6-15-2024

Statistical Modeling of Knee Morphology and Material Properties Considering Diverse Populations

Gabrielle Jeannine Kindy

Statistical Modeling of Knee Morphology and Material Properties Considering Diverse Populations

Abstract

Total Knee Arthroplasty (TKA) is a widely performed surgical procedure aimed at alleviating pain and restoring function in patients with severe knee osteoarthritis. Despite its general success, disparities in postoperative outcomes have been observed across different racial and ethnic groups, with minority populations often experiencing less favorable results. One potential avenue for improving the generalizability of orthopaedic implants is using Statistical Shape and Intensity Models (SSIMs), which can be used to incorporate patient variability directly into the orthopaedic medical device development workflow through population-based finite element analysis.

This work aimed to construct an SSIM from a diverse subject set, incorporating male and female subjects of various ages from Asian, Hispanic, Black or African American, Native American, and White racial or ethnic groups using a novel registration method. Through t-tests and Analysis of Variance (ANOVA), significant differences in both shape and material properties across these demographic groups were detected. Notably, the analysis revealed literature-supported differences in bone size between the sexes and changes in bone material quality with age. While significant differences in bone morphology and bone quality among racial and ethnic groups were observed, further validation with a more balanced and robust training set is needed to confirm these findings. Finally, an innovative application to facilitate the utilization of these findings in the development of orthopedic devices was created. This work represents a significant step towards greater inclusivity and personalized care in orthopedic device development.

Document Type

Masters Thesis

Degree Name

M.S. in Mechanical Engineering

First Advisor

Peter J. Laz

Second Advisor

Chadd W. Clary

Third Advisor

Charles S. Reichardt

Keywords

Anatomic variability, Diverse populations, Knee joint, Statistical intensity modelling, Statistical shape modelling, Total knee arthroplasty

Subject Categories

Biological Engineering | Biomechanical Engineering | Biomedical Engineering and Bioengineering | Engineering | Mechanical Engineering | Statistics and Probability

Statistical Modeling of Knee Morphology and Material Properties Considering Diverse Populations

A Thesis

Presented to

the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Gabrielle Jeannine Kindy

June 2024

Dr. Peter J. Laz

Author: Gabrielle Jeannine Kindy
Title: Statistical Modeling of Knee Morphology and Material Properties Considering Diverse Populations
Advisor: Dr. Peter J. Laz
Degree Date: June 2024

Abstract

Total Knee Arthroplasty (TKA) is a widely performed surgical procedure aimed at alleviating pain and restoring function in patients with severe knee osteoarthritis. Despite its general success, disparities in postoperative outcomes have been observed across different racial and ethnic groups, with minority populations often experiencing less favorable results. One potential avenue for improving the generalizability of orthopaedic implants is using Statistical Shape and Intensity Models (SSIMs), which can be used to incorporate patient variability directly into the orthopaedic medical device development workflow through population-based finite element analysis.

This work aimed to construct an SSIM from a diverse subject set, incorporating male and female subjects of various ages from Asian, Hispanic, Black or African American, Native American, and White racial or ethnic groups using a novel registration method. Through t-tests and Analysis of Variance (ANOVA), significant differences in both shape and material properties across these demographic groups were detected. Notably, the analysis revealed literature-supported differences in bone size between the sexes and changes in bone material quality with age. While significant differences in bone morphology and bone quality among racial and ethnic groups were observed, further validation with a more balanced and robust training set is needed to confirm these findings. Finally, an innovative application to facilitate the utilization of these findings in

the development of orthopedic devices was created. This work represents a significant step towards greater inclusivity and personalized care in orthopedic device development.

Acknowledgements

My journey has been marked by challenges and triumphs, and I am deeply grateful to the individuals whose support has shaped my path and helped me reach this milestone.

To my supervisor, Dr. Peter Laz, thank you not only for the guidance and mentorship, but for your patience as I figured out my path in life. Having someone who believes in you makes all the difference. I am also deeply grateful to my thesis committee members, Dr. Chadd Clary and Dr. Chip Reichardt, whose influence on my academic journey extends far beyond the classroom. I also wish to acknowledge the graduate assistantship support from DePuy Synthes and the Mechanical and Materials Engineering Department.

To Ola Alsaadi – a sincere thank you for your invaluable assistance in segmenting subjects for this research. Collaboration with you laid the groundwork for this project, and I am profoundly grateful for your contributions. To my other colleagues – I do not consider myself a "work friend" kind of person, but it's hard not to be with people like you. Thank you for the environment you create when we come to work every day. To my friends – Rylan, Brandon, Arrison, and Jalen – I know I am not always very present in my personal relationships, and I am grateful for your unwavering support despite this. Having friends who "get" you is a true blessing.

To my parents – with each passing day, I grow increasingly grateful for the effort you poured into raising me. Thank you for providing an environment that was both demanding and unconditionally loving. I am who I am today because of you. To my sister Alexie - thank you for being my confidante, my partner in mischief, and my unwavering supporter through thick and thin. Life is better with a sister.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 An Overview of the Clinical Landscape

1.1.1 Osteoarthritis

Osteoarthritis (OA), once thought to affect only the articular cartilage of diarthrodial joints, is now recognized as a group of overlapping joint disorders involving the entire joint. OA is classified by its causes as either primary, which lacks a known cause and is associated with risk factors such as increasing age, genetics and, as recently suggested, low grade systemic inflammation, or secondary, which arises due to known causative factors such as trauma, surgery on the joint structures, or abnormal joints or joint loading (Martel-Pelletier et al. 2016). The common thread linking the disorders of OA is a failure of joint repair following alterations in one or more of the joint tissues that ultimately leads to joint failure.

OA typically progresses slowly over several years. Among the structural damages to the joint are cartilage degradation, osteophyte formation, subchondral bone changes, and meniscal alterations, which manifest as symptoms like pain, morning stiffness, and joint crepitus (a grating sensation during joint movement) (Martel-Pelletier et al. 2016), with pain being the most prevalent symptom (Perrot 2015). Pouli et al. conducted 24 structured interviews with 17 women and 7 men with physician-diagnosed knee OA with the aim of gaining insight into the daily experience of people living with knee OA. Pain emerged as the focal point of the subjects' OA experience, with many expressing a

constant awareness of it, describing it as a perpetual presence. Their accounts highlighted how pain dominates their lives, intertwining with their very existence and becoming their primary concern (Pouli et al. 2014). Mary (pseudonym), a 72-year-old participant, poignantly described her struggle with knee OA:

> "I describe the pain as living with misery ...it makes you feel miserable ...makes you feel that if this is all ...it is for the rest of your life ...and I have always been a really independent person and now I've got to be dependent on someone else ...and that's been killing me" (Mary, 72 years) (Pouli et al. 2014).

Mary's words encapsulate the profound impact of knee OA on individuals' lives. With lived experiences like these, and estimates projecting the annual incidence of symptomatic knee OA at 240 per 100,000 patients per year (Varacallo, Luo, and Johanson, n.d.), it is perhaps unsurprising that osteoarthritis has a significant impact on disability in adults. In the Global Burden of Disease Study 2010, musculoskeletal disorders like OA were among the top causes of years lived with disability (YLDs) among all conditions. Knee OA was among the top 25 causes of YLDs globally (Varacallo, Luo, and Johanson, n.d.). In 2019 about 528 million people worldwide were living with osteoarthritis, an increase of 113% since 1990 (Long et al. 2022). As populations age and obesity rates increase, prevalence of OA is projected to continue rising (Leifer, Katz, and Losina 2022).

Despite its widespread impact, to date, OA remains a challenging condition to manage (Martel-Pelletier et al. 2016). Currently, medical management primarily focuses on relieving symptoms, as there's no pharmacologic therapy available to prevent disease onset or progression (Shahid and Singh 2016). Non-pharmacological interventions play a crucial role and often involve education, exercise, weight management with realistic

goals, and regular monitoring. These approaches can be complemented with pharmacological interventions targeting pain relief. Options include topical treatments like capsaicin, oral medications such as NSAIDs, paracetamol, selective COX2 inhibitors, and opioids for patients unable to tolerate NSAIDs. Injectable treatments like intra-articular corticosteroids and hyaluronic acid are also available, though the efficacy of the latter is subject to debate (Martel-Pelletier et al. 2016).

1.1.2 Total Knee Arthroplasty

When non-surgical options prove ineffective for treating patients with severe end-stage OA, surgical options become necessary to maintain function and alleviate pain (Shahid and Singh 2016). Total Knee Arthroplasty (TKA) is the most common surgical intervention for knee OA (Martel-Pelletier et al. 2016), as it delivers reliable outcomes for OA patients, relieving pain, restoring function, and enhancing quality of life (Varacallo, Luo, and Johanson, n.d.). Recently, there has been a renewed interest in using Unilateral Knee Replacements (UKAs) to treat patients with end stage unicompartmental OA (typically medial) with the aim of preserving more of the knee's natural anatomy. However, the procedure's adoption has been limited by its higher revision rates relative to TKA (Atik, Hangody, and Turan 2023). Broadly, both procedures entail removing and replacing damaged compartments of the knee joint with artificial components—UKA replacing one and TKA replacing all compartments. Annually, approximately 400,000 primary knee replacement surgeries are conducted in the US, with the primary clinical diagnosis associated with this procedure being primary, end-stage, tri-compartmental osteoarthritis (Varacallo, Luo, and Johanson, n.d.).

TKA is widely acknowledged as an effective treatment for end-stage OA, providing rapid and significant relief from pain, improved functional status, and enhanced overall health-related quality of life in approximately 90% of patients who undergo the procedure ("NIH Consensus Statement on Total Knee Replacement" 2003). The success of TKA in addressing what is anticipated to become a more prevalent condition suggests an anticipated surge in its usage. Between 1991 and 2010, the annual primary TKA volume in the US Medicare population alone surged by 161.5%, escalating from over 93,000 to more than 226,000 cases (Varacallo, Luo, and Johanson, n.d.). Projections indicate that this upward trend will persist; compared to 2005 rates, annual TKA utilization in the USA is forecasted to soar by 673% by the year 2030 (Kurtz et al. 2007).

Given the current and projected prevalence of TKA in treating OA, it is essential to address the procedure's limitations. Despite being a dependable and consistently successful surgery for patients with severe knee OA, studies indicate that as many as 1 in 5 patients who undergo primary TKA are still dissatisfied with the outcome (Varacallo, Luo, and Johanson, n.d.). This suggests that TKA is not achieving its goal of relieving pain and restoring function in a substantial proportion of patients. Patient dissatisfaction post-TKA is linked to a range of factors, including unmet expectations, certain preoperative conditions like older age and living alone, and postoperative complications necessitating hospital readmissions (Bourne et al. 2010).

While research has shown that sex does not significantly impact the outcomes of TKA (Ritter et al. 2008), differences in outcomes following TKA have been observed across different racial and ethnic groups, with minority groups often experiencing

disproportionately poorer results (Shahid and Singh 2016). In a comprehensive review examining racial/ethnic disparities in Total Joint Arthroplasty utilization and outcomes, Shahid et al. found that in the US, Black patients undergoing TKA faced higher rates of hospital readmissions, longer hospital stays, increased non-infection and infection-related complications, worse postoperative functional status, more pain, and elevated revision rates compared to White patients (Shahid and Singh 2016). The results of a review by Rudisill et al. concurred, finding also that Black patients had higher rates of non-home discharges and mortality following TKA compared to White patients (Rudisill et al. 2023). These disparities persist even within equal-access healthcare systems; Ibrahim et al. revealed that Black veterans receiving TKA through the Veterans Health Administration (VHA) experienced higher complication rates than their White counterparts, despite the VHA offering care to veterans regardless of race, ethnicity, income, or other socio-economic factors (Ibrahim et al. 2005). Hispanic patients also exhibited higher rates of complications, prolonged hospital stays, and non-home discharges following TKA relative to other patient groups, while Asian patients showed lower complication rates, but a higher likelihood of non-home discharge compared to White patients. Additionally, studies involving American Indian/Alaska Native (AIAN) and Pacific Islander patient cohorts indicated longer hospital stays and, for AIAN patients, higher inpatient mortality rates compared to White patients (Rudisill et al. 2023).

1.1.3 Lack of Diversity in Medical Device Development

The higher rates of complications and less significant improvements in patient-reported outcomes among minority groups likely stem from a multitude of complex factors. These may include patient factors (e.g. presence of medical comorbidities, status of preoperative mental health, socioeconomic status), provider and healthcare system factors (e.g. provider bias, provider and healthcare system quality, disparities in healthcare coverage), and societal factors (e.g. institutional and structural racism, racial residential segregation) (Hu et al. 2022). Another potential contributor to disparities in orthopedic outcomes is a lack of research equity. Despite efforts to address the underrepresentation of women and historically marginalized populations in U.S.-based clinical trials and research by federal agencies like the NIH Office of Research on Women's Health and the FDA Office of Minority Health and Health Equity, underrepresentation of these groups persists across various medical fields and diseases, including cardiology, oncology, Alzheimer's Disease, and HIV/AIDS (Committee on Improving the Representation of Women and Underrepresented Minorities in Clinical Trials and Research et al. 2022). Orthopaedics is no exception. In a systematic review of randomized controlled trials in three leading orthopaedic journals, Cwalina et al. found that the racial and ethnic distribution of participants skewed heavily towards white individuals, with fewer Black, Hispanic, and Asian participants compared to their representation in the general population according to the 2019 United States Census (Cwalina et al. 2022). Similar findings were reported by Issa et al. in trials specifically focused on high-risk orthopaedic medical devices, where White participants were

overrepresented, while Black, Asian, and Hispanic participants were significantly underrepresented (Issa et al. 2023). Given that research has shown that groups often excluded from clinical trials may have unique disease presentations or health conditions that impact their response to treatments (Committee on Improving the Representation of Women and Underrepresented Minorities in Clinical Trials and Research et al. 2022), an absence of racial and ethnic diversity in orthopaedic clinical trials could restrict the generalizability of orthopaedic solutions like TKA to minority populations and perpetuate disparities in treatment use, complications, and functional outcomes (Cwalina et al. 2022).

Clinical trials, such as those used to validate medical devices like the implants utilized in TKA, represent just one stage in the medical device design process where diversity, or lack thereof, can have significant implications. Alongside the Validation step, which assesses if the right product has been built after the device is designed, the highly regulated medical device development process also includes a pre-clinical Verification step, questioning whether the right product is being built during the design phase itself. Methods for design verification encompass risk analysis, rapid prototyping, and finite element analysis (FEA) (Aitchison et al. 2009).

FEA has played a crucial role in orthopedic biomechanics for over four decades, aiding in the understanding of bone-implant systems' behavior to facilitate implant design and pre-clinical testing. However, traditional FEA approaches like comparative or parametric study designs often rely on models based on a single representative bone or joint (Taylor and Prendergast 2015), neglecting population variability and limiting the

generalizability of results (Galloway et al. 2013). Recent advancements have introduced statistical methods, notably probabilistic analyses, which enable the simultaneous exploration of multiple design and environmental parameters, including patient-to-patient variability (Taylor and Prendergast 2015).

Earlier attempts to incorporate patient variability in FE models involved manually modeling a small cohort of subjects or scaling the size and/or material properties of a single bone. An alternative approach employs statistical shape and/or intensity models, which are statistical representations of the morphology and material properties of the bone segment or joint. These models can be used to generate hundreds to thousands of synthetic instances for population-based FE models based on a smaller training set of bones or joints (Taylor and Prendergast 2015). Importantly, the demographic makeup of this training set influences anatomic variability in the synthetic subject pool, emphasizing the necessity of considering the diversity of the training set for the probabilistic analyses to yield generalizable results.

Including demographic diversity in research requires careful consideration of responsible practices. Unlike physiological variables such as age and sex, racial and ethnic identities are socially constructed and historically fluid concepts, lacking biological basis (Cwalina et al. 2022). Most scholars across disciplines like evolutionary biology and anthropology agree that racial distinctions are not genetically discrete, reliably measured, or scientifically meaningful in a biological context (Smedley and Smedley 2005). Therefore, when including racial and ethnic diversity in scientific research, particularly in studies aiming to highlight differences between groups, it is

imperative to avoid the errors of 19th-century racialized science, which viewed races as phenotypically distinct subdivisions of the human species based on ancestral geographic origins (Smedley and Smedley 2005). The Canadian Medical Association Journal underscores this in its updated guidance on reporting race and ethnicity in research articles, emphasizing that race should not be presented as a surrogate for biological or genetic variation (Stanbrook and Salami 2023).

However, this does not negate the importance of including racially and ethnically diverse subjects and reporting their demographics. While race lacks biological significance, social race remains a significant determinant of access to societal resources and barriers to full inclusion, as with the presence of racial and ethnic disparities in health care. The fact of inequality means that race and ethnicity continue to matter in important ways (Smedley and Smedley 2005). Consequently, representation by self-identified race and ethnicity remains crucial for the generalizability of study findings (Committee on Improving the Representation of Women and Underrepresented Minorities in Clinical Trials and Research et al. 2022). Researchers should therefore strive to ensure diversity in their subject population while adhering to established best practices outlined by reputable sources. This includes reporting race and ethnicity alongside other pertinent biological and social factors to gain a comprehensive understanding of study results (Cwalina et al. 2022; Stanbrook and Salami 2023). By doing so, researchers ensure consideration and understanding of all critical factors, both genetic and non-genetic, essential for generalizing findings to the broader population (Committee on Improving the

Representation of Women and Underrepresented Minorities in Clinical Trials and Research et al. 2022).

1.2 Objectives

As the incidence of knee OA continues to rise, the demand for effective and personalized treatment options becomes increasingly critical. Despite the success of TKA in many patients, significant challenges remain, including disparities in outcomes across different demographic groups. To address these issues, it is essential to understand variation in the population and to incorporate demographic diversity into the development and validation of orthopedic devices, ensuring that treatments are effective across varied populations. Accordingly, the objectives of this thesis were:

1) to develop a Statistical Shape and Intensity Model (SSIM) to quantify shape and material property variation in the femur and tibia in a diverse population using a novel general regression neural network-based registration workflow,
2) to use this model to investigate the impact of gender, race, and age on knee shape and material properties,
3) to build upon a previously developed graphic user interface (GUI) to facilitate the use of both the SSM and SIM portions of the model to support medical device development workflows.

1.3 Thesis Organization

Chapter 2 provides an overview of the literature related to bone morphology and material properties considering factors of gender, age, and ethnicity. Chapter 3 provides background on the development of statistical shape and intensity models. Chapter 4 describes generalized methods that were utilized to create the statistical shape and intensity models; some methods were previously developed, while others were advanced in the current work. Chapter 5 shares and discusses the results related to the morphological outcomes of the model, while Chapter 6 focuses on the material property outcomes. Chapter 7 discusses a GUI tool to support interfacing with the models, while Chapter 8 captures overall conclusions, limitations, and potential future work.

2. Literature Review

2.1 Overview

The literature review section delves into three key areas concerning knee morphology and material properties: sexual dimorphism, age-related differences, and racial/ethnic disparities, providing a comprehensive overview of current research in these domains, highlighting key findings, and identifying gaps in knowledge.

2.2 Sexual Dimorphism in Knee Morphology and Material Properties

Many studies have investigated the role of gender in knee morphology, which has led to the identification of now well-established anatomic differences in the knees of males and females. For example, females have been shown to have narrower mediolateral to anteroposterior aspect ratios, less pronounced anterior condyles, and greater quadriceps angle (Kim et al. 2017). Traditionally, these investigations into sexual dimorphism in knee morphology have focused on various anatomical features, including distance features, angular features, and curvature features. Asseln et al. conducted a comprehensive study aiming to compare the anatomical features of the femur and tibia between the sexes, utilizing a dataset of 412 femurs and tibias (248 female and 164 male). Their findings underscored the presence of gender-specific differences in certain parameters, emphasizing the need for parametric implant designs tailored separately for the AP and ML directions to provide adequate coverage for all patients (Asseln et al. 2018).

Statistical shape models have also been used to assess sexual dimorphism in knee morphology. Wise et al. explored the association between sex and the shape of outlines of the distal femur and proximal tibia derived from 2D radiographs using a statistical shape model. Their study, involving 339 femurs and 340 tibias without osteoarthritis, confirmed previous observations of sexual dimorphism in knee shape and extended the understanding of such differences through an analysis centered on 2D shape rather than anatomic measurements (Wise et al. 2016). Audenaert et al. expanded the investigation using statistical shape models capturing full 3D morphology. Their findings emphasized significant differences in knee width between male and female samples (Audenaert et al. 2019). Importantly, the agreement between the results of feature-based studies and statistical model-based studies lends validity to the statistical shape modeling approach in investigating morphological variation between populations in the knee (Wise et al. 2016).

In addition to morphological differences, researchers have also investigated sexual dimorphism in bone quality, particularly within the realm of understanding osteoporosis. The interplay of body size adds complexity to studying sex-related bone mineral density (BMD) variations, with males generally being larger than females. An analysis of NHANES data revealed that when correcting for differences in height and weight, femur BMD disparities between males and females in younger adults were nullified. However, in older adults, sex-related differences in BMD persisted even after such statistical corrections. Other studies have reported sex-differences in bone material properties following size correction, with Nieves et al. reporting greater BMD at the hip and distal tibia, as well as greater tibial cortical thickness in males. Conversely, several other

studies have reported that gender-related differences in bone mass disappear following correction for body size in both adults and children (Nieves et al. 2005). Given the importance of bone quality to the early fixation and long-term survivability of TKA implants, characterization of the material property variation in the bones of the knee remains a valuable area of investigation.

Despite this, there has been limited utilization of statistical intensity modeling to characterize variations in material properties within the femur and tibia. Bah et al. employed this approach to develop gender-based statistical shape and intensity models of the femur. Subsequent analysis involving the generation of 1000 new male and female geometries, respectively, revealed significant differences in gross anatomical measurements between genders, particularly in clinically relevant parameters such as femoral canal definition, bone densities, femoral head offsets, anteversion, among others (Bah et al. 2015). Although this study primarily focused on the proximal femur, it demonstrates and validates the process of using statistical intensity modeling to explore variations in material properties. Bruce et al. contributed to this line of inquiry by characterizing geometry and density variations within the tibial-fibular complex using statistical shape and appearance models. The results of their model were leveraged to conduct finite element analyses, shedding light on factors contributing to sex disparities in stress fractures. Their findings revealed differences in cortical density and cortical thickness between male and female tibias (Bruce et al. 2022; Bruce and Edwards 2023). To the author's knowledge, no work has yet investigated sex-related bone quality differences in both the proximal tibia and femur.

2.3 Age-Related Differences in Knee Morphology and Material Properties

In the context of age, significant changes in bone size have been documented. A comprehensive cross-sectional study involving 373 men and 323 women aged 20 to 97 years revealed that both sexes experienced an increase in bone area by approximately 15% over their lifetimes, suggesting ongoing periosteal bone growth throughout adult life (Riggs et al. 2004). This change in size has often confounded studies like that by Han et al, who analyzed 2D morphological parameters derived from Magnetic Resonance images of 535 patients (273 males and 262 females). The study identified significant age-related differences in several femoral parameters: femoral width, the distance from the distal and posterior cartilage surface to the medial/lateral epicondyle, medial posterior condylar offset, and posterior condylar angle (all $P < 0.001$) but was not able to determine whether these changes were related to shape or size for many parameters. No significant differences were found in lateral posterior condylar offset and medial/lateral tibial slopes. Additionally, a significant interaction between gender and age groups was noted in most parameters (Han et al. 2016).

