Are Metric Methods Really User-Friendly?
A Methodological Study of Sex Estimation Techniques for the Talus and Calcaneus

by

Grace Chan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Public Issues Anthropology

Waterloo, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Skeletal sex is most commonly estimated using the pelvis and the skull. These elements, however, are not always available in archaeological and forensic situations as they may be missing or damaged as a result of burial practices or poor preservation. Anthropologists have developed sex estimation methods that utilize other skeletal elements, and many of these alternative methods rely on statistical analyses of bone metrics. Because metric methods are seen as more objective and less dependent on examiner experience, most have not undergone the independent validation to which morphologic methods have been subjected. The purpose of this research is to validate two previously developed metric methods for the talus and the calcaneus using a different population than the ones on which these methods were developed, and explore potential issues of precision and validity when these methods are applied by external users. This thesis recommends several areas for improvement in the development and publication of metric methods, including the necessity for more external validation studies, greater standardization of variables and methodology, an increased use of probabilistic estimates, and a re-evaluation of how symmetry and error are conceptualized and assessed.

Key Words: anthropology, bioarchaeology, sex estimation, metric methods, talus, calcaneus

Acknowledgments

First and foremost, I am deeply grateful to my supervisor, Dr. Maria Liston, for her support throughout my time at UW. Although I regret that we were never able to travel to Greece together, my experience in this program has been extremely positive due to her guidance and very kind encouragement.

I would also like to especially thank Dr. Alexis Dolphin for her care and support throughout my time in this program, and particularly for her assistance in my transition to a new project when my plans had to change due to Covid-19 travel restrictions.

I am so grateful to my thesis committee composed of Dr. Bonnie Glencross, Dr. Alexis Dolphin, and Dr. Maria Liston for their time and support in helping me produce my thesis. I would like to thank Dr. Maria Liston for acting as a second observer for data collection, as well as Jeff Coffin for his assistance in the lab. I would also like to thank Dr. Jennifer Liu, Dr. Adrienne Lo, Dr. Alexis Dolphin, and Dr. Maria Liston for their guidance throughout the coursework terms of this program.

I am grateful for the financial support that has been provided to me by the UW Anthropology Department, the UW President's Graduate Scholarship, the family of Iris Yuzdepski through the Iris Yuzdepski Award, and the Social Sciences and Humanities Research Council through the Canada Graduate Scholarship.

I would like to thank my friends, family, and graduate cohort for their support throughout this journey. In particular, I would like to thank my partner, Peter, and my parents, Homan and Jenny, for being there for me always. I would also like to especially thank Laurel McInnis for her assistance and company in the lab and throughout this process, and Kaithleen Tan for her consistent support and encouragement.

Finally, I am grateful to the individuals of the Church of St. George in Gurat, who have made this study possible.

Table of Contents

List of Figures

List of Tables

Prologue: Changes Due to Covid-19

The original goal of this thesis was to estimate the sex distribution of a Byzantine Greek cemetery containing over 200 individuals. As the upper bodies of many individuals were exhumed prior to archaeological discovery, sex was to be estimated from the talus and calcaneus, which are abundant in number and generally well-preserved in this population. Due to travel restrictions surrounding Covid-19 as well as the occurrence of wildfires in Greece, the original project was unable to proceed. As a result, the intended methods were applied to a sample of archaeological skeletons upon generous permission from Dr. Alexis Dolphin. The skeletons were excavated from the site of the Church of St. George in Gurat, France, and are currently stored within the Ancient and Contemporary Environmental Bioindicators Laboratory (ACEBioLab) at the University of Waterloo.

Chapter One: Introduction to the Cave-Church at Gurat and Ethical Considerations

1.1 Chapter Introduction

Bioarchaeology is an area of biological anthropology that studies human experiences across space and time through integrating evidence from human remains with contextual understandings of environmental and cultural processes. Bioarchaeological conclusions, therefore, may provide insights on topics including past demographics, health, and behaviour (Martin, Harrod, & Pérez, 2013; Stojanowski & Duncan, 2015). A bioarchaeological approach, however, also necessitates an appreciation of the current public issues in which the researcher and the archaeological remains are situated, as these contexts impact interpretations of the past. My thesis analyzes the application of two relatively new sex estimation methods to human remains from the Church of St. George in Gurat, France, dated to approximately 1390 AD. This chapter will introduce the historical context of the site and research, the ethical considerations surrounding the use of human remains in research, and the state of archaeological remains legislation in France.

1.2 Historical and Site Context

The tiny village of Gurat in southwestern France is perched atop a limestone cliff overlooking the Lixonne river valley. Hidden within the cliff face, beneath the village and the road, are a series of carved grottos and a church. The cave-church is one of a number of rock-cut ecclesiastical sites found across southwestern France and other European countries. These sites are thought to have been part of a cultural tradition of communities practicing group hermitism and monasticism across Europe (Gervers, 1967). The hermitical movement in this area rose during the 6^{th} to 8^{th} centuries under the influence of figures including Pope Gregory the Great and Charlemagne, persisting until the Wars of Religion in the 16th century (Gervers, 1967). Although the cave-church of Gurat, known locally as the "Church of St. George", is not referenced in the local archives, historical evidence demonstrates that the geographic region near Gurat was associated with monastic activity for an extended time. The site's hidden location within the cliff provided the community seclusion from the bustle of daily life while its proximity to an important medieval road and pilgrimage routes afforded inhabitants the ability to easily access the outside world (Gervers, 1967).

It is hypothesized that a hermitical community arose in the grottos of this site several centuries prior to the carving of the church. The church remained in use by the hermitical community throughout the Hundred Years' War (1337-1453) and the Black Death until the Wars of Religion (1562-1598), during which the community was dispersed. Following the late 16th century, the church and grottos were used for various secular purposes, including as a refugee shelter, an iron foundry, and, ultimately, a garbage dump (Gervers & Gervers, 1975).

The excavation, led by Michael and Veronica Gervers, took place between 1965 and 1974. Although the general lack of stratigraphy (due to the nature of the rock-cut site) presents challenges in dating the occupation, analysis of pottery and coins found at the site support that its occupation by the hermitical community included the 12th to 14th centuries (Franklin & Gervers, 1978; Gervers, 1967; Gervers & Gervers, 1975).

As two larger churches in the area surrounding Gurat were associated with necropolises or cemeteries, it was unsurprising that skeletal remains were found at the site. The cemetery found at Gurat was unlike the others, however, as the burials were not located inside the church but rather on a ledge directly in front of the church. The remains of 18 individuals were excavated during the 1973 and 1974 field seasons, but it is estimated that over 200 individuals are still interred at the site (Meijer et al., 2019). Charcoal fragments found in one of the burials (Grave Hb) have been carbon-dated to approximately 1390 AD, although the buried individuals may not have been contemporaneous (Franklin & Gervers, 1978; Meijer et al., 2019)

The archaeological study of ecclesiastical buildings is rooted in architectural conservation projects of the nineteenth century. Thus, the field of monastic and church archaeology was initially focused on describing patterns in the physical forms of churches and situating these churches within the wider landscape. In the last half-century, however, archaeologists have shifted toward an increasingly holistic approach in understanding the topographic, symbolic, and economic contexts of these spaces, leading to new and diverse research questions (Gilchrist, 2014). The continued study of this site and its individuals will therefore contribute to scholarly understandings of the hermitical tradition that flourished across Europe in the Middle Ages, and the interpretation of archaeological evidence at this site will provide insights and answers which cannot be found in documentary sources (Gervers & Gervers, 1975).

1.3 Biological Anthropology in France

The present-day analysis of archaeological human remains is impacted by the processes that have shaped biological anthropology and archaeology in France. Research on skeletal remains in France began in the 19th century, rooted in the philosophical school of naturalism (Michel & Charlier, 2011). Skeletal analysis initially focused on understanding human evolution largely through the study of skulls, influenced by the belief that all pertinent information that can be learned from a skeleton is presented in the skull. These priorities, and the associated study of cephalic indices and other race-based differences, perpetuated harmful and discriminatory stereotypes until the 1970s (Michel & Charlier, 2011). Such stereotypes were not unique to France and are widely reflected in the early study of skeletal remains across Europe and the Americas.

Within the last 50 years, the trajectory of French biological anthropology has mirrored that of other countries in Europe and America in shifting away from racist and essentialist ideas and toward a more comprehensive interpretation of past populations' cultural and biological evolution (Caspari, 2003; Walker, 2000). French schools for biological anthropology developed in the 1980s and 1990s and began to explore a more anthropological bioarchaeology. For example, Henri Duday's *archaeothanatologie*, the study of the body within its burial environment, examines the interactions of biological and cultural components of death in context (Tiesler, 2011). The increased use of archaeological material and methods, along with growing multi-disciplinary cooperation in analyzing multiple lines of evidence such as taphonomic changes, funerary practices, DNA, isotopes, and entomological remains, has vastly broadened the research pursuits of biological anthropologists across France (Michel & Charlier, 2011).

1.4 Defining the Public and the Public Issue

The expertise gained by bioarchaeologists through the study of the biocultural experience across both space and time is relevant to many topics of public debate (Stojanowski & Duncan, 2015). It is also apparent, however, from the history of biological anthropology, that research on skeletal remains can perpetuate harm in the present day. Therefore, it is important to consider the public issues relevant to this research and the impacts of such issues. The "public" impacted by bioarchaeological research may include groups who are biologically, geographically, or culturally related to the population being studied (Pullman, 2018). Although much is still unknown about the Gurat individuals, the existence of a general geographic relationship between these individuals and the present-day population of France is clear. Additionally, because the interpretation of antiquity contributes to the building of cultural memory and identity, the public may also include future peoples of France.

Given the necessity of obtaining and analyzing human remains in bioarchaeological inquiries, it stands to reason that the ethics of using human remains for research is an important public issue. The ethics of humans remains research is a complicated topic in bioarchaeology, however, due to the temporal and often spatial separation of the researcher from the people being studied. In the case of this project, this separation is further compounded by the remains having been excavated by other researchers in the 1960s and 1970s, given changes that have occurred in France's legislation surrounding archaeological remains in the time between the first excavation in 1965 and the research occurring in 2021. Additionally, between 1965 and 2021, an increased concern with the ethical aspect of human remains research has compelled biological anthropologists to consider ethical issues surrounding the excavation and study of human remains (Davadie & Koehler, 2021; Michel & Charlier 2011). A discussion of the ethics of using human remains therefore raises larger questions about ownership, access rights, and moral responsibility that concern a public beyond the people of France.

1.4a Ethical Considerations in Human Remains Research

Much of the change in the discipline's approach toward the excavation and study of human remains has been brought about by anthropologists in countries with a history of colonialism, where researchers from the dominant group have a history of excavating and studying the remains of oppressed peoples (Buikstra, 2006; Klesert & Powell, 1993; Sayer, 2009;

Schreiber et al., 2019). As a result, professional organizations such as the American Association of Biological Anthropologists and the Canadian Association of Biological Anthropology have put forth codes of ethics that outline the primary responsibility of anthropologists as preventing harm to the people with whom they work (AAPA, 2003; CAPA, 2019). It is important to note, however, that not all research on human remains involves such a power differential (DeWitte, 2015) and, without a power imbalance between those approving the research and the peoples being researched, there are fewer barriers to ensuring this research is ethical.

The absence of a postcolonial context in this project, however, is only one ethical consideration amongst many. There has been substantial debate within bioarchaeology regarding who can give consent for the study of past peoples (Holm, 2001; Pullman, 2018) and whether universal principles for the ethical treatment of human remains should exist (Pullman, 2014; Walker, 2000). Additionally, although discussions of anthropological ethics often focus on the effects on contemporary living peoples, it is important to consider that research can also harm the dead (Scarre, 2006; Tarlow, 2006). Concepts of personhood differ across time periods and cultures, and such conceptualizations are reflected in how human remains are viewed and engaged with by relatives and descendants (Alfonso & Powell, Buikstra, 2006; Walker, 2000). For example, some cultures view the post-mortem body as another stage in the life cycle and, therefore, human remains must be treated in certain ways (Buikstra, 2006). Even though some cultures view the body as invulnerable after death, human remains may carry symbolic significance to the living and must therefore be treated appropriately (Walker, 2000).

The complexities of the ethics involved in human remains research far exceed the scope of this thesis, but it is clear that anthropologists must acknowledge the messy history of the discipline and consider the impacts of their research in the applicable historical and cultural contexts.

1.4b The Status of Archaeological Human Remains in France

Especially as a foreign researcher, it is important that this project respects the context of French archaeological remains and the laws that govern them. During excavations in the 1960s and 1970s, permissions to excavate the site and remove the remains from France were provided by several stakeholders including the owner of the land on which the site is located, the mayor of Gurat, the Direction Régionale des Antiquités Historiques, and the Conservation Régionale des Batiments de France de Poitou Charentes (Gervers & Gervers, 1975). Since the 1980s, however, France has seen an increased number of excavations due to growing rehabilitation and construction projects, spurring several changes in legislation surrounding the ownership and legal status of archaeological remains. Because public concerns are constantly in flux (Stojanowski & Duncan, 2015), it is important to re-evaluate the context of these archaeological human remains in order to truly address the public issues surrounding this research.

In contrast with the controversy surrounding Native American and medieval Jewish graves, there has been little debate over the archaeological study of Christian graves (DeWitte, 2015; Gilchrist, 2014). In France, there does not exist any specific regulations pertaining to religious considerations, nor are religious authorities involved in the large majority of excavations. Research excavations are authorized by the regional archaeological department under the Direction Régionale des Affaires Culturelles. No separate license is needed to excavate archaeological human remains, and the discovery of burials does not need to be reported to the authorities. Prior to 2004, archaeological human remains did not have any special legal status and were thus governed under Law no. 41-4011 which provided regulations for archaeological excavations. Law no. 2001-44, passed in Jan 2001, discusses archaeological artifacts but also does not provide any differentiation between human remains and other material culture (Michel & Charlier, 2011). This lack of differentiation is also seen in the legislation of several other European countries, including Italy (Sayer, 2009) and Greece (Eliopoulos et al., 2011).

In 2004, France's Ministry of Culture released a new standard operating procedure for the treatment of items recovered from archaeological fieldwork which distinguished between "archaeological finds" and "natural and biological materials". Although archaeological finds could continue to be divided between the state and private landowner(s), all natural and biological materials were declared to be under State jurisdiction (Davadie & Koehler, 2021; Michel & Charlier, 2011). In 2011, the Ministry of Culture introduced a standardized inventory with conservation recommendations for various finds. Despite an increase in regulations, however, the country still lacks the infrastructure to adequately accommodate human remains. Although some remains are stored in museums close to the site at which they were found, not all museums are willing or able to store large numbers of remains (Davadie & Koehler, 2021). There is also no specific legislation for the reinterment or disposal of human remains, resulting in some unfortunate instances of remains being deposited at public waste sites (Michel & Charlier, 2011).

France's legislation permits human remains to be used in research for the goals of knowledge and education. Human remains or samples of human remains may also be sent abroad for analysis upon an agreement being reached between the relevant authorities in both countries, and only occasionally is an export declaration required (Michel & Charlier, 2011). Thus, the current study of human remains obtained during the 1970s is still in keeping with France's present-day legislation.

The Centre for Conservation and Study of Alsace (CCE) was established in 2016 with the goal of improving the conservation and management of its regional archaeological materials, including human remains. In 2020, recognizing an increased demand for access to human remains for research purposes, the CCE implemented a set of protocols for destructive analyses to protect the sustainability of these materials (Davadie & Koehler, 2021). Although the study of the human remains used in this project is not bound by these protocols, it is part of my ethical responsibility as a foreign researcher to recognize and aid the more recent goals of French archaeology by using minimally destructive methods.

1.5 Proposed Venue for Publication

I intend to prepare Chapter Two of my MA thesis for submission to the International Journal of Osteoarchaeology. The journal publishes population-level studies regarding methodology, paleopathology, or biomolecular analyses relevant to the study of human and animal bones in archaeological contexts (Martin & Bendrey, 2021). My methodological study, which compares the application of two sex estimation methods developed for European populations, will contribute to further understandings of how metric methods may be developed and described to maximize their validity and usability for scholars other than the original authors. Furthermore, other interested audiences including the Gurat community may access my thesis through the University of Waterloo's repository (UWSpace).

Chapter Two: Methodological Considerations in the Application of Two Metric Sex Estimation Techniques for the Tarsals

2.1 Chapter Introduction

The purpose of this thesis is to assess the application of two metric sex estimation methods for the talus and calcaneus to a different population than the ones on which these methods were originally developed.

Estimations of sex, age-at-death, and ancestry provide important information for interpreting the past at both the individual and population levels (Agarwal, 2012; Martin et al., 2013). Biological sex is commonly estimated from skeletons through the visual analysis of morphologic traits in the pelvis and the skull. These methods have generally yielded high accuracies and are most commonly employed in osteological analysis (Buikstra & Ubelaker, 1994; Phenice 1969). The pelvis and skull, however, are not always available in archaeological and forensic situations as they may be missing or damaged as a result of taphonomic changes or burial practices. Thus, biological anthropologists have explored additional sex estimation methods using other skeletal elements, including various long bones (Colman et al., 2018; Kranioti & Michalodimitrakis, 2009; Tallman & Blanton, 2020), the scapula and clavicle (Hudson et al., 2015, Papaioannou et al., 2011), the vertebrae and sacrum (Rozendaal et al., 2020; Zhan et al., 2018), metacarpals (Barrio et al., 2006), and tarsals (Mahakkanukrauh et al., 2014).

Many sex estimation methods which use elements other than the skull and pelvis are metric, as sexual dimorphism of these elements is not easily distinguished visually (Abd-Elaleem et al., 2012). Although metric methods have several unique limitations compared to morphologic methods such as population specificity, the requirement for relatively intact and complete elements, and the more time-consuming nature, it is generally accepted that metric methods are more objective (Phenice, 1969). Because of this objectivity, metric methods are assumed to be easier to apply and less dependent on examiner experience. Therefore, most have not undergone the same scrutiny and validation to which morphologic methods have been subjected. Examples of external validation tests for morphologic methods include those by Irurita Olivares and Aleman Aguilera (2016), Lovell (1989), Meindl et al. (1985), Vance et al. (2011), and Weiss (1972), but external validation studies of this nature are uncommon for metric methods. Most studies using pre-existing metric methods apply previously established variables and methodologies to develop new equations for different populations (e.g., Peckmann et al., 2015a, 2015b). It is possible, therefore, that metric methods may not be as repeatable and valid as is generally accepted, as there are few independent studies that have tested the application of previously developed metric methods. This thesis compares methodologies and results to explore possible sources of error in external users' interpretation and application of two metric methods, and discusses potential best practices for future authors in the development and publication of these metric methods to enhance their usability.

2.2 Background

2.2a The Use of Tarsals for Sex Estimation

Bones with relatively smaller surface area to volume ratios undergo reduced exposure to taphonomic processes and thus are more likely to be recovered intact, which is crucial for many metric methods (Mountrakis et al., 2010). The tarsals have drawn the attention of anthropologists developing sex estimation methods due to their compact structure and density (Saldías et al., 2016; Scott et al., 2017; Wilbur, 1998).

The sexual dimorphism of the human foot, evident in the differences between the average shoe sizes of males and females, has also been explored and confirmed in scholarly literature (Domjanic et al., 2015; Fessler et al., 2005; Krishan et al., 2015). Steele (1976) was amongst the first to demonstrate sex-related differences in the tarsals, using 120 documented individuals from the Smithsonian's Terry Collection to develop discriminant function equations for metric sex estimation on the talus and calcaneus. Following Steele's (1976) study, the late 20th century and early 21st century have seen remarkable growth in the exploration of metric sex estimation methods using the tarsals. These methods have been explored on a number of different populations, including Egyptian (Abd-Elaleem et al., 2012), South African (Bidmos & Asala, 2004; Bidmos & Dayal, 2003), Spanish (Saldías et al., 2016), Italian (Gualdi-Russo, 2007; Introna et al., 1997), Korean (Lee et al., 2012), Thai (Mahakkanukrauh et al., 2014), and Greek (Peckmann et al., 2015a; Peckmann et al., 2015b) populations. While the bulk of the studies in this area have used historical or contemporary skeletons from documented collections, some methods have been developed using prehistoric skeletons for whom sex had been estimated independently by previous researchers using pelvic and cranial traits (Barrett et al., 2001; Murphy, 2002; Wilbur, 1998). Consistent findings of dimorphism in the tarsals of populations across space and time demonstrates that the study of sex estimation using the tarsals is valuable.

Although some studies have explored sexual dimorphism in the navicular, cuboid, and cuneiforms (Harris & Case, 2012; Saldías et al., 2016), an overwhelming majority of sex estimation methods for the tarsals have focused on the talus and/or calcaneus. As the largest tarsals, the talus and calcaneus are theoretically most likely to be found in excavations.

Anthropologists have also explained the use of the talus and calcaneus on the basis that all populations exhibit some sexual dimorphism in weight, and the important role of these bones in weight bearing and weight transmission would thus reflect this dimorphism (Barrett et al., 2001; Harris & Case, 2012; Mahakkanukrauh et al., 2014, Nathena et al., 2017). The relationship between sex and weight, however, can be complicated. For example, a history of adolescent obesity has been found to be positively correlated with pelvis breadth in adulthood (Novak et al., 2020). This raises questions about the basis of sexual dimorphism and what is truly evaluated by metric sex estimation methods.

2.2b The Theoretical Basis of Sex Estimation

Sexual dimorphism results from an interaction of intrinsic and extrinsic factors which affect the human skeleton throughout life. Intrinsic factors are those controlled by genetics, while extrinsic factors include diet and nutrition, activity levels, body mass, and the biomechanical effects of load bearing, locomotion, and muscle usage. These factors interact to create dimorphic traits; for example, the growth of an individual's bones is impacted by both genetics and nutrition. Moreover, a trait that is genetically controlled may have biomechanical consequences which enhance the expression of that trait as the skeleton develops (Moore, 2013). Biological anthropologists have attempted to understand the effects of these factors on the skeleton and utilize this knowledge to estimate sex from skeletal remains. Methods of sex estimation tend to evaluate either the size and robusticity of features or pelvic traits related to the functions of childbirth. The former includes both morphologic (e.g., the visual assessment of cranial traits) and metric methods (Ubelaker & DeGaglia, 2017).

At its core, metric sex estimation considers an element's size in comparison to the size of the same element in other males and females of the population. Male skeletal elements are, on average, larger than female skeletal elements, although the bell-shaped curves overlap in the middle (Moore, 2013; White & Folkens, 2005; Ubelaker & DeGaglia, 2017). Size-related approaches, however, must carefully consider the important factor of population variation. A plethora of literature, from anthropology and other scientific disciplines, has demonstrated that human growth and development, and the resulting expression of sexual dimorphism in the skeleton, vary considerably across the globe (Eleveth & Tanner, 1990; Perez & Monteiro, 2009; Relethford, 1994; Ruff, 2002; Schafer & Black, 2005). While population variation also impacts the assessment of morphological traits, its effects are particularly reflected in the methodology of size-related (and particularly metric) techniques. Thus, many researchers in recent decades have sought to develop sex estimation techniques for various skeletal elements using populations that have not been studied extensively in prior literature (Ubelaker & DeGaglia, 2017).

