Estimating Sex from the Human Scapula: A Validation Study of the Five- and Two-variable models and FORDISC 3.0 in two White European populations

by Ian Carter Bell

A Thesis Submitted to Saint Mary's University, Halifax, Nova Scotia,
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Applied Science
July 26, 2013, Halifax, Nova Scotia

Approved: Dr. Tanya Peckmann
Supervisor
Department of Anthropology &
Forensic Sciences Program

Approved: Dr. Joseph Parish
External Examiner
Department of Anthropology and
Sociology
Cape Breton Univeristy

Approved: Michelle MacKay
Supervisory Committee Member
Department of Biology

Approved: Dr. Matthew Bowes
Supervisory Committee Member
Nova Scotia Medical Examiner
Service

Approved: Dr. Jeremy Lundholm
Graduate Studies Representative

Date: July 26, 2013

Dedicated to my two brothers, Nathan and Brian, who taught me the value of hard work,
the dedication to one's career, and the ties that bind a family

ABSTRACT

Estimating Sex from the Human Scapula: A Validation Study of the Five- and Two-variable models and FORDISC 3.0 in two White European populations

by Ian Carter Bell

The objectives were to understand the relationship between biological sex and estimated sex from the scapula of two White European populations based on metric analyses of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005). This research provided alternative methods for estimating sex from the scapula based on metric analyses of the height, breadth, and calculated area of the glenoid cavity. Three hundred and thirty-five left scapulae from the Athens Collection and the William Bass Collection were measured. The results of the study produced three results: the five- and two-variable models are accurate methods for estimating sex over White European populations, FORDISC 3.0 is an accurate methodology for estimating sex and, the glenoid cavity can be used as an accurate osteometric characteristic for estimating sex from the scapula and the difference in accuracy rates between similar populations groups are not significant.

July 26, 2013

ACKNOWLEDGMENTS

There were a number of individuals that could make this research based Master's thesis possible and if my feeble, fried, and, otherwise, forgetful mind should miss someone, I do apologize in advance. However, I am grateful for your help.

I would like to thank my thesis committee supervisor, Dr. Tanya Peckmann, for fostering my love for forensic anthropology and teaching me how to become a better person and a better scientist. Dr. Peckmann has been teaching me about forensic anthropology and its importance within our community and I have learned so much from her that I will forever be in her debt.

Next, I would like to thank my two supporting thesis committee advisors, Prof. Michelle MacKay and Dr. Matthew Bowes, who worked diligently to prepare my thesis to a style that was professional, clear, and comprehensible to an academic standard and, for that, I am very thankful.

I would like to thank Drs. Lee Meadow Jantz and Soritis Manolis, curators of the William H. Bass and Athens Collection, respectively, for allowing me to use their skeletal reference collections to accomplish the goals of this current research. Without their approval this thesis would not have been possible.

I would like to thank Kathleen Hauther, of the University of Tennessee, and Maria-Eleni Chovalopoulou, of the University of Athens, who so graciously volunteered their time and energy to be my inter-observer measurers. They were not only a huge asset to this investigation but they also helped me maneuverer through the strange streets of their respective cities. I would also like to thank my photography/artistic colleague, James Neish, for his assistance with finalizing the scapular photographs in this study.

Lastly, I would like to thank the ever patient and always helpful Dr. Susan Meek, who was so helpful in almost every aspect of this thesis, especially when dealing with data collection, data analyses, and data interpretation. Her help on this current research was immeasurable and I will always be grateful for that.

TABLE OF CONTENTS

ABSTRACT......i

ACKNOWLEDGMENTS...... ii

LIST OF TABLES...... viii

LIST OF FIGURES......xi

CHAPTER 1: INTRODUCTION......1

1.1 Objectives......1

1.2 Concepts of identity in physical anthropology 2

1.3 Historical background for estimation of sex from human skeletal remains..4

1.3.1 Estimating sex from the pelvis 5

1.3.2 Estimating sex from the skull......9

1.3.3 Estimating sex from the long bones......11

1.3.4 Estimating sex from the metacarpals, carpals, metatarsals and tarsals.......13

1.3.5 Estimating sex from the scapula......17

1.3.6 Estimating sex from other postcranial skeletal elements 23

1.4 Osteological collections utilized for this research......24

1.4.1 University of Athens Human Skeletal Reference Collection......24

1.4.2 William Bass Donated Skeletal Collection......25

CHAPTER 2: MATERIALS AND METHODS 26

2.1 Skeletal materials utilized for this research......26

2.2 Methods......27

2.2.1 Dabbs and Moore-Jansen (2010) five-variable and two-variable models......27

2.2.2 FORDISC 3.0 by Jantz and Ousley (2005)......33

2.2.3 The creation of the models for "Macaluso's Hypothesis" for estimating sex from glenoid cavity measurements 36

2.3 Statistical analyses......39

CHAPTER 3: RESULTS **44**

3.1 The five- and two-variable models and FORDISC 3.0 validation study......**44**

3.2 Descriptive statistics of the measurements for the five- and two-variable models and FORDISC 3.0 validation study **45**

3.2 Results of Dabbs and Moore-Jansen (2010) validation study **47**

3.2.1 Two-sample t-test vs. Mann-Whitney U test 47

3.2.2 Sexual dimorphic variation between males and females of the Athens and Tennessee samples 48

3.2.3 Accuracy of the two-variable and five-variable models 49

3.2.4 Chi-squared proportion tests of the five-variable and two-variable models...51

3.2.5 Statistical variation between same sex measurements: comparing the Athens and Tennessee populations 53

3.2.6 Statistical variation between same sex measurements: comparing the Athens and "Cleveland" populations 55

3.2.7 Statistical variation between same sex measurements: comparing the Tennessee and "Cleveland" populations 57

3.2.8 Variation of standard deviation using f-tests 59

3.3 Results of the FORDISC 3.0 validation study **62**

3.3.1 FORDISC 3.0 analyses...62

3.3.2 Accuracy of FORDISC 3.0 when applied to the Athens and Tennessee population groups...63

3.3.3 Chi-squared test results for comparing the accuracy of the two-variable model with FORDISC 3.0....64

3.3.4 Descriptive statistics of the discriminate function analyses of the Athens population group 65

3.3.5 Descriptive statistics of the discriminate function analyses of the Tennessee population group 67

3.3.6 Low typical probability..69

3.4 Intra- and inter- observer error for the five- and two-variable models and FORDISC 3.0 validation study 70

3.5 Pearson's correlation test for age at death of the Athens and Tennessee population groups for the five- and two-variable models and FORDISC 3.0. 71

3.6 Individuals used for the creation of the models for "Macaluso's Hypothesis" for estimating sex from the glenoid cavity. 73

3.7 Descriptive statistics for the creation of the models for "Macaluso's Hypothesis" 73

3.8 Results of the logistic regression analyses for the creation of the models for "Macaluso's Hypothesis" 75

3.8.1 Sexual dimorphic variation between males and females of the Athens and Tennessee population groups 75

3.8.2 Statistical variation between same sex measurements: Athens and Tennessee 76

3.8.3 Binary logistic regression analysis on the Athens population. 78

3.8.4 Hosmer and Lemeshow test for goodness-of-fit on Athens predictive models 80

3.8.5 Classification plots on Athens predictive models 81

3.8.6 Direct vs. Stepwise equations of the Athens population group 81

3.8.7 Binary logistic regression analysis on the Tennessee population. 82

3.8.8 Hosmer and Lemeshow test for goodness-of-fit on Tennessee predictive models 83

3.8.9 Classification plots on the Tennessee predictive models 84

3.8.10 Direct vs. Stepwise equations of the Tennessee population group 84

3.8.11 Binary logistic regression analysis on the Combined Population groups (Athens and Tennessee). 85

3.8.12 Hosmer and Lemeshow test for goodness-of-fit on Combined Population groups (Athens and Tennessee). 86

3.8.13 Classification plots on the Combined Population groups predictive models 87

3.8.14 Direct vs. Stepwise equations of the Combined Population groups............88

3.8.15 Chi-squared proportion tests of the classification accuracies of the models for "Macaluso's Hypothesis" (Athens, Tennessee and Combined Population).........89

3.9 Pearson's correlation test for age at death in the creation of the models for "Macaluso's Hypothesis"..90

3.10 Intra- and inter-observer error in the creation of the models for "Macaluso's Hypothesis"..91

CHAPTER 4: DISCUSSION 93

4.1 Context of the current project...93

4.2 Dabbs and Moore-Jansen (2010) validation study on the five- and two-variable models and population diversity 94

4.3 FORDISC 3.0 validation study and population diversity...........................104

4.4 The models for "Macaluso's Hypothesis" for estimating sex from the glenoid cavity. 110

CHAPTER 5: CONCLUSION...117

REFERENCES: 124

Appendix A: Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) 134

Appendix B: Measurements used for the creation of the models for "Macaluso's Hypothesis"...154

LIST OF TABLES

Table 2.1: Equations for the Dabbs and Moore-Jansen (2010) models for estimating sex from the scapula. 28

Table 2.2: Description of measurements for the five-variable model (Adapted from Dabbs and Moore-Jansen, 2010). 28

Table 2.3: Description of measurements for the two-variable model (Adapted from Dabbs and Moore-Jansen, 2010). 28

Table 2.4: Description of measurements for the FORDISC 3.0 program (Adapted from Buikstra and Ubleaker, 1994). 35

Table 2.5: Description of measurements used for the models for “Macaluso’s Hypothesis” for estimating sex of the glenoid cavity (Adapted from Macaluso, 2010)......37

Table 3.1: Sex and age of individuals used for the five- and two-variable models and FORDISC 3.0 validation study 45

Table 3. 2: Athens descriptive statistics....46

Table 3. 3: Tennessee descriptive statistics 47

Table 3.4: Tests performed for validating sexual dimorphism of scapular measurements for males and females of the Athens population....49

Table 3.5: Tests performed for validating sexual dimorphism of scapular measurements for males and females of the Tennessee population....49

Table 3.6: Accuracy classification of the five-variable (5-VM) and two-variable (2-VM) models for the Athens and Tennessee population groups....51

Table 3.7: Chi-squared proportion tests on the classification accuracies of the five-variable (5-VM) and two-variable (2-VM) models 52

Table 3.8: Tests performed for measurement differences between the male individuals of the Athens and Tennessee populations 53

Table 3. 9: Tests performed for measurement differences between the female individuals of the Athens and Tennessee populations....54

Table 3.10: T-tests performed for measurement differences between the male individuals of the Athens population and the “Cleveland” sample 55

Table 3.11: T-tests performed for measurement differences between the female individuals of the Athens population and the “Cleveland” sample................56

Table 3. 12: T-tests performed for measurement differences between the male individuals of the Tennessee population and the “Cleveland” sample.............................57

Table 3.13: T-tests performed for measurement differences between the female individuals of the Tennessee population and the “Cleveland” sample 58

Table 3.14: F-tests for variation significance between the measurements of the Athens population and the “Cleveland” sample..60

Table 3.15: F-tests for variation significance between the measurements of the Tennessee population and the “Cleveland” sample..61

Table 3.16: Accuracy classification of the FORDISC 3.0 analysis for the Athens and Tennessee population groups...64

Table 3.17: Chi-squared proportion test used between the accuracy of the two-variable model and FORDISC 3.0 65

Table 3.18: Intra- and Inter-observer error bias for the five- and two-variable models and the FORDISC 3.0 validation study. 71

Table 3.19: Pearson’s correlation test for age at death of the Athens and Tennessee Population groups for the five- and two-variable models and FORDISC 3.0 72

Table 3.20: Sex and age of individuals for the creation of the models for “Macaluso’s Hypothesis”...73

Table 3.21: Athens descriptive statistics for the creation of the models for “Macaluso’s Hypothesis”...75

Table 3.22: Tennessee descriptive statistics for the creation of the models for “Macaluso’s Hypothesis”...75

Table 3.23: Tests performed for validating sexual dimorphism between males and females of the Athens population..76

Table 3.24: Tests performed for validating sexual dimorphism between males and females of the Tennessee population..76

Table 3.25: Tests performed for measurement differences between the male individuals of the Athens and Tennessee populations 77

Table 3.26: Tests performed for measurement differences between the female individuals of the Athens and Tennessee populations 78

Table 3.27: Binary logistic regression analysis on the Athens population 80

Table 3.28: Hosmer and Lemeshow test for goodness-of-fit on Athens predictive models 80

Table 3.29: Classification plots on the Athens predictive models 81

Table 3.30: Binary logistic regression analysis on the Tennessee population..................83

Table 3.31: Hosmer and Lemeshow test for goodness-of-fit on Tennessee predictive models 83

Table 3.32: Classification plots on the Tennessee predictive models...............................84

Table 3.33: Binary logistic regression analysis on the Combined Population groups...86

Table 3.34: Hosmer and Lemeshow test for goodness-of-fit on the Combined Population groups predictive models 87

Table 3.35: Classification plots on the Combined Population groups predictive models 88

Table 3.36: Chi-squared proportion tests on the classification accuracies of the models for “Macaluso’s Hypothesis” (Athens, Tennessee and Combined Population) ...89

Table 3.37: Pearson’s correlation test for age at death of the Athens and Tennessee Population groups in the creation of the models for “Macaluso’s Hypothesis” 91

Table 3.38: Inter- and Intra- observer error bias for the creation of the models for “Macaluso’s Hypothesis” 92

LIST OF FIGURES

Figure 2.1: Posterior view of the left scapula. The measurements are (1) XLS, (2) XBS, and (3) XHS. The measurements with the asterisk (*) are used for the two-variable model. XHS is used in both the five- and two-variable models.......29

Figure 2.2: Anterior view of the left scapula. The measurement is (4) CSV. The black dot indicates the midpoint between the inferior margin of the glenoid fossa prominence and the spinous axis (the point where the measurement is actually taken).........30

Figure 2.3: Photograph of coordinate caliper used to measure the osteometric point CSV. 31

Figure 2.4: Left lateral view of the left scapula. The measurements are (5) HAX and (6) TLB. 32

Figure 2.5: Posterior view of the left scapula. The measurements for FORDISC 3.0 analysis are (1) XBS and (2) XHS 35

Figure 2.6: Left lateral view of the glenoid cavity. (1) The maximum height of the glenoid cavity. Scale in 10 mm 38

Figure 2.7: Left lateral view of the glenoid cavity. (2) The maximum breadth of the glenoid cavity. Scale in 10 mm 39

Figure 3.1: All Athens individuals who were classified as “male” by FORDISC 3.0. Males (M) are in blue diamonds and females (F) are in red squares. 66

Figure 3.2: All Athens individuals who were classified as “female” by FORDISC 3.0. Males (M) are in blue diamonds and females (F) are in red squares. 67

Figure 3.3: All Tennessee individuals who were classified as “male” by FORDISC 3.0. Males (M) are in blue diamonds and females (F) are in red squares. 68

Figure 3.4: All Tennessee individuals who were classified as “female” by FORDISC 3.0. Males (M) are in blue diamonds and females (F) are in red squares. 69

CHAPTER 1: INTRODUCTION

1.1 Objectives

There is a need within death investigation to find accurate and reliable methodologies to assist with the estimation of sex for unknown human remains. Understanding human anatomy, and its associated diversity, is important to how forensic anthropologists estimate sex from human skeletal remains. Not only does sex constitute an important part of the biological profile, but it also provides the underpinning for a better understanding of other elements of the biological profile. For example, the estimation of stature and age at death are sex dependent. Almost all bones within the human skeleton have been shown to be good predictors of sex for the forensic anthropologist. Within a forensic and archaeological context, however, degradation of the skeletal material may render these bones unusable to the investigator. Research to investigate new methods for estimating sex from other skeletal elements is vital.

The objectives of this thesis are as follows:

1. To understand the relationship between biological sex and estimated sex from the scapula based on metric analyses of two White European population groups from North America and Europe.
2. To test the accuracy and reliability of two scapular methodologies published by Dabbs and Moore-Jensen (2010) and Jantz and Ousley (2005).
3. To provide alternative methods for estimating sex from the scapula, of White European populations, based on metric analyses of the glenoid cavity.

1.2 Concepts of identity in physical anthropology

'Race,' is "a culturally assigned category that is based on having a set of inherited biological traits" (Ember and Ember 1998: 380). This older concept suggests that humans can be classified into groups using physical characteristics, that these physical characteristics are inherited from one generation to the next, and that these groups are by nature unequal and, therefore, can be ranked in order of superiority (Corcos 1997:1). This definition encompasses a social and cultural perspective of human classification that is outmoded.

Human classification and the notion of racial typology began with classifying humans from phenotypic properties, such as skin colour. However, as early as the 19th century, racial typology began to spread to other scientific discourses, including physical anthropology. Craniometry is the study of the size and shape of the skull and its analysis led many scientists to believe that an individual with a larger skull had higher brain function and, therefore, was obviously superior to other 'races' (Corcos 1997). This notion was later discredited by anthropologist Franz Boas in the 20th century because it would be impossible for certain behaviours to be classified by 'racial' typologies (e.g. White Europeans were not actually smarter based on larger brain sizes) (Molnar 1983: 15). However, during the 20th century, anthropologists began to question the use of the term 'race' and its overall applicability to scientific investigation. Therefore, 'race' will not be used in this thesis to differentiate between groups of people.

In an attempt to alleviate the negative connotations associated with human 'racial' classification, but still trying to explain human variation, the term ethnicity was created; ethnicity, or ethnic groups, are groups of people distinguished by cultural similarities, such as beliefs, values, habits, customs, religion, history, and language (Kottak 2007: 59). Ethnicity, however, still uses the social and cultural ideas that relate to an individual's, or a group of individuals', identity (Kottak 2007). These ethnic groups share cultural beliefs, customs, and norms that are part of their common background. Often 'race' and ethnicity are used interchangeably without a clear distinction between the two. Kottak (2007: 62) gives an example of the term "Hispanic" which is an ethnic group but used in the United States as a 'racial' classification. "Hispanics" can be of any 'race' but the cultural norm that unites them is a common cultural style of Spanish-speaking individuals, regardless of whether they are White European or Black African (Kottak 2007).   Ethnicity does not take into account the biological traits that are used in a forensic anthropological investigation and therefore, will not be used in this thesis as a means to differentiate between groups of people.

Estimating biological diversity, or ancestry, has been one of the cornerstones of forensic anthropological investigation since the 19th century. Biological diversity was described in terms of 'race.' Ancestry replaced 'race' and ethnicity in scientific discourse because its definition is founded on the principles of human biology rather than cultural differences. Jorde and Wooding (2004: S30) describe ancestry as, "a more subtle and complex description of an individual's genetic makeup than is race. This is in part a consequence of the continual mixing and migration of human populations throughout history." Because ancestry denotes the genetic relationships and population diversity across the globe, whereas 'race' and ethnicity do not, ancestry will be used in this thesis to describe groups of people of different biological and physical traits.

Ancestry is a term that is commonly used in forensic anthropology when a group of individuals share similar biological traits. However, when a group of people share a geographical location they also share similarities with physical properties, including skeletal biology. Population, therefore, is described as, “a number of individuals who possess a large number of [genetic] characteristics in common, though with some degree of variation” (Molnar 1983: 43). Because these population groups share a geographical locale they create a “gene pool.” This gene pool is “the total aggregate of genes in a population at any one time” (Campbell and Reece 2005: 455). However, these similarities in gene frequency are attributed to the total mean of a population rather than an average, so there is still some degree of population diversity. This diversity is often expressed over a wide range of population groups in neighbouring countries causing a gradient in physical attributes (Molnar 1983). This contributes to global population diversity. Therefore, population and population groups will be used in this thesis to describe geographically distinct groups of individuals.

1.3 Historical background for estimation of sex from human skeletal remains

Almost all bones of the human skeleton have been used to estimate the sex of an individual. The reason is that within a forensic and archaeological context the degradation of bone could sometimes render some skeletal elements unusable. There are two methods used to estimate sex, metric and morphological analyses. These types of methodologies need to be tested and retested for accuracy and reliability, especially those methods that are deemed population specific. Only methods with the highest level of accuracy should be used within a medico-legal context to ensure admissibility of evidence within a court of law.

1.3.1 Estimating sex from the pelvis

In 1969, T.W. Phenice developed a method for estimating sex from the morphological characteristics of the pelvis. This research stemmed from previous studies on the sexual dimorphic traits of the pelvis by Washburn (1948).

The Phenice Method involves visually comparing three aspects of the pelvis: the ventral arc, the subpubic concavity, and the medial aspect of the ischio-pubic ramus. The benefits of this new method were to allow researchers to accurately and reliably estimate the sex of the individual using the os coxae. Two hundred and seventy-five individuals were tested with this method and an accuracy rate of 95% was obtained. It has been shown that some limitations of the results could be from the age of the individual and the biological affinity of the individual (Phenice 1969).

Lovell (1989) tested the Phenice Method on a modern population of 50 individuals. Twelve participants were used to score each of the 50 pubic bones. All of the pubic bones tested were from individuals of White European descent between 52 and 92 years of age. The results of the test showed an accuracy rate of 83% compared to the 95% previously recorded by Phenice. The reliability of the method, i.e. the consistency of accurate classification, was high. However, one of the errors discovered in this research was that as the individual`s age increases the accuracy of the Phenice Method decreases (Lovell 1989).

In 2002, Bruzek developed a method for estimating the sex of an individual from different morphological characteristics of the pelvis. This research stemmed from previous studies on sexual dimorphism of the pelvis by Ferembach and colleagues (1980), Işcan and Derrick (1984), and Phenice (1969). The five characteristics that were visually assessed were the: preauricular surface, greater sciatic notch, form of the composite arch (the anterior arm of the auricular surface), morphology of the inferior pelvis (ischiopubic ramus), and ischiopubic proportions (proportion of the length of the ishium and pubis). Each of these characteristics is significantly sexually dimorphic and the researcher designates male, female, or indeterminate for each characteristic. This method produced an accuracy rate was 98% when evaluating the os coxae (Bruzek 2002).

Population specific traits of the pelvis have also been investigated. In 2003, Patriquin and colleagues observed sexually dimorphic characteristics between South African Whites and Blacks. They examined five visual characteristics of the pelvis: the shape of the greater sciatic notch, subpubic concavity, ischiopubic ramus roughness, orientation of the ischial tuberosity, and the pubic bone shape. Each characteristic was classified as male, female, or intermediate. The researchers found significant sexually dimorphic differences in the pelvis. When comparing their results to other studies they noticed significant sexually dimorphic differences between African American and South African Black populations (Patriquin et al. 2003).

Another study on population diversity was conducted by Walker in 2005. This was a test of estimating sex from the greater sciatic notch, which was developed by Buikstra and Ubelaker (1994). The method was tested on skeletal remains of White European and Black African descent. Walker (2005) reported that the accuracy of the procedure did not decrease between and within population groups. However, Walker (2005) suggested that the age at death of the individual could affect the accuracy of this methodology as younger individuals tended to have more feminine appearing sciatic notches.

In 1941, Letterman conducted a study on the morphology of the greater sciatic notch and its relationship to both sex and ancestry. Letterman (1941) measured the width and height of the greater sciatic notch of individuals from White European and Black African populations. This research showed that this area of the pelvis was sexually dimorphic when subjected to metric analyses. Also, the research found that there were sexually dimorphic differences of the width and height of the greater sciatic notch between White European and Black African ancestral groups.

Flander (1978) studied the sacrum to develop a method for estimating the sex of an individual. Five measurements of the sacrum were taken and subjected to univariate statistical analyses. Those measurements were the mid-ventral line, anterior breadth, maximum articular surface, mid-ventral curve length, and transverse diameters of the S1 body. A discriminate function formula was created from the measurements. The accuracy rate of this method for identifying the sex of an unknown individual ranged from 80% to 94%.

Arsuaga and Carretero (1994) and Gonzalez and colleagues (2009) used multivariate statistical analyses of the pelvis to investigate sexual dimorphism. Both of these studies utilized the skeletal collections at the Museu Antropologico de Coimbra in Portugal. Arsuaga and Carretero (1994) discovered that female pelvic bones were relatively larger with respect to the pelvic inlet. Also, females exhibited a broader sciatic notch. Gonzalez and colleagues (2009) used geometric and morphometric techniques, along with discriminate function analyses, to develop a method for estimating the sex of an individual. This method involved two sexually dimorphic characteristics: the greater sciatic notch and the ischiopubic complex. These two areas were evaluated using targeted landmarks from two-dimensional photographs. The research found that there were marked differences between the sexes with regard to both the shape and size of the sciatic notch and ischiopubic complex. The researchers used this information to formulate a methodology to estimate the sex of an individual using multivariate analyses. When tested against a sample set of individuals from their original skeletal population, the accuracy rate of this method ranged between 90.1% and 93.4%.

Murphy (2000) and Benazzi and colleagues (2008) used metric analyses of the acetabulum to estimate the sex of unknown human remains. Murphy (2000) used skeletal remains from a prehistoric New Zealand population and Benazzi et al. (2008) used Italian remains from the University of Bologna. Murphy (2000) measured the maximum diameter of the acetabulum while Benazzi et al. (2008) measured the perimeter and total area of the acetabulum using digital photographs. In both studies, the measurements were subjected to discriminate function analyses and formulae were derived for estimating the sex of an unknown individual. The accuracy rate for Murphy’s (2000) research was 86.2%. The accuracy rate for Benazzi et al. (2008) ranged from 85.2% to 86.2%. Both of these studies showed estimation of sex to be population specific and that the discriminate function analyses should be recalculated for that specific population to obtain higher levels of accuracy.

1.3.2 Estimating sex from the skull

Many of the methodologies for estimating sex from the skull have originated from early investigations of skeletal analyses by Giles and Elliot (1963), Krogman and Işcan (1986) and Stewart (1979). Buikstra and Ubelaker (1994) and France (1998) developed methodologies for sex estimation based on visible changes in the skull's features. Those features included the mastoid process shape, nuchal crest size, browridge shape, frontal bone angle, supraorbital margin shape, supraorbital ridge shape, and chin size. Each of these characteristics was shown to be sexually dimorphic and a method for scoring their physical changes was developed. This allowed the investigator to visually estimate the sex of an individual based on the skull.

In 1998, Konigsberg and Hens used visual characteristics of the skull and used multivariate cumulative probit models to help estimate the sex of an individual. The sexually dimorphic features evaluated were the superciliary arch form, chin form, mastoid process size, supraorbital margin shape, and nuchal cresting. From these characteristics, logistic regression analyses were used to create single indicator and multivariate indictor models to estimate the sex of an unknown individual. The overall accuracy rate of this method was 81%. However, the overall rate of accurately estimating males was considerable higher than that of females.

Noren and colleagues (2005) and Lynnerup and colleagues (2005) used the petrous part of the temporal bone to estimate the sex of an individual. Noren and colleagues (2005) examined the angle between the lateral part of the internal auditory canal and the medial surface of petrous part of the temporal bone. They tested this method against 113 petrous bones with known sex. The researchers found a significant correlation between sex and the angle of this bone. The accuracy rate of this method is 83.2%.

Lynerup and colleagues (2005) studied the diameter of the internal opening of the acoustic canal in the petrous bone. They measured the diameter of 113 left petrous bones. The results suggested a small measurement difference between males and females. Unfortunately, when the predictive value of this method was tested with inter- and intra-observer error, there is an accuracy rate of 70.0%.

In 1996, Loth and Henneberg developed methods from the mandible to estimate sex. The researchers discovered that males have a distinct angulation of the posterior border of the mandibular ramus, which may be related to development because it only manifests consistently after adolescence. However, in many of the females, the posterior ramus kept the same shape as seen in the juvenile population. When this morphological characteristic was tested in a blind study, the accuracy rate was as high as 99.0% in predicting the sex of an unknown mandible.

Byers (2008) promotes the idea that the skull is the second best indicator for estimating the sex of an individual next to the human pelvis. However, Spradely and Jantz (2010) conducted a study to test the accuracy rates and reliability of sex estimation methodologies from the skull and postcranial elements and found Byers statement to be incorrect. The researchers studied 11 postcranial bone methodologies for estimating the sex of an unknown individual; the bones included the clavicle, scapula, humerus, radius, ulna, sacrum, os coxae, femur, tibia, and fibula. Postcranial bone measurements were then tested against a metric analysis of the skull to determine which measurements were more sexually dimorphic. The researchers discovered that the humerus and radius were the best indictors of sex with an accuracy rate between 93.8% and 94.3%, respectively. The cranium had an accuracy level of 90.0%. All other postcranial elements had an accuracy level between 92.0% and 94.0%. Spradely and Jantz (2010) showed that the postcranial bones provided a better estimation of sex than when only using the skull when using multivariate metric analyses.

1.3.3 Estimating sex from the long bones

Black (1978), MacLaughlin and Bruce (1985), and Safont and colleagues (2000) used bone circumference of the femur to estimate sex from human remains. Safont and colleagues (2000) also included the radius, ulna, and humerus within their analyses. The data from each study were subjected to discriminate function analyses. Black (1978) found that the length of the femur was a more accurate indicator of sex than the circumference of the femoral head. MacLaughlin and Bruce (1985) discovered that the maximum anteroposterior diameter of the femoral shaft was more sexually dimorphic than the midshaft circumference. Safont and colleagues (2000) observed that the circumference of the radius, ulna, and humerus could more accurately estimate the sex of an individual than the femur. They concluded that there was more mechanical stress on the radii, ulnae, and humeri, created distinct sexually dimorphic differences between males and females in the Mediterranean population they examined (Safont et al. 2000).

Charisi and collegues (2011) examined sexually dimorphic traits present in the radii, ulnae, and humeri using metric analyses from a modern Greek skeletal population. The maximum length and epiphyseal widths were measured and then subjected to discriminate function analyses. The results of the study showed sexual dimorphism between the three long bones. The right humerus had the highest accuracy rate (95.7%)

and left ulna had the lowest accuracy rate (90.3%). However, the results were shown to be population specific and the authors suggested that the method should only be used to estimate sex from a modern Greek population.

In 1999, Rogers used four visual characteristics of the posterior distal humerus to create a methodology for estimating the sex of an unknown individual. Those four characteristics were the trochlear constriction, trochlear symmetry, olecranon fossa shape and depth and angle of the medial epicondyle. This method was developed on the Grant Skeletal Collection at the University of Toronto, which consists mainly of White European males. When the author tested this method on skeletal collections in New Mexico and Tennessee the combined accuracy rates were 92% for correctly identifying males and females. Falys and colleagues (2005) conducted a blind study of the Rogers' method on a skeletal collection in Great Britain and found that when all traits were combined to estimate sex an overall accuracy rate of 79.1% was achieved.

Işcan and Shihai (1995), Işcan and colleagues (1998) and King and colleagues (1998) examined sexual dimorphism of the arm and leg bones in different Asian population groups. Işcan and Shihai (1995) and King and colleagues (1998) conducted a study on femora from Thai and Chinese populations. Işcan and colleagues (1998) examined the right humerus in three Asian populations: Chinese, Japanese, and Thai. In each study, the bones were measured from each population group and subjected to both stepwise and discriminate function analyses. Işcan and colleagues (1998) showed that the humerus is sexually dimorphic in all three populations; the Chinese group showed the least sexual dimorphism and the Japanese and Thai groups showed the most sexual dimorphism. In order to create a standard for estimating sex based on the humerus, a discriminate function formula was created for each population group. The authors concluded that, although all three groups were of Asian origin, there was a need for three different discriminate function formulae, i.e. the formulae were population specific.

Srivastava and collegues (2012) conducted a study on estimating sex using femora from a contemporary North Indian population group. Eight parameters were measured and analyzed by discriminant function analyses. The accuracy rate of sex prediction ranged from 70.5% to 83.6%. Also, Milner and Boldsen (2012) examined sexual dimorphism of humeral and femoral head diameters in a contemporary White European population. However, this study did not use discriminate function or logistic regression analyses to develop a methodology to estimate sex. Rather, the authors used probability ratios to help determine whether a group of skeletal remains deviated from fixed measurements of humeral and femoral head diameters. If the unknown individual's humeral or femoral head diameters were in a range above or below those fixed measurements then the individual was classified as either male or female. Their results are used in disaster related fatalities in which there could be an overrepresentation of one sex.

1.3.4 Estimating sex from the metacarpals, carpals, metatarsals and tarsals.

Studies on estimation of sex based on metacarpal measurements have been developed by several researchers (Falsetti, 1995; Scheuer and Elkington, 1993; Smith, 1996; Stojanowski, 1999). All authors used metric measurements from metacarpals one to five to derive equations for the estimation of sex of an unknown individual. However, Smith (1996) also used hand phalanges to create an equation to estimate sex for left and right hands as well as between two population groups: Black African and White

European. The accuracy rates for Scheuer and Elkington (1993), whose sample came from contemporary White European cadaver specimens, ranged from 74.0% to 94.0%. The accuracy rates for assigning both ancestry and sex for the Smith (1996) study ranged from 72.0% to 89.0%. Falsetti (1995) and Stojanowski (1999) both used contemporary White European and Black African skeletal collections and had accuracy rates from 79.0% to 92.0%.

Case and Ross (2007), Lazenby (1994) and Zanella and Brown (2003) validated the research by Falsetti (1995), Scheuer and Elkington (1993) and Stojanowski (1999). The overall results of each study suggested that the accuracy rates that were reported by the original researcher varied considerably when tested on a different population sample. Zanella and Brown (2003) used a contemporary White European cadaveric sample and showed that the methodologies created by Falsetti (1995) and Scheuer and Elkington (1993) had accuracy rates lower than those originally reported. Lazenby (1994) used a 19th century White European population and found that Scheuer and Elkington's (1993) methodology more correctly classified males than females.

Barrio and colleagues (2006), Khanpetan and colleagues (2012) and Manolis and colleagues (2009) created population specific metacarpal sex estimation methodologies. Each study used similar methodologies by Falsetti (1995), Scheuer and Elkington (1993) and Stojanowski (1999), however, Barrio et al. (2006) used a contemporary Spanish population, Khanpetan et al. (2012) used a contemporary Thai population and Manolis et al. (2009) used a contemporary Greek population. The accuracy rates for each methodology were: 81% to 91% for the Barrio et al. (2006), 83.2% to 89.8% for the Khanpetan et al. (2012) and 83.7% to 89.7% for the Manolis et al. (2009).

A preliminary study on the metric methodology for estimating sex from the carpal bones was conducted by Sulzmann and colleagues (2008). The authors used a White European historic 18th and 19th century cemetery population. Each carpal bone was assigned four to nine measurements based on its size and shape and from those measurements discriminate function equations were created to estimate the sex of the individual. The accuracy rates for these equations ranged between 64.6% and 88.6%.

Mastrangelo et al. (2011a) and Mastrangelo et al. (2011b) conducted studies on population specific methodologies for estimating sex from the carpal bones. Mastrangelo and colleagues (2011a) used a 20th century Spanish population and Mastrangelo and colleagues (2011b) used a contemporary Mexican population. Following the methodologies outlined by Sulzmann and colleagues (2008), both studies created discriminate function models for the targeted population group within their study. The accuracy rates for Mastrangelo and colleagues (2011a) and Mastrangelo and colleagues (2011b) ranged between 88.2% to 98.1% and 81.3% to 92.3%, respectively.

Robling and Ubelaker (1997) and Smith (1997) used metric analyses of the metatarsals to develop estimation of sex methodologies for unknown individuals. Robling and Ubelaker (1997) used contemporary White European and Black African individuals. They developed discriminate function models from individual carpal bones and by combining measurements from all five of the carpal bones. The accuracy rates for the Robling and Ubelaker (1997) study ranged between 83.0% and 100.0%. Smith (1997) also used foot phalanges to create an equation to estimate sex for left and right feet as well as between two population groups: Black African and White European. The accuracy rates for assigning both ancestry and sex for the Smith (1997) study ranged from 70.0% to 84.0%.

Mountrakis and colleagues (2010) developed a population specific discriminate function model for the metatarsals. They used a contemporary Greek population group. The results suggested that the metatarsal bones for this population group were highly sexually dimorphic with an accuracy rate from 80.7% to 90.1%.

Steele (1976), Introna and colleagues (1997), and Gualdi-Russo (2007) used metric analyses of the tarsals to develop estimation of sex methodologies for unknown individuals. Steele (1976) used Black African and White European individuals and measured the talus and calcaneus to develop his discriminate function models. The accuracy rates for Steele (1976) ranged from 79.0% to 89.0%. Gualdi-Russo (2007) and Introna and colleagues (1997) used a contemporary Italian sample to develop their methodologies using multivariate discriminate function analyses. However, Introna and colleagues (1997) used only measurements of the calcaneus and Gualdi-Russo (2007) measured the talus and calcaneus. The accuracy rates for assigning the correct sex for Introna and colleagues (1997) was 85.0%. The accuracy rates for the Gualdi-Russo (2007) study ranged from 87.9% to 95.7%.

Bidmos and Asala (2003), Bidmos and Asala (2004) and Bidmos and Dayal (2004) examined sexual dimorphism of the tarsal bones in South African populations. Bidmos and Asala (2003) measured the calcaneus of South African White individuals and Bidmos and Asala (2004) measured the calcaneus of South African Black individuals. The authors' developed ancestry specific discriminate function models from nine parameters of the calcaneus. The average accuracy rate for the Bidmos and Asala (2003) study ranged from 73.0% to 86.0% and the Bidmos and Asala (2004) study ranged from 79.0% to 86.0%. Bidmos and Dayal (2004) measured nine parameters of the talus in South African Black individuals to validate previous discriminate functions equations for estimating sex and developed their own methodology for estimating sex from the talus. The average accuracy rate for the Bidmos and Dayal (2004) study ranged from 80.0% to 89.0%.