Morphometric analyses, such as Principal Component Analysis (PCA), have been instrumental in distinguishing age-related shape differences independent of size. For example, Li et al. Found significant differences in the 3D morphometric measurements of the distal femur between the age groups in a Chinese Han population (K. Li et al. 2018a). This trend proved true across racial and ethnic groups, with Cavaignac et al. finding analogous results in a population of White subjects (Cavaignac et al. 2017).

The decline in bone quality with age is well-documented, with changes in volumetric bone mineral density (vBMD) playing a critical role. The aforementioned cross-sectional study from Riggs et al. found marked decreases in vBMD with age. Cortical vBMD remained stable until approximately 50 years of age before declining in both sexes, while trabecular vBMD began to decrease even in young adulthood, with continual loss observed throughout life. These decreases in vBMD were more pronounced in women than in men, aligning with the effects of menopause and the subsequent near-complete estrogen deficiency in women (Riggs et al. 2004). Despite extensive research on age-related changes in bone quality, statistical intensity models have yet to be leveraged to comprehensively investigate these changes. This novel analytical approach could provide fresh insights into how bone material properties evolve with age.

2.4 Racial and Ethnic Differences in Knee Morphology and Material Properties

To date, investigations into knee morphology have primarily focused on gender-related differences, with comparatively less emphasis on race and ethnicity. Existing literature primarily focusses on variations in morphometric data within or between White and Asian cohorts (Kim et al. 2017), with studies like those by Vaidya et al. (Vaidya et al. 2000), Li et al. (K. Li et al. 2018b), and Kwak et al. (Kwak et al. 2012) contributing to this discourse by shedding light on knee morphology variations in Indian, Chinese, and Korean populations, respectively.

To consolidate the findings of individual studies like those mentioned above, Kim et al. conducted a systematic review of the PubMed database to investigate the differences in morphologic features of the distal femur and proximal tibia among and within various

racial and ethnic groups. Their analysis of 23 studies involving East Asian patients, 11 studies with White patients, three studies of Indian patients, and three studies of Black patients unveiled key differences in knee morphology relevant to TKA. They found that White patients had larger femoral anterior-posterior measurements and a smaller femoral aspect ratio compared to East Asian patients. They also demonstrated that White patients had a larger tibial aspect ratio compared to Black patients (Kim et al. 2017). Perhaps most importantly, this work highlighted the scarcity of studies detailing anthropometric variations across historically underrepresented populations, with the authors noting that Black and Indian populations were underrepresented in their work, and African and Middle Eastern populations were not represented at all due to a lack of available data (Kim et al. 2017). While the work did not include other groups present in the US population such as Hispanic, Native American, and Pacific Islander groups, it is likely that they are similarly underrepresented in works like these, hindering a comprehensive analysis of ethnic and racial variation in knee morphology.

The authors of the aforementioned review note the importance of aligning TKA components with resected bony surfaces to minimize complications and extend implant survival. They note also that it is "conceivable" that ethnic and racial variability in knee morphology could potentially lead to suboptimal outcomes if not adequately addressed (Kim et al. 2017). Other studies assert a more direct link between ethnic and racial morphological variability and disparities in TKA clinical outcomes. Ho et al., in a study comparing the morphology of Asian knees with currently available TKA prostheses in Asia, suggested that the existing prosthetic designs may not accurately match the femoral

aspect ratio of Asian knees, thereby resulting in mediolateral component overhang among Chinese patients (Ho, Cheng, and Liau 2006).

As with those investigating the effects of gender, previous studies have mainly focused on comparing morphometric measurements across different racial and ethnic groups to understand variations in knee shape. The use of statistical models, in addition to providing a comprehensive 3D analysis of shape, provides the advantage of also supporting implant design through instance generation for pre-clinical FE modeling. However, statistical shape models considering diverse populations are limited, with Mahfouz et al. standing out as the only such model considering the knee. Their study compared 1000 adult knees from African American (n = 80), East Asian (n = 80), and White patients of European descent (n = 840) with the aim of identifying shape differences not only between racial and ethnic groups but also between genders within each race. They found (unsurprisingly) that males tended to have larger knees across all races and ethnicities. Moreover, African American females displayed distinct characteristics, such as a deeper patellar groove compared to their Caucasian counterparts, while African American males exhibited larger femoral and tibial dimensions compared to East Asian males (Mahfouz et al. 2015).

Beyond morphology, emerging evidence suggests differences in bone mineral density (BMD) across racial and ethnic groups. Non-Hispanic Black adults, for instance, exhibit higher BMD and lower rates of bone loss compared to other ethnicities. US Asian adults, on the other hand, may display lower BMD compared to non-Hispanic White adults, although data remains limited and inconclusive in some cases. It has also been noted that

the directionality of these differences may vary by skeletal site. For example, Mexican American adults showed higher mean BMD of the proximal femur compared with Non-Hispanic White adults, despite having a lower mean BMD for the total body and other sites (Noel, Santos, and Wright 2021). Despite these findings, there has been little research using statistical models to explore material property variation in the bones of the knee across ethnic and racial groups. To the author's knowledge, no statistical intensity models of the knee have been created from a racially and ethnically diverse training set, nor has any such model been leveraged to investigate the material properties of the tibia and femur across these groups, although Bah et al. noted that these methods can be used to assess anatomical variation between racial and ethnic groups, if training datasets can be accessed (Bah et al. 2015). Given the importance of bone quality to the success of TKA implants, this highlights a key area for future research to explore.

3. Theory and Background Related to Statistical Shape and Intensity Modeling

3.1 Overview

This section explores the fundamental concepts involved in the of construction of statistical shape and intensity models defining shape, shape registration, material property registration, dimension reduction, and statistical model evaluation, elucidating their significance and methodologies.

3.2 Defining Shape

Decisions about objects are often made using their sizes and shapes. For example, a patient's tumor might be treated using different protocols depending on its size. Similarly, a facial recognition program might allow (or deny) someone access depending on the shape of their face. Size and shape analysis has relevance in a wide variety of disciplines including biology, chemistry, medicine, image analysis, archaeology, bioinformatics, geology, particle science, genetics, geography, law, pharmacy, and physiotherapy (Dryden and Mardia 2016). The analysis of size and shape is an area with far-reaching implications and developing methods for doing so is therefore of significant importance.

In practice, the analysis of size and shape requires a way of describing shape, some notion of distance between two shapes, and methods for the statistical analysis of shape (Dryden and Mardia 2016). We begin first with the concept of shape and how it is defined and measured. In this text the definition of Shape by Dryden and Mardia is

adopted; “Shape: All of the geometrical information that remains when location, scale, and rotational effects are removed from an object” (Dryden and Mardia 2016). In other words, two objects have the same shape if they can be translated, rotated, and scaled to each other so that they match exactly. This concept is illustrated in Figure 3-1, where each of the hands depicted is a transformation (consisting of a translation, scale, and/or rotation) of the other.

Figure 3-1: Four hand shapes with different locations, scales, and rotations. Taken from Stegmann and Gomez (Stegmann and Gomez, n.d.).

A practical way to describe the shape of an object is through a finite set of points on the surface of the object in question (Figure 3-2). These points, which summarize key geometrical information, are referred to as landmarks. The landmarks that describe a shape can be defined in several ways. Scientific landmarks are assigned to points that meaningfully correspond between objects (Dryden and Mardia 2016). For example, if the shape of an eye is being described, anatomical points like the center of the pupil or the

inner corner of the eye might be used as scientific landmarks. Mathematical landmarks correspond to points located on an object according to some mathematical or geometrical property like the extremum of a curvature. If more landmarks are required than can be obtained through scientific or mathematical methods, pseudo-landmarks can be constructed on an object, either around the boundary of the object or in between scientific or mathematical landmarks (Stegmann and Gomez, n.d.).

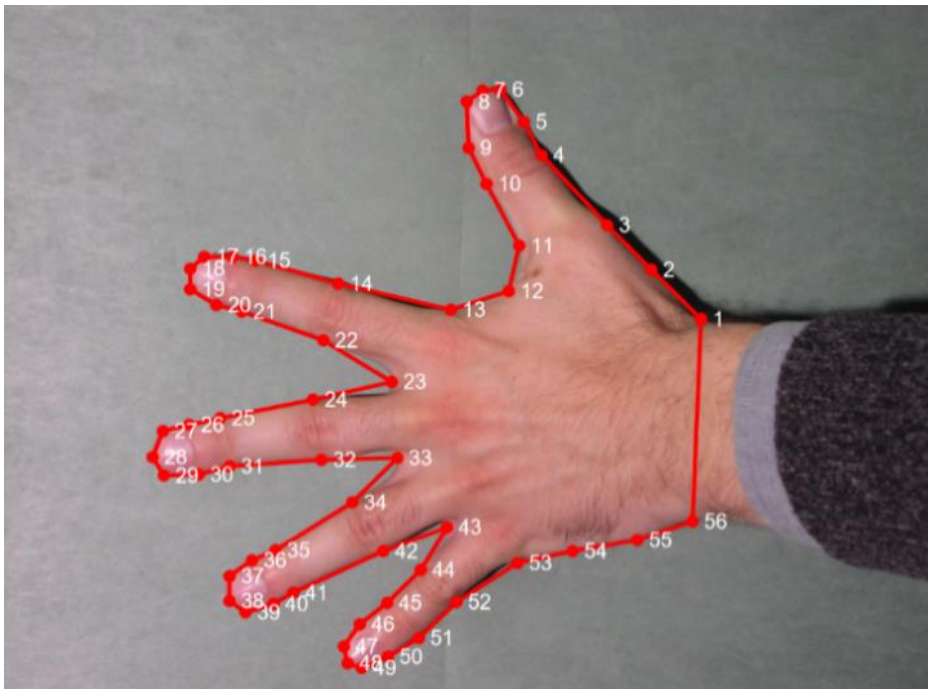

Figure 3-2: Example of a set of landmarks describing the shape of a hand. Taken from Stegmann and Gomez (Stegmann and Gomez, n.d.).

The landmarks that describe a shape can be acquired through manual point identification. However, this method is impractical when defining shapes through many landmarks or when dealing with highly deformable objects. Instead, these point sets are

generally acquired using imaging devices such as 3D scanners, manual modeling software, or from images processed via segmentation or isosurfacing (van Kaick et al. 2011). Once these landmarks are established and their coordinates are identified, the shape can be described mathematically by a point set. For example, an n-point shape in k dimensions could be described by concatenating each dimension into an nk-vector. The vector representation for a planar (k = 2) shape S would be:

$$\mathbf{S = [x_1, y_1, x_2, y_2, \ldots, x_n, y_n]^T}$$

where x and y are the 2D coordinates of each of the n points describing the shape (van Kaick et al. 2011).

3.3 Shape Registration

While defining shape has its applications (i.e. in facial recognition), often the ultimate goal is to depict or describe the size and shape changes in a class of objects (Dryden and Mardia 2016). Before extracting this information from a set of shapes, some essential pre-processing steps must be completed. The goal of this preprocessing, termed registration, is to assign meaningful correspondences between the point sets describing the shapes in a class (Myronenko and Xubo Song 2010). The problem of registration can be generally stated as: "Given input shapes **S1**, **S2**,…,**SN**, find a meaningful relation **R** between their elements" (van Kaick et al. 2011).

Defining a meaningful correspondence between shapes is the challenging part of registration. In some cases, the geometry of the shape can provide enough information and correspondence can be established between elements of two or more shapes that

possess similar structure in terms of shape and local context (van Kaick et al. 2011), as with the top of the fin or the eye of the fish in Figure 3-3. In other cases, however, the geometry and number of corresponding parts of the shape, as well as the topology between these elements, may vary greatly. Correspondence problems between these types of shapes involve understanding both the structure (and possibly function) of the shapes at the local and global levels (van Kaick et al. 2011).

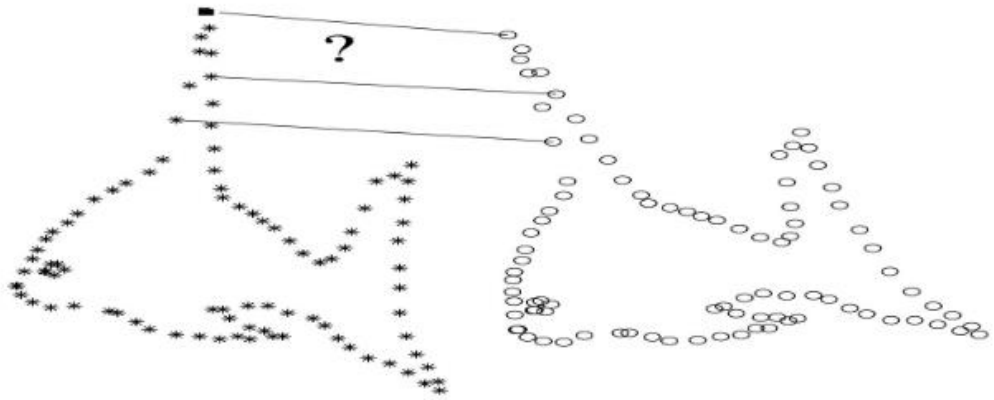

Figure 3-3: Point Set Registration. Taken from van Kaick et al (van Kaick et al. 2011).

Registration can be further differentiated based on the type of correspondence needed. First, depending on the inclusion properties of the relation **R**, the type of registration can be categorized either as full or partial. A full correspondence is defined for the entirety of the shape, while a partial correspondence is defined only for a subset of elements. The density of the relation is another way in which registration can be classified. In a dense correspondence, the relation is defined for all elements of a shape. In contrast, a sparse correspondence is defined for only a small number of pre-selected elements. Finally, group correspondence involves computing correspondences between the shapes in a

group. This can be done either by registering all the pairs of shapes in the group or by registering all the shapes simultaneously, a process that is less straightforward but has its advantages (van Kaick et al. 2011).

Once the type of correspondence problem is defined, correspondence can be established. Correspondence can be obtained directly through feature matching, or it can be obtained through rigid or non-rigid alignment. In feature matching, the similarity between pairs of feature points is estimated by optimizing an objective function composed of two terms: one which maximizes the similarity between shape descriptors and one which minimizes distortion between the shapes (Figure 3-4b).This method can be applied whenever it is possible to compute a set of descriptors for the shape's elements. An alternative method for establishing correspondence involves first aligning the shapes and then deriving correspondence from the proximity of the aligned elements (Figure 3-4c and Figure 3-4d) (van Kaick et al. 2011). In this project, a dense correspondence problem between a group of full shapes defined by point elements is considered. The correspondence between these shapes is obtained through alignment, a process discussed in depth in the next section. This correspondence problem will hereon be referred to as 'Registration'.

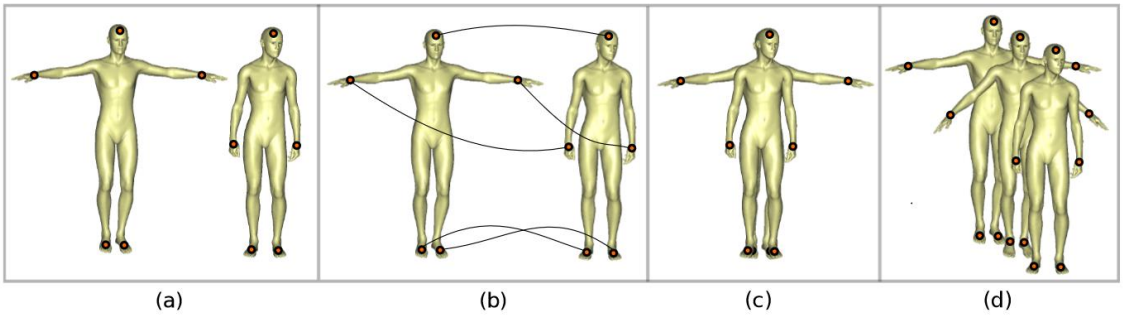

Figure 3-4: Correspondence between the shapes shown in (a) could be established using feature matching (b) or by first rigid body aligning the shapes (c) and then non-rigidly deforming one onto the other (d). Taken from van Kaick et al. (van Kaick et al. 2011).

3.3.1 Alignment

3.3.1.1 Rigid Alignment

Registration begins by establishing a coordinate reference with respect to position, scale, and rotation. In practice, this coordinate reference is usually defined using the template point set – often the median shape within the set of shapes being compared. The remaining point sets, termed instances, are rigidly aligned to the template, bringing them into shape space (Dryden and Mardia 2016). Per van Kaick et al., the problem of rigid alignment can be posed as: find the rigid transformation that maximizes the number of points in $\mathbf{S}_1$ that align to the points in $\mathbf{S}_2$ (van Kaick et al. 2011). To understand this definition, two terms must be further defined. Rigid transformation is defined as a transformation in which the distances between points are preserved and no deformation occurs (Figure 3-6). It is parametrized by the variables of rigid body motion: translation and rotation (Golyanik 2020). Alignment is defined as a certain threshold of proximity between points; when two points are close enough, they can be considered aligned with each other. A measure of alignment can be given either through the largest common point

set method, in which the best alignment is considered as the one with the largest number of points aligned or through geometric distance, which minimizes the sum of squared distances between points and their nearest neighbor (van Kaick et al. 2011). Thus, rigid alignment can be thought of as the process of translating and rotating one shape so that its points most closely match the points of the second shape.

The Iterative Closest Point algorithm (ICP) is a classical method for rigid registration first proposed in 1992 by Besl and McKay (Besl and McKay 1992). Broadly, ICP consists of two main steps: a correspondence step and an alignment step. In the correspondence step, ICP establishes point-to-point relationships between two sets of points (Figure 3-5a) by pairing each point in one set with the closest point in the other set (Figure 3-5b). This initial matching forms the basis for subsequent alignment. Following correspondence, the alignment step seeks to minimize the distances between corresponding points by finding an optimal transformation. This transformation is then applied to adjust the positions of the points (Figure 3-5c). The correspondence and alignment steps are then repeated ("iterated") until the corresponding point to point distances converge within a specific tolerance (Figure 3-5d) (Cyrill Stachniss 2020).

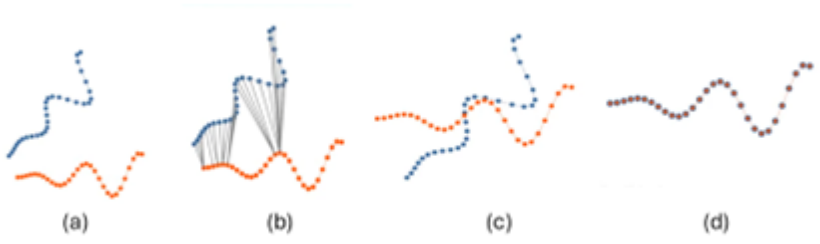

Figure 3-5: Steps involved in the Iterative Closest Point algorithm. Adapted from Kim (Jangseob Kim 2020).

This iterative nature of ICP ensures convergence, making the algorithm widely adopted for its simplicity and reliability. However, the limitations of ICP require practical considerations. The success of ICP relies strongly on a good initial alignment, and the algorithm is sensitive to outliers and partial overlaps in the data (Biham 2022). Refining the data to remove outliers and preprocessing the data into a consistent local coordinate system can help mitigate issues with the implementation of ICP.

3.3.1.2 Nonrigid Alignment

Once they are aligned in shape space, the correspondence between point sets can be established. This is accomplished through non-rigid alignment (Golyanik 2020), informally known as “deformation” or “morphing”. The simplest nonrigid transformation is affine, which allows for anisotropic scaling and skews (Myronenko and Xubo Song 2010). Affine transformations preserve collinearity and ratios of distances between points on a line and can be thought of as the linear “stretching” of a shape (Figure 3-6). However, when there are more than k + 1 landmarks in k dimensions, affine transformations will not yield an exact fit and a non-affine shape deformation is required (Dryden and Mardia 2016). Non-affine shape transformations allow for arbitrary deformations within the constraint that point topology is preserved (Figure 3-6). This constraint prevents intersections between the displacements of the points and self-intersections of the surfaces (or volumes) represented by point sets (Golyanik 2020).

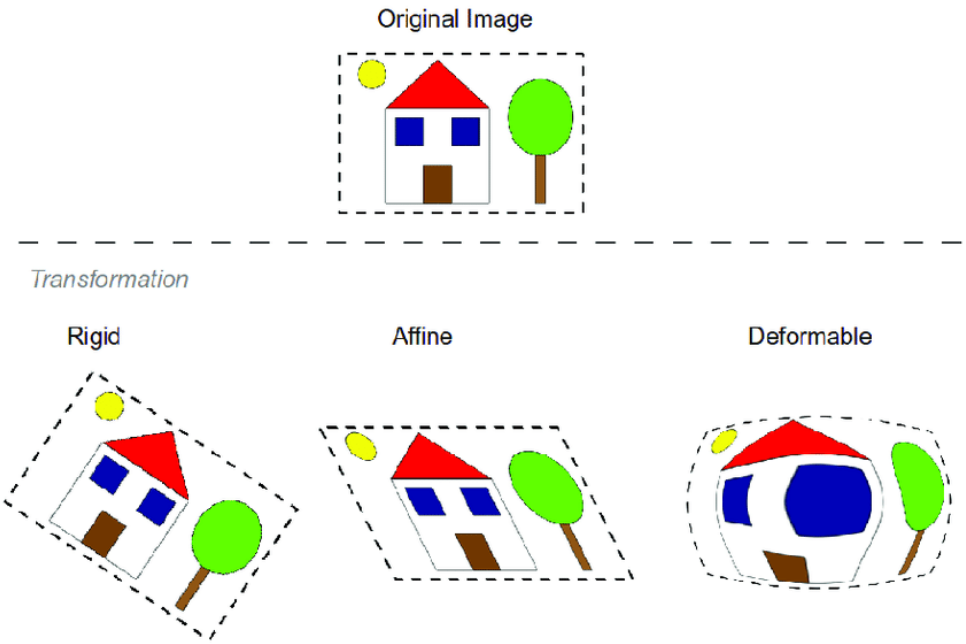

Figure 3-6: Types of deformation used in point set registration. Taken from Temerinac-Ott (Temerinac-Ott 2012).

There are various approaches to non-affine shape deformation, including the use of thin-plate splines, transformation grids, finite element methods, biorthogonal grids, and kriging (Dryden and Mardia 2016). Two deformation methods are of interest to this project: the Coherent Point Drift (CPD) Algorithm and Generalized Regression Neural Networks (GRNN).

CPD approaches the problem of non-rigid deformation by treating the point sets as probability distributions, with one point set representing the Gaussian Mixture Model (GMM) centroids and the other representing the data points. It seeks a smooth transformation that maximizes the likelihood of the GMM centroid given the data point. It does this iteratively through the Expectation-Maximization (EM) algorithm, where in

each iteration, it estimates the probability of correspondence between points and adjusts the transformation parameters accordingly. A key aspect of CPD is its ability to enforce coherent movement of data as a group, preserving the topological structure of point sets and ensuring smooth deformation (Myronenko and Xubo Song 2010). CPD's strength lies in its ability to reliably handle complex transformations of large data sets. Its probabilistic framework allows it to incorporate uncertainty in the data, making it robust even in the face of noise, outliers, and missing points. However, its accuracy comes with tradeoffs in terms of large RAM requirements and computational time. To mitigate these issues, a low-rank matrix approximation can be employed, reducing computational time with only a moderate loss of accuracy (Myronenko and Xubo Song 2010).