Understanding that the body forms throughout the lifetime is not only key to this study's methodological considerations, but also to meaningfully interpreting the results and implications of this study and any other research on skeletal sex estimation. Many researchers who work with sex estimation methods differentiate sex from gender using the biological versus cultural distinction, resolving to only address the biological component of sex within their studies (Wesp, 2017). Such a simplistic dichotomy, however, ignores the biocultural and plastic nature of skeletal tissue, and the resulting complexity of sex as it is reflected in the skeleton (Agarwal, 2016; Joyce, 2017). Instead, researchers must acknowledge that traits of sexual dimorphism reflected in the skeleton are constructed throughout the life course by an interplay of biological and cultural factors that have transformed throughout space and time (Agarwal, 2016; Lock & Nguyen, 2010; Wesp, 2017).

The fallacy of the oft-used biological versus cultural distinction is even more relevant to present-day work in anthropology because of the ongoing discourse in both the public and academic spheres regarding sex and gender, and the oppressive impacts this discourse may have on marginalized groups. The idea of a sex binary is still quite pervasive in 21st century Western academia, particularly in research on skeletal biology (Joyce, 2017), and it is clear that academic work on sex and gender, even on archaeological individuals, has public impacts on present-day views of sex and gender (Power, 2020). Therefore, although sex estimation is undeniably valuable for bioarchaeological research, researchers addressing questions of sex differences in past populations have the responsibility to recognize that sex is not determinant, universal, or binary, and incorporate this into their interpretations of data (Agarwal, 2012; Joyce, 2017).

The terms "sexing", "sex determination", "sex assessment", and "sex estimation" are commonly used interchangeably in biological anthropology. Spradley and Jantz (2011) differentiate between sex estimation and sex assessment on a methodological basis, associating them with metric and morphological methods, respectively. It is argued that metric methods can produce sex estimates because they are accompanied by estimated classification rates and other statistics, while morphologic methods produce sex assessments because they are more subjective and do not include statistics. Wesp (2017) also discusses the use of this terminology, but prefers the use of sex estimation for all methods, maintaining that "estimation" is a more accurate reflection of the process as all such analyses are based on a proxy for sex/gender identity, which is inextricably linked to cultural understandings. Wesp (2017) argues that using the term sex estimation allows anthropologists to acknowledge the limitations and assumptions inherent to their analyses and avoid misleading others, especially those outside of the discipline. In contributing to academic literature that seeks to be respectful of sex and gender identities in both the past and present, this thesis will use the term “sex estimation”.

2.2c The Gurat Individuals

18 individuals were excavated from the burials at the Gurat cave-church found during the 1973-1974 field seasons (Meijer et al., 2019). The limestone graves were cut from a ledge directly east of the church and the burials were oriented eastward, consistent with the tradition of Christian burials (Gervers & Gervers, 1975). The excavated burials are composed of both primary and secondary burials, and some consist of several individuals buried together. Additionally, some of the burials were likely disturbed during the 16th century War of Religion, as evidenced by the presence of stone and cement-like material from a destroyed wall within three of the graves (Clements & Gruspier, n.d.).

In the decades following excavation, several attempts were made to analyze these remains (Clements & Gruspier, n.d.; Gaherty et al., 1978) but the results were preliminary and never published. A blood group study was also performed, although the sex estimations from this study differed markedly from the results derived from the preliminary skeletal analysis (Gervers, 1976). The skeletal remains were also used in studies unrelated to understanding the Gurat site, such as a diagenesis study by Hancock and colleagues (1987) and an unpublished density study (Seymour, 2019). Additionally, samples have been taken from various bones of at least five of the skeletons for unknown purposes (Seymour, 2019).

Recent projects have produced more insights on the Gurat individuals. Meijer and colleagues’ (2019) strontium isotope analysis of 14 Gurat individuals showed that the individuals were all migrants to the site, but had lived there for at least several years before death. As well, additional estimates of age-at-death and sex (Meijer, 2018) and the creation of osteobiographies (Seymour, 2019) have provided further information on many of the individuals. The relative completeness of many of these individuals, with the presence of the skull and pelvis for sex estimations using traditional methods, makes this sample appropriate for a methodological study.

2.3 Materials and Methods

2.3a Sample

This study utilized all available tali and calcanei from the excavated Gurat individuals. The sample was composed of 11 tali (five paired, one unpaired) and one calcaneus (Table 2). Unfortunately, many calcanei had been removed from the sample for a bone density study in the 1970s which was never completed, and the bones were not returned (Seymour, 2019).

Age-at-death estimates (Appendix D.2) were obtained through examination of the auricular surface and cranial suture fusion (Buikstra & Ubelaker, 1994). The sample is mostly composed of middle-aged adults between the ages of 35-50, with the exception of Gu 6 (a juvenile), Gu 3 (a younger adult), and Gu 7 (an older adult). Individual Gu 6 is a juvenile, estimated to be between the ages of 14-16 at death based on epiphyseal fusion (Meijer, 2018; Seymour, 2019). Given the methodological aims of this study and the already limited sample size, it was reasonable to include Gu 6 in the data collection. The analysis of results, however, will consider that this individual is a juvenile.

Limitations of the small sample size are addressed in section 2.5c.

2.3b Data Collection

The methods applied to this sample were drawn from Peckmann and colleagues (2015a, 2015b) and Curate and colleagues (2021), and are hereafter referred to as Method 1 and Method 2, respectively. Method 1 was developed by Peckmann and colleagues (2015a, 2015b) using a 20th century Greek sample from the University of Athens Human Skeletal Reference Collection. Method 1 provides discriminant function equations for application by external users. Method 2 was developed by Curate and colleagues (2021) using a 19th-20th century Portuguese sample from the Coimbra Identified Skeletal Collection. Method 2 uses logistic regression and support vector machine models, and is applied using a custom web tool designed by the authors. These methods were selected as, out of the existing methods which use the talus and calcaneus for sex estimation, the Greek and Portuguese populations are of the greatest geographic proximity to France. Given that genetics are an important factor in sexual dimorphism, these methods may be more likely to be applicable to a French sample compared to methods developed using American, Thai, Korean, or South African populations. More information on the samples used for the development of these methods can be found in Appendix A.2.

For Method 1, measurements of nine variables were collected from each talus as described by Peckmann and colleagues (2015b). Written descriptions of the variables were obtained from Steele (1976) (cited by Peckmann, 2015b) and Peckmann (pers. comm. 2021, Appendix B.1). The variables were: talar length (TL), head-neck length (HNL), trochlear length (TrL), length of the posterior articular facet for the calcaneus (LPAS), talar width (TW), trochlear breadth (TrB), breadth of the posterior articular facet for the calcaneus (BPAS), head height (HH), and talar height (TH). The measurements of talar length (TL), trochlear length (TrL), length of the posterior articular facet for the calcaneus (LPAS), talar width (TW), trochlear breadth (TrB), and breadth of the posterior articular facet (BPAS) were collected using digital sliding calipers (with a precision of 0.01mm), while the measurements of head-neck length (HNL) and talar height (TH) were collected with coordinate calipers (with a precision of 0.5mm) as sliding calipers could not capture these measurements accurately due to the bone's morphology. Additionally, measurements for trochlear length (TrL) and head height (HH) were repeated with adjustable calipers as the plane and/or points of measurement intended by the authors was unclear from the descriptions.

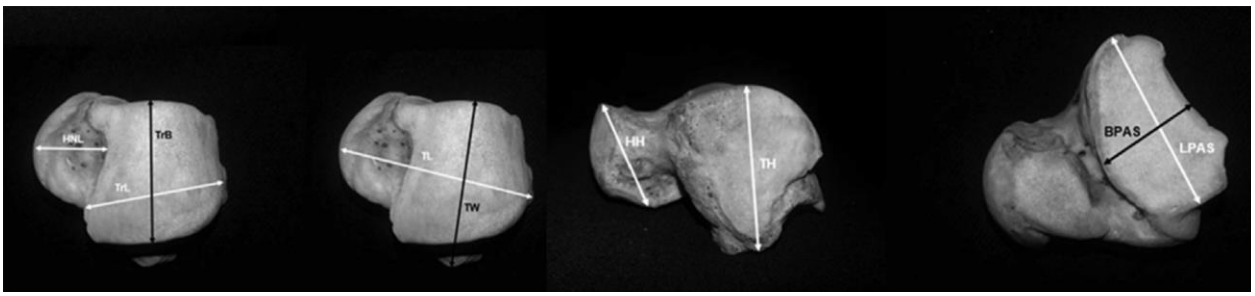


Figure 1: *Variables for Talus, Method 1. Image from Peckmann et al., 2015b*

For the calcaneus, measurements were collected for nine variables as described by Peckmann and colleagues (2015a). The variables were: maximum length (MAXL), load arm length (LAL), minimum breadth (MINB), middle breadth (MIDB), body height (BH), maximum height (MAXH), dorsal articular facet length (DAFL), dorsal articular facet breadth (DAFB), and cuboidal facet height (CFH). All measurements were taken with digital sliding calipers apart from MAXH, for which coordinate calipers were used. All measurements of load arm length (LAL) and body height (BH) were repeated using coordinate calipers.

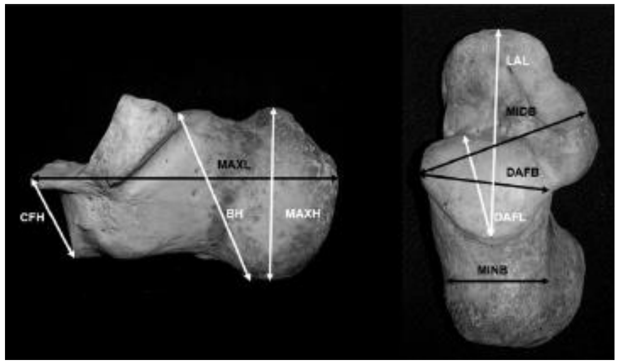


Figure 2: *Variables for Calcaneus, Method 1. Image from Peckmann et al., 2015a.*

For Method 2, variables were measured according to Curate and colleagues (2021) and written descriptions for all variables were obtained from the Instructions tab on the online tool *osteomics.com/CalcTalus/* (Appendix B.2). Six measurements were collected from the talus: maximum length of the talus (TM1), width of the talus (TM2), body height of the talus (TM3), maximum body height of the talus (TM3a), trochlear length (TM4), and trochlear width (TM5). Nine measurements were collected from the calcaneus: calcaneus maximum length (CM1), calcaneus length (CM1a), load arm width (CM2), load arm width (CLAL), body height (CM4), body length (CM5), maximum body height (CMBH), tuber calcanei height (CM7), and tuber calcanei width (CM8). TM3, TM3a, and CM4 were collected using coordinate calipers, and all other measurements were collected using digital sliding calipers.

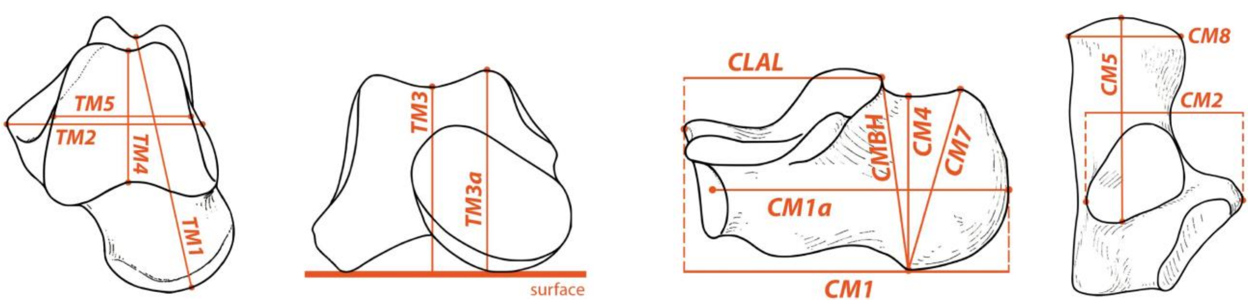


Figure 3: *Variables for Talus and Calcaneus, Method 2. Image from Curate et al., 2021.*

As the sample is composed of archaeological bones, it was expected that the bones may be damaged or incomplete. For both methods, variables were not measured if damage to the bone affected the measurement (see Appendix B.3). Therefore, analysis of the data also excluded any statistical functions involving damaged measurements.

Three trials were conducted for each method. Trial 1 and Trial 2 were conducted by Observer A (author) on two different days, and Trial 3 was conducted by Observer B (an experienced anthropologist) on a separate day. Observer B was provided with the measurement descriptions (Curate et al., 2021, Peckmann 2015a, 2015b, pers. comm. 2021;), the instructions available on osteomics.com/CalcTalus/, and directions regarding which calipers were to be used for each measurement, but did not receive any further instruction on how to take each measurement. Trial 3 does not exclude measurements affected by damage, so all of the statistical functions in each method were performed.

For comparative purposes, sex was also estimated by Observer A using traditional morphologic methods (Buikstra & Ubelaker, 1994). The pelvis was used where available as it is widely seen as a more universal indicator of sex due to its function in childbirth (Moore, 2013), while cranial morphology varies across populations (Buikstra & Ubelaker, 1994) and is affected by aging (Meindl et al., 1985). Moreover, sex estimates from the pelvis have been found to be more accurate (Meindl et al., 1985) and the accuracy of sex estimates from the skull has been challenged (Spradley & Jantz, 2011). The following traits were assessed on each pelvic bone where observable, according to Buikstra and Ubelaker (1994): ventral arc, subpubic concavity, ischiopubic ramus ridge, greater sciatic notch, and preauricular sulcus. As the pelvic bones of Gu 3 were too damaged to assess, sex was estimated using cranial morphology. The sex estimates obtained using the morphologic methods, as well as sex estimates by previous researchers, are found in Appendix D.2.

To prevent bias, the metric data obtained using methods 1 and 2 was not analyzed until both the metric and morphologic methods were completed.

For Method 1, the collected data were inputted into the multivariate discriminant function equations developed by Peckmann and colleagues (2015a, 2015b), with scores greater than 0 resulting in an estimate of male and scores less than 0 resulting in an estimate of female. Functions 1-4 are direct multivariate discriminant equations differentiated by the type(s) of variable used in each equation. Function 1 uses all variables. Function 2, composed of length variables, utilizes talar length (TL), head-neck length (HNL), trochlear length (TrL), and length of the posterior articular facet (LPAS) for the talus and maximum length (MAXL), load arm length (LAL), and dorsal articular facet length (DAFL) for the calcaneus. Function 3, for breadth variables, utilizes talar width (TW), trochlear breadth (TrB), and breadth of the articular facet (BPAS) for the talus and minimum breadth (MINB), middle breadth (MIDB), and dorsal articular facet breadth (DAFB) for the calcaneus. Function 4, for height variables, uses head height (HH) and talar height (TH) for the talus and maximum height (MAXH), body height (BH), and cuboidal facet height (CFH) for the calcaneus. Function 5 is a stepwise multivariate discriminant function equation that uses talar width (TW) and talar length (TL) for the talus, and maximum length (MAXL), dorsal articular facet length (DAFB), and cuboidal facet height (CFH) for the calcaneus. Although the general consensus in the literature is that multivariate functions are less biased for sex discrimination than univariate functions, it has been found that some univariate functions have comparable accuracy to multivariate functions (Curate et al., 2021; Steele, 1976) and are thus worth exploring due to their potential use in forensic and archaeological situations. The univariate functions identified by Peckmann and colleagues (2015a:379, 2015b:16) to be the most accurate for each sex in their original and cross-validated groups (talar length [TL], length of the posterior articular facet [LPAS], talar width [TW], trochlear length [TrL], and trochlear breadth [TrB]) were also applied to these data. A measurement less than the provided sectioning point was estimated to be female, while a measurement greater than the provided sectioning point was estimated to be male.

For Method 2, the collected data were inputted into the applicable statistical models on osteomics.com/CalcTalus/, a web application developed by the authors of Method 2 (Curate et al., 2021). See Appendix 4.B for more information about the CalcTalus tool. The applicable measurements were entered into the web application and a probabilistic estimate was provided for each model. The method provides linear regression (LR) and support vector machine (SVM) models that analyze tali and calcanei separately or together. No combined models (using both the talus and calcaneus) were run as it is unclear whether Gu DC talus and Gu DC calcaneus belong to the same individual. Three logistic regression and three support vector machine models were run for the talus, and two logistic regression and two support vector machine models were run for the calcaneus.

The measurements used by each model are provided in Appendix B.5. Of note for Method 2 is that the article by Curate and colleagues (2021) as well as the *Instructions* tab on *osteomics.com/CalcTalus/* describe 9 calcaneal measurements and 6 talar measurements, but some of these measurements are not used in any model available on the website (including the combined models not used in this study). The unused variables for the calcaneus are: CM1 (calcaneus maximum length), CM2 (load arm width), CLAL (load arm length), CM4 (body height), and CM7 (tuber calcanei length). The unused variables for the talus are: TM2 (width of the talus), TM3 (body height of the talus), and TM3a (maximum body height of the talus). It appears that these variables are unnecessarily included in the instructions, as they serve no purpose during the implementation of the models, but no explanation was provided by the authors.

2.3c Statistical Analysis

All statistical analyses were conducted on Microsoft Excel version 2109. Descriptive statistics consisting of the maximum, minimum, mean, standard deviation, and range were obtained for each method. Intraobserver error was calculated between trials 1 and 2 in each method, and interobserver error was calculated between trials 1 and 3 in each method. Calculations of error included the mean and median absolute differences (in mm), the mean difference relative to mean size (%), and the technical error of measurement (TEM). The formula for TEM is

$$\text{TEM} = \sqrt{\frac{\sum D^2}{2N}}$$

in which D is the difference between the two measurements and N is the total number of subjects (Harris & Case, 2012).

The differences between the left and right bones of the same individual were also compared for each measurement. The data obtained from Trials 1 and 2 were averaged, and the data for each variable were assessed for normality using QQ plots. The difference between measurements taken from the left and the right sides was then assessed using paired t-tests and a qualitative comparison of results.

2.4 Results

Table 1: *Summary of Sex Estimations by Method*

Individual	Bone	Method 1 *Peckmann et al., 2015a/b* Trial 1 (Obs. A)	Method 1 Trial 2 (Obs. A)	Method 1 Trial 3 (Obs. B)	Method 2 *Curate et al., 2021* Trial 1 (Obs. A)	Method 2 Trial 2 (Obs. A)	Method 2 Trial 3 (Obs. B)	*Buikstra & Ubelaker, 1994* Pelvis (Obs. A)	Notes
Gu 1	L Talus	F *	F *	F	F	F	F	F	
	R Talus	F	F *	F	F	F	F		
Gu 2	L Talus	M *	M *	M *	M	M	M	M	*Healed fracture – see Appendix B.6*
	R Talus	M *	M *	M *	M	M	M		
Gu 3	L Talus	IND *	IND *	M *	M	M	M	n/a	*Pelvis fragmented. Features for estimation damaged or missing. Cranial morphology indicates possible F.*
	R Talus	IND *	M *	M *	M	M	M		
Gu 6	L Talus	F	F	F *	F	F	F	Possible F	*Juvenile (many unfused long bone epiphyses)*
	R Talus	F	F	F	F	F	F		
Gu 7	L Talus	M *	M *	M *	M	M	M	M	
	R Talus	M *	M *	M *	M	M	M		
Gu DC	R Talus	IND *	M *	M *	M	M	M	Gu 8 and 9 are both Possible M	*Belongs to Gu 8 or 9*
	L Calcaneus	M	M *	F *	M	M	M		*Belongs to Gu 8 or 9*

Legend:
Obs. = observer
F = estimated female
M = estimated male
* = not all functions have provided the same result but the majority (greater than 50%) have provided the same result. See Tables 2 and 3 and Appendix C for more details.
IND = indeterminate. An estimate could not be provided as the same number of male and female results were provided. See Appendix C for more details.
n/a = could not be assessed (missing or damaged)

2.4a Results for Method 1

The final sex estimate for each bone (results in Table 1) is derived from the results of the multivariate and univariate functions developed by Peckmann and colleagues (2015a, 2015b). A detailed distribution of the results of each equation can be found in Appendix C.1. In many cases, the equations for a particular bone did not all yield the same result, so the final estimate was determined by the majority (i.e. the result indicated by greater than 50% of the equations used). In cases where the different equations resulted in an equal number of male and female estimates for the same bone, the estimate was marked "indeterminate".

Any equations involving variables taken with both digital sliding calipers and coordinate calipers were run twice, and all but one yielded the same result upon replacement. The only function that yielded a different result was Function 4 (height variables) for the DC calcaneus in Trial 3. The function, which uses body height (BH) and cuboidal facet height (CFH), gave a "Female" result when using BH taken with sliding calipers and "Male" when using BH taken with coordinate calipers. As there was a difference, both results are represented in the tally for Trial 3 in Table 3.

Indeterminate results were obtained for Trials 1 and 2 on Gu 3's left talus and for Trial 1 on Gu 3's right talus. Several measurements were not taken due to damage, resulting in some functions not being calculated. The functions that were calculated showed conflicting results, particularly between the functions deemed most accurate in Peckmann and colleagues' (2015b) article.

Gu 6 is estimated by Method 1 to be female. In practice, however, neither sex can be excluded as it is known that Gu 6 is a juvenile and was therefore not fully grown at the time of death.

It is unknown whether Gu DC talus and Gu DC calcaneus belong to the same individual, as Gu 8 was buried directly above Gu 9 in Grave DC. As such, the estimated sexes for the Gu DC talus and the GU DC calcaneus were compared to the morphological sex estimates for both Gu 8 and Gu 9 (Appendix D.2).

Table 2: *Summary of Results Obtained from Multivariate and Univariate Discriminant Functions for the Talus in Method 1. F : M denotes the number of female results to male results.*

F = Female, M = Male, IND = Indeterminate. Equations with measurements excluded due to damage resulted in N/A and are not counted in this table.