Harris and Case (2012) conducted a study to determine which of the seven tarsals would demonstrate the greatest sexual dimorphism and which could be used for accurate sex determination. Eighteen measurements were obtained from the tarsals of contemporary White European males and females. Logistic regression analyses were performed to create equations for sex discrimination. The average accuracy rate for the Harris and Case (2012) study ranged from 88.0% to 92.0%.

1.3.5 Estimating sex from the scapula

One of the first metric studies conducted on the variation of the human scapula was published in 1887 by a medical professor named Thomas Dwight. Dwight collected statistics on scapular indices for people of different ancestries, Native American, White European, and Black African. The two scapular indices were defined as the breadth of the scapula and the infra-spinous index. Dwight used scapular remains from the skeletal collection at the Harvard Medical School, the Boston Society of Natural History, and the Peabody Museum of Archaeology. Assumptions were made regarding the biological affinities of some of the individuals within the study. For example, human remains from Kentucky and California were presumed to be Native American in origin and were labeled “Mound Builders” because they were part of museum collections (Dwight 1887: 629).

Dwight compared the two scapular indices of the “Mound Builder” population to 113 White European scapulae. He concluded that both indices were much smaller in the “Mound Builder” group than in the White European group which he attributed to the individuals’ occupation and health (Dwight 1887). Although estimating sex was not specifically addressed in this study, his research revealed that variation in the scapula could be related to differences in biological sex.

In 1956, Bainbridge and Genovese-Tarazaga examined human scapulae for differences related to sexual dimorphism. The authors employed morphological and metric analyses of the scapula. The morphological characteristics assessed in this study were the costal facets, shape of the glenoid cavity, angle of the axillary border, form of the supraspinous fossa and suprascapular notch. The results were inconclusive for the estimation of sex based on morphological assessment of the scapula. However, the metric analyses proved to be significant for the estimation of sex from the scapula. The metric measurements included the maximum length of the scapula, breadth of the scapula, maximum length of the spine, length of the axillary border, maximum width of the process of teres major, maximum length of the coracoid, length of the glenoid cavity, breadth of the glenoid cavity, maximum and minimum length of the crest of the spine and width of the axillary border. The results showed that the breadth of the glenoid fossa, the maximum breadth of the scapula, and the maximum length of the scapula is sexual dimorphism. The breadth of the glenoid cavity was a significant variable for estimating sex from the scapula. However, only utilizing the glenoid breadth measurement produced a large percentage of individuals as “indeterminate”, i.e. unclassified as either male or female.

Bainbridge and Genovese-Tarazaga (1956) also used the sexually dimorphic characteristics of the scapula to create a method for estimating sex. This method examined the deviation of an established mean of six different measurements of the scapula. The measurements include: breadth of the glenoid fossa, maximum breadth of the scapula, maximum length of the scapula, width of the axillary border, maximum length of the spine and length of the axillary border. The measurement obtained from each area of the scapula has an upper and lower limit from the standard deviation of the mean. If the measurement reaches or exceeds the upper limit then it is classified as male, if the measurement reaches or falls below the lower limit then it is classified to be female, and if the measurement was between the male and female limit it would be considered as "indeterminate".

In 1979, Stewart re-examined Dwight's (1887) original method for estimating sex from the human scapula. Stewart (1979) measured the maximum length of the scapula and the maximum length of the glenoid cavity. Stewart's results are similar to Dwight's findings in that the maximum lengths of female scapulae rarely surpass 14 cm and that the maximum length of male scapulae rarely falls below 17 cm. However, Stewart's (1979) study had limitations in that those individuals who fell between 14 cm and 17 cm were classified as indeterminate. Stewart (1979) also re-examined the length of the glenoid cavity and its relationship to estimating sex. The author measured the glenoid cavity of males and females and discovered that no glenoid cavity that had a total length of 3.6 cm or greater was female and that no glenoid cavity measuring less than 3.6 cm was male.

In 1994, Di Vella and colleagues used a contemporary Italian skeletal population and measured seven areas of the scapula to examine sexually dimorphic traits. The seven measurements were the maximum length of the scapula, maximum breadth of the scapula, maximum distance between the acromion and coracoid, maximum length of the acromion, maximum length of the coracoid and the length and breadth of the glenoid cavity. The researchers conducted multivariate discriminate function analyses on the three most sexually dimorphic measurements (maximum distance between the acromion and coracoid, maximum length of the coracoid, and length of the glenoid cavity) and achieved an accuracy rate of 95% on classifying an individual as either male or female.

Prescher and Klumpen (1995) examined the total area of the glenoid cavity and its relationship to estimating sex from human skeletal remains. The investigator applied adhesive tape to the glenoid cavity and then cut around the edges to outline the total area of the cavity. This adhesive tape, once cut, was removed and placed onto a piece of paper, which was scanned and analyzed using a computer program. The researchers found that glenoid cavities larger than 9.57 cm^2 would be estimated as male and scapulae smaller than 6.83 cm^2 would be estimated as female. The researchers suggested that those individuals who fell between 6.83 cm^2 and 9.57 cm^2 be classified as indeterminate.

In 1997, Prescher and Klumpen conducted a second study, which examined the morphological characteristics of the glenoid cavity. They noticed that a large number of scapulae had a “pear-shaped notch” within the cavity and they wanted to know if this phenomenon was sexually dimorphic in nature. Unfortunately, they found that the pear-shaped notch of the glenoid cavity was not due to sex differences but rather due to an anatomical anomaly. The authors suggested that the notch was related to the glenoid labrum and its tendon that stretches across to the subscapularis muscle in the shoulder joint.

Murphy (2002) and Ozer and colleagues (2006) used scapular measurements and statistical analyses to develop estimation of sex methodologies for unknown individuals. Murphy (2002) used a prehistoric New Zealand population and Ozer and colleagues (2006) used a medieval bone collection from East Anatolia. All studies employed the length and breadth of the glenoid cavity but Ozer and colleagues (2006) also used the maximum length and breadth of the scapula. These measurements were then subjected to discriminate function analyses to illustrate which characteristics were more sexually dimorphic. The measurements were also subjected to logistic regression analyses, which were then used to formulate an equation to estimate the sex of an individual for that specific population. Once a logistic regression equation was formulated for each population, the accuracy rates increased for estimating the sex of individuals within the specific population. The research concluded that high accuracy rates for estimating sex were population specific.

Frutos (2002) and Papaioannou and colleagues (2012) used scapular measurements and clavicle measurements to develop estimation of sex methodologies for unknown individuals. Frutos (2002) used a contemporary Guatemalan population and Papaioannou and colleagues (2012) used a contemporary Cretan population. All studies employed the length and breadth of the glenoid cavity but Papaioannou and colleagues (2012) also used the maximum length of the scapular spine. These measurements were then subjected to discriminate function analyses to illustrate which characteristics were more sexually dimorphic. The measurements were also subjected to logistic regression analyses, which were then used to formulate an equation to estimate the sex of an individual for that specific population. Once a logistic regression equation was formulated for each population, the accuracy rates increased for estimating the sex of individuals within the specific population. The research concluded that high accuracy rates for estimating sex were population specific.

In 2010, Macaluso examined the total area of the glenoid cavity to estimate sex using digital photographs and standard calliper measurements. The methodologies utilized were similar to those of Prescher and Klumpen (1995). However, Macaluso's statistical treatment of the data made the methodology population specific. Macaluso (2010) used the skeletal remains of 120 contemporary Black South African individuals. The author used the software program "ImageJ" to measure digital photographs of the glenoid cavity of each individual. The four measurements used were the height and breadth of the cavity, the perimeter, and total area of the cavity. Macaluso (2010) formulated a logistic regression equation to measure sexual dimorphism of the glenoid cavity using those four measurements. However, the logistic regression equation can only be used for contemporary Black South African populations. The author also measured the height and breadth of the glenoid cavity by using a standard sliding caliper. He reported that the measurements obtained by the two techniques, i.e. digital photographs and sliding caliper measurements were not statistically different. Macaluso (2010) also suggested that one of the most sexually dimorphic measurements, the area of the glenoid cavity, could be obtained by multiplying the height and breadth of the cavity but that method was not tested in his study.

Dabbs and Moore-Jansen (2010) developed two discriminate function models from 23 measurable sexually dimorphic characteristics for estimating sex from the human scapula. The models were developed on contemporary White European and Black African population groups. One model was developed from the five most sexually dimorphic traits: maximum length of the scapular spine, maximum length of the scapula, maximum breadth of the scapula, height of the glenoid prominence, lateral curvature, and thickness of the lateral border. This was called the five-variable model. A second model was created using discriminate function analyses but only using two of the original 23 variables: maximum length of the scapula and maximum breadth of the scapula. The researchers developed this second model as sometimes all five variables were not available due to the diagenesis of bone. This was called the two-variable model.

When Dabbs and Moore-Jansen (2010) tested the five- and two-variable models against a contemporary cadaveric sample of unknown ancestry the accuracies of the five-variable and two-variable models in identifying the sex of an individual ranged from 71.4% to 88.9%. Dabbs and Moore-Jansen (2010) gave four explanations as to the limitations of their models, which were: population diversity, age distribution of the sample, bilateral asymmetry between left and right scapulae, and changes in the size of the scapula based on contemporary and historic population groups.

1.3.6 Estimating sex from other postcranial skeletal elements

Other bones of the human skeleton have also been used to estimate sex of unknown individuals. Most of these methodologies are population specific and have not been tested on other populations. Kim et al. (2006), Kindschuh et al. (2010), and Miller et al. (1998) used the hyoid and found its accuracy rate to be 82.0%- 88.0%. Cologlu et al. (1998), Işcan (1985), Ramadan et al. (2010), and Wiredu et al. (1999) used the sternal ribs ends and found their accuracy rate to be 74.0% to 90.0%. Marino (1995), Marlow and

Pastor (2011), and Wescott (2000) used the vertebral column and found its accuracy rate to be 75.0% to 85.0%.

1.4 Osteological collections utilized for this research

1.4.1 University of Athens Human Skeletal Reference Collection

The University of Athens Human Skeletal Reference Collection (The Athens Collection) consists of 250 individuals. The Athens Collection is housed at the Department of Animal and Human Physiology, at the University of Athens, Greece. The individuals were acquired from cemeteries within the area of Athens. According to funerary customs of Greece, individuals are buried for a period of three to five years and then exhumed to be placed in ossuaries. Living members of the deceased must pay “rent” to keep their loved ones in the ossuary otherwise the skeletons are placed in a large underground pit located in the cemetery (Eliopoulos et al. 2007). Since the 1990’s, skeletal remains have been donated to the University of Athens through a legal agreement with the municipalities; the skeletal remains of deceased individuals whose family members are unable to pay “rent” for their tomb are donated to the University of Athens. Complete demographics about each individual are known as death certificates provide information on sex, age, cause of death, occupation, and place of birth therefore, providing the most accurate comparative sample from which to study human variation (Eliopoulis et al. 2007). The collection represents individuals who have lived within the last half of the 20th century. This collection was chosen for this current research project because it represents a contemporary White European population from Greece.

1.4.2 William Bass Donated Skeletal Collection

The William Bass Donated Skeletal Collection consists of 900 individuals. The collection is housed at the Department of Anthropology, the University of Tennessee in Knoxville. The individuals are donated by the families of the deceased or willed by the deceased themselves prior to death. Some individuals were donated by the medical examiner's office in Tennessee. The collection consists of males and females, subadults and adults, from infancy to older adulthood, of White European, Black African, Asian and Hispanic ancestry. This collection was chosen for this thesis project because it represents a contemporary White European population from North America.

CHAPTER 2: MATERIALS AND METHODS

2.1 Skeletal materials utilized for this research

In this study, 335 individual scapulae from two skeletal reference collections, the Athens Collection and the William Bass Collection, were measured. Only left scapulae of White European individuals were measured as by standards set by other researchers and to maintain anatomical consistency (Builkstra and Ubelaker, 1994; Stewart, 1979). Only scapulae of adult individuals (ages 20+ years old) were used as subadult scapulae have not reached their maximum size. Damaged or remodelled scapulae were excluded from the study. Post-mortem damage consisted of scapulae that exhibited chipping or wear on any of the scapular landmarks. Ante-mortem remodelling was present primarily within the glenoid cavity. This remodelling consisted of osteophytes, which would create larger measurements taken from the glenoid cavity so these individuals were removed from the study.

The University of Athens Human Skeletal Reference Collection (The Athens Collection) consists of 250 individuals. The Athens Collection is housed at the Department of Animal and Human Physiology, at the University of Athens, Greece. Of the 250 individuals present in the Athens Collection, 77 males and 50 females were used for the Dabbs and Moore-Jensen (2010) validation study and the FORDISC 3.0 validation study (Jantz and Ousley 1993). There were 95 males and 74 females used for the new methodologies (the models for “Macaluso’s Hypothesis”), created by the present author, for estimating sex from the glenoid cavity.

The William Bass Donated Skeletal Collection (The Tennessee Collection) consists of 900 individuals. The collection is housed at Department of Anthropology, University of Tennessee in Knoxville. Of the 900 individuals in the Tennessee Collection, 94 males and 76 females were used for the Dabbs and Moore-Jensen (2010) validation study and the FORDISC 3.0 validation study (Jantz and Ousley 2005). There were 92 males and 74 females used for the new methodologies, created by this author, for estimating sex from the glenoid cavity (the models for “Macaluso’s Hypothesis”).

2.2 Methods

2.2.1 Dabbs and Moore-Jansen (2010) five-variable and two-variable models

The five-variable and two-variable models are two metric methodologies for estimating sex from the scapula (Table 2.1). These models were generated using six of the most sexually dimorphic measurements from a list of 23 scapular measurements (Tables 2.2 and 2.3; Figures 2.1- 2.4). The composition of the skeletal material used to create these equations consisted of 804 North American White and Black populations from the Cleveland Museum of Natural History Skeletal Reference Collection. This sample used by Dabbs and Moore-Jansen (2010) will be labeled the “Cleveland” sample for the rest of this thesis.

This study followed the same methodologies as those outlined in Dabbs and Moore-Jansen (2010). Those same six measurements were taken from the left scapula of each individual, from both the Athens and Tennessee skeletal reference collections, and inputted into the equations from Table 2.1. The classification accuracy of each model was determined by its potential to correctly classify an individual as being either male or female.

Table 2.1: Equations for the Dabbs and Moore-Jansen (2010) models for estimating sex from the scapula.

Model	Equation
Five-variable	Sex* = (0.136 x XLS) + (0.117 x XHS) + (0.541 x HAX) + (0.296 x CSV) + (0.904 x TLB) – 66.186
Two-variable	Sex* = (0.212 x XBS) + (0.201 x XHS) – 51.425

*Sex = >0 individual is male; <0 individual is female. Result in y-value

Table 2.2: Description of measurements for the five-variable model (Adapted from Dabbs and Moore-Jansen, 2010).

Measurement	Description (recorded to the nearest tenth of a millimeter)
Maximum Length of Spine (XLS)	Sliding calipers were used to measure from the medial margin of the scapula at the spinous axis to the most lateral point on the scapular spine.
Maximum Length of Scapula (XHS)	Sliding calipers were used to measure from the superior point on the superior angle to the most inferior point on the inferior angle.
Height of Glenoid Prominence (HAX)	Spreading calipers were used to measure from the superior margin of the glenoid prominence to the inferior margin of the glenoid prominence.
Lateral Curvature (CSV)	Coordinate calipers were used to measure the distance from parallel at the midpoint between the inferior margin of the glenoid fossa prominence and the spinous axis.
Thickness of Lateral Border (TLB)	Sliding calipers were used to measure the thickness of the border at the midpoint between the inferior margin of the glenoid prominence and the inferior angle. The measurement should be taken perpendicular to the scapular body.

Table 2.3: Description of measurements for the two-variable model (Adapted from Dabbs and Moore-Jansen, 2010).

Measurement	Description (recorded to the nearest tenth of a millimeter)
Maximum Breadth of the Scapula (XBS)	Sliding calipers were used to measure from the lateral surface of the glenoid dorsal cavity to the spinous axis.
Maximum Length of Scapula (XHS)	Sliding calipers were used to measure from the superior point on the superior angle to the most inferior point on the inferior angle.

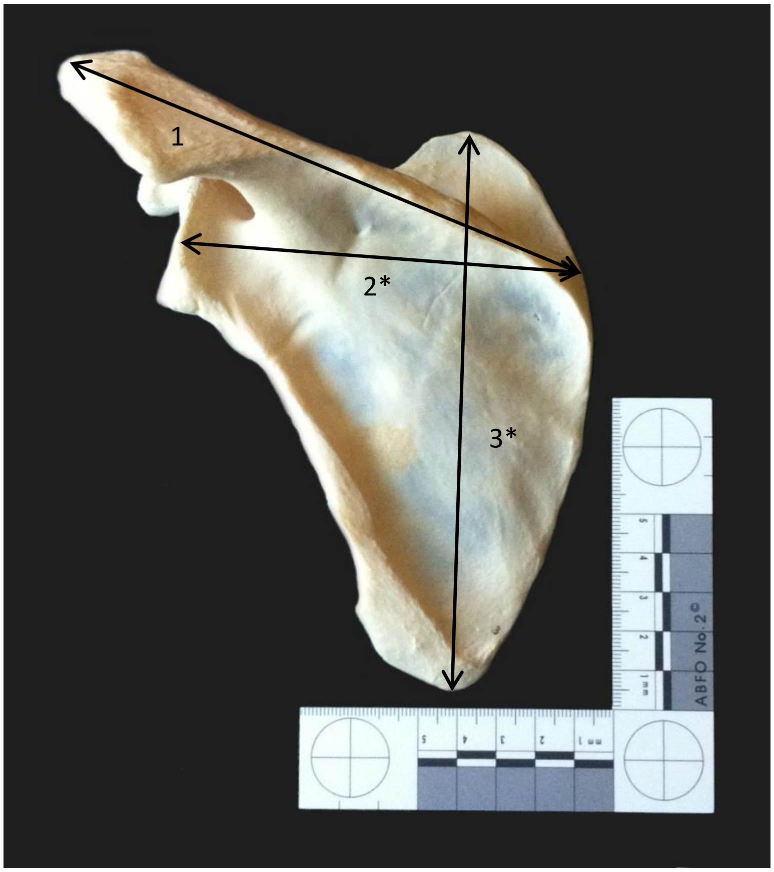

Figure 2.1: Posterior view of the left scapula. The measurements are (1) XLS, (2) XBS, and (3) XHS. The measurements with the asterisk (*) are used for the two-variable model. XHS is used in both the five- and two-variable models. (Photo by Ian Bell and James Neish).

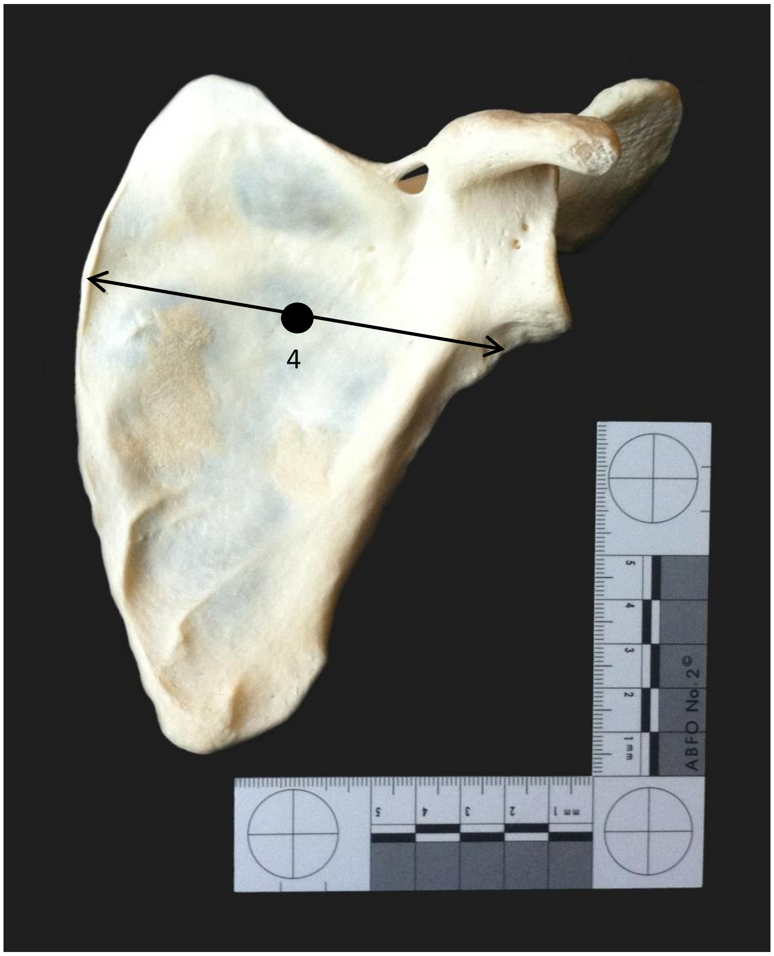

Figure 2.2: Anterior view of the left scapula. The measurement is (4) CSV. The black dot indicates the midpoint between the inferior margin of the glenoid fossa prominence and the spinous axis (the point where the measurement is actually taken). (Photo by Ian Bell and James Neish).

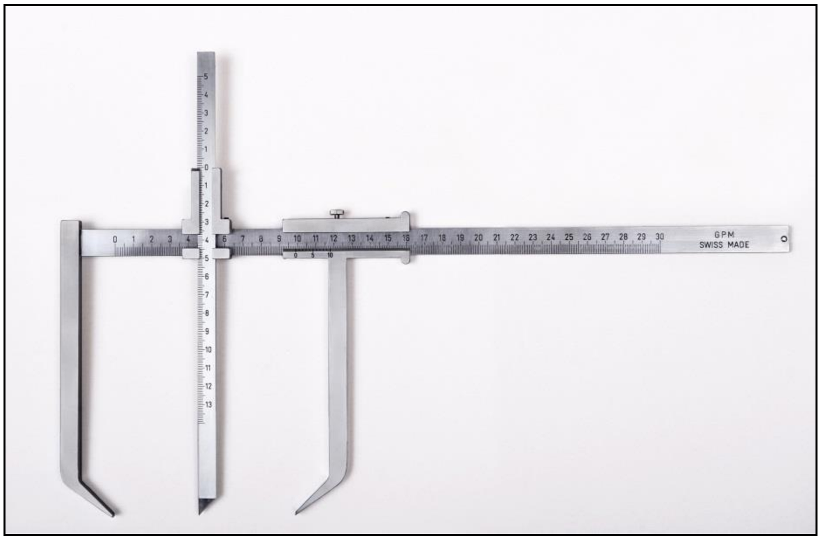

Figure 2.3: Photograph of coordinate caliper used to measure the osteometric point CSV[1].

---

[1] Retrieved on August 4th, 2012 from http://www.antropolog-instrument.ru/catalog.php.

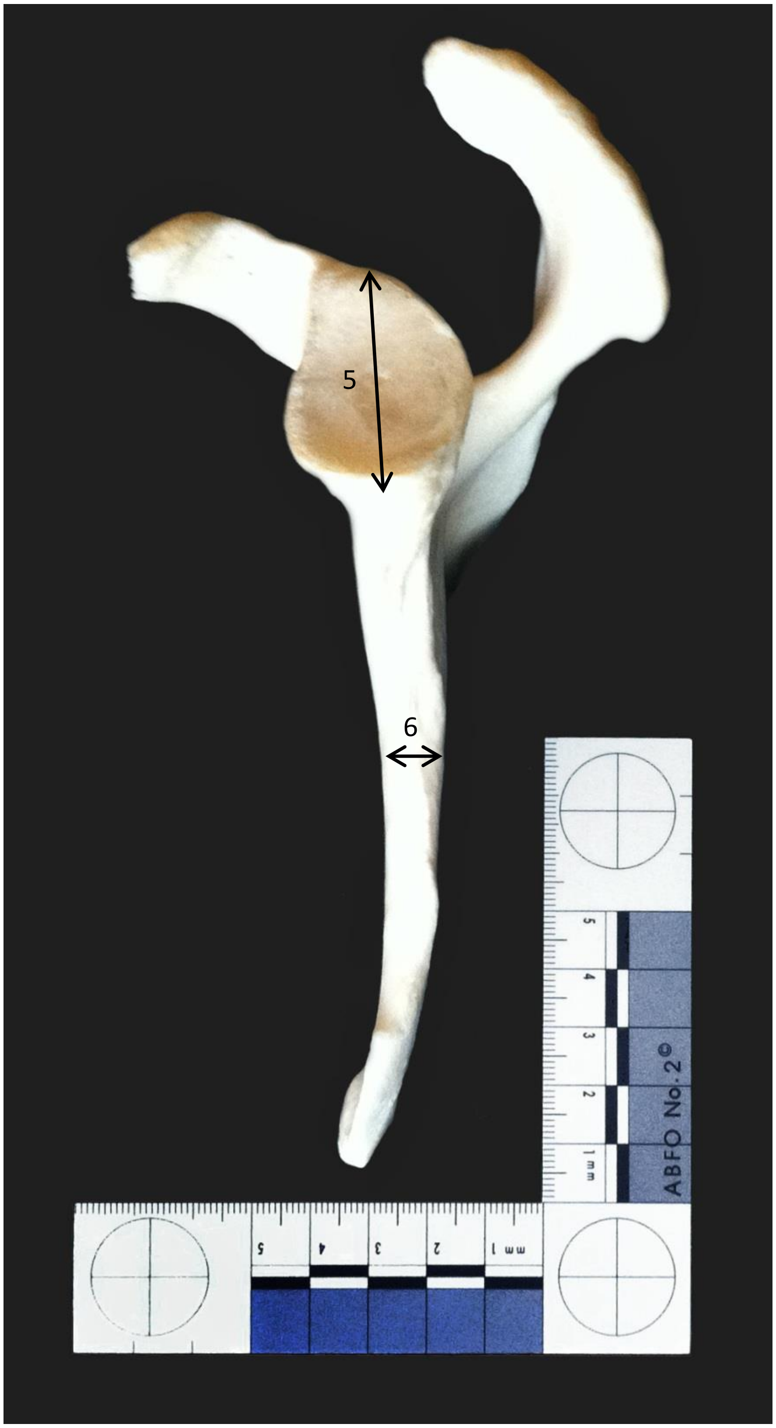

Figure 2.4: Left lateral view of the left scapula. The measurements are (5) HAX and (6) TLB. (Photo by Ian Bell and James Neish).

2.2.2 FORDISC 3.0 by Jantz and Ousley (2005)

FORDISC 3.0 is a computer program that assists in the identification of unknown skeletonized individuals. It was developed by Stephen Ousley and Richard Jantz at the University of Tennessee. This program performs two primary functions. Firstly, by using cranial and postcranial measurements it classifies unknown individuals into ancestry and sex groups by using multiple discriminant function analyses (Sanders, 2002). Secondly, this program compares profiles of known individuals from its database to unknown skeletal elements to aid in the identification of the unknown individuals.

This program allows external data (the height and breadth of the scapula) to be inputted into its discriminant function formulae in which it generates a result for the classification of male or female. The outputted data relates to the number of individuals cross-referenced, the accuracy of the classification formulae, the posterior probability and the typicality probability for all groups sampled in the Forensic Data Bank. The posterior probability is defined as the likelihood that the inputted sample data are similar to other individuals within a group or category. The typicality probability determines how similar or typical the inputted sample data are to the FORDISC 3.0 database. The closer both probabilities are to 1.0 the stronger the interpretation of the results. However, if the posterior probability of a test is very high and the typicality probability is low, the calculated results are still considered accurate for classifying an individual into their appropriate sex or ancestry because the likelihood that the individual is male or female is not determined by typical probability.

In some cases, tests are run multiple times to eliminate outlier groups which could affect the posterior and typicality probabilities. Tests are re-run until the result is close to a minimum of 80% accuracy. Also, it is important to note that FORDISC 3.0 is biased when it performs post-cranial ancestry analyses as it only uses White European and Black African groups as comparative populations.

In this study, scapulae from both the Athens and Tennessee collections were used to evaluate the FORDISC 3.0 software. The data were collected using standard measurements from Buikstra & Ubelaker (1994) and entered into the FORDISC 3.0 program (Table 2.4 and Figure 2.5). The interface of FORDISC 3.0 allows the user to select ancestral and sex categories for the analyses of post-cranial elements. FORDISC 3.0 has four categories: “Black Females,” “Black Males,” “White Females,” and “White Males.” The ancestry of all individuals in this study was known to be White European from two geographically different population groups. Therefore, the author only selected a White European population group (i.e. “White Females” and “White Males”) within the FORDISC 3.0 program for the post-cranial sex estimation analyses. This was done to eliminate the outliers that could confound the data and misclassify an individual by ancestry. By eliminating the Black African category (i.e. “Black Females” and “Black Males”) in the FORDISC 3.0 analyses, the results reflected the accuracy of the sex estimation discriminate functions of FORDISC 3.0 and not the classification of the ancestry discriminate functions.

Table 2.4: Description of measurements for the FORDISC 3.0 program (Adapted from Buikstra and Ubleaker, 1994).

Measurement*	**Description (recorded to the nearest tenth of a millimeter)**
Scapula: Height (XHS)	Sliding calipers were used to measure the direct distance from the most superior point of the cranial angle to the most inferior point on the caudal angle.
Scapula: Breadth (XBS)**	Sliding calipers were used to measure the distance from the midpoint on the dorsal border of the glenoid fossa to midway between the ridges of the scapular spine on the vertebral border.

*The measurements for the two-variable model and the measurements for the FORDISC 3.0 program are identical. Therefore, the acronym XHS will be used for the height of the scapula and the XBS will be used for the breadth of the scapula in both the two-variable model and the FORDISC 3.0 analyses.

**In this study, a sliding caliper was used for this measurement. Buikstra and Ubleaker (1994) cite the use of spreading calipers for this measurement however using sliding calipers is also an appropriate tool, and does not change the accuracy rate, for this measurement (Dabbs , Jantz, Ubelaker, personal communication, 2012)

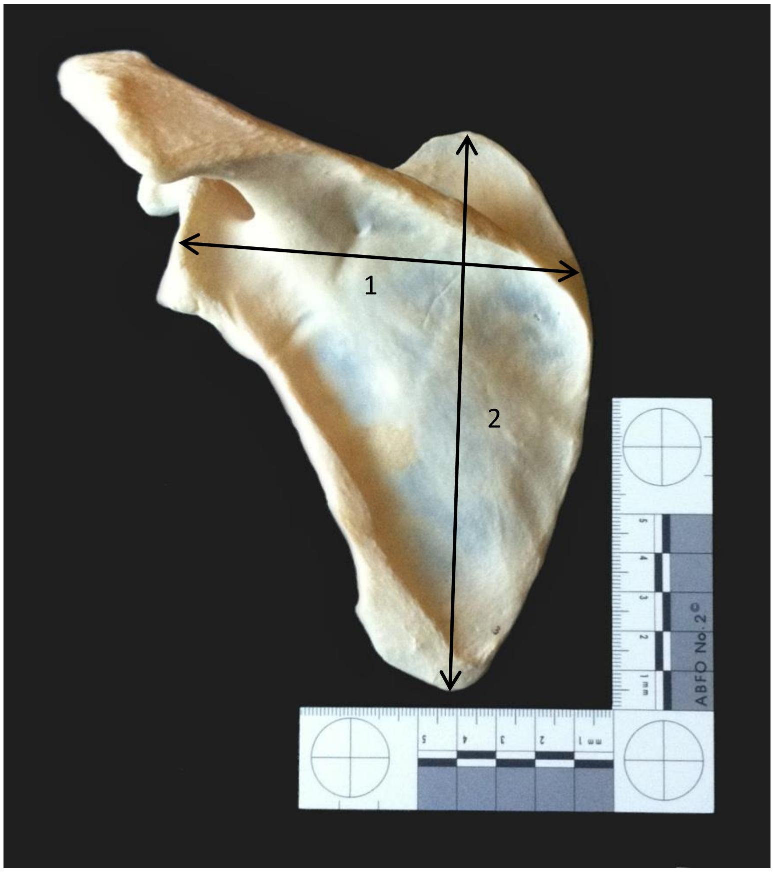

Figure 2.5: Posterior view of the left scapula. The measurements for FORDISC 3.0 analysis are (1) XBS and (2) XHS. (Photo by Ian Bell and James Neish).

2.2.3 The creation of the models for "Macaluso's Hypothesis" for estimating sex from glenoid cavity measurements

Since Dwight (1894), the glenoid cavity has been used to estimate sex of unknown human remains. Recently, Macaluso (2010) used digital photographs of the glenoid fossa to estimate the sex of unknown Black South African populations. Macaluso (2010) estimated sex by collecting measurements from the height, breadth, area and perimeter of the glenoid cavity using the program "ImageJ." Although "ImageJ" is a free software program offered by the National Institutes of Health, the user requires training on how to properly take measurements from this software. Macaluso (2010) suggested that the area of the glenoid cavity could be obtained without using "ImageJ". If the height and breadth of the glenoid cavity were measured with sliding calipers, directly from the skeletal remains, then the calculated area of the glenoid could be obtained by multiplying the height and the breadth. He found that the area obtained by using "ImageJ" and the calculated area obtained directly from the skeletal remains had the same predictive accuracies for estimating the sex of an individual. To test this hypothesis, the current study examined the sexually dimorphic nature of the measured height, breadth, and calculated area of the glenoid cavity to establish new methods for estimating sex from the scapula in White European populations without using the program "ImageJ". The calculated area is not the true area of the glenoid fossa, in this study the calculated area is being used as a mathematical surrogate to confirm or dispute Macaluso's hypothesis that conventional measurements taken from the specimen can be used for sex determination.

The height and breadth of each left scapula were measured and the calculated area determined (Table 2.5; Figures 2.6 and 2.7). Binary logistic regression analysis was then used to create new methods for estimating sex from the glenoid cavity from the Athens and Tennessee collections. These new methods for estimating sex will be called the models for “Macaluso’s Hypothesis” for the remainder of the thesis.

Table 2.5: Description of measurements used for the models for “Macaluso’s Hypothesis” for estimating sex of the glenoid cavity (Adapted from Macaluso, 2010).

Measurement	**Description (recorded to the nearest tenth of a millimeter)**
Height of the glenoid cavity	Sliding calipers were used to measure the maximum distance across the glenoid cavity perpendicular to the anteroposterior axis
Breadth of the glenoid cavity	Sliding calipers were used to measure the maximum distance across the glenoid cavity measured at a right angle to the axis of the length of the glenoid cavity
Calculated area of the glenoid cavity	The height of the glenoid cavity was multiplied by the breadth of the glenoid cavity*.

*The calculated area is not the true area of the glenoid fossa but rather a mathematical surrogate in the shape of a rectangle.

Figure 2.6: Left lateral view of the glenoid cavity. (1) The maximum height of the glenoid cavity. Scale in 10 mm. (Photo by Ian Bell and James Neish).

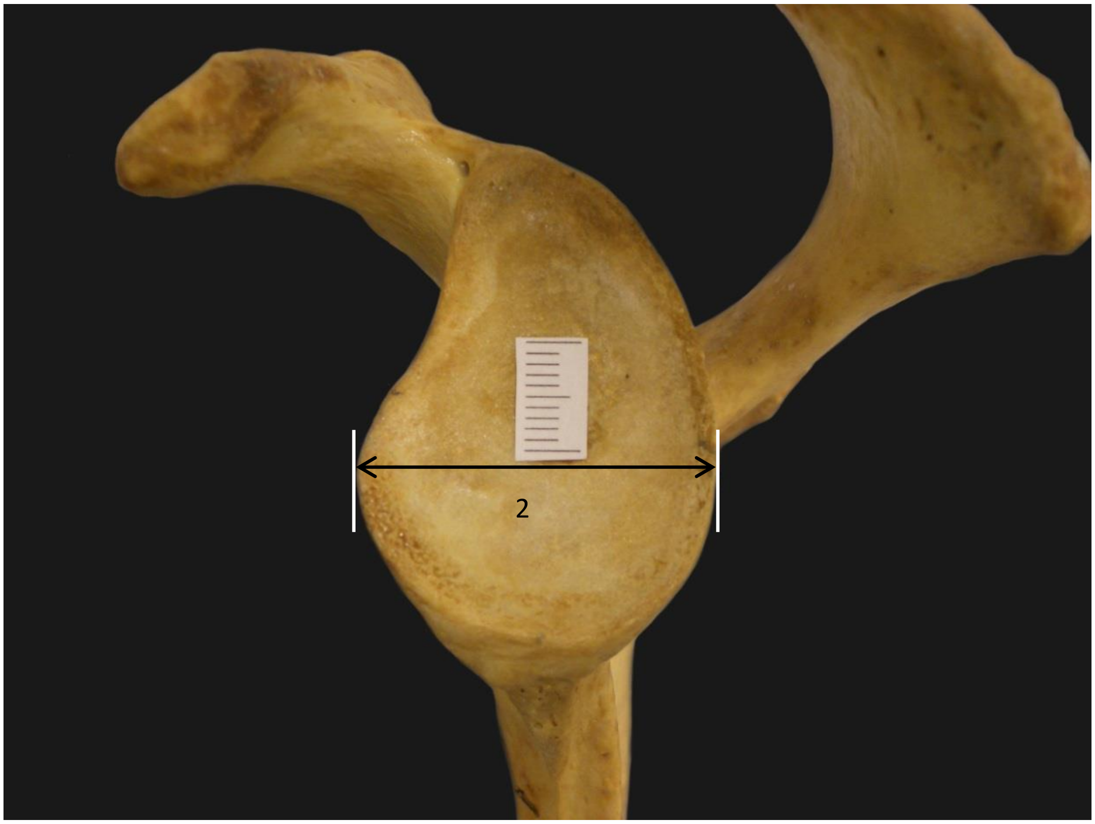

Figure 2.7: Left lateral view of the glenoid cavity. (2) The maximum breadth of the glenoid cavity. Scale in 10 mm. (Photo by Ian Bell and James Neish)

2.3 Statistical analyses

In the current study, all statistical analyses were performed using MiniTab 16.0 statistical software package or the Statistical Package for Social Science 17.0 (SPSS). MiniTab 16.0 was used for the two-sample t-tests, Mann-Whitney U tests, paired t-tests, and f-tests. However, SPSS was used for the binary logistic regression analyses because this software has a more user-friendly interface when calculating complex multivariate analyses (Acocke 2005).