Radial Basis Function (RBF) Networks have also been used to successfully address non-rigid deformation problems (Zhang, Ackland, and Fernandez 2018; Qingqiong Deng, Zhou, and Wu 2010), having demonstrated the ability to deform meshes with errors that are significantly smaller than the mesh resolution, even when dealing with complex anatomical structures (Zhang, Ackland, and Fernandez 2018). Generalized Regression Neural Networks (GRNN) are a type of artificial neural network belonging to the RBF family. These networks have gained attention in addressing non-rigid deformation challenges, particularly in mesh deformation tasks. Recently, Andreassen et al. proposed a GRNN-based method for deformation, which demonstrated improved performance compared to existing RBF algorithms. This approach tackles non-rigid deformation through two main steps: correspondence and deformation. Like ICP, correspondence is established by proximity, employing methods like k-nearest neighbor (KNN) or projected

distance search to find the nearest points between two sets. A displacement field created from the correspondence is then used to train a GRNN, which predicts desired nodal displacements. Following application of the nodal displacements, this iterative process continues until convergence is achieved. By initially applying this process to downsampled point sets and then using the results to deform the original denser meshes, the algorithm can efficiently handle high-density meshes with minimal computational overhead (Andreassen et al. 2024). While the decreased computational time of the GRNN-based approach presents a significant advantage over CPD, especially in the context of meshes with 3D element types, it is worth noting the user involvement necessary to tune parameters when using this algorithm.

3.3.2 Establishing Correspondence

Once the template has been deformed onto the instance, proximity between points on each shape can be used as a measure of correspondence (van Kaick et al. 2011). In other words, the template points correspond to whichever instance points are closest to them post-deformation. Ultimately, the goal of obtaining correspondence is to describe the shape of each object using a consistent set of landmarks, allowing for each of the shapes to be described in a mathematically consistent manner. This process yields a matrix whose columns correspond to the S-vectors describing the shape of each respective object in the set. This matrix is termed the Shape Register.

3.4 Material Property Registration

Until now, the construction of statistical models has been discussed primarily in the context of shape, which tends to be more straightforward to grasp. However, the foundational concept of creating consistent descriptions for comparison is equally applicable to bone material properties. Correspondence between the internal locations of a bone (or any shape) is assigned by deforming a volumetric mesh with internal nodes onto each instance (Grassi, Väänänen, and Isaksson 2021). The elements of this mesh serve an analogous role to the points that describe the surface of a shape. For each specimen, the pixel or voxel intensities from the Computed Tomography (CT) scan, given in Hounsfield Units (HU), corresponding to each element are captured and processed. This process results in a matrix where each column consistently describes the material properties of a subject (Grassi, Väänänen, and Isaksson 2021): the Material Property Register.

In the realm of Statistical Intensity Models and other FEA related applications, bone material properties are typically represented as bone mineral density (BMD), which measures the amount of bone mineral per volume of bone (Grassi, Väänänen, and Isaksson 2021). To assign these material properties from a CT scan, the relationship between BMD and HU must be established. This process is commonly referred to as CT Scan Calibration.

The gold standard of CT scan calibration involves the use of a calibration phantom — a physical object containing known concentrations of substances like calcium hydroxyapatite (CaCO3 or CaHA) or hydrogen dipotassium phosphate (K2HPO4 or

KHP) (Figure 3-7a). This phantom is scanned alongside the patient (Figure 3-7b). By comparing the known BMD values of mineral samples within the phantom with their corresponding HU values obtained from the CT scan, a calibration curve is generated. This curve serves as a reference for determining the BMD values of bone structures in the scan (Eggermont et al. 2019).

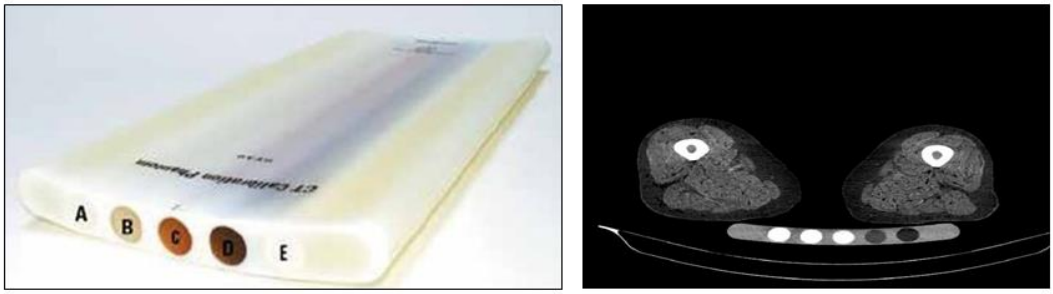

Figure 3-7: (a) CT calibration phantom. Taken from Montaseri et al. (Montaseri, Alinaghizadeh, and Mahdavi 2012). (b) Transverse slice of a CT scan taken with a calibration phantom

Unfortunately, the use of calibration phantoms in clinical CT scans is not a widespread practice due to their limited availability and high cost. To address this issue, a phantomless calibration method can be employed to assign material properties from scans lacking a phantom. Several approaches have been proposed for phantomless calibration. Some studies utilize a calibration function derived from a separate scan containing a calibration phantom. Others determine calibration factors based on CT scans with a calibration phantom and then apply these factors to scans without phantoms. In these cases, BMD can be calculated using a regression model based on previous phantom calibration (Eggermont et al. 2019).

Another method, like that proposed by Eggermont et al., utilizes scan-specific HU values of external air and tissues like fat and muscle for calibration. This involves extracting peaks for air, fat, and muscle tissue from a histogram of the HU within a standardized region of interest encompassing the patient's leg and surrounding air. These CT peaks are then linearly fitted to reference "BMD" values of the corresponding tissues to obtain a calibration function (Eggermont et al. 2019). Studies comparing phantom calibration with phantomless methods have demonstrated comparable results (Eggermont et al. 2019), making phantomless calibration a valuable tool for integrating clinical CT scans into statistical intensity models.

3.5 Dimension Reduction

Consider now, for a moment, the number of points you would need to accurately describe a shape. For a simple two-dimensional (2D) shape like a triangle, you may only need a few (read: three) points. However, when considering an irregular three-dimensional (3D) shape, like the Stanford Bunny in Figure 3-8, tens or even hundreds of thousands of points may be needed. Considering that a shape register consists of (n points x k dimensions) x p shapes, it is clear how comparisons across multiple instances of a complex shape can quickly result in a large dataset.

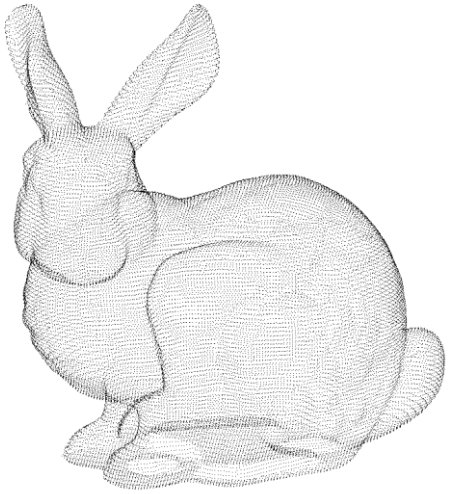

Figure 3-8: The "Stanford Bunny" as described by a point cloud. Taken from the Standford 3D Scanning Repository ("The Stanford 3D Scanning Repository," n.d.).

Understanding and interpreting the contents of a large dataset can be overwhelming, impractical, or even impossible. In such cases, reducing the dataset's size can be helpful, although care must be taken to minimize information loss. One option for reducing a dataset is to choose a subset of the original variables that adequately represent the entire pool. An alternative strategy is to build a smaller number of new variables from the original ones, so that each of the new variables is constructed from the original ones. Although this approach may seem less intuitive than choosing a subset, it offers the advantage of achieving a greater reduction in dimensionality for the same amount of information loss (Jolliffe 1990).

Principal Component Analysis (PCA) is the simplest of these variable-building techniques, often forming the mathematical basis for dimension reduction of the large datasets involved in Statistical Shape and Intensity Models. Its simplicity lies in the fact

that each new variable, known as a principal component (PC), is a linear combination of the original variables. The first principal component is the linear function that captures the maximum possible variance in the data. Subsequently, the second PC is derived to maximize variance while ensuring it is uncorrelated with the first PC. Similarly, the third PC maximizes variance while being uncorrelated with the first and second PCs, and so forth (Jolliffe 1990). Practically, each of these PCs represents a dominant pattern of variation in the original data. In the context of shape, each PC signifies a distinct "way" in which the shape varies across the dataset, with the first PC explaining the most significant way the shape changes and subsequent PCs representing other ways the shape changes in order of decreasing importance. The results are analogous when considering material properties.

3.6 Statistical Model Evaluation

The accuracy of statistical models is dependent on the quality and accuracy of the data used to create them, which is impacted by the image resolution and quality, and the segmentation and registration processes. The performance of these models is typically assessed with compactness, a measure of how much variability is explained in early modes, and with a leave-one-out evaluation, which assess the ability of the model to describe the behavior of a left-out-subject. Building on these prior measures, Audenaert et al. recently introduced a comprehensive set of metrics to assess the effectiveness of statistical models:

- Model compactness: Compactness refers to the variability or dispersion in the shape data captured by the statistical model. A compact model suggests that the

shape variations within the dataset are well-represented by a small number of principal components or shape modes. Compactness is quantified as the cumulative explained variance of the Mth eigenmode obtained from the model's covariance matrix decomposition.

- Model accuracy: Considering the ability to represent instances within the training set, this test determines the minimum number of PCs needed to reproduce an instance within a specified accuracy threshold, answering questions about how much variance is sufficient to build a "good" model. It is computed as the average absolute difference between the model description and the original instance for each training set subject.
- Model Generalizability: Like the leave-one-out evaluation, this test evaluates whether the sample size of the training set is adequate by assessing how well the models can classify unseen data beyond the training set. It involves a series of leave-one-out tests on the training data, reporting the average reconstruction error of the unseen instance.
- Model Specificity: Specificity measures the realism of new instances randomly generated by the statistical model. It is assessed by generating a large set (typically 1000) of virtual instances and computing the difference between each virtual instance and its respective closest real sample in the training set.

The integration of these evaluation metrics enhances understanding of the capabilities and limitations of statistical models generated, giving credibility to the claim that they

represent a given population and ensuring that they serve their intended purpose in practical applications, like medical device development (Audenaert et al. 2019).

4. General Methods

4.1 Overview

This section provides a comprehensive overview of the methodologies used in constructing and analyzing the statistical shape and intensity model. It covers subject selection, scan processing, shape and material property registration, and post-processing. While this work builds on previously developed methods, novel methods are highlighted, including the development of the New Mexico Decedent Image Database (NMDID)-to-register pipeline and the GRNN-based registration method.

4.2 Subject Selection

The New Mexico Decedent Image Database (NMDID) is a comprehensive dataset containing whole-body CT scans and associated metadata for over 15,000 individuals who died in New Mexico between 2010 and 2017 (Edgar et al. 2020). This dataset offers an invaluable resource for accessing a diverse population of scans. For this work, an effort was made to gather a sample representative of the United States population (Barton et al., n.d.), including equal number of males and females of various ages from Black or African American, Native American, Asian, and Hispanic backgrounds. A sample of White subjects were obtained from the training set of an SSIM previously developed at the University of Denver (Bayoglu et al. 2020). The NMDID race and ethnicity metrics were self-reported by the decedents in the 2010 census. It is important to note that for the purposes of this work, "Hispanic" refers to decedents who identified as racially Hispanic,

reflecting the usage in the 2010 census where many New Mexicans selected Hispanic as an "Other" race option (Edgar et al. 2020). The inclusion criteria for this study required subjects to be aged between 25 and 80 years, with a body mass index (BMI) of less than 40, based on cadaver height and weight. Subjects were excluded if they had a recorded history of cancers, tumors, or malignancies, any recorded injuries to the lower limbs, a manner of death that could result in lower limb injuries, cadaver decomposition, or severe osteoarthritis (OA).

4.3 Preliminary Scan Assessment

Each subject file downloaded from the NMDID database contained 26 CT scans. Among these, three scans held importance for this project: the “thin bone torso” and “thin bone lower extremity” scans, utilized to acquire subject bony geometry, and the “thin, soft-tissue lower extremity” scans, used to obtain the subject bony material properties. Before processing, a preliminary assessment of each subject's respective scans was conducted to check for factors that may render the subject unsuitable for our model’s purposes, including the presence of moderate to severe osteoarthritis in the knee joints, intraosseous cannulae (typically in the proximal tibia) or orthopedic implants in the lower extremity under consideration, or scan artifacts within regions of interest. Both of the subject’s lower extremities were considered for use. Only one leg was included per subject. However, if neither leg proved appropriate for use, the subject was subsequently excluded from further analysis.

4.4 Scan Screening and Material Property Calibration

Pre-processing started by calibrating the "thin soft-tissue lower extremity" scan. Since phantoms aren't included in scans from the NMDID database, we utilized a phantomless air-fat-muscle calibration method, similar to the one described by Eggermont et al. (Eggermont et al. 2019). The process was automated using a MATLAB (MathWorks, Natick, MA) script. First, we imported DICOM files. Then, nine slices were chosen for calibration. The first slice was the most superior one of the lower extremities that did not contain buttock or genital, followed by eight slices inferior to it. On each of these selected slices, a square region of interest was defined, encompassing the tissue of one leg and some surrounding air (Figure 4-1a). These regions were combined into a volume of interest (Figure 4-1b).

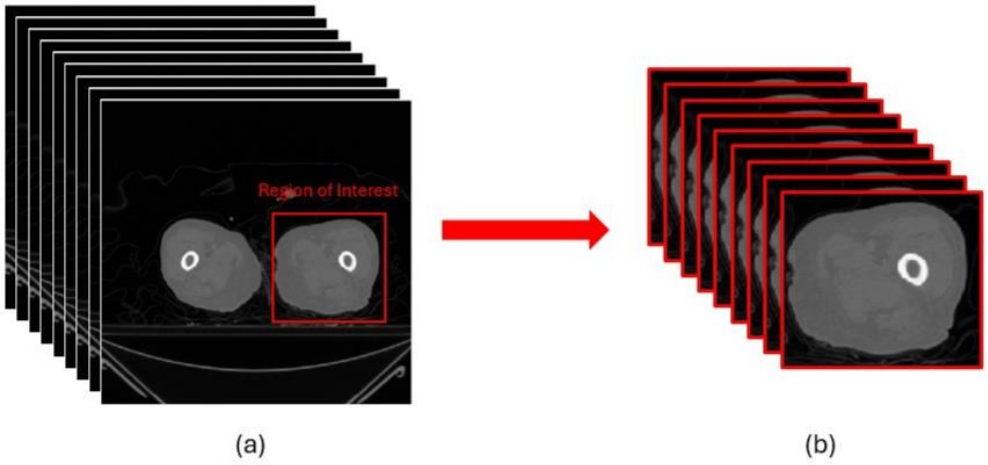

Figure 4-1: A visual depiction of the (a) square region of interest selected on each slice and (b) the resulting volume of interest from which the scan is calibrated.

A combined histogram of all Hounsfield units (HU) in the volume of interest was generated to identify peaks for air, fat, and muscle tissue (Figure 4-2a). These values

were then linearly fitted to the reference "BMD" values for air, fat, and muscle as outlined in Eggermont et al. (840, -80, and 30 respectively) (Eggermont et al. 2019), resulting in a calibration curve from which the relationship between HU and BMD could be derived (Figure 4-2b).

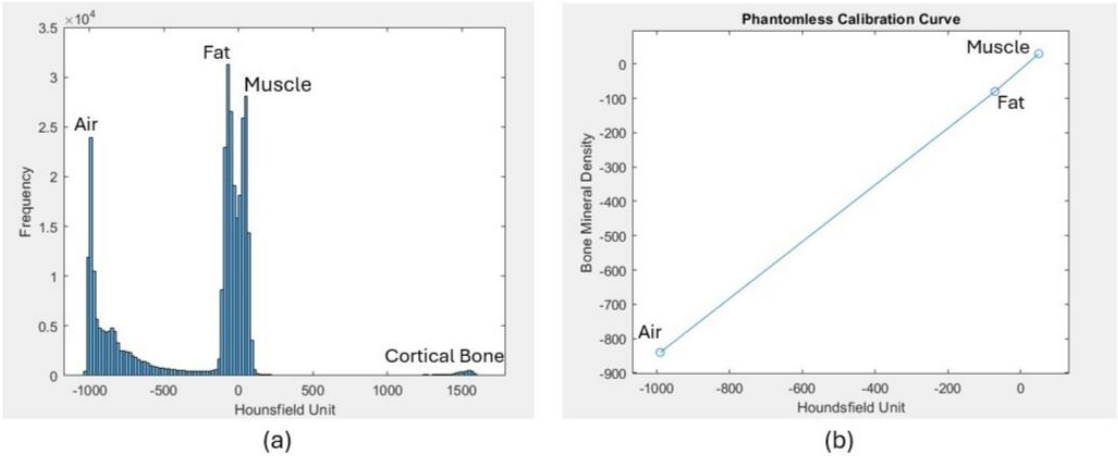

Figure 4-2: (a) An example of a combined histogram of all HU in a volume of interest and (b) the resulting air-muscle-fat calibration curve.

The phantomless calibration method was compared to a traditional phantom-based calibration approach by applying both methods to calibrate a representative scan. Although both methods yielded generally comparable results, small differences were observed in the calibration curves, especially in regions with higher pixel intensity (Figure 4-3). The maximum difference between the BMD values generated for each pixel of the scan using both methods was 96 kg/m^3, occurring in the area of highest bone density (Figure 4-4b). This discrepancy is expected, as the phantomless calibration method exhibited progressively higher BMD values with increasing pixel intensity

compared to the traditional approach (Figure 4-4a). Consequently, it is anticipated that the phantomless calibration method will provide denser readings, particularly in the cortical regions of the bone. The average difference across all pixel densities in the scan was found to be 28 kg/m^3 (Figure 4-4b), which was deemed acceptable. Intersubject differences in the calibration curves were also observed using the air-fat-muscle approach and similar in magnitude to the differences between the phantom and phantomless calibration approaches.

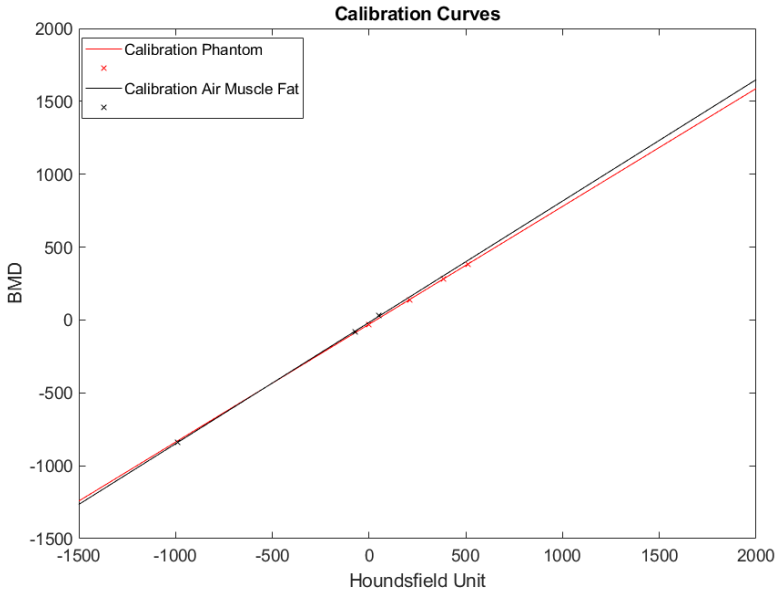

Figure 4-3: Calibration curves generated for a representative scan using both a phantom and phantomless calibration method.

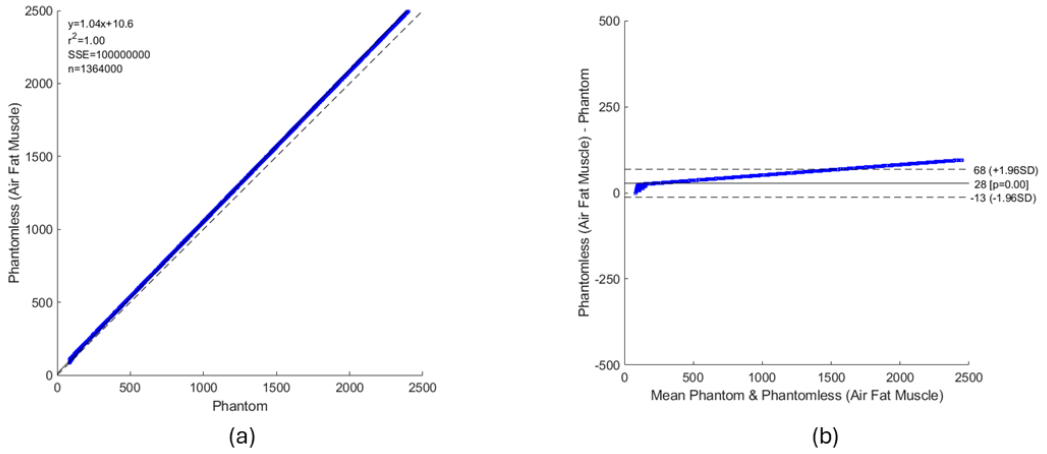

Figure 4-4: (a) Correlation Plot and (b) Bland-Altman Plot comparing the BMD values derived from the phantom and phantomless calibration methods for all pixels in a representative scan.

Calibration was performed as the initial step of pre-processing to avoid wasting time processing subjects whose scans could not be calibrated due to homogeneity in HU of their fat and muscle. While the use of a soft tissue-specific scan helped mitigate this issue, it did not eliminate it. These subjects were excluded from further analysis.

4.5 Segmentation

After evaluating and calibrating the scan, the subject's bony geometry was segmented from the "thin bone torso" (proximal femur) and "thin bone lower extremity" (distal femur, tibia, distal fibula) scans using ScanIP software (Synopsys, Mountain View, CA). CT image thresholds were adjusted interactively to isolate the desired tissue, with any identified gaps manually reconnected and artifacts removed. Subsequently, the cavities were filled, and smoothed masks for each bone were generated. These masks underwent

visual inspection to identify and address any artifacts. They were then exported from ScanIP as STL files.

4.6 Landmark Assignment

The location of nine landmarks necessary for the creation of local coordinate systems for the femur and tibia were identified in HyperMesh (Altair, Troy, MI). These landmarks provided both local descriptions around the knee, such as the femoral epicondyles and the tibial plateaus, and global descriptions of overall alignment, including the hip center and ankle center. The center of the femoral head was approximated as the center of a sphere surface generated from four manually selected points on its surface (Figure 4-5). To identify the medial and lateral epicondyles of the femur, the distal femur was examined in the coronal plane, with these landmarks determined as the most medial and lateral points, respectively, on the bone's surface (Figure 4-6). The most anterior and posterior edges of the tibial medial and lateral plateau were identified through visual inspection of the proximal tibia (Figure 4-7). The ankle center was defined as the midpoint of a line between the most medial point on the medial malleolus of the distal tibia and the most lateral point on the lateral malleolus of the distal fibula when viewed in the coronal plane (Figure 4-8). Additionally, although not utilized in coordinate system generation, the femur anterior cortex point, marking the transition from convex to concave curvature on the anterior surface of the distal femur, was also identified (Figure 4-6). The author notes that due to her limited experience in manual landmark identification, there might be significant variability in the recorded location of this point. For each instance, landmark coordinates were documented in a text file and saved alongside the original HyperMesh

file for future reference. It is important to consider potential error resulting from manual landmark identification if utilizing this data for purposes beyond local coordinate system establishment. Following this, subjects were grouped into “batches” for processing. Each "batch" folder housed STL files, landmark text documents, and DICOMs for the included subjects, alongside an Excel file containing metadata for all subjects in the batch.