Individual	Side	Trial 1 (F : M)	Trial 1 Sex Estimate	Trial 2 (F : M)	Trial 2 Sex Estimate	Trial 3 (F : M)	Trial 3 Sex Estimate
Gu 1	**Left**	9:1	F	9:1	F	10:0	F
	Right	10:0	F	9:1	F	10:0	F
Gu 2	**Left**	1:9	M	1:9	M	1:9	M
	Right	1:9	M	1:9	M	1:9	M
Gu 3	**Left**	2:2	IND	2:2	IND	4:6	M
	Right	5:5	IND	4:6	M	3:7	M
Gu 6	**Left**	10:0	F	10:0	F	9:1	F
	Right	4:0	F	4:0	F	10:0	F
Gu 7	**Left**	4:6	M	3:7	M	3:7	M
	Right	3:7	M	2:8	M	2:8	M
Gu DC	**Right**	4:4	IND	2:6	M	4:6	M

Table 3: *Summary of Female to Male Results Obtained from Multivariate and Univariate Discriminant Functions for the Calcaneus in Method 1. F : M denotes the number of female results to male results.*

F = Female, M = Male, IND = indeterminate. Equations with measurements excluded due to damage resulted in N/A and are not counted in this table.

Individual	Side	Trial 1 (F : M)	Trial 1 Sex Estimate	Trial 2 (F : M)	Trial 2 Sex Estimate	Trial 3 (F : M)	Trial 3 Sex Estimate
Gu DC	**Left**	4:0	M	3:1	M	6:3	F

The results of each equation were compared for each bone between trials 1 and 2 and between trials 1 and 3. For the talus, six differences were identified between the results of trials 1 and 2, and seven differences were identified between the results of trials 2 and 3. For the calcaneus, one difference was identified between the results of trials 1 and 2, and three differences were identified between the results of trials 2 and 3. A detailed distribution of the discrepancies is available in Appendix C.1.

Descriptive statistics (Tables 4 and 5) were calculated for the data from the tali only, as there was only one calcaneus within the sample. Similarly, intraobserver error and interobserver error (Tables 6 and 7) were only calculated for the data from the talus. There was more interobserver error than intraobserver error for each variable with the exception of the breadth of the posterior articular facet (BPAS), which had a higher median and mean absolute difference between trials 1 and 2 than between trials 1 and 3, and talar width (TW), which had a higher median absolute difference. The variables with the lowest TEM between trials 1 and 2 were talar length (TL) and trochlear breadth (TrB), and the variables with the lowest TEM between trials 1 and 3 were trochlear breadth (TrB) and head-neck length (HNL). The variables with the highest TEM between trials 1 and 2 (intraobserver) were length of the posterior articular facet (LPAS) and breadth of the posterior articular facet (BPAS), which is reflected in the conflicting results between trials 1 and 2 for the Univariate LPAS function (Appendix C.1, Table 20). The variables with the highest TEM between trials 1 and 3 (interobserver) were head height (HH) and talar height (TH), which may explain the additional conflict between the trials for Gu 1 Left in Function 4 (which uses height variables) (Appendix C.1, Table 21). Despite the higher TEM and mean/median differences for interobserver error compared to intraobserver error, the comparisons between talus results for Trial 1 vs. Trial 2 and Trial 2 vs. Trial 3 show relatively equal numbers of disagreements.

Table 4: *Descriptive Statistics for Trials 1 and 2 of Method 1 (Observer A only)*

Variable	TL	HNL	TrL	LPAS	TW	TrB	BPAS	HH	TH
Max (mm)	58.98	27	38.63	38.03	49.49	37.12	26.28	33.3	39.5
Min (mm)	46.65	21	27.42	25.47	35.72	26.21	17.76	23.97	29
Range (mm)	12.33	6	11.21	12.56	13.77	10.91	8.52	9.33	10.5
Mean (mm)	53.13	23.34	33.02	30.95	41.80	32.35	21.74	28.12	34.61
SD	1.85	3.91	3.73	4.35	3.80	2.79	3.30	3.56	1.41

Table 5: *Descriptive Statistics for Trials 1, 2, and 3 of Method 1*

Variable	TL	HNL	TrL	LPAS	TW	TrB	BPAS	HH	TH
Max (mm)	64.53	27	42.29	39.43	49.75	37.12	26.53	33.31	46.5
Min (mm)	46.65	21	27.28	25.47	35.72	26.21	17.76	22.66	29
Range (mm)	17.88	6	15.01	13.96	14.03	10.91	8.77	10.65	17.5
Mean (mm)	53.38	23.50	33.17	31.21	41.74	32.48	22.00	27.79	35.15
SD	4.90	1.85	4.04	3.65	4.39	3.72	2.89	3.51	4.18

Table 6: *Intraobserver Error (Trials 1 and 2) for Method 1*

Variable	TL	HNL	TrL	LPAS	TW	TrB	BPAS	HH	TH
TEM (mm)	0.15	0.46	0.37	0.85	0.53	0.10	0.82	0.54	0.49
Median Absolute Difference (mm)	0.12	0.50	0.11	0.38	0.66	0.14	0.57	0.545	0.50
Mean Absolute Difference (mm)	0.16	0.59	0.34	0.87	0.63	0.13	0.92	0.64	0.59
Mean Difference as a Percent of Mean Size (%)	0.31	2.53	1.03	2.80	1.50	0.41	4.21	2.26	1.71

Table 7: *Interobserver Error (Trials 1 and 3) for Method 1*

Variable	TL	HNL	TrL	LPAS	TW	TrB	BPAS	HH	TH
TEM (mm)	1.28	0.60	1.17	1.15	0.78	0.45	0.92	2.46	1.80
Median Absolute Difference (mm)	0.24	0.50	0.53	1.58	0.55	0.42	0.26	2.39	1.50
Mean Absolute Difference (mm)	0.85	0.73	1.02	1.50	0.85	0.49	0.83	2.84	1.91
Mean Difference as a Percent of Mean Size (%)	1.58	3.08	3.08	4.79	2.03	1.51	3.76	10.38	5.37

From the QQ plots created for each talar variable (Appendix C.2), a normal distribution could generally be assumed. None of the paired t-tests (Table 8) demonstrate a statistically significant difference (α = 0.05) between measurements of the same variable taken from the left and right bones of the same individual. The effect of symmetry was also assessed qualitatively based on the agreement of left and right estimates for each multivariate and univariate function with two disagreements in trials 1 and 2, and three in trial 3. All three trials resulted in a disagreement in the estimate between the left and right tali of Gu 7 in Function 2, and trials 2 and 3 resulted in a disagreement for Gu 3 in the univariate function for TrL. See more in Appendix C.3.

Table 8: *Results of Paired T-tests Comparing Method 1 Measurements From the Left and Right Tali of the Same Individual*

Variable	TL	HNL	TrL	LPAS	TW	TrB	BPAS	HH	TH
p-value from paired t-test	0.191	0.280	0.376	0.676	0.308	0.479	0.151	0.274	0.426
df	4	4	4	3	3	4	4	4	4

2.4b Results for Method 2

The sex estimates resulting from Method 2 are shown in Table 1. Although the exact probabilities resulting from each model differed, all models used for the same individual generally agreed on the same estimate. The results for each individual are provided in Appendix C.4.

Models LR_T3 and SVM_T3 were not run for Gu 3 in trials 1 and 2 as TM5 was excluded due to damage. In all three trials, Gu 6 was estimated to have a high probability of being female. As this individual is a juvenile, neither sex can be excluded.

For the vast majority of models, either the male or female probability was greater than 0.90. Model SVM_T3 for Gu 1 (Figure 4) is an example of one of the only models in which the sex that makes up the majority of the probability (female, in this case) has a lower probability relative to the sex that makes up the majority in the other models. In conjunction with the probabilities provided by the other models, however, one could reasonably estimate the sex of this individual to be female.

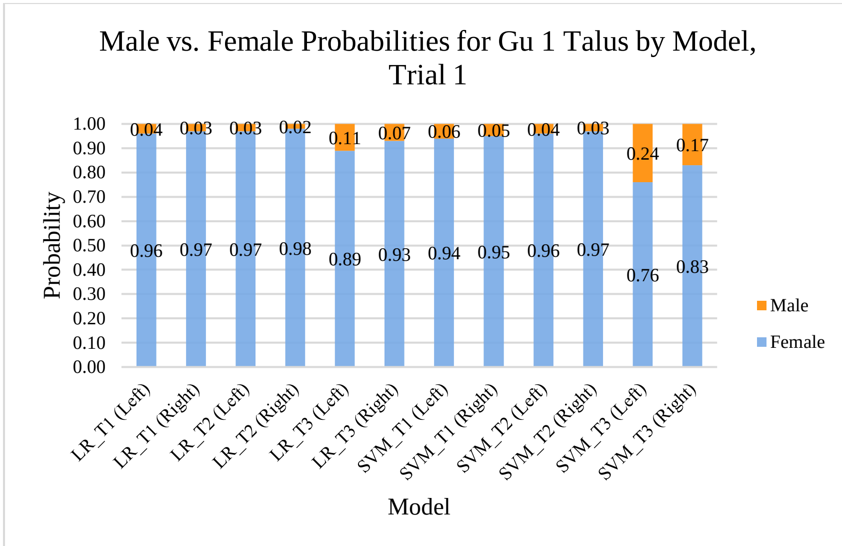


Figure 4: *Male vs. Female Probabilities for Gu 1 Talus by Model, Trial 1. Probabilities obtained from osteomics.com/CalcTalus/, figure by author.*

The results for Gu DC calcaneus (Figure 5) also show that the estimate (male) is supported by lower probabilities in LR_C1 and SVM_C1. This is because both models use only the measurement CMBH (maximum body height), but this measurement was impacted by damage to the posterior articular facet on the Gu DC calcaneus. Although the measurement was impacted by damage, it was still taken as it was required for every model applicable to this bone. In the event that this bone yielded a female estimate, neither sex could be excluded as it would be unknown whether the full measurement would yield a male estimate. These models, however, show that the individual is more probably male, and an estimate of male is also strongly supported by LR_C2 and SVM_C2, which incorporate CM8 (tuber calcanei width) into their predictions.

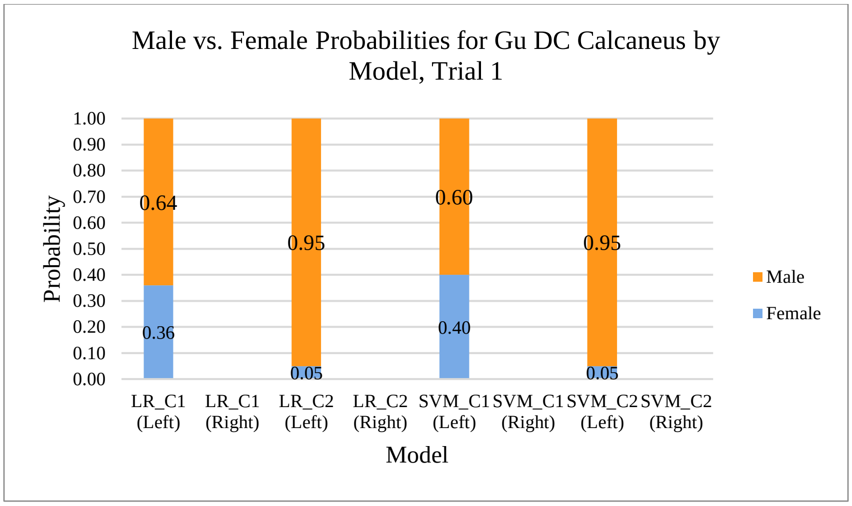


Figure 5: *Male vs. Female Probabilities for Gu DC Calcaneus by Model, Trial 1. Probabilities obtained from osteomics.com/CalcTalus/, figure by author.*

Descriptive statistics, intraobserver error, and interobserver error were calculated using data from the talus only as there was only one calcaneus in the sample. For each variable, interobserver error was higher than intraobserver error. TM1 (maximum length of the talus) and TM2 (width of the talus) demonstrated the lowest error for both intraobserver and interobserver comparisons. The greatest error was found in TM3 (body height of the talus) and TM3a (maximum body height of the talus), which were not used in any of the models applied to this sample. Although the probabilities obtained from any given model differed between the three trials, an overall sex estimate was clear for each individual and all trials resulted in the same final estimate for each bone (Table 2) with the exception of the indeterminate results.

Table 9: *Descriptive Statistics for Trials 1 and 2 of Method 2 (Observer A)*

Variable	**TM1**	**TM2**	**TM3**	**TM3a**	**TM4**	**TM5**
Max (mm)	59.01	45.29	38	41	40.61	37.17
Min (mm)	46.94	35.6	27.5	30	27.8	25.05
Range (mm)	12.07	9.69	10.5	11	12.81	12.12
Mean (mm)	53.21	40.95	33.30	35.25	33.43	31.99
SD	4.67	3.67	3.25	3.57	4.14	4.23

Table 10: *Descriptive Statistics for Trials 1, 2, and 3 of Method 2*

Variable	TM1	TM2	TM3	TM3a	TM4	TM5
Max (mm)	59.01	45.29	38	42	40.61	37.17
Min (mm)	46.94	35.4	27.5	28.5	25.24	25.05
Range (mm)	12.07	9.89	10.5	13.5	15.37	12.12
Mean (mm)	53.22	40.98	33.42	35.64	33.15	32.14
SD	4.54	3.63	3.52	3.96	4.08	4.18

Table 11: *Intraobserver Error (Between Trials 1 and 2) for Method 2*

Variable	TM1	TM2	TM3	TM3a	TM4	TM5
TEM (mm)	0.12	0.28	0.44	0.51	0.31	0.34
Median Absolute Difference (mm)	0.06	0.29	0.50	0.50	0.12	0.39
Mean Absolute Difference (mm)	0.12	0.34	0.50	0.50	0.29	0.39
Mean Difference as a Percent of Mean Size (%)	0.22	0.28	1.50	1.42	0.88	1.21

Table 12: *Interobserver Error (Between Trials 1 and 3) for Method 2*

Variable	TM1	TM2	TM3	TM3a	TM4	TM5
TEM (mm)	0.29	0.55	1.59	2.07	1.12	0.55
Median Absolute Difference (mm)	0.24	0.61	2.00	2.50	0.49	0.60
Mean Absolute Difference (mm)	0.33	0.65	1.95	2.73	1.03	0.69
Mean Difference as Percent of Mean Size (%)	0.62	0.80	5.83	7.60	3.12	2.15

Measurements from the left and right bones of the same individual were compared for all variables except TM2 (width of the talus), as this measurement was excluded for at least one bone in three out of the five individuals. Based on the QQ plots created for each variable (Appendix C.5), the data for each of the remaining variables could be assumed to be normally distributed. None of the paired t-tests (Table 13) demonstrate a statistically significant difference ($\alpha = 0.05$) although TM3a (maximum body height of the talus) may be considered to be approaching significance. Given that TM3a was found to have high interobserver and intraobserver error, however, it is not convincing that this indicates anything meaningful about left vs. right symmetry particularly in light of the small sample size. Additionally, through visual assessment of the probability graphs produced by Method 2, it is clear that differences between the left and right bones of the same individuals, whether due to observer error or asymmetry, are not substantial enough to cause different sex estimates for the left and right bones in the same individual.

Table 13: *Results of Paired T-tests Comparing Method 2 Measurements From the Left and Right Tali of the Same Individual*

Variable	**TM1**	**TM3**	**TM3a**	**TM4**	**TM5**
p-value from paired t-test	0.344	0.235	0.099	0.846	0.716
df	4	4	4	4	3

2.5 Discussion

2.5a Discussion of Methodology

Metric methods are widely accepted to be more user-friendly than many morphologic methods because their accuracy is less dependent on the examiner's expertise and previous experience, but few independent studies have been conducted to support this claim. In this study, it was found that the application of metric methods still necessitated familiarity with bone features and anatomical terminology, and application became much faster with practice. Moreover, it is evident that not all metric methods are equally user-friendly; the user-friendliness of a method is affected by factors such as the clarity of the variables communicated via written explanations and photos or diagrams, as well as the practical ease of measuring the variables.

Depending on the clarity of the written descriptions and diagrams provided by the author(s) of a method, users may have to make judgment calls when taking measurements. For instance, discrepancies were found between several variable descriptions (TrL, TL, BH) provided for Method 1 (Peckmann et al., 2015a, 2015b, pers. comm 2021) and their corresponding images (see Appendix D.1). Additionally, the clarity of images or diagrams can impact the user's understanding of a variable. For example, the posterior point of the talar length variable (TL in Method 1, TM1 in Method 2) clearly extends to the groove for the *M. flexor hallucis longus* muscle in Method 2's diagram but is unclear in the Method 1 diagram (Figure 6). Peckmann and colleagues' (2015a) article also does not provide written descriptions of the measurements, thus necessitating heavy reliance on images unless a user were to contact the author(s) for the descriptions as was the case for this study. Since some features may be difficult to identify in photographs from certain perspectives, future authors may consider showing variables from multiple angles (such as with the use of rotatable 3D images) or using a mixture of diagrams and photos to depict their measurements.

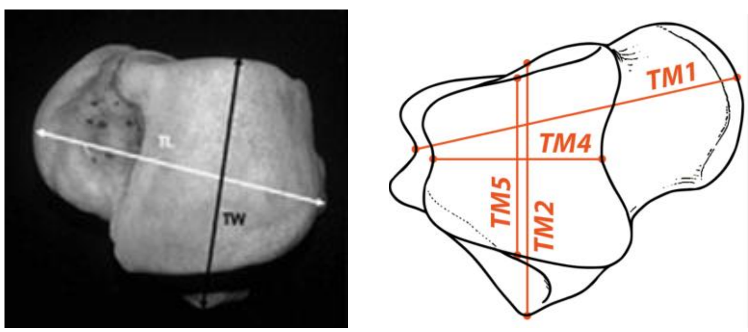


Figure 6: *Comparison of Talar Length Variable Diagrams*
6a (left): *Image of Talar Length (TL) Variable from Method 1, Peckmann et al. (2015b).*
6b (right): *Image of Maximum Length of Talus (TM1) Variable from Method 2 Curate et al. (2021)*

Another discrepancy between various tarsal sex estimation methods is seen in the instruments used to take the measurements. Specifically, Peckmann and colleagues (2015b) state that all measurements were taken using a digital vernier caliper, while Bidmos and Dayal (2003), who use the same variables, state that all measurements were taken with a digital vernier caliper with the exception of talar height (TH), head-neck length (HNL), and head height (HH), which were taken using a manual vernier caliper. In the description provided by Peckmann (2021, pers. comm.), however, head-neck length (HNL) was specified to be taken using a sliding caliper with moving arms. More detailed descriptions of which instruments should be used for each variable and how they should be used are necessary to decrease the number of judgment calls required of users.

In both methods, the plane on which variables should be measured was not always clearly described or demonstrated in the photos. An example of this is seen for the variable trochlear length, which is only shown from a superior perspective in each method (Figures 7a and 7b). Method 1 (Peckmann et al., 2015b) describes the variable as the "distance between the two intersecting points of the middle sagittal curvature of the trochlea with the anterior and posterior edge of the superior facet". Method 2, (Curate et al., 2021) describes the variable as the "maximum length of the trochlear articular surface on the midline measured parallel to the sagittal axis of the trochlea". Given the anteroposterior curvature of the trochlea, the posterior point is vertically inferior to the superior point. Therefore, the variable would result in two different measurements if taken using sliding calipers (shown in blue) versus coordinate calipers (yellow) (Figure 7c). Similar issues can be avoided with the use of more informative descriptions and diagrams.

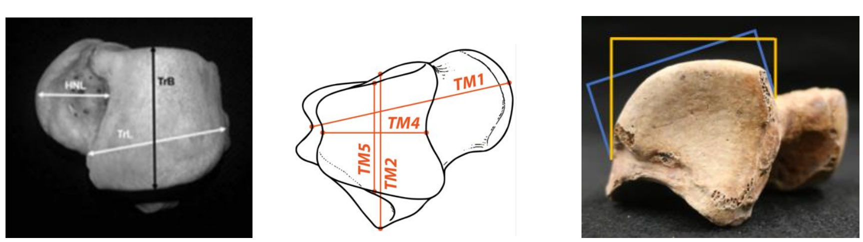


Figure 7: *Trochlear Length Variable*
7a (left): *Image of Trochlear Length (TrL) Variable from Method 1, Peckmann et al. (2015b).*
7b (centre): *Image of Trochlear Length (TM4) from Method 2, Curate et al. (2021).*
7c (right): *Image of Two Hypothetical Measurements Using Sliding Calipers (blue) and Coordinate Calipers (yellow). Photograph by author.*

Some variables were inherently more difficult to measure, and this is reflected in the heightened error (Tables 6, 7, 11, 12). For example, the variable of talar head height (HH) in Method 1 was difficult to measure using digital calipers because the superior and inferior points between which this measurement is taken are often diagonal from one another on a curved surface (Figure 8), leading to shifting as the calipers close in.

Figure 8: *Image of Gu 3 Talus. Annotations (in orange) demonstrate the superior and inferior points of the head between which variable HH is measured.*

Moreover, the use of coordinate calipers to measure height variables between a point on the talus and the surface on which the bone was resting was challenging as the scale on the calipers had to be held precisely perpendicular to the surface. A similar issue was found with the measurement of head-neck length (HNL) in Method 1, although for HNL the scale on the coordinate calipers was held parallel to the surface on which the bone was resting. Lastly, wear or damage to the posterior articular facet in some cases made it difficult to identify the edges of the facet for the length of the posterior articular facet (LPAS) and breadth of the posterior articular facet (BPAS) measurements in Method 1. In general, Method 1 contained more variables that were difficult to measure compared to Method 2.

On a broader level, the application of these two methods demonstrates that, to truly be user-friendly, there needs to be greater standardization in which variables are used and how variables are described. In reviewing several previously developed methods for the talus and calcaneus (Abd-elaleem et al., 2012; Curate et al., 2021; DiMichele & Spradley, 2012; Lee et al., 2012, Gualdi-Russo, 2007; Peckmann et al., 2015a, 2015b, Mahakkanukrauh et al., 2014; Scott et al., 2017), it is evident that the variables used often differ from one method to the next. Although some variables such as length, width, and height of the talus and calcaneus are used across many studies, some studies introduce new variables such as minimum inferior interarticular distance for the talus (Mahakkanukrauh et al., 2014) and posterior circumference for the calcaneus (DiMichele & Spradley, 2012).

Most studies using the talus and/or calcaneus tend to cite one or more of the following in describing their measurements: Bidmos and Dayal (2003, 2004), Bidmos and Asala (2003, 2004), Harris & Case (2012), Martin (1928), Martin & Knussman (1988), and Steele (1976). Variable descriptions, however, can be both literally and figuratively lost in translation. Martin (1928) and Martin and Knussman (1988) describe the variables in German, and scholars not fluent in German would therefore be reliant on diagrams and translations if using a method that cites one of these sources and does not provide its own descriptions. Even the most commonly used measurements, such as talar length, are defined in various ways. For example, Peckmann and colleagues (2015b) measure talar length according to Steele (1976) (between the most anterior point on the head and the groove for the M. flexor hallucis longus), while Lee and colleagues (2012) include the lateral tubercle of the talus as per Bidmos and Dayal's (2003) method. Such discrepancies become problematic for user interpretation if the authors cite more than one source for the variables but do not explicitly link each variable to a citation, particularly when in-article descriptions and diagrams are missing or unclear.

