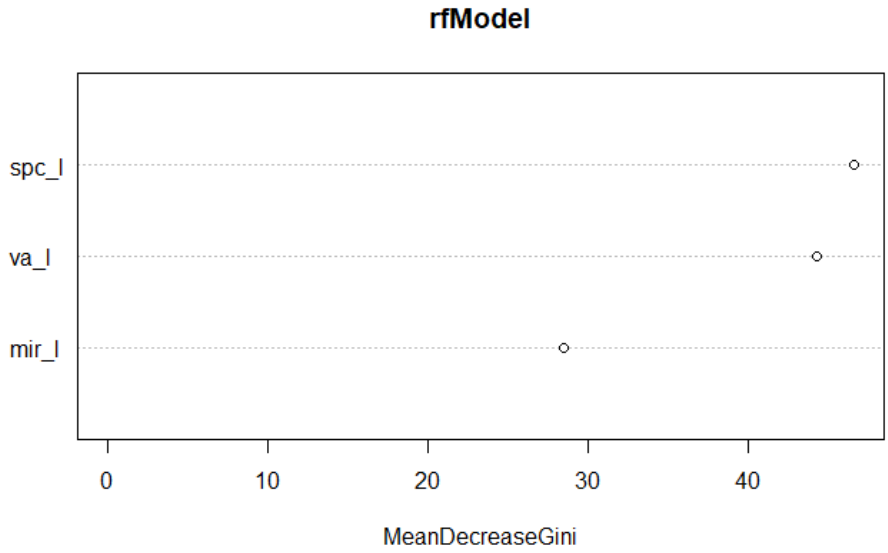


Figure F1. Visualisation of the variable importance in the pooled sample RFM

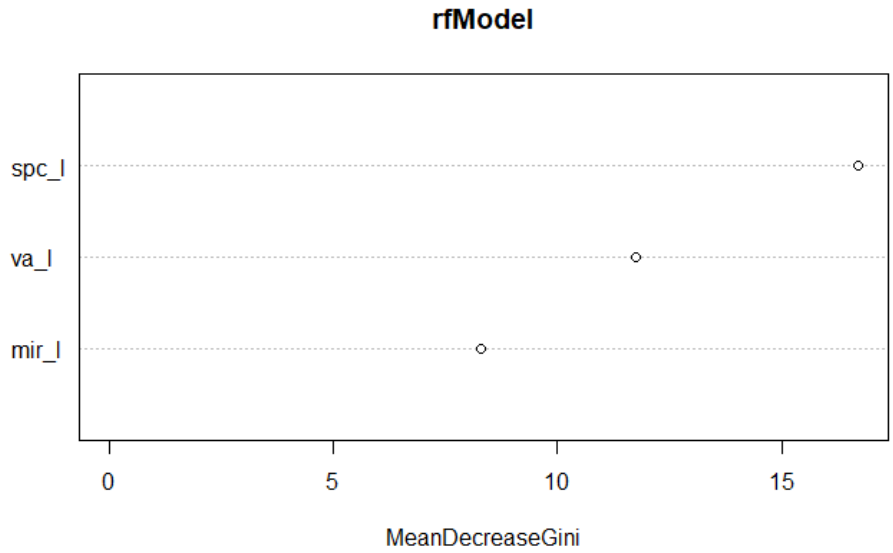


Figure F2. Visualisation of the variable importance in middle-aged sample RFM

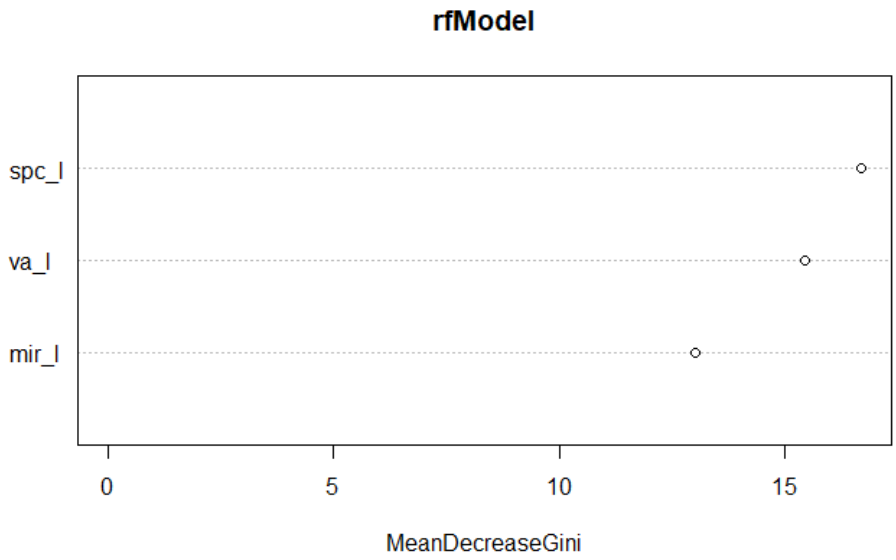


Figure F3. Visualisation of the variable importance in the younger sample RFM

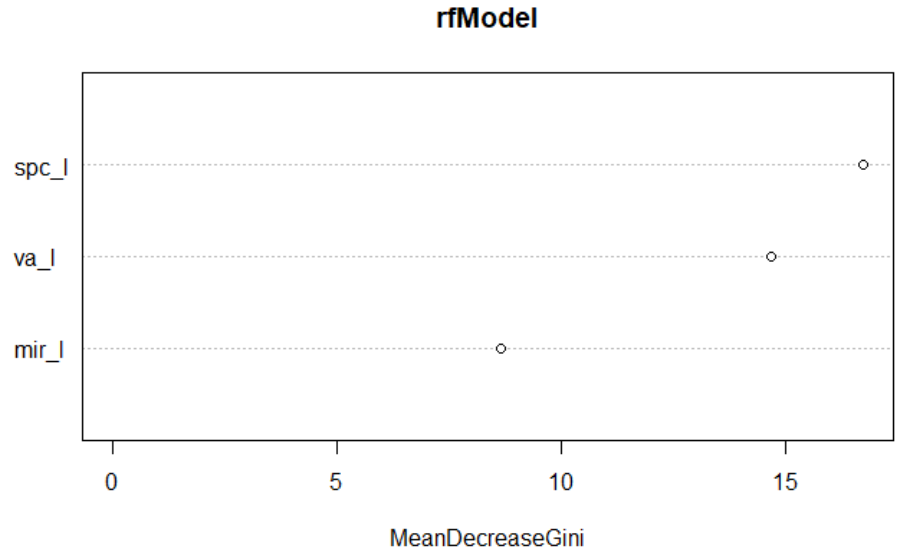


Figure F4. Visualisation of the variable importance in the older sample RFM

Walker (2008) Traits Population Estimation RFM

Due to the results of the Kruskal Wallis test showing significant differences between WSAs and BSAs a Population Estimation RFM was run to analyse the possible outcomes. The accuracy of the Population Prediction RFM was 84.31% is beyond what would be expected by random chance The variable importance scale of the model illustrated that the nuchal crest showed the greatest level of importance, followed by the mental eminence, the mastoid process, the supra-orbital margin and then lastly the glabella (Figure F5). The testing model, which was the 25% of the data that was not utilised to build the model, comprised of 73 black South Africans, 29 white South Africans (Table F1**Table F3.** Percent correct for the function when

divided into population groups.). Of the 73 BSAs 70 correctly classified as BSA and 3 misclassified as WSA, leaving a correct classification rate of 95.89% for the BSAs. Furthermore, of the 29 WSAs, 16 correctly classified as WSA and 13 misclassified as BSA, leaving a correct classification rate of 60.69% for the WSAs.

Table F1. Percent correct for the function when divided into population groups.

RFM Function	Percent Correct (%)		
	BSA	WSA	Total
Population Estimation	95.89	60.69	84.31