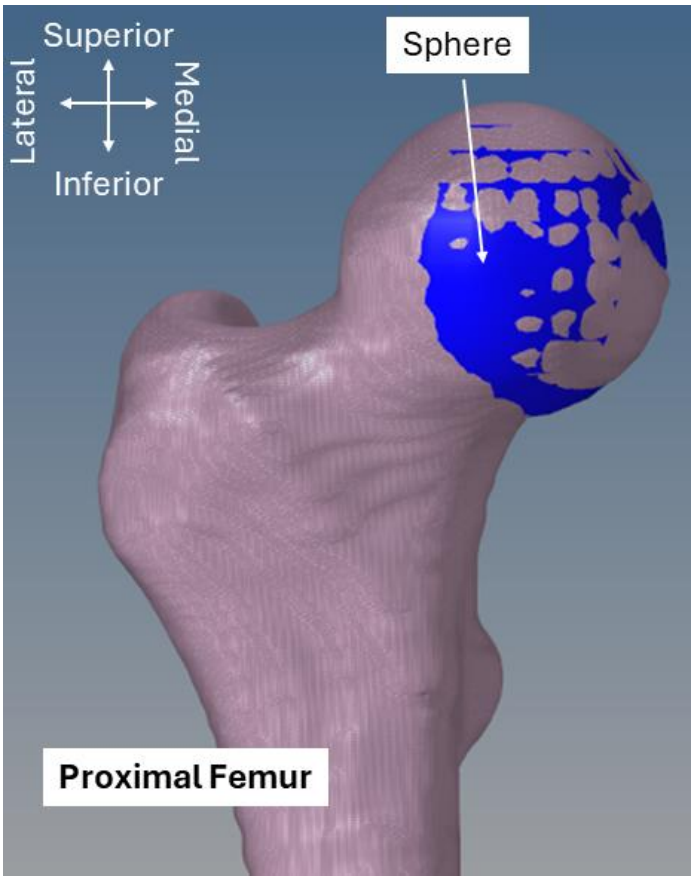

Figure 4-5: A proximal femur viewed in the coronal plane with a sphere whose center approximates the center of the femoral head.

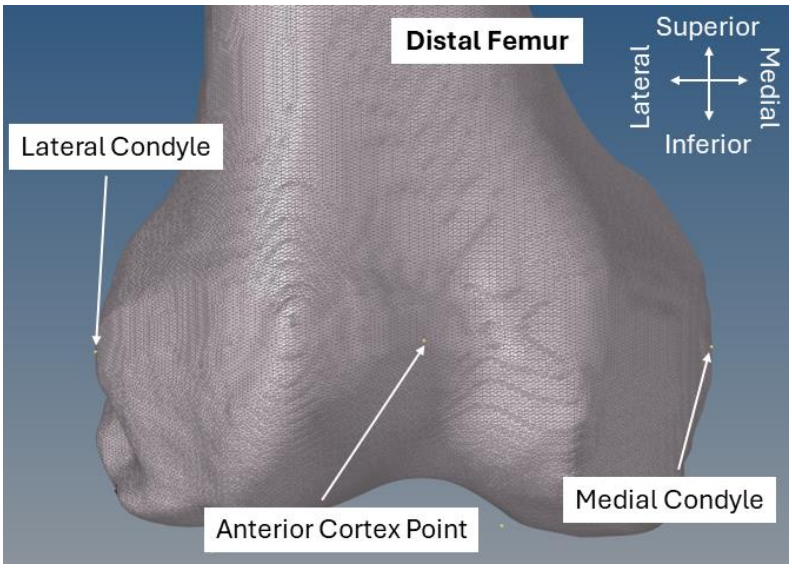

Figure 4-6: A distal femur viewed in the coronal plane with the anterior cortex point, lateral, and medial condyle landmarks identified.

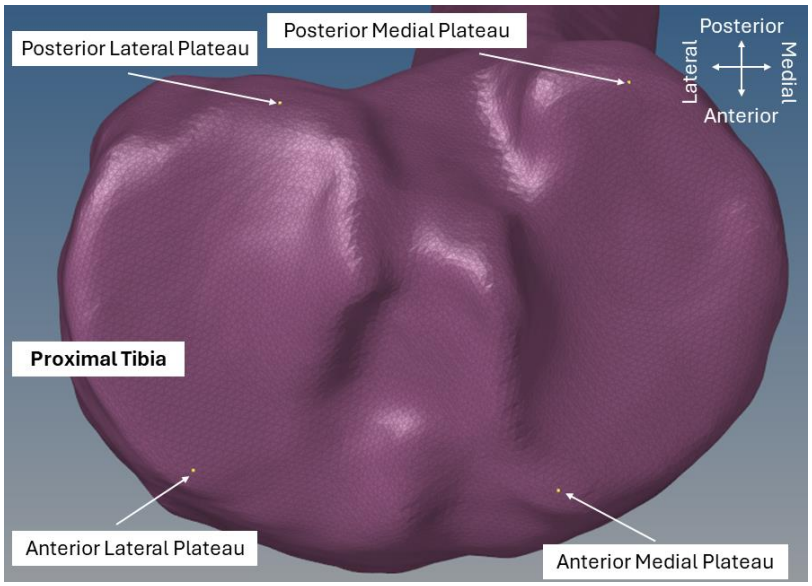

Figure 4-7: A proximal tibia viewed in the transverse plane with the posterior and anterior landmarks identified on the medial and lateral plateaus.

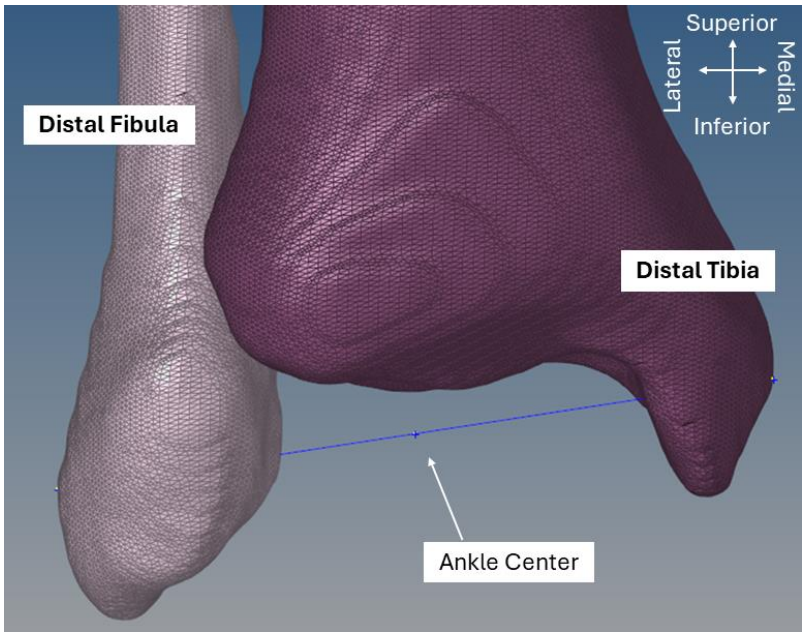

Figure 4-8: A distal fibula and tibia viewed in the coronal plane with the ankle center identified.

4.7 Creation of Local Coordinate Systems

In a process fully automated in a previously developed MATLAB script, local anatomical coordinate systems were created for each of the tibias and femurs in the batch using the previously identified landmarks. Left instances and their landmarks were mirrored across the sagittal plane. The tibial anatomical coordinate system for each instance was established by first defining the medial and lateral plateau centers were calculated as the midpoint between the medial-anterior (1) and medial-posterior (2) and lateral-anterior (3) and lateral-posterior (4) landmarks, respectively. The origin was defined as the midpoint between the two centers and the mediolateral (ML) x-axis as the vector from medial plateau center to lateral plateau center. The superior-inferior (SI) z-axis was defined from the estimated ankle joint center (5) to the origin and pointed superiorly. The anteroposterior (AP) y-axis was obtained by crossing the x- and z-axes.

The femoral anatomical coordinate system was established based on three bony landmarks that were extracted from the template surface geometry: hip joint center (1), medial epicondyle (2), and lateral epicondyle (3). The origin was defined as the midpoint between the epicondyle landmarks, and the mediolateral (ML) x-axis as the vector from medial epicondyle to lateral condyle. The superior-inferior (SI) z-axis was defined as the vector passing through the origin and hip joint center and pointed superiorly. The anteroposterior (AP) y-axis was obtained by crossing the z- and x-axes. The nodes describing the surfaces of the tibia and femur, imported into MATLAB as the previously exported STLs, were transformed from the global CT coordinate system into their respective anatomical coordinate system along with their corresponding landmarks. The transformation matrices defining the transformation between the CT and anatomical coordinate systems were retained for future use. This script also processed the relevant metadata from the excel sheet into MATLAB, storing it within the same organized structure containing the transformed surface data, transformed landmark data, and transformation matrices, hereon known as the Prepared Training Set Structure.

4.8 Template Selection and Meshing

In a previously built SSIM (Bayoglu et al. 2020), median right tibia and femur were selected as the templates and aligned in their local coordinate systems as described above. Tibial template meshes were constructed for surfaces (triangular (tri), 0.95 ± 0.16 mm edge length, 20,296 nodes and 40,588 tri elements) and solids (tetrahedral (tet), 0.96 ± 0.14 mm edge length, 239,903 nodes and 1,364,343 tetrahedral elements). The diaphysis of the template geometry was resected at a ratio of 1.38 between SI length and ML

condylar width. Femoral template meshes were constructed for surfaces (2D, 0.82 ± 0.19 mm edge length, 49,767 nodes and 99,530 tri elements) and solids (3D, 1.03 ± 0.16 mm edge length, 418,569 nodes and 2,337,654 tetrahedral elements). The diaphysis of the template geometry was resected at a ratio of 1.74 between SI length and ML condylar width. The same tibial and femoral template meshes were used in the construction of the current SSIM.

4.9 Shape Registration

Shape registration was also conducted in batches using a MATLAB script to fully automate the process, employing separate scripts for femurs and tibias. The template (data as an Abaqus .inp file) and instance (Prepared Training Set Structure) tri meshes were imported into the MATLAB workspace. Since our model is concerned only with the proximal and distal portions of the tibia and femur, respectively, the template resection ratio was applied to trim the shaft of the instance. This step ensured precise establishment of the 1-to-1 correspondence in the shaft region of the bone. While the entire tibia was segmented in the NMDID subjects, there were instances where femurs lacked "sufficient shaft" to meet the required aspect ratio due to segmentation being done across two scans. These cases were subsequently excluded from further analysis.

After adjusting the instance to the correct SI length, an ICP-based algorithm (Kroon 2024) was used to first rigidly align the instance to the template and then affinely deform the template onto the instance. Following the methods outlined by Andreassen et al. (Andreassen et al. 2024), a Generalized Regression Neural Network (GRNN)-based algorithm completed the morphing of the template onto the instance. Initially, the

algorithm generated reduced point clouds of both the template and the instance by down sampling each mesh. Then, the reduced template point cloud was morphed onto the reduced instance point cloud through an iterative correspondence and deformation process. A GRNN, trained on the nodal displacements between the original reduced template and its final morph, was utilized to deform the original template mesh. This deformed template mesh was further morphed onto the original instance mesh until convergence between the meshes was achieved. The nodal coordinates of the final deformed template mesh were recorded as the registered instance tri mesh. Morphing error was quantified as the root mean square (RMS) error of the distances between the deformed template points and their respective closest point in the instance mesh.

Following that, another GRNN was trained on the nodal displacements between the original template triangular mesh and its final morph. This GRNN was then used to predict the nodal displacements of the template tetrahedral (tet) mesh, yielding a one-to-one correspondence between the triangular mesh and the surface nodes on the tetrahedral mesh. This deformation process was carried out incrementally, with continuous updates and retraining of the GRNN to ensure the preservation of element integrity throughout the process and in the final product. Additionally, to reduce the memory requirements of the algorithm, the tet mesh was divided into four quadrants, which were morphed one at a time before being recombined into the final morphed tetrahedral mesh, whose nodal coordinates were recorded as the registered instance tetrahedral mesh. The morphing algorithm can be visualized in Figure 4-10. Morphing error was quantified as the root mean square (RMS) error of the distances between the surface points of the deformed

template tetrahedral mesh and their respective closest point in the instance tri mesh. Mesh quality was evaluated using MATLAB's *meshQuality* function, which returns a row vector of values ranging from 0 to 1 representing the shape quality of all elements in the mesh, with 1 indicating optimal element shape. The final mesh quality was reported as the number of elements with a shape quality less than or equal to 0.5. The surface register, tetrahedral register, original triangular and tetrahedral template meshes, landmarks, coordinate system transformations, and metadata were saved in an organized structure hereon known as the Shape Register.

Figure 4-9: GRNN-based algorithm used to morph the triangular and tetrahedral meshes.

4.10 Material Property Registration

Material properties were derived from the CT scan of each instance and were assigned to the respective registered tetrahedral meshes using a previously developed MATLAB script. Registered meshes were first transformed back into global CT space using the established transformations. Material properties for each element were reported as apparent bone mineral density. To calculate this, the intensity of the voxel nearest to each node was identified and averaged for the nodes in each element. The average intensity was then converted to BMD using the subject-specific conversion of HU to BMD determined during the scan calibration process and then further converted to apparent density using the established relationships shown in the Equations 4-1 to 4-3 below:

$$BMD = a + b \cdot HU \qquad \text{Equation (4-1)}$$

$$\rho_{ash} = k_{ash} \cdot BMD + l_{ash} \qquad \text{Equation (4-2)}$$

$$\rho_{app} = \frac{\rho_{ash}}{m_{app}} \qquad \text{Equation (4-3)}$$

where *HU* is Hounsfield unit, *BMD* is bone mineral density, *a* and *b* are calibration constants, ρ_{ash} and ρ_{app} denote the ash and apparent densities, respectively. Conversion factors k_{ash}, l_{ash}, and m_{app} were adapted from existing literature (J. H. Keyak, Lee, and Skinner 1994; Schileo et al. 2007; Joyce H Keyak et al. 2005) and are shown in Table 4-1. As with the nodal coordinate information, the element material properties for each instance were combined into a matrix: the Material Property Register. Figure 4-10 visualizes the entire registration process from segmentation to Material Property Register creation.

Table 4-1: Conversion factors used to convert HU to BMD in the Tibia and Femur.

Conversion Factors				
	Tibia		Femur	
	Value	Source	Value	Source
k_{ash}	0.953	Keyak et al. 1994	0.887	Keyak et al. 2005
l_{ash}	45.7	Keyak et al. 1994	63.3	Keyak et al. 2005
m_{app}	0.55	Keyak et al. 1994	0.593	Schileo et al. 2008

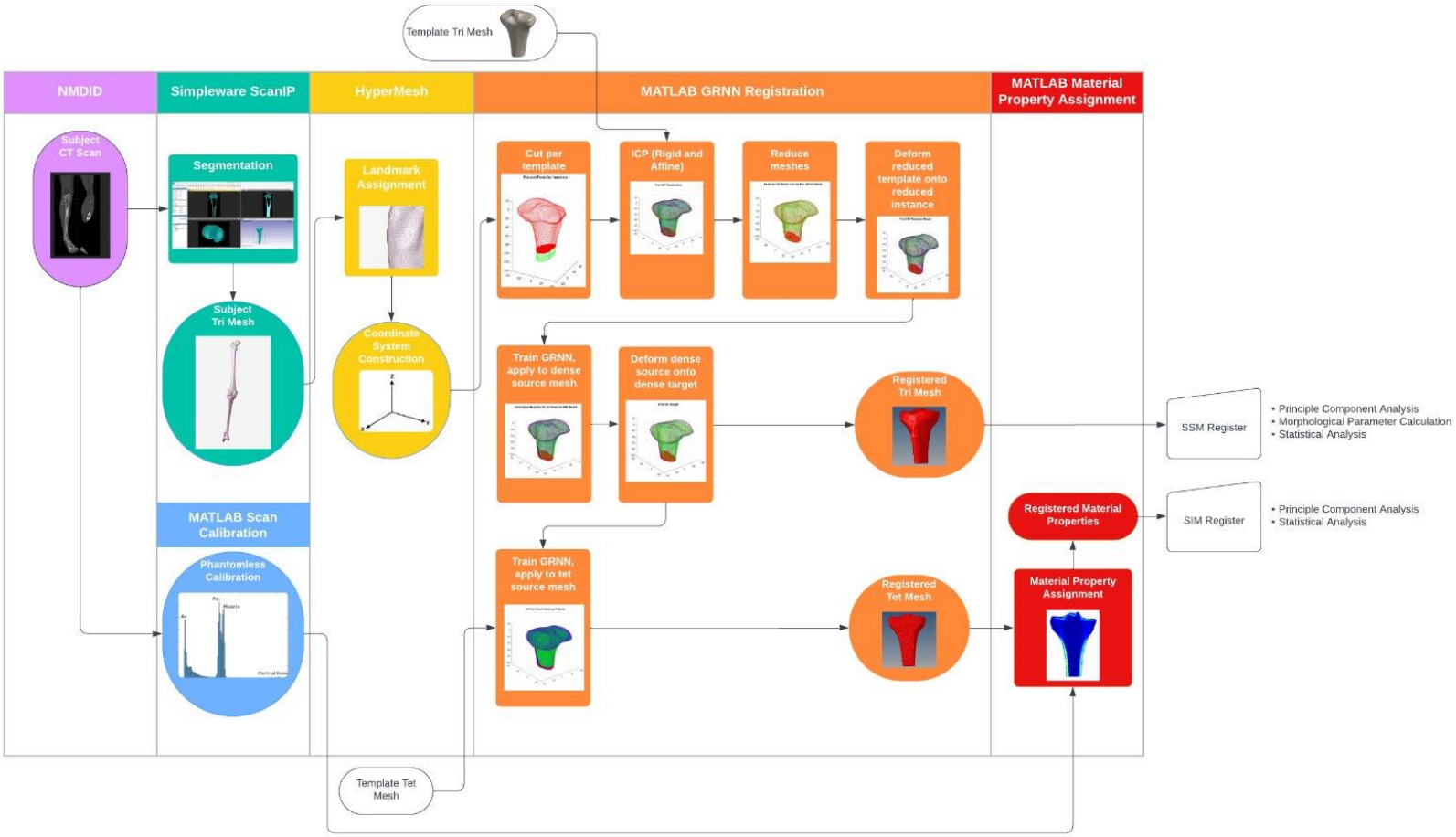

Figure 4-10: Visualization of the registration process from subject segmentation (left) to register creation (right).

4.11 Morphological Parameter Calculation

Morphological parameters were calculated for each tibia and femur instance using fully automated, previously developed MATLAB codes based on work from Mahfouz et al. (Mahfouz et al. 2012) and Ma et al. (Ma et al. 2017). Parameters calculated for the tibia and femur are visualized in Figure 4-11 and Figure 4-12, respectively. These measures are further defined in Table 4-2.

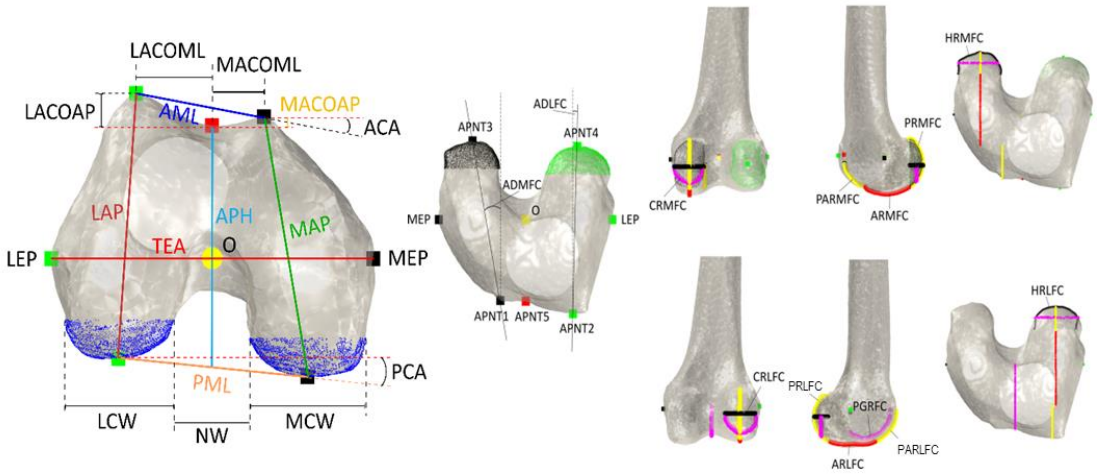

Figure 4-11: Morphological parameters calculated for the femur. "O" indicates the origin. Table 4-2 lists abbreviation and definitions for the illustrated parameters. Taken from Bayoglu et al. (Bayoglu et al. 2020).

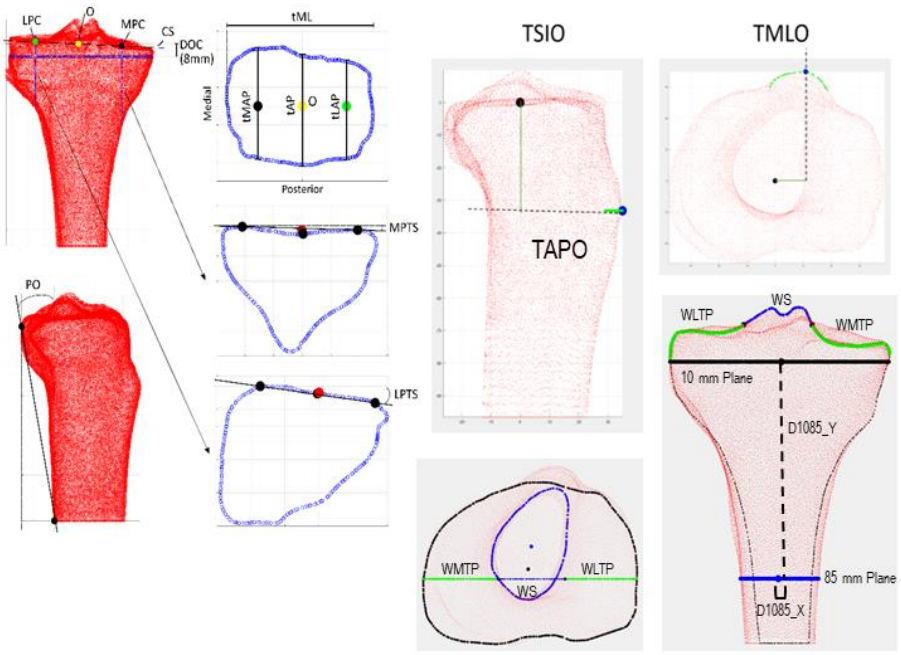

Figure 4-12: Morphological parameters calculated for the tibia. "O" indicates the origin. Table 4-2 lists abbreviation and definitions for the illustrated parameters. Taken from Bayoglu et al. (Bayoglu et al. 2020).