Given limitations in word count for published articles, images often become essential to a user's understanding and application of a method. It is evident, however, that the quality and consistency of images can vary. An example of a pictorial discrepancy between articles using the same variable names is shown in Figures 9a and 9b.

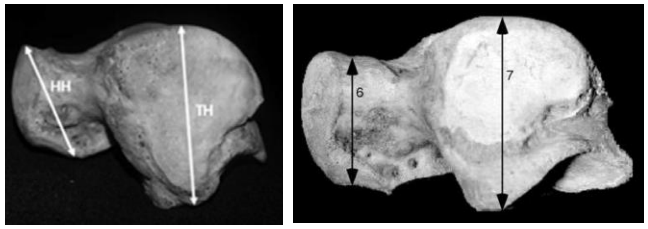


Figure 9: *Head Height Variable Diagram Comparison*
9a (left): *Variable of Head Height (arrow HH) in Method 1, Peckmann et al., (2015b).*
9b (right): *Variable of Head Height (arrow 6) in Bidmos & Dayal, (2004).*

Apart from posing issues to ease of understanding and application, a lack of standardization for variables is problematic when researchers seek to use their own data in other methods or vice versa. For example, to investigate the population specificity of metric methods on the talus, Peckmann and colleagues (2015b) compared their data with those from three other studies which assessed the same variables on different populations: Koreans (Lee et al., 2012), White South Africans (Bidmos & Dayal, 2003), and Black South Africans (Bidmos and Dayal (2004). Peckmann and colleagues (2015b) found significant differences in the mean values of the variables when comparing the Greek data to the Korean, White South African, and Black South African populations, as well as much lower accuracy rates when entering their own (Greek) data into the functions developed by the other authors, leading the authors to conclude that this is evidence for population specificity of the method. As detailed above, however, there are at least two methodological differences evident between the Peckmann and colleagues article and some of the articles with which they compare, which raises the question of whether differences found between the studies can be entirely attributed to population variation.

In summary, there is a need for greater standardization of common variables and more informative descriptions in terms of written definitions, diagrams, and citations for each variable. Increased standardization would contribute to the user-friendliness of these methods, the ability for authors to meaningfully compare data and results across methods, as well as the ability for researchers to quickly apply the same data to multiple methods if the population group of the individual(s) is uncertain. A publication of standards for the more recent metric methods, akin to Buikstra and Ubelaker's (1994) publication, accompanied by a database of 3D images demonstrating the variables, would be useful. This call for standardization should not discourage researchers from exploring new variables that may enhance the use of these methods, but rather provide a basis and an exemplar for future investigations.

2.5.b Discussion of Results

Sex Estimates

The final sex estimates obtained from Method 1 and Method 2 are generally in agreement with each other and with the morphologic estimates performed in this study and two previous studies (Meijer, 2018; Seymour, 2019) (Appendix D.2).

In Method 1, Function 1 (which uses all nine talar variables) was found by Peckmann and colleagues (2015b) to be most accurate for their original group of Greek tali but, interestingly, resulted in "Female" for each individual in every trial of this study. Moreover, in Peckmann and colleagues' cross-validated group, systematic error was found in that all of the functions resulted in 100% accuracy for males and lower accuracies for females. This means that the functions would never misclassify a true male but would sometimes misclassify true females, but the same is not seen in the results of this study (with the assumption that the morphologic sex estimations are accurate). All of the males in this study received a female result from at least one function and often more, while the females received very few male results. These discrepancies suggest that (barring any methodological differences between this study and Peckmann and colleagues') the method developed for contemporary Greeks is not entirely suitable for the archaeological French skeletons, which supports Peckmann and colleagues' (2015b) conclusion that population variation impacts the results of this method. Such variation may result from intrinsic factors (e.g., genetics) or extrinsic factors (e.g., the effect of modern shoes on tarsal morphology). This study also suggests, however, that if the results of multiple functions are considered, accurate sex estimates may still be obtained even when a method is applied to individuals who do not belong to the intended population. More research using larger samples is needed to substantiate this claim.

In Method 2, the various models used for the same individual showed much greater agreement with each other compared to the functions in Method 1, so it was much simpler to arrive at a final sex estimate for each individual. Additionally, the probabilistic nature of Method 2's estimates allow for trends in the data to be more easily seen. Since Method 1 uses sectioning points in each function to place individuals into a discrete "male" or "female" category, the estimate does not account for how close the raw score of the multivariate or univariate function is to the other estimate. As demonstrated in Appendix D.3, the Function 1 raw results covered a wide range and some were quite close to 0, but all individuals were estimated to be female as all scores were ultimately below 0. The use of discrete categories in metric sex estimation may obfuscate variation that actually exists, and create confusion when different functions provide varying results for more borderline individuals. In this regard, Method 2's probabilistic estimates are advantageous as they allow users to make more informed decisions about the strength of an estimate (e.g., 0.94 female and 0.06 male vs. 0.60 female and 0.40 male) and, in turn, how much weight to give this estimate in their interpretation of the data. Therefore, it may be beneficial for authors developing metric methods in the future to explore the use of probabilistic estimates which allow users to make more informed decisions about their results.

Error

In evaluating error, the variables measured using coordinate calipers (head-neck length [HNL] and talar height [TH] in Method 1, body height [TM3] and maximum body height [TM3a] in Method 2) unsurprisingly showed relatively high error due to the lower precision of the coordinate calipers (0.5mm) compared to the digital sliding calipers (which measured to the nearest 0.01mm).

For both methods, interobserver error was generally higher than intraobserver error. According to standards in forensic anthropology, a TEM of less than 1mm is acceptable (Colman et al., 2018; Papaioannou et al., 2011). In the intraobserver comparison, all variables in each method had a TEM under 1mm. In the interobserver comparison for Method 1, five out of nine variables had a TEM greater than 1mm (talar length [TL], trochlear length [TrL], length of the posterior articular facet for the calcaneus [LPAS], talar height [TH], and he HH), one of which was greater than 2mm (HH at 2.46). In the interobserver comparison for Method 2, three variables out of six had a TEM greater than 1mm (body height [TM3], maximum body height [TM3a], trochlear length [TM4]), one of which was greater than 2mm (TM3a at 2.07). By these measures, the methods exhibit comparable error as approximately half of the variables in each method resulted in an interobserver TEM above the acceptable threshold. It is important to note that Method 2 does not actually use variables TM3 and TM3a (which happen to have the highest error) in any of its statistical models. Method 2 is also less meaningfully affected by error due to the probabilistic nature of its estimates, while the error in Method 1 caused conflicting male/female results in several functions with more borderline raw scores.

In analyzing error, Peckmann and colleagues performed paired t-tests and stated that “the results of the intra- and inter-observer error analyses show that none of the comparisons were statistically significant therefore there were no differences within or between observers” (2015a:381, 2015b:17). The findings of this current study, however, raise questions regarding what conclusions can be drawn from significance testing. The results challenge the notion that a lack of statistical significance is sufficient to conclude that error has no effect, but more studies with larger samples are required to evaluate the relationship between statistical significance and the effects of error on results. This issue is relevant to a greater debate within the discipline (and research in general) regarding how significance testing should be used and interpreted (Amrhein et al., 2019; Nuzzo, 2014; Smith, 2018, 2020; Weiss, 2011). Furthermore, the sources of error found in this study demonstrate the necessity for independent validation studies using observers who were not affiliated with the creation of the method and are relying solely upon the instructions that are available to external users of the method.

Symmetry

The issue of bilateral symmetry is often addressed by developers of metric methods, and it impacts the methodology used in the development of these methods. Although it is generally accepted that the lower limbs are not perfectly symmetrical, the extent to which various authors consider this asymmetry as relevant to metric methods has resulted in different approaches to addressing symmetry. For example, Curate and colleagues (2021) acknowledged asymmetry by citing previous literature (Gualdi-Russo, 2007; Silva, 1995) and averaged the measurements in the left and right bones to develop their method. Peckmann and colleagues (2015a, 2015b) compared measurements taken from a subsample of paired bones and found no statistically significant differences, thus proceeding to develop their methods using measurements from left bones only and substituting a right bone when the left of the pair was missing or damaged.

It seems that although most authors acknowledge the issue of asymmetry in developing their methodology, few test their developed methods for the impacts of bilateral asymmetry. A study by Michiue and colleagues (2018), however, which developed a sex estimation method for the patella noted bilateral differences in the accuracy of their method for both males and females. This thesis sought to uncover more insights into the impact (or lack thereof) of asymmetry on previously developed metric methods. For Method 1, no statistically significant differences were found between measurements for the same variables on the left and right bones, but a qualitative assessment of the results of each function revealed several disagreements in each trial between the left and right bones in the same pair although the final estimates (accounting for results from all functions) for the left and right sides remained in agreement (Appendix C.3). The variables of Method 2 also did not exhibit statistically significant bilateral differences and any small differences in the probabilistic estimates provided for the left and right bones in each pair were not substantial enough to cause different sex estimates.

Harris and Case (2012) found that the variables with significant side asymmetry generally also had the greatest error. In this study, the result of the paired t-test for variable TM3a was relatively low ($p = 0.099$) but this variable also exhibited high intraobserver and interobserver error. This demonstrates that the effects of asymmetry and error may be conflated, and raises an interesting question about how error should be accounted for in further investigations of asymmetry.

Given the conflicting views in how bilateral asymmetry should be addressed in the development of metric methods and the paucity of independent validation studies that explore the tangible impacts of asymmetry on developed methods, it is clear that further investigation into the impacts of asymmetry is required. Future research should seek to evaluate the presence of impacts from error and symmetry on sex estimates and explore the relationship between such impacts and the presence of statistical significance.

2.5c Limitations

Many of the limitations in this study are related to the small sample size. As there was only one calcaneus, statistical analyses comparing error and symmetry could not be performed. Moreover, Steele (1976) found that while functions for the talus were more accurate than for the calcaneus, the most accurate functions use both bones. A lack of paired tali and calcanei and the resulting inability to use the combined models for Method 2 prevented an investigation of this aspect. Furthermore, as the sexes of the individuals are not known and the sample size is small, accuracy rates cannot be compared with the original studies. This study, therefore, cannot make any conclusive statements regarding whether the methods are suitable for a French population. The small sample size also limits the conclusions that can be drawn from analyses of error and symmetry.

Another limitation of this study is secular change, as the methods that have been applied to the archaeological population were developed using contemporary skeletons. Secular change in height and weight across the last several centuries has been demonstrated in contemporary Greek and Portuguese populations (Cardoso, 2008; Papadimitriou et al., 2008; Weisensee & Jantz, 2011). Moreover, it has been found in an American population that lower limb and distal bones are more susceptible to secular change (Jantz & Jantz, 1999). As the tarsals are distal bones of the lower limb, and their sexual dimorphism is hypothesized to reflect dimorphism in weight, further research into the effect of secular change on the use of metric methods for the archaeological tali and calcanei is required.

2.6 Conclusion

Ubelaker and DeGaglia (2017) state that insights into variation can be gathered when methods are tested on samples and populations other than the ones with which they were developed. Through this study's test of two metric methods on a French population, it has become evident that greater standardization is required in both the development and communication of metric methods for greater comparability between studies and in an effort to reduce user error. This study also raises questions regarding how symmetry and error can be conceptualized and evaluated, and draws attention to the necessity for independent validation studies of metric methods. Future research in this area should validate methods on larger samples of archaeological and contemporary remains. Additional future directions may include further study into the impacts and extents of population specificity as well as the impact of various pathologies, such as antemortem fractures, on the use of metric methods.

References

Abd-elaleem SA, Abd-elhameed M, Ewis AA. 2011. Talus measurements as a diagnostic tool for sexual dimorphism in Egyptian population. *Journal of forensic and legal medicine* **19**(2): 70-76. DOI:10.1016/j.jflm.2011.12.003

Agarwal SC. 2012. The past of sex, gender, and health: Bioarchaeology of the aging skeleton. *American Anthropologist* **114**(2): 322-335. DOI:10.1111/j.1548-1433.2012.01428.x

Agarwal SC. 2016. Bone morphologies and histories: Life course approaches in bioarchaeology. *American Journal of Physical Anthropology* **159**: 130-149. DOI:10.1002/ajpa.22905

American Association of Physical Anthropologists. 2003. Code of Ethics of the American Association of Physical Anthropologists (Approved by the AAPA Membership at the annual business meeting on April 25, 2003). https://physanth.org/documents/3/ethics.pdf

Amrhein V, Greenland S, McShane B. 2019. Retire statistical significance. *Nature* **567**: 305–307.

Barrett C, Cavallari W, Sciulli PW. 2001. Estimation of sex from the talus in prehistoric Native Americans. *Collegium Antropologicum* **25**(1): 13-9.

Barrio PA, Trancho GJ, Sánchez JA. 2006. Metacarpal sexual determination in a Spanish population. *Journal of Forensic Sciences* **51**(5): 990-995. DOI:10.1111/j.1556-4029.2006.00237.x

Bidmos MA, Asala SA. 2003. Discriminant function sexing of the calcaneus of the South African Whites. *Journal of Forensic Sciences* **48**: 1213e8.

Bidmos MA, Asala SA. 2004. Sexual dimorphism of the calcaneus of South African blacks. *Journal of Forensic Sciences* **49**(3): 446-450. DOI:10.1520/JFS2003254

Bidmos MA, Dayal MR. 2003. Sex determination from the talus of South African whites by discriminant function analysis. *The American Journal of Forensic Medicine and Pathology* **24**(4): 322-328. DOI:10.1097/01.paf.0000098507.78553.4a

Bidmos MA, Dayal MR. 2004. Further evidence to show population specificity of discriminant function equations for sex determination using the talus of South African blacks. J Forensic Sci 49:1–6. DOI:10.1520/jfs2003431

Buikstra JE. 2006. Chapter 15 Repatriation and Bioarchaeology: Challenges and Opportunities. In *Bioarchaeology: The Contextual Analysis of Human Remains,* Buikstra JE, Beck LA (eds.). Amsterdam: Academic Press an imprint of Elsevier: 389-415.

Buikstra J, Ubelaker D. 1994. *Standards for Data Collection from Human Skeletal Remains.* Arkansas Research Series No 44.

The Canadian Association for Physical Anthropology. 2019. Code of Ethics. https://capa-acap.net/sites/default/files/basic-page/capa-acap_code_of_ethics_-_jan_2019.pdf

Cardoso HFV. 2008. Secular changes in body height and weight of Portuguese boys over one century. *American Journal of Human Biology* **20**(3): 270-277. DOI:10.1002/ajhb.20710

Caspari R. 2003. From types to populations: A century of race, physical anthropology, and the American anthropology association. *American Anthropologist* **105**(1): 65-76.

Clements V, Gruspier K. n.d. An Analysis of skeletal material from the rock-cut church at Gurat. *Unpublished draft.*

Colman KL, Janssen MCL, Stull KE, van Rijn RR, Oostra RJ, de Boer HH, van der Merwe AE. 2018. Dutch population specific sex estimation formulae using the proximal femur. *Forensic Science International* **286**: 268.e1-268.e8. DOI:10.1016/j.forsciint.2017.12.029

Curate F, d'Oliveira Coelho J, Silva AM. 2021. CalcTalus: an online decision support system for the estimation of sex with the calcaneus and talus. *Archaeological and Anthropological Sciences* **13**: 74. DOI:10.1007/s12520-021-01327-y

Davadie A, Koehler H. 2021. Human bones, archaeology and interdisciplinary research in France: From excavation to conservation for research. *Ethics, Medicine, and Public Health* **18**: 100675. DOI:10.1016/j.jemep.2021.100675

DeWitte SN. 2015. Bioarchaeology and the ethics of research using human skeletal remains: Bioarchaeology and ethics. *History Compass* **13**(1): 10-19. DOI:10.1111/hic3.12213

Domjanic J, Seidler H, Mitteroecker P. 2015. A combined morphometric analysis of foot form and its association with sex, stature, and body mass. *American Journal of Physical Anthropology* **157**(4):582-591. DOI:10.1002/ajpa.22752

Eleveth PB, Tanner JM. 1990. *Worldwide Variation in Human Growth* (2nd ed.). Cambridge University Press: Cambridge.

Eliopoulos C, Moraitis K., Vanna V, Manolis, S. 2011. Greece/ΕΛΛΑΔΑ. In *The Routledge Handbook of Archaeological Remains and Legislation: An international guide to laws and practice in the excavation and treatment of archaeological human remains,* Marquez-Grant N, Fibiger L (eds.). Routledge: London; 173-183.

Franklin, U and Gervers, M. 1978. Technical analysis of coinage from the medieval rock-cut church at Gurat, France. *Journal of the International Institute for Conservation* **3**(2): 2-6.

Fessler DMT, Haley KJ, Lal RD. 2005. Sexual dimorphism in foot length proportionate to stature. *Annals of Human Biology* **32**(1):44-59. DOI:10.1080/03014460400027581

Gaherty C, Gretenhart CEC, Gretenhart FH. 1978. Skeletal remains from a medieval cemetery, Gurat, Charente. *Unpublished draft.*

Gervers M, Gervers V. 1975. Cliffs, caves and hermits: The rock-cut church at Gurat. *Rotunda* **VIII**: 5–13.

Gervers M. 1967. The Cave Church at Gurat (Charente): Preliminary Report. *Gesta* **6**: 10-20.

Gervers M. 1976. Église rupestre de Gurat: compte-rendu de fouille, 1974. *Mémoires de la Société Archéologique et Historique de la Charente* **1975/76**: 135-156

Gilchrist R. 2014. Monastic and Church Archaeology. *Annual Review of Anthropology* **43**(1): 235-250. DOI:10.1146/annurev-anthro-102313-025845

Gualdi-Russo E. 2006. Sex determination from the talus and calcaneus measurements. *Forensic science International* **171**(2): 151-156. DOI:10.1016/j.forsciint.2006.10.014

Harris SM, Case DT. 2012. Sexual dimorphism in the tarsal bones: implications for sex determination. *Journal of Forensic Sciences* **57**(2): 295-305. DOI:10.1111/j.1556-4029.2011.02004.x

Holm S. 2001. The Privacy of Tutankhamen – Utilising the genetic information in stored tissue samples. *Theoretical Medicine and Bioethics* **22**(5): 437-449. DOI:10.1023/A:1013010918460

Hudson A, Peckmann TR, Logar CJ, Meek S. 2015. Sex determination in a contemporary Mexican population using the scapula. *Journal of Forensic and Legal Medicine* **37**: 91-96. DOI:10.1016/j.jflm.2015.11.006

Introna J, Di Vella G, Campobasso CP, Dragone M. 1997. Sex determination by discriminant analysis of calcanei measurements. *Journal of forensic sciences* **42**(4): 725-728.

Irurita Olivares J, Alemán Aguilera I. 2016. Validation of the sex estimation method elaborated by Schutkowski in the Granada Osteological Collection of identified infant and young children: Analysis of the controversy between the different ways of analyzing and interpreting the results. *International Journal of Legal Medicine* **130:**1623-1632.

Jantz LM, Jantz R. 1999. Secular change in long bone length and proportion in the United States, 1800-1970. *American Journal of Physical Anthropology*. DOI:10.1002/(SICI)1096-8644(199909)110:1<57::AID-AJPA5>3.0.CO;2-1

Joyce RA. 2017. Chapter One: Sex, Gender, and Anthropology: Moving Bioarchaeology Outside the Subdiscipline. In *Exploring Sex and Gender in Bioarchaeology,* Agarwal SC, Wesp JK (eds.). University of New Mexico Press: Albuquerque, NM; 1-12.

Kranioti EF, Michalodimitrakis M. 2009. Sexual dimorphism of the humerus in contemporary Cretans – A population-specific study and a review of the literature. *Journal of Forensic Sciences* **54**(5): 996-1000. DOI:10.1111/j.1556-4029.2009.01103.x

Krishan K, Kanchan T, Passi N, DiMaggio JA. 2015. Sexual dimorphism in foot length ratios among North Indian adolescents. *Journal of Forensic and Legal Medicine* **36**(Complete): 96-101.

Lee UY, Han SH, Park DK, Kim YS, Kim DI, Chung IH, Chun, MH. 2012. Sex determination from the talus of Koreans by discriminant function analysis. *Journal of Forensic Sciences* **57**(1): 166-171.

Lock M, Nguyen VK. 2010. *An anthropology of biomedicine*. Wiley-Blackwell.

Lovell NC. 1989. Test of Phenice's technique for determining sex from the os pubis. *American Journal of Physical Anthropology* **79**(1): 117-120.

Mahakkanukrauh P, Praneatpolgrang S, Ruengdit S, Singsuwan P, Duangto P, Case DT. 2014. Sex estimation from the talus in a Thai population. *Forensic Science International* **240**: 152.e1-152.e8. DOI:10.1016/j.forsciint.2014.04.001

Martin DL, Bendrey R. 2021. International Journal of Osteoarchaeology: Description. Retrieved March 1, 2021. wiley.com/en-us/International+Journal+of+Osteoarchaeology-p-9780JN

Martin DL, Harrod RP, Pérez VR. 2013. *Bioarchaeology an integrated approach to working with human remains*. Springer: New York.

Martin R. 1928. *Lehrbuch der anthropologie* (2nd ed.). Fischer: Jena.

Martin R, Knussmann R. Anthropologie. 1988. *Handbuch der Vergleichendan. Biologie des Menschen.*: Gustav Frischer: Stuttgart.

Meijer J, Dolphin AE, Yakymchuk C, Gervers M. 2019. Interpreting medieval mobility from burials at the rock-hewn church of St. Georges, Gurat (France): Insights from strontium isotope analysis of bones and teeth. *International Journal of Osteoarchaeology* **29**(4): 574-583. DOI:10.1002/oa.2753

Meijer J. 2018. Reconstructing the origins of the medieval monastic inhabitants of a cave church in Gurat, France, using strontium isotope analysis. *UW Space.*

Meindl RS, Lovejoy CO, Mensforth RP, Carlos LD. 1985. Accuracy and direction of error in the sexing of the skeleton: Implications for paleodemography. *American Journal of Physical Anthropology* **68**(1): 79-85. DOI:10.1002/ajpa.1330680108

Michel J, Charlier P. 2011. France. In *The Routledge Handbook of Archaeological Remains and Legislation: An international guide to laws and practice in the excavation and treatment of archaeological human remains,* Marquez-Grant N, Fibiger L (eds.). Routledge: London; 151-162.

Michiue T, Hishmat AM, Oritani S, Miyamoto K, Amin MF, Ishikawa T, Maeda H. 2018. Virtual computed tomography morphometry of the patella for estimation of sex using postmortem Japanese adult data in forensic identification. *Forensic Science International* **285:** 206.e1-206.e6. DOI:10.1016/j.forsciint.2017.11.029

Moore M. 2013. Sex estimation and assessment. In *Research Methods in Human Skeletal Biology*, DiGangi E, Moore M (eds.). Elsevier: Amsterdam; 91-116.