A test for normality was performed in order to determine what statistical analyses would be most appropriate for this data. A normal distribution occurs when the results are normally distributed over a bell curve and are not the result of chance outcomes (Moore 2010). Two-sample t-tests and Mann-Whitney U tests were performed on both males and females of the Athens and Tennessee collections to statistically evaluate the five- and two-variable methodologies, FORDISC 3.0, and the models for "Macaluso's Hypothesis". Two-sample t-tests and Mann-Whitney U tests were then used to assess the sexual dimorphic credibility of the measurements of the five- and two-variable models, as well as, to assess whether these measurements are different between population groups. This will help assess accuracy of the models between individuals of similar ancestral groups. Two-sample t-tests and Mann-Whitney U tests were also used to assess the sexually dimorphic credibility of the height, breadth, and calculated area of the glenoid cavity for the models for "Macaluso's Hypothesis" for estimating sex.

A methodology can be considered reliable if the measurements for that methodology are taken correctly or consistently. To ensure measurement accuracy and avoid measurement bias, intra- and inter- observer measurement error was examined. A reliable method is dependable and will give the statically similar result every time. Intra-observer error is the within group bias in which the researcher measures a specific variable two or more times to avoid measurement error. Inter- observer error is the between group bias in which the researcher employs one or more outsiders to measure a specific variable to avoid measurement error. Intra- and inter- observer measurement errors were examined using paired t-tests. Paired t-test are used to compare two sets of means where one set of data can be paired with a second set of data, usually before and after an experiment (Moore 2010). The purpose of using paired t-tests for this study was to evaluate measurement reliability between the five-variable model measurements, two-variable model measurements, FORDISC 3.0 measurements, and the models for

“Macaluso’s Hypothesis”. The author re-measured the five-variable model measurements, two-variable model measurements, FORDISC 3.0 measurements, and the measurements on the models for “Macaluso’s Hypothesis” on a subsample of 30 individuals from both the Athens and Tennessee population groups. Also, the author had an assistant, with forensic anthropology experience, re-measure the same subsample of 30 scapulae from both the Athens and Tennessee collections. Paired t-tests were performed on the original data and the re-measured data.

A chi-squared proportions test was used to examine if there was a significant difference between the accuracies of the two methodologies, i.e. the two-variable model and the five-variable model or the two-variable model and the FORDISC 3.0 program, within each population group. The chi-squared distribution tests the difference among two or more proportions (Moore 2010). The chi-squared tests analyze the proportions of accuracy, those individuals who were correctly classified and those individuals who were not, between the Athens and Tennessee population samples. The chi-squared tests analyzed the proportions of accuracy between the Athens and Tennessee population samples and the “Cleveland” sample; the original calibration sample used by Dabbs and Moore-Jansen (2010) is called the “Cleveland” sample. By using chi-squared tests, the results show which model, the five- or the two-variable model, is more accurate for estimating sex from the scapula for each population group.

Multivariate discriminate function analyses employ multiple variables that are used to determine group membership (e.g., males or females). If the difference between the measurements' standard deviation is too great, the individuals inputted into the model’s equation could be misclassified. To test for differences in standard deviation between measurements of the five-variable model and two-variable model, f-tests were performed. These tests were used to examine if any statistical similarities, or differences, existed between measurement variations by evaluating standard deviation. The Athens and Tennessee population groups were tested against the original standard deviations of the “Cleveland” sample. If statistical differences occur between the standard deviation of the measurements it would have an effect on the accuracy of sex classification of the five- and two-variable models for both the Athens and Tennessee populations.

Logistic regression analysis was used to create the equations for the models for “Macaluso’s Hypothesis” for estimating sex from the glenoid cavity for the Athens and Tennessee population groups. Logistic regression is the impact of multiple independent variables presented simultaneously to predict membership into one or the other of two dependent variable categories (Burns and Burns 2008: 569). The glenoid cavity measurements (height, breadth, and calculated area of the glenoid fossa) are used as the independent variables that help predict membership into either maleness or femaleness. This is done by the creation of a coefficient (β) for each independent variable. This coefficient measures the independent variables’ (height, breadth, and calculated area of the glenoid fossa) contribution to the variations in the dependent variables’ (males or females) outcome. A backward stepwise procedure is conducted to help discriminate those independent variables that have the most statistical strength for creating a model that can allocate group membership (males and females) based on the most sexually dimorphic measurements. This is done through the subtraction of independent variables (i.e., one at a time) to the backward stepwise model in order to create the best model for differentiating between males and females. The goal is to use the most sexually dimorphic measurements to create the best models for estimating sex.

Pearson's Correlation tests were performed on all measurements of the five- and two-variable models and all of the glenoid cavity measurements for the models for "Macaluso's Hypothesis". This test was used to establish whether or not there was a correlation between age-at-death and measurement size within the Athens and Tennessee population groups. The Pearson's Correlation test is used to compare the positive or negative correlation between two variables. The two variables are the age of the individual at the time of their death and independent variables (all individual measurements for the five-, two-, and the models for "Macaluso's Hypothesis"). The purpose for using the Pearson's Correlation test is to test if the age of the individual affects the accuracy of estimating sex when using the five- and two-variable models. The Pearson's Correlation also tested if age affects the height, breadth, and calculated area of the glenoid fossa for both population groups when creating the models for "Macaluso's Hypothesis.

CHAPTER 3: RESULTS

This research aimed to test the accuracy of three sex estimation methods on White European populations living in different geographic areas. The objectives of this project were:

1. To understand the relationship between biological sex and estimated sex from the scapula based on metric analyses of two White European population groups from North America and Europe.
2. To test the accuracy of the five- and two-variable models by Dabbs and Moore-Jensen (2010) and the FORDISC 3.0 program by Jantz and Ousley (2005) for estimating sex from two White European populations.
3. To develop new methodologies for estimating sex from the glenoid cavity in two White European population groups.

3.1 The five- and two-variable models and FORDISC 3.0 validation study.

Two hundred and ninety-seven individuals from two skeletal reference collections, the Athens Collection and the William Bass Collection, were measured. Only left scapulae of White European individuals were measured as per standards set by other researchers and to maintain anatomical consistency (Buikstra and Ubelaker 1994). The collections represent individuals who lived within the last half of the twentieth century. These collections were chosen because they represent contemporary White European populations from two geographically distinct areas. All individuals were born around, or after, the year 1940. Table 3.1 shows the descriptive statistics for each population sample. The original calibration sample by Dabbs and Moore-Jansen (2010) will now be labeled the "Cleveland" sample for simplicity.

Table 3.1: Sex and age of individuals used for the five- and two-variable models and FORDISC 3.0 validation study

Sex	Population	N	Mean (Years)	SD (Years)	Median (Years)	Minimum (Years)	Maximum (Years)
Male	Athens	77	57.5	17.4	59	25	94
	Tennessee	94	47.5	9.1	49	19	62
Female	Athens	50	61.1	17.2	63	22	85
	Tennessee	76	52.6	7.9	53	31	67

3.2 Descriptive statistics of the measurements for the five- and two-variable models and FORDISC 3.0 validation study

The five-variable and two-variable models are two metric methodologies for estimating sex from the scapula that were developed by Dabbs and Moore-Jansen (2010). These models were created using six of the most sexually dimorphic measurements from a list of 23 scapular measurements. Those six measurements are: maximum length of the spine (XLS), maximum length of the scapula (XHS), maximum breadth of the scapula (XBS), height of the glenoid prominence (HAX), depth of the lateral curvature (CSV), and the thickness of the lateral border (TLB). The specific measurements used for the five-variable model are: XLS, XHS, HAX, CSV, and TLB. The specific measurements used for the two-variable model are: XHS and XBS (Dabbs and Moore-Jansen 2010).

The measurements used for the FORDISC 3.0 analysis are the scapular height (XHS) and scapular breadth (XBS) (Jantz and Ousley 2005). These measurements are similar to those of the two-variable model except that the FORDISC 3.0 methodology requires the necessary computer program to run the analysis. Tables 3.2 and 3.3 show the descriptive statistics (mean, standard deviation (SD), minimum measurement length (MinL), and maximum measurement length (MaxL)) for each measurement used in the Dabbs and Moore Jansen (2010) and the Jantz and Ousley (2005) validation studies. Table 3.2 shows the data for the Athens population group and Table 3.3 shows the data for Tennessee population group.

Table 3. 2: Athens descriptive statistics

Sex	Measurement	Mean (mm)	SD (mm)	MinL (mm)	MaxL (mm)
Males	XLS*	142.10	8.28	121.20	162.08
	XHS*^	159.90	9.23	134.88	184.90
	XBS^	106.70	6.73	91.55	123.90
	HAX*	42.16	2.75	36.00	47.50
	CSV*	6.51	2.34	0.50	14.00
	TLB*	9.74	1.22	7.15	12.60
Females	XLS*	125.90	7.25	112.52	149.58
	XHS*^	137.70	8.17	119.84	151.08
	XBS^	94.84	5.67	83.70	115.30
	HAX*	37.39	2.01	34.00	44.50
	CSV*	5.46	1.83	1.00	10.50
	TLB*	8.01	1.21	5.80	11.50

*These measurements were used for the five-variable model.
^These measurements were used for the two-variable model and the FORDISC 3.0 analysis.

Table 3. 3: Tennessee descriptive statistics

Sex	Measurement	Mean (mm)	SD (mm)	MinL (mm)	MaxL (mm)
Males	XLS*	144.61	7.02	126.60	163.08
	XHS*^	163.13	9.23	146.28	186.62
	XBS^	107.99	5.79	92.75	122.45
	HAX*	42.66	2.33	38.00	49.00
	CSV*	5.26	2.26	0.00	13.00
	TLB*	10.24	1.53	7.15	14.50
Females	XLS*	127.33	6.35	115.10	143.72
	XHS*^	141.32	7.34	123.24	154.92
	XBS^	95.83	5.06	86.20	108.20
	HAX*	37.05	1.85	33.00	42.00
	CSV*	3.95	2.12	0.00	9.00
	TLB*	7.91	1.20	5.55	11.70

*These measurements were used for the five-variable model.
^These measurements were used for the two-variable model and the FORDISC 3.0 analysis.

3.2 Results of Dabbs and Moore-Jansen (2010) validation study

3.2.1 Two-sample t-test vs. Mann-Whitney U test

To statistically evaluate the data for the Dabbs and Moore-Jansen (2010) validation study, two-sample t-tests and Mann-Whitney U tests were performed. These tests are used to compare the similarities or differences between two variables or, in this case, to determine if two measurements are statistically different. The difference between the two tests is that the t-test is used when data are normally distributed over the density curve (parametric) and the Mann-Whitney U test is used when the data are not normally distributed (non-parametric). Each measurement, for both sexes in the two population groups, was evaluated for normality using MiniTab 16. If the data were normally distributed then a t-test was performed and if the data were not normally distributed then a Mann-Whitney U test was used. These tests were used to evaluate whether the measurements for the five- and two-variable models were sexually dimorphic, as well as, to determine whether these measurements were different between population groups. This helped assess the accuracy of the models between individuals of similar ancestral groups.

3.2.2 Sexual dimorphic variation between males and females of the Athens and Tennessee samples

Dabbs and Moore-Jansen (2010) used six of the most sexually dimorphic measurements from a list of 23 scapular measurements. To test if these measurements were sexually dimorphic for the Athens and Tennessee population groups, two-sample t-tests and Mann-Whitney U tests were performed on each measurement. Tables 3.4 and 3.5 show the individual tests performed on the measurements used in the two-variable and five-variable models. These statistical analyses were performed to examine whether each measurement, within each population group, was sexually dimorphic. The results show that within each population group all measurements were sexually dimorphic ($p<0.01$). Therefore, each measurement has the potential to have good predictive value for classifying males and females in both population groups using the five-variable and two-variable models.

Table 3.4: Tests performed for validating sexual dimorphism of scapular measurements for males and females of the Athens population

Measurement	Test	T-Value (t-Test)	U-Value (Mann-Whitney)	P-Values*
XLS	T-Test	11.60	N/A	0.000*
XHS	T-Test	14.22	N/A	0.000*
XBS	T-Test	10.70	N/A	0.000*
HAX	T-Test	12.46	N/A	0.000*
CSV	Mann-Whitney	N/A	5463.50	0.008*
TLB	T-Test	7.84	N/A	0.000*

*Significance was established as $p < 0.01$

Table 3.5: Tests performed for validating sexual dimorphism of scapular measurements for males and females of the Tennessee population

Measurement	Test	T-Value (T-Test)	U-Value (Mann-Whitney)	P-Values*
XLS	T-Test	16.83	N/A	0.000*
XHS	T-Test	17.17	N/A	0.000*
XBS	T-Test	14.60	N/A	0.000*
HAX	Mann-Whitney	N/A	11403.50	0.000*
CSV	T-Test	3.88	N/A	0.000*
TLB	Mann-Whitney	N/A	10833.50	0.000*

*Significance was established as $p < 0.01$

3.2.3 Accuracy of the two-variable and five-variable models

Table 3.6 shows the accuracy of the two-variable and five-variable models for males and females of the Athens and Tennessee samples. Table 3.6 shows the number of individuals that were correctly classified, the number that were incorrectly classified, and the percentage of those accuracies. In the Athens sample, males were more accurately classified using the five-variable model (94.8%) whereas females were classified with the same accuracy (94.0%) using both the five- and the two-variable models. Overall, classification accuracy of the Athens population group shows that the five-variable model has a greater accuracy (94.5%) than the two-variable model (89.7%) in classifying both males and females. The sex bias ratio percentage shows how biased the model is toward classifying one sex over the other. For the Athens population, the five-variable model correctly classified males more often than females by 0.8%. The two-variable model classified females more accurately than males by 7.0%. In the Tennessee sample, both males and females were more accurately classified using the five-variable model (96.8% and 94.7%, respectfully). Overall, classification accuracy of the Tennessee population group shows that the five-variable model had a greater accuracy (95.9%) than the two-variable model (89.4%) in classifying both males and females. The sex bias ratio for the Tennessee population correctly classified males more often than females by 2.1% for the five-variable model and classified males more often than females by 11.8% for the two-variable model. These accuracies show overall percentages without any correlation between the population groups and because each population group has a different number of individuals, the accuracy reflects those numbers. To understand which model is more accurate with each population group and sex, a chi-squared proportion test was used.

Table 3.6: Accuracy classification of the five-variable (5-VM) and two-variable (2-VM) models for the Athens and Tennessee population groups

Sample	(N)	Correct (N)		Incorrect (N)		Percentage Accuracy (%)		Sex Bias* (%)	
		5-VM	*2-VM*	*5-VM*	*2-VM*	*5-VM*	*2-VM*	*5-VM*	*2-VM*
Athens - Males	77	73	67	4	10	94.8	87.0	0.8	-7.0
Athens - Females	50	47	47	3	3	94.0	94.0		
Athens - Total	**127**	**120**	**114**	**7**	**13**	**94.5**	**89.7**		
Tennessee-Males	94	91	89	3	5	96.8	94.7	2.1	11.8
Tennessee-Females	76	72	63	4	13	94.7	82.9		
Tennessee-Total	**170**	**163**	**152**	**7**	**18**	**95.9**	**89.4**		

*Sex bias % = % males correctly classified - % females correctly classified

3.2.4 Chi-squared proportion tests of the five-variable and two-variable models

A chi-squared proportions test was used to examine if there was a significant difference between the accuracies of the five-variable model and the two-variable model, between each population group. The chi-squared distribution tests the difference among two or more proportions (Moore 2010). Table 3.7 shows the results of the chi-squared tests analyzing the proportions of accuracy, i.e. those individuals who were correctly classified and those individuals who were not, between the Athens and Tennessee population samples. Also, Table 3.7 shows the resulting p-values of the chi-squared tests that analyzed not only the proportions of accuracy between the Athens and Tennessee populations but also the “Cleveland” sample. These results indicate which model is more accurate for estimating sex in each population group.

Table 3.7: Chi-squared proportion tests on the classification accuracies of the five-variable (5-VM) and two-variable (2-VM) models

Sample	Sex	*5-VM*		*2-VM*	
		Chi-squared value	***P-value***	***Chi-squared value***	***P-value****
Athens- Tennessee	Male	0.433	0.511	3.11	0.078
Athens-Tennessee	Female	0.031	0.860	3.355	0.067
Athens-Tennessee	**Total**	**0.315**	**0.575**	**0.010**	**0.922**
Athens-“Cleveland”	Male	0.005	0.946	0.330	0.566
Athens-“Cleveland”	Female	1.262	0.261	0.028	0.867
Athens-“Cleveland”	**Total**	**0.449**	**0.503**	**0.308**	**0.579**
Tennessee-“Cleveland”	Male	0.574	0.449	2.599	0.107
Tennessee-“Cleveland”	Female	1.024	0.311	11.160	0.001*
Tennessee-“Cleveland”	**Total**	**0.002**	**0.966**	**0.586**	**0.444**

*Significance was established as $p < 0.05$

All of the resulting p-values were greater than $p=0.05$ except for one. This shows that all proportions that were analyzed were statistically similar except for the proportions between the Tennessee and “Cleveland” female sample ($p=0.001$). These results indicate that the population groups measured, whether it was the Athens and Tennessee, the Athens and “Cleveland”, or the Tennessee and “Cleveland”, have accuracy proportions that are similar for the five- and two-variable models. This is important because these results show whether or not these models can be used with similar degrees of accuracy for estimating sex as described by the Dabbs and Moore-Jansen’s (2010) original method.

The resulting p-value, for the female analyses between the Tennessee and "Cleveland" samples, was statistically different at $p=0.001$. What these results indicate is that White North American females are being misclassified more often with the two-variable model. To understand why this phenomenon is occurring, the specific measurements, for each population group, need to be compared.

3.2.5 Statistical variation between same sex measurements: comparing the Athens and Tennessee populations

Two sample t-tests and Mann-Whitney U tests were performed on same sex measurements, i.e. male measurements of one population compared to male measurements of a second population or female measurements of one population compared to female measurements of a second population, to determine if there were any measurement differences between the two populations. Tables 3.8 and 3.9 show the resulting p-values of the statistical variation tests comparing the measurements of males and females between the Athens and Tennessee population samples.

Two measurements in Table 3.8, the maximum breadth of the scapula (XBS) and the height of the glenoid prominence (HAX), have resulting p-values greater than p=0.05. These measurements are therefore statistically similar within both the Athens and Tennessee groups. All other measurements had resulting p-values lower than p=0.05 illustrating that these measurements were statistically different between the two population groups.

Table 3.8: Tests performed for measurement differences between the male individuals of the Athens and Tennessee populations

Measurement	Test	T-Value (t-test)	U-Value (Mann-Whitney)	P-Values*
XLS	T-Test	-2.09	N/A	0.038*
XHS	T-Test	-2.25	N/A	0.026*
XBS	T-Test	-1.33	N/A	0.187
HAX	T-Test	-0.02	N/A	0.982
CSV	T-Test	3.54	N/A	0.001*
TLB	Mann-Whitney	N/A	5943.0	0.035*

*Significance was established as $p < 0.05$

Four measurements in Table 3.9, the maximum length of the spine (XLS), the maximum breadth of the scapula (XBS), the height of the glenoid prominence (HAX), and the thickness of the lateral border (TLB), had resulting p-values greater than p=0.05. The other two measurements had resulting p-values lower than p=0.05 illustrating that these measurements were statistically different between the two population groups.

Table 3. 9: Tests performed for measurement differences between the female individuals of the Athens and Tennessee populations

Measurement	Test	T-Value (t-test)	U-Value (Mann-Whitney)	P-Values*
XLS	T-Test	-1.10	N/A	0.275
XHS	T-Test	-2.15	N/A	0.014*
XBS	T-Test	-1.00	N/A	0.318
HAX	Mann-Whitney	N/A	3369.5	0.333
CSV	Mann-Whitney	N/A	7628.5	0.000*
TLB	T-Test	0.42	N/A	0.677

*Significance was established as $p < 0.05$

The overall results from Tables 3.8 and 3.9 show that there are some measurement variations between the males and females of the Athens and Tennessee population groups. These results show that the males and females of these two population groups vary biologically based on geography with regard to these specific measurements. Classification accuracy, however, is still high based on the percentiles calculated in Table 3.4 and the chi-squared tests calculated between the Athens and Tennessee males and females presented in Table 3.5.

3.2.6 Statistical variation between same sex measurements: comparing the Athens and "Cleveland" populations

Two sample t-tests were performed on same sex measurements, i.e. male measurements of one population compared to male measurements of another population and female measurements of one population compared to female measurements of another population, to determine if there were any significant differences between the measurements of the Athens population group and the "Cleveland" sample. Tables 3.10 and 3.11 show the resulting p-values of the statistical variation tests comparing the measurements of males and females between the Athens population sample and the "Cleveland" sample.

Three measurements in Table 3.10, the maximum length of the spine (XLS), the maximum breadth of the scapula (XBS) and the depth of the lateral curvature (CSV), have resulting p-values greater than p=0.05. These measurements are therefore statistically similar within the Athens population group and the "Cleveland" sample. All other measurements had resulting p-values lower than p=0.05 illustrating that these measurements were statistically different between the Athens population group and the "Cleveland" sample.

Table 3.10: T-tests performed for measurement differences between the male individuals of the Athens population and the "Cleveland" sample

Measurement	**T-Value**	**P-Values***
XLS	0.76	0.450
XHS	-2.26	0.026*
XBS	-0.23	0.817
HAX	4.49	0.000*
CSV	0.37	0.710
TLB	-3.16	0.002*

*Significance was established as p< 0.05

Four measurements in Table 3.11, the maximum length of the spine (XLS), the maximum height of the scapula (XHS), the maximum breadth of the scapula (XBS), and the thickness of the lateral border (TLB), had resulting p-values greater than p=0.05. All other measurements had resulting p-values lower than p=0.05 illustrating that these measurements were statistically different between the Athens population group and the "Cleveland" sample.

Table 3.11: T-tests performed for measurement differences between the female individuals of the Athens population and the "Cleveland" sample

Measurement	**T-Value**	**P-Values***
XLS	0.61	0.541
XHS	-1.97	0.053
XBS	-0.28	0.782
HAX	9.57	0.000*
CSV	3.11	0.003*
TLB	0.76	0.449

*Significance was established as $p< 0.05$

The overall results from Tables 3.10 and 3.11 show that there are some measurement variations between the males and females of the Athens population group and the "Cleveland" sample. Even though measurement variation existed between the Athens and "Cleveland" sample, the classification accuracy rate was not affected. This is illustrated by the percentiles tabulated in table 3.4 and the chi-squared tests calculated between males and females of the Athens and "Cleveland" samples shown in Table 3.5.

3.2.7 Statistical variation between same sex measurements: comparing the Tennessee and "Cleveland" populations

Two sample t-tests were performed on same sex measurements, i.e. males compared to males or females compared to females, to determine if there were any significant differences between the measurements of the Tennessee population group and the "Cleveland" sample. Tables 3.12 and 3.13 show the resulting p-values of the statistical variation tests comparing the measurements of males and females between the Tennessee population sample and the measurements of the "Cleveland" sample.

Three measurements in Table 3.12, the maximum height of the scapula (XHS), the maximum breadth of the scapula (XBS), and the thickness of the lateral border (TLB), had resulting p-values greater than p=0.05. These measurements are therefore statistically similar between the Tennessee population group and the "Cleveland" sample. All other measurements had resulting p-values lower than p=0.05 illustrating that these measurements were statistically different between the Tennessee population group and the "Cleveland" sample.

Table 3. 12: T-tests performed for measurement differences between the male individuals of the Tennessee population and the "Cleveland" sample

Measurement	T-Value	P-Values*
XLS	-4.02	0.000*
XHS	-0.61	0.545
XBS	-1.66	0.100
HAX	-5.49	0.000*
CSV	4.30	0.000*
TLB	-0.06	0.954

*Significance was established as p< 0.05

Three measurements in Table 3.13, the maximum height of the scapula (XHS), the maximum breadth of the scapula (XBS), and the thickness of the lateral border (TLB), had resulting p-values greater than p=0.05. All other measurements had resulting p-values lower than p=0.05 illustrating that these measurements were statistically different between the Tennessee population group and the "Cleveland" sample.

Table 3.13: T-tests performed for measurement differences between the female individuals of the Tennessee population and the "Cleveland" sample

Measurement	T-Value	P-Values*
XLS	-2.51	0.013*
XHS	-1.13	0.261
XBS	-1.13	0.259
HAX	-10.73	0.000*
CSV	2.15	0.033*
TLB	-0.26	0.794

*Significance was established as $p < 0.05$

The overall results show that there are some measurement variations between the males and females of the Tennessee population group and the "Cleveland" sample (Tables 3.12 and 3.13). Even though measurement variations occurred between the males and females of these two samples, classification accuracy of the five-variable model was not affected. Tables 3.4 and 3.5 indicate that the percent of accuracy are the lowest for the Tennessee females. The chi-squared test for the two-variable model showed a significant difference in accuracy rates between the females of the Tennessee population group and the "Cleveland" sample. However, the results in Table 3.13 show that the two measurements employed in the two-variable model (Maximum height of the scapula (XHS) and maximum breadth of the scapula (XBS)) are statistically similar between the females of the Tennessee population group and the "Cleveland" sample. This indicates that measurement variation between the Tennessee females and the “Cleveland” females are not related to the overall difference in the accuracy rates for estimating sex in the two-variable model.

3.2.8 Variation of standard deviation using f-tests

The standard deviations of the measurements themselves are another factor that affects the accuracy rates of the five- and two-variable models. To assess the differences between the standard deviation of the measurements a series of f-tests were performed on all measurements from all population groups (Athens, Tennessee, and “Cleveland”). T-tests and Mann-Whitney U tests are used to determine similarities or differences between the averages of two measurements, however f-tests are used to measure the standard deviation, or how much the data deviates from that average, between two measurements.

Multivariate discriminate function analyses employ multiple variables or measurements that are used to determine group membership (e.g., males or females). The range of measurement variation, or standard deviation, between two samples could have an overall effect on the accuracy rates of the five- and two-variable models in estimating sex. To test for differences in standard deviation between measurements f-tests were performed. These tests were used to determine if any statistical similarities, or differences, existed between measurement variations by evaluating the standard deviation. The Athens and Tennessee population groups were each tested against the “Cleveland” sample. If statistical differences occur between the standard deviation of the measurements it could affect the accuracy of sex classification of the five- and two-variable models for either the Athens or Tennessee population group. Because the five-and two-variable models were developed on a specific White European and Black African American population group from the Cleveland Museum of Natural History, the range of variation between the measurements could affect the accuracy of group membership. Tables 3.14 and 3.15 show the resulting p-values of the f-tests performed on the standard deviations of the measurements used in the Dabbs and Moore-Jansen's (2010) five- and two-variable models. Table 3.14 shows the resulting p-values of the f-tests comparing the males and females of the Athens population group to the "Cleveland" sample. Table 3.15 shows the resulting p-values of the f-tests comparing the males and females of the Tennessee population group to the "Cleveland" sample.

Table 3.14: F-tests for variation significance between the measurements of the Athens population and the "Cleveland" sample

Sex	Measurement	F-value	P-Value*
Males	XLS	0.91	0.541
	XHS	1.09	0.646
	XBS	0.84	0.285
	HAX	1.39	0.080
	CSV	1.27	0.197
	TLB	1.37	0.090
Females	XLS	0.92	0.670
	XHS	1.14	0.594
	XBS	0.86	0.455
	HAX	0.97	0.850
	CSV	1.54	0.070
	TLB	0.84	0.394

*Significance was established at $p<0.05$

Table 3.15: F-tests for variation significance between the measurements of the Tennessee population and the "Cleveland" sample

Sex	Measurement	F-Value	P-Value*
Males	XLS	1.26	0.174
	XHS	1.09	0.610
	XBS	1.13	0.470
	HAX	1.93	0.000*
	CSV	1.36	0.068
	TLB	0.87	0.377
Females	XLS	1.20	0.346
	XHS	1.41	0.077
	XBS	1.08	0.702
	HAX	1.15	0.489
	CSV	1.15	0.486
	TLB	0.86	0.372

*Significance was established at $p<0.05$

All resulting p-values presented in Table 3.14 are statistically similar to each other. This suggests that there are no differences between the standard deviations of the measurements between the Athens population group and the "Cleveland" sample. One measurement in Table 3.15, the height of the glenoid prominence (HAX), for the Tennessee males has a p-value lower than $p=0.05$. This shows that there is a difference between the standard deviation of that measurement between the males of the Tennessee population group and the "Cleveland" sample. The f-tests show that the range of variation between each measurement of the five- and two-variable models, in the Athens and Tennessee population groups, are not statistically different from the "Cleveland" sample. Therefore, the range of measurement variation between the "Cleveland" sample and those individuals of the Athens and Tennessee population groups is not a factor in the accuracy rates of the five- and two-variable models when applied to the Athens and Tennessee population groups.

3.3 Results of the FORDISC 3.0 validation study

3.3.1 FORDISC 3.0 analyses

Classification accuracy was examined by how precise FORDISC 3.0 was at assigning group membership (i.e. males and females). Because FORDISC 3.0 and the two-variable model use the same measurements (XHS and XBS), this project assessed which methodology was more accurate in classifying males and females from the scapula. Classification accuracy was tested by how well FORDISC 3.0 classified individuals that are geographically distinct (i.e. an Athens population group versus a Tennessee population group). To do this, posterior and typical probabilities were used to assess how likely an individual was correctly classified when that individual's scapular measurements were analyzed with the FORDISC 3.0 program.

FORDISC 3.0 is a metric analysis program that contains a database of skeletal measurements from the Forensic Data Bank, which it uses to make and test custom discriminant functions for estimation of ancestry, sex, and stature (Ousley and Jantz, 2005). The program was developed by Steven Ousley and Richard Jantz at the University of Tennessee. Using standard postcranial measurements of the scapula, the program can classify unknown individuals into ancestry and sex groups by using multiple discriminant function analyses and comparing it to the database. In this research, ancestry is a known variable, however, one of the purposes of this is study is to examine the accuracy and reliability of the FORDISC 3.0 program for estimating sex when employing skeletal remains from a similar biological ancestry but from different geographic locations.

FORDISC 3.0 uses two measurements of the scapula to estimate sex: scapular height (XHS) and scapular breadth (XBS) (Jantz and Ousley, 2005). The process of classifying individuals as either male or female is determined by the results of the posterior and typical probabilities. The posterior probability is defined as the likelihood that the inputted sample data is similar to other individuals within a group or category. The typicality probability determines how similar the inputted sample data is to the database. The closer both probabilities are to 1.0 the stronger the interpretation of the results. However, if the posterior probability of a test is high and the typicality probability is low, the classification of the individual's sex is still considered correct because the likelihood that the individual is male or female is not determined by typical probability.

3.3.2 Accuracy of FORDISC 3.0 when applied to the Athens and Tennessee population groups

Table 3.16 shows the accuracy of the FORDISC 3.0 analysis on both the Athens and Tennessee samples for males and females. Table 3.16 shows the number of individuals that were correctly classified, the numbers that were incorrectly classified, and the percentage of those accuracies. In the Athens sample, females (98.0%) were more often correctly classified than males (81.8%) with a sex bias ratio of classifying females over males by 16.2%. In the Tennessee sample, females (97.3%) were more often correctly classified than males (88.3%) with a sex bias ratio of classifying females over males by 9.0%. Overall classification accuracy of the Tennessee population group showed a higher rate of classification (92.3%) as compared to the Athens population group (88.2%).

Table 3.16: Accuracy classification of the FORDISC 3.0 analysis for the Athens and Tennessee population groups

Sample	Total (N)	Correct (N)	Incorrect (N)	Percentage Accuracy (%)	Sex Bias Ratio (%)*
Athens - Males	77	63	14	81.8	-16.2
Athens - Females	50	49	1	98.0	
Athens – Total	**127**	**112**	**15**	**88.2**	
Tennessee-Males	94	83	11	88.3	- 9.0
Tennessee-Females	76	74	2	97.3	
Tennessee - Total	**170**	**157**	**13**	**92.3**	

*Sex bias % = % males correctly classified - % females correctly classified

3.3.3 Chi-squared test results for comparing the accuracy of the two-variable model with FORDISC 3.0.

A chi-squared proportions test was used to examine if there was a significant difference between the accuracies of the FORDISC 3.0 program and the two-variable model for estimating sex from the scapula within the Athens and Tennessee population groups. The chi-squared distribution tests the differences between two or more proportions (Moore, 2010).

Table 3.17 shows the results of the chi-squared tests analyzing the proportions of accuracy, i.e. those individuals who were correctly classified, between the two-variable model and FORDISC 3.0, for each population group. Most of the resulting p-values were all greater than p=0.05. This shows that those accuracy proportions were statistically similar. This indicates that there were no differences between the accuracies of the two models, for estimating sex, within those population groups. The only exception was the analysis of females in the Tennessee population group. The resulting p-value was 0.003, which suggests those proportions were statistically different. These data, along with the percentiles presented in Tables 3.4 and 3.16, indicate that the two-variable model was less accurate in correctly classifying females in the Tennessee population group than was the FORDISC 3.0 methodology.

Table 3.17: Chi-squared proportion test used between the accuracy of the two-variable model and FORDISC 3.0

Sample	**Sex**	**Chi-Squared Value**	**Chi-Squared P-Value***
Athens	Male	0.790	0.374
Athens	Female	1.042	0.307
Athens	**Total**	**0.161**	**0.689**
Tennessee	Male	2.459	0.117
Tennessee	Female	8.950	0.003*
Tennessee	**Total**	**0.887**	**0.346**

*Significance was established at $p<0.05$

3.3.4 Descriptive statistics of the discriminate function analyses of the Athens population group

Figures 3.1 and 3.2 show the scatter plot results of the typical and posterior probability p-values of all the individuals in the Athens population group as classified by FORDISC 3.0. Figure 3.1 shows all of the individuals who were classified as male by FORDISC 3.0. Only one female individual was classified as a male in this population group. Figure 3.2 shows all of the individuals that were classified as female by FORDISC 3.0. Fourteen male individuals were classified as female. A large majority of females had a posterior probability greater than 0.95, which suggests a larger accuracy classification for estimation of sex from the scapula in females.

Overall, almost all of the individuals in the Athens population group were classified correctly as is illustrated by the high levels of posterior probability. However, for both males and females, the typical probability ranges from very low to very high. This indicates that although the individuals are being properly classified, FORDISC 3.0 has difficulty placing these individuals into their respective sex groups. The scapular measurements taken from the Athens population group are different than those used by the Forensic Data Bank and the Athens Collection is not incorporated into the Forensic Data Bank. Therefore, these factors may account for the problems with sex classification. If the typical probability were higher the classification reliability of FORDISC 3.0 would be greater.