Table 4-2: Morphological parameters calculated for the tibia and femur instances. Taken from Bayoglu et al. (Bayoglu et al. 2020).

Abbreviation	Measurement	Definition
TIBIA PARAMETERS		
tML	Mediolateral length of the proximal tibia plateau	Length of the transverse section taken on the proximal tibia along the ML axis. The transverse section is perpendicular to the SI axis, and the depth of cut is 8 mm below the highest plateau center.
tAP	Anteroposterior length of the proximal tibia plateau	Length of the transverse section taken on the proximal tibia along the AP axis.
tMAP	Anteroposterior length of the medial proximal tibia plateau	Length of the line drawn on the transverse section passing through the estimated center of the medial tibia plateau and directing along the AP axis.
tLAP	Anteroposterior length of the lateral proximal tibia plateau	Length of the line drawn on the transverse section passing through the estimated center of the lateral tibia plateau and directing along the AP axis.
MPTS	Medial tibial posterior slope	Angle between the tangent line to the medial plateau and the AP axis projected on the sagittal plane.
LPTS	Lateral tibial posterior slope	Angle between the tangent line to the lateral plateau and the AP axis projected on the sagittal plane.

CS	Coronal slope	Angle between the line passing through medial and lateral plateau centers and the ML axis projected on the coronal plane.
PO	Posterior overhang	Angle between the most posterior points of the proximal and distal tibia projected on the sagittal plane.
AS	Asymmetry	Ratio of the anteroposterior lengths of the medial and lateral proximal tibia plateaus.
TMLO	Tubercle mediolateral offset	Mediolateral distance between the most anterior point of the proximal tibia (tubercle) and the origin.
TAPO	Tubercle anteroposterior offset	Anteroposterior distance between the most anterior point of the proximal tibia (tubercle) and the origin.
TSIO	Tubercle superior-inferior offset	Superior-inferior distance between the most anterior point of the proximal tibia (tubercle) and the origin.
WS	Tibial Spine Width	Distance between 1mm medially from Lateral Center Point and 1mm laterally from Medial Center point in ML direction, along Tibial Spine.
WMTP	Medial Plateau Width	Distance from the most medial point of the tibial spine to the most medial point of the tibia.
WLTP	Lateral Plateau Width	Distance from the most lateral point of the tibial spine to the most lateral point of the tibia.
FEMUR PARAMETERS		
TEA	Transepicondylar axis length	Distance between the medial and lateral epicondyles
APH	Anteroposterior height	Distance between the most anterior aspect of the cortex and the midpoint between the most posterior points on the medial and lateral condyles
MAP	Medial anteroposterior height	Distance between the most anterior and posterior aspects of the medial condyle
LAP	Lateral anteroposterior height	Distance between the most anterior and posterior aspects of the lateral condyle
AML	Anterior mediolateral length	Distance between the two most anterior aspects of the medial and lateral condyles
PML	Posterior mediolateral length	Distance between the two most posterior aspects of the medial and lateral condyles
MCW	Medial condylar mediolateral width	Length of the medial condyle in the mediolateral direction
LCW	Lateral condylar mediolateral width	Length of the lateral condyle in the mediolateral direction
MACOAP	Medial flange anteroposterior offset	Anteroposterior offset between the most anterior aspect of the medial condyle and the most anterior aspect of the cortex
LACOAP	Lateral flange anteroposterior offset	Anteroposterior offset between the most anterior aspect of the lateral condyle and the most anterior aspect of the cortex
MACOML	Medial flange mediolateral offset	Mediolateral offset between the most anterior aspect of the medial condyle and the most anterior aspect of the cortex
LACOML	Lateral flange mediolateral offset	Mediolateral offset between the most anterior aspect of the lateral condyle and the most anterior aspect of the cortex
PCA	Posterior condylar angle	Angle between a line passing through the most posterior aspects of the medial and lateral condyles and the ML axis projected on the transverse plane
ACA	Anterior condylar angle	Angle between a line passing through the most anterior aspects of the medial and lateral condyles and the ML axis projected on the transverse plane
NW	Notch width	Intercondylar notch width

PRMFC	Posterior radius of medial femoral condyle	Radius along the posterior portion of the medial femoral condyle
ARMFC	Anterior radius of medial femoral condyle	Radius along the anterior portion of the medial femoral condyle
PARMFC	Patella radius of medial femoral condyle	Radius along the patellar portion of the medial femoral condyle
HRMFC	Horizontal radius of medial femoral condyle	Radius along the horizontal portion of the medial femoral condyle
CRMFC	Coronal radius of medial femoral condyle	Radius along the coronal portion of the medial femoral condyle
PRLFC	Posterior radius of lateral femoral condyle	Radius along the posterior portion of the lateral femoral condyle
ARLFC	Anterior radius of lateral femoral condyle	Radius along the anterior portion of the lateral femoral condyle
PARLFC	Patella radius of lateral femoral condyle	Radius along the patellar portion of the lateral femoral condyle
HRLFC	Horizontal radius of lateral femoral condyle	Radius along the horizontal portion of the lateral femoral condyle
CRLFC	Coronal radius of lateral femoral condyle	Radius along the coronal portion of the lateral femoral condyle
PGRFC	Patella groove radius of femoral condyles	Radius along the patellar groove of the femoral condyles
ADMFC	Angle of divergence medial femoral condyle	Angle found for divergence of the medial femoral condyle
ADLFC	Angle of divergence of lateral femoral condyle	Angle found for divergence of the lateral femoral condyle

4.12 SSIM Creation

The Shape Registers, Material Property Registers, and morphological parameters from all batches were consolidated into two SSIM superstructures: one for the Tibia and another for the Femur. These SSIM superstructures include triangular and tetrahedral template data, a triangular shape register (in both matrix and column formats), a tetrahedral shape register (also in both matrix and column formats), and a material property register, as well as mesh quality parameters, morphological parameters, landmark coordinates, transformation matrices, and metadata for all instances. From here, principal component analysis was performed individually on both the tri surface registers and material property registers, yielding Statistical Shape Models and Statistical Intensity Models of the Proximal Tibia and Distal Femur, respectively.

4.13 Result Generation and Statistical Analysis

For each model, metrics including accuracy, compactness, generalizability, and specificity were computed across various levels of principal component (PC) inclusion to evaluate model quality (Audenaert et al. 2019). Visualization of the first four modes of shape and intensity variation in each model was achieved by adjusting the average geometry or material properties within +/- 3.0 standard deviations. Additionally, correlation coefficients were determined to explore the relationships between modes of variation and morphological parameters, among different morphological parameters themselves, and between the first ten shape and intensity modes.

To quantify interpopulation variances in PC scores and morphological parameters, unpaired two-sample t-tests were conducted across different sex and race groups, including Male/Female (M/F), Asian/Black (A/B), Asian/Hispanic (A/H), Asian/Native American (A/NA), Asian/White (A/W), Black/Hispanic (B/H), Black/Native American (B/NA), Black/White (B/W), Hispanic/Native American (H/NA), Hispanic/White (H/W), and Native American/White (NA/W). Subjects were further categorized into age groups: "young" (ages 25-39), "middle" (ages 40-59), and "old" (ages 60-80), with subsequent two-sample t-tests performed to compare young age/middle age (YA/MA), young age/old age (YA/OA), and middle age/old age (MA/OA). Box plots were employed to visually represent distributions of the first four PC scores and select morphological parameters.

Further, an Analysis of Variance (ANOVA), a statistical method for identifying significant differences among group means, was also conducted to investigate inter-group differences. The ANOVA method compares the variance within each group to the

variance between the groups. If the variance between the groups is significantly greater than the variance within the groups, it suggests that there are meaningful differences between the group means (Kim 2014). This method allows for the investigation of inter-group differences while accounting for imbalances in population sizes. The independent variables considered in our ANOVA included sex, race, age, height, and BMI, with the latter three being continuous variables.

5. Shape Morphology Variation Considering Diverse Populations

5.1 Introduction

While patient sex does not significantly impact the outcomes of Total Knee Arthroplasty (TKA) (Ritter et al. 2008), differences in outcomes following TKA have been observed across different racial and ethnic groups, with minority groups often experiencing disproportionately poorer results (Shahid and Singh 2016). The higher rates of complications and less significant improvements in patient-reported outcomes among minority groups likely stem from a multitude of complex factors, including patient factors, provider and healthcare system factors, and societal factors (Hu et al. 2022).

One potential avenue for improving the generalizability of orthopaedic implants is through pre-clinical evaluations that consider patient variability. Finite element analysis (FEA) is a common tool utilized in the pre-clinical design phase of orthopedic implants to gain insights into bone-implant system behavior. One method to incorporate patient variability in FEA models is by employing statistical shape models (SSMs), which are statistical representations of the morphology of the bone segment or joint. These models help capture variability within a training set of bones or joints, enabling the generation of numerous synthetic instances to support population-based FE models (Taylor and Prendergast 2015). Importantly, the demographic makeup of this training set influences anatomic variability in the synthetic subject pool, emphasizing the necessity of

considering the diversity of the training set for the probabilistic analyses to yield results representative of the entire population.

While SSMs have been used with success to compare the morphology of the tibia and femur between the sexes (Audenaert et al. 2019; Wise et al. 2016; Bah et al. 2015), the difficulty of obtaining population-specific CT scans has limited the construction of SSMs based on ethnically and racially diverse training sets. Mahfouz et al. stands out as one of few models considering a large diverse population of knees. Their study compared 1000 adult knees from African American (n = 80), East Asian (n = 80), and White patients of European descent (n = 840) using an SSM and identified shape differences not only between racial and ethnic groups but also between genders within each race. African American females displayed distinct characteristics, such as a deeper patellar groove compared to their Caucasian counterparts, while African American males exhibited larger femoral and tibial dimensions compared to East Asian males. Moreover, they found that males tended to have larger knees across all races and ethnicities (Mahfouz et al. 2015).

With the goal of continuing to improve the representation of diverse populations, the present study focuses on the development of a statistical shape and intensity model of the proximal tibia and distal femur based on a diverse library of CT scans, encompassing individuals of both sexes from Black or African American, Native American, Hispanic, Asian, and White backgrounds. In addition to presenting the model outcomes, key anatomic parameters are measured and compared between both sex and racial and ethnic groups.

5.2 Methods

5.2.1 Subject Selection

The New Mexico Decedent Image Database (NMDID) is a comprehensive dataset containing whole-body CT scans and associated metadata for over 15,000 individuals who died in New Mexico between 2010 and 2017 (Edgar et al. 2020). This dataset offers an invaluable resource for accessing a diverse population of scans. For this work, an effort was made to gather a sample representative of the United States population (Barton et al., n.d.), including equal numbers of males and females of various ages from Black or African American, Native American, Asian, and Hispanic backgrounds. A sample of White subjects were obtained from the training set of an SSIM previously developed at the University of Denver (Bayoglu et al. 2020). The NMDID race and ethnicity metrics were self-reported by the decedents in the 2010 census. It is important to note that for the purposes of this work, "Hispanic" refers to decedents who identified as racially Hispanic, reflecting the usage in the 2010 census where many New Mexicans selected Hispanic as an "Other" race option (Edgar et al. 2020). The inclusion criteria for this study required subjects to be aged between 25 and 80 years, with a BMI of less than 40, based on cadaver height and weight. Subjects were excluded if they had a recorded history of cancers, tumors, or malignancies, any recorded injuries to the lower limbs, a manner of death that could result in lower limb injuries, cadaver decomposition, or severe osteoarthritis.

5.2.2 Bone Model Generation

Each subject file downloaded from the NMDID database contained 26 CT scans. Among these, three scans held importance for this project: the “thin bone torso” and “thin bone lower extremity” scans. Before processing, a preliminary assessment of each subject's respective scans was conducted to check for factors that may render the subject unsuitable for our model’s purposes, including the presence of moderate to severe osteoarthritis in the knee joints, intraosseous cannulae (typically in the proximal tibia) or orthopedic implants in the lower extremity under consideration, or scan artifacts within regions of interest, or movement artifacts between the torso and lower extremity scans. Both of the subject’s lower extremities were considered for use. Only one leg was included per subject. However, if neither leg proved appropriate for use, the subject was subsequently excluded from further analysis.

Subject's bony geometry was then segmented from the "thin bone torso" (proximal femur) and "thin bone lower extremity" (distal femur, tibia, distal fibula) scans using ScanIP software (Synopsys, Mountain View, CA) and the location of nine key landmarks on the femur (hip center, femur medial and lateral epicondyles) and tibia (anterior medial plateau, posterior medial plateau, anterior lateral plateau, posterior lateral plateau, ankle center) were identified in HyperMesh (Altair, Troy, MI). Left instances were mirrored across the sagittal plane. Local anatomical coordinate systems were created using the landmarks for each femur and tibia in an automated MATLAB script. The nodes describing the surfaces of the tibia and femur were transformed from the global CT coordinate system into their respective anatomical coordinate system along with their

corresponding landmarks. The transformation matrices defining the transformation between the CT and anatomical coordinate systems were retained for future use.

5.2.3 Template Selection and Meshing

Using a workflow similar to Fugit et al. (in review), median right tibia and femur were selected as the templates and aligned in their local coordinate systems as described above. A template mesh was constructed for tibial surfaces (triangular (tri), 0.95 ± 0.16 mm edge length, 20,296 nodes and 40,588 tri elements) and femoral surfaces (triangular, 0.82 ± 0.19 mm edge length, 49,767 nodes and 99,530 tri elements). The diaphyses of the tibial and femoral template geometries were resected at a ratio of 1.38 and 1.74, respectively, between SI length and ML condylar width.

5.2.4 Mesh Registration and Morphing

Registration of the instances was fully automated in a MATLAB script, employing separate scripts for femurs and tibias. The shafts of the instances were first resected per the template aspect ratio. While the entire tibia was segmented in the NMDID subjects, there were instances where femurs lacked "sufficient shaft" to meet the required aspect ratio due to segmentation being done across two scans. These cases were subsequently excluded from further analysis.

After adjusting the instance to the correct Superior-Inferior length, an Iterative Closest Point-based algorithm (Kroon 2024) was used to first rigidly align the instance to the template and then affinely deform the template onto the instance. Following the methods outlined by Andreassen et al., a Generalized Regression Neural Network (GRNN)-based algorithm completed the morphing of the template onto the instance

(Andreassen et al. 2024), with nodal coordinates of the final deformed template mesh recorded as the registered instance tri mesh. Each instance served as a row in a matrix R of the form: $R = [X_1, X_2, \ldots, X_n]$, where X_i is a single vector of length $3n$ that defines the i th instance and contains the nodal coordinates (x, y, z) of the n nodes describing the registered instance. Morphing error was quantified as the root mean square (RMS) error of the distances between the deformed template points and their respective closest point in the instance mesh.

5.2.5 Model Generation

Principal component analysis was performed on the registers, yielding SSMs of the proximal tibia and distal femur, respectively. Metrics of accuracy, compactness, generalizability, and specificity were computed across various levels of principal component (PC) inclusion to evaluate SSM model quality (Audenaert et al. 2019).

5.2.6 Anatomic Parameter Measurements

Morphological parameters. Similar to those reported by Mahfouz et al. (Mahfouz et al. 2012) and Ma et al. (Ma et al. 2017), were calculated for each tibia and femur instance using fully automated MATLAB scripts. Parameters calculated for the tibia and femur are visualized in Figure 5-1 and Figure 5-2, respectively. These measures are further defined in Table 5-1.

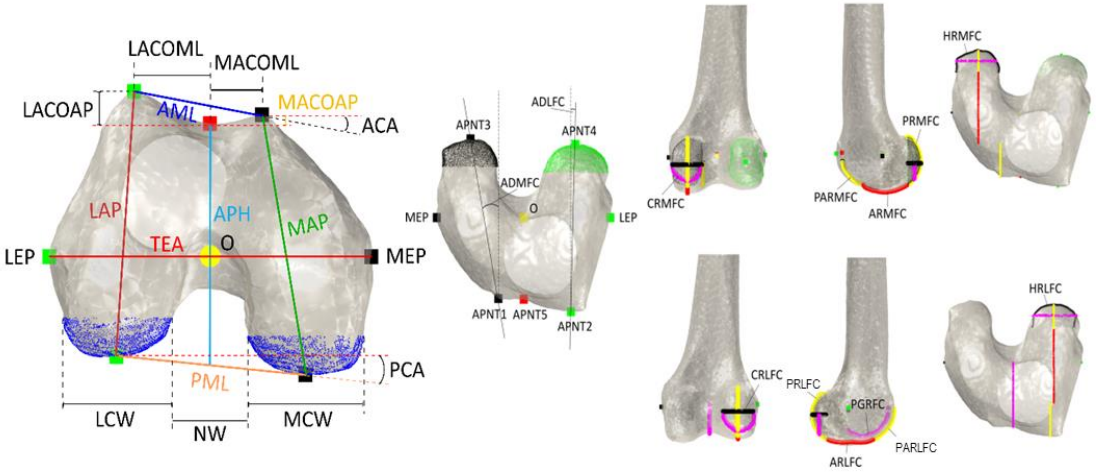

Figure 5-1: Morphological parameters calculated for the femur. "O" indicates the origin. Table 5-1 lists abbreviation and definitions for the illustrated parameters. Taken from Bayoglu et al. (Bayoglu et al. 2020).

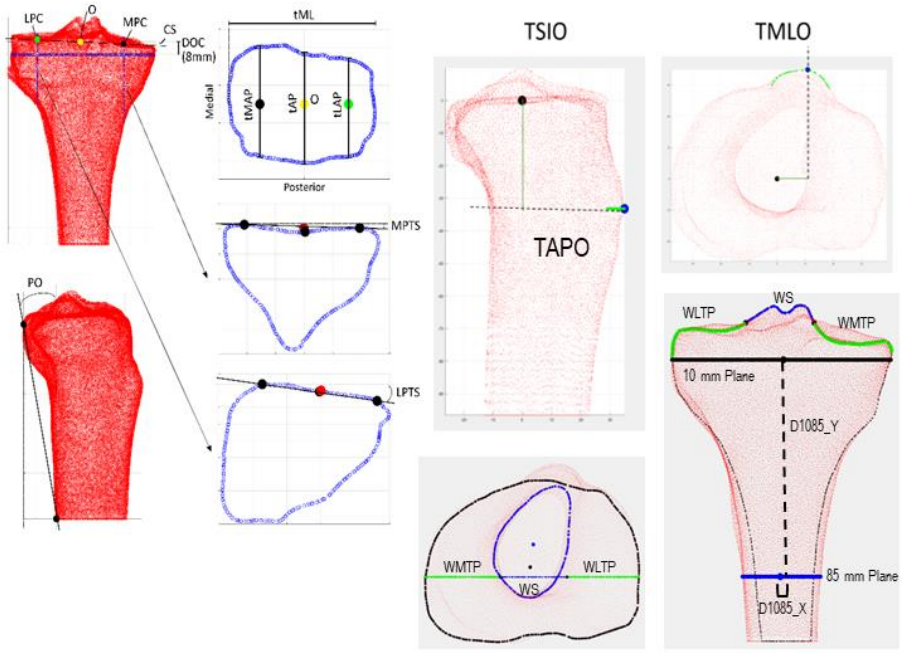

Figure 5-2: Morphological parameters calculated for the tibia. "O" indicates the origin. Table 5-1 lists abbreviation and definitions for the illustrated parameters. Taken from Bayoglu et al. (Bayoglu et al. 2020).

Table 5-1: Morphological parameters calculated for the tibia and femur instances. Taken from Bayoglu et al. (Bayoglu et al. 2020).

Abbreviation	Measurement	Definition
TIBIA PARAMETERS		
tML	Mediolateral length of the proximal tibia plateau	Length of the transverse section taken on the proximal tibia along the ML axis. The transverse section is perpendicular to the SI axis, and the depth of cut is 8 mm below the highest plateau center.
tAP	Anteroposterior length of the proximal tibia plateau	Length of the transverse section taken on the proximal tibia along the AP axis.
tMAP	Anteroposterior length of the medial proximal tibia plateau	Length of the line drawn on the transverse section passing through the estimated center of the medial tibia plateau and directing along the AP axis.
tLAP	Anteroposterior length of the lateral proximal tibia plateau	Length of the line drawn on the transverse section passing through the estimated center of the lateral tibia plateau and directing along the AP axis.
MPTS	Medial tibial posterior slope	Angle between the tangent line to the medial plateau and the AP axis projected on the sagittal plane.

LPTS	Lateral tibial posterior slope	Angle between the tangent line to the lateral plateau and the AP axis projected on the sagittal plane.
CS	Coronal slope	Angle between the line passing through medial and lateral plateau centers and the ML axis projected on the coronal plane.
PO	Posterior overhang	Angle between the most posterior points of the proximal and distal tibia projected on the sagittal plane.
AS	Asymmetry	Ratio of the anteroposterior lengths of the medial and lateral proximal tibia plateaus.
TMLO	Tubercle mediolateral offset	Mediolateral distance between the most anterior point of the proximal tibia (tubercle) and the origin.
TAPO	Tubercle anteroposterior offset	Anteroposterior distance between the most anterior point of the proximal tibia (tubercle) and the origin.
TSIO	Tubercle superior-inferior offset	Superior-inferior distance between the most anterior point of the proximal tibia (tubercle) and the origin.
WS	Tibial Spine Width	Distance between 1mm medially from Lateral Center Point and 1mm laterally from Medial Center point in ML direction, along Tibial Spine.
WMTP	Medial Plateau Width	Distance from the most medial point of the tibial spine to the most medial point of the tibia.
WLTP	Lateral Plateau Width	Distance from the most lateral point of the tibial spine to the most lateral point of the tibia.
FEMUR PARAMETERS		
TEA	Transepicondylar axis length	Distance between the medial and lateral epicondyles
APH	Anteroposterior height	Distance between the most anterior aspect of the cortex and the midpoint between the most posterior points on the medial and lateral condyles
MAP	Medial anteroposterior height	Distance between the most anterior and posterior aspects of the medial condyle
LAP	Lateral anteroposterior height	Distance between the most anterior and posterior aspects of the lateral condyle
AML	Anterior mediolateral length	Distance between the two most anterior aspects of the medial and lateral condyles
PML	Posterior mediolateral length	Distance between the two most posterior aspects of the medial and lateral condyles
MCW	Medial condylar mediolateral width	Length of the medial condyle in the mediolateral direction
LCW	Lateral condylar mediolateral width	Length of the lateral condyle in the mediolateral direction
MACOAP	Medial flange anteroposterior offset	Anteroposterior offset between the most anterior aspect of the medial condyle and the most anterior aspect of the cortex
LACOAP	Lateral flange anteroposterior offset	Anteroposterior offset between the most anterior aspect of the lateral condyle and the most anterior aspect of the cortex
MACOML	Medial flange mediolateral offset	Mediolateral offset between the most anterior aspect of the medial condyle and the most anterior aspect of the cortex
LACOML	Lateral flange mediolateral offset	Mediolateral offset between the most anterior aspect of the lateral condyle and the most anterior aspect of the cortex
PCA	Posterior condylar angle	Angle between a line passing through the most posterior aspects of the medial and lateral condyles and the ML axis projected on the transverse plane

ACA	Anterior condylar angle	Angle between a line passing through the most anterior aspects of the medial and lateral condyles and the ML axis projected on the transverse plane
NW	Notch width	Intercondylar notch width
PRMFC	Posterior radius of medial femoral condyle	Radius along the posterior portion of the medial femoral condyle
ARMFC	Anterior radius of medial femoral condyle	Radius along the anterior portion of the medial femoral condyle
PARMFC	Patella radius of medial femoral condyle	Radius along the patellar portion of the medial femoral condyle
HRMFC	Horizontal radius of medial femoral condyle	Radius along the horizontal portion of the medial femoral condyle
CRMFC	Coronal radius of medial femoral condyle	Radius along the coronal portion of the medial femoral condyle
PRLFC	Posterior radius of lateral femoral condyle	Radius along the posterior portion of the lateral femoral condyle
ARLFC	Anterior radius of lateral femoral condyle	Radius along the anterior portion of the lateral femoral condyle
PARLFC	Patella radius of lateral femoral condyle	Radius along the patellar portion of the lateral femoral condyle
HRLFC	Horizontal radius of lateral femoral condyle	Radius along the horizontal portion of the lateral femoral condyle
CRLFC	Coronal radius of lateral femoral condyle	Radius along the coronal portion of the lateral femoral condyle
PGRFC	Patella groove radius of femoral condyles	Radius along the patellar groove of the femoral condyles
ADMFC	Angle of divergence medial femoral condyle	Angle found for divergence of the medial femoral condyle
ADLFC	Angle of divergence of lateral femoral condyle	Angle found for divergence of the lateral femoral condyle

5.2.7 Result Generation and Statistical Analysis

Visualization of the first four modes of shape and intensity variation in each model was achieved by adjusting the average geometry or material properties within +/- 3.0 standard deviations. Correlation coefficients were determined to explore the relationships between modes of variation and morphological parameters and the among different morphological parameters themselves.