Mountrakis C, Eliopoulos C, Koilias CG, Manolis SK. 2010. Sex determination using metatarsal osteometrics from the Athens collection. *Forensic Science International* **200**(1): 178.e1-178.e7. DOI:10.1016/j.forsciint.2010.03.041

Murphy AM. 2005. The articular surfaces of the hindfoot: sex assessment of prehistoric New Zealand polynesian skeletal remains. *Forensic Science International* **151**(1): 19-22.

Nathena D, Michopoulou E, Kranioti EF. 2017. Sexual dimorphism of the calcaneus in contemporary Cretans. *Forensic Science International* **277**: 260.e1-260.e8. DOI:10.1016/j.forsciint.2017.04.005

Nuzzo R. 2014. Statistical errors: P values, the “gold standard” of statistical validity, are not as reliable as many scientists assume. *Nature (London)* **506**(7487): 150-152.

Papadimitriou A, Fytanidis G, Douros K, Papadimitriou DT, Nicolaidou P, Fretzayas A. 2008. Greek young men grow taller. *Acta Pædiatrica* **97**(8): 1105-1107. DOI:10.1111/j.1651-2227.2008.00855.x

Papaioannou VA, Kranioti EF, Joveneaux P, Nathena D, Michalodimitrakis M. 2011. Sexual dimorphism of the scapula and the clavicle in a contemporary Greek population: Applications in forensic identification. *Forensic Science International* **217**(1), 231.e1-231.e7. DOI:10.1016/j.forsciint.2011.11.010

Peckmann TR, Orr K, Meek S, Manolis SK. 2015a. Sex determination from the calcaneus in a 20th century Greek population using discriminant function analysis. *Science & Justice* **55**(6): 377-382. DOI:10.1016/j.scijus.2015.04.009

Peckmann TR, Orr K, Meek S, Manolis SK. 2015b. Sex determination from the talus in a contemporary Greek population using discriminant function analysis. *Journal of Forensic and Legal Medicine* **33**: 14-19. DOI:10.1016/j.jflm.2015.03.011

Perez SU, Monteiro LR. 2009. Nonrandom factors in modern human morphological diversification: a study of craniofacial variation in southern South American populations. *Evolution* **63**: 978–993.

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. *American Journal of Physical Anthropology* **30**(2): 297-301.

Power M. 2020. Non-binary and intersex visibility and erasure in Roman archaeology. *Theoretical Roman Archaeology Journal* **3**(1): 11, 1-19. DOI:10.16995/traj.422

Pullman D. 2014. There are universal ethical principles that should govern the conduct of medicine and research worldwide. In *Contemporary Debates in Bioethics*, Caplan AL, Arp R (eds.). Chichester: John Wiley & Sons: 17-26.

Pullman D. 2018. Bioarchaeology, bioethics, and the Beothuk. *American Anthropologist* **120**(1): 11-23. DOI:10.1111/aman.12970

Relethford JH. 1994. Craniometric variation among modern human populations, *American Journal of Physical Anthropology* **95**: 53-62.

Rozendaal AS, Scott S, Peckmann TR, Meek S. 2020. Estimating sex from the seven cervical vertebrae: An analysis of two European skeletal populations. *Forensic Science International* **306**: 110072. DOI:https://doi-org.proxy.lib.uwaterloo.ca/10.1016/j.forsciint.2019.110072

Ruff C. 2002. Variation in human body size and shape. *Annual Review of Anthropology* **31** (2002) :211-232.

Saldías E, Malgosa A, Jordana X, Isidro A. Sex estimation from the navicular bone in Spanish contemporary skeletal collections. *Forensic science international*. 2016;267:229.e1-229.e6. doi:10.1016/j.forsciint.2016.08.002

Sayer D. 2009. Is there a crisis facing British burial archaeology?. *Antiquity* **83**(319): 199-166. DOI:10.1017/S0003598X00098203

Scarre G. 2006. Chapter 11: Can archaeology harm the dead?. In *The Ethics of Archaeology: Philosophical Perspectives on Archaeological Practice*, Scarre C, Scarre G (eds.). Cambridge University Press: Cambridge; 181-198.

Schafer MC, Black SM. 2005. Comparison of ages of epiphyseal union in North American and Bosnian skeletal material. *Journal of Forensic Sciences* **50**: 777-784.

Scott S, Ruengdit S, Peckmann TR, Mahakkanukrauh P. 2017. Sex estimation from measurements of the calcaneus: Applications for personal identification in Thailand. *Forensic Science International* **278**: 405.e1-405.e8. DOI:10.1016/j.forsciint.2017.06.035

Seymour G. 2019. Reconstructing the life histories of the individuals buried in the rock-cut cave church of St. Georges, in Gurat, France. *UW Space.*

Smith RJ. 2018. The continuing misuse of null hypothesis significance testing in biological anthropology. *American Journal of Physical Anthropology* **166**(1): 236-245. DOI:10.1002/ajpa.23399

Smith RJ. 2020. P > .05: The incorrect interpretation of “not significant” results is a significant problem. *American Journal of Physical Anthropology* **172**(4): 521-527. DOI:10.1002/ajpa.24092

Spradley MK, Jantz RL. 2011. Sex estimation in forensic anthropology: Skull versus postcranial elements. *Journal of Forensic Sciences* **56**(2): 289-296. DOI:10.1111/j.1556-4029.2010.01635.x

Steele DG. 1976. The estimation of sex on the basis of the talus and calcaneus. *American Journal of Physical Anthropology* **45**(3): 581-588. DOI:10.1002/ajpa.1330450323

Stojanowski CM, Duncan WN. 2015. Engaging bodies in public imagination: bioarchaeology as social science, science, and humanities. *American Journal of Human Biology* **27**(1): 51-60.

Tallman SD, Blanton AI. 2020. Distal humerus morphological variation and sex estimation in modern Thai individuals. *Journal of Forensic Sciences* **65**(2): 361-371. DOI:10.1111/1556-4029.14218

Tarlow S. 2006. Chapter 12: Archaeological ethics and the people of the past. In *The Ethics of Archaeology: Philosophical Perspectives on Archaeological Practice*, Scarre C, Scarre G (eds.). Cambridge University Press: Cambridge; 199-216.

Ubelaker DH, DeGaglia CM. 2017. Population variation in skeletal sexual dimorphism. *Forensic Science International* **278**: 407.e1-407.e7. DOI:10.1016/j.forsciint.2017.06.012

Vance VL, Steyn M, L'Abbe EN. 2011. Nonmetric sex determination from the distal and posterior humerus in Black and White South Africans. *Journal of Forensic Sciences* **56**(3):710-3.

Walker PL. 2000. Chapter 1 Bioarchaeological Ethics: A historical perspective on the value of human remains. In *Biological Anthropology of the Human Skeleton*, Katzenberg MA, Saunders SR (eds.). Wiley-Liss, Inc; 3-39.

Weisensee KE, Jantz RL. 2011. Secular changes in craniofacial morphology of the portuguese using geometric morphometrics. *American journal of physical anthropology* **145(**4): 548-559. DOI:10.1002/ajpa.21531

Weiss KM. 1972. On the systematic bias in skeletal sexing. *American Journal of Physical Anthropology* **37**(2): 239-250.

Weiss, K. (2011). The 5% Solution. *Evolutionary Anthropology* **20**(3): 81–84. https://doi.org/10.1002/evan.20304

Wesp J. K. 2017. Chapter Five: Embodying Sex/Gender Systems in Bioarchaeological Research. In *Exploring Sex and Gender in Bioarchaeology,* Agarwal SC, Wesp JK (eds.). University of New Mexico Press: Albuquerque, NM; 99-126.

White TD, Folkens PA. 2005. Ch. 19.4: Determination of Sex. In *The Human Bone Manual.* Elsevier Academic Press: Burlington, MA.

Wilbur AK. 1998. The utility of hand and foot bones for the determination of sex and the estimation of stature in a prehistoric population from west-central Illinois. *International Journal of Osteoarchaeology* **8**(3): 180-191. DOI:10.1002/(SICI)1099-1212(199805/06)8:3<180::AID-OA421>3.0.CO;2-D

Zhan M, Fan F, Qiu L, Peng Z, Zhang K, Deng Z. 2018. Estimation of stature and sex from sacrum and coccyx measurements by multidetector computed tomography in Chinese. *Legal Medicine (Tokyo, Japan)* **34**: 21-26. DOI:10.1016/j.legalmed.2018.07.003

Appendices

APPENDIX A: Background and Sample Information

A.1 Plan Drawing of Grave Locations

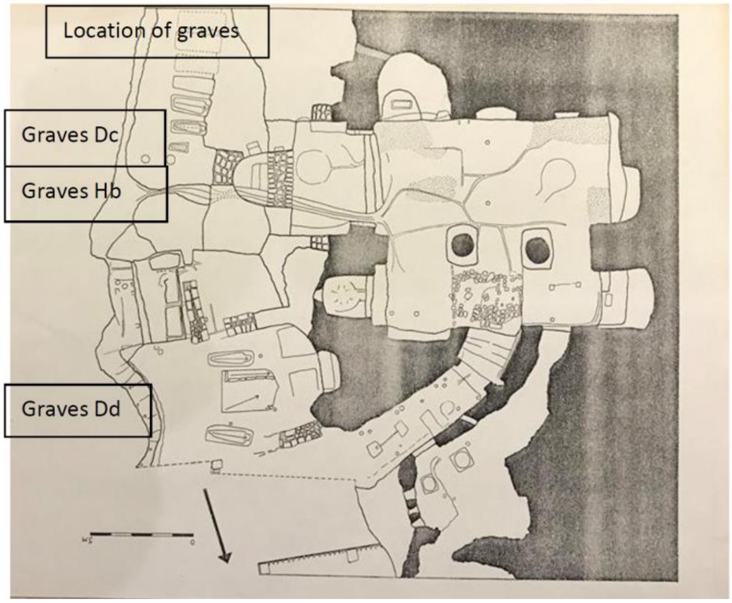


***Figure 10**: Location of Graves at St. George's Church (From Seymour, 2019. Modified image from M. Gervers.)*

A.2 Background Information for Methods 1 and 2

***Table 14**: Information for Skeletal Collections Used in Development of Peckmann et al., 2015a/b (Method 1) and Curate et al., 2021 (Method 2)*

	Peckmann et al., 2015a/b (Method 1)	**Curate et al., 2021 (Method 2)**
Collection	University of Athens Human Skeletal Reference Collection	Coimbra Identified Skeletal Collection, University of Coimbra
Population/Ancestry	Greek	Portuguese Nationals, mostly from Central Portugal
Sample Size	198	180
Sex Distribution	103 males, 95 females	87 males, 93 females
Age-at-death Range	20-99	20-89
Birth/Death Dates Range	Year of death between 1960-1996	Year of birth between 1827-1913. Year of death between 1910-1936.
Other Information:	Death certificates provide demographics, including occupation	Majority from underprivileged socioeconomic backgrounds

APPENDIX B: Methods

B.1 Definitions of Variables Used in Method 1

<u>Calcaneal variables provided in Peckmann et al., 2015a</u>

Maximum length (MAXL): This is the linear distance between the most anterior point of the calcaneus and the most posterior point on the calcaneal tuberosity.

Load arm length (LAL): The linear distance between the most anterior point on the calcaneus and the most posterior point on the posterior articular facet.

Minimum breadth (MINB): This is the minimum distance between the medial and lateral surfaces of the body of the calcaneus.

Middle breadth (MIDB): This is the linear distance between the most lateral point on the posterior articular facet and the most medial point on the sustentaculum tali.

Body height (BH): This is the linear distance between the superior and the inferior surfaces of the body of the calcaneus taken in the coronal plane, at the midpoint between the most posterior point of the posterior articular facet and the most anterior point of the calcaneal tuberosity.

Maximum height (MAXH): This is the distance between the most superior and the most inferior points on the calcaneal tuberosity.

Dorsal articular facet length (DAFL): This is the linear distance between the most posterior and the most anterior points on the posterior articular facet of the calcaneus.

Dorsal articular facet breadth (DAFB): This is taken from the most medial to the most lateral points on the posterior articular facet.

Cuboidal facet height (CFH): This is the linear distance between the most superior and the most inferior points on the cuboidal articular facet.

Load arm length (LAL): The linear distance between the most anterior point on the calcaneus and the most posterior point on the posterior articular facet.

Body height (BH): This is the linear distance between the superior and the inferior surfaces of the body of the calcaneus taken in the coronal plane, at the midpoint between the most posterior point of the posterior articular facet and the most anterior point of the calcaneal tuberosity.

<u>Talar variables provided by Peckmann (2021, pers. comm.)</u>

Height of the trochlea (Talar Height; TH) in Steele 1976: the maximum height of the body in an inferior/superior plane. The measurement is taken by placing the talus on a flat surface and then determining the most superior point of the articular surface for the tibia. The superior point is generally along the medial rim of the facet.

Talar length (TL) in Steele 1976: the maximum length of the trochlear surface in an anterior/posterior plane

The following are translations from Martin 1914 (SLC = sliding caliper):

Height of the head (HH) in Martin 1914- distance of the most distant point of the middle longitudinal curvature of the navicular articular facet from a straight line which connects the endpoints of that curve *the talus was placed on a flat surface to determine the inferiormost and superiormost points on the head

Width of the talus (TW): Projective distance between the tip of the lateral process of the talus from the medial edge of the talus in the transversal level of the trochlea. The transversal edge intersects the tip of the lateral process and is perpendicular to the sagittal level of the talus. The tips of the SLC must touch the horizontal surface on which the bone is resting.

Length of trochlea (TrL): Distance between the two intersecting points of the middle sagittal curvature of the trochlea with the anterior and posterior edge of the superior facet

Breadth of the trochlea (TrB): Distance between the lateral from the medial edge of the superior facet of the trochlea in the transversal level SLC.

Head–neck length (HNL): = length of the caput and collum: projective distance between the most distant point of the navicular articular facet from the anterior end of the middle sagittal curvature of the trochlea, measured in the longitudinal axis of the collum and projected on the surface on which the bone rests. SCL with moving arms

Length of the posterior articular facet for the calcaneus (LPAS): straight-line distance between the two endpoints of the middle longitudinal curvature of the posterior articular facet for the calcaneus from each other. The points are exactly at the edge of the articular facets

Breadth of the posterior articular facet for the calcaneus (BPAS): straight-line distance perpendicular to the length.

B.2 Definitions of Variables Used in Method 2

Provided on osteomics.com/CalcTalus/

///////// Calcaneal measurements

CM1 - Maximum length
Projected line from the most posterior point of the tuberosity of the calcaneus to the most anterior/superior point of the cuboidal facet (Martin, 1928).

CM1a - Length
Projected line from the most posterior point of the tuberosity of the calcaneus to the mid-point of the cuboidal facet (Martin, 1928)

CM2 - Load arm width
Transverse projected line perpendicular to the long axis from the most lateral point of the posterior articular surface, to the most medial point of the sustentaculum tali (Martin, 1928).

CLAL - Load arm length
Projected line from the most posterior point of the posterior articular surface for the talus, to the most anterior/superior point of the cuboidal facet (Slal, according to Steele, 1976).

CM4 - Body height
Projected height from the most inferior point of the tuberosity of the calcaneus to the most inferior point of the body of the calcaneus (Martin, 1928).

CM5 - Body length
Projected line from the most superior point of the tuber calcanei to the most distally point of the posterior facet of the calcaneus, measured parallel to the sagittal axis of the bone (Martin, 1928).

CMBH - Maximum body height
Greatest projected height of the calcaneus, and measured from the most inferior point of the tuberosity of the calcaneus to the most superior point of the posterior facet of the calcaneus (Smh, according to Steele, 1976).

CM7 - Tuber calcanei length
Greatest projected height of the tuber calcanei, measured from the most superior point of the tuber calcanei to the most inferior point of the *Processus medialis tuberis calcanei* (Martin, 1928).

CM8 - Tuber calcanei width
Projected line laterally/medially of the tuber calcanei, perpendicular to the sagittal plane (Martin, 1928).

///////// Talar measurements

TM1 - Maximum length [redacted]
From the *M. flexor hallucis longus* groove to the most anterior point on the head measured parallel to the sagittal axis of the trochlea (Martin, 1928).

TM2 - Width of the talus [redacted]
Maximum projected line laterally/medially perpendicular to the sagittal plane. The lateral point is the most lateral point on the articular surface for the lateral maleolus and the line generally bisects the articular surface for the tibia slightly forward of the midpoint (Martin, 1928).

TM3 - Body height [redacted]
Height of the body along the sagittal plane, taking above the articular surface for the tibia. The measurement is taken by placing the talus on a flat surface (Martin, 1928).

TM3a - Maximum body height [redacted]
Maximum height of the body in an inferior/medially perpendicular to the sagittal plane. The measurement is taken by placing the talus on a flat surface and then determining the most superior point of the articular surface for the tibia. The superior point is generally along the medial rim of the facet (Martin, 1928).

TM4 - Trochlear length [redacted]
Maximum length of the trochlear articular surface on the midline measured parallel to the sagittal axis of the trochlea (Martin, 1928).

TM5 - Trochlear width [redacted]
From the mid-lateral edge of the trochlea to its mid-medial edge measured perpendicular to the sagittal plane of the trochlea (Martin, 1928).

B.3: Measurements Excluded from Analysis Due to Damage

***Table 15**: Measurements Excluded from Analysis due to Damage*

Method	Bone	Measurement
1 (Peckmann et al., 2015a)	Gu DC Left Calcaneus	LAL
1 (Peckmann et al., 2015a)	Gu DC Left Calcaneus	DAFL
1 (Peckmann et al., 2015a)	Gu DC Left Calcaneus	CFH
1 (Peckmann et al., 2015b)	Gu 3 Left Talus	LPAS
1 (Peckmann et al., 2015b)	Gu 3 Left Talus	TW
1 (Peckmann et al., 2015b)	Gu 6 Right Talus	LPAS
1 (Peckmann et al., 2015b)	Gu 6 Right Talus	TW
1 (Peckmann et al., 2015b)	Gu DC Right Talus	HH
2 (Curate et al., 2021)	Gu 2 Left	TM2
2 (Curate et al., 2021)	Gu 3 Left	TM2
2 (Curate et al., 2021)	Gu 6 Right	TM2
2 (Curate et al., 2021)	Gu 3 Left	TM5

B.4: More on the CalcTalus Tool (osteomics.com/CalcTalus/)

CalcTalus is a web application developed by Francisco Curate, Joao d'Oliveira Coelho, and Ana Maria Silva (2021). The authors developed an original metric sex estimation method using a Portuguese population and created CalcTalus as an accompanying web tool to make their method more easily accessible to the public. Using CalcTalus, users measure the talus and calcaneus according to the *Instructions* tab (Appendix B.2). On the *Analysis* tab, users will select a statistical model and input the measurements for each applicable variable. CalcTalus then provides a probabilistic prediction (probability of being female and probability of being male) based on the measurements inputted.

The two types of statistical models used by Curate and colleagues (2021) are logistic regression (LR) and support vector machine (SVM). Each type has multiple models that can be selected from the drop down menu in the upper left, and these models use different variables from the talus, the calcaneus, or both the talus and calcaneus.

CalcTalus is one of several decision support systems for human osteology housed on the website osteomics.com, created by David Senhora Navega and Joao d'Oliveira Coelho.

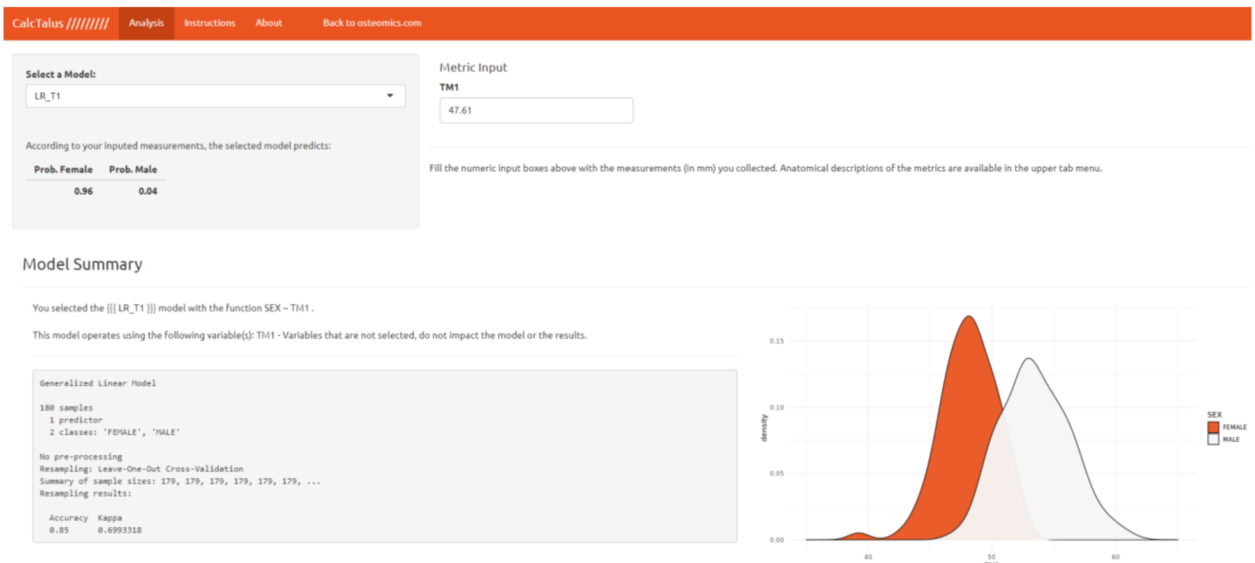


Figure 11*: User Interface on the CalcTalus Tool Used in Method 2 (osteomics.com/CalcTalus/).*

Pictured is the output for Gu 1 Left Talus (trial 1) using Model LR_T1 (logistic regression, talus 1).

B.5: Statistical Models in Method 2 (Curate et al., 2021)

***Table 16**: List of Models and Measurements Used in Method 2 (Curate et al., 2021)*

Bone	Model	Measurements Used
Talus	LR_T1	TM1
Talus	LR_T2	TM1, TM4
Talus	LR_T3	TM1, TM4, TM5
Talus	SVM_T1	TM1
Talus	SVM_T2	TM1, TM4
Talus	SVM_T3	TM1, TM4, TM5
Calcaneus	LR_C1	CMBH
Calcaneus	LR_C2	CMBH, CM8
Calcaneus	SVM_C1	CMBH
Calcaneus	SVM_C2	CMBH, CM8

LR = Logistic Regression, SVM = Support Vector Machine.