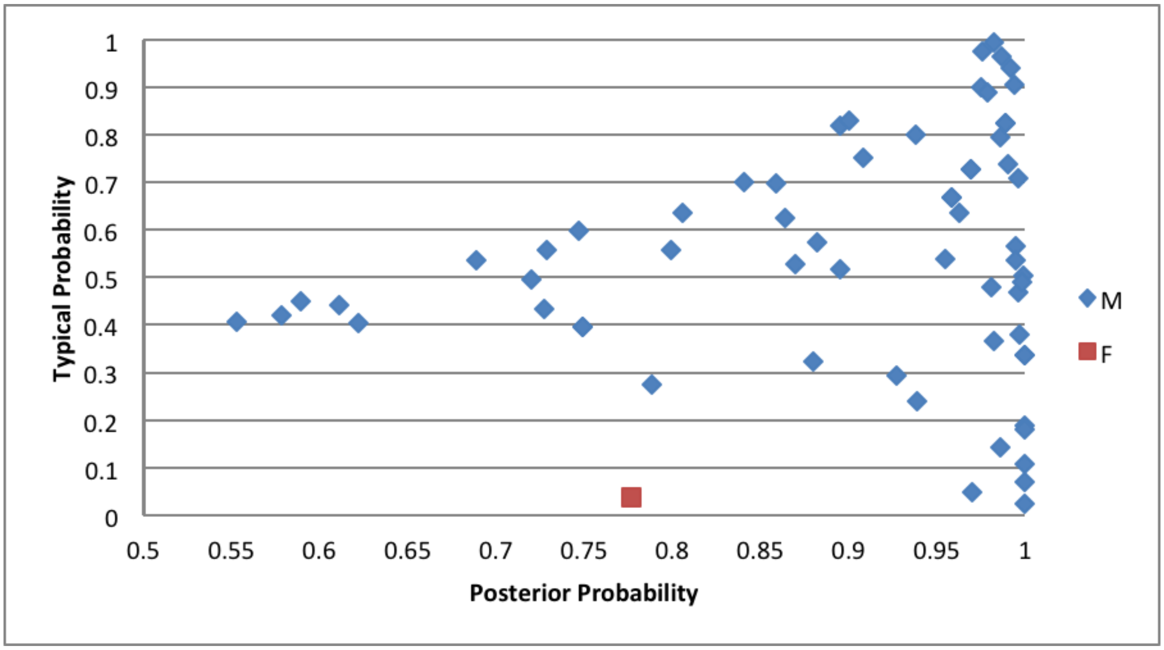

Figure 3.1: All Athens individuals who were classified as "male" by FORDISC 3.0. Males (M) are in blue diamonds and females (F) are in red squares

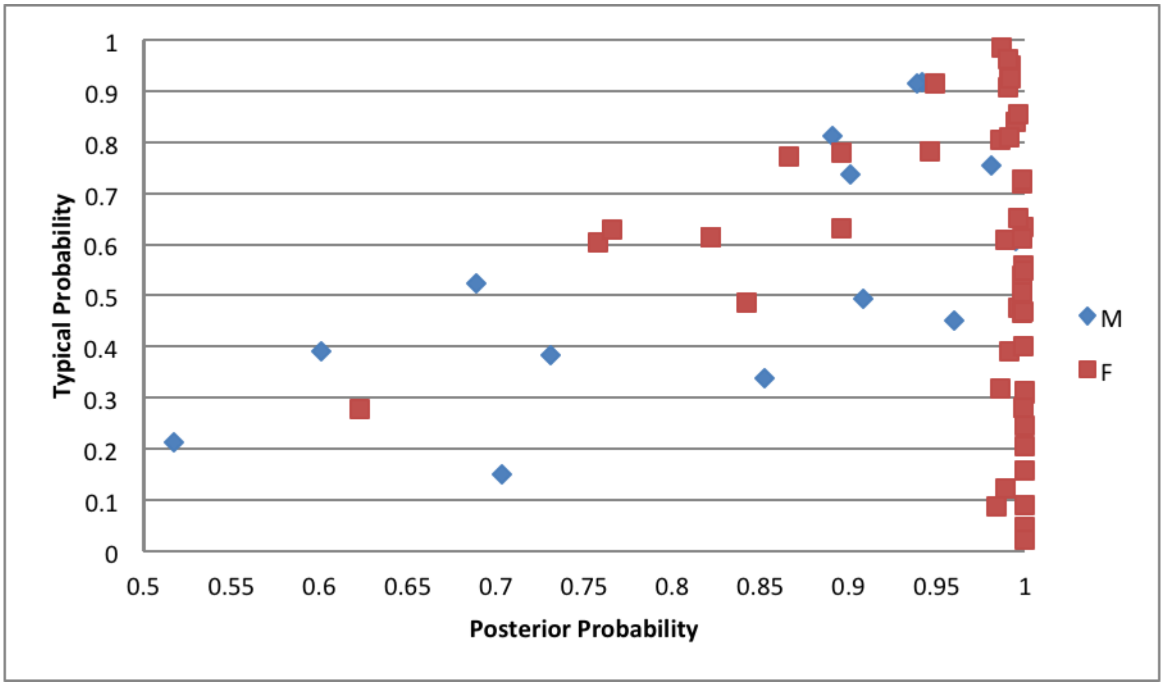

Figure 3.2: All Athens individuals who were classified as "female" by FORDISC 3.0. Males (M) are in blue diamonds and females (F) are in red squares

3.3.5 Descriptive statistics of the discriminate function analyses of the Tennessee population group

Figures 3.3 and 3.4 show the scatter plot results of the typical and posterior probability p-values of all the individuals in the Tennessee population group as classified by FORDISC 3.0. Figure 3.3 shows all of the individuals who were classified as male by FORDISC 3.0. Only two female individuals were classified as male in this population group. Figure 3.4 shows all of the individuals that were classified as female by FORDISC 3.0. Eight male individuals were classified as female. Figure 3.4 also illustrates that a large majority of females had a posterior probability greater than 0.95, which suggests a greater accuracy classification for estimation of sex from the scapula. This is not unexpected because the skeletal individuals used in the Tennessee sample are also incorporated as part of the skeletal sample used for the Forensic Data Bank.

Overall, almost all of the individuals in the Tennessee population group are being classified correctly as is illustrated by the high levels of posterior probability. However, for both males and females, the typical probability ranges from moderate to very high. This means that the individuals are being properly classified by FORDISC 3.0. These individuals are being placed into their correct male and female categories because the scapular measurements used to calculate sex from the scapula for the Tennessee population are similar to the scapular measurements employed by the FORDISC 3.0 program, which come from the Forensic Data Bank.

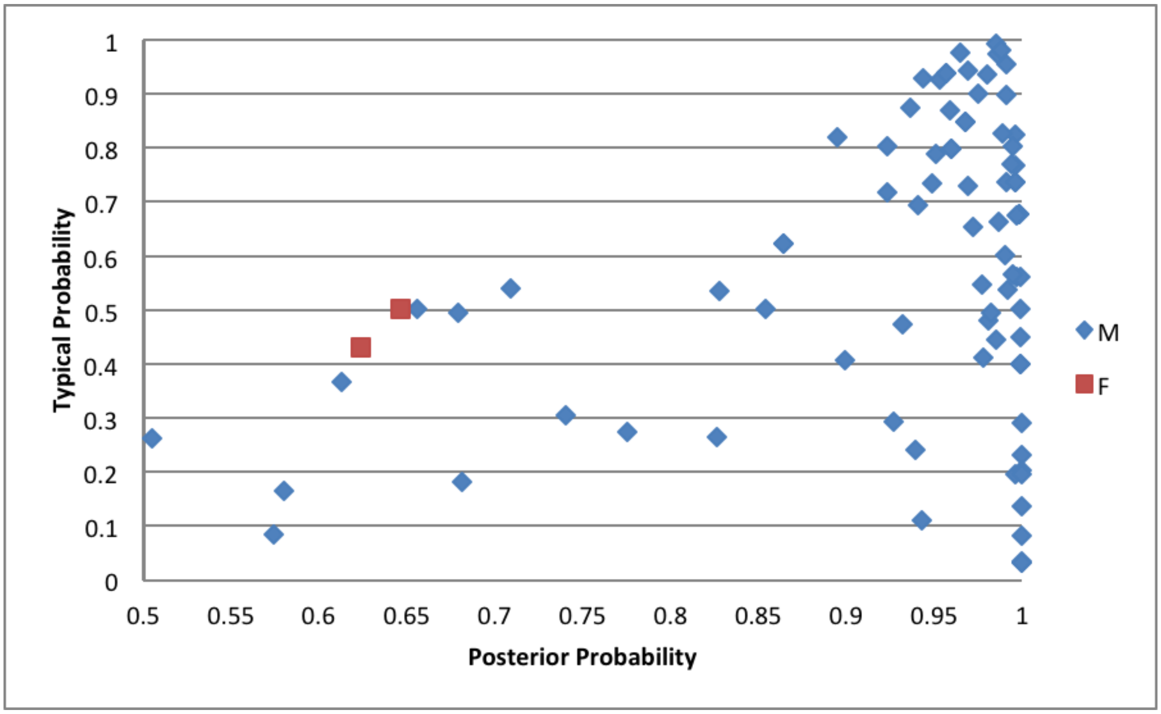

Figure 3.3: All Tennessee individuals who were classified as "male" by FORDISC 3.0. Males (M) are in blue diamonds and females (F) are in red squares

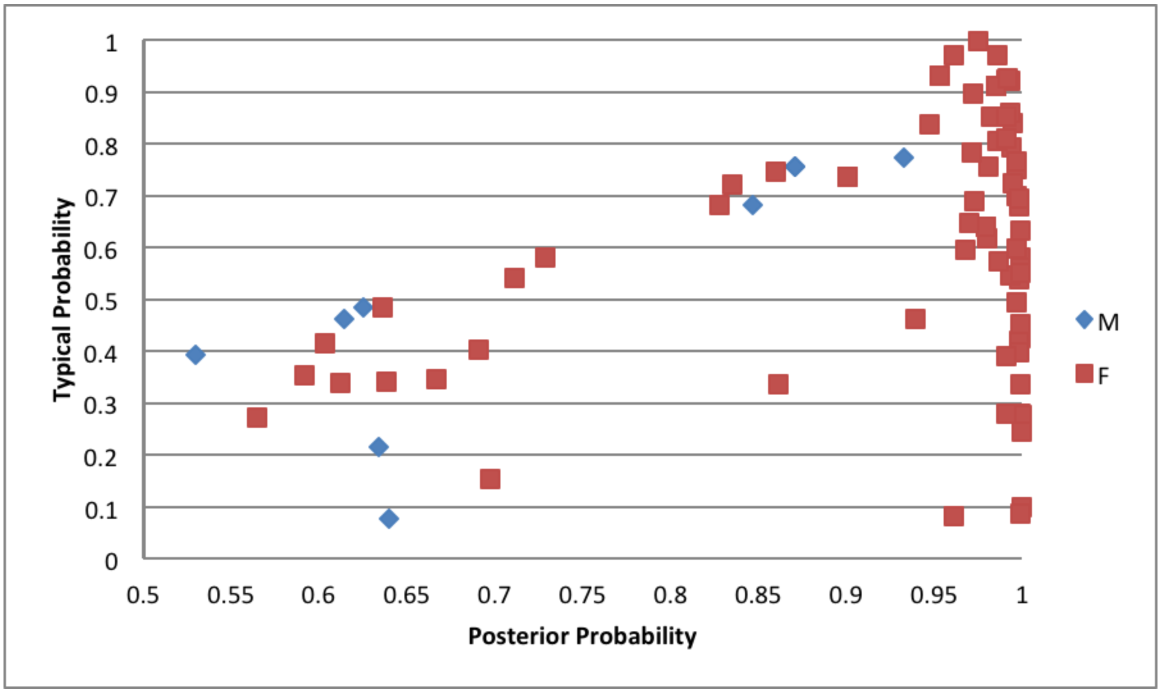

Figure 3.4: All Tennessee individuals who were classified as "female" by FORDISC 3.0. Males (M) are in blue diamonds and females (F) are in red squares

3.3.6 Low typical probability

The typical probability determines how similar or typical the data is to the database. If a typical probability falls below 0.05 then the accuracy of FORDISC 3.0 to correctly classify that individual is significantly low. This is because the value of the measurements of an individual greatly falls outside the range of the database to which it is being compared. Low typical probability was exhibited for several individuals from both population groups. In the Athens population group, two males and two females had typical probability values less than 0.05 (Figures 3.1 and 3.2). In the Tennessee population group, only one male individual had a typical probability less than 0.05 (Figure 3.3).

3.4 Intra- and inter- observer error for the five- and two-variable models and FORDISC 3.0 validation study

To ensure measurement accuracy and avoid measurement bias, intra- and inter-observer measurement error was examined. The purpose of these paired t-tests was to evaluate measurement reliability in estimating sex from the scapula between the five-variable model, two-variable model, and FORDISC 3.0 measurements. A methodology is considered unreliable if the measurements for that methodology are not repeatable. Intra-observer error is the differences between interpretations of one individual making two or more observations of the same phenomenon. To test this, the author repeated the five- and two-variable model methods on a sub-sample of 30 individuals from both the Athens and Tennessee population groups. Paired t-tests, which examined the paired statistical differences of the methods' measurements, were performed on the original data and the re-measured data. The results of those paired t-tests are presented in Table 3.18. The resulting p-values show no significant differences between the measurements obtained for the five- and two-variable models and the FORDISC 3.0 methodology (Table 3.18).

Inter-observer error is the differences between interpretations of two or more individuals making observations of the same phenomenon. In this study, the author had an assistant, with forensic anthropology experience, re-measure the same subsample of 30 scapulae from both the Athens and Tennessee populations. The measurements were then analyzed for variation using paired t-tests. Table 3.18 shows the statistical paired t-tests for the inter-observer error for the five- and two-variable models and the FORDISC 3.0 validation study. The resulting p-values illustrate that three measurements (maximum length of spine (XBS), lateral curvature (CSV), and thickness of lateral border (TLB)) for the Athens collection and one measurement (lateral curvature (CSV)) for the Tennessee sample had p-values lower than p=0.05. This shows that the values of those measurements are statistically different than the original data set. This suggests that the reliability of obtaining those measurements is not consistent.

Table 3.18: Intra- and Inter-observer error bias for the five- and two-variable models and the FORDISC 3.0 validation study.

Sample	Measurement	T-Value	Intra-observer Data (P-Value)*	T-Value	Inter-observer Data (P-Value)*
Athens	XLS	-0.17	0.870	-1.81	0.080
	XHS	1.58	0.125	-0.81	0.426
	XBS	1.82	0.078	2.49	0.019*
	HAX	1.88	0.070	-0.99	0.332
	CSV	1.68	0.103	3.29	0.003*
	TLB	1.22	0.231	3.37	0.002*
Tennessee	XLS	0.00	0.999	0.24	0.814
	XHS	-0.01	0.989	-0.06	0.953
	XBS	0.03	0.973	0.79	0.438
	HAX	0.44	0.665	0.47	0.641
	CSV	-0.08	0.935	2.09	0.046*
	TLB	-0.12	0.906	0.17	0.869

*Significance was established at $p<0.05$

3.5 Pearson's correlation test for age at death of the Athens and Tennessee population groups for the five- and two-variable models and FORDISC 3.0.

A Pearson's correlation test was performed on all the measurements used in the five- and two-variable models and the FORDISC 3.0 validation study. This was to establish whether or not there was a correlation between age at death and measurement size within the Athens and Tennessee population groups. Table 3.19 shows the resulting p-values of the Pearson's correlation test performed on both the Athens and Tennessee population samples. In table 3.19, the Athens population sample had only three p-values less than p=0.05, which were height of the glenoid prominence (HAX) and lateral curvature (CSV) in males and height of glenoid prominence (HAX) in females. All other p-values were greater than p=0.05, which indicates no correlation with age. For all but three variables in the Athens population sample, the age of the individual at death does not have an effect on the size of the variable. For example, if an individual is extremely young (early 20 years) or extremely old (65+ years) the size of that measurement is directly affected by the age of that individual. The Tennessee population group had all but one p-value greater than p=0.05, which was maximum breadth of the scapula (XBS) in males. For all but one variable (maximum breadth of the scapula (XBS)), in the Tennessee population sample, the age of the individual at death does not have an effect on the size of the variable, i.e. the size of the measurement is unaffected by the age of the individual. The overall results of the Pearson's correlation tests suggest that the age of the individual does not play a significant role in the accuracy of the five- and two-variable models and FORDISC 3.0.

Table 3.19: Pearson's correlation test for age at death of the Athens and Tennessee Population groups for the five- and two-variable models and FORDISC 3.0

Population	**Measurement**	**Male**		**Female**	
		R-values	**P-values***	**R-values**	**P-values***
Athens	XLS	-0.020	0.863	0.093	0.533
	XHS	-0.005	0.966	-0.017	0.908
	XBS	0.067	0.564	0.057	0.701
	HAX	0.248	0.029*	0.300	0.040*
	CSV	0.270	0.017*	0.264	0.072
	TLB	-0.065	0.572	-0.143	0.336
Tennessee	XLS	0.151	0.148	0.080	0.494
	XHS	0.051	0.629	-0.057	0.623
	XBS	0.213	0.039*	0.137	0.238
	HAX	0.074	0.479	0.126	0.277
	CSV	0.058	0.577	0.176	0.129
	TLB	-0.027	0.796	-0.079	0.495

*Significance was established p<0.05

3.6 Individuals used for the creation of the models for "Macaluso's Hypothesis" for estimating sex from the glenoid cavity.

In this study, 335 individuals from two skeletal reference collections, the Athens Collection and the Tennessee Collection, were measured. Only left scapulae of White European individuals were measured as per standards published by other researchers (Buikstra and Ubelaker, 1994). This study examined a contemporary skeletal collection, i.e. all individuals utilized were born around or after the year 1940. Therefore, the research examined how geographical distance between similar population groups, living at the same temporal period, affected estimating sex from the scapula. Table 3.20 shows the descriptive statistics for the individuals used for this study.

Table 3.20: Sex and age of individuals for the creation of the models for "Macaluso's Hypothesis"

Sex	Sample	N	Mean (Years)	SD (Years)	Median (Years)	Minimum (Years)	Maximum (Years)
Male	Athens	95	54.1	19.3	55.5	20	94
	Tennessee	92	47.4	9.1	49	19	62
Female	Athens	74	58.2	18.8	60	20	87
	Tennessee	74	52.5	8.1	53	31	67

3.7 Descriptive statistics for the creation of the models for "Macaluso's Hypothesis"

Macaluso (2010) estimated sex from the scapula by collecting measurements from the height, breadth, area and perimeter of the glenoid cavity using the program "ImageJ".

Although “ImageJ” is a free software program offered by the National Institutes of Health, the program requires training on how to properly take measurements using this software. Macaluso (2010) suggested that the area of the glenoid cavity could be obtained without “ImageJ”. If the height and breadth of the glenoid cavity were taken with sliding calipers then the calculated area of the glenoid cavity could be obtained by multiplying the height and the breadth. In his study, Macaluso (2010) found that the area of the glenoid cavity obtained by using “ImageJ” and the calculated area obtained by using sliding calipers had the same predictive qualities for estimating the sex of an individual. To test this hypothesis, the current study examined the sexually dimorphic nature of the measured height, breadth, and calculated area of the glenoid cavity. From that the current study established new methods for estimating sex from the scapula in two White European population groups. These new methods for estimating sex will be called the models for “Macaluso’s Hypothesis” for the remainder of the thesis.

Tables 3.21 and 3.22 show the descriptive statistics used for the glenoid cavity study. Tables 3.21 and 3.22 show the mean, standard deviation, minimum measurement length, and maximum measurement length within the Athens and Tennessee population groups.

Table 3.21: Athens descriptive statistics for the creation of the models for "Macaluso's Hypothesis"

Sex	Measurement	Mean (mm)	Standard Deviation (SD)	Minimum (mm)	Maximum (mm)
Males	Height	38.10	2.90	28.20	43.25
	Breadth	28.86	2.57	21.3	36.45
	Area	1104.58	165.51	649.65	1567.35
Females	Height	33.50	1.85	29.95	39.90
	Breadth	24.40	2.01	20.25	30.40
	Area	819.841	104.55	613.97	1127.84

Table 3.22: Tennessee descriptive statistics for the creation of the models for "Macaluso's Hypothesis"

Sex	Measurement	Mean (mm)	Standard Deviation (SD)	Minimum (mm)	Maximum (mm)
Males	Height	38.99	2.14	34.10	44.35
	Breadth	30.22	1.93	23.20	34.80
	Area	1181.08	125.04	818.96	1527.86
Females	Height	33.74	1.85	29.10	37.65
	Breadth	25.53	1.83	22.40	31.30
	Area	863.37	94.21	669.18	1159.66

3.8 Results of the logistic regression analyses for the creation of the models for "Macaluso's Hypothesis"

3.8.1 Sexual dimorphic variation between males and females of the Athens and Tennessee population groups

In order to create a new methodology for estimating sex, each measurement (i.e. height, breadth and area) in the study needs to be sexually dimorphic. Two-sample t-tests and Mann-Whitney U tests were performed on each measurement between males and females. Tables 3.23 and 3.24 show the individual tests performed on the measurements between males and females. The difference between the two tests is that the t-test is used when data are normally distributed over the density curve (parametric) and the Mann-Whitney U test is used when the data are not normally distributed (non-parametric). Each measurement, for both sexes in the two population groups, was evaluated for normality using MiniTab 16. If the data were normally distributed then a t-test was performed and if the data were not normally distributed then a Mann-Whitney U test was used. These tests were performed to show that each measurement, within each population group, was sexually dimorphic. All resulting p-values were less than p=0.001. This illustrates that each measurement has the potential to have good predictive value for classifying males and females in both population groups using the glenoid cavity measurements.

Table 3.23: Tests performed for validating sexual dimorphism between males and females of the Athens population

Measurement	**Test**	**U-Value**	**P-Values***
Height	Mann-Whitney	10882.00	0.000*
Breadth	Mann-Whitney	11001.50	0.000*
Area	Mann-Whitney	11051.00	0.000*

*Significance was established as p< 0.001

Table 3.24: Tests performed for validating sexual dimorphism between males and females of the Tennessee population

Measurement	**Test**	**T-Value**	**P-Values***
Height	T-Test	16.90	0.000*
Breadth	T-Test	16.01	0.000*
Area	T-Test	18.66	0.000*

*Significance was established as p< 0.001

3.8.2 Statistical variation between same sex measurements: Athens and Tennessee

As the Athens and Tennessee populations are representative of the same ancestry, i.e. White European, combining the two groups to create one logistic regression equation for a broader, more global, methodology for White Europeans was attempted. To determine if the two population samples could be combined, two-sample t-tests and Mann-Whitney U tests were performed on each measurement between males and females. Tables 3.25 and 3.26 show the resulting p-values of the statistical variation tests within the Athens and Tennessee population samples.

All three measurements in Table 3.26, the height, the breadth, and the calculated area of the glenoid cavity, have resulting p-values less than p=0.05. This illustrates that those measurements for estimating sex from the scapula are statistically different between the Athens and Tennessee population groups.

One measurement in Table 3.26, the height of the glenoid cavity, had a resulting p-value greater than p=0.05. This shows that the measurement is statistically similar within those population groups. However, all other measurements had resulting p-values lower than p=0.05, showing significant difference within these two population groups. Therefore in addition to combining the two population groups to create a larger sample size, two separate logistic regression models were created for each population group (Athens and Tennessee) and the predicted accuracy rates of those models were examined collectively.

Table 3.25: Tests performed for measurement differences between the male individuals of the Athens and Tennessee populations

Measurement	Test	T-Value	U-Value	P-Values*
Height	T-Test	-2.40	N/A	0.017*
Breadth	Mann-Whitney	N/A	7257.50	0.000*
Area	T-Test	-3.57	N/A	0.000*

*Significance was established as p< 0.05

Table 3.26: Tests performed for measurement differences between the female individuals of the Athens and Tennessee populations

Measurement	Test	T-Value	U-Value	P-Values*
Height	Mann-Whitney	N/A	5116.50	0.129
Breadth	T-Test	-3.59	N/A	0.000*
Area	Mann-Whitney	N/A	4703.00	0.002*

*Significance was established as $p < 0.05$

3.8.3 Binary logistic regression analysis on the Athens population.

Logistic regression analysis was used to create the new equations for estimating sex from the glenoid cavity for the Athens and Tennessee population groups. Logistic regression is the application of multiple independent variables presented simultaneously to predict membership into one or the other of two dependent variable categories (Burns and Burns 2008: 569). The glenoid cavity measurements (height, breadth, and calculated area) are used as the independent variables to help predict membership into either males or females. This is done by the creation of a coefficient (β) for each independent variable. This coefficient measures the independent variables' (height, breadth, and calculated area of glenoid fossa) contribution to the variations in the dependent variables' (males or females) outcome. A backward stepwise procedure is conducted to help discriminate those independent variables that have the most statistical strength for creating a model that can allocate group membership (males and females) based on the most sexually dimorphic measurements. This is done through the subtraction of independent variables (i.e. one at a time) to the backward stepwise model in order to create the best model for differentiating between males and females. The goal is to use the most sexually dimorphic measurements to create the best models for estimating sex.

Table 3.27 shows the resulting coefficient β of the direct and backward stepwise analyses performed to create the equations for estimating sex from the height, breadth, and calculated area of the glenoid cavity. The “direct” analysis incorporates all three variables within that equation along with a constant. The other two equations, Step 2 and Step 3, are the results of a backward stepwise function that calculates new equations by removing one or more variables that could potentially increase the accuracy of the model. These binary logistic regression equations follow the formula: $y = (\beta_0) + (\beta_1)(X_1) + (\beta_2)(X_2) + (\beta_3)(X_3)$, where y = predicted sex, β_0 is the constant, and $(\beta_x)(X_x)$ is the coefficient β created by the model multiplied by the measured variable. If y is greater than 0.50 then the individual is predicated to be male. If y is less than 0.50 then the individual is predicated to be female. The equations for the Athens population group are:

Direct:
y(predicted sex) = (Constant [6.322]) – (Coefficient β for height [0.651]) * (measured height) - (Coefficient β for breadth [0.598]) * (measured breadth) + (Coefficient β for calculated area [0.035]) * (calculated area)

Step 2:
y(predicted sex) = - (Constant [9.805]) – (Coefficient β for height [0.183]) * (measured height) + (Coefficient β for calculated area [0.018]) * (calculated area)

Step 3:
y(predicted sex) = - (Constant [13.195]) + (Coefficient β for calculated area [0.014]) * (calculated area)

Table 3.27: Binary logistic regression analysis on the Athens population

		Coefficient β	**Standard Error**	**Sectioning Point**
Direct	**Height**	-0.651	1.490	0.50
	Breadth	-0.598	1.882	0.50
	Area	0.035	0.055	0.50
	Constant	6.322	50.841	0.50
Step 2*	**Height**	-0.183	0.208	0.50
	Area	0.018	0.004	0.50
	Constant	-9.805	4.208	0.50
Step 3**	**Area**	0.014	0.002	0.50
	Constant	-13.195	1.904	0.50

*Variable not selected for backward stepwise analysis: breadth
**Variable not selected for backward stepwise analysis: height and breadth

3.8.4 Hosmer and Lemeshow test for goodness-of-fit on Athens predictive models

Table 3.28 shows the resulting p-values of the Hosmer and Lemeshow goodness-of-fit test for the predictive models outlined in Table 3.27. The Hosmer and Lemeshow analysis tests for observed and expected outcomes of the logistic regression model. The p-values of the test for the Athens population group are all greater than $p=0.05$. This indicates that for all the models (Direct, Step 2, and Step 3) the numbers of observed individuals, both male and female, are not significantly different than those predicted by the model. Thus, all three models have good predictive value.

Table 3.28: Hosmer and Lemeshow test for goodness-of-fit on Athens predictive models

	Chi-Squared	**Degrees of Freedom**	**P-value***
Direct	8.801	8	0.359
Step 2	8.693	8	0.369
Step 3	5.373	8	0.717

*Significance was established at $p<0.05$

3.8.5 Classification plots on Athens predictive models

Table 3.29 shows the resulting percentages, as well as, the sex bias percentages of the models outlined in Table 3.27. Table 3.29 also shows the number of individuals correctly and incorrectly classified by the models outlined in Table 3.27. Table 3.29 shows that both the Direct and Step 3 model had an accuracy percentage of 87.6%, while the Step 2 model had the highest predicted accuracy percentage of 88.8%. Although, Step 2 had the highest percentage in predicted accuracy it also had the highest sex bias percentage, which indicates that this model may misclassify more females over males.

Table 3.29: Classification plots on the Athens predictive models

Observed			**Predicted**			**Sex Bias (%)***
			Code		**Percentage Correct (%)**	
			0.00 (Females) (N)	**1.00 (Males) (N)**		
Direct	**Code**	**0.00**	65	9	87.8	-0.4
		1.00	12	83	87.4	
Overall Percentage					**87.6**	
Step 2	**Code**	**0.00**	64	10	86.5	4.0
		1.00	9	86	90.5	
Overall Percentage					**88.8**	
Step 3	**Code**	**0.00**	64	10	86.5	1.9
		1.00	11	84	88.4	
Overall Percentage					**87.6**	

*Sex bias% = % males correctly classified - % females correctly classified

3.8.6 Direct vs. Stepwise equations of the Athens population group

The result of the logistic regression analysis has produced three separate equations. Although each equation can be used to estimate sex, based on the Hosmer and

Lemeshow test, the only one that should be used is the equation with the least amount of sex bias. This is to avoid an equation that may misclassify the sexes or more accurately classify one sex over the other. Although a Stepwise equation may have a higher predicted percentage based on the classification plots, the Direct method should be used because it has the lowest sex bias ratio with a similar amount of accuracy.

3.8.7 Binary logistic regression analysis on the Tennessee population.

Table 3.30 shows the resulting equations of the Direct and backward Stepwise analyses performed to create the equations for estimating sex from the height, breadth, and calculated area of the glenoid cavity for the Tennessee population. The Direct analysis incorporates all three variables within that equation along with a constant. The other equation, Step 2, is the result of a backward stepwise function that calculates new equations by removing one or more variables that could potentially increase the accuracy of the model. These binary logistic regression equations follow the formula: $y = (\beta_0) + (\beta_1)(X_1) + (\beta_2)(X_2) + (\beta_3)(X_3)$, where y = predicted sex, β_0 is the constant, and $(\beta_x)(X_x)$ is the coefficient β created by the model multiplied by the measured variable. If y is greater than 0.50 then the individual is predicated to be male. If y is less than 0.50 then the individual is predicated to be female. The equations for the Tennessee population group are:

Direct:
y(predicted sex) = (Constant [52.774]) – (Coefficient β for height [1.982]) * (measured height) - (Coefficient β for breadth [3.568]) * (measured breadth) + (Coefficient β for calculated area [0.118]) * (calculated area)

Step 2:
y(predicted sex) = - (Constant [18.627]) – (Coefficient β for breadth [0.969]) * (measured breadth) + (Coefficient β for calculated area [0.045]) * (calculated area)

Table 3.30: Binary logistic regression analysis on the Tennessee population

		Coefficient β	**Standard Error**	**Sectioning Point**
Direct	**Height**	-1.982	4.735	0.50
	Breadth	-3.568	6.251	0.50
	Area	0.118	0.174	0.50
	Constant	52.774	170.199	0.50
Step 2*	**Breadth**	-0.969	0.506	0.50
	Area	0.045	0.011	0.50
	Constant	-18.627	6.733	0.50

*Variable not selected for backward stepwise analysis: height

3.8.8 Hosmer and Lemeshow test for goodness-of-fit on Tennessee predictive models

Table 3.31 shows the resulting p-values of the Hosmer and Lemeshow goodness-of-fit test for the predictive models outlined in Table 3.30. The Hosmer and Lemeshow analysis tests for observed and expected outcomes of the logistic regression model. The p-values of the test for the Tennessee population group are all greater than p=0.05. This indicates that for all the models (Direct and Step 2) the numbers of observed individuals, both male and female, are not significantly different than those predicted by the model. Thus, both models have good predictive value.

Table 3.31: Hosmer and Lemeshow test for goodness-of-fit on Tennessee predictive models

	Chi-Squared	**Degrees of Freedom**	**P-value***
Direct	6.143	8	0.631
Step 2	5.022	8	0.755

*Significance was established at $p<0.05$

3.8.9 Classification plots on the Tennessee predictive models

Table 3.32 shows the resulting percentages, as well as, the sex bias percentages of the models outlined in Table 3.30. Table 3.32 also shows the number of individuals correctly and incorrectly classified by the models outlined in table 3.30. Table 3.32 shows that the Direct and Step 2 models have an accuracy percentage of 94.6% and 94.0%, respectively. However, Step 2 has the lowest percentage in accuracy and highest sex bias percentage, which indicates that this model will misclassify females over males. The Direct model has a sex bias percentage of 0.0%, which indicates that it will not misclassify one sex more often than the other.

Table 3.32: Classification plots on the Tennessee predictive models

Observed			Predicted			Sex Bias (%)*
			Code		Percentage Correct (%)	
			0.00 (Females)	1.00 (Males)		
Direct	**Code**	**0.00**	70	4	94.6	0.0
		1.00	5	87	94.6	
Overall Percentage					**94.6**	
Step 2	**Code**	**0.00**	69	5	93.2	1.4
		1.00	5	87	94.6	
Overall Percentage					**94.0**	

*Sex bias% = % males correctly classified - % females correctly classified

3.8.10 Direct vs. Stepwise equations of the Tennessee population group

The result of the logistic regression analysis has produced two separate equations. Although each equation can be used to estimate sex, based on the Hosmer and Lemeshow test, the only one that should be used is the equation with the least amount of sex bias. This is to avoid any misclassification of sexes. The Stepwise equation does not have a higher predicted percentage of accurately classifying sex in the Tennessee population based on the classification plots. Therefore, the Direct method should be used because it has the lowest sex bias ratio and the highest percentage of classification accuracy.

3.8.11 Binary logistic regression analysis on the Combined Population groups (Athens and Tennessee).

Table 3.33 shows the resulting equations of the Direct and backward Stepwise analyses performed to create the equations for estimating sex from the height, breadth, and calculated area of the glenoid cavity from the pooled samples of the Athens and Tennessee populations. The reason this was done was to examine if predicted accuracy rates would be affected by combining the two populations, since both represent European population groups. Even though there are statistical differences in the size of the glenoid cavity, in males and females, between the Athens and Tennessee populations the author wanted to combine the populations to create a larger sample size and test if the accuracy rates increased or decreased. The Direct analysis incorporates all three variables within that equation along with a constant. The Direct analysis incorporates all three variables within that equation along with a constant. The other equations, Step 2 and Step 3, are the result of a backward stepwise function that calculates new equations by removing one or more variables that could potentially increase the accuracy of the model. These binary logistic regression equations follow the formula: $y = (\beta_0) + (\beta_1)(X_1) + (\beta_2)(X_2) + (\beta_3)(X_3)$, where y = predicted sex, β_0 is the constant, and $(\beta_x)(X_x)$ is the coefficient β created by the model multiplied by the measured variable. If y is greater than 0.50 then the individual is predicated to be male. If y is less than 0.50 then the individual is predicated to be female. The equations for the Combined Population groups are:

Direct:
y(predicted sex) = (Constant [54.323]) – (Coefficient β for height [1.954]) * (measured height) - (Coefficient β for breadth [2.725]) * (measured breadth) + (Coefficient β for calculated area [0.092]) * (calculated area)

Step 2:
y(predicted sex) = - (Constant [13.767]) – (Coefficient β for breadth [0.211]) * (measured breadth) + (Coefficient β for calculated area [0.020]) * (calculated area)

Step 3:
y(predicted sex) = - (Constant [16.265]) + (Coefficient β for calculated area [0.017]) * (calculated area)

Table 3.33: Binary logistic regression analysis on the Combined Population groups

		Coefficient β	**Standard Error**	**Sectioning Point**
Direct	**Height**	-1.954	1.251	0.50
	Breadth	-2.725	1.625	0.50
	Area	0.092	0.047	0.50
	Constant	54.323	43.295	0.50
Step 2*	**Breadth**	-0.211	0.225	0.50
	Area	0.020	0.004	0.50
	Constant	-13.767	3.096	0.50
Step 3**	**Area**	0.017	0.002	0.50
	Constant	-16.265	1.725	0.50

*Variable not selected for backward stepwise analysis: height
**Variables not selected for backward stepwise analysis: breadth and height

3.8.12 Hosmer and Lemeshow test for goodness-of-fit on Combined Population groups (Athens and Tennessee).

Table 3.34 shows the resulting p-values of the Hosmer and Lemeshow goodness-of-fit test for the predictive models outlined in Table 3.33. The Hosmer and Lemeshow analysis tests for observed and expected outcomes of the logistic regression model. The p-values of the test for the Combined Population groups are all greater than p=0.05. This indicates that for all the models (Direct, Step 2, and Step 3) the numbers of observed individuals, both male and female, are not significantly different than those predicted by the model. Thus, both models have good predictive value.

Table 3.34: Hosmer and Lemeshow test for goodness-of-fit on the Combined Population groups predictive models

	Chi-Squared	Degrees of Freedom	P-value*
Direct	6.143	8	0.726
Step 2	5.022	8	0.149
Step 3	7.219	8	0.513

*Significance was established at p<0.05

3.8.13 Classification plots on the Combined Population groups predictive models

Table 3.35 shows the resulting percentages, as well as, the sex bias percentages of the models outlined in Table 3.33. Table 3.35 also shows the number of individuals correctly and incorrectly classified by the models outlined in Table 3.33. Table 3.35 shows that both the Direct and Step 2 models had an accuracy percentage of 89.0%, while the Step 3 model had the highest predicted accuracy percentage of 88.7%. The Direct and Step 2 had the highest percentage in predicted accuracy. The Direct model had the lowest sex bias percentage, which indicates that this model does not misclassify more females over males then Step 2 or Step 3.