To quantify interpopulation variances in PC scores and morphological parameters, unpaired two-sample t-tests were conducted across different sex and race groups,

including Male/Female (M/F), Asian/Black (A/B), Asian/Hispanic (A/H), Asian/Native American (A/NA), Asian/White (A/W), Black/Hispanic (B/H), Black/Native American (B/NA), Black/White (B/W), Hispanic/Native American (H/NA), Hispanic/White (H/W), and Native American/White (NA/W). Subjects were further categorized into age groups: "young" (ages 25-39), "middle" (ages 40-59), and "old" (ages 60-79), with subsequent two-sample t-tests performed to compare young age/middle age (YA/MA), young age/old age (YA/OA), and middle age/old age (MA/OA). Box plots were employed to visually represent distributions of the first four PC scores and select morphological parameters.

Further, a series of Analysis of Variance (ANOVA) tests were conducted to investigate inter-group differences while accounting for imbalances in population sizes. The independent variables included sex, race, age, height, and BMI, with the latter three being continuous variables. Although the results of this statistical test are preliminary, this analysis is important when considering multiple independent variables that could potentially impact bone morphology.

5.3 Results

5.3.1 Femur

5.3.1.1 Subject Demographics

Initially, an equal number of scans were identified for each demographic category, ensuring balanced representation by sex, race, and a range of ages. However, the training set was unevenly reduced due to issues with scan quality, such as limited

field of view or poor tissue differentiation for material property assignment, and the presence of severe osteoarthritis or osteophytes.

The resulting training data for the femur represented 111 cases, 55 female and 56 male, with an average age of 52.8 (+/- 17.3) and 54.7 (+/- 19.1) years, respectively. Subjects obtained from the NMDID database (n = 70) came from Asian (n = 13), Black or African American (n = 16), Hispanic (n = 18), and Native American (n = 10) backgrounds. The remaining subjects (n = 41) included White (n = 38), Hispanic (n = 1), and Pacific Islander (n = 1) subjects were derived from a cadaveric training set previously developed at the University of Denver. The distribution of the overall subject demographics is shown in Table 5-2. Subject age, height, and BMI data by demographic are shown in Figure 5-3, Figure 5-4, and Figure 5-5, respectively.

Table 5-2: Subject demographics for the Femur SSM training set.

	Asian	Black or African American	Hispanic	Native American	Pacific Islander	White	Total
Female	7	12	8	10	0	18	55
Age (Years)	45.7 +/- 19.3	44.3 +/- 12.9	45.1 +/- 17.6	46.5 +/- 10.4		68.2 +/- 12.5	52.8 +/- 17.3
Height (cm)	161.6 +/- 7.3	164.4 +/- 7.9	160.7 +/- 7.7	166.2 +/- 9.6		161.7 +/- 5.6	163.0 +/- 7. 4
BMI (kg/m^2)	25.5 +/- 7.6	27.8 +/- 5.7	27.0 +/- 5.6	27.8 +/- 5.5		23.7 +/- 5.9	26.0 +/- 6.0
Male	6	4	11	14	1	20	56
Age (Years)	47 +/- 10.0	50.5 +/- 16.8	41.4 +/- 13.7	39.3 +/- 11.7	77 +/- 0	74.9 +/- 6.3	54.7 +/- 19.1
Height (cm)	171.3 +/- 8.7	174.6 +/- 4.4	172.5 +/- 4.1	175.7 +/- 9.4	165.1 +/- 0	174.4 +/- 8.1	173.9 +/- 7.6
BMI (kg/m2)	25.9 +/- 3.3	27.7 +/- 4.7	28.0 +/- 5.6	29.3 +/- 5.7	29.3 +/- 0	20.5 +/- 4.6	25.4 +/- 6.1
Total	13	16	19	24	1	38	111
Age (Years)	46.3 +/- 15.1	45.8 +/- 13.7	42.9 +/- 15.1	42.3 +/- 11.5	77 +/- 0	71.7 +/- 10.2	53.8 +/- 18.2
Height (cm)	166.1 +/- 9.2	167.0 +/- 8.3	167.5 +/- 8.3	171.8 +/- 10.4	165.1 +/- 0	168.4 +/- 9.4	168.5 +/- 9.3
BMI (kg/m2)	25.7 +/- 5.8	27.8 +/- 5.3	27.5 +/- 5.5	28.6 +/- 5.5	29.3 +/- 0	22.0 +/- 5.4	25.7 +/- 6.1

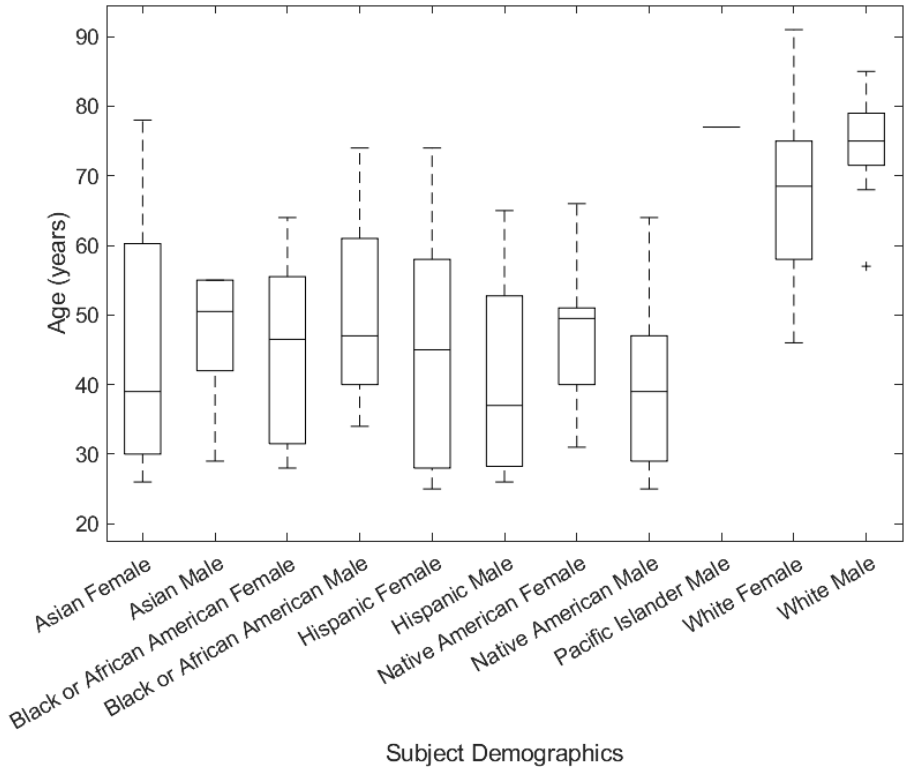

Figure 5-3: Distribution of subject age by demographic.

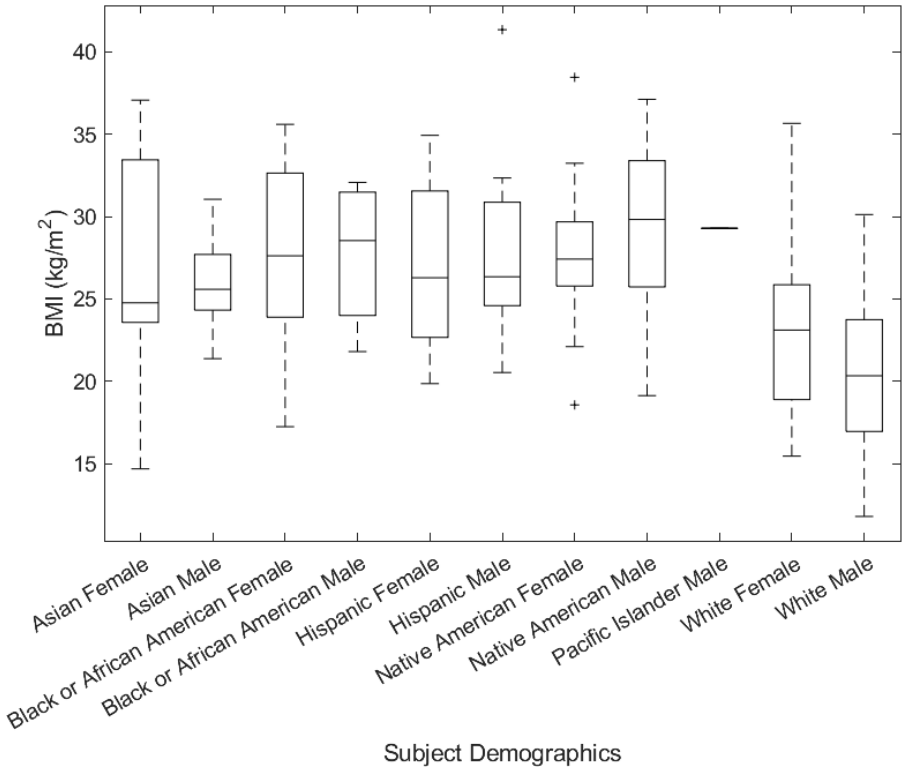

Figure 5-4: Distribution of subject BMI by demographic.

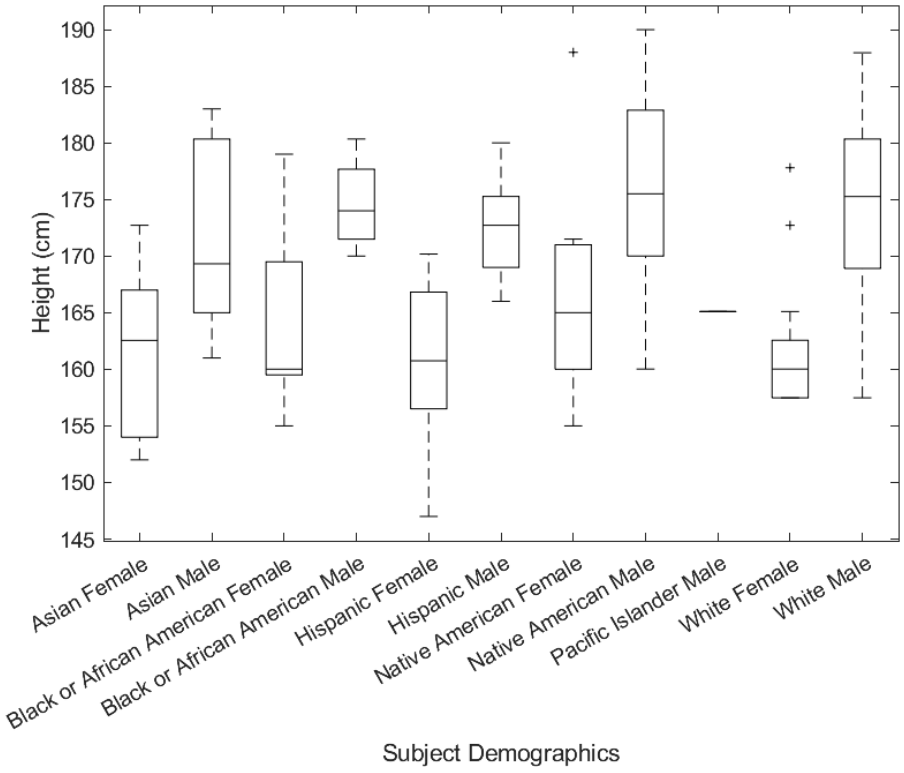

Figure 5-5: Distribution of subject height by demographic.

5.3.1.2 Model Quality

Measures of accuracy, compactness, generalizability, and specificity were computed to assess model quality. The accuracy of the Femur SSM improved with the inclusion of a greater number of principal components (Figure 5-6). Achieving a clinically relevant submillimeter distance error of 0.6 mm required the incorporation of five principal components. Accuracy achieved by inclusion of the first ten principal components is outlined in Table 5-3. Figure 5-7 depicts the cumulative compactness of the Femur SSM, indicating that seven principal components were necessary to encompass at least 95% of the model variance. The distribution of variance across the first ten principal components

is reported in Table 5-3. The generalizability of the model with increasing numbers of training set samples is shown in Figure 5-8. With 111 instances, the generalizability and specificity of the model were found to be 0.33 +/- 0.06 mm and 0.94+/-0.15 mm, respectively (Table 5-4).

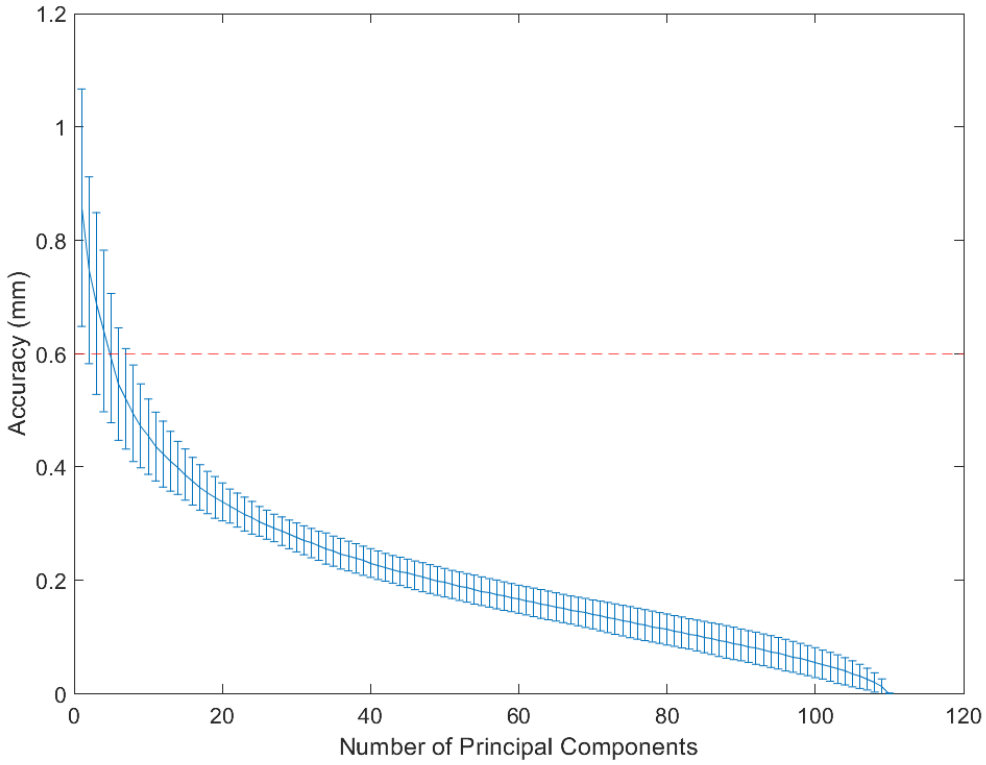

Figure 5-6: Accuracy of the Femur SSM with the progressive addition of principal components to the model.

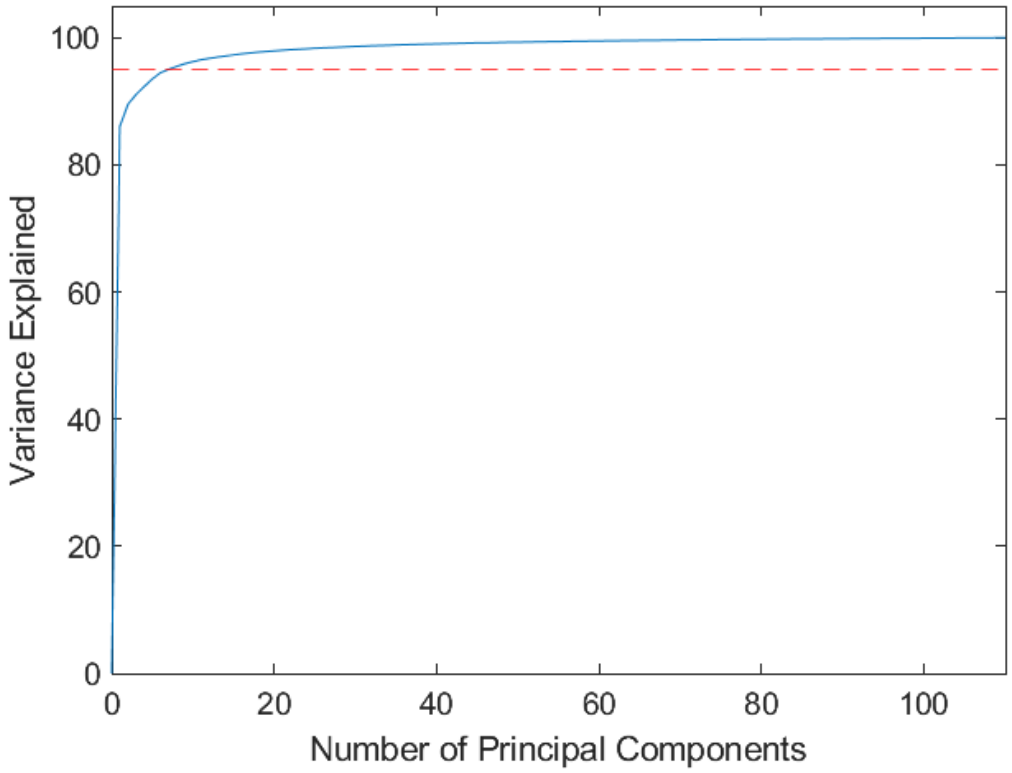

Figure 5-7: Cumulative compactness of the Femur SSM with increasing number of PC modes included.

Table 5-3: Accuracy, variance explained, and cumulative compactness for the first ten PCs of the Femur SSM. Desired accuracy and compactness thresholds are highlighted.

Num PCs	Accuracy (mm)	Variance Explained (%)	Cumulative (%)
1	0.86 +/- 0.21	86.10	86.10
2	0.75 +/- 0.17	3.46	89.56
3	0.69 +/- 0.16	1.53	91.09
4	0.64 +/- 0.14	1.24	92.34
5	0.59 +/- 0.11	1.18	93.52
6	0.55 +/- 0.10	0.99	94.51
7	0.52 +/- 0.09	0.53	95.04
8	0.49 +/- 0.09	0.47	95.51
9	0.47 +/- 0.07	0.41	95.92
10	0.45 +/- 0.07	0.34	96.26

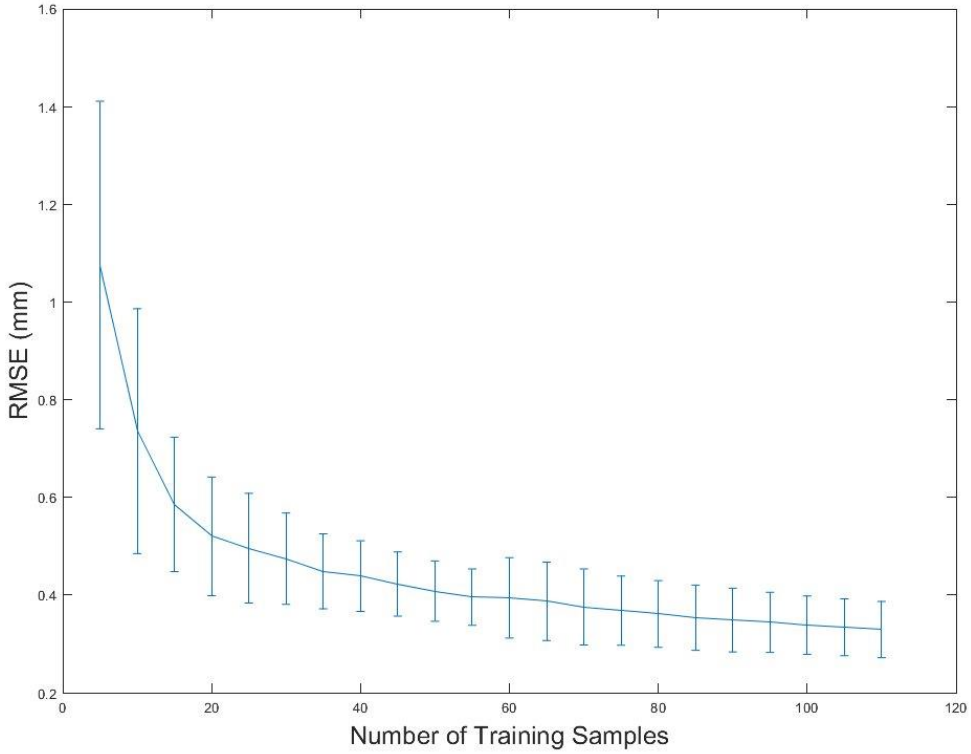

Figure 5-8: Generalizability of the Femur SSM with the progressive addition of training samples to the register.

Table 5-4: Generalizability and specificity for the Femur SSM.

Number of Instances	111
Generalizability (mm)	0.33 +/- 0.06
Specificity (mm)	0.94 +/- 0.15

5.3.1.3 Model Outcomes

The first four modes of shape variation of the Femur SSM are visualized by perturbing the mean geometry at +/- 3.0 standard deviations. Correlation coefficients

between modes of variation and morphological parameters (Table 5-5) and between morphological parameters (Table 5-6) were calculated.