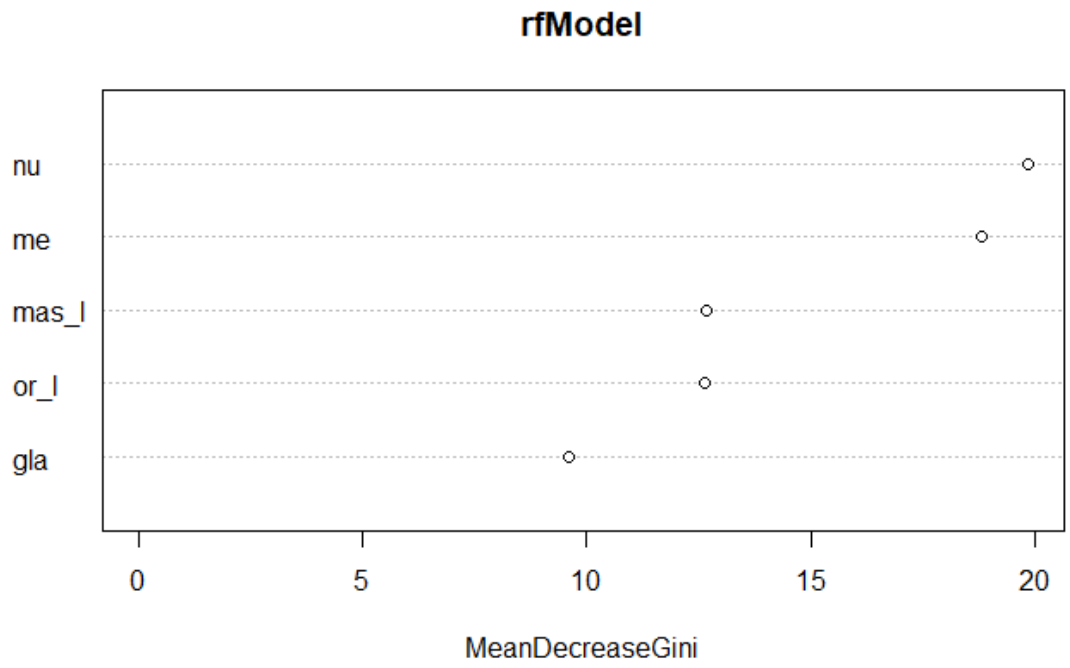


Figure F5. Visualisation of the variable importance in the Population RFM for the Walker Traits

Table F2. Variable importance for the population estimation RFM.

Walker (2008) Trait	Variable importance
Gla	9.616963
Mas	12.666302
Or	12.638978
Nu	19.864699

Me	18.803610

Phenice (1969) Traits Population Prediction RFM

The accuracy of the Population Prediction RFM was 67.82% which is beyond what would be expected by random chance. The variable importance scale of the model illustrated that the subpubic contour showed the greatest level of importance, followed by the medial aspect of the ischiopubic ramus, and then lastly the ventral arc (Figure F6; Table F4). The testing model, which was the 25% of the data that was not utilised to build the model, comprised of 70 black South Africans, 37 white South Africans (Table F3**Table F3.** Percent correct for the function when divided into population groups.). Of the 70 BSAs 68 correctly classified as BSA and 2 misclassified as WSA, leaving a correct classification rate of 97.14% for the BSAs. Furthermore, of the 37 WSAs, 4 correctly classified as WSA and 33 misclassified as BSA, leaving a correct classification rate of 10.81% for the WSAs.

Table F3. Percent correct for the function when divided into population groups.

RFM Function	Percent Correct (%)		
	BSA	WSA	Total
Population Estimation	97.14	10.81	67.29

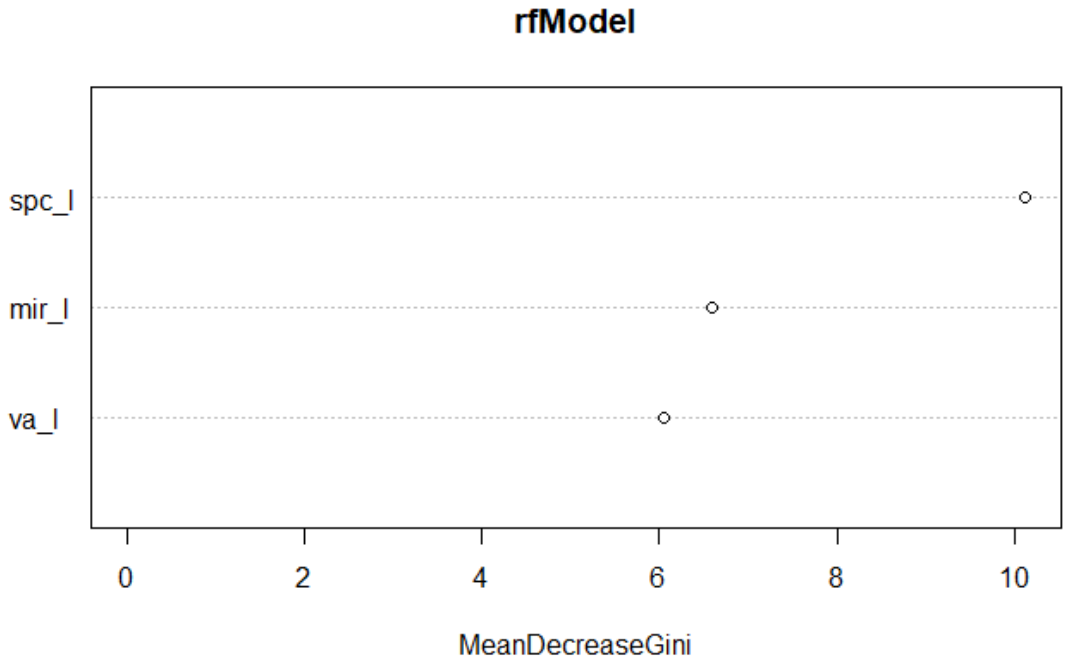


Figure F6. Visualisation of the variable importance for the population estimation RFM

Table F4. Variable importance for the population estimation RFM.

Phenice (1969) Trait	Variable importance
VA	6.041616
MIR	6.594131
SPC	10.124451