Table 3.35: Classification plots on the Combined Population groups predictive models

Observed			**Predicted**			**Sex Bias (%)***
			Code		**Percentage Correct (%)**	
			0.00 (Females) (N)	**1.00 (Males) (N)**		
Direct	**Code**	**0.00**	132	16	89.2	-0.4
		1.00	21	166	88.8	
Overall Percentage					**89.0**	
Step 2	**Code**	**0.00**	128	20	86.5	4.4
		1.00	17	170	90.9	
Overall Percentage					**89.0**	
Step 3	**Code**	**0.00**	128	20	86.5	3.9
		1.00	18	169	90.4	
Overall Percentage					**88.7**	

*Sex bias% = % males correctly classified - % females correctly classified

3.8.14 Direct vs. Stepwise equations of the Combined Population groups

The result of the logistic regression analysis has produced three separate equations. Although each equation can be used to estimate sex, based on the Hosmer and Lemeshow test, the only one that should be used is the equation with the least amount of sex bias. This is to avoid any misclassification of sexes. The Stepwise equations have similar predicted percentage of accurately classifying sex in the Combined Population groups based on the classification plots. However, the Direct method should be used because it has the lowest sex bias ratio and the highest percentage of classification accuracy.

3.8.15 Chi-squared proportion tests of the classification accuracies of the models for "Macaluso's Hypothesis" (Athens, Tennessee and Combined Population)

A chi-squared proportions test was used to examine if there was a significant difference between the accuracies of the Direct Methods for the Athens, Tennessee, and Combined Population models for "Macaluso's Hypothesis". The chi-squared distribution tests the difference among two or more proportions (Moore 2010). Table shows the results of the chi-squared tests analyzing the proportions of accuracy, i.e. those individuals who were correctly classified and those individuals who were not, between the Athens, Tennessee, and combined population samples. These results indicate which model is more accurate for estimating sex in each population group.

Table 3.36: Chi-squared proportion tests on the classification accuracies of the models for "Macaluso's Hypothesis" (Athens, Tennessee and Combined Population)

Models	**Sex**	***Chi-squared value***	***P-value***
Athens- Tennessee	Male	2.929	0.087
Athens-Tennessee	Female	2.108	0.147
Athens-combined population	Male	0.120	0.729
Athens- combined population	Female	0.090	0.764
Tennessee- combined population	Male	2.451	0.117
Tennessee- combined population	Female	1.758	0.185

*Significance was established as $p< 0.05$

All of the resulting p-values were greater than $p=0.05$. This shows that all proportions that were analyzed were statistically similar. These results indicate that the population groups measured, whether it was the Athens, Tennessee, or combined population sample have accuracy proportions that are similar for each of the "Bell Models" that were created in this current study. This is important because these results show that these models can be used with similar degrees of accuracy for estimating sex.

3.9 Pearson's correlation test for age at death in the creation of the models for "Macaluso's Hypothesis"

A Pearson's correlation test was performed on all glenoid cavity measurements for the new methodology for estimating sex from the scapula. This was to establish whether or not there was a correlation between age at death and glenoid cavity size within the Athens and Tennessee population groups. Table 3.37 shows the resulting p-values of the Pearson's correlation test performed on both the Athens and Tennessee population samples. The purpose was to test if age affects glenoid cavity height, breadth, and calculated area for both population groups. In Table 3.37, the Athens population sample had only one p-value greater than p=0.05, which was the height of the glenoid cavity for males. All other p-values were less than p=0.05, which indicates a correlation with age. For all but one variable in the Athens population, the age at death of the individual has an effect on the size of the variable. For example, if an individual is extremely young (early 20 years) or extremely old (65+ years) the size of that measurement is directly affected by the age of that individual. The Tennessee population group had all but one p-value greater than p=0.05, which was the breadth of the glenoid cavity for males. This indicates that, for both males and females, all other variables do not show a correlation with age. For all but one variable, in the Tennessee population sample, the age at death of the individual does not have an effect on the size of the variable. This means that the size of the measurement is unaffected by the age of the individual. The overall results of the Pearson's correlation tests suggest that the age of the individual could play a role in the accuracy of the glenoid cavity models, especially for the Athens population group.

Table 3.37: Pearson's correlation test for age at death of the Athens and Tennessee Population groups in the creation of the models for "Macaluso's Hypothesis"

Population	Sex	R-values			P-values*		
		Height	Breadth	Area	Height	Breadth	Area
Athens	Males	0.068	0.309	0.220	0.514	0.002*	0.033*
	Females	0.342	0.386	0.400	0.004*	0.001*	0.001*
Tennessee	Males	0.019	0.275	0.177	0.861	0.008*	0.091
	Females	-0.019	0.212	0.125	0.870	0.069	0.287

*Significance was established $p<0.05$

3.10 Intra- and inter-observer error in the creation of the models for "Macaluso's Hypothesis"

To ensure measurement repeatability and avoid measurement bias, intra- and inter- observer measurement error was examined. The purpose of these paired t-tests was to evaluate measurement reliability between the Athens and Tennessee population groups. To test this, the author re-measured the two variables (height and breadth) of the glenoid cavity on a sub-sample of 30 individuals from both the Athens and Tennessee population groups. Paired t-tests were performed on the original data and the re-measured data. Table 3.38 shows the resulting p-values of the statistical paired t-tests for intra-observer error. None of the p-values were less than $p=0.05$, which indicates that there were no statistical differences between the measurements.

Inter- observer error was also examined in the current study on the new methodology for estimating sex from the scapula. The author had an assistant, with forensic anthropology experience, re-measure the same sub-sample of 30 scapulae. Then the measurements were analyzed for variation using paired t-tests. Table 3.38 shows the resulting p-values of the statistical paired t-tests for intra-observer error. None of the p-values were less than p=0.05, which indicates that there were no statistical differences between the measurements.

Table 3.38: Inter- and Intra- observer error bias for the creation of the models for "Macaluso's Hypothesis"

Sample	Sub-sample Initial Data	Inter-observer Data		Intra-observer Data	
		T-Value	P-Value*	T-Value	P-Value*
Athens	Height	-1.61	0.119	-1.78	0.086
	Breadth	-1.67	0.105	-1.59	0.123
Tennessee	Height	0.90	0.373	-1.54	0.133
	Breadth	-0.61	0.547	1.01	0.323

*Significance was established at p<0.05

CHAPTER 4: DISCUSSION

4.1 Context of the current project

The usefulness and utility of scientific methodologies can be classified as: accurate and reliable, accurate and unreliable, inaccurate and reliable, and inaccurate and unreliable (Blanchard 2006). These classifications are important in forensic science because they increase our understanding of the usefulness of the methodology within a medico-legal context. Having methodologies that are inaccurate or unreliable calls into question their usability within a court of law. Many of the bones of the human skeleton have been used to estimate the sex of unknown human remains. These types of methodologies need to be tested and retested for accuracy and reliability, especially those methods that are shown to be population specific.

The objectives of this thesis are as follows: (1) to understand the relationship between biological sex and estimated sex from the scapula based on metric analyses on two White European population groups from North America and Europe; (2) to test the accuracy and reliability of two scapular methodologies published by Dabbs and Moore-Jensen (2010) and Jantz and Ousley (1993) and; (3) to provide an alternative method for estimating sex from the scapula, of White European populations, based on metric analyses of the glenoid cavity.

4.2 Dabbs and Moore-Jansen (2010) validation study on the five- and two-variable models and population diversity

The accuracy of any sex estimation model is determined by its potential to correctly classify an individual as being either male or female. This study examined whether the five- and two-variable models were accurate when applied to a different population group from which they were created. It also examined which model, the five- or two-variable, more accurately classified group membership when it was applied to this new population sample. The results indicate that the five-variable model correctly classified both males and females of the Athens and Tennessee population group more often than the two-variable model (Table 3.6). The reason for the classification differences between the five- and two-variable models, in the Athens and Tennessee population groups, was because of the number of variables utilized in the models themselves. Bronowski and Long (1952) showed that discriminate functions equations with multivariate analysis allow for a high level of distinction between two groups. The more variables in the equation allow for a sharper distinction of group membership. However, the researchers do warn that the accuracy of the model will not increase if too many variables are added into the model. Therefore, the lower accuracy rates of the two-variable model, in both population groups, is due in part because there were not enough measurements incorporated into that model to classify group membership as either male or female. The higher accuracy rates of the five-variable model can be attributed to the use of five sexually dimorphic measurements identified by Dabbs and Moore-Jansen (2010). These results are reinforced by Dabbs and Moore-Jansen's (2010) original study which showed that the two-variable model had a lower accuracy rate than the five-variable model for estimating sex in their population group.

Dabbs and Moore-Jansen's (2010) original study reported a total combined (i.e. males and females) accuracy rate of 92.5% for the five-variable model (males = 89.8%, females = 96.8%) and 92.5% (males = 91.5%, females = 93.6%) for the two-variable model when tested on a smaller subsample (N=80) of individuals; this subsample is called the "Cleveland" sample. The overall classification accuracy of the five- and two-variable models when tested on a separate cadaveric sample (N=32) from Wichita State University was 84.4% (males = 88.9%, females = 78.6%) for the five-variable model and 81.3% (males = 88.9%, females = 71.4%) for the two-variable model (Dabbs and Moore-Jansen 2010). In the current study, when the five-variable model, for both the Athens and Tennessee groups, was compared to the "Cleveland" sample there were similar accuracy classifications rates, with the total (male and female) accuracy rates being 94.5% (Athens population) and 95.9% (Tennessee population). With the two-variable model both the Athens and Tennessee population groups showed lowered combined total accuracy classifications with those rates being 89.7% (Athens population) and 89.4% (Tennessee population). When the Athens and Tennessee results of the current study are compared to the Wichita State cadaver sample, the five- and two-variable models had a higher accuracy rates for sex classification. The accuracy rates of the five- and two-variable models in the current study are higher than those previously reported by Dabbs and Moore-Jansen (2010). This indicates that these models can be used on a much larger White European population group, specifically from North America and Europe.

Spradley and Jantz (2011) conducted a study on using the postcranial skeleton to estimate sex. The researchers used the height and breadth of the scapula and had an overall accuracy that was 93.04% in “American Whites.” The researchers concluded that multivariate analysis, as opposed to univariate analysis, of postcranial elements, including the scapula, produces a higher level of accuracy in the classification of males and females than the skull. In fact, the scapula was one of the most sexually dimorphic bones when combined with multivariate analyses in “American White” individuals (Spradley and Jantz 2011). This indicates that the more variables within a discriminate function model the more accurate the classification can be when estimating sex from the scapula. The current research shows that the five-variable model, although statistically similar in accuracy rates, is slightly more accurate than the two-variable model.

Dabbs (2010) and Papaioannou and colleagues (2012) examined sexual dimorphism with emphasis on population diversity. They examined population specific methodologies with high percentages of accuracy utilizing a different combination of scapular measurements than those employed by the five- and two-variable models. Dabbs (2010) used an ancient Egyptian population to develop models for estimating sex from the scapula. In the ancient Egyptian model, the five most sexually dimorphic measurements were: maximum length of the scapula (XHS), maximum length of the spine (XLS), breadth of the infraspinous body (BXB), height of the glenoid fossa (HAX), and breadth of the glenoid fossa (BCB). The accuracy rates of the Egyptian model for the Dabbs (2010) study were between 84% and 88%. Two of the measurements for the Egyptian model (breadth of the infraspinous body and breadth of the glenoid fossa) differ from the contemporary five- and two-variable models developed by Dabbs and Moore-Jansen (2010) because the independent t-tests in Dabbs’ (2010) ancient Egyptian study showed that those measurements for that Egyptian population group were the most sexually dimorphic. This indicates that discriminate function models use a variety of metric combinations that may work for one population group may not always produce the same metric combinations for another population group, whether contemporary or historic. However, they may still produce high accuracy rates for the same skeletal element.

Similarly, Papaioannou and colleagues (2012) used only the maximum length of the scapular spine, the maximum length of the glenoid cavity height and the glenoid cavity breadth and had a combined overall classification accuracy of 95.9% with a sex bias ratio of correctly classifying females over males by 1.9%. Papaioannou and colleagues' (2012) study showed that the best discriminate function model to estimate sex on a contemporary Creatian population involved two measurements: the maximum length of the spine and breadth of the glenoid cavity. The researchers concluded that sexually dimorphic differences occur at the individual level rather than as a representation of the general population. This does not mean that similar population groups are not different however, the degree of sexual dimorphism, particularly in the scapula, is driven by individual muscular development and activity (Hrdlička 1942). Even though there are occupational stressors and environmental influences that shape body development, finding combinations of metric variables from the scapula to produce the highest accuracy for estimating sex within similar population groups is possible. The current study shows that the Athens and Tennessee population groups have no statistical differences in accuracy rates when using either the five- and two-variable models by Dabbs and Moore-Jansen (2010) (Table 3.7). The current study showed that the five- and two-variable models achieve a high level of accuracy in both the Athens and Tennessee populations. This illustrates that discriminate function equations can be used within populations of similar ancestry if the right combination of measurements is employed.

Other research on another post-cranial element has shown high accuracy rates in similar population groups in different geographical areas. Steyn and Patriquin (2009) conducted a study on population-specific sex determination from the pelvis. The researchers utilized three population groups from Crete and South Africa. They discovered that population differences do not have an overall effect on the accuracy rates when estimating the sex from the pelvis. Steyn and Patriquin (2009) concluded that population specific formulae may not be necessary based on the distinct sexual dimorphism of the pelvis. However, the researchers state, "[t]he same is most probably not true for other, less dimorphic bones of the post-cranial skeleton." (Steyn and Patriquin 2009: 113.e3). However, the researchers do not outline what constitutes a "less dimorphic bone", but in the current study, the author has determined that the scapula is indeed a very sexually dimorphic post-cranial skeletal element. The high accuracy rate for estimating sex from the scapula using two White European population groups (North American Whites and Greeks) in the current study correlates with the overall conclusion of Steyn and Patriquin's (2009) study in that there is a need for the elimination for population-specific methods for determining sex, which can now be extended to the scapula. Also, Steyn and Patriquin (2009) suggested that the high accuracy rates between population groups were due to certain pelvic characteristics that were highly sexually dimorphic. This indicates that skeletal elements that are highly sexually dimorphic may have certain characteristics, whether metric or non-metric, that contribute only to the differentiation of biological sex rather than population diversity. The current study, the five- and two-variable models have very high accuracy rates that are statistically similar to each other (based on the chi-squared tests) between population groups. This indicates that these measurements reflect the differentiation between sex rather than the differentiation of populations, and therefore, have no impact on the accuracy rate of the sex estimation models (the five- and two variable).

The age of the individual is also a factor that contributes to the accuracy of the five- and two-variable models for estimating sex from the scapula. Dabbs and Moore-Jansen (2012) conducted a study on age changes of the scapula from the Hamman-Todd Collection at the Cleveland Museum of Natural History ("Cleveland" sample). Twenty-three measurements were obtained from each scapula and statistically analyzed for age related changes in individuals between the ages of 19 and 93 years of age. All measurements for the five- and two-variable models developed by Dabbs and Moore-Jansen (2010) for estimating sex were analyzed. With advanced age, the ventral curvature of the scapula, in "white males", increased while the overall scapular length decreased. In the current study, the Pearson correlation tests (Table 3.19) show that in the Athens population, age had an effect on the lateral curvature (CSV), which is the distance between the sub fossa and spinous axis. The lateral curvature measurement had a curvature that was greater in the Athens population than in the Tennessee population (Table 3.8). However, the males in the Tennessee population group had statistically smaller lateral curvatures than the males in the "Cleveland" sample. Dabbs and Moore-Jansen (2012) suggest that the greater curvature change, with advanced age, is due to occupational stress since, "the area of greatest gracility in the scapula is the supraspinous fossa" (Dabbs and Moore-Jansen 2012: 375). This area is prone to muscular stress which results in the scapula being subjected to the forces of gravity. Also scapular curvature may explain why the Athens males had a statistically smaller maximum length of the scapula (XHS) than the males in the "Cleveland" sample (Table 3.10). This type of advanced curvature also leads to a decrease in scapular length, which affects the measurements of the two-variable model. This factor may help explain why the two-variable model is more accurate for estimating sex in the Tennessee male population group than in the Athens male population group.

The female individuals, in both the Athens and Tennessee groups, did not exhibit increased curvature due to age as was present in the male scapulae. However, the Pearson's correlation tests in Table 3.19 did show that the height of the glenoid prominence (HAX) was affected by age in the Athens females, Athens males and Tennessee males. Dabbs and Moore-Jansen (2012) and Hrdlička (1942) noted an increase in ossification around the glenoid cavity with advanced age. Although, Dabbs and Moore-Jansen (2012) illustrated that this increase in ossification could relate to an overall increase in breadth of the scapula. The variation tests of the current study show no statistical differences between the breadth of scapula (XBS) in all three populations (Athens, Tennessee, and "Cleveland"). Therefore, height of the glenoid prominence did not impact the accuracy of the five- and two-variable models in the Athens and Tennessee population group, however, the lateral curvature measurement could have potentially been an age-related measurement for a lowered accuracy rate of the Athens male individuals.

Measurement reliability, or the ability to reproduce the same measurement value, is another factor that may contribute to the accuracy of the five- and two-variable models. The inter-observer bias has an effect on the accuracy of the measurements obtained for the five- and two-variable models. The results of the paired t-tests for the intra-observer bias showed no significant differences between the measurements obtained for the five- and two-variable models. However, the results for the inter-observer test showed that some of the measurements for the five- and two-variable model were consistently unreliable when the original measurements were compared against the measurements taken by the second researchers (Table 3.18). The resulting p-values showed three measurements (the breadth of scapula, the lateral curvature, and the thickness of the lateral border) for the Athens collection and one measurement (the lateral curvature) for the Tennessee collection had p-values lower than 0.05. This indicates that those measurements were statistically different when attempting to reproduced the same measurement value. These measurements, especially the lateral curvature and the thickness of the lateral border, have particularly unclear landmarks that could have resulted in the lower p-values. The thickness of the lateral border requires the measurement of the thickness of the lateral border at the midpoint between the inferior margin of the glenoid prominence and the inferior angle. How to obtain that midpoint, however, is not clearly defined by Dabbs and Moore-Jansen (2010). The lateral curvature also uses a coordinate caliper. It is difficult to obtain consistently accurate measurements with this tool. Measurements that are unreliably taken could affect the accuracy rate of the entire method, which could result in the methodology being challenged in the court of law. Some researchers have revised previous methodologies to make them more accurate and reliable. For example, Blanchard (2006) conducted a study on sex estimation methodological of the pelvis and found that the Bruzek (2002) methodology was less accurate and less reliable than previously reported by the original study. This was determined through inter- and intra-observer analyses and when Blanchard (2006) revised the methodology by removing some of the more erroneous variables in the Bruzek (2002) methodology, the accuracy of the model increased. Erroneous variables could lead to the exclusion of a methodology in an expert witness testimony. However, in recent years, criteria for forensic specialists have been outlined by the courts in several legal cases, mainly Daubert, Kumho, and Mohan rulings.

Daubert, Kumho, and Mohan criteria require forensic specialists to substantiate their assertions with scientifically tested methods and with probability assessments. This has promoted an improvement, and a stronger focus on, quantitative methods for hypothesis testing and probability estimation. These developments have changed not only forensic anthropology practice and methodologies, but also ‘‘the standards by which the profession determine(s) what should count as an admissible problem, or as a legitimate problem–solution’’ (Kuhn, 1970: 6). The Daubert criteria in 1993 stipulated that any evidence or methodology brought into court must: “(1) be (and has been) tested using the scientific method, (2) the technique has been subject to peer review, preferably in the form of publication in peer reviewed literature, (3) there are consistently and reliably applied professional standards and known or potential error rates for the technique, and (4) consider general acceptance within the relevant scientific community” (Christensen 2004: 2). The Kumho ruling in 1999 took these Daubert criteria one step further and concluded that expert witnesses can base their evidence on their own observations and experiences as relevant to the case, that all of their evidence should be evaluated with the same “level of rigor”, and the Daubert criteria are guidelines and may not be applicable to every case. (Grivas and Komar 2008: 772). Mohan is a Canadian ruling in which expert witness testimony has to have: “relevance, necessity in assisting the trier of fact, the absence of any exclusionary rule, and the qualifications of the expert” (Rogers 2004: 2).

These rulings (Daubert, Kumho, and Mohan) affect all forensic specialists when testifying in court, but the relevance they have for forensic anthropologists are to the accuracy and reliability of identification methodologies. Page et al. (2011) examined judiciary rulings and challenges to exclude or limit expert testimony in the light of the Daubert and Kumho criteria. The researchers found that out of 541 cases involving forensic expert testimony between 1993 and 2008, there were 81 cases in which the forensic evidence was excluded and/or limited. Out of those 81 cases, 51 cases were excluded because of reliability issues (e.g. the conduct of the witness, the accuracy and reliability of the methodology presented, and the underlying premises of their conclusions were not proven) (Page et al. 2011). Page et al. (2011: 1183) states, “this study reveals that the reliability of forensic identification sciences is still suffering criticism in the courts, and is responsible for the majority of exclusions or limitations of such evidence”. Also, Grivas and Komar (2008) specifically looked at forensic anthropology identification techniques and their inclusion into expert witness testimony under the Daubert and Kumho criteria. These researches found methodologies of estimating age at death, sex, and stature of unknown individuals are subject to the Daubert guidelines. Although these types of methodologies have measureable accuracies and documented error rates, the researchers found that they are not entirely objective. The data set and sample on which these methodologies were created (i.e., their skeletal reference sample), as well as the observer’s experience with the methodology are often challenged in court (Grivas and Komar 2008). This becomes increasingly important when discussing the results of the current study. The five- and two-variable models by Dabbs and Moore-Jansen (2010) need to be scientifically tested through a peer-reviewed process or they will face problems of accuracy and reliability within a legal standing. In the current study, having the five-and two-variable models tested on outside population groups allows these methodologies to be creditable methodologies through the scientific process of objectivity.

The results of the current study show that the five- and two-variable models adhere to the Daubert, Kumho, and Mohan criteria when estimating sex from an unknown individual. However, the paired t-tests in Table 3.18 show that some of the measurements, mainly the lateral curvature and thickness of the lateral border, can be unreliably taken if the observer using the methodology is inexperienced with the methodology of the five-variable model. To conform to the Daubert, Kumho, and Mohan criteria, the observer using the five-variable model needs to make it clear that they have adequate knowledge and understanding of the methodology and that those measurements (the lateral curvature and thickness of the lateral border) were reliably taken, since this is one of the main challenges facing expert witness testimony in forensic anthropology.

4.3 FORDISC 3.0 validation study and population diversity

The current study examined the accuracy rate of the FORISC 3.0 by Jantz and Ousley (2005) program for estimating sex from the scapula on two White European population groups. Also, the current study examined if the two-variable model by Dabbs and Moore-Jansen (2010) more accurately determines sex than FORDISC 3.0 since the two methodologies use the same measurements. The accuracy rates for the five-variable model were not compared to FORDISC 3.0 because it does not use the same measurements as FORDISC 3.0 for estimating sex from the scapula.

With regard to population diversity, the current study showed that the measurement the maximum height of the scapula, which is used in the FORDISC 3.0 discriminate function model, was statistically different for males and females between the Athens and Tennessee population groups (Tables 3.8 and 3.9). The other measurement used in the FORDISC 3.0 methodology, the maximum breadth of the scapula, was not significantly different between the two population groups in either sex (males or females). Although there are slight differences in the size of the measurements between population groups, this does not have an overall effect on the accuracy of the two methodologies (FORDISC 3.0 and the two-variable model) for estimating sex from the scapula.

When the discriminate functions of FORDISC 3.0 were applied to the left scapula of a population group from within the Forensic Data Bank, i.e. the Tennessee population group, and from outside the Forensic Data Bank, i.e. the Athens population group, there were no significant differences in the accuracy rates (Table 3.16) for estimating sex. The overall, combined male and female, accuracy rate for the Athens population group was 88.2% and for the Tennessee population group it was 92.3%. Also, p-values between typical and posterior probability of all the individuals in the Athens and Tennessee population group, as classified by FORDISC 3.0, show that almost all of the individuals are being classified correctly with high levels of posterior probability (Figures 3.1-3.4). The significance of these results is that posterior probability tells how an unknown individual compares to the range of variation in the reference data set, i.e. the Forensic Data Bank, which is used to develop the discriminant functions of FORDISC 3.0. A p-value of 0.05 or less indicates that the unknown individual is outside the range of variation from the reference data set to which the unknown individual has been assigned by FORDISC 3.0. All of the individuals in this study had posterior probabilities higher than 0.05, therefore, the results show high levels of accuracy in the allocation of group membership. However, typical probability expands on the posterior probability in that the discriminant functions of FORDISC 3.0 will allocate the unknown individual to one of the groups (male or female) regardless of whether that unknown individual is a member of the group or not. For example, the unknown remains may be of a biological male but classified by FORDISC 3.0 as a female because the individual does not show the pattern of sexual dimorphism (based on the measurements inputted into the program) present in the reference data set (Albanese 2012).

Guyomarc'h and Bruzek (2011) and Jantz and Ousley (2005) suggest that when typical probability values fall below 0.05 the accuracy of that classification can be unreliable and should be excluded. Even though the posterior probability of the individual gives a correct sex classification, the low typical probability suggests that the individual, or individuals, fall outside the range of variation for FORDISC 3.0 to correctly classify the sex of the unknown person. Low typical probability does not mean that the individual is misclassified; it just means that the individual falls outside the pattern of variation present within the Forensic Data Bank. In the Athens population group, two males and two females had typical probability values less than 0.05 and in the Tennessee population group only one male individual had a typical probability less than 0.05 (Figures 3.1-3.4). The results from the current research suggest that, when estimating sex from the scapula, the FORDISC 3.0 program is accurate when applied to geographically similar population groups since the posterior probability values were high and very few individuals displayed low typical probability.

In the current study only fourteen males and one female were misclassified in the Athens population group and eleven males and two females were misclassified in the Tennessee population group (Table 3.16). The reason for this misclassification and a lower, but not statistically significant, accuracy rate between the Athens and Tennessee population groups is a result of how FORDISC 3.0 creates its discriminate functions. No studies have been conducted on FORDISC 3.0's ability to estimate sex, specifically on the scapula, especially when trying to evaluate its applicability to populations outside the Forensic Data Bank. However, several studies have examined the strength of FORDISC 3.0 in estimating sex from other skeletal elements.

In 1998, Ousley and Jantz examined secular changes in the length of postcranial elements from different population groups from the nineteenth and twentieth centuries. They noticed that, when the lengths of long bones of White European and Black African males were compared, homogenization was occurring from the nineteenth to the twentieth century. This created accuracy rate problems for previous discriminate functions that estimated sex and ancestry from postcranial elements because contemporary population groups were no longer as biologically distinct as previously thought. The FORDISC program, which uses data from the Forensic Data Bank, was created to address the accuracy rate issues of the discriminate function models that were derived from nineteenth century populations within the United States (Giles and Elliot, 1962; Giles and Elliot, 1963; Işcan and Cotton, 1990). Although the FORDISC program is constantly being updated and more skeletal material from other reference collections is added to the database, there are accuracy rate problems that still persist when estimating sex on unknown individuals from outside the Forensic Data Bank.

Ramsthaler and colleagues (2007) conducted a study of cranial measurements from a Germanic population and used FORDISC 3.0 to estimate sex. They cited that FORDISC 3.0 was unreliable when trying to estimate sex. The study also compared the results from FORDISC 3.0 to another methodology involving a visual assessment of cranial features for estimating sex. The morphological assessment had a higher level of accuracy than FORDISC 3.0 for estimating sex from the German skeletons. Ramsthaler and colleagues (2007) concluded FORDISC 3.0 was unreliable because their population group was not part of the reference data set, i.e. their Germanic population was outside the Forensic Data Bank reference collection. Guyomarc'h and Bruzek (2011) also used FORDISC to estimate sex on population groups outside the Forensic Data Bank. Guyomarc'h and Bruzek (2011) used French and Thai reference collections and 12 cranial measurements to assess the accuracy of FORDISC 3.0. They concluded that FORDSIC 3.0 had an accuracy rate ranging from 52% to 77% depending on the ancestral groups selected during analyses. The results indicate that with regard to FORDISC 3.0 and estimating sex through cranial analyses, the accuracy of the program is limited to those individuals found within the Forensic Data Bank.

However, in the current study these misclassification issues are not present in the postcranial skeleton, specifically the scapula, in the population group outside the Forensic Data Bank, i.e. the Athens population group. This reinforces that with regard to estimating sex from the scapula using FORDISC 3.0, the program can be used on geographically similar population groups. Although the accuracy rate differences were minor, the current study illustrates that population diversity may not be a significant factor to the accuracy rates for estimating sex using FORDISC 3.0 on outside populations groups as previous research concluded, especially for the scapula.

When comparing the accuracy rates of the FORDISC 3.0 program and the two-variable model each methodology has its own strengths and weaknesses. In the Athens population, FORDISC 3.0 correctly classified females (98.0%) more often than males (81.8%). In the Tennessee sample, FORDISC 3.0 correctly classified females (97.3%) more often than males (88.3%) (Table 3.16). When comparing the FORDISC 3.0 methodology with the two-variable model, the males in the Athens sample were more accurately classified using the two-variable model (87.0%) whereas females were more correctly classified with the FORDISC 3.0 methodology (98.0%). In the Tennessee sample, the males were more accurately classified using the two-variable model (94.7%) whereas females were more accurately classified with the FORDISC 3.0 methodology (97.3%). Also, the chi-squared proportion test shows that there was a significant difference between the accuracy proportions of the Tennessee females when comparing the FORDISC 3.0 program and the two-variable model (Table 3.17). This shows that, in the Tennessee population group, the two-variable model correctly classifies females less often than the FORDISC 3.0 program. However, both FORDISC 3.0 and the two-variable model have high accuracy rates, despite minor statistical differences. The reason for this classification difference is that the reference sample for FORDISC 3.0, the Forensic Data Bank, is more similar to the Tennessee population group than to the Athens population. Therefore, FORDISC 3.0 correctly identifies females within the Tennessee population group more often. Although the independent t-tests and Mann-Whitney U-tests showed no statistical difference between the Tennessee population group measurements (the maximum height of the scapula and the maximum breadth of the scapula) and the "Cleveland" sample (from which the two-variable model was created), the classification accuracy rate was still low in the females of the Tennessee population group. The reason more females were being misclassified may be due to a small sample size for this group (i.e. N= 76)

Although, the two-variable model was less accurate in classifying females within the Tennessee population group the model is designed to eliminate the computational human errors that are associated with FORDISC 3.0. The two-variable model uses the same two measurements as FORDISC 3.0 but the two-variable model uses only one equation to estimate sex, which is more user-friendly. The advantages of using the two-variable model over the FORDISC 3.0 program are the utility of estimating sex from the scapula without the proprietary software, which can be less available to some forensic anthropologist. In terms of accuracy rates, FORDISC 3.0 is still the more accurate program for estimating sex in a female North American population group, specifically a Tennessee population group. However the FORDISC 3.0 program had a sex bias ratio that was double the sex bias ratio of the two-variable model in the Athens population group. The sex bias ratio was high for the FORDISC 3.0 analyses in the Athens population group and, for that reason, the two-variable model would be a better methodology when estimating sex from the scapula. The sex bias ratio is an indicator of how much a methodology estimates one sex over the other. If the sex bias ratio is closer to zero then the methodology can be considered equally discriminatory, since the end result is discriminating between bilateral group memberships (i.e. if it is not male, it is female). Since the FORDISC 3.0 program has a high sex bias ratio then that indicates that more males are being misclassified as females, which is related to the discriminate function model within FORDISC 3.0 and how it allocates group membership.

4.4 The models for "Macaluso's Hypothesis" for estimating sex from the glenoid cavity.

The third objective of this study was to provide an alternative method for estimating sex from the scapula of White European populations, based on metric analyses of the glenoid cavity. The current project showed that the height, breadth, and calculated area of the glenoid cavity were sexually dimorphic between males and females. All results showed that each measurement (height, breadth, and calculated area) had good predictive value for classifying males and females in both the Athens and Tennessee populations. Other studies for estimating sex from the scapula have also shown that the glenoid cavity is a good predictor of sex (Macaluso 2010; Papaioannou et al. 2012). Macaluso (2010) used the height, breadth, perimeter, and area of the glenoid cavity to estimate sex from the scapula of black South African skeletons. Macaluso (2010) found that the area of the glenoid cavity obtained by the software “ImageJ” and the area obtained by using hand held calipers (and multiplying height and breadth) had the same predictive qualities for estimating the sex of an individual. For the current study, the author decided to test that technique (multiply height and breadth to get calculated area) and create a new methodology for estimating sex from the glenoid cavity in White European population groups, i.e. the Athens and Tennessee skeletal collection. Papaioannou and colleagues (2012) used a Cretan skeletal sample to estimate sex from the scapula. The researchers found that the height and breadth of the glenoid cavity were two excellent predictors for estimating sex from the scapula in a Southeastern European population group. From those results they were able to create a population-specific discriminate function model for a modern Cretan population group.

Logistic regression analysis was used to create three models to estimate sex from the glenoid cavity: one model for the Athens population, one for the Tennessee population, and one for the combined populations model. Although discriminant function analysis is widely used in estimation of sex methodologies, logistic regression analysis was used because it is considered more accommodating than discriminate function analysis when analyzing small sample sizes (Acharya et al. 2010). Logistic regression analysis is “more flexible in its assumptions [and] it can handle both discrete and continuous variables, which need not be normally distributed, linearly related, or of equal variance” (Acharya et al. 2010: 200). This allows for the independent variables (height, breadth, and calculated area of the glenoid cavity) to predict group membership (i.e. males and females) with a similar predictive accuracy as a discriminate function model, a smaller sample size, and less variation between independent variables. For the rest of this chapter the Athens and Tennessee logistic regression models for estimating sex will be cited as the (Athens, Tennessee and Combined Population) models for “Macaluso’s Hypothesis” for simplicity.

The Athens and Tennessee populations were combined into one White European population sample even though the resulting p-values of the glenoid cavity measurements suggested that the two populations were statistically different with regard to the height, breadth, and calculated area. However, the results from the previous research conducted on population diversity may indicate that population specific formulas for sex estimation may not be necessary. Steyn and Patriquin (2009) in their study of different population groups of White Europeans and Black Africans of Greece, South Africa, and North America showed that larger sample sizes of similar ancestries could “eradicate the need for population-specific formulae” since larger sample sizes tend to remove minor population differences (Steyn and Patriqun 2009:113.e3). This resulted in the creation of three logistic regression models for estimating sex from the glenoid cavity for the two population groups (Athens and Tennessee).

Churchill et al. (2001) and Merrill et al. (2009) examined the size of the glenoid cavity and the differences between males and females. Both studies utilized White European and Black African scapula from the Hamman-Todd Skeletal Collection in Cleveland. Although their research goals were not to estimate biological sex for a forensic anthropology purpose, their results showed that the height and breadth of the glenoid cavity were highly sexually dimorphic. Also, interestingly, the results of Churchill et al.

(2001) showed that there were no size differences within the height and breadth of the glenoid cavity between White Europeans and Black Africans. These results correlate well with the current study on estimating sex from the glenoid cavity. Since the Athens and Tennessee population groups are of similar ancestries the need separate the two populations into two specific sex estimation methodologies is not necessary.

The accuracy rates of the (Athens, Tennessee and Combined Population) models for "Macaluso's Hypothesis" (Tables 3.29, 3.32 And 3.35) and the chi-squared proportion tests of the three models (Table 3.36) in the current study illustrate little difference between the two populations with regard to estimating sex from the glenoid fossa. The predicted accuracy of the Athens model for "Macaluso's Hypothesis" was 87.6% utilizing the Direct method (with all three sexually dimorphic measurements). The results of the Tennessee model for "Macaluso's Hypothesis" had an accuracy rate percentage of 94.6% utilizing the Direct method (with all three sexually dimorphic measurements). However the combined population model had a predicated accuracy rate of 89.0% utilizing the Direct method (with all three sexually dimorphic measurements). This is slightly better, though not statistically different, than the Athens model for "Macaluso's Hypothesis". The chi-squared proportion tests of all three models (the Athens, Tennessee and Combined Population models for "Macaluso's Hypothesis") in Table 3.36 showed that all three models had statistically similar accuracy rates. What makes the combined population model slightly more accurate is that the sample size is larger. The increased number of individuals in the combined population model helps give the model statistical strength when creating the logistic regression equation. These results indicate that population specific equations for estimating sex from the scapula are just as accurate if similar contemporary population groups are pooled together. Accuracy rates obtained from the Athens, Tennessee and Combined Population models for “Macaluso’s Hypothesis” are comparable to other studies on estimating sex from the glenoid cavity (Macaluso 2010; Papaioannou et al. 2012). Macaluso (2010) had accuracy rates ranging between of 83.3% to 90.0% and Papaioannou and colleagues (2012) had an accuracy rate of 92%.