B.6: Photos of Gurat Remains

Figure 12: *Healed Antemortem Fracture on Gu 2 Left Talus. Superior (left) and inferior (right) views. Photos by author.*

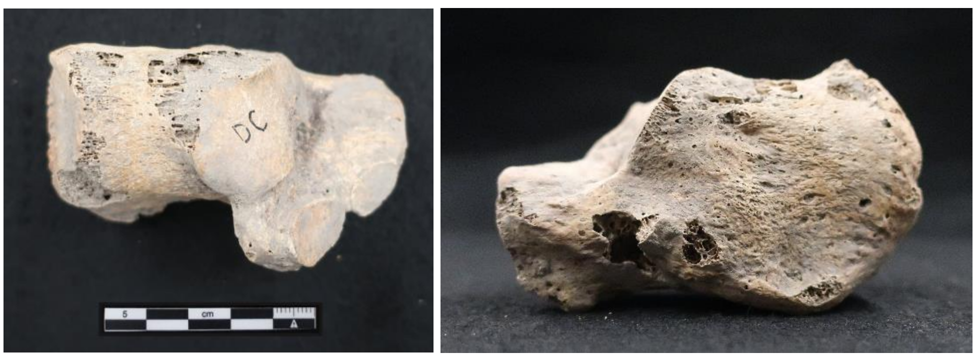


Figure 13: *Gu DC Calcaneus. Superior (left) and lateral (right) views. Photos by author.*

Figure 14: *Gu 3 Left Talus. Superior (left) and lateral (right) views showing damage to the lateral process. Photos by author.*

For trials 1 and 2, the measurement of TW (talar width) was not taken on Gu 3 left talus due to this damage

APPENDIX C: Results

C.1: Results for Method 1

N/A = not calculated: least one of the variables in this function was excluded due to damage. Highlighted rows indicate the bones that had the most conflicting results.

Table 17: *Results for Method 1, Trial 1 (talus)*

Individual	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables)	Function 5 (Stepwise: TW and TL)	Univariate TL	Univariate LPAS	Univariate TW	Univariate TrL	Univariate TrB
Gu 1	**Left**	Female	Female	Female	Male	Female	Female	Female	Female	Female	Female
	Right	Female	Female	Female	Female	Female	Female	Female	Female	Female	Female
Gu 2	**Left**	Female	Male	Male	Male	Male	Male	Male	Male	Male	Male
	Right	Female	Male	Male	Male	Male	Male	Male	Male	Male	Male
Gu 3	**Left**	N/A	N/A	N/A	Male	N/A	Female	N/A	N/A	Female	Male
	Right	Female	Female	Male	Male	Male	Female	Female	Male	Female	Male
Gu 6	**Left**	Female	Female	Female	Female	Female	Female	Female	Female	Female	Female
	Right	N/A	N/A	N/A	Female	N/A	Female	N/A	N/A	Female	Female
Gu 7	**Left**	Female	Female	Male	Male	Male	Female	Female	Male	Male	Male
	Right	Female	Male	Male	Male	Male	Female	Female	Male	Male	Male
Gu DC	**Right**	N/A	Female	Male	N/A	Female	Female	Female	Male	Male	Male

Table 18: *Results for Method 1, Trial 2 (talus)*

Individual	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables)	Function 5 (Stepwise: TW and TL)	Univariate TL	Univariat e LPAS	Univariate TW	Univariate TrL	Univariate TrB
Gu 1	**Left**	Female	Female	Female	Male	Female	Female	Female	Female	Female	Female
	Right	Female	Female	Female	Male	Female	Female	Female	Female	Female	Female
Gu 2	**Left**	Female	Male	Male	Male	Male	Male	Male	Male	Male	Male
	Right	Female	Male	Male	Male	Male	Male	Male	Male	Male	Male
Gu 3	**Left**	N/A	N/A	N/A	Male	N/A	Female	N/A	N/A	Female	Male
	Right	Female	Female	Male	Male	Male	Female	Female	Male	Male	Male
Gu 6	**Left**	Female	Female	Female	Female	Female	Female	Female	Female	Female	Female
	Right	N/A	N/A	N/A	Female	N/A	Female	N/A	N/A	Female	Female
Gu 7	**Left**	Female	Female	Male	Male	Male	Female	Male	Male	Male	Male
	Right	Female	Male	Male	Male	Male	Female	Male	Male	Male	Male
Gu DC	**Right**	N/A	Female	Male	N/A	Male	Female	Male	Male	Male	Male

Table 19: *Results for Method 1, Trial 3 (talus)*

Individual	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables)	Function 5 (Stepwise: TW and TL)	Univariate TL	Univariate LPAS	Univariate TW	Univariate TrL	Univariate TrB
Gu 1	**Left**	Female	Female	Female	Female	Female	Female	Female	Female	Female	Female
	Right	Female	Female	Female	Female	Female	Female	Female	Female	Female	Female
Gu 2	**Left**	Female	Male	Male	Male	Male	Male	Male	Male	Male	Male
	Right	Female	Male	Male	Male	Male	Male	Male	Male	Male	Male
Gu 3	**Left**	Female	Female	Male	Male	Male	Female	Male	Male	Female	Male
	Right	Female	Female	Male	Male	Male	Female	Male	Male	Male	Male
Gu 6	**Left**	Female	Female	Female	Male	Female	Female	Female	Female	Female	Female
	Right	Female	Female	Female	Female	Female	Female	Female	Female	Female	Female
Gu 7	**Left**	Female	Female	Male	Male	Male	Female	Male	Male	Male	Male
	Right	Female	Male	Male	Male	Male	Female	Male	Male	Male	Male
Gu DC	**Right**	Female	Female	Male	Male	Male	Female	Female	Male	Male	Male

***Table 20**: Comparison of Talus Results Between Trials 1 and 2. "TRUE" indicates that the estimates were consistent between the trials and "FALSE" indicates different estimates*

Individual	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables)	Function 5 (Stepwise)	Univariate TL	Univariate LPAS	Univariate TW	Univariate TrL	Univariate TrB
Gu 1	**Left**	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
	Right	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 2	**Left**	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
	Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 3	**Left**	N/A	N/A	N/A	TRUE	N/A	TRUE	N/A	N/A	TRUE	TRUE
	Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
Gu 6	**Left**	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
	Right	N/A	N/A	N/A	TRUE	N/A	TRUE	N/A	N/A	TRUE	TRUE
Gu 7	**Left**	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
	Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
Gu DC	**Right**	N/A	TRUE	TRUE	N/A	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE

***Table 21:** Comparison of Talus Results Between Trials 1 and 3. "TRUE" indicates that the estimates were consistent between the trials and "FALSE" indicates different estimates.*

Individual	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables)	Function 5 (Stepwise)	Univariate TL	Univariate LPAS	Univariate TW	Univariate TrL	Univariate TrB
Gu 1	**Left**	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
	Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 2	**Left**	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
	Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 3	**Left**	N/A	N/A	N/A	TRUE	N/A	TRUE	N/A	N/A	TRUE	TRUE
	Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE
Gu 6	**Left**	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
	Right	N/A	N/A	N/A	TRUE	N/A	TRUE	N/A	N/A	TRUE	TRUE
Gu 7	**Left**	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
	Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
Gu DC	**Right**	N/A	TRUE	TRUE	N/A	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE

Table 22: *Results for Method 1, Trial 1 (Calcaneus)*

Individual	Side	Function 3 (Breadth variables)	Univariate MAXL	Univariate MIDB	Univariate DAFB
Gu DC	Left	Male	Male	Male	Male

Table 23: *Results for Method 1, Trial 2 (Calcaneus)*

Individual	Side	Function 3 (Breadth variables)	Univariate MAXL	Univariate MIDB	Univariate DAFB
Gu DC	Left	Male	Male	Male	Female

Table 24: *Results for Method 1, Trial 3 (Calcaneus)*

This table contains a greater number of discriminant functions than for trials 1 and 2 as damaged measurements were not excluded during data collection. Function 4 is included twice because the result changed when using BH taken with coordinate calipers as compared to the digital sliding calipers.

Individual	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables) – BH taken using digital calipers	Function 4 (Height variables) – BH taken using coordinate calipers	Function 5 (Stepwise)	Univariate MAXL	Univariate MIDB	Univariate DAFB
Gu DC	Left	Female	Male	Female	Female	Male	Female	Male	Female	Female

Table 25: *Comparison of Calcaneus Results from Trials 1 and 2 for Method 1. "TRUE" indicates that estimates were consistent between the trials and "FALSE" indicates a different estimate*

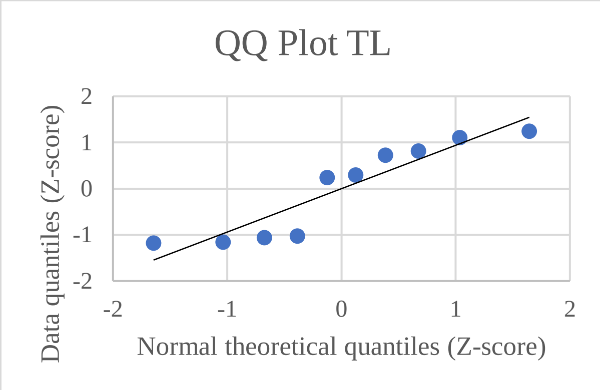
Individual	Side	Function 3 (Breadth Variables)	Univariate MAXL	Univariate MIDB	Univariate DAFB
Gu DC	Left	TRUE	TRUE	TRUE	FALSE

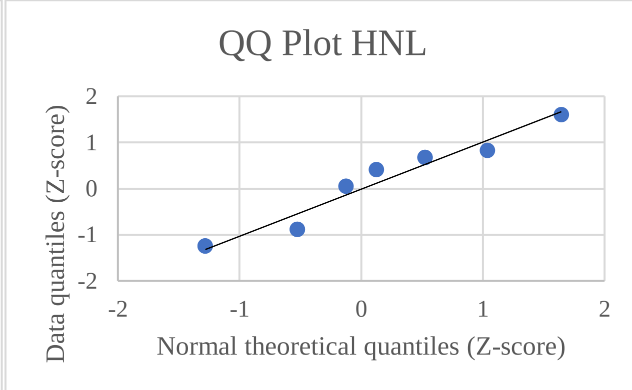
Table 26: *Comparison of Calcaneus Results from Trials 1 and 3 for Method 1, where "TRUE" indicates that results are consistent across all trials and "FALSE" indicates conflicts between the results of different trials.*

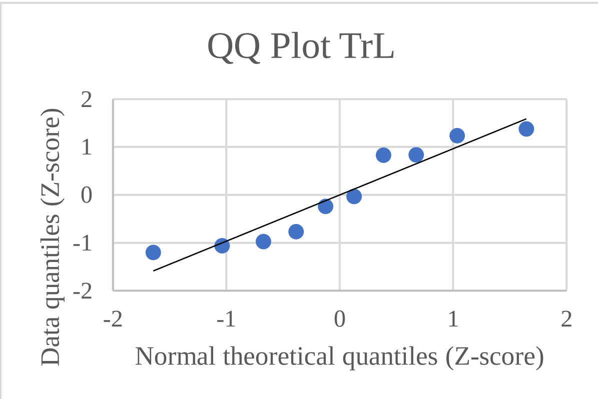
N/A = not compared, as Trial 1 excluded a variable in this function due to damage.

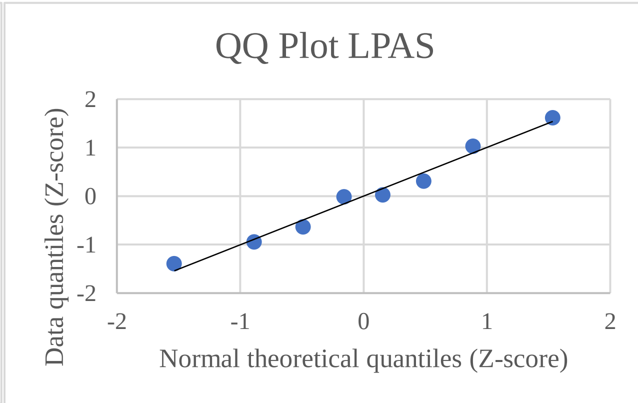
Individual	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables) – BH taken using digital calipers	Function 4 (Height variables) – BH taken using coordinate calipers	Function 5 (Stepwise)	Univariate MAXL	Univariate MIDB	Univariate DAFB
Gu DC	Left	N/A	N/A	FALSE	N/A	N/A	N/A	TRUE	FALSE	FALSE

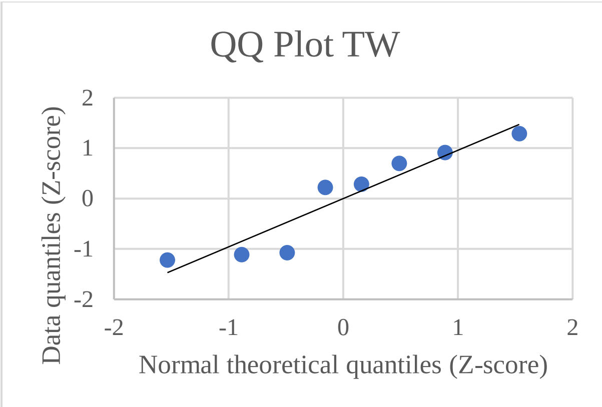
C.2: QQ Plots for Variables in Method 1

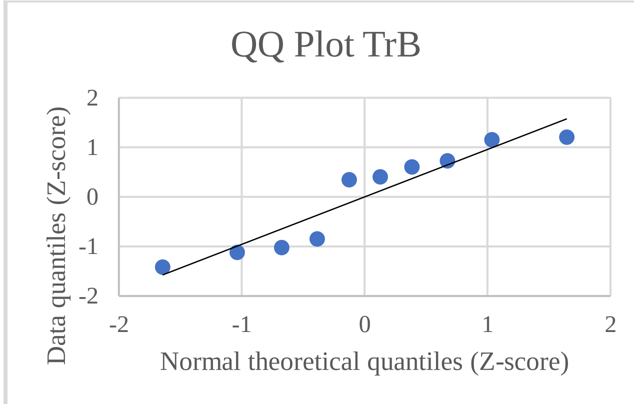


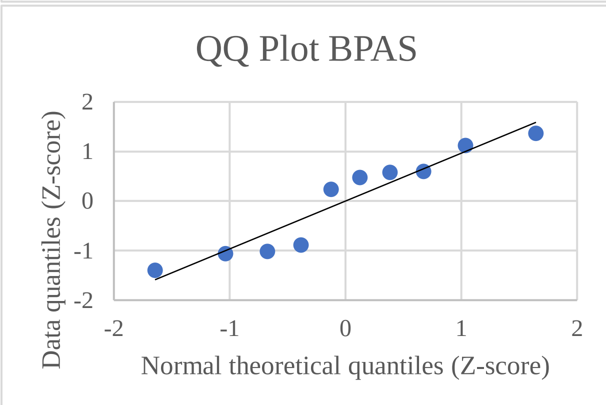


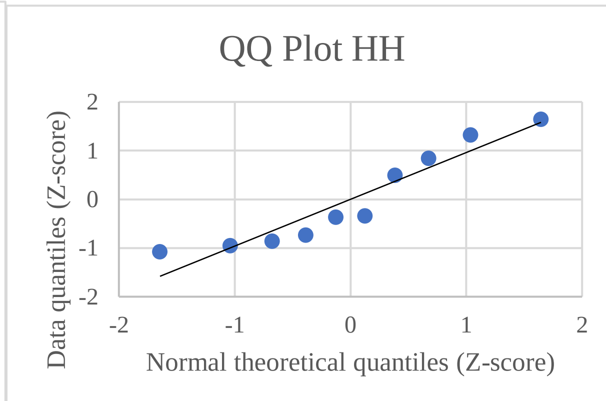


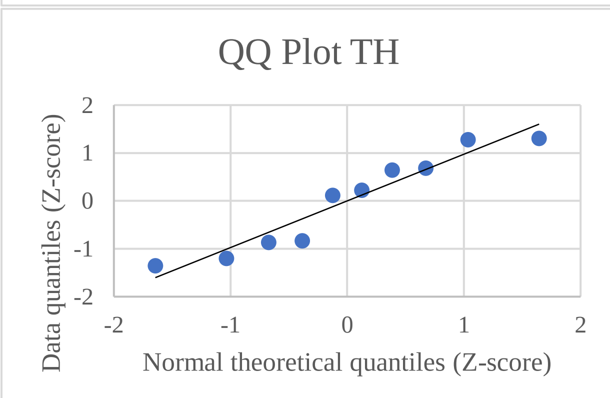



QQ Plot TW
Data quantiles (Z-score)
2
1
0
-1
-2
-2
-1
0
1
2
Normal theoretical quantiles (Z-score)


QQ Plot TrB
Data quantiles (Z-score)
2
1
0
-1
-2
-2
-1
0
1
2
Normal theoretical quantiles (Z-score)


QQ Plot BPAS
Data quantiles (Z-score)
2
1
0
-1
-2
-2
-1
0
1
2
Normal theoretical quantiles (Z-score)


QQ Plot HH
Data quantiles (Z-score)
2
1
0
-1
-2
-2
-1
0
1
2
Normal theoretical quantiles (Z-score)


QQ Plot TH
Data quantiles (Z-score)
2
1
0
-1
-2
-2
-1
0
1
2
Normal theoretical quantiles (Z-score)

C.3: Qualitative Assessment of Symmetry

Table 27: *Left/Right Agreement per Individual, Trial 1*

TRIAL 1	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables)	Function 5 (Stepwise)	Univariate TL	Univariate LPAS	Univariate TW	Univariate TrL	Univariate TrB
Gu 1	Left Right	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 2	Left Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 3	Left Right	N/A	N/A	N/A	TRUE	N/A	TRUE	N/A	N/A	TRUE	TRUE
Gu 6	Left Right	N/A	N/A	N/A	TRUE	N/A	TRUE	N/A	N/A	TRUE	TRUE
Gu 7	Left Right	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

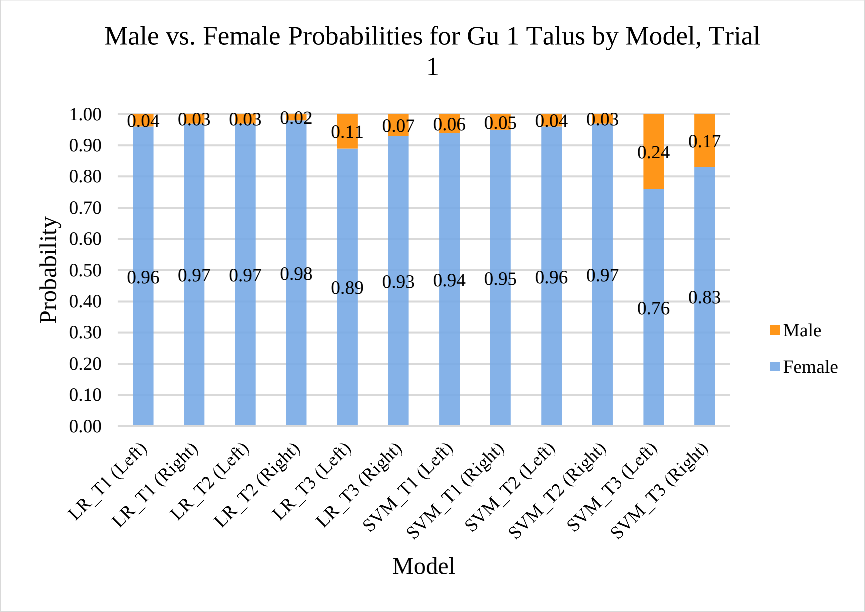
Table 28: *Left/Right Agreement per Individual, Trial 2*

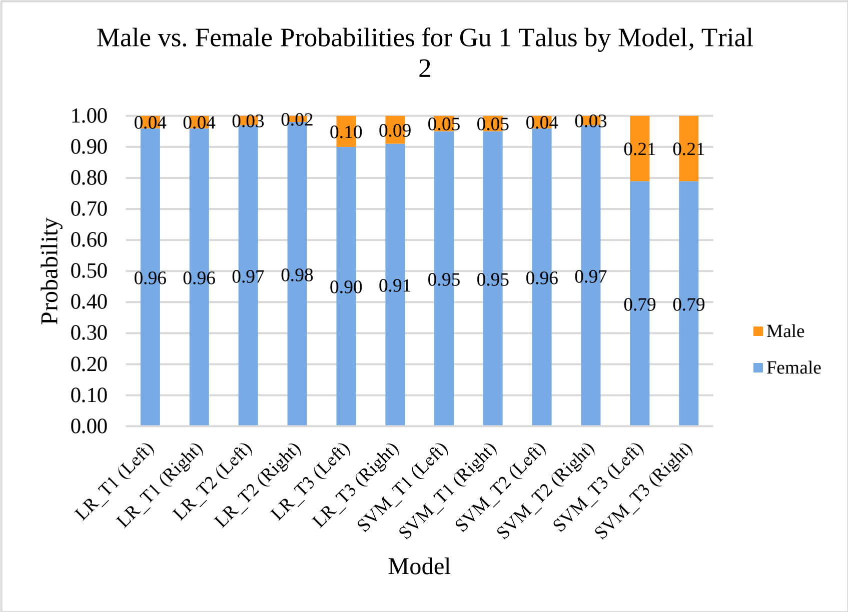
TRIAL 2	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables)	Function 5 (Stepwise)	Univariate TL	Univariate LPAS	Univariate TW	Univariate TrL	Univariate TrB
Gu 1	Left Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 2	Left Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 3	Left Right	N/A	N/A	N/A	TRUE	N/A	TRUE	N/A	N/A	FALSE	TRUE
Gu 6	Left Right	N/A	N/A	N/A	TRUE	N/A	TRUE	N/A	N/A	TRUE	TRUE
Gu 7	Left Right	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

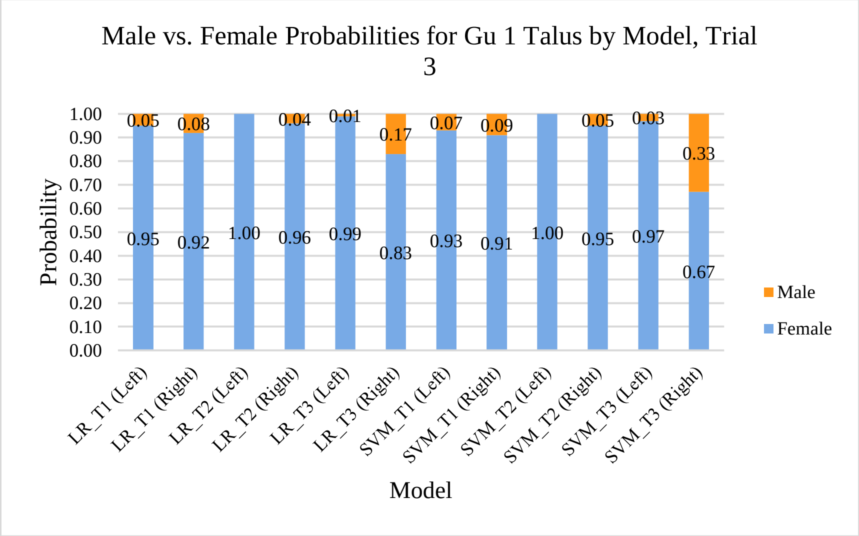
Table 29: *Left/Right Agreement per Individual, Trial 3*