Mode 1 (Figure 5-9) described uniform scaling and was highly correlated with transepicondylar axis length; medial, lateral and centerline anteroposterior height; anterior and posterior mediolateral length; medial and lateral condylar mediolateral width; medial flange of anteroposterior offset; anterior, patellar, and coronal radii of the medial femoral condyle; coronal, horizontal, and patellar groove radii of the lateral femoral condyle, and the angle of divergence of the medial femoral condyle (Table 5-5). Mode 2 (Figure 5-10) described a change in the radius in the region of transition from the metaphysis to diaphysis and a complementary change in shaft size. It was not highly correlated with any morphological parameters but was significantly correlated ($p < 0.05$) with midline and lateral posterior height, medial condylar mediolateral width, notch width, and the angle of divergence of the lateral femoral condyle (Table 5-5). Mode 3 (Figure 5-11) described a change in the anteroposterior size of the condyles, independent of the shaft size and width of the condyles, as well as capturing a possible change in distal femur angle. Mode 3 was not highly correlated with any morphological parameters but was significantly correlated ($p < 0.05$) with medial, lateral, and centerline anteroposterior height, the patellar radius of the lateral femoral condyle, and the angle of divergence of the medial and lateral condyles (Table 5-5). Mode 4 (Figure 5-12) described primarily the curvature of the posterior portion of the condyles in the transverse plane (either towards the midline or away from it). While Mode 4 was not highly correlated with any morphological parameters, it was significantly correlated ($p < 0.05$)

with anterior and posterior mediolateral length, the medial condylar mediolateral width, the lateral flange mediolateral offset, the posterior and anterior condylar angles, and the posterior radius of the medial femoral condyle (Table 5-5).

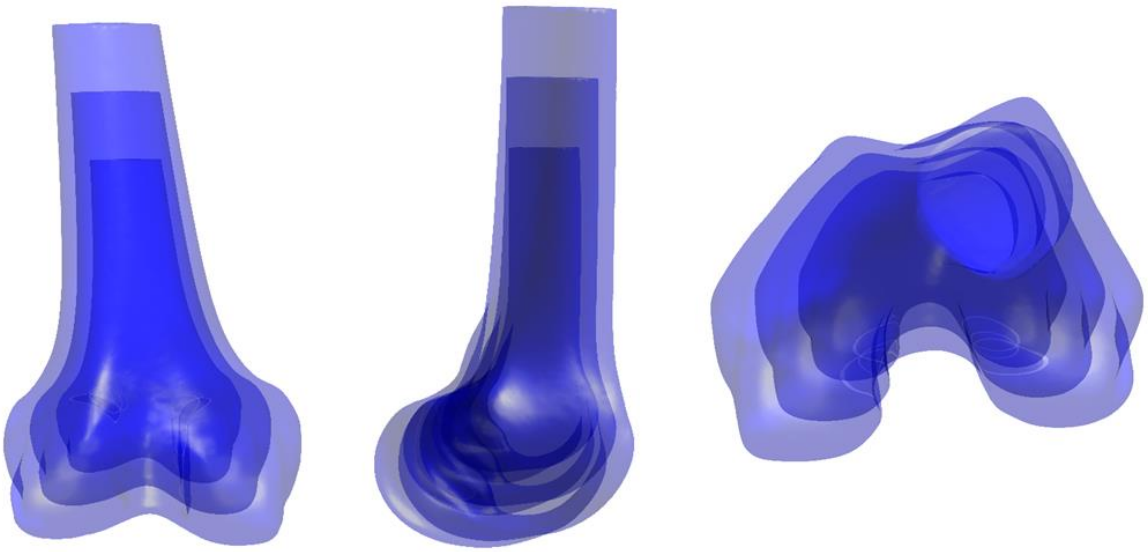

Figure 5-9: Shape variation in Mode 1 of the Femur SSM. Variation is shown to +/- three standard deviations from the mean.

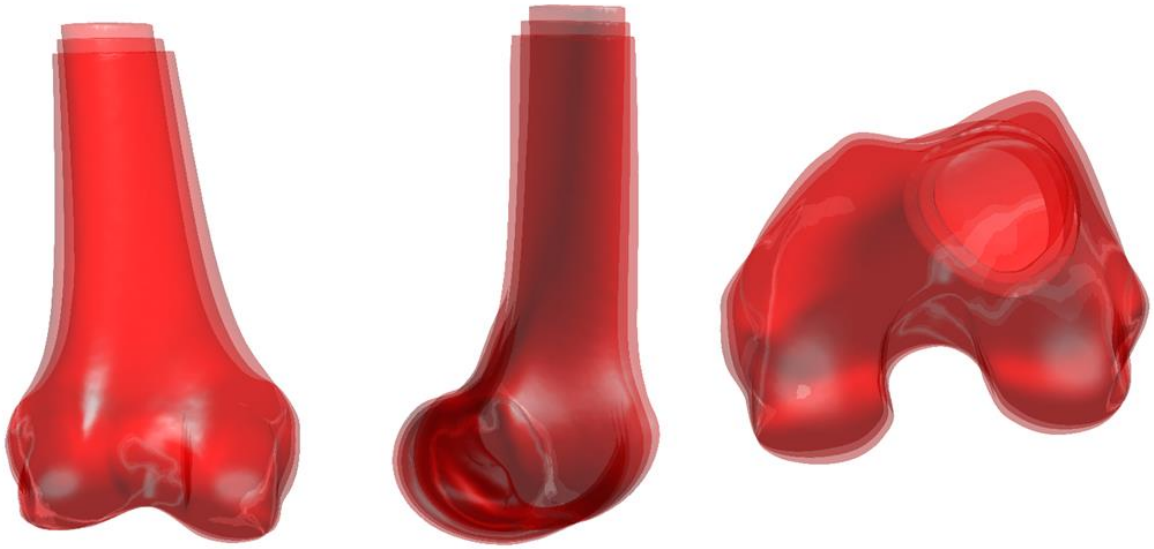

Figure 5-10: Shape variation in Mode 2 of the Femur SSM. Variation is shown to +/- three standard deviations from the mean.

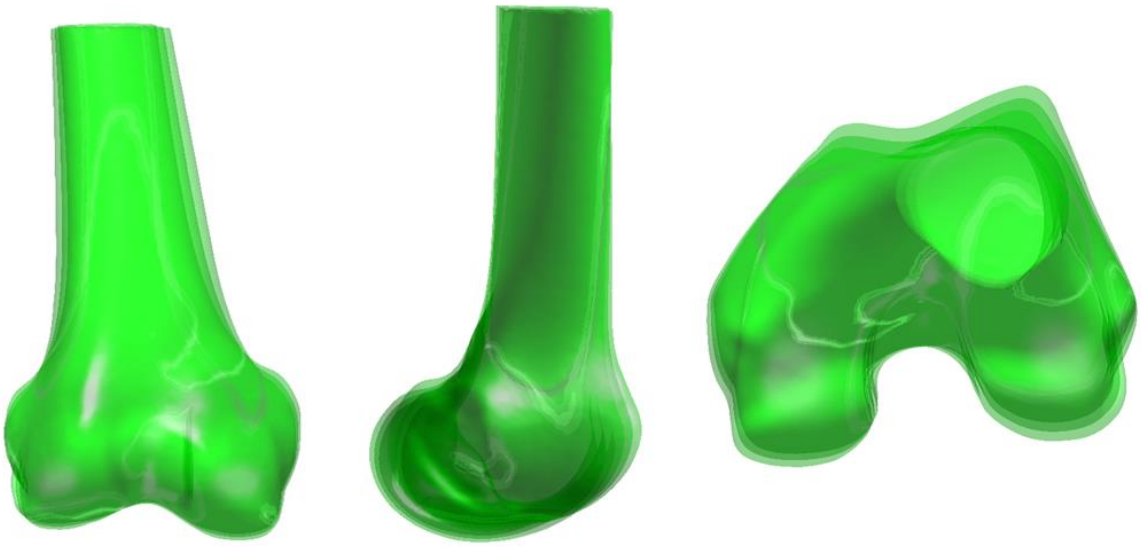

Figure 5-11: Shape variation in Mode 3 of the Femur SSM. Variation is shown to +/- three standard deviations from the mean.

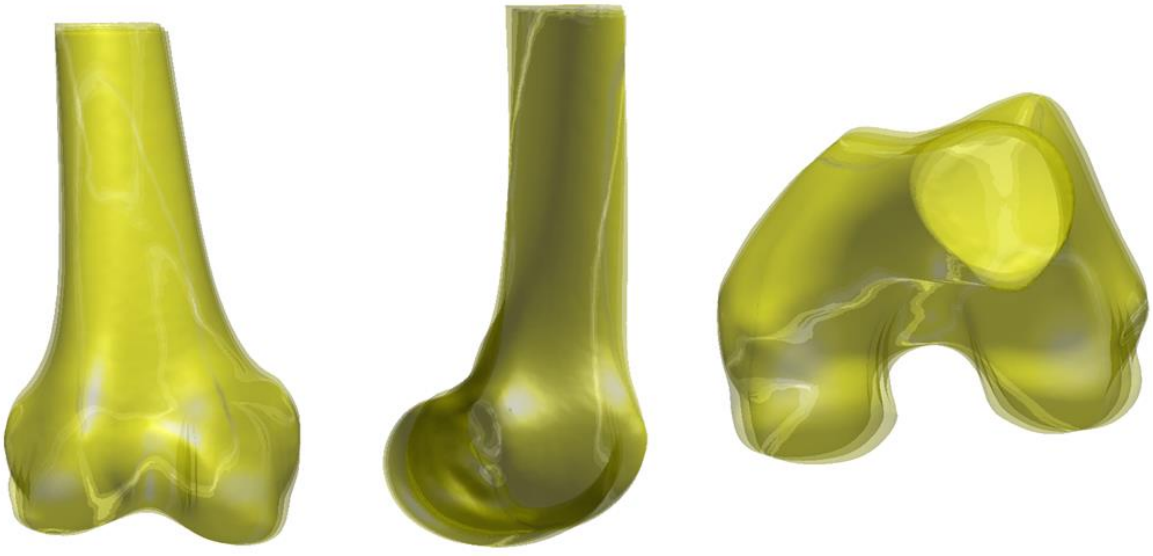

Figure 5-12: Shape variation in Mode 4 of the Femur SSM. Variation is shown to +/- three standard deviations from the mean.

Table 5-5: Correlation coefficients (R) between PC scores and morphological parameters. Only statistically significant ($p<0.05$) correlations are shown. High correlations ($R> 0.65$) are bolded.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
TEA	**0.99**						
APH	**0.87**	0.21	0.26				
MAP	**0.90**		0.26				
LAP	**0.89**	0.23	0.27				
AML	**0.68**			0.28			
PML	**0.93**			0.24			
MCW	**0.65**	0.36		0.19			
LCW	**0.79**						
MACOAP	**0.86**						
LACOAP					-0.26	-0.22	
MACOML	0.48				0.29		
LACOML	0.53			0.20	0.25		
PCA	0.55			0.22	-0.33	-0.23	
ACA	0.28			-0.19	0.56	0.26	
NW		0.21			0.54	0.28	
PRMFC	0.56			0.41			-0.23
ARMFC	**0.86**						
PARMFC	**0.68**						
HRMFC	0.60						
CRMFC	**0.78**						
PRLFC	**0.69**						
ARLFC	**0.82**					0.22	
PARLFC			0.33		-0.31	-0.31	0.21
HRLFC	**0.65**						
CRLFC	**0.86**						
PGRFC	**0.82**						
ADMFC	**0.70**		0.19				-0.22
ADLFC		-0.28	0.19		-0.26	-0.21	

Table 5-6: Correlation coefficients (R) between morphological parameters. Only statistically significant ($p<0.05$) correlations are shown. High correlations ($R> 0.65$) are bolded.

	TEA	APH	MAP	LAP	AML	PML	MCW	LCW	MACOAP	LACOAP	MACOML	LACOML	PCA	ACA	NW	PRMFC	ARMFC	PARMFC	HRMFC	CRMFC	PRLFC	ARLFC	PARLFC	HRLFC	CRLFC	PGRFC
TEA																										
APH	**0.85**																									
MAP	**0.89**	**0.94**																								
LAP	**0.87**	**0.91**	**0.92**																							
AML	**0.68**	0.48	0.52	0.63																						
PML	**0.92**	**0.80**	**0.84**	**0.82**	**0.68**																					
MCW	0.61	0.62	**0.66**	**0.67**	0.60	**0.66**																				
LCW	**0.77**	**0.69**	**0.70**	**0.72**	0.63	**0.70**	0.46																			
MACOAP	**0.86**	**0.76**	**0.81**	**0.80**	**0.67**	**0.86**	0.61	**0.76**																		
LACOAP					0.27			0.19																		
MACOML	0.47	0.26	0.38	0.53	**0.70**	0.48	0.46	0.43	0.51	0.39																
LACOML	0.51	0.33	0.37	0.52	**0.87**	0.54	0.52	0.51	0.53		**0.73**															
PCA	0.57	0.46	0.48	0.43	0.63	0.52	0.35	0.47	0.52	0.42	0.19															
ACA	0.27	0.22	0.35			0.27	0.33		0.33	-0.24	0.32	0.30														
NW				0.22						-0.59	0.42	0.25	-0.49	0.48												
PRMFC	0.58	0.47	0.48	0.46	0.35	**0.72**	0.42		0.39			0.20	0.38													
ARMFC	**0.85**	**0.84**	**0.86**	**0.85**	0.58	**0.80**	**0.65**	**0.70**	**0.80**		0.39	0.45	0.45	0.29		0.44										
PARMFC	**0.65**	0.64	**0.66**	**0.66**	0.38	**0.66**	0.41	0.63	0.64		0.38	0.31	0.28	0.21		0.32	0.57									
HRMFC	0.60	0.61	**0.66**	0.58	0.31	0.56	0.47	0.59	0.61		0.22	0.21	0.36			0.25	0.58	0.51								
CRMFC	**0.75**	**0.69**	**0.70**	**0.72**	0.59	**0.70**	0.46	**0.97**	**0.74**	0.19	0.42	0.47	0.46				**0.71**	**0.67**	0.60							
PRLFC	**0.66**	0.55	0.54	0.61	0.57	0.60	0.38	**0.90**	0.63	0.25	0.39	0.45	0.46				0.56	0.57	0.46	**0.90**						
ARLFC	**0.79**	**0.79**	**0.81**	**0.82**	0.55	**0.79**	**0.65**	**0.66**	**0.79**		0.44	0.43	0.42	0.27		0.43	**0.84**	0.58	0.59	0.67	0.54					
PARLFC											-0.23	-0.26	0.20	-0.29	-0.27											
HRLFC	**0.65**	0.61	0.63	**0.70**	0.59	0.61	0.46	0.58	0.59	0.21	0.57	0.54	0.32	0.22		0.30	0.58	0.49	0.38	0.55	0.53	0.57				
CRLFC	**0.87**	**0.76**	**0.80**	**0.78**	0.64	**0.86**	0.59	**0.72**	**0.89**		0.50	0.51	0.50	0.31		0.46	**0.75**	0.64	0.54	**0.71**	0.64	**0.71**		0.62		
PGRFC	**0.82**	**0.72**	**0.78**	**0.74**	0.61	**0.79**	0.52	**0.75**	**0.87**		0.43	0.45	0.54	0.28		0.39	**0.74**	0.58	0.58	**0.73**	0.64	**0.66**		0.57	**0.88**	
ADMFC	**0.68**	**0.81**	**0.73**	**0.69**	0.31	**0.67**	0.40	0.60	0.61				0.34			0.43	0.63	0.61	0.55	0.61	0.53	0.63		0.46	0.63	0.57
ADLFC				-0.22	-0.55		-0.27	-0.27			-0.57	**-0.82**		-0.32	-0.30	0.26				-0.23	-0.23		0.39	-0.35		

5.3.1.4 Interpopulation Comparisons

The sex and racial distribution of the scores pertaining to the first four principal components are depicted in Figure 5-13 (PC1 and PC2) and Figure 5-14 (PC3 and PC4). To quantify interpopulation differences in PC scores and morphological parameters, unpaired two-sample t-tests and a series of ANOVA tests were conducted. Table 5-7 displays the p-values resulting from t-tests with significant differences ($p < 0.05$) between two populations. Boxplots illustrating the distribution of PCs one through four and select morphological populations across sex, race, and age categories are shown in Figure 5-15 through Figure 5-41. Significant differences ($p < 0.05$) in these plots are indicated by a red bar. Table 5-8 displays significant p-values ($p < 0.05$) resulting from a series of ANOVA tests for each PC or morphological parameter considering sex, race, age, height, and BMI.

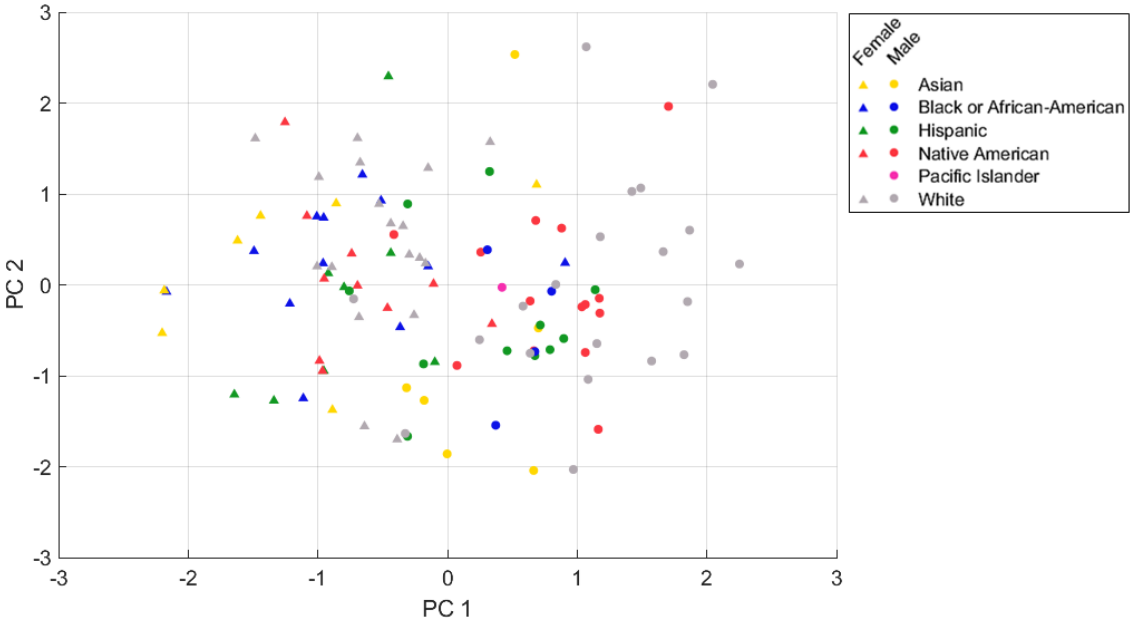

Figure 5-13: Sex and racial distribution of PC 1 and PC 2 scores of the Femur SSM.

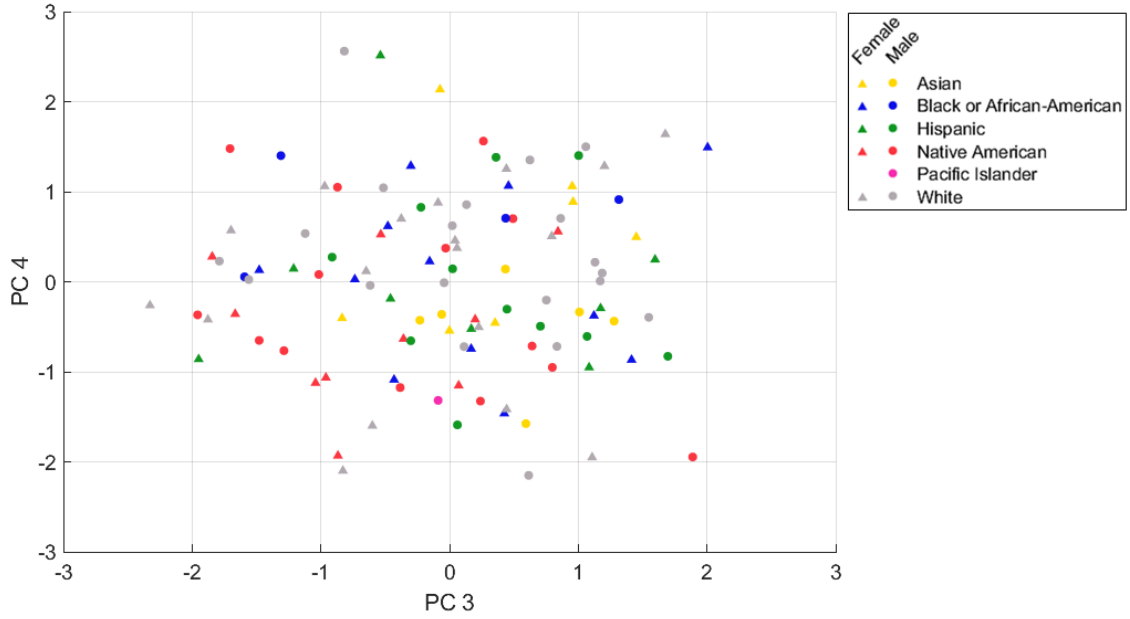

Figure 5-14: Sex and racial distribution of PC 3 and PC 4 scores of the Femur SSM.

Table 5-7: P-values derived from unpaired two-sample t-tests indicating significant differences ($p < 0.05$) between populations. Highly significant differences ($p < 0.001$) are shown in bold.

	M/F	A/B	A/H	A/NA	A/W	B/H	B/NA	B/W	H/NA	H/W	NA/W	YA/MA	YA/OA	MA/OA
PC 1	**5.06E-21**			3.60E-02	1.09E-02		3.18E-02	8.55E-03						
PC 2	3.50E-02											4.91E-02		
PC 3				6.98E-03					4.09E-02					
PC 4														
PC 5					4.95E-02									
PC 6											3.61E-02			
PC 7														
TEA	**1.14E-22**			4.92E-02	2.31E-02		4.25E-02	1.87E-02						
APH	**8.56E-12**				1.11E-02			2.51E-02						
MAP	**4.42E-13**				2.44E-02			4.63E-02						
LAP	**2.11E-12**				9.50E-03									
AML	**1.11E-12**			2.01E-02										
PML	**7.76E-18**				1.52E-02			1.77E-02						
MCW	**2.52E-07**													
LCW	**1.44E-13**				1.92E-02			2.78E-03					3.36E-03	1.68E-03
MACOAP	**2.08E-20**							1.64E-02						4.32E-02
LACOAP														
MACOML	**1.83E-05**													
LACOML	**1.12E-05**			1.76E-02					5.77E-03					
PCA	**3.97E-12**													
ACA														
NW										4.29E-02				
PRMFC	**3.31E-06**				3.31E-02									
ARMFC	**1.06E-13**				6.33E-03			2.56E-03			3.47E-02			
PARMFC	**5.89E-09**				2.69E-02			1.18E-03		3.30E-02	4.87E-03			
HRMFC	**7.57E-08**				4.94E-02						1.38E-02			
CRMFC	**1.58E-12**				8.16E-03			1.07E-03			3.34E-02		4.04E-03	1.81E-03
PRLFC	**3.59E-13**				1.98E-02	4.24E-02	3.27E-02	4.80E-03					1.56E-02	4.73E-03
ARLFC	**3.43E-13**							9.63E-03			2.18E-02			
PARLFC														
HRLFC	**6.44E-08**													
CRLFC	**8.31E-17**							1.42E-02						
PGRFC	**2.28E-16**				4.92E-02			1.98E-02						
ADMFC	**3.85E-06**				4.34E-03		3.87E-02	**1.61E-04**			1.94E-02			3.15E-02
ADLFC				1.65E-02					1.22E-03		9.58E-03			

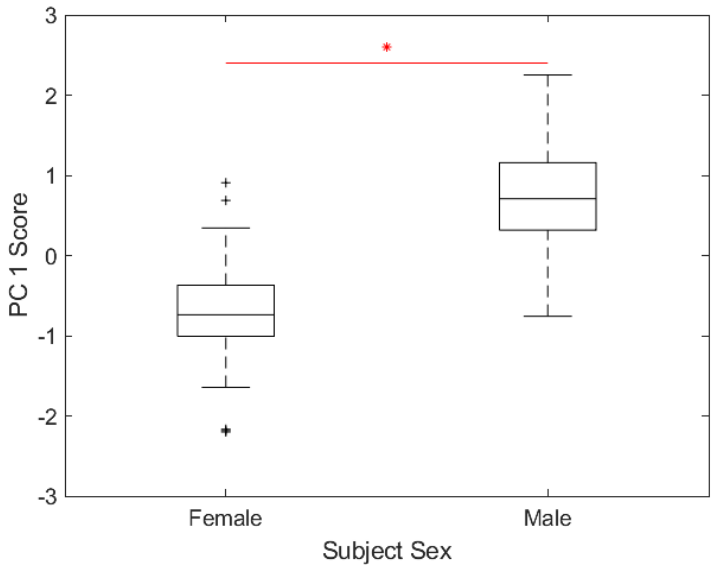

Figure 5-15: Femur SSM PC 1 score by subject sex.