To ensure measurement accuracy and avoid measurement bias, intra- and inter-observer measurement error was examined. The researcher examined intra-observer error by re-measuring the two variables (height and breadth) for the glenoid cavity on a sub-sample of 30 individuals from both the Athens and Tennessee population groups. Also, the results of the paired t-tests for the inter-observer error showed no significant differences between the measurements obtained for the glenoid fossa on a sub-sample of 30 individuals which were re-measured by a another observer. The overall results showed that the measurements for these new glenoid cavity methodologies (all three models for “Macaluso’s Hypothesis”) can be accurately reproduced. The significance of these findings is that the measurements can be reliably taken. As mentioned previously, measurement reliability of any new estimation of sex methodology is one of the many challenges that forensic anthropologists face in the court room since the Daubert, Kumho, and Mohan rulings (Page et al. 2011; Grivas and Komar 2008; Christensen 2004; Rogers 2004). The current study illustrates that all three models for “Macaluso’s Hypothesis” can be reliably utilized by forensic anthropologists. One limitation to the models however, is that the models for “Macaluso’s Hypothesis” have not been tested on an outside population group. Further research needs to test these models on other contemporary White European population groups in North America and Europe. In this way the validity of the accuracy rates of the models for “Macaluso’s Hypothesis” can be proven.

Also, another potential limitation to the models for “Macaluso’s Hypothesis” could be the age of the individual. Hrdlička (1942) and Humphrey (1998) both conducted studies on age related changes in the scapula. They concluded that factors such as activity, occupation, division of labour, and diet could affect the scapular size and dimension since the scapula is “almost totally dependent on the muscles which are attached to it, [therefore] …. the ultimate form which the body of the bone achieves is of functional nature and due to muscular activity” (Hrdlička 1942: 85). Dabbs (2011) also concluded that as individuals age the height of the scapula, especially in males, decreased. With advanced age, the ventral curvature of the scapula increased while the overall scapular length decreases. Also, as an individual increases in age the ossification of bony material around the glenoid cavity increased the size of the cavity, therefore, affecting the overall height and breadth of the glenoid cavity.

A Pearson’s correlation test was performed on all scapular measurements for the new models for “Macaluso’s Hypothesis”. This was to establish whether or not there was a correlation between age-at-death and glenoid cavity size within the Athens and Tennessee population groups. The two variables within this analysis are the age of the individual at the time of their death and the three independent variables (height, breadth and calculated area) used for the logistic regression analysis. For all but one variable in the Athens population sample, the age of the individual at death had an effect on the size of the variable. The Tennessee population group had only one variable that was correlated with age.

The overall results of the Pearson’s correlation tests suggest that the age of the individual may play a role in the accuracy of the glenoid cavity models, especially for the Athens population group. Dabbs and Moore-Jansen (2012) found that the glenoid cavity increases in size with advanced age due to age related changes such as the ossification of the ligaments surrounding that area. Some of the individuals utilized in the Athens population group were much older than those of the Tennessee population group for both males and females. This is a reason why the individuals in the Athens model for “Macaluso’s Hypothesis” had a lower predicted accuracy than those individuals of the Tennessee model for “Macaluso’s Hypothesis” However, future research needs to examine the age of an individual and its effect on the accuracy of the glenoid cavity as a predictor of sex.

CHAPTER 5: CONCLUSION

The objectives of this thesis were as follows:

1. To understand the relationship between biological sex and estimated sex from the scapula based on metric analyses of two White European population groups from North America and Europe.
2. To test the accuracy and reliability of two scapular methodologies published by Dabbs and Moore-Jensen (2010) and Jantz and Ousley (2005).
3. To provide an alternative method for estimating sex from the scapula, of White European populations, based on metric analyses of the glenoid cavity.

In this study, 307 individual scapulae from two skeletal reference collections, the Athens Collection and the William Bass Collection (i.e. the Tennessee collection), were measured. Only left scapulae of White European individuals were measured as per standards set by other researchers and to maintain anatomical consistency (Builkstra and Ubelaker, 1994; Stewart, 1979). Only scapulae of adult individuals (ages 20+ years) were used as subadult scapulae have not reached their maximum size. Damaged or remodelled scapulae were excluded from the study. The five-variable and two-variable models are two metric methodologies for estimating sex from the scapula developed by Dabbs and Moore-Jansen (2010). The classification accuracy of each model was determined by its potential to correctly classify an individual as being either male or female.

The results of the Dabbs and Moore-Jansen (2010) validation study showed that in the Athens population group, males were accurately classified using the five-variable model (94.8%) whereas females were classified with the same accuracy (94%) using both the five- and the two-variable models. Combined male and female classification accuracy of the Athens population group showed that the five-variable model had a greater accuracy (94.5%) than the two-variable model (89.7%). The sex bias ratio percentage was examined. This illustrates how biased the model was toward classifying one sex over the other. For the Athens population, the five-variable model correctly classified males more often than females by 0.8%. The two-variable model classified females more accurately than males by 7.0%.

In the Tennessee sample, both males (96.8%) and females (94.7%) were accurately classified using the five-variable model. Combined male and female classification accuracy rates of the Tennessee population group showed that the five-variable model had a greater accuracy (95.9%) than the two-variable model (89.4%). The sex bias ratio for the Tennessee population correctly classified males more often than females by 2.1% (using the five-variable model) and classified males more often than females by 11.8% (using the two-variable model).

The scapular measurements used in the five-and two-variable models were highly sexually dimorphic and they contributed to the differentiation of biological sex. In the current study, the five- and two-variable model had very high accuracy rates between population groups that reflect the differentiation between sex rather than reflect the differentiation between population groups. This means that the accuracy of the five-and two-variable models were not determined by which population group it was used on but rather on the sexually dimorphic scapular measurements within the models themselves. The current study analyzing the five- and two-variable models by Dabbs and Moore-Jansen (2010) indicate that the five- and two-variable models can achieve a high level of accuracy between contemporary White European population groups, specifically Greek and North American populations. This proves that discriminate function models can be used on similar population groups if the right combination of measurements is used. Future research should explore the possibility of using the five- and two-variable models on other ancestries to test their accuracy and reliability for estimating sex from the scapula.

In this study, scapulae from both the Athens and Tennessee collections were used to evaluate the FORDISC 3.0 program created by Jantz and Ousley (2005). The results of the Jantz and Ousley (2005) validation study showed that the FORDISC 3.0 analyses on the Athens sample correctly classified females (98.0%) more often than males (81.8%) with a sex bias ratio of 16.2% for correctly classifying females over males. In the Tennessee sample, females (97.3%) were more correctly classified than males (88.3%) with a sex bias ratio of 9.0% for correctly classifying females over males. The Tennessee population group showed a higher rate of combined male and females accuracy classification (92.3%) than the Athens population group (88.2%).

In the current study only fourteen males and one female were misclassified in the Athens population group and eleven males and two females were misclassified in the Tennessee population group. The reason for this misclassification and a lowered, but not statistically significant, accuracy rate difference between the Athens and Tennessee population groups was a result of how FORDISC 3.0 creates its discriminate functions. Very few studies have been conducted on the ability of FORDISC 3.0 to estimate sex from the scapula, especially when trying to evaluate its applicability to skeletal populations outside the Forensic Data Bank. However, this study showed that the FORDISC 3.0 program can be used across more geographically distinct White European population groups outside the Forensic Data Bank. Although the accuracy rate differences were insignificant, the current study illustrates that population differences may not be as damaging to the accuracy rates for the estimation of sex using FORDISC 3.0 on outside populations groups as previous research concluded, especially for the scapula.

Also, FORDISC 3.0 and the two-variable model by Dabbs and Moore-Jansen (2010) were compared to determine which methodology was more accurate and reliable for sex estimation. In the Athens sample, males were more accurately classified using the two-variable model (87.0%) whereas females were more correctly classified with the FORDISC 3.0 methodology (98.0%). In the Tennessee sample, males were more accurately classified using the two-variable model (94.7%) whereas females were more correctly classified with the FORDISC 3.0 methodology (97.3%).

The current study showed that the FORDISC 3.0 program underestimates the sex classification of males more frequently than females when applied to the Athens and Tennessee population groups. It was also shown that the two-variable model underestimates the sex classification of females in the Tennessee population and underestimates the sex classification of males in the Athens population group. The advantages of using the two-variable model over the FORDISC 3.0 program are the utility of estimating sex from the scapula without purchasing the FORDISC proprietary software, which can be an expense not affordable to some forensic anthropologist, e.g. those working in developing nations. In terms of accuracy rates, FORDISC 3.0 is still the more accurate program for estimating sex in a female White North American population group, specifically the Tennessee population group. However, when the FORDISC 3.0 program was applied to the Athens population group, it produced a sex bias ratio double than what was produced when the two-variable model was applied to the sample. Although, the sex bias ratio was high for the FORDISC 3.0 analyses in the Athens population group, the two-variable model is a slightly more accurate methodology when estimating sex from the scapula. Future research should examine how the FORDISC 3.0 program allocates group membership (i.e. sex and/or ancestry) from postcranial remains (e.g. the upper limbs and lower limbs) for population groups that are not included in the Forensic Data Bank. This would show the applicability of the FORDISC 3.0 program on a global scale and address the question of population-specific formulae for estimating sex from postcranial skeletal remains.

Logistic regression analysis was used to create three models to estimate sex from the glenoid cavity: one model for the Athens population, one for the Tennessee population, and a combined population model. To increase the sample size and to increase the statistical strength of the logistic regression equation, the author combined the two sample populations, i.e. Athens and Tennessee, into one White European population sample. Even though the results showed that the glenoid cavity measurements were different between the two populations, the results from the previous research by Steyn and Patriquin (2009) indicate that population specific formulas for sex estimation may not be necessary between similar population groups. They concluded that larger sample sizes tend to remove minor population differences. The author of the current study examined how that might affect the accuracy rates of estimation of sex methodologies when combining similar populations groups (mainly the Athens and Tennessee populations). This resulted in the creation of three logistic regression models for estimating sex from the glenoid cavity for the two population groups (Athens and Tennessee), the third being a combined population logistic regression model.

The results of “Macaluso’s Hypothesis” to estimate sex from the glenoid cavity created three new methodologies to estimate sex in two different population groups. The independent t-tests and the Mann-Whitney U-tests between the males and females of both population groups illustrated that each variable (height, breadth, and calculated area of the glenoid cavity) had good predictive value for classifying males and females in both the Athens and Tennessee population groups. Accuracy rates of the models for “Macaluso’s Hypothesis” (Athens, Tennessee, and Combined Populations) were comparable to other studies for estimating sex from the glenoid cavity (Macaluso 2010; Papaioannou et al. 2012). Macaluso (2010) had accuracy rates ranging between of 83.3% to 90.0% and Papaioannou and colleagues (2012) had an accuracy rate of 92%. The predicted accuracy of the Athens model was 87.6% utilizing the Direct method (with all three sexually dimorphic measurements). The results of the Tennessee model showed that the model has an accuracy rate percentage of 94.6% utilizing the Direct method (with all three sexually dimorphic measurements).

However the Combined Population model had a predicated accuracy rate of 89.0% utilizing the Direct method (with all three sexually dimorphic measurements). This is slightly better, though not statistically different, than the Athens model. The chi-squared proportion tests of all three models (the Athens, Tennessee and Combined Population models for “Macaluso’s Hypothesis”) showed that all three models had statistically similar accuracy rates. What makes the combined population model slightly more accurate is that the sample size is larger. The increased number of individuals in the combined population model helps give the model statistical strength when creating the logistic regression equation. These results indicate that population specific equations for estimating sex from the scapula are just as accurate if similar contemporary population groups are pooled together.

The results of the current study on estimating sex from the scapula have produced three significant results: (1) the five- and two-variable model are accurate methods for estimating sex over White European populations, (2) FORDISC 3.0 is an accurate methodology for estimating sex from the scapula in populations from *outside* the Forensic Data Bank and, (3) the glenoid cavity can be used as an accurate methodology for estimating sex from the scapula and the difference in accuracy rates between similar populations groups are not significant. All of these results suggest that population diversity and the need for population specific formulae for estimating sex from the scapula are not necessary. Accurate methodologies for estimating sex from the scapula can be achieved by combining data from similar contemporary populations rather than creating equations for the separate populations.

REFERENCES:

Acharya, A. B., S. Prabhu, and M. V. Muddapur

2010 Odontometric sex assessment from logistic regression analysis. *International Journal of Legal Medicine* 125(2):199-204

Acocke, A. C.

2005 SAS, Stata, SPSS: A Comparison. *Journal of Marriage and Family* 67: 1093–1101.

Arsuaga, J.L., and J.M. Carretero

1994 Multivariate analysis of the sexual dimorphism of the hip bone in a modern population and in early hominids. *American Journal of Physical Anthropology* 93(2):241-257.

Bainbridge, D. and S. G. Genovese-Tarazaga

1956 A study of sex differences in the scapula. *Journal of the Royal Anthropological Institute* 86: 109–13.

Barrio, P. A., G. J. Trancho, and J. A. Sánchez

2006 Metacarpal sexual determination in a Spanish population. *Journal of Forensic Sciences* 51(5):990-995.

Benazzi, S., C. Maestri, S. Parisini, F. Vecchi, and G. Gruppioni

2008 Sex assessment from the acetabular rim by means of image analysis. *Forensic Science International* 180(1):58.e1- 58.e3.

Bidmos, M. A., and S. A. Asala

2003 Discriminant function sexing of the calcaneus of the South African whites. *Journal of Forensic Sciences* 48(6):1213-8.

Bidmos, M. A., and S. A. Asala

2004 Sexual dimorphism of the calcaneus of South African blacks. *Journal of Forensic Sciences* 49(3):446-50.

Bidmos, M. A., and M. R. Dayal

2004 Further evidence to show population specificity of discriminant function equations for sex determination using the talus of South African blacks. *Journal of Forensic Sciences* 49(6):1165-1170.

Black, T.K.

1978 A new method for assessing the sex of fragmentary skeletal remains: Femoral shaft circumference. *American Journal of Physical Anthropology* 48(2):227-232.

Bronowski, J., and W. M. Long.

1952 Statistics of discrimination in anthropology**.** *American Journal of Physical Anthropology* 10(4):385–394.

Bruzek, J.

2002 A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology* 117(2):157-168.

Buikstra, J.E., and D. Ubelaker

1994 Standards for data collection from human skeletal remains: Proceedings of a seminar at the field museum of natural history. Fayetteville: Arkansas Archeological Survey Research Series No. 44.

Burns, R. B. and R. A. Burns

2008 *Business Research Methods and Statistics Using SPSS.* Los Angeles; London: Sage. Chapter 24: 569-588 retrieved from http://www.uk.sagepub.com/burns/website%20material/Chapter%2024%20-%20Logistic%20regression.pdf

Byers, Steven N.

2008 *Introduction to Forensic Anthropology, 3rd ed.* Boston: Pearson Education, Inc.

Campbell, Neil A., and Jane B. Reece

2005 *Biology, 7th ed.* San Francisco: Pearson Education, Inc.

Case, D. T. and A. H. Ross

2007 Sex determination from hand and foot bone lengths. *Journal of Forensic Sciences* 52(2):264-70.

Charisi, D., C. Eliopoulos, V. Vanna, C. G. Koilias, and S. K. Manolis

2011 Sexual Dimorphism of the Arm Bones in a Modern Greek Population. *Journal of Forensic Sciences* 56(1): 10–18.

Christensen, A. M.

2004 The impact of Daubert: implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). *Journal of Forensic Science* 49(3):427-30.

Churchill, R.S., J.J. Brems, and H. Kotschi

2001 Glenoid size, inclination, and version: an anatomic study. *Journal of Shoulder Elbow Surgery* 10(4):327-332.

Cologlu A. S., M. Y. Iscan, M. F. Yavuz, and H. Sari

1998 Sex determination from the ribs of contemporary Turks. *Journal of Forensic Science* 43(2):273-276.

Corcos, Alain F.
1997 *The Myth of Human Races*. East Lansing: Michigan State University Press.

Dabbs, G. R.
2010 Sex determination using the scapula in New Kingdom skeletons from Tell El-Amarna. *Homo* 61(6):413-420.

Dabbs, G.R., and P. H. Moore-Jansen
2010 A method for estimating sex using metric analysis of the scapula. *Journal of Forensic Science* 55(1): 149–152.

DiVella, G., C. P. Campobasso, M. Dragone, and F. Introna
1994 Skeletal sex determination by scapular measurements. *Boll. Soc. Ital. Biol. Sper*. 70: 299–30.

Dwight, T.
1887 The Range of Variation of the Human Shoulder-Blade. *The American Naturalist* 21(7): 627-638.

Dwight, T.
1894 The range and significance of variation in the human skeleton. *Boston Medicine and Surgery Journal* 131, 73–76, 97–101.

Eliopoulos, C., A. Lagia, and S. Manolis
2007 A modern, documented human skeletal collection from Greece. *Homo* 58(3):221-8.

Ember, C. R., and M. Ember
1998 Anthropology: A Brief Introduction, 3rd ed. New Jersey: Prentice Hall, Inc.

Ferembach, D., I. *Schwidetzky, and M. Stloukal*
1980 *Recommendations for age and sex diagnoses of skeleton. Journal of Human Evolution 9:517–549.*

Flander, L.B.
1978 Univariate and multivariate methods for sexing the sacrum. *American Journal of Physical Anthropology* 49(1):103-110.

Falsetti, A. B.
1995 Sex assessment from metacarpals of the human hand. *Journal of Forensic Sciences* 40(5):774-776.

Falys, Ceri G., H. Schutkowski, and D. A. Weston
2005 The Distal Humerus—A Blind Test of Rogers' Sexing Technique Using a Documented Skeletal Collection. *Journal of Forensic Sciences* 50(6): 1-5

France, D. L.
1998 Observation and metric analysis of sex in the skeleton. In: Reichs, K. (Ed.) *Forensic Osteology: Advances in the Identification of Human Remains, 2nd Ed.* Charles C. Thomas, Springfield pp. 218-228.

Frutos, L. R.
2002 Determination of sex from the clavicle and scapula in a Guatemalan contemporary rural indigenous population. *American Journal of Forensic Medicine and Pathology* 23(3):284-8

Giles, E. and O. Elliot
1962 Race Identification from Cranial Measurements. *Journal of Forensic Sciences* 7: 147-157.
1963 Sex Determination by Discriminant Function Analysis of Crania. *American Journal of Physical Anthropology* 21: 53-68.

Gonzalez P.N., Bernal V., Perez S.I.
2009 Geometric morphometric approach to sex estimation of human pelvis. *Forensic Science International* 189:68-74.

Grivas, C.R. and D.A. Komar
2008 Kumho, Daubert, and the nature of scientific inquiry: implications for forensic anthropology. *Journal of Forensic Science* 53(4):771-6.

Gualdi-Russo, E.
2007 Sex determination from the talus and calcaneus measurements. *Forensic Science International* 171(2-3):151-6.

Guyomarc'h P., and J Bruzek
2011 Accuracy and reliability in sex determination from skulls: a comparison of Fordisc® 3.0 and the discriminant function analysis *Forensic Science International* 208(1-3):180.e1-e6.

Harris, S. M. and D. T. Case
2012 Sexual Dimorphism in the Tarsal Bones: Implications for Sex Determination. *Journal of Forensic Sciences* 57: 295–305.

Hrdlicˇka, A.
1942 The Adult Scapula: Additional Observations and Measurements. *American Journal of Physical Anthropology* 29: 363–415.

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

Introna, F. Jr., G. Di Vella, C. P. Campobasso, and M. Dragone
1997 Sex determination by discriminant analysis of calcanei measurements. *Journal of Forensic Sciences* 42(4):725-8.

Işcan, M. Y.
1985 Osteometric analysis of sexual dimorphism in the sternal end of the rib. *Journal of Forensic Sciences* 30(4): 1090–1099.

Iscan, M.Y., and T.S. Cotton.
1990 Osteometric assessment of racial affinity from multiple sites in the postcranial skeleton. In: Gill, G.W. and Rhine, S. (eds) *Skeletal Attribution of Race: Methods for Forensic Anthropology*. Maxwell Museum of Anthropology, Anthropological Papers No. 4

Işcan, M. Y., and K. Derrick
1984 Determination of sex from the sacroiliac joint: A visual assessment technique. *Florida Scientist* 47(2):94-98.

Işcan, M.Y., and D Shihai
1995 Sexual dimorphism in the Chinese femur. *Forensic Science International* 74:79-87.

Işcan, M. Y., S. R. Loth, C. A. King, D. Shihai, and M. Yoshino
1998 Sexual dimorphism in the humerus: A comparative analysis of Chinese, Japanese, and Thais. *Forensic Science International* 98:17-29.

Jorde, Lynn B., and Stephen P. Wooding
2004 Genetic Variation, Classification, and "Race." Nature Genetics 36(11):S28-S33.

Lazenby RA.
1994 Identification of sex from metacarpals: effect of side asymmetry. *Journal of Forensic Sciences* 39(5):1188-1194.

Letterman, G.S.
1941 The greater sciatic notch in American whites and Negroes. *American Journal of Physical Anthropology* 28(1):99-116.

Loth, S.R., and M. Henneberg
1996 Mandibular ramus flexure: A new morphologic indicator of sexual dimorphism in the human skeleton. *American Journal of Physical Anthropology* 99(3):473-485.

Lovell, N. C.
1989 Test of phenice's technique for determinant sex from the os pubis. *American Journal of Physical Anthropology* 79(1):117-120.

Lynnerup, N., M. Schulz, A. Madelung, and M. Graw
2006 Diameter of the human internal acoustic meatus and sex determination. *International Journal of Osteoarchaeology* 16:118-123.

Khanpetch, P., S. Prasitwattanseree, D. T. Case, and P. Mahakkanukrauh
2012 Determination of sex from the metacarpals in a Thai population. *Forensic Science International* 217(1-3):229.e1-e8.

Kindschuh, S. C., T. L. Dupras, and L. W. Cowgill
2010 Determination of sex from the hyoid bone. *American Journal of Physical Anthropology* 143(2):279-84.

Kim D. I., U. Y. Lee, D. K. Park, Y. S. Kim, K. H. Han, K. H. Kim, and S. H. Han
2006 Morphometric of the hyoid bone for human sex determination from digital photographs. *Journal of Forensic Sciences* 51:979–984

King, C. A., M. Y. Işcan, and S. R. Loth
1998 Metric and comparative analysis of sexual dimorphism in the Thai femur. *Journal of Forensic Science* 43(5):954-958.

Konigsberg, L.W., and S. M. Hens
1998 Use of ordinal categorical variables in skeletal assessment of sex from the cranium. *American Journal of Physical Anthropology* 107:97- 112.

Kottak, Conrad Phillip
2007 *Mirror for Humanity, 5th ed.* New York, NY: McGraw-Hill, Inc.

Krogman, W., and M. Y. Işcan
1986 *The human skeleton in forensic medicine*. Springfield, IL: Charles C. Thomas.

Kuhn T.
1970. *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Macaluso, J. P.
2010 Sex determination from the glenoid cavity in black South Africans: morphometric analysis of digital photographs. *International Journal of Legal Medicine*, retrieved online January 2, 2011 from http://www.springerlink.com/content/f554606h4k173ln6/.

MacLaughlin, S. M., and M. F. Bruce
1985 A simple univariate technique for determining sex from fragmentary femora: Its application to a Scottish short cist population. *American Journal of Physical Anthropology* 67(4): 413–417.

Manolis S. K., C. Eliopoulos, C. G. Koilias, and S. C. Fox
2009 Sex determination using metacarpal biometric data from the Athens Collection. *Forensic Science International* 193(1-3):130.e1-6.

Marino, E. A.
1995 Sex estimation using the first cervical vertebra. *American Journal of Physical Anthropology* 97(2):127-33.

Marlow E. J., and R. F. Pastor
2011 Sex determination using the second cervical vertebra--A test of the method. *Journal of Forensic Sciences* 56(1):165-9.

Mastrangelo, P., S. De Luca, I. Alemán, and M. C. Botella
2011 Sex assessment from the carpals bones: discriminant function analysis in a 20th century Spanish sample. *Forensic Science International* 206(1-3):216.e1-e10.

Mastrangelo, P., S. De Luca, and G. Sánchez-Mejorada
2011 Sex assessment from carpals bones: discriminant function analysis in a contemporary Mexican sample. *Forensic Science International* 209(1-3):196.e1-e15.

Merrill, A., K. Guzman, and S.L. Miller
2009 Gender differences in glenoid anatomy: an anatomic study. *Surgical Radiology Anatomy* 31(3):183-189.

Miller, K. W., P. L. Walker, and R. L. O'Halloran
1998 Age and sex-related variation in hyoid bone morphology. *Journal of Forensic Sciences* 43(6):1138-1143.

Milner, G. R. and J. L. Boldsen
2012 Humeral and Femoral Head Diameters in Recent White American Skeletons. *Journal of Forensic Sciences* 57(1): 35–40.

Molnar, Stephen
1983 *Human Variation: Races, Types, and Ethnic Groups.* New Jersey: Prentice-Hall.

Moore, David S.
1996 *The Basic Practice of Statistics, 5th edition*. New York: W. H. Freeman and Company.

Mountrakis, C., C. Eliopoulos, C. G. Koilias, S. K. Manolis
2010 Sex determination using metatarsal osteometrics from the Athens collection. *Forensic Science International* 200(1-3):178.e1-e7.

Murphy A. M. C.

2000 The acetabulum: Sex assessment of prehistoric New Zealand polynesian innominates. *Forensic Science International* 108(1):39-43.

Murphy, A. M. C.

2002 Articular surfaces of the pectoral girdle: sex assessment of prehistoric New Zealand Polynesian skeletal remains. *Forensic Science International* 125: 134–136.

Norén, A., N. Lynnerup, A. Czarnetzki, and M. Graw

2005 Lateral angle: A method for sexing using the petrous bone. *American Journal of Physical Anthropology* 128:318-323.

Ousley, S. D., and R. L. Jantz

1998 The Forensic Data Bank: Documenting Skeletal Trends in the United States. In: Reichs, K. (Ed.) *Forensic Osteology: Advances in the Identification of Human Remains (2nd Ed.),* Charles C. Thomas, Springfield pp. 441-459.

2005 FORDISC 3.0: Personal Computer Forensic Discriminant Functions. University of Tennessee.

Özer, I., K. Katayama, M. Sa˘gir, and E. Gülec

2006 Sex determination using the scapula in medieval skeletons from East Anatolia. *Collegium Antropologicum* 30: 415–419.

Page, M., J. Taylor, and M. Blenkin

2011 Forensic identification science evidence since Daubert: Part I--A quantitative analysis of the exclusion of forensic identification science evidence. *Journal of Forensic Sciences* 56(5):1180-1184.

Papaioannou, V. A., E. F. Kranioti, P. Joveneaux, D. Nathena, and M. Michalodimitrakis

2012 Sexual dimorphism of the scapula and the clavicle in a contemporary Greek population: applications in forensic identification. *Forensic Science International* 217(1-3):231.e1-e7.

Patriquin M. L., S. R. Loth, and M. Steyn

2003 Sexually dimorphic pelvic morphology in south African whites and blacks. *Homo* 53(3):255-262.

Phenice T.W.

1969 A newly developed method of sexing the os pubis. *American Journal of Physical Anthropology* 30:297-302.

Prescher, A., and T. Klümpen

1995 Does the area of the glenoid cavity of the scapula show sexual dimorphism? *Journal of Anatomy* 186: 223–22.

Prescher, A., and T. Klümpen
1997 The glenoid notch and its relation to the shape of the glenoid cavity of the scapula. *Journal of Anatomy* 190: 457-460.

Ramsthaler, F., K. Kreutz, and M. A. Verhoff
2007 Accuracy of metric sex analysis of skeletal remains using Fordisc based on a recent skull collection. *International Journal of Legal Medicine* 121(6):477-82.

Ramadan S. U., N. Türkmen, N. A. Dolgun, D. Gökharman, R. G. Menezes, M. Kacar, and U. Koşar
2010 Sex determination from measurements of the sternum and fourth rib using multislice computed tomography of the chest. *Forensic Science International* 197(1-3):120.e1-e5.

Robling A. G. and D. H. Ubelaker
1997 Sex estimation from the metatarsals. *Journal of Forensic Sciences* 42(6):1062-1069.

Rogers, T. L.
1999 A visual method of determining the sex of skeletal remains using the distal humerus. *Journal of Forensic Sciences* 44(1):57–60.

Rogers, T. L. and T. T. Allard
2004 Expert testimony and positive identification of human remains through cranial suture patterns. *Journal of Forensic Sciences* 49(2):203-7.

Safont, S., A. Malgosa, and M. E. Subira
2000 Sex assessment on the basis of long bone circumference. *American Journal of Physical Anthropology* 113(3):317-328.

Scheuer, J. L., and N. M. Elkington
1993 Sex determination from metacarpals and the first proximal phalanx. *Journal of Forensic Sciences* 38(4):769-78.

Smith, S. L.
1996 Attribution of hand bones to sex and population groups. *Journal of Forensic Sciences* 41(3):469-77.
1997 Attribution of foot bones to sex and population groups. *Journal of Forensic Sciences* 42(2):186-95.

Spradley, M. K. and L. R. Jantz
2011 Sex Estimation in Forensic Anthropology: Skull Versus Postcranial Elements. *Journal of Forensic Sciences* 56(2): 289–29.

Srivastava, R., V. Saini, R. K. Rai, S. Pandey, and S. K. Tripathi
2012 A study of sexual dimorphism in the femur among North Indians. *Journal of Forensic Sciences* 57(1):19-23

Steele, D. G.
1976 The estimation of sex on the basis of the talus and calcaneus. *American Journal of Physical Anthropology* 45(3 pt. 2):581-588.

Stewart, T. D.
1979 *Essentials of Forensic Anthropology*. Charles C. Thomas Publisher; Springfield, IL.

Steyn, M. and M.L.Patriquin
2009 Osteometric sex determination from the pelvis--does population specificity matter? *Forensic Science International* 191(1-3):113.e1-5.

Stojanowski, C. M.
1999 Sexing potential of fragmentary and pathological metacarpals. *American Journal of Physical Anthropology* 109(2):245-52.

Sulzmann, C. E., J. L. Buckberry, and R. F. Pastor
2008 The utility of carpals for sex assessment: a preliminary study. *American Journal of Physical Anthropology* 135(3):252-62.

Walker, P.
2005 Greater sciatic notch morphology: Sex, age, and population differences. *American Journal of Physical Anthropology* 127(4):385-391.

Washburn, S. L.
1948 Sex differences in the pubic bone. *American Journal of Physical Anthropology* 6: 199–208.

Wescott, D. J.
2000 Sex variation in the second cervical vertebra. *Journal of Forensic Sciences* 45(2):462-466.

Williams, F. L., R. L. Belcher, and G. J. Armelagos
2005 Forensic misclassification of Ancient Nubian crania: implications for assumptions about human variation**.** *Current Anthropology* 46, 340 – 346.

Wiredu, E. K., R. Kumoji, R. Seshadri, and R. B. Biritwum
1999 Osteometric analysis of sexual dimorphism in the sternal end of the rib in a West African population. *Journal of Forensic Sciences* 44(5):921-925.

Zanella, V. P., and T. M. Brown
2003 Testing the validity of metacarpal use in sex assessment of human skeletal remains. *Journal of Forensic Sciences* 48(1):17-20.

Appendix A: Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005)

Appendix A1: Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Athens individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
WLH	001	M	85	132.40	156.80	100.65	45.50	5.50	10.45	5.8563	M	1.4296	M	M	0.622	0.403
WLH	002	M	64	121.20	156.80	93.55	47.00	2.50	12.40	6.0194	M	-0.0756	F	F	0.703	0.150
WLH	006	M	62	148.54	162.00	112.90	42.50	7.00	7.15	4.49754	M	5.0718	M	M	0.986	0.796
WLH	008	M	60	142.36	152.00	108.30	41.00	9.50	8.40	3.54556	M	2.0866	M	M	0.749	0.397
WLH	009	M	76	151.20	158.20	111.60	42.50	7.50	9.45	6.6419	M	4.0324	M	M	0.958	0.669
WLH	010	F	68	127.65	139.10	94.10	35.00	10.50	5.80	-5.2647	F	-3.5167	F	F	0.992	0.951
WLH	011	M	82	141.98	158.34	106.40	42.00	6.50	9.45	4.83786	M	2.95814	M	M	0.895	0.820
WLH	012	F	84	123.60	138.32	92.55	37.50	4.50	8.75	-3.66346	F	-4.00208	F	F	0.995	0.840
WLH	013	F	49	126.70	132.16	95.60	36.00	8.00	8.35	-4.09968	F	-4.59364	F	F	0.998	0.538
WLH	014	M	65	143.40	162.30	109.55	44.00	7.50	10.50	7.8215	M	4.4219	M	M	0.976	0.975
WLH	016	M	36	137.86	154.52	101.50	39.00	5.50	10.70	3.0416	M	1.15152	M	M	0.553	0.407
WLH	020	M	67	132.74	161.66	101.70	41.00	8.00	8.85	3.33026	M	2.62906	M	M	0.870	0.529
WLH	021	M	76	139.10	167.56	110.30	41.50	7.00	9.95	5.85442	M	5.63816	M	M	0.994	0.907
WLH	022	M	94	137.54	149.84	104.80	40.50	7.50	9.40	2.67882	M	0.91044	M	F	0.601	0.390
WLH	023	M	48	132.08	155.30	101.50	43.50	10.50	11.15	6.66808	M	1.3083	M	M	0.611	0.442
WLH	024	M	87	144.64	154.26	113.30	46.00	8.50	9.90	7.88506	M	3.60086	M	M	0.927	0.294
WLH	027	M	65	133.54	148.00	96.30	42.00	6.50	11.95	4.74024	M	-1.2614	F	F	0.901	0.737
WLH	028	F	-	112.76	130.00	84.95	37.50	5.50	7.60	-6.85474	F	-7.2856	F	F	1.000	0.157
WLH	030	F	66	131.28	150.62	99.45	44.50	7.50	10.80	5.34832	M	-0.06698	F	F	0.758	0.604
WLH	031	M	60	129.70	134.88	97.00	40.00	8.00	9.90	0.19176	M	-3.75012	F	F	0.995	0.606
WLH	032	F	44	124.70	145.62	96.15	40.00	5.00	9.05	-0.88806	F	-1.77158	F	F	0.949	0.916

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A1 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Athens individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
WLH	033	F	72	121.88	134.52	91.85	38.50	4.00	6.65	-5.84738	F	-4.91428	F	F	0.999	0.633
WLH	034	F	81	137.56	138.24	98.90	40.00	9.00	6.45	-1.16896	F	-2.67196	F	F	0.986	0.804
WLH	036	F	63	128.82	123.62	95.20	34.50	10.00	6.95	-6.29564	F	-6.39498	F	F	1.000	0.090
WLH	037	F	44	115.38	144.36	88.60	37.50	4.50	9.55	-3.3515	F	-3.62544	F	F	0.991	0.390
WLH	038	M	43	143.40	156.00	103.15	46.00	9.50	8.70	7.1312	M	1.7988	M	M	0.747	0.599
WLH	043	M	55	130.96	153.40	96.70	42.00	3.50	9.50	1.91836	M	-0.0912	F	F	0.731	0.383
WLH	045	M	57	149.38	164.10	115.65	43.50	8.50	8.70	7.24368	M	6.0769	M	M	0.995	0.565
WLH	046	M	60	146.28	176.50	113.45	46.50	6.50	12.30	12.55828	M	8.1029	M	M	1.000	0.338
WLH	048	M	50	149.28	163.28	112.25	41.00	7.50	10.20	6.84164	M	5.19128	M	M	0.989	0.826
WLH	049	M	56	150.66	164.34	117.15	41.50	7.50	10.30	7.51424	M	6.44314	M	M	0.997	0.379
WLH	050	M	51	146.34	162.76	104.35	43.50	6.50	10.75	7.93466	M	3.41196	M	M	0.938	0.801
WLH	051	M	49	146.82	152.00	108.15	40.00	5.00	8.00	1.91752	M	2.0548	M	M	0.749	0.397
WLH	053	F	63	128.74	140.80	95.95	36.00	4.50	7.85	-4.29936	F	-2.7828	F	F	0.987	0.986
WLH	054	M	77	140.90	148.14	107.20	41.00	3.00	9.25	1.73978	M	1.07754	M	F	0.517	0.213
WLH	055	M	58	145.20	164.12	108.80	43.50	4.50	9.95	6.62354	M	4.62872	M	M	0.982	0.994
WLH	057	F	-	119.70	135.50	87.95	34.00	5.00	7.90	-7.0377	F	-5.5441	F	F	0.999	0.401
WLH	061	F	54	134.74	136.24	99.95	39.00	6.00	6.65	-3.03468	F	-2.85136	F	F	0.989	0.610
WLH	063	M	79	138.12	154.68	107.60	42.50	4.50	10.45	4.46718	M	2.47688	M	M	0.799	0.558
WLH	064	M	65	141.28	161.38	98.90	44.00	3.00	10.35	5.95794	M	1.97918	M	M	0.788	0.275
WLH	067	M	56	158.22	157.26	118.25	42.50	9.50	7.95	6.72264	M	5.25326	M	M	0.986	0.142

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A1 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Athens individuals.