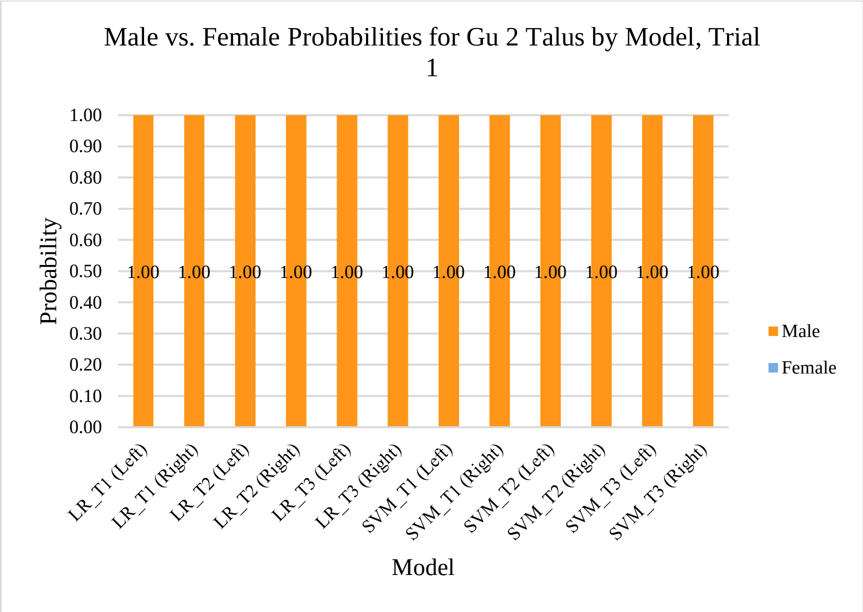
TRIAL 3	Side	Function 1 (All 9 variables)	Function 2 (Length variables)	Function 3 (Breadth variables)	Function 4 (Height variables)	Function 5 (Stepwise)	Univariate TL	Univariate LPAS	Univariate TW	Univariate TrL	Univariate TrB
Gu 1	Left Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 2	Left Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 3	Left Right	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
Gu 6	Left Right	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Gu 7	Left Right	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

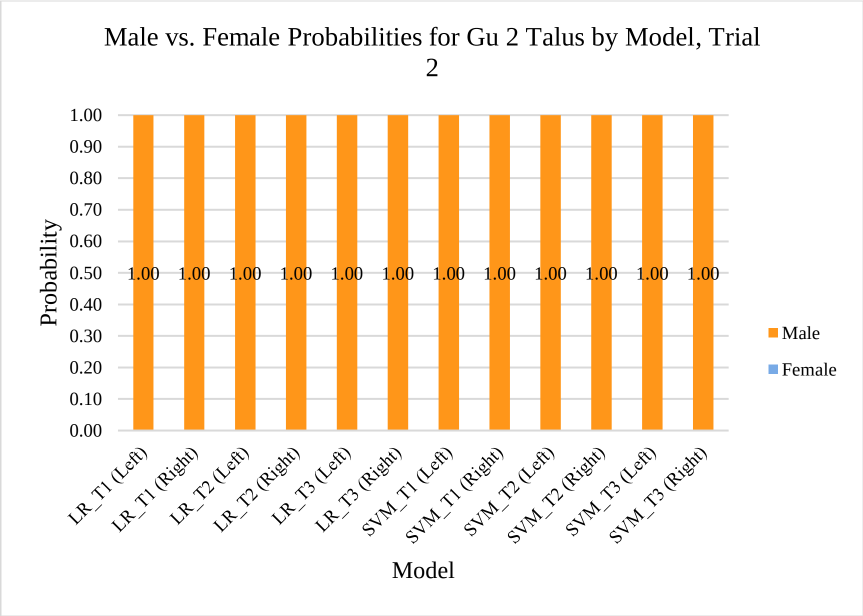
C.4: Results for Method 2

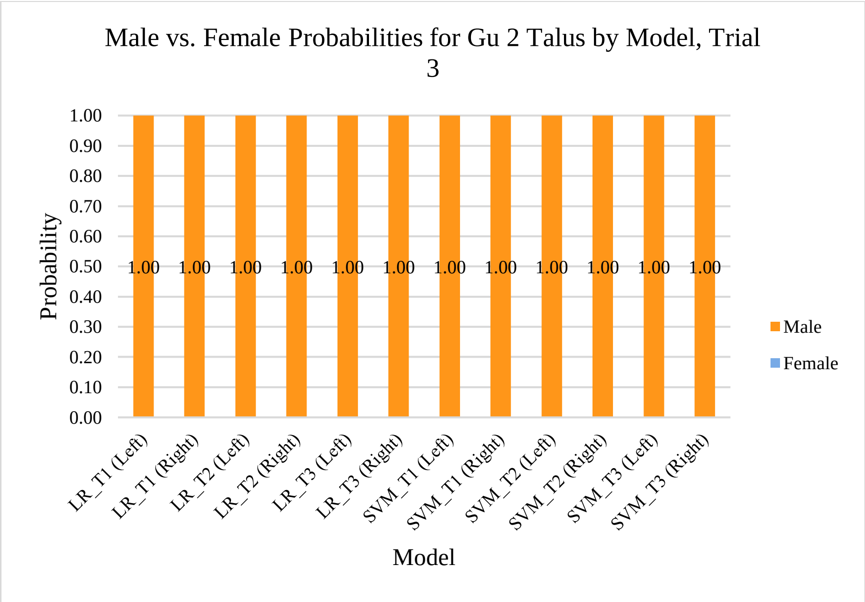


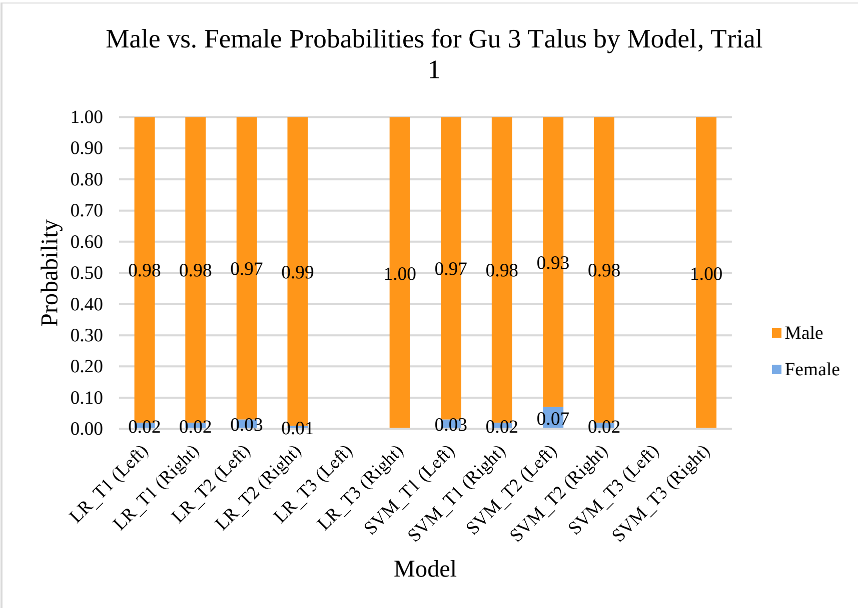


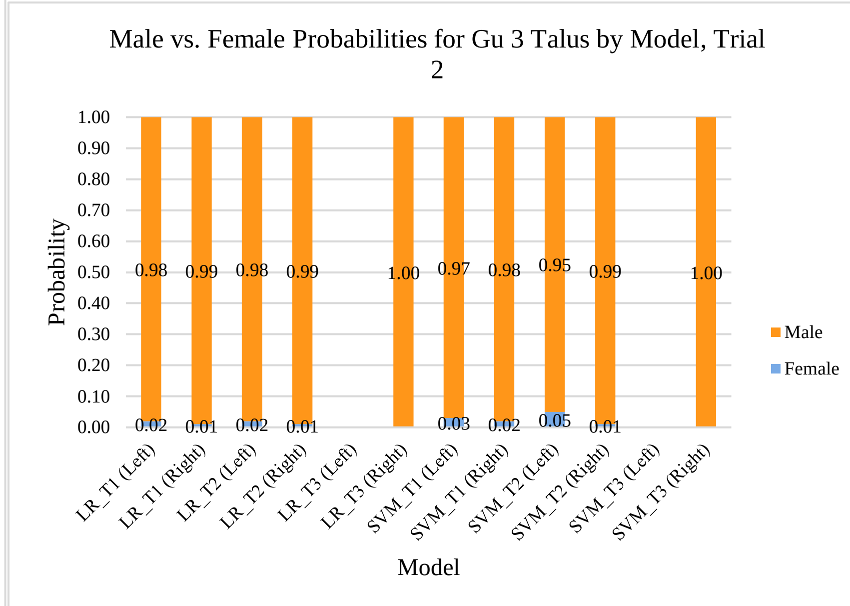


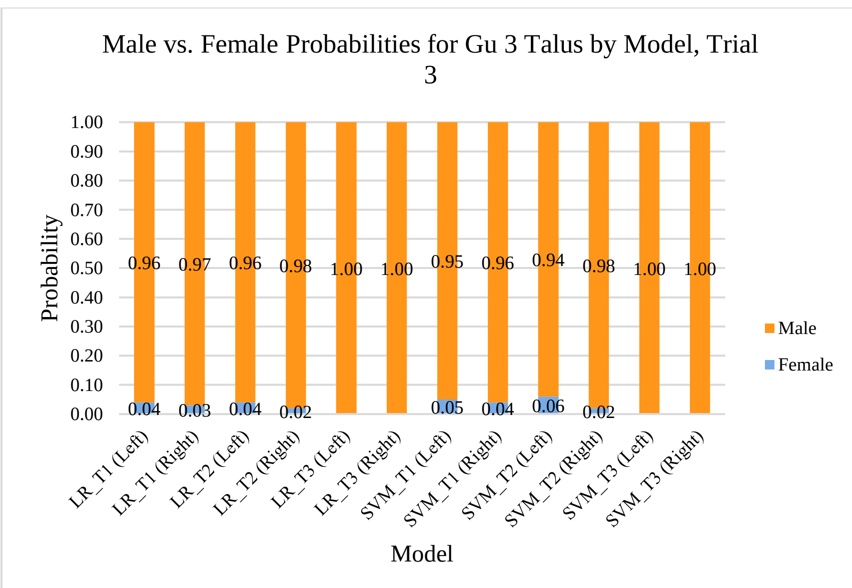

Male vs. Female Probabilities for Gu 2 Talus by Model, Trial 1
Probability
Model
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Male
Female

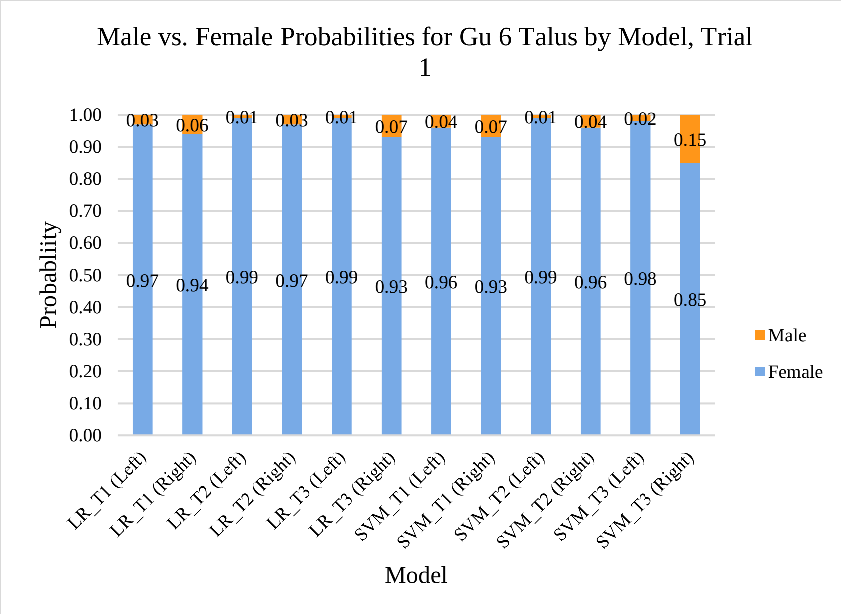

Male vs. Female Probabilities for Gu 2 Talus by Model, Trial 2
Probability
Model
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Male
Female

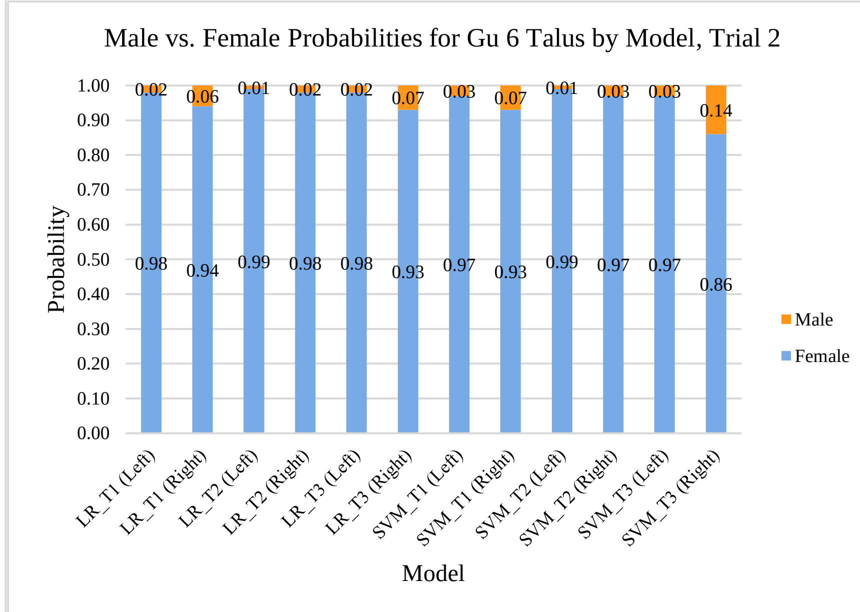

Male vs. Female Probabilities for Gu 2 Talus by Model, Trial 3
Probability
Model
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Male
Female

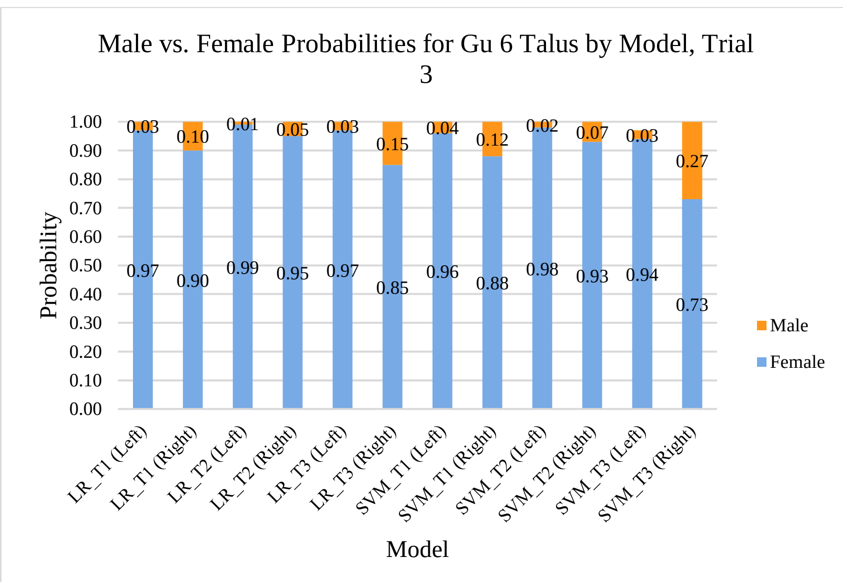

Male vs. Female Probabilities for Gu 3 Talus by Model, Trial 1
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.98 0.98 0.97 0.99 1.00 0.97 0.98 0.93 0.98 1.00
0.02 0.02 0.03 0.01 0.03 0.02 0.07 0.02
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Model
Male
Female

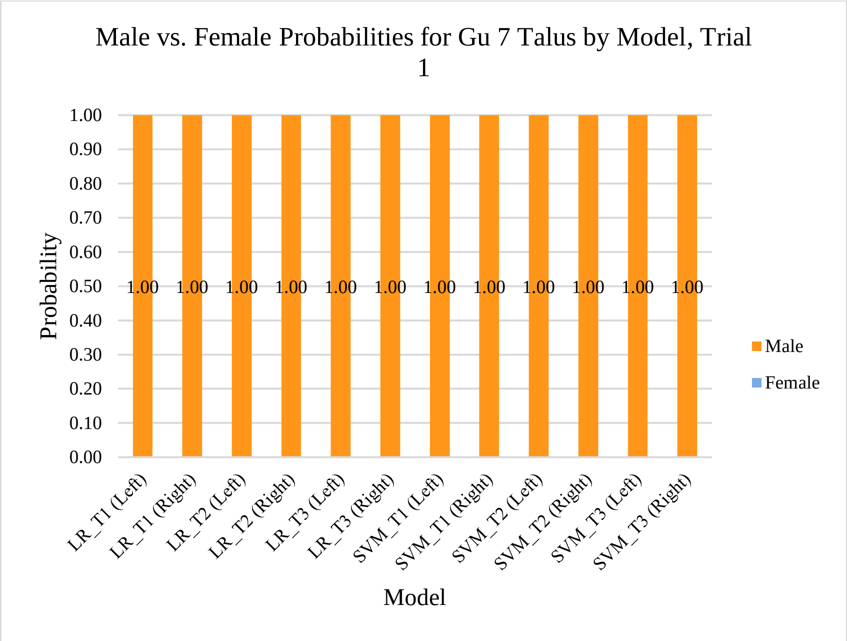

Male vs. Female Probabilities for Gu 3 Talus by Model, Trial 2
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.98 0.99 0.98 0.99 1.00 0.97 0.98 0.95 0.99 1.00
0.02 0.01 0.02 0.01 0.03 0.02 0.05 0.01
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Model
Male
Female

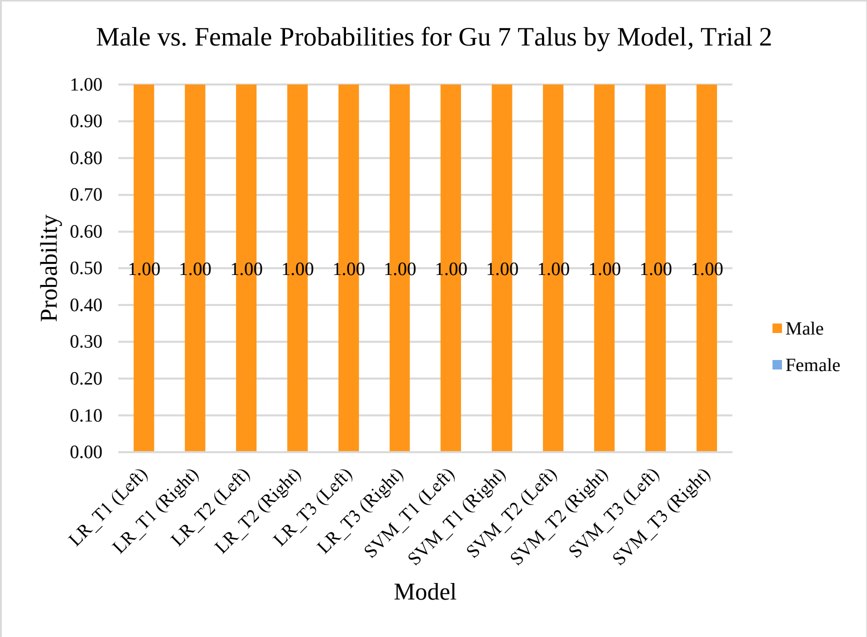

Male vs. Female Probabilities for Gu 3 Talus by Model, Trial 3
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.96 0.97 0.96 0.98 1.00 1.00 0.95 0.96 0.94 0.98 1.00 1.00
0.04 0.03 0.04 0.02 0.05 0.04 0.06 0.02
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Model
Male
Female

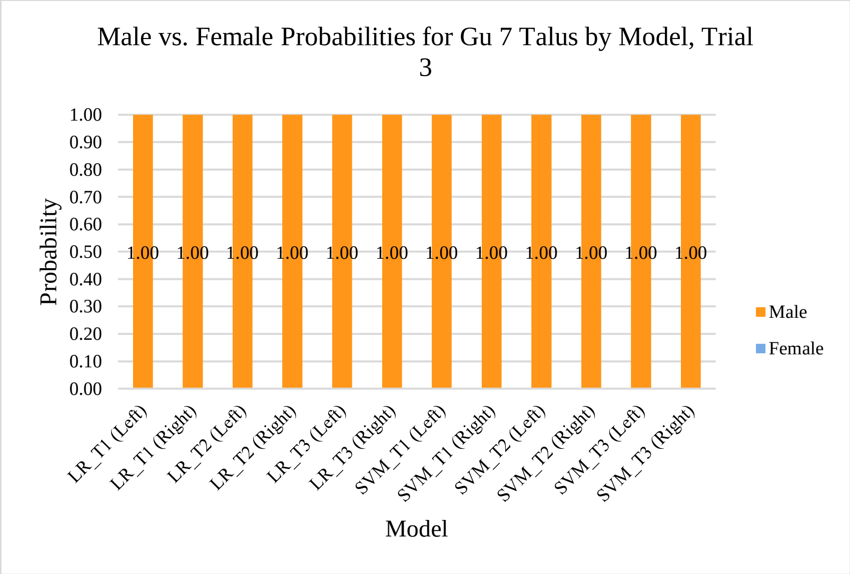

Male vs. Female Probabilities for Gu 6 Talus by Model, Trial 1
Probablity
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.03 0.06 0.01 0.03 0.01 0.07 0.04 0.07 0.01 0.04 0.02 0.15
0.97 0.94 0.99 0.97 0.99 0.93 0.96 0.93 0.99 0.96 0.98 0.85
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Model
Male
Female