				Measurements (mm)						Five-variable Model Results		Two-variable Model Results		FORDISC 3.0 Results		
															Probabilities (p-values)	
Sample^		**Sex**	**Age***	**XLS**	**XHS**	**XBS**	**HAX**	**CSV**	**TLB**	**Equation**	**Sex**	**Equation**	**Sex**	**Predicated Sex**	**Posterior**	**Typical**
WLH	069	M	80	142.08	166.34	110.45	46.00	9.50	10.05	9.38186	M	5.42474	M	M	0.992	0.940
WLH	070	M	48	140.46	154.52	104.75	37.50	3.00	10.10	1.3013	M	1.84052	M	M	0.689	0.537
WLH	072	M	27	142.70	165.28	104.85	40.50	5.00	11.30	6.16466	M	4.02448	M	M	0.969	0.729
ABH	073	M	62	152.24	164.50	116.45	43.50	10.50	8.20	7.81944	M	6.3269	M	M	0.996	0.469
ABH	077	F	54	141.20	147.28	106.85	36.00	5.50	7.70	-1.68624	F	0.83048	M	F	0.623	0.277
ABH	078	M	43	140.30	161.82	110.15	37.00	4.00	10.10	2.15914	M	4.45262	M	M	0.975	0.901
ABH	083	M	81	137.52	165.38	106.70	45.50	5.00	9.95	6.95648	M	4.43678	M	M	0.979	0.889
ABH	086	F	61	117.16	136.64	91.70	35.50	4.50	10.10	-4.59746	F	-4.51996	F	F	0.998	0.719
ABH	090	F	72	127.76	142.10	100.30	39.00	7.00	6.10	-3.49954	F	-1.5993	F	F	0.946	0.782
ABH	092	F	54	125.80	134.62	96.75	37.00	5.50	8.75	-3.77166	F	-3.85538	F	F	0.996	0.652
ABH	095	F	37	123.24	126.48	90.00	37.00	6.00	7.50	-6.0542	F	-6.92252	F	F	1.000	0.205
ABH	097	F	46	132.50	147.22	99.65	38.50	5.50	7.95	-1.29796	F	-0.70798	F	F	0.866	0.771
ABH	098	F	82	123.10	132.00	93.60	36.00	6.00	6.65	-6.7368	F	-5.0498	F	F	0.999	0.560
ABH	099	F	70	124.26	128.62	93.35	36.00	1.00	7.15	-8.0025	F	-5.78218	F	F	1.000	0.307
ABH	102	F	58	129.08	129.70	97.80	34.00	4.50	8.10	-6.40782	F	-4.6217	F	F	0.999	0.280
ABH	105	M	78	161.38	171.68	122.90	47.50	8.50	9.85	12.96614	M	9.13748	M	M	1.000	0.108
ABH	106	M	46	154.58	159.18	114.70	43.00	8.50	10.25	8.50594	M	4.88658	M	M	0.981	0.480
ABH	109	M	68	135.38	146.68	99.85	42.00	3.50	8.60	0.91964	M	-0.77412	F	F	0.891	0.813
ABH	114	F	80	123.10	146.88	88.35	42.00	6.50	7.70	-0.65264	F	-3.17192	F	F	0.986	0.318
ABH	119	F	85	125.88	127.54	94.25	39.50	5.50	9.30	-2.73944	F	-5.80846	F	F	1.000	0.245

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A1 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Athens individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						**Five-variable Model Results**		**Two-variable Model Results**			**Probabilities (p-values)**	
Sample^		**Sex**	**Age***	**XLS**	**XHS**	**XBS**	**HAX**	**CSV**	**TLB**	**Equation**	**Sex**	**Equation**	**Sex**	**Predicated Sex**	**Posterior**	**Typical**
ABH	121	F	27	118.10	135.80	92.40	36.00	2.50	8.55	-6.2906	F	-4.5404	F	F	0.998	0.726
ABH	122	F	78	125.60	138.12	96.80	36.00	3.00	11.50	-2.18436	F	-3.14128	F	F	0.990	0.907
ABH	123	F	71	129.38	140.60	93.20	37.50	5.50	9.05	-2.04342	F	-3.406	F	F	0.991	0.924
ABH	125	M	50	137.40	154.32	105.20	43.50	8.50	10.90	6.45894	M	1.89572	M	M	0.729	0.559
ABH	126	M	74	135.78	169.00	102.80	47.00	4.50	9.50	7.40008	M	4.3376	M	M	0.982	0.366
ABH	127	M	29	144.10	171.84	106.95	45.00	4.50	9.40	7.69148	M	5.78824	M	M	0.995	0.536
ABH	128	F	77	131.08	139.82	95.85	38.00	4.00	10.15	-1.08258	F	-3.00098	F	F	0.990	0.963
ABH	129	M	65	136.28	149.80	102.40	41.00	4.50	10.40	2.78928	M	0.3936	M	F	0.689	0.523
ABH	130	M	65	126.22	144.32	98.20	37.00	8.00	8.40	-2.15604	F	-1.59828	F	F	0.942	0.918
ABH	131	F	46	118.44	141.52	88.55	36.00	5.00	9.00	-4.42832	F	-4.20688	F	F	0.996	0.476
ABH	132	F	74	149.58	147.56	115.30	38.50	6.50	7.55	0.9991	M	2.67816	M	M	0.777	0.038
ABH	135	M	34	160.96	153.74	119.55	44.00	8.50	8.80	7.96734	M	4.82134	M	M	0.970	0.048
ABH	139	M	44	130.34	148.16	91.55	37.00	4.00	8.65	-2.10444	F	-2.23624	F	F	0.960	0.450
ABH	141	M	48	143.76	153.34	103.00	42.00	5.00	10.20	4.72894	M	1.23234	M	M	0.589	0.451
ABH	143	F	79	117.80	129.90	89.05	37.00	6.00	8.30	-5.6707	F	-6.4365	F	F	1.000	0.312
ABH	144	F	60	121.08	133.94	95.00	38.00	5.00	8.20	-4.59734	F	-4.36306	F	F	0.998	0.611
ABH	145	M	80	148.04	157.24	108.00	44.00	6.50	9.10	6.29892	M	3.07624	M	M	0.908	0.753
ABH	147	M	84	144.60	179.72	110.80	47.00	8.50	11.55	12.89104	M	8.18832	M	M	1.000	0.187
ABH	148	M	71	150.10	164.28	110.15	45.50	8.50	9.60	9.25826	M	4.94708	M	M	0.987	0.966
ABH	149	M	66	137.52	155.60	106.40	43.00	6.50	12.10	6.84732	M	2.4074	M	M	0.806	0.635

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A1 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Athens individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
ABH	151	F	63	139.38	149.54	100.85	39.50	7.00	7.10	0.12576	M	0.01274	M	F	0.766	0.630
ABH	155	M	74	162.08	174.00	116.10	45.00	9.50	9.00	11.50788	M	8.1622	M	M	1.000	0.338
ABH	156	M	59	139.42	156.48	106.40	42.50	7.50	11.30	6.51098	M	2.58428	M	M	0.841	0.702
ABH	157	M	64	138.54	162.62	101.75	42.50	6.00	9.50	5.03848	M	2.83262	M	M	0.895	0.518
ABH	158	M	78	144.02	155.26	109.95	42.00	11.00	8.50	5.22814	M	3.09166	M	M	0.882	0.573
ABH	162	F	69	126.68	137.00	97.65	38.00	5.50	6.90	-4.50492	F	-3.1862	F	F	0.991	0.809
ABH	168	M	66	140.68	165.08	104.70	37.50	6.50	8.05	1.74954	M	3.95248	M	M	0.969	0.729
ABH	169	F	84	128.64	150.74	97.35	38.00	8.00	7.05	-1.75518	F	-0.48806	F	F	0.822	0.613
ABH	172	M	88	153.52	179.02	115.85	46.00	8.50	8.85	11.04046	M	9.11822	M	M	1.000	0.181
ABH	173	M	58	151.28	154.22	114.05	42.00	9.00	10.95	7.71662	M	3.75182	M	M	0.939	0.240
ABH	176	F	81	121.62	137.36	93.70	38.00	5.50	7.10	-4.97016	F	-3.95124	F	F	0.996	0.856
ABH	177	M	65	144.32	165.44	102.90	42.50	14.00	8.00	7.1665	M	3.64324	M	M	0.955	0.538
ABH	182	M	55	141.60	159.90	103.50	44.00	5.50	9.55	5.8451	M	2.6569	M	M	0.859	0.698
ABH	183	F	59	125.18	132.96	92.60	39.00	4.00	6.70	-5.2654	F	-5.06884	F	F	0.999	0.550
ABH	184	F	72	112.52	122.46	84.40	38.00	4.50	7.80	-7.61426	F	-8.91774	F	F	1.000	0.048
ABH	185	F	72	126.32	151.08	95.40	37.50	2.50	7.80	-3.25142	F	-0.83312	F	F	0.842	0.487
ABH	186	M	26	149.00	158.48	111.30	41.00	6.50	9.15	4.99676	M	4.02508	M	M	0.958	0.669
ABH	188	F	38	115.34	119.84	83.70	35.00	3.50	6.30	-10.8123	F	-9.59276	F	F	1.000	0.021
ABH	191	M	32	149.72	163.48	112.50	41.50	6.00	9.00	5.66658	M	5.28448	M	M	0.989	0.826
ABH	192	M	27	147.28	173.18	113.25	46.00	3.50	9.90	8.97774	M	7.39318	M	M	0.999	0.505

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A1 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Athens individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
ABH	193	M	25	139.22	152.88	104.15	46.00	4.50	8.40	4.44648	M	1.38368	M	M	0.578	0.420
ABH	198	F	45	127.70	138.84	95.60	38.00	5.00	8.15	-3.16892	F	-3.25096	F	F	0.992	0.926
ABH	199	M	26	159.52	174.82	123.90	47.50	4.50	12.60	14.38256	M	9.98062	M	M	1.000	0.070
ABH	200	M	43	145.12	173.14	108.70	43.50	5.00	10.75	8.5392	M	6.42054	M	M	0.998	0.490
ABH	205	F	47	126.38	131.56	95.20	34.50	6.00	8.30	-5.6621	F	-4.79904	F	F	0.998	0.467
ABH	209	M	44	138.14	168.78	106.05	45.00	6.50	10.95	8.5161	M	4.98238	M	M	0.990	0.738
ABH	210	M	43	131.40	158.60	100.40	41.00	5.50	11.20	4.1744	M	1.7384	M	M	0.727	0.435
ABH	213	M	32	139.78	184.90	105.30	42.00	5.50	9.55	7.44058	M	8.0635	M	M	1.000	0.024
ABH	217	F	22	131.32	145.00	100.40	38.50	3.50	8.75	-1.58698	F	-0.9952	F	F	0.896	0.779
ABH	223	F	38	130.80	132.34	96.20	35.00	6.00	7.56	-5.36818	F	-4.43026	F	F	0.998	0.507
ABH	225	F	-	128.04	149.38	95.00	39.00	6.00	8.60	-0.6457	F	-1.25962	F	F	0.896	0.632
ABH	228	M	64	136.48	153.92	106.30	43.00	4.50	8.60	2.75332	M	2.04852	M	M	0.720	0.497
ABH	231	M	33	136.78	159.52	105.55	41.00	7.00	9.15	3.60452	M	3.01512	M	M	0.900	0.829
ABH	232	F	33	128.94	150.96	97.35	35.50	3.50	9.35	-2.29394	F	-0.44384	F	F	0.822	0.613
ABH	233	F	85	122.94	131.62	91.90	39.00	5.50	7.30	-4.74042	F	-5.48658	F	F	0.999	0.468
ABH	234	F	50	120.14	128.90	95.10	36.50	5.50	7.70	-6.43036	F	-5.3549	F	F	0.999	0.280
ABH	235	M	43	134.40	163.00	99.00	39.50	6.00	10.35	3.6653	M	2.326	M	M	0.880	0.323
ABH	236	M	59	133.06	145.88	97.00	39.00	4.00	8.00	-1.50688	F	-1.53912	F	F	0.939	0.915
ABH	237	M	48	137.88	162.20	99.45	45.00	3.00	8.85	4.77648	M	2.2606	M	F	0.852	0.339
ABH	239	M	37	140.56	142.04	103.90	37.00	9.00	9.70	0.99864	M	-0.84816	F	F	0.908	0.494

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A1 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Athens individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
ABH	240	F	58	115.20	149.06	85.40	36.00	4.00	9.35	-3.96638	F	-3.35914	F	F	0.984	0.088
ABH	241	M	75	136.14	160.84	102.80	43.50	8.50	8.95	5.28762	M	2.69744	M	M	0.864	0.624
ABH	244	F	79	132.64	132.44	104.15	37.00	8.50	6.85	-3.92608	F	-2.72476	F	F	0.989	0.122
ABH	246	M	30	126.78	144.30	92.40	36.00	0.50	7.15	-5.97322	F	-2.8319	F	F	0.981	0.755
ABH	247	M	68	148.26	170.92	108.55	46.00	5.50	8.85	8.4894	M	5.94252	M	M	0.996	0.708
ABH	250	M	38	141.58	165.48	103.85	43.00	7.00	11.90	8.52264	M	3.85268	M	M	0.963	0.635

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A2: Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Tennessee individuals.

				Measurements (mm)						Five-variable Model Results		Two-variable Model Results		FORDISC 3.0 Results		
															Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
UT	01-03	M	47	138.04	155.46	102.30	39.50	6.00	9.15	2.19336	M	1.51006	M	M	0.656	0.501
UT	01-05	M	44	150.52	163.72	112.55	44.00	5.00	9.90	7.67356	M	5.34332	M	M	0.989	0.826
UT	01-87	M	39	145.66	163.80	105.40	44.00	3.00	9.65	6.20396	M	3.8436	M	M	0.959	0.870
UT	01-95	M	42	146.88	172.28	108.70	45.00	5.00	10.55	9.30864	M	6.24768	M	M	0.997	0.562
UT	02-02	M	46	146.88	159.84	110.00	40.00	6.00	9.45	4.44976	M	4.02284	M	M	0.960	0.798
UT	02-08	F	65	121.16	143.34	98.45	38.00	7.50	6.95	-3.87666	F	-1.74226	F	F	0.953	0.931
UT	02-89	M	36	144.58	162.96	106.35	46.00	7.00	10.45	8.948	M	3.87616	M	M	0.957	0.939
UT	03-00	M	43	149.64	163.30	113.90	43.00	11.50	7.95	7.12494	M	5.5451	M	M	0.991	0.737
UT	03-06	F	52	131.10	135.60	96.25	36.50	4.00	7.70	-4.5999	F	-3.7644	F	F	0.995	0.724
UT	04-00	M	56	144.58	175.44	110.00	47.50	3.50	10.85	10.54526	M	7.15844	M	M	0.999	0.401
UT	04-02	F	60	131.06	149.90	98.55	36.00	8.00	6.50	-3.10354	F	-0.4025	F	F	0.828	0.683
UT	04-06	F	58	120.12	130.06	92.90	34.00	2.50	8.70	-7.63386	F	-5.58814	F	F	0.999	0.425
UT	04-96	M	55	158.98	159.72	114.55	46.00	3.00	9.50	8.48452	M	4.96332	M	M	0.981	0.480
UT	04-97	M	33	142.34	154.42	103.70	41.50	7.00	9.85	4.66728	M	1.59782	M	M	0.939	0.240
UT	04-99	M	57	141.38	164.52	105.00	43.50	6.00	12.35	8.76442	M	3.90352	M	M	0.968	0.849
UT	05-99	M	38	159.32	162.04	110.90	41.00	4.50	9.40	6.4508	M	4.65584	M	M	0.980	0.936
UT	06-08	F	61	127.56	154.02	98.70	37.50	2.50	8.85	-1.7896	F	0.45742	M	F	0.592	0.353
UT	07-00	M	38	137.74	146.60	101.30	38.50	10.00	7.55	0.31254	M	-0.4828	F	F	0.847	0.683
UT	07-01	F	50	120.90	138.70	94.35	39.00	2.00	8.85	-3.8243	F	-3.5441	F	F	0.993	0.920
UT	07-02	M	59	148.74	149.38	106.60	41.00	5.00	11.50	5.5771	M	1.19958	M	F	0.505	0.262
UT	07-05	M	40	136.34	160.50	94.50	42.50	2.50	9.30	3.27444	M	0.8695	M	M	0.574	0.084

^UT = University of Tennessee; *Age in years

Appendix A2 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Tennessee individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
UT	07-94	M	41	143.42	169.82	103.95	40.50	4.50	9.65	5.15416	M	4.74622	M	M	0.985	0.444
UT	08-03	M	51	146.08	171.58	110.00	44.00	8.00	11.60	10.41414	M	6.38258	M	M	0.998	0.678
UT	08-04	M	57	144.04	160.92	108.15	40.00	5.00	8.20	2.76388	M	3.84772	M	M	0.953	0.927
UT	08-07	F	57	116.58	138.22	87.00	37.00	3.00	7.10	-6.83598	F	-5.19878	F	F	0.998	0.420
UT	08-87	M	25	134.94	161.90	105.75	44.00	3.50	9.45	4.49094	M	3.5359	M	M	0.936	0.875
UT	08-93	M	52	141.18	161.72	104.30	41.50	6.50	9.55	4.94442	M	3.19232	M	M	0.923	0.803
UT	08-98	M	36	130.06	147.60	95.35	41.00	5.50	7.60	-0.54924	F	-1.5432	F	F	0.933	0.774
UT	09-00	F	43	123.60	140.86	93.70	36.00	3.50	6.95	-6.10098	F	-3.24774	F	F	0.991	0.924
UT	09-03	M	51	145.62	168.48	103.65	43.50	5.50	12.75	10.01798	M	4.41328	M	M	0.982	0.496
UT	09-93	M	42	146.96	166.00	113.40	44.00	13.00	9.95	9.86936	M	5.9818	M	M	0.996	0.768
UT	09-97	M	56	135.30	168.84	102.40	39.50	3.00	9.55	2.85978	M	4.22064	M	M	0.978	0.412
UT	09-99	F	54	132.76	151.92	99.30	37.00	6.00	8.10	-1.2406	F	0.16252	M	F	0.711	0.540
UT	100-06	F	57	140.04	148.92	106.10	38.00	4.50	7.40	-1.13732	F	1.00112	M	F	0.565	0.272
UT	10-03	M	49	144.44	165.32	108.65	45.00	5.50	9.35	7.22568	M	4.83812	M	M	0.986	0.975
UT	10-05	M	46	149.76	159.12	112.35	42.00	5.50	11.45	7.4992	M	4.37632	M	M	0.972	0.653
UT	10-07	F	50	124.02	138.02	98.75	37.00	2.50	7.85	-5.31754	F	-2.74798	F	F	0.986	0.804
UT	101-06	F	60	134.02	138.34	100.15	42.00	6.00	7.85	-0.1791	F	-2.38686	F	F	0.979	0.641
UT	107-06	F	54	120.74	145.04	87.80	36.00	2.50	8.85	-4.57928	F	-3.65836	F	F	0.991	0.280
UT	107-08	F	52	120.84	134.10	88.05	37.00	6.50	7.65	-5.20546	F	-5.8043	F	F	0.999	0.452
UT	109-07	F	48	134.94	147.38	105.40	39.00	4.50	7.55	-1.3345	F	0.54318	M	F	0.667	0.347
UT	10-91	M	35	143.78	150.06	102.55	45.00	6.50	12.05	8.0873	M	0.47766	M	F	0.636	0.484

^UT = University of Tennessee; *Age in years

Appendix A2 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Tennessee individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
UT	11-00	M	55	147.18	154.76	113.00	43.00	7.50	10.00	6.4604	M	3.63776	M	M	0.927	0.294
UT	11-03	F	47	125.50	144.62	92.70	37.00	5.50	8.10	-3.23006	F	-2.70398	F	F	0.981	0.755
UT	11-04	F	54	120.78	134.80	91.70	37.50	1.50	8.60	-5.48242	F	-4.8898	F	F	0.999	0.633
UT	111-07	F	50	132.64	148.74	96.45	37.00	0.50	6.65	-4.56778	F	-1.08086	F	F	0.901	0.737
UT	112-07	F	64	128.80	141.36	92.45	37.00	4.50	9.50	-2.19308	F	-3.41224	F	F	0.991	0.851
UT	12-01	M	50	140.04	169.58	105.35	45.00	2.50	9.95	6.7801	M	4.99478	M	M	0.990	0.601
UT	12-04	F	60	132.94	138.58	100.40	34.50	5.00	7.90	-4.6062	F	-2.28562	F	F	0.979	0.641
UT	12-98	M	46	135.38	160.08	102.65	41.00	1.00	12.90	5.09364	M	2.51288	M	M	0.864	0.624
UT	13-00	M	44	140.70	168.24	105.95	42.00	1.00	10.50	5.14328	M	4.85264	M	M	0.987	0.664
UT	13-03	M	48	137.66	146.28	100.90	41.50	3.50	12.85	4.75442	M	-0.63192	F	F	0.871	0.755
UT	13-88	M	31	152.98	171.98	110.70	46.00	4.50	8.75	8.86894	M	6.61138	M	M	0.998	0.678
UT	13-91	M	34	150.04	177.30	111.40	41.00	4.50	11.75	9.09854	M	7.8291	M	M	1.000	0.291
UT	14-03	M	50	138.38	165.10	109.65	38.00	1.50	10.95	2.85118	M	5.0059	M	M	0.988	0.981
UT	14-04	M	19	143.74	160.92	107.10	42.00	3.00	11.35	6.06068	M	3.62512	M	M	0.944	0.928
UT	14-90	M	37	141.34	148.80	109.30	40.00	4.50	7.15	-0.11856	F	1.6554	M	M	0.580	0.164
UT	14-93	M	32	126.60	158.70	92.75	42.00	3.00	10.15	2.3851	M	0.1367	M	F	0.640	0.076
UT	15-06	F	59	124.28	146.24	95.70	40.50	5.00	9.75	0.03066	M	-1.74236	F	F	0.947	0.837
UT	16-98	M	58	151.92	177.12	117.55	47.50	9.00	10.90	13.41326	M	9.09672	M	M	1.000	0.203
UT	17-01	M	51	158.68	182.40	122.45	43.00	2.00	11.30	10.80548	M	11.1968	M	M	1.000	0.035
UT	17-02	F	50	127.60	139.82	97.15	36.50	3.50	10.00	-2.65096	F	-2.72538	F	F	0.985	0.912
UT	17-05	F	58	127.56	134.16	94.65	42.00	2.50	8.05	-2.40192	F	-4.39304	F	F	0.997	0.698

^UT = University of Tennessee; *Age in years

Appendix A2 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Tennessee individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						**Five-variable Model Results**		**Two-variable Model Results**			**Probabilities (p-values)**	
Sample^		**Sex**	**Age***	**XLS**	**XHS**	**XBS**	**HAX**	**CSV**	**TLB**	**Equation**	**Sex**	**Equation**	**Sex**	**Predicated Sex**	**Posterior**	**Typical**
UT	17-06	F	50	134.14	148.52	99.75	37.00	5.50	7.95	-1.73432	F	-0.42548	F	F	0.835	0.721
UT	17-99	M	56	133.64	152.06	100.45	40.00	8.50	10.05	3.02126	M	0.43446	M	F	0.614	0.461
UT	18-03	F	47	121.80	141.86	94.50	35.00	3.50	6.30	-7.35738	F	-2.87714	F	F	0.986	0.970
UT	18-04	F	44	128.64	153.22	97.70	37.50	2.00	11.00	0.05928	M	0.08462	M	F	0.691	0.404
UT	18-99	M	55	148.42	174.26	109.30	42.00	6.50	10.65	8.66114	M	6.77286	M	M	0.999	0.449
UT	19-03	M	55	147.94	167.72	114.10	44.00	8.00	8.75	7.63908	M	6.47592	M	M	0.997	0.675
UT	19-04	F	60	119.58	142.90	89.45	40.00	1.50	7.45	-4.38502	F	-3.7387	F	F	0.993	0.547
UT	19-92	M	27	135.06	151.50	101.30	42.00	4.50	9.75	2.77566	M	0.5021	M	F	0.625	0.485
UT	20-03	F	44	122.06	137.96	91.55	35.00	3.00	7.55	-6.79632	F	-4.28644	F	F	0.997	0.750
UT	20-08	F	62	130.90	139.86	100.50	40.50	6.00	6.30	-2.63828	F	-2.00714	F	F	0.973	0.688
UT	21-08	F	65	123.40	127.02	94.70	33.50	9.00	8.40	-6.16116	F	-5.81758	F	F	1.000	0.245
UT	21-95	M	50	147.62	166.40	109.55	43.00	3.00	9.45	6.05292	M	5.246	M	M	0.991	0.954
UT	21-98	M	52	146.88	149.94	110.30	42.00	4.50	9.45	3.92946	M	2.09654	M	M	0.681	0.181
UT	22-00	M	57	143.54	151.68	106.95	41.50	6.50	7.15	1.9211	M	1.73608	M	M	0.613	0.367
UT	22-01	M	47	151.90	170.80	109.20	42.00	7.50	9.25	7.76	M	6.0562	M	M	0.996	0.737
UT	22-02	M	50	140.78	153.36	105.40	45.00	5.50	9.00	5.0122	M	1.74516	M	M	0.679	0.494
UT	22-03	M	48	141.96	163.64	103.65	44.00	4.00	10.40	6.65604	M	3.44044	M	M	0.941	0.695
UT	22-08	F	50	128.16	142.52	96.15	37.00	1.00	7.10	-5.35	F	-2.39468	F	F	0.975	0.997
UT	22-95	M	41	150.80	171.50	115.70	43.00	4.00	10.40	8.2369	M	7.5749	M	M	0.999	0.501
UT	24-02	M	52	147.36	160.30	102.80	43.00	6.50	8.40	5.39066	M	2.5889	M	M	0.864	0.624
UT	24-99	M	49	150.80	158.64	110.15	46.00	5.50	11.70	9.97448	M	3.81344	M	M	0.949	0.735

^UT = University of Tennessee; *Age in years

Appendix A2 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Tennessee individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		**Sex**	**Age***	**XLS**	**XHS**	**XBS**	**HAX**	**CSV**	**TLB**	**Equation**	**Sex**	**Equation**	**Sex**	**Predicated Sex**	**Posterior**	**Typical**
UT	25-01	M	48	145.98	146.70	107.90	40.00	6.50	8.60	2.16958	M	0.9365	M	F	0.634	0.216
UT	25-02	M	42	149.78	169.68	109.90	44.00	7.50	9.80	8.91984	M	5.97948	M	M	0.995	0.803
UT	25-05	F	51	120.96	145.52	91.55	36.00	1.50	8.75	-4.8796	F	-2.76688	F	F	0.980	0.617
UT	25-06	F	44	132.26	153.34	99.80	37.00	5.50	5.55	-3.59566	F	0.55394	M	F	0.603	0.414
UT	26-01	M	54	144.00	167.54	108.20	43.50	6.00	11.00	8.25368	M	5.18894	M	M	0.991	0.897
UT	26-03	M	49	138.44	167.72	103.80	39.00	5.00	11.45	5.19488	M	4.29232	M	M	0.977	0.546
UT	27-03	M	46	133.94	168.64	97.95	44.50	5.50	14.50	10.57122	M	3.23704	M	M	0.943	0.110
UT	27-07	F	45	130.04	134.46	93.50	39.00	4.50	7.40	-3.64814	F	-4.57654	F	F	0.998	0.694
UT	27-93	M	39	158.42	180.70	116.90	46.50	4.00	9.70	11.61032	M	9.6785	M	M	1.000	0.137
UT	28-90	F	45	128.10	149.38	92.90	38.00	2.00	8.20	-2.72414	F	-1.70482	F	F	0.939	0.461
UT	29-00	M	39	151.42	167.88	112.35	41.50	4.00	10.00	6.72458	M	6.13708	M	M	0.996	0.825
UT	29-03	F	59	127.40	132.56	95.40	35.50	4.00	5.95	-7.58178	F	-4.55564	F	F	0.998	0.538
UT	29-04	M	34	152.04	178.86	116.10	43.00	7.00	10.50	10.24506	M	9.13906	M	M	1.000	0.197
UT	29-93	M	56	152.78	162.46	119.80	42.50	4.50	7.45	4.6592	M	6.62706	M	M	0.996	0.196
UT	30-02	M	61	134.56	146.62	100.95	41.00	8.00	9.60	2.4961	M	-0.55298	F	F	0.871	0.755
UT	30-04	M	59	157.94	186.62	117.80	45.00	6.00	13.85	15.76978	M	11.05922	M	M	1.000	0.031
UT	30-93	M	46	153.74	176.88	117.55	44.50	3.00	11.45	10.7309	M	9.04848	M	M	1.000	0.232
UT	31-05	F	51	122.38	135.54	91.70	39.00	6.00	8.30	-3.30594	F	-4.74106	F	F	0.998	0.680
UT	31-07	F	67	118.90	139.66	90.65	35.00	3.50	9.55	-5.07118	F	-4.13554	F	F	0.985	0.912
UT	33-02	M	39	141.08	162.60	107.30	40.00	3.00	9.55	3.18628	M	4.0052	M	M	0.965	0.976
UT	33-03	F	52	127.42	136.10	96.85	38.50	2.00	7.20	-5.00388	F	-3.5367	F	F	0.994	0.792

^UT = University of Tennessee; *Age in years

Appendix A2 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Tennessee individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
UT	34-02	M	58	152.76	161.84	110.70	41.50	4.00	8.40	4.75374	M	4.57324	M	M	0.975	0.901
UT	34-93	M	51	143.82	164.84	109.95	42.50	6.00	9.10	5.6547	M	5.01724	M	M	0.985	0.993
UT	35-02	F	55	139.38	135.96	107.50	38.00	8.50	6.95	-1.9662	F	-1.30704	F	F	0.961	0.082
UT	35-03	M	62	138.82	151.32	103.35	39.00	1.50	8.80	-0.10384	F	0.90052	M	F	0.530	0.392
UT	35-07	F	46	127.54	139.82	97.05	35.00	7.50	7.50	-4.54662	F	-2.74658	F	F	0.985	0.912
UT	36-01	M	49	144.54	154.60	108.55	40.00	7.50	10.10	4.55004	M	2.6622	M	M	0.828	0.536
UT	36-05	M	41	145.66	170.18	109.85	43.00	7.50	14.50	12.12582	M	6.06938	M	M	0.996	0.737
UT	37-03	M	43	135.40	146.78	100.50	40.00	5.50	9.85	1.57406	M	-0.61622	F	F	0.871	0.755
UT	37-07	F	57	120.48	139.58	89.00	37.00	8.00	8.30	-3.58166	F	-4.50142	F	F	0.997	0.599
UT	38-04	M	54	146.34	169.28	117.30	49.00	0.00	10.70	9.7038	M	7.46788	M	M	0.999	0.400
UT	38-05	M	35	145.30	151.36	109.30	39.00	3.50	10.25	2.68492	M	2.16996	M	M	0.740	0.306
UT	39-01	F	36	126.88	148.64	96.60	38.50	0.00	8.50	-3.02694	F	-1.06916	F	F	0.901	0.737
UT	39-03	F	52	133.34	139.14	101.55	36.00	2.00	7.35	-5.05998	F	-1.92926	F	F	0.968	0.595
UT	39-04	M	60	141.72	154.86	109.05	43.00	5.50	9.05	4.27874	M	2.82046	M	M	0.854	0.501
UT	40.08	F	33	119.54	132.02	92.02	35.00	2.00	6.80	-8.80802	F	-5.38074	F	F	0.999	0.550
UT	40-03	M	60	151.08	171.64	114.15	46.00	4.00	9.00	8.64876	M	7.27444	M	M	0.999	0.561
UT	40-06	F	51	140.48	152.06	105.00	41.00	1.00	9.90	2.1369	M	1.39906	M	M	0.624	0.431
UT	41-01	F	58	123.68	133.42	91.95	38.00	0.50	8.60	-5.27498	F	-5.11418	F	F	0.999	0.581
UT	41-07	F	37	123.24	135.06	94.65	36.00	6.00	7.00	-6.04334	F	-4.21214	F	F	0.997	0.765
UT	42-05	M	42	137.26	164.50	101.50	42.00	8.50	13.20	8.89866	M	3.1575	M	M	0.932	0.474
UT	42-06	F	56	126.40	143.80	97.45	36.50	2.50	9.40	-3.1869	F	-1.8618	F	F	0.961	0.970

^UT = University of Tennessee; *Age in years

Appendix A2 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Tennessee individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
UT	43-01	M	59	144.42	159.30	110.10	42.00	5.50	9.15	4.71482	M	3.9355	M	M	0.960	0.798
UT	44-02	M	60	142.26	170.40	105.10	46.50	5.50	7.95	7.06946	M	5.1066	M	M	0.992	0.537
UT	44-03	M	46	137.12	163.84	100.25	44.00	5.00	12.50	8.2156	M	2.75984	M	M	0.899	0.407
UT	44-04	M	39	142.96	162.10	106.95	44.00	7.00	11.90	8.85586	M	3.8305	M	M	0.957	0.939
UT	44-05	M	51	150.14	186.60	112.40	40.00	6.50	11.60	10.11564	M	9.9104	M	M	1.000	0.033
UT	45-04	M	54	163.08	178.64	121.15	48.00	5.00	12.70	15.82256	M	10.16544	M	M	1.000	0.083
UT	50-03	M	62	144.18	161.70	109.20	39.00	9.00	11.70	6.68118	M	4.2271	M	M	0.969	0.944
UT	50-04	M	51	136.18	164.56	105.95	43.50	2.50	9.65	4.5851	M	4.11296	M	M	0.968	0.849
UT	52-03	M	55	147.66	164.84	115.70	41.50	9.00	10.30	7.60874	M	6.23624	M	M	0.995	0.565
UT	53-06	F	54	115.10	130.16	87.15	36.00	3.00	8.10	-7.61728	F	-6.78704	F	F	1.000	0.277
UT	54-03	M	54	145.20	165.20	113.75	45.00	5.50	12.15	9.8462	M	5.8952	M	M	0.994	0.769
UT	54-05	F	54	129.08	138.82	92.95	34.00	1.50	8.05	-6.27398	F	-3.81678	F	F	0.995	0.840
UT	55-06	F	66	127.94	139.88	98.90	37.00	2.00	7.60	-4.9408	F	-2.34232	F	F	0.982	0.853
UT	55-07	F	51	126.18	131.14	97.65	36.50	4.50	7.45	-5.86884	F	-4.36406	F	F	0.998	0.398
UT	56-07	F	57	122.26	138.48	95.40	38.00	2.50	6.30	-6.36328	F	-3.36572	F	F	0.992	0.926
UT	56-08	F	57	139.98	148.24	105.85	35.50	4.50	9.35	-0.81474	F	0.81144	M	F	0.612	0.338
UT	57-03	M	55	143.06	165.14	104.55	45.00	4.50	11.75	8.89054	M	3.93274	M	M	0.969	0.729
UT	57-05	F	60	122.26	144.60	90.50	39.00	1.00	9.00	-3.10944	F	-3.1744	F	F	0.987	0.572
UT	58-06	F	51	127.18	132.80	93.15	37.00	5.00	7.30	-5.25572	F	-4.9844	F	F	0.999	0.560
UT	59-04	M	48	131.60	162.42	98.75	41.00	5.50	9.75	3.33774	M	2.15642	M	M	0.826	0.264
UT	59-05	M	53	138.82	163.84	104.80	42.00	2.00	9.40	3.6744	M	3.72444	M	M	0.951	0.788
UT	61-04	M	50	145.12	157.40	109.20	39.00	7.00	10.50	4.62912	M	3.3628	M	M	0.923	0.718

^UT = University of Tennessee; *Age in years

Appendix A2 (continued): Results of the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) on Tennessee individuals.

														FORDISC 3.0 Results		
				Measurements (mm)						Five-variable Model Results		Two-variable Model Results			Probabilities (p-values)	
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	Equation	Sex	Equation	Sex	Predicated Sex	Posterior	Typical
UT	61-05	F	55	127.50	145.94	93.05	38.50	5.00	8.50	-1.77852	F	-2.36446	F	F	0.971	0.783
UT	61-08	F	48	124.28	146.98	92.90	36.00	3.50	8.45	-3.93646	F	-2.18722	F	F	0.970	0.647
UT	63-03	F	58	140.28	144.62	108.20	36.00	3.00	9.05	-1.64218	F	0.58202	M	F	0.697	0.153
UT	63-04	M	61	156.48	151.34	110.70	42.50	5.00	9.55	5.90776	M	2.46274	M	M	0.775	0.273
UT	64-04	M	53	136.34	156.48	102.35	44.00	2.50	9.40	3.706	M	1.72568	M	M	0.709	0.539
UT	68-06	F	54	128.44	140.16	92.75	37.00	6.50	6.00	-4.95444	F	-3.58984	F	F	0.993	0.860
UT	68-07	F	42	132.56	154.62	97.30	35.50	3.50	11.70	0.751	M	0.28122	M	F	0.638	0.341
UT	69-04	F	62	128.14	129.48	97.85	37.00	4.50	7.25	-5.7068	F	-4.65532	F	F	0.999	0.280
UT	69-06	F	45	129.48	148.78	98.85	37.00	5.50	7.75	-2.51846	F	-0.56402	F	F	0.860	0.746
UT	70-06	M	42	143.54	158.24	106.05	41.00	5.00	10.70	5.18332	M	2.86384	M	M	0.895	0.820
UT	72-08	F	55	128.50	144.46	94.80	40.00	5.00	8.65	-0.86858	F	-2.29094	F	F	0.972	0.896
UT	73-08	F	60	136.46	150.26	102.35	36.50	8.50	6.75	-1.68252	F	0.47546	M	F	0.636	0.484
UT	74-06	F	42	132.16	143.74	98.95	36.50	3.50	6.85	-4.41976	F	-1.55586	F	F	0.953	0.931
UT	77-07	F	36	118.10	123.24	88.45	35.00	2.50	7.80	-8.97912	F	-7.90236	F	F	1.000	0.099
UT	78-06	F	49	137.06	149.28	101.55	39.00	5.50	7.75	-0.34708	F	0.10888	M	F	0.729	0.581
UT	79-05	F	59	121.58	140.58	88.60	33.00	1.50	7.65	-7.99066	F	-4.38522	F	F	0.997	0.494
UT	82-07	F	31	118.12	136.14	86.20	34.00	3.50	10.20	-5.5425	F	-5.78646	F	F	0.999	0.337
UT	84-08	F	46	118.04	144.12	88.75	36.00	4.50	8.80	-4.50732	F	-3.64188	F	F	0.991	0.390
UT	85-08	F	47	129.02	137.12	97.80	38.00	5.50	8.25	-2.95224	F	-3.13028	F	F	0.991	0.809
UT	92-05	F	47	143.72	154.92	103.60	39.00	4.50	5.70	-0.93064	F	1.67712	M	M	0.646	0.503
UT	97-07	F	66	132.04	126.06	100.05	36.00	6.50	6.45	-6.24874	F	-4.87634	F	F	0.999	0.087
UT	98-07	F	66	124.96	152.88	93.60	36.00	2.00	6.15	-5.67688	F	-0.85292	F	F	0.861	0.335

^UT = University of Tennessee; *Age in years

Appendix A3: Inter- and Intra-observer measurements for the Athens population for the five- and two-variable models by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005)

				Inter-observer measurements (mm)						**Intra-observer measurements (mm)**					
Sample^		**Sex**	**Age***	**XLS**	**XHS**	**XBS**	**HAX**	**CSV**	**TLB**	**XLS**	**XHS**	**XBS**	**HAX**	**CSV**	**TLB**
WLH	011	M	82	142.16	158.48	106.00	43.00	7.50	9.20	142.48	158.70	106.10	41.50	7.50	9.75
WLH	020	M	67	133.12	161.68	101.20	40.00	7.50	8.65	134.62	161.60	101.40	40.00	9.50	8.25
WLH	022	M	94	137.66	149.76	104.10	40.50	8.00	9.60	137.52	149.42	104.00	40.00	8.25	9.60
WLH	023	M	48	133.52	154.88	100.25	43.00	9.00	11.25	133.72	155.08	100.45	42.50	10.25	11.00
WLH	069	M	80	143.16	165.76	110.05	45.00	10.00	10.25	144.02	166.98	111.15	45.00	8.00	9.85
WLH	072	M	27	143.08	165.28	104.20	39.00	4.50	11.40	142.70	165.30	104.15	41.50	4.50	11.20
ABH	073	M	62	158.34	163.86	116.30	43.00	8.50	8.15	159.56	163.30	116.60	45.00	7.00	7.90
ABH	077	F	54	144.80	147.26	107.35	36.50	8.00	7.45	146.18	147.56	101.90	39.00	6.50	7.70
ABH	086	F	61	116.22	136.44	91.25	35.00	4.00	10.50	116.88	136.50	91.90	36.00	3.75	9.60
ABH	090	F	72	126.58	142.26	98.55	39.50	6.00	6.50	127.20	142.28	98.55	38.00	6.00	5.25
ABH	092	F	54	124.78	134.58	96.10	36.00	3.50	8.90	125.00	134.52	95.50	37.00	2.75	7.85
ABH	097	F	46	131.64	147.12	98.25	38.00	5.50	7.25	132.96	147.30	98.55	39.00	5.75	7.85
ABH	098	F	82	122.56	131.84	90.75	35.00	5.00	6.10	123.06	131.58	90.70	36.00	4.00	6.00
ABH	102	F	58	129.00	129.50	97.00	35.00	5.00	6.80	129.50	130.00	95.90	36.00	2.75	5.60
ABH	106	M	46	153.92	159.48	114.50	42.50	6.50	11.10	154.46	159.72	114.20	43.00	6.25	10.95
ABH	114	F	80	121.50	145.22	88.20	41.00	6.50	7.30	122.40	147.72	88.70	42.00	4.75	6.80
ABH	121	F	27	117.50	135.70	93.00	35.50	2.00	8.70	119.04	135.86	92.15	36.50	2.00	8.65
ABH	122	F	78	124.80	138.38	96.50	37.00	4.00	10.80	124.86	138.36	96.50	37.00	2.25	9.75
ABH	130	M	65	126.14	144.24	98.10	37.50	7.50	8.75	126.64	144.36	97.60	39.00	7.75	8.90
ABH	131	F	46	117.88	141.62	88.20	36.50	4.50	8.65	118.28	141.46	88.40	36.50	5.50	7.30
ABH	132	F	74	149.20	146.98	115.55	39.00	6.00	7.40	149.34	147.70	115.95	40.50	7.25	7.35
ABH	144	F	60	120.48	134.58	94.55	37.50	5.00	7.85	121.38	135.10	94.85	39.00	4.50	7.55

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A3 (continued): Inter- and Intra-observer measurements for the five- and two-variable model by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) for the Athens population

				Inter-observer measurements (mm)						Intra-observer measurements (mm)					
Sample^		Sex	Age*	XLS	XHS	XBS	HAX	CSV	TLB	XLS	XHS	XBS	HAX	CSV	TLB
ABH	145	M	80	148.24	157.46	109.50	45.00	6.50	9.20	148.34	157.70	108.85	43.50	5.50	8.55
ABH	156	M	59	139.16	156.54	106.95	42.00	5.50	11.10	139.88	156.36	106.15	41.00	5.75	10.55
ABH	157	M	64	138.02	162.76	102.10	42.00	6.50	9.25	138.44	162.64	101.80	41.50	4.50	9.00
ABH	173	M	58	150.86	154.16	114.80	42.00	9.00	10.50	151.12	154.22	114.00	42.50	7.75	9.85
ABH	183	F	59	124.76	132.84	92.80	38.50	4.50	6.60	125.30	132.90	92.90	38.00	4.00	6.70
ABH	185	F	72	126.20	150.90	95.50	37.00	2.50	8.05	125.30	150.56	95.50	37.00	3.50	8.40
ABH	192	M	27	146.96	173.04	113.55	45.00	3.50	10.15	147.18	172.98	113.50	46.00	1.25	9.60
ABH	193	M	25	139.04	152.78	104.30	45.00	3.00	8.35	139.52	152.96	103.95	45.50	3.00	8.30

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix A4: Inter- and Intra-observer measurements for the five- and two-variable model by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005)for the Tennessee population

Sample^		Sex	Age*	Inter-observer measurements (mm)						Intra-observer measurements (mm)					
				XLS	XHS	XBS	HAX	CSV	TLB	XLS	XHS	XBS	HAX	CSV	TLB
UT	01-87	M	39	145.68	163.94	105.55	43.50	3.50	9.90	146.20	163.80	105.55	44.00	3.50	9.55
UT	02-89	M	36	144.90	162.96	106.10	45.50	8.50	10.20	144.46	162.96	105.55	45.00	8.50	9.85
UT	03-06	F	52	130.78	135.58	96.15	35.50	3.00	7.65	128.52	135.48	95.35	37.00	2.00	7.75
UT	04-06	F	58	119.90	130.12	92.17	35.00	2.50	8.90	120.50	129.78	92.15	38.00	2.50	7.95
UT	07-01	F	50	121.28	138.92	94.40	39.00	1.50	8.30	119.82	138.60	93.00	38.00	2.00	6.90
UT	100-06	F	57	139.82	149.06	105.95	38.00	4.50	7.50	139.00	149.30	105.35	41.00	4.00	7.60
UT	101-06	F	60	134.70	138.02	100.00	41.50	6.00	7.70	133.16	138.36	99.55	42.00	6.50	7.80
UT	11-03	F	47	125.80	144.64	92.95	37.00	6.00	8.30	125.18	144.70	91.40	34.00	5.50	8.10
UT	12-01	M	50	139.60	169.74	105.65	45.00	2.00	9.70	142.14	169.76	105.35	43.00	2.50	10.00
UT	13-88	M	31	152.04	171.76	110.45	45.00	4.50	8.85	152.18	172.98	110.05	45.00	5.00	9.05
UT	14-90	M	37	141.12	148.84	109.55	40.00	5.00	7.10	141.52	148.92	109.25	41.00	4.50	7.35
UT	14-93	M	32	126.54	158.74	92.10	41.00	3.50	9.85	125.60	158.80	91.55	41.00	1.50	10.00
UT	16-98	M	58	151.46	177.44	117.85	47.50	9.50	10.85	151.52	176.54	117.35	48.00	7.00	11.15
UT	17-01	M	51	158.34	182.42	122.80	43.50	2.00	10.95	160.10	181.42	120.00	44.00	2.00	11.20
UT	17-05	F	58	127.14	134.20	94.00	41.00	2.00	8.15	128.42	134.00	94.55	41.00	1.50	7.00
UT	17-06	F	50	133.22	148.60	99.90	37.00	6.00	7.90	133.52	148.22	99.80	37.00	5.50	7.80
UT	18-03	F	47	121.98	141.88	94.45	35.50	3.50	6.20	121.04	141.58	94.75	29.00	2.00	7.00
UT	24-99	M	49	150.66	159.02	109.85	46.00	5.00	11.70	150.78	159.90	107.55	44.00	5.00	10.85
UT	25-02	M	42	150.10	169.88	109.50	44.00	8.00	9.90	154.42	169.72	109.85	44.00	7.50	9.45
UT	25-05	F	51	121.18	145.32	91.55	35.50	1.50	9.05	119.20	145.62	90.55	37.00	1.00	8.55
UT	27-93	M	39	158.26	181.04	116.80	46.50	3.50	9.75	153.10	180.70	116.75	42.00	3.00	10.90
UT	28-90	F	45	128.24	149.46	93.10	38.00	2.00	8.10	130.60	149.38	92.85	39.00	1.00	8.95

^UT = University of Tennessee; *Age in years

Appendix A4 (continued): Inter- and Intra-observer measurements for the five- and two-variable model by Dabbs and Moore-Jansen (2010) and FORDISC 3.0 by Jantz and Ousley (2005) for the Tennessee population

Sample^		Sex	Age*	Inter-observer measurements (in mm)						Intra-observer measurements (in mm)					
				XLS	XHS	XBS	HAX	CSV	TLB	XLS	XHS	XBS	HAX	CSV	TLB
UT	29-03	F	59	128.16	132.02	95.55	36.00	4.50	5.85	128.22	135.54	95.65	32.00	3.00	6.20
UT	29-93	M	56	152.50	162.56	119.65	42.00	4.00	7.65	153.82	162.62	119.85	43.00	5.50	7.15
UT	58-06	F	51	127.06	132.78	93.20	37.50	3.50	7.25	127.10	132.76	92.90	38.00	2.50	7.80
UT	61-05	F	55	127.98	145.88	92.90	39.00	5.50	9.55	127.26	145.90	92.80	42.00	6.00	8.60
UT	63-03	F	58	141.36	144.12	108.50	35.50	3.00	9.05	137.56	143.52	108.00	36.00	2.50	9.75
UT	69-04	F	62	128.74	129.42	97.75	37.00	4.50	7.55	126.98	129.50	98.55	38.00	2.50	7.55
UT	79-05	F	59	121.84	140.60	89.45	34.00	2.50	7.65	120.16	140.20	88.95	33.00	1.50	7.35
UT	92-05	F	47	143.30	154.94	103.70	39.00	3.50	5.75	143.70	155.00	101.20	41.00	2.50	6.25

^UT = University of Tennessee; *Age in years

Appendix B: Measurements used for the creation of the models for "Macaluso's Hypothesis"

Appendix B1: Measurements used in the Athens population

				Measurements (mm)		
Sample^		Sex	Age*	Height	Breadth	Calculated Area
WLH	001	M	85	42.85	33.55	1,437.62
WLH	002	M	64	38.60	29.30	1,130.98
WLH	005	F	67	30.50	28.50	869.25
WLH	006	M	62	37.90	31.00	1,174.90
WLH	008	M	60	39.50	30.05	1,186.98
WLH	009	M	76	40.00	29.30	1,172.00
WLH	010	F	68	32.75	21.55	705.76
WLH	011	M	82	39.55	28.80	1,139.04
WLH	012	F	84	33.50	24.80	830.80
WLH	013	F	49	32.55	23.30	758.42
WLH	014	M	65	40.05	32.20	1,289.61
WLH	016	M	36	38.20	28.60	1,092.52
WLH	020	M	67	36.30	26.45	960.14
WLH	021	M	76	37.80	27.50	1,039.50
WLH	022	M	94	35.70	27.90	996.03
WLH	023	M	48	39.30	29.75	1,169.18
WLH	024	M	87	40.90	32.00	1,308.80
WLH	026	M	46	34.15	29.40	1,004.01
WLH	027	M	65	37.50	28.15	1,055.63
WLH	028	F	-	33.50	23.30	780.55
WLH	030	F	66	39.90	27.60	1,101.24
WLH	031	M	60	35.85	26.45	948.23
WLH	032	F	44	36.20	25.70	930.34
WLH	033	F	72	34.80	24.50	852.60
WLH	034	F	81	38.00	28.15	1,069.70
WLH	036	F	63	32.30	23.70	765.51
WLH	037	F	44	33.20	24.00	796.80
WLH	038	M	43	39.65	28.00	1,110.20
WLH	040	F	71	33.70	23.10	778.47
WLH	041	F	27	32.50	23.00	747.50
WLH	043	M	55	39.90	26.70	1,065.33
WLH	044	M	64	33.85	27.05	915.64
WLH	045	M	57	38.30	28.60	1,095.38
WLH	046	M	60	42.25	31.75	1,341.44
WLH	048	M	50	36.95	27.75	1,025.36
WLH	051	M	49	37.40	28.25	1,056.55

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix B1 (continued): Measurements used in the Athens population

				Measurements (mm)		
Sample^		**Sex**	**Age***	**Height**	**Breadth**	**Calculated Area**
WLH	052	F	82	34.15	27.85	951.08
WLH	053	F	63	33.10	25.80	853.98
WLH	055	M	58	40.40	33.25	1,343.30
WLH	057	F	-	31.60	22.45	709.42
WLH	060	F	45	31.50	23.75	748.13
WLH	062	M	41	40.95	28.75	1,177.31
WLH	063	M	79	40.00	32.35	1,294.00
WLH	064	M	65	41.10	32.00	1,315.20
WLH	067	M	56	38.20	27.25	1,040.95
WLH	068	M	26	43.25	29.05	1,256.41
WLH	069	M	80	37.95	31.40	1,191.63
WLH	070	M	48	35.50	29.25	1,038.38
WLH	071	M	-	37.40	26.55	992.97
WLH	072	M	27	35.50	27.50	976.25
ABH	073	M	62	39.75	33.50	1,331.63
ABH	074	M	26	38.00	29.65	1,126.70
ABH	077	F	54	33.00	26.20	864.60
ABH	078	M	43	36.90	27.50	1,014.75
ABH	079	F	51	33.85	24.85	841.17
ABH	082	F	48	32.75	24.00	786.00
ABH	083	M	81	40.60	35.95	1,459.57
ABH	084	F	65	32.50	22.40	728.00
ABH	086	F	61	32.20	25.35	816.27
ABH	087	M	36	37.65	25.70	967.61
ABH	088	F	35	31.45	22.90	720.21
ABH	090	F	72	37.10	30.40	1,127.84
ABH	091	F	51	31.40	22.25	698.65
ABH	092	F	54	34.10	24.00	818.40
ABH	093	F	47	33.15	24.15	800.57
ABH	095	F	37	32.30	22.20	717.06
ABH	096	F	33	30.90	20.25	625.73
ABH	097	F	46	34.15	22.60	771.79
ABH	098	F	82	30.75	24.50	753.38
ABH	099	F	70	32.40	23.40	758.16
ABH	100	M	64	37.70	27.25	1,027.33
ABH	102	F	58	31.00	23.90	740.90

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix B1 (continued): Measurements used in the Athens population

				Measurements (mm)		
Sample^		Sex	Age*	Height	Breadth	Calculated Area
ABH	103	M	24	39.60	27.90	1,104.84
ABH	104	F	57	34.95	22.55	788.12
ABH	105	M	78	38.10	30.20	1,150.62
ABH	106	M	46	31.60	27.35	864.26
ABH	107	M	68	37.35	29.75	1,111.16
ABH	108	M	28	42.10	31.20	1,313.52
ABH	109	M	68	32.95	26.70	879.77
ABH	111	M	44	39.65	28.20	1,118.13
ABH	112	F	56	31.10	23.40	727.74
ABH	114	F	80	36.80	26.25	966.00
ABH	119	F	85	34.90	23.05	804.45
ABH	121	F	27	32.60	21.85	712.31
ABH	122	F	78	32.40	24.10	780.84
ABH	123	F	71	32.80	24.00	787.20
ABH	125	M	50	39.00	28.45	1,109.55
ABH	126	M	74	41.50	28.40	1,178.60
ABH	127	M	29	40.90	28.75	1,175.88
ABH	128	F	77	34.95	25.75	899.96
ABH	129	M	65	36.10	26.75	965.68
ABH	130	M	65	33.00	24.65	813.45
ABH	131	F	46	33.00	23.40	772.20
ABH	132	F	74	34.35	26.75	918.86
ABH	135	M	34	38.15	28.10	1,072.02
ABH	137	F	41	33.00	22.20	732.60
ABH	139	M	44	32.25	21.50	693.38
ABH	141	M	48	36.10	24.65	889.87
ABH	143	F	79	37.25	29.10	1,083.98
ABH	144	F	60	34.05	24.35	829.12
ABH	145	M	80	41.30	29.00	1,197.70
ABH	146	M	55	37.95	28.35	1,075.88
ABH	147	M	84	43.00	36.45	1,567.35
ABH	148	M	71	41.75	30.95	1,292.16
ABH	149	M	66	39.55	31.90	1,261.65
ABH	151	F	63	35.90	25.50	915.45
ABH	152	M	81	28.20	27.65	779.73
ABH	155	M	74	41.80	30.85	1,289.53

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix B1 (continued): Measurements used the Athens population

Sample^		Sex	Age*	Measurements (mm)		
				Height	Breadth	Calculated Area
ABH	156	M	59	36.60	28.10	1,028.46
ABH	157	M	64	36.95	29.10	1,075.25
ABH	158	M	78	36.95	24.30	897.89
ABH	159	F	79	33.50	25.15	842.53
ABH	162	F	69	31.90	24.40	778.36
ABH	166	F	69	35.50	24.85	882.18
ABH	168	M	66	35.10	26.40	926.64
ABH	169	F	84	33.70	24.45	823.97
ABH	172	M	88	42.95	31.10	1,335.75
ABH	173	M	58	37.20	27.55	1,024.86
ABH	176	F	81	34.20	24.60	841.32
ABH	177	M	65	37.55	27.50	1,032.63
ABH	179	F	87	33.05	23.50	776.68
ABH	180	F	81	32.40	22.20	719.28
ABH	181	M	94	34.00	28.95	984.30
ABH	182	M	55	37.40	29.35	1,097.69
ABH	183	F	59	33.35	24.55	818.74
ABH	185	F	72	33.60	26.05	875.28
ABH	186	M	26	35.55	28.20	1,002.51
ABH	188	F	38	29.95	20.50	613.98
ABH	190	F	20	33.30	22.70	755.91
ABH	191	M	32	38.95	29.95	1,166.55
ABH	192	M	27	42.35	30.00	1,270.50
ABH	193	M	25	40.50	28.20	1,142.10
ABH	194	F	35	33.00	21.95	724.35
ABH	197	M	36	37.10	28.15	1,044.37
ABH	198	F	45	33.25	25.75	856.19
ABH	199	M	26	41.05	31.25	1,282.81
ABH	200	M	43	39.75	29.25	1,162.69
ABH	202	F	26	31.85	22.15	705.48
ABH	207	M	37	39.65	27.40	1,086.41
ABH	209	M	44	39.30	32.55	1,279.22
ABH	210	M	43	38.20	28.80	1,100.16
ABH	211	M	28	36.55	25.95	948.47
ABH	213	M	32	39.55	30.75	1,216.16
ABH	214	M	33	37.80	28.15	1,064.07

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix B1 (continued): Measurements used in the Athens population

Sample^		Sex	Age*	Measurements (mm)		
				Height	Breadth	Calculated Area
ABH	215	F	32	33.10	24.75	819.23
ABH	216	F	24	32.75	25.75	843.31
ABH	217	F	22	35.65	28.90	1,030.29
ABH	218	M	29	39.55	28.05	1,109.38
ABH	220	F	35	32.05	21.25	681.06
ABH	221	M	25	35.95	27.30	981.44
ABH	223	F	38	31.00	24.65	764.15
ABH	225	F	-	35.55	25.50	906.53
ABH	226	F	68	33.55	24.75	830.36
ABH	231	M	33	39.60	29.30	1,160.28
ABH	232	F	33	34.25	23.65	810.01
ABH	233	F	85	33.50	24.25	812.38
ABH	234	F	50	34.05	24.35	829.12
ABH	235	M	43	36.95	27.85	1,029.06
ABH	236	M	59	34.55	26.65	920.76
ABH	237	M	48	41.80	34.55	1,444.19
ABH	238	F	79	37.00	26.90	995.30
ABH	239	M	37	35.50	26.25	931.88
ABH	240	F	58	33.25	24.05	799.66
ABH	241	M	75	38.50	30.60	1,178.10
ABH	244	F	79	36.40	27.35	995.54
ABH	245	M	20	30.50	21.30	649.65
ABH	246	M	30	32.00	25.30	809.60
ABH	247	M	68	41.35	31.80	1,314.93
ABH	250	M	38	37.35	28.45	1,062.61

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix B2: Measurements used in the Tennessee population

Sample^		Sex	Age*	Measurements (mm)		
				Height	Breadth	Calculated Area
UT	01-03	M	47	37.00	29.00	1,073.00
UT	01-05	M	44	39.40	29.40	1,158.36
UT	01-87	M	39	41.85	28.30	1,184.36
UT	01-95	M	42	43.05	33.35	1,435.72
UT	02-02	M	46	39.10	30.95	1,210.15
UT	02-08	F	65	35.05	26.35	923.57
UT	02-89	M	36	42.40	31.10	1,318.64
UT	03-00	M	43	40.15	28.65	1,150.30
UT	03-06	F	52	32.10	23.70	760.77
UT	04-00	M	56	43.35	31.65	1,372.03
UT	04-02	F	60	33.35	24.65	822.08
UT	04-06	F	58	29.35	22.80	669.18
UT	04-96	M	55	43.35	32.00	1,387.20
UT	04-97	M	36	37.40	29.65	1,108.91
UT	05-99	M	38	37.45	29.50	1,104.78
UT	06-08	F	61	35.15	27.35	961.35
UT	07-00	M	38	37.00	26.55	982.35
UT	07-01	F	50	37.10	26.60	986.86
UT	07-02	M	59	36.65	29.90	1,095.84
UT	07-05	M	40	39.00	29.30	1,142.70
UT	07-94	M	41	36.95	29.70	1,097.42
UT	08-04	M	57	37.70	28.85	1,087.65
UT	08-07	F	57	32.80	24.10	790.48
UT	08-87	M	25	40.85	30.65	1,252.05
UT	08-93	M	52	37.95	30.35	1,151.78
UT	08-98	M	36	38.20	26.80	1,023.76
UT	09-00	F	43	33.30	25.40	845.82
UT	09-03	M	51	41.10	31.95	1,313.15
UT	09-93	M	42	36.15	29.20	1,055.58
UT	09-97	M	56	36.45	30.45	1,109.90
UT	09-99	F	54	35.30	24.90	878.97
UT	100-06	F	57	35.00	26.85	939.75
UT	10-03	M	49	40.75	30.40	1,238.80
UT	10-05	M	46	37.50	31.10	1,166.25

^UT = University of Tennessee; *Age in years

Appendix B2 (continued): Measurements used in the Tennessee population

Sample^		Sex	Age*	Measurements (mm)		
				Height	Breadth	Calculated Area
UT	10-07	F	50	33.60	23.90	803.04
UT	101-06	F	60	36.20	27.35	990.07
UT	107-06	F	54	33.60	22.80	766.08
UT	107-08	F	52	33.90	26.15	886.49
UT	109-07	F	48	36.55	26.50	968.58
UT	10-91	M	35	40.65	29.95	1,217.47
UT	11-00	M	55	38.75	29.20	1,131.50
UT	11-03	F	47	34.50	24.75	853.88
UT	11-04	F	54	31.65	23.45	742.19
UT	111-07	F	50	33.90	26.75	906.83
UT	112-07	F	64	33.95	22.60	767.27
UT	12-01	M	50	37.20	32.70	1,216.44
UT	12-04	F	60	31.45	26.40	830.28
UT	12-98	M	46	38.95	30.20	1,176.29
UT	13-00	M	44	38.00	30.95	1,176.10
UT	13-03	M	48	37.60	29.05	1,092.28
UT	13-88	M	31	42.60	31.40	1,337.64
UT	13-91	M	34	39.05	32.45	1,267.17
UT	14-03	M	50	36.00	28.60	1,029.60
UT	14-04	M	19	39.15	28.35	1,109.90
UT	14-90	M	37	36.35	27.15	986.90
UT	14-93	M	32	36.85	26.70	983.90
UT	15-06	F	59	33.85	27.20	920.72
UT	16-98	M	58	43.60	33.25	1,449.70
UT	17-01	M	51	40.40	32.95	1,331.18
UT	17-02	F	50	32.90	25.40	835.66
UT	17-05	F	58	37.30	27.30	1,018.29
UT	17-06	F	50	33.50	26.70	894.45
UT	17-99	M	56	38.50	31.50	1,212.75
UT	18-03	F	47	31.15	24.20	753.83
UT	18-04	F	44	35.20	27.60	971.52
UT	18-99	M	55	38.80	30.75	1,193.10
UT	19-03	M	55	39.25	30.10	1,181.43
UT	19-92	M	27	39.20	27.25	1,068.20
UT	20-03	F	44	32.15	23.60	758.74
UT	20-08	F	62	33.95	27.65	938.72

^UT = University of Tennessee; *Age in years

Appendix B2 (continued): Measurements used in the Tennessee population

				Measurements (mm)		
Sample^		Sex	Age*	Height	Breadth	Calculated Area
UT	21-08	F	65	30.95	24.85	769.11
UT	21-95	M	50	40.80	29.65	1,209.72
UT	21-98	M	52	38.75	31.05	1,203.19
UT	22-00	M	57	39.45	31.20	1,230.84
UT	22-01	M	47	40.00	33.40	1,336.00
UT	22-02	M	50	40.85	31.80	1,299.03
UT	22-03	M	48	38.00	29.65	1,126.70
UT	22-08	F	50	30.60	25.75	787.95
UT	22-95	M	41	39.30	31.70	1,245.81
UT	24-02	M	52	42.20	30.80	1,299.76
UT	24-99	M	49	41.45	30.90	1,280.81
UT	25-01	M	48	37.35	28.25	1,055.14
UT	25-02	M	42	40.85	30.10	1,229.59
UT	25-05	F	51	31.60	25.60	808.96
UT	25-06	F	44	33.50	27.80	931.30
UT	26-01	M	56	40.45	31.75	1,284.29
UT	26-03	M	49	35.26	28.75	1,013.73
UT	27-03	M	46	40.60	33.40	1,356.04
UT	27-07	F	45	35.45	24.75	877.39
UT	27-93	M	39	43.30	32.15	1,392.10
UT	28-90	F	45	35.95	25.60	920.32
UT	29-00	M	39	38.90	29.90	1,163.11
UT	29-03	F	59	32.50	24.00	780.00
UT	29-04	M	34	37.45	30.05	1,125.37
UT	29-93	M	56	40.20	30.60	1,230.12
UT	30-02	M	61	37.50	28.95	1,085.63
UT	30-04	M	59	40.80	31.90	1,301.52
UT	30-93	M	46	40.80	30.40	1,240.32
UT	31-07	F	67	33.00	24.40	805.20
UT	33-02	M	39	37.20	27.70	1,030.44
UT	33-03	F	52	33.95	26.35	894.58
UT	34-02	M	58	38.75	28.60	1,108.25
UT	34-93	M	51	40.05	30.80	1,233.54
UT	35-02	F	55	35.55	28.05	997.18
UT	35-03	M	62	34.10	27.95	953.10
UT	35-07	F	46	32.95	25.50	840.23

^UT = University of Tennessee; *Age in years

Appendix B2 (continued): Measurements used in the Tennessee population

Sample^		Sex	Age*	Measurements (mm)		
				Height	Breadth	Calculated Area
UT	36-01	M	49	37.25	29.25	1,089.56
UT	36-05	M	41	38.05	31.45	1,196.67
UT	37-03	M	43	36.60	29.50	1,079.70
UT	37-07	F	57	35.25	26.95	949.99
UT	38-04	M	54	40.90	34.50	1,411.05
UT	38-05	M	35	35.90	27.15	974.69
UT	39-01	F	36	37.65	25.30	952.55
UT	39-03	F	52	33.65	24.80	834.52
UT	39-04	M	60	37.55	30.10	1,130.26
UT	40-03	M	60	40.30	32.65	1,315.80
UT	40-06	F	51	37.05	31.30	1,159.67
UT	40-08	F	33	31.50	22.75	716.63
UT	41-01	F	58	34.85	24.90	867.77
UT	41-07	F	37	33.35	22.40	747.04
UT	42-05	M	42	37.95	28.25	1,072.09
UT	42-06	F	56	34.60	29.00	1,003.40
UT	43-01	M	59	39.95	32.55	1,300.37
UT	44-02	M	60	40.70	29.20	1,188.44
UT	44-03	M	46	37.50	31.20	1,170.00
UT	44-04	M	39	38.35	30.65	1,175.43
UT	44-05	M	51	37.20	31.65	1,177.38
UT	45-04	M	54	44.35	34.45	1,527.86
UT	50-03	M	62	35.95	30.30	1,089.29
UT	50-04	M	51	39.25	31.15	1,222.64
UT	52-03	M	55	38.50	31.60	1,216.60
UT	53-06	F	54	33.80	29.35	992.03
UT	54-03	M	54	40.10	34.80	1,395.48
UT	54-05	F	54	31.60	24.80	783.68
UT	55-06	F	66	34.45	24.30	837.14
UT	55-07	F	51	33.30	25.25	840.83
UT	56-07	F	57	34.85	24.40	850.34
UT	56-08	F	57	33.30	27.75	924.08
UT	57-03	M	55	41.85	28.85	1,207.37
UT	57-05	F	60	36.05	27.30	984.17
UT	58-06	F	51	31.65	24.15	764.35
UT	59-04	M	48	35.30	23.20	818.96
UT	59-05	M	53	39.10	28.95	1,131.95

^UT = University of Tennessee; *Age in years

Appendix B2 (continued): Measurements used in the Tennessee population

Sample^		Sex	Age*	Measurements (mm)		
				Height	Breadth	Calculated Area
UT	61-04	M	50	36.50	29.25	1,067.63
UT	61-05	F	55	35.95	24.50	880.78
UT	61-08	F	48	34.00	22.40	761.60
UT	63-03	F	58	34.25	25.45	871.66
UT	63-04	M	61	38.20	29.90	1,142.18
UT	64-04	M	53	36.25	27.80	1,007.75
UT	68-06	F	54	34.30	25.40	871.22
UT	68-07	F	42	34.35	26.20	899.97
UT	69-04	F	62	31.45	24.65	775.24
UT	69-06	F	45	31.85	22.55	718.22
UT	70-06	M	42	38.25	29.65	1,134.11
UT	72-08	F	55	35.35	26.30	929.71
UT	73-08	F	60	31.70	26.00	824.20
UT	74-06	F	42	35.05	24.45	856.97
UT	77-07	F	36	32.95	24.50	807.28
UT	78-06	F	49	34.25	28.55	977.84
UT	79-05	F	59	29.10	23.50	683.85
UT	82-07	F	31	29.80	23.15	689.87
UT	84-08	F	46	35.05	23.50	823.68
UT	85-08	F	47	36.10	28.15	1,016.22
UT	92-05	F	47	35.13	25.50	895.82
UT	97-07	F	66	32.80	27.65	906.92
UT	98-07	F	66	33.30	26.80	892.44

^UT = University of Tennessee; *Age in years

Appendix B3: Inter- and Intra-observer measurements of the models for “Macaluso’s Hypothesis” from the glenoid cavity for the Athens population

				Inter-observer measurements (mm)		Intra-observer measurements (mm)	
Sample^		**Sex**	**Age***	**Height**	**Breadth**	**Height**	**Breadth**
WLH	011	M	82	39.30	28.00	39.00	28.90
WLH	020	M	67	35.75	26.70	34.35	26.35
WLH	022	M	94	36.70	27.50	36.70	27.00
WLH	023	M	48	39.20	29.75	37.95	29.55
WLH	064	M	65	40.60	31.50	43.00	32.20
WLH	069	M	80	38.30	31.70	39.85	31.20
ABH	072	M	27	36.55	27.25	35.10	27.50
ABH	073	M	62	39.55	33.85	39.05	33.00
ABH	077	F	54	33.45	26.25	33.10	25.80
ABH	086	F	61	32.95	25.80	32.40	25.95
ABH	090	F	72	37.05	30.00	36.35	30.00
ABH	092	F	54	33.35	24.50	33.80	24.85
ABH	097	F	46	35.20	22.95	34.95	22.10
ABH	098	F	82	32.40	24.65	31.15	24.70
ABH	102	F	58	32.25	24.00	30.65	25.00
ABH	105	M	78	39.90	30.80	38.70	30.70
ABH	106	M	46	31.00	27.50	36.70	27.55
ABH	114	F	80	37.40	26.50	38.55	27.25
ABH	121	F	27	32.50	22.00	32.30	21.30
ABH	122	F	78	32.55	24.70	31.45	24.00
ABH	130	M	65	34.00	24.75	33.60	24.75
ABH	131	F	46	33.50	23.70	33.60	23.65
ABH	132	F	74	34.90	27.10	34.45	27.35
ABH	144	F	60	34.75	24.50	34.45	25.65
ABH	145	M	80	38.75	28.50	41.85	30.00
ABH	155	M	74	41.85	31.00	42.55	30.10
ABH	156	M	59	36.85	28.40	37.65	28.75
ABH	157	M	64	37.10	29.20	37.05	29.20
ABH	183	F	59	33.40	24.65	35.60	25.30
ABH	185	F	72	33.30	26.20	33.75	26.15

^ WLH = Weiner Lab Human, ABH = Athens Biology Human; *Age in years

Appendix B4: Inter- and Intra-observer measurements of the models for "Macaluso's Hypothesis" from the glenoid cavity for the Tennessee population

				Inter-observer measurements (mm)		Intra-observer measurements (mm)	
Sample^		Sex	Age*	Height	Breadth	Height	Breadth
UT	01-87	M	39	41.85	28.50	42.20	28.55
UT	02-89	M	36	41.80	31.05	42.95	29.90
UT	03-06	F	52	32.00	24.20	33.15	24.05
UT	04-06	F	58	30.05	22.75	29.35	22.90
UT	07-01	F	50	37.35	26.75	37.30	26.65
UT	100-06	F	57	34.50	26.65	34.65	26.90
UT	101-06	F	60	35.60	27.40	36.40	27.80
UT	11-03	F	47	34.25	24.45	35.15	25.95
UT	12-01	M	50	36.90	32.85	38.00	33.10
UT	13-88	M	31	41.65	31.35	43.25	30.50
UT	14-90	M	37	35.75	27.10	35.25	26.75
UT	14-93	M	32	36.35	26.70	37.20	26.55
UT	16-98	M	58	43.20	32.75	42.60	32.35
UT	17-01	M	51	40.60	32.90	45.00	33.00
UT	17-05	F	58	37.75	27.40	35.60	22.55
UT	17-06	F	50	34.10	26.65	31.90	26.55
UT	18-03	F	47	31.00	24.20	32.15	24.20
UT	24-99	M	49	41.20	30.40	40.45	30.30
UT	25-02	M	42	40.50	30.25	40.15	29.75
UT	25-05	F	51	31.70	26.10	33.90	25.50
UT	27-93	M	39	43.35	32.10	42.45	32.60
UT	28-90	F	45	35.35	25.35	35.35	25.25
UT	29-03	F	59	32.80	24.15	37.65	24.45
UT	29-93	M	56	40.30	30.75	39.60	31.00
UT	58-06	F	51	32.30	24.20	31.85	24.25
UT	61-05	F	55	35.95	24.55	35.25	24.40
UT	63-03	F	58	34.60	25.35	34.65	25.00
UT	69-04	F	62	31.90	24.70	35.00	25.00
UT	79-05	F	59	29.25	23.80	33.65	23.65
UT	92-05	F	47	34.80	25.95	34.10	25.65

^ UT = University of Tennessee; *Age in years