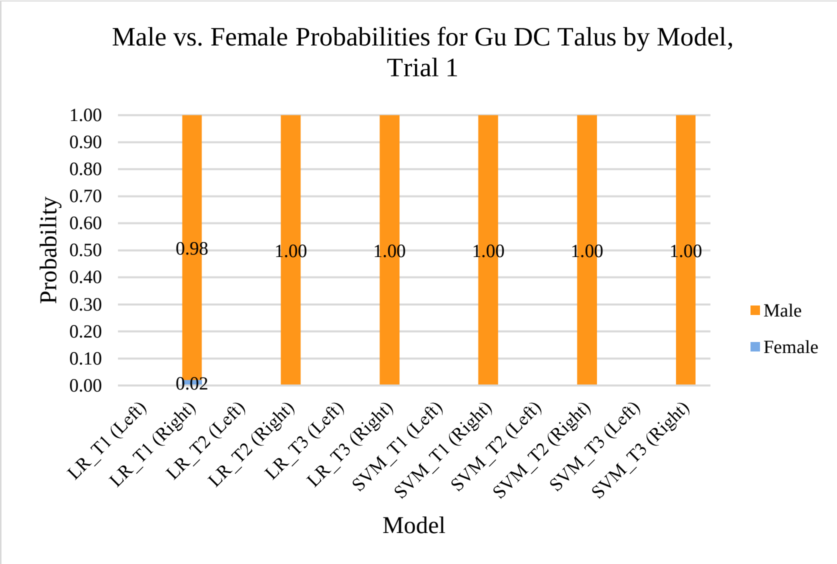

Male vs. Female Probabilities for Gu 6 Talus by Model, Trial 2
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.02 0.06 0.01 0.02 0.02 0.07 0.03 0.07 0.01 0.03 0.03 0.14
0.98 0.94 0.99 0.98 0.98 0.93 0.97 0.93 0.99 0.97 0.97 0.86
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Model
Male
Female

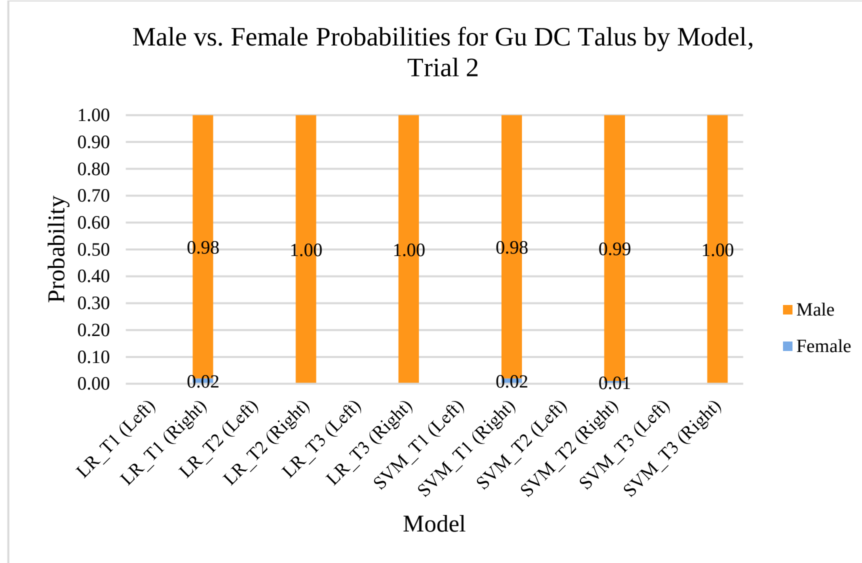

Male vs. Female Probabilities for Gu 6 Talus by Model, Trial 3
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.03 0.10 0.01 0.05 0.03 0.15 0.04 0.12 0.02 0.07 0.03 0.27
0.97 0.90 0.99 0.95 0.97 0.85 0.96 0.88 0.98 0.93 0.94 0.73
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
Model
Male
Female

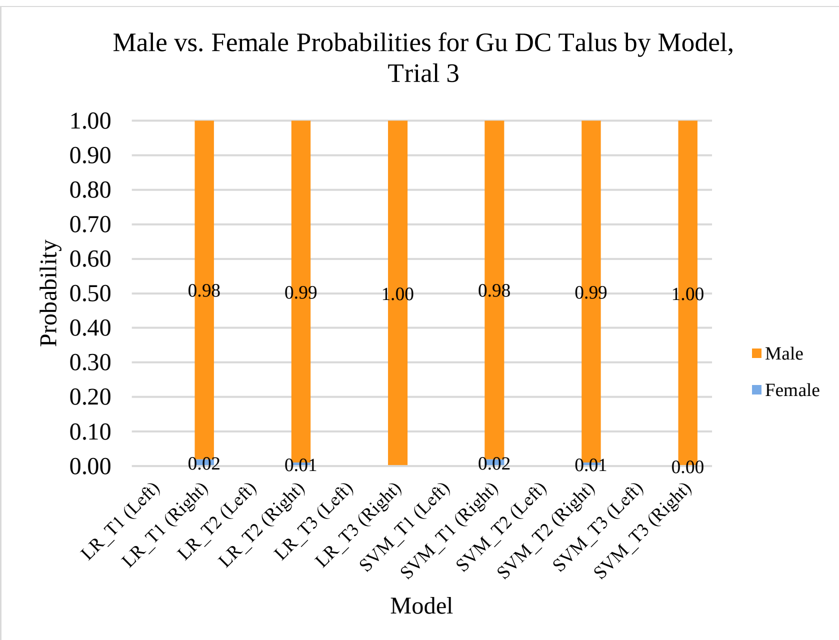

Male vs. Female Probabilities for Gu 7 Talus by Model, Trial 1
Probability
Model
Male
Female
1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)

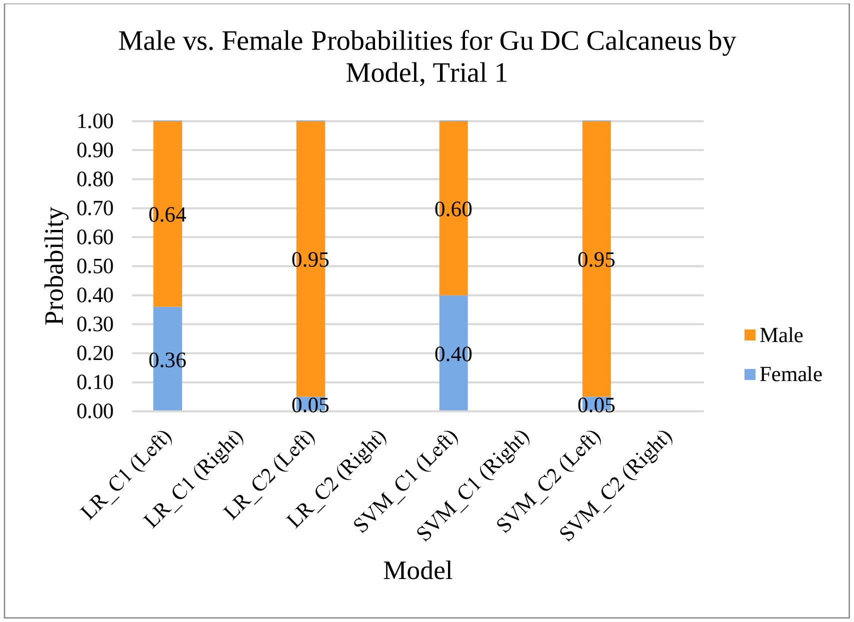

Male vs. Female Probabilities for Gu 7 Talus by Model, Trial 2
Probability
Model
Male
Female
1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)

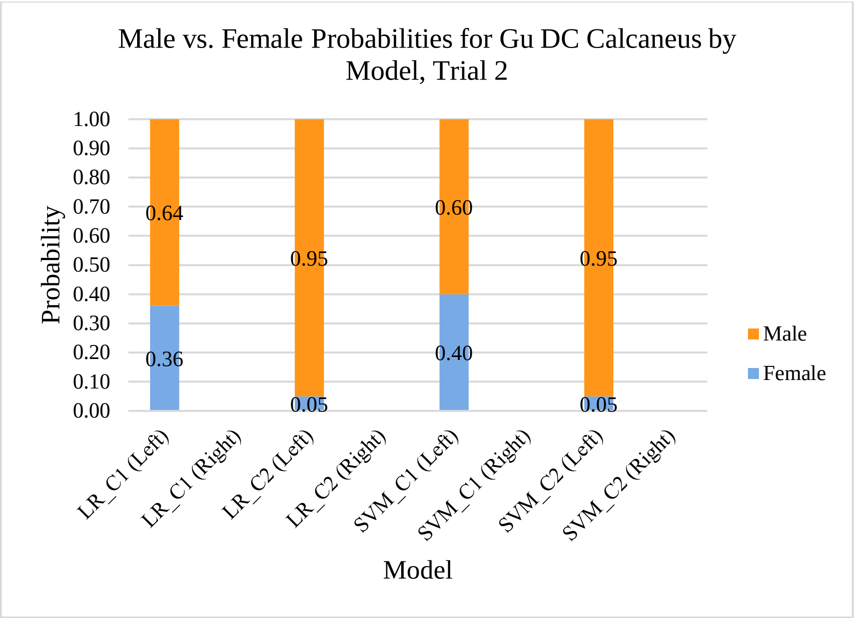

Male vs. Female Probabilities for Gu 7 Talus by Model, Trial 3
Probability
Model
Male
Female
1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)


Male vs. Female Probabilities for Gu DC Talus by Model, Trial 1
Probability
Model
Male
Female
LR_T1 (Left)
LR_T1 (Right)
LR_T2 (Left)
LR_T2 (Right)
LR_T3 (Left)
LR_T3 (Right)
SVM_T1 (Left)
SVM_T1 (Right)
SVM_T2 (Left)
SVM_T2 (Right)
SVM_T3 (Left)
SVM_T3 (Right)
0.98
1.00
1.00
1.00
1.00
1.00
0.02

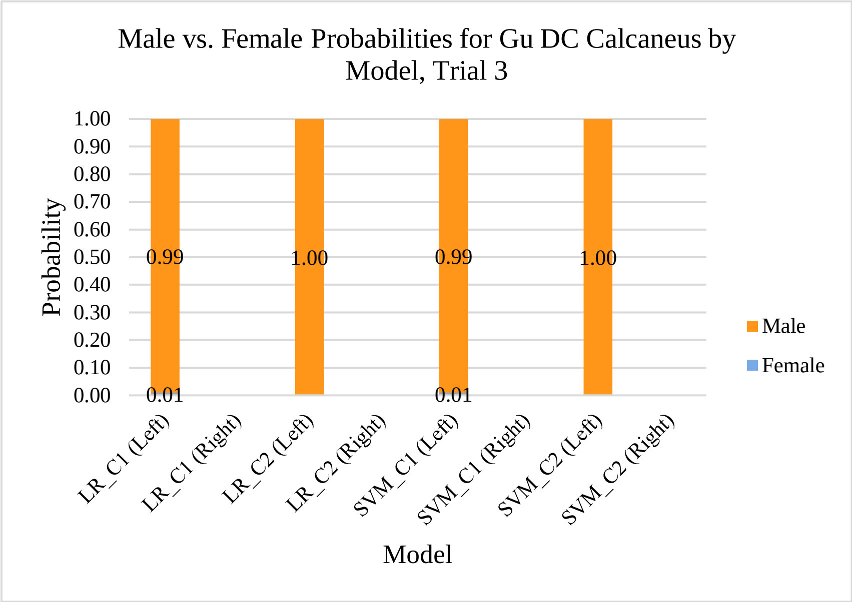

Male vs. Female Probabilities for Gu DC Talus by Model, Trial 2
Probability
Model
Male
Female
0.98
1.00
1.00
0.98
0.99
1.00
0.02
0.02
0.01


Male vs. Female Probabilities for Gu DC Talus by Model, Trial 3
Probability
Model
Male
Female
0.98
0.99
1.00
0.98
0.99
1.00
0.02
0.01
0.02
0.01
0.00

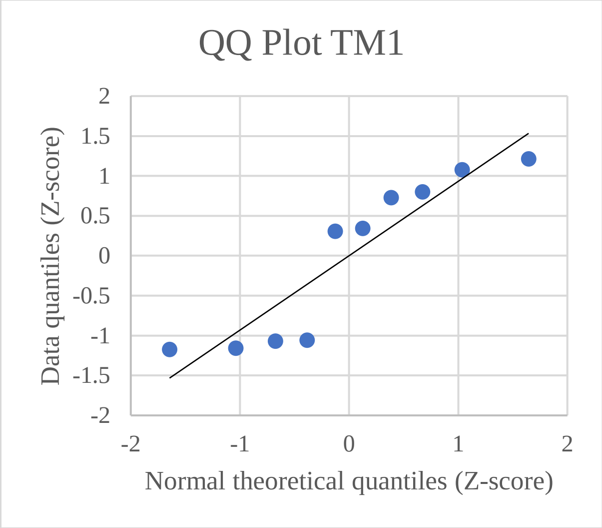

Male vs. Female Probabilities for Gu DC Calcaneus by Model, Trial 1
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.64
0.36
0.95
0.05
0.60
0.40
0.95
0.05
LR_C1 (Left)
LR_C1 (Right)
LR_C2 (Left)
LR_C2 (Right)
SVM_C1 (Left)
SVM_C1 (Right)
SVM_C2 (Left)
SVM_C2 (Right)
Model
Male
Female

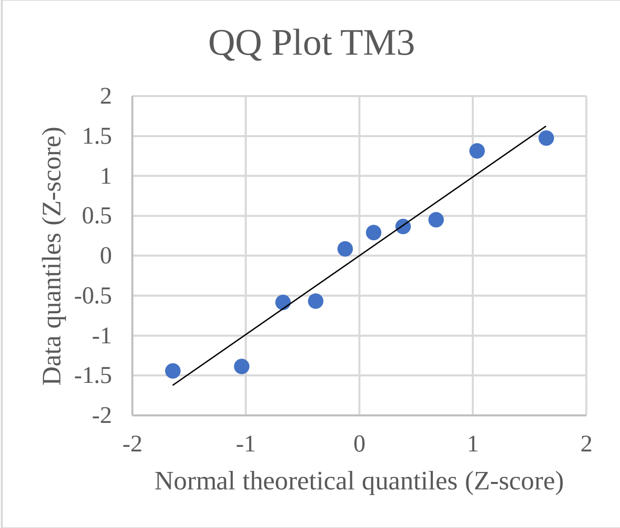

Male vs. Female Probabilities for Gu DC Calcaneus by Model, Trial 2
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.64
0.36
0.95
0.05
0.60
0.40
0.95
0.05
LR_C1 (Left)
LR_C1 (Right)
LR_C2 (Left)
LR_C2 (Right)
SVM_C1 (Left)
SVM_C1 (Right)
SVM_C2 (Left)
SVM_C2 (Right)
Model
Male
Female

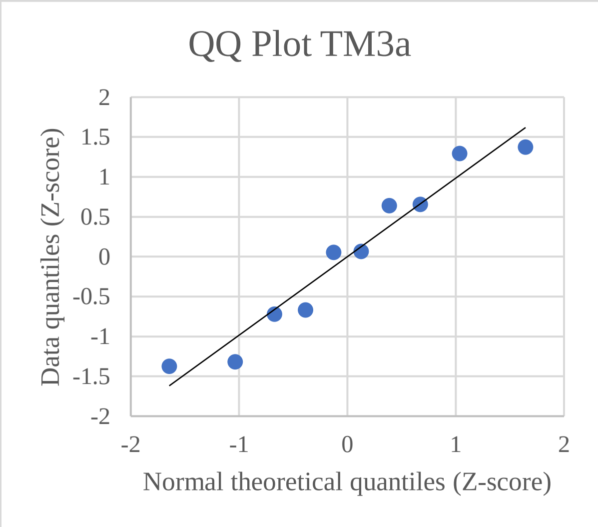
C

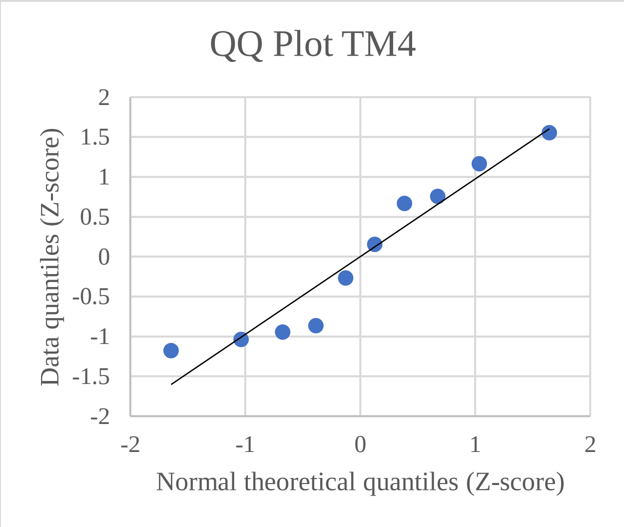

Male vs. Female Probabilities for Gu DC Calcaneus by Model, Trial 3
Probability
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.99
0.01
1.00
0.99
0.01
1.00
LR_C1 (Left)
LR_C1 (Right)
LR_C2 (Left)
LR_C2 (Right)
SVM_C1 (Left)
SVM_C1 (Right)
SVM_C2 (Left)
SVM_C2 (Right)
Model
Male
Female

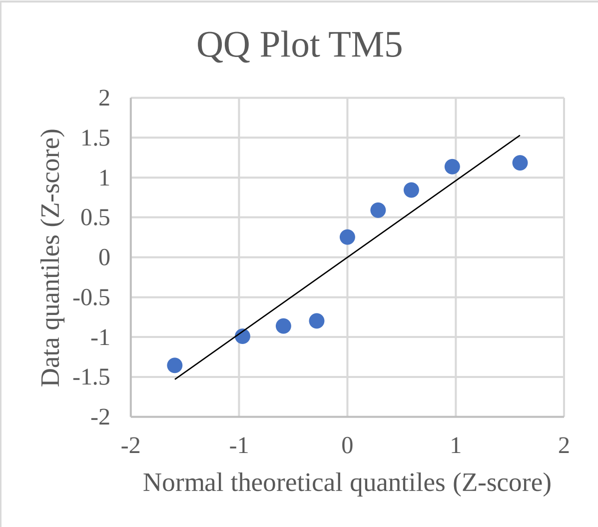
C.5: QQ Plots for Method 2











APPENDIX D: Discussion

D.1: Discrepancies Between Variable Descriptions and Photos in Method 1

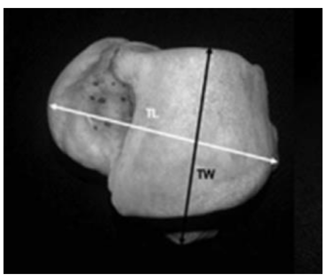
Talar Measurements:

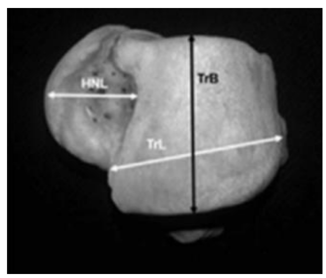
Talar Length (TL): in Steele 1976: the maximum length of the trochlear surface in an anterior/posterior plane (Peckmann 2021, pers. comm.)

- Explanation of discrepancy: Steele (1976) defines maximum length as "the projected line from the sulcus for the flexor hallucis longus muscle at the posterior aspect of the talus to the most anterior point on the articular surface for the navicular."
- The description by Steele was followed for data collection in all three trials of Method 1, as this description also corresponded to the image provided in Peckmann et al., 2015b

Trochlear Length (TrL): distance between the two intersecting points of the middle sagittal curvature of the trochlea with the anterior and posterior edge of the superior facet (Peckmann 2021, pers. comm.)

- Explanation of discrepancy: The anterior point of this variable as depicted in the image seems to be too lateral, but this may be due to morphologic variation between tali in terms of where the middle sagittal curvature is located. TrL was taken at the deepest point of the sagittal curvature.





Images from Peckmann et al., 2015b.

Calcaneal Measurements:

Body height (BH): This is the linear distance between the superior and the inferior surfaces of the body of the calcaneus taken in the coronal plane, at the midpoint between the most posterior point of the posterior articular facet and the most anterior point of the calcaneal tuberosity. (Peckmann et al., 2015a)

- Explanation of discrepancy: the superior point is described as "the midpoint between the most posterior point of the posterior articular facet and the most anterior point of the calcaneal tuberosity", but the diagram depicts it as the most posterior point of the posterior articular facet. The measurement accurate to the description is annotated in orange.
- The written description was followed for data collection in all three trials of Method 1

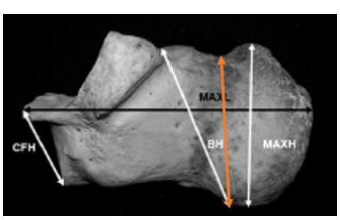


Image from Peckmann et al., 2015a. Annotation by author.

D.2: Age and Sex Estimates

Table 30: *Age-at-Death and Sex Estimates, Including Comparison of Sex Estimates Obtained in Current Study (Method 1, Method 2, Pelvic Morphology) With Estimates from Previous Work on the Gurat Individuals*

M = Male, F = Female, IND = Indeterminate

Individual	**Grave**	**Age Estimate (Auricular Surface)**	**Age Estimate (Cranial Suture Fusion)**	**Method 1** *(Peckmann et al., 2015a/b)*	**Method 2** *(Curate et al., 2021)*	**Pelvic Morphology** *(Buikstra & Ubelaker, 1994)*	**Seymour thesis** (2019)	**Meijer thesis** (2018)
Gu 1	HA	35-44	n/a	F	F	F	F	F
Gu 2	HB	35-44	Middle	M	M	M	M	M
Gu 3	HB	n/a*	Young	IND/M	M	n/a**	M	Possible M
Gu 6 (juvenile)	JA	n/a***		F	F	Possible F	F	F
Gu 7	DC	40-59	Middle-Old	M	M	M	M	Possible M
Gu DC (Talus)	DC	35-39	Middle	M	M	Gu 8 and 9 both Possible M	Gu 8 and 9 both M	Gu 8 and 9 both Probable M
Gu DC (Calcaneus)	DC	45-49	Middle	M	M			

*Although the auricular surface was not available for observation in Gu 3, unfused clavicles indicate that age at death was between 20-30 years

** The pelvis was too damaged for sex estimation, so cranial morphology was used (Buikstra & Ubelaker, 1994) and yielded a result of "probable female"

***Gu 6 is estimated to be between 14-16 yrs at death based on epiphyseal fusion (Meijer, 2018; Seymour, 2019)

n/a = not observed

D.3: Function 1 Raw Results and Estimates for Trial 3

Table 31: *Function 1 Raw Results and Estimates for Trial 3*

Skeleton	Side	Function 1 Raw Result	Function 1 Sex Estimate
Gu 1	**Left**	-4.55676	Female
Gu 1	**Right**	-4.44544	Female
Gu 2	**Left**	-1.80888	Female
Gu 2	**Right**	-0.89863	Female
Gu 3	**Left**	-3.29381	Female
Gu 3	**Right**	-3.30626	Female
Gu 6	**Left**	-4.72497	Female
Gu 6	**Right**	-4.67087	Female
Gu 7	**Left**	-2.70301	Female
Gu 7	**Right**	-2.18143	Female
Gu DC	**Right**	-2.50663	Female