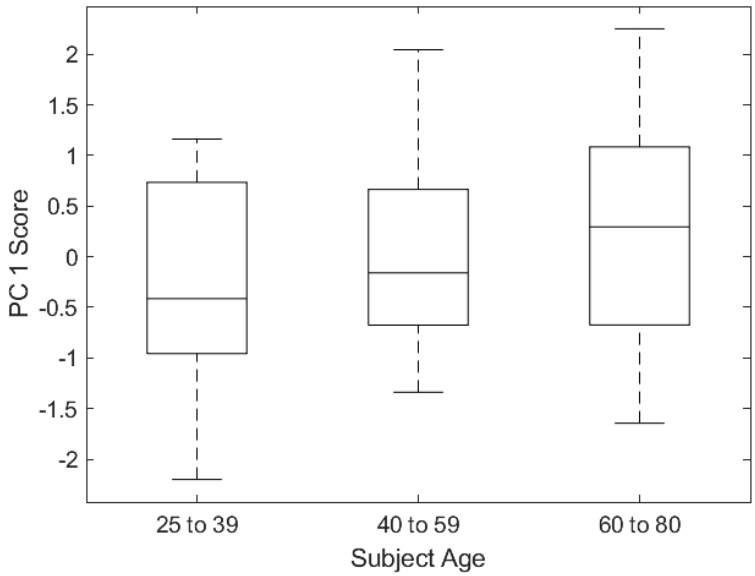

Figure 5-16: Femur SSM PC 1 score by subject age.

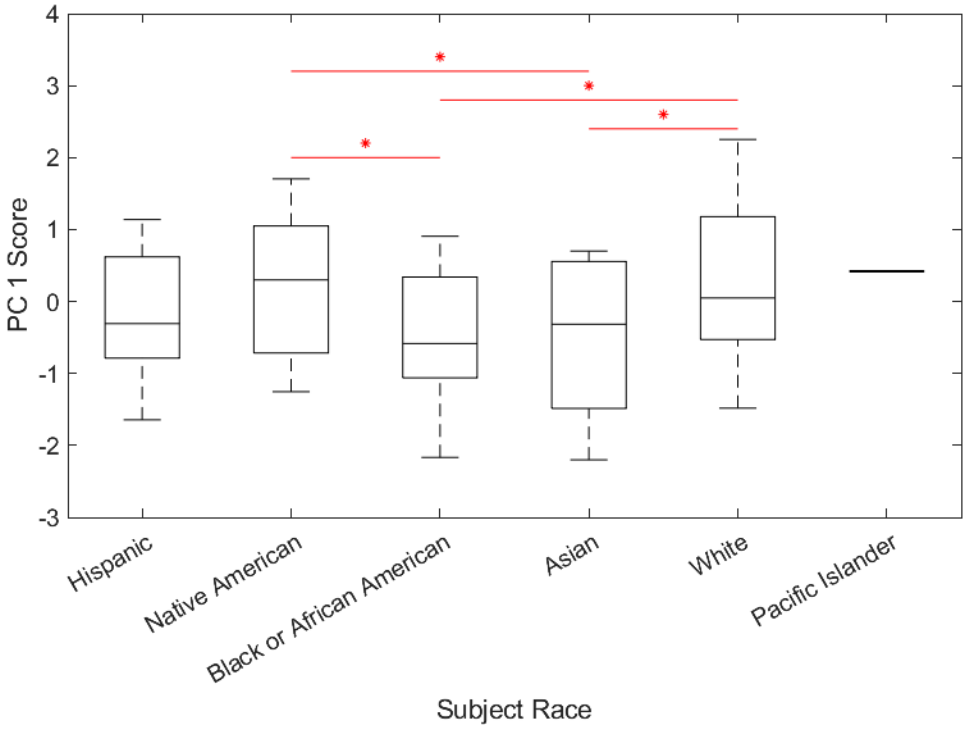

Figure 5-17: Femur SSM PC 1 score by subject race.

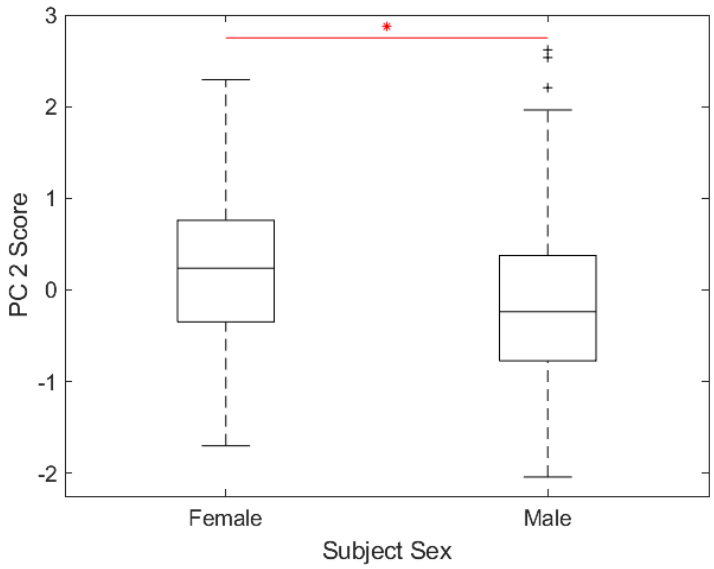

Figure 5-18: Femur SSM PC 2 score by subject sex.

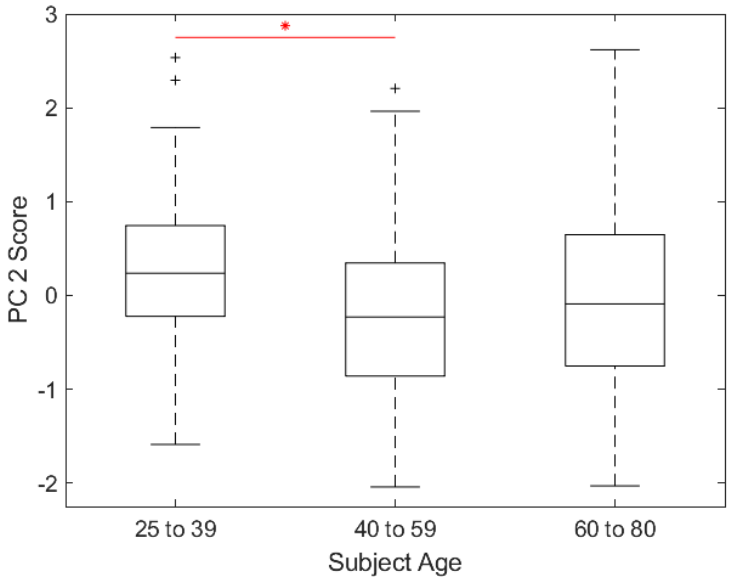

Figure 5-19: Femur SSM PC 2 score by subject age.

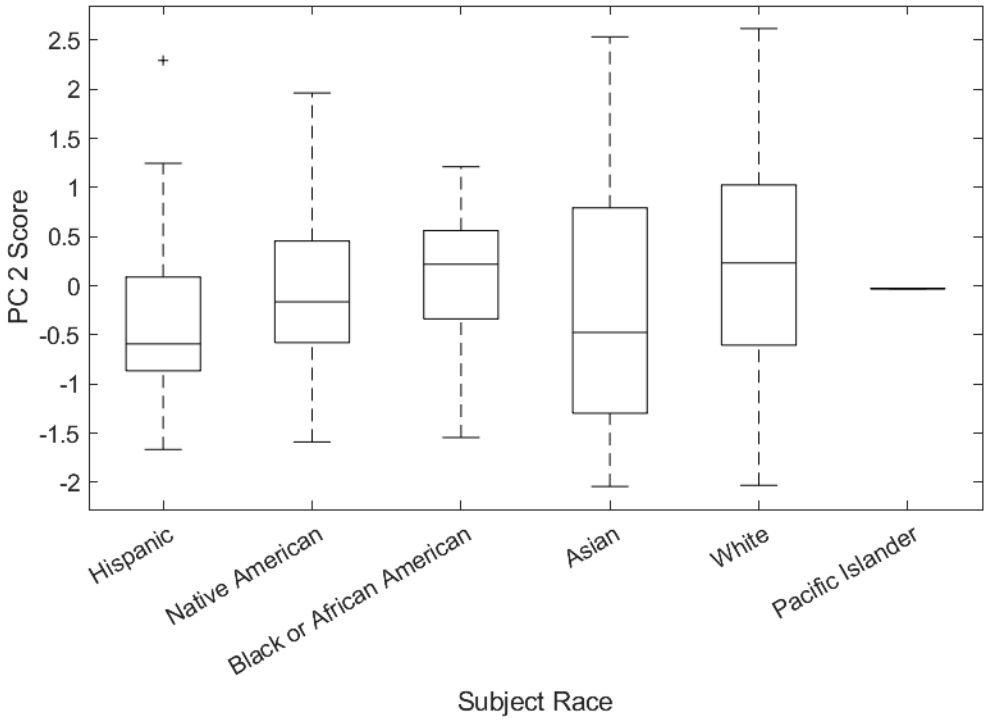

Figure 5-20: Femur SSM PC 2 score by subject race.

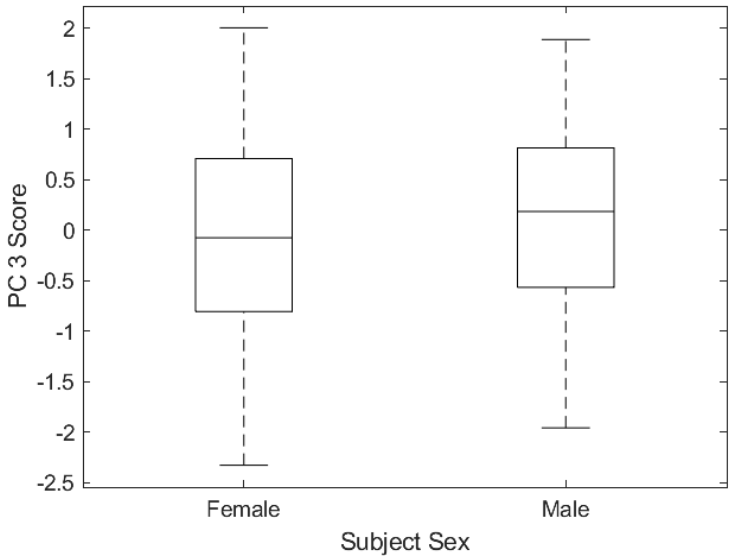

Figure 5-21: Femur SSM PC 3 score by subject sex.

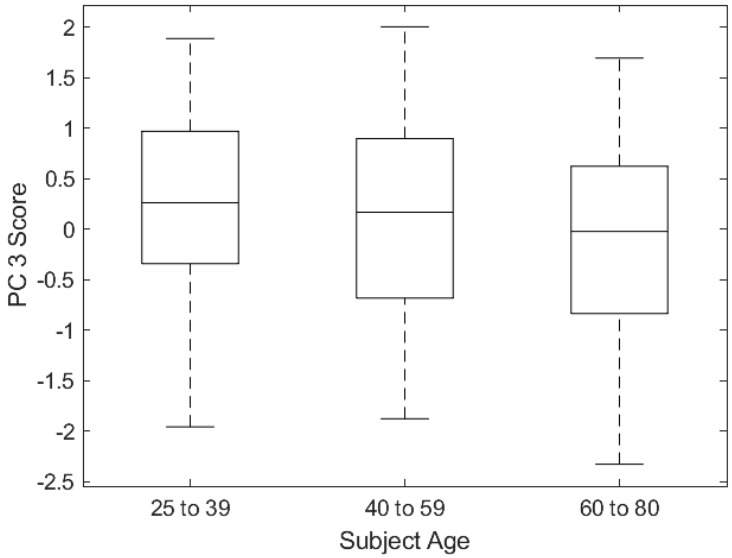

Figure 5-22: Femur SSM PC 3 score by subject age.

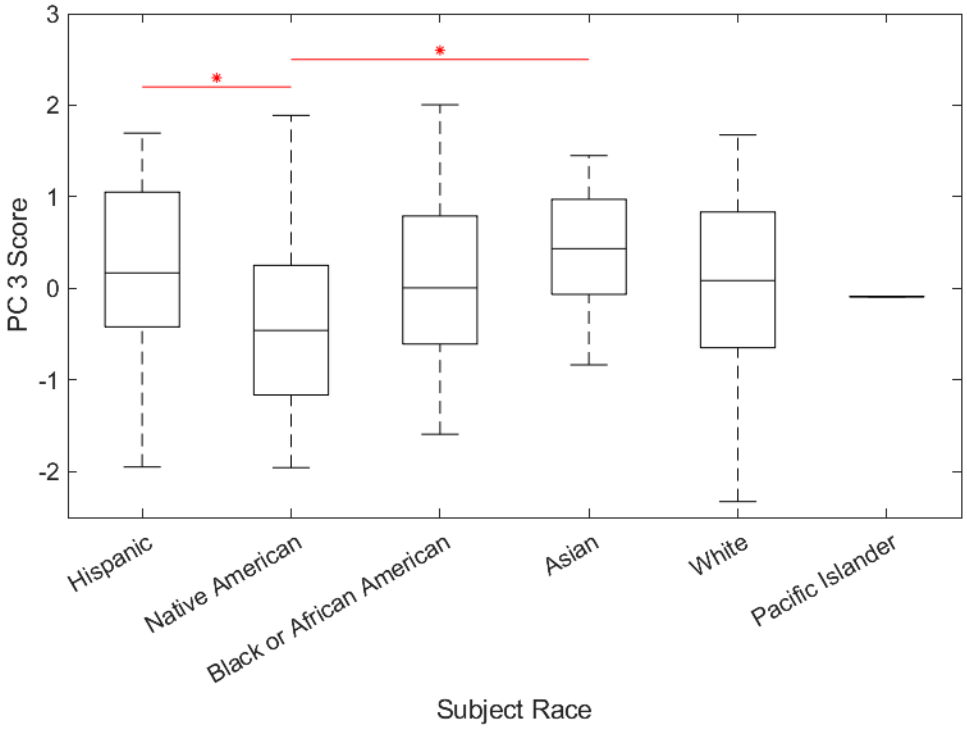

Figure 5-23: Femur SSM PC 3 score by subject race.

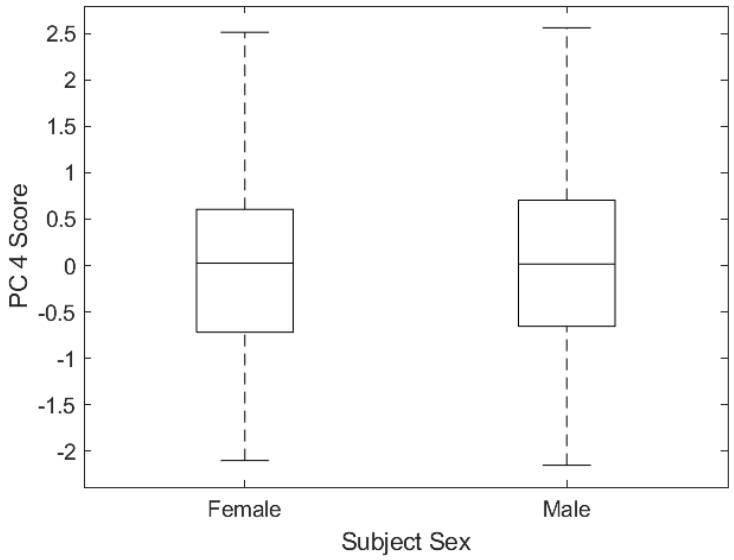

Figure 5-24: Femur SSM PC 4 score by subject sex.

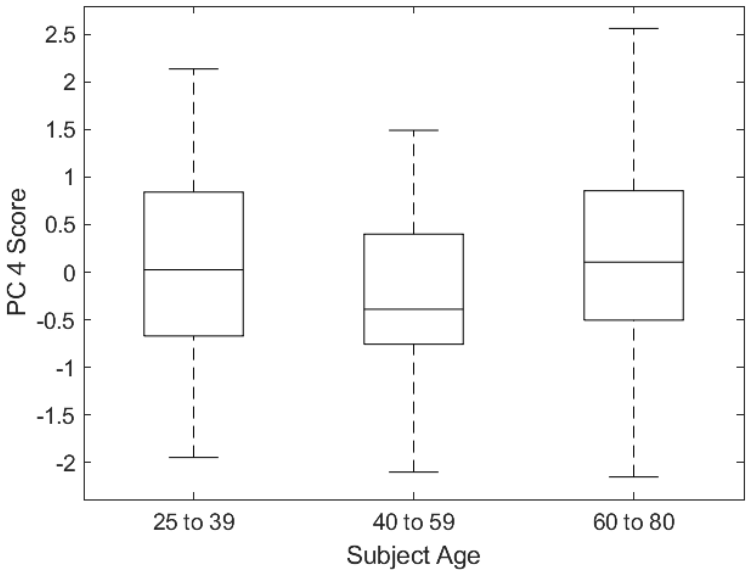

Figure 5-25: Femur SSM PC 4 score by subject age.

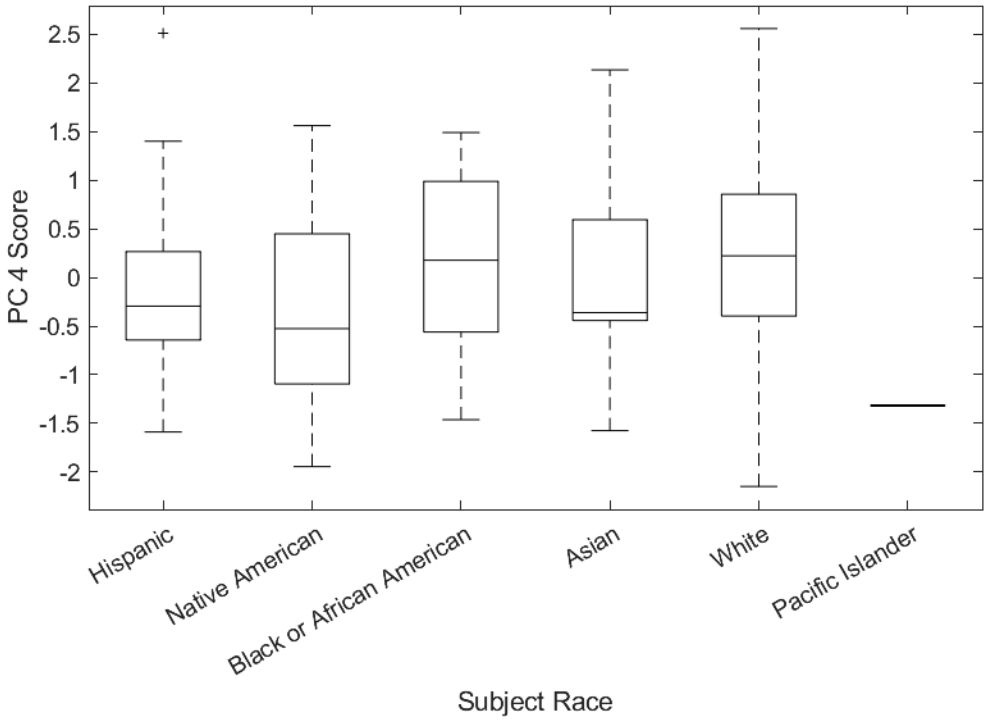

Figure 5-26: Femur SSM PC 4 score by subject race.

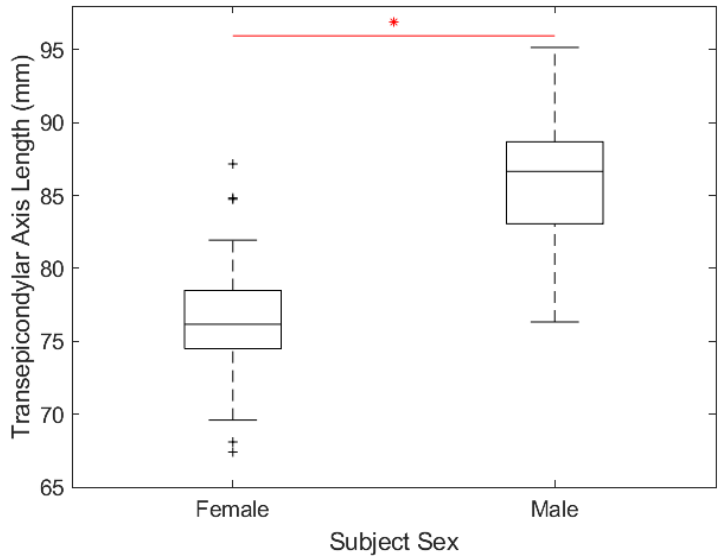

Figure 5-27: Transepicondylar axis length by subject sex.

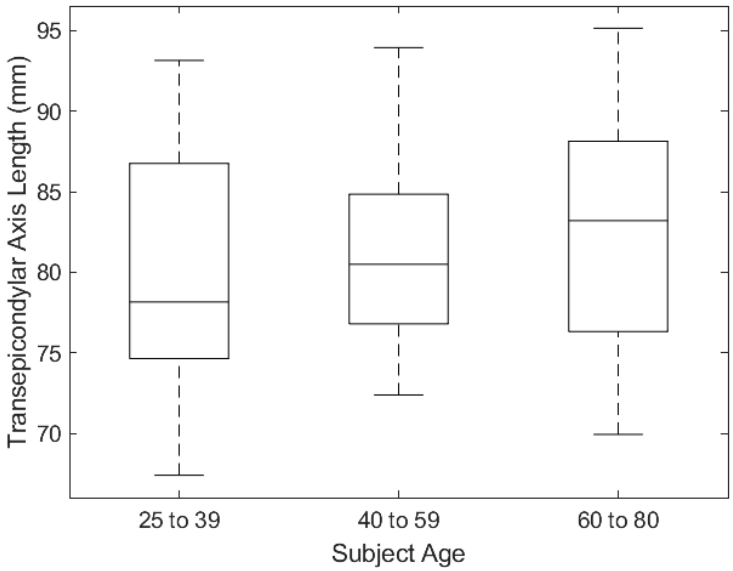

Figure 5-28: Transepicondylar axis length by subject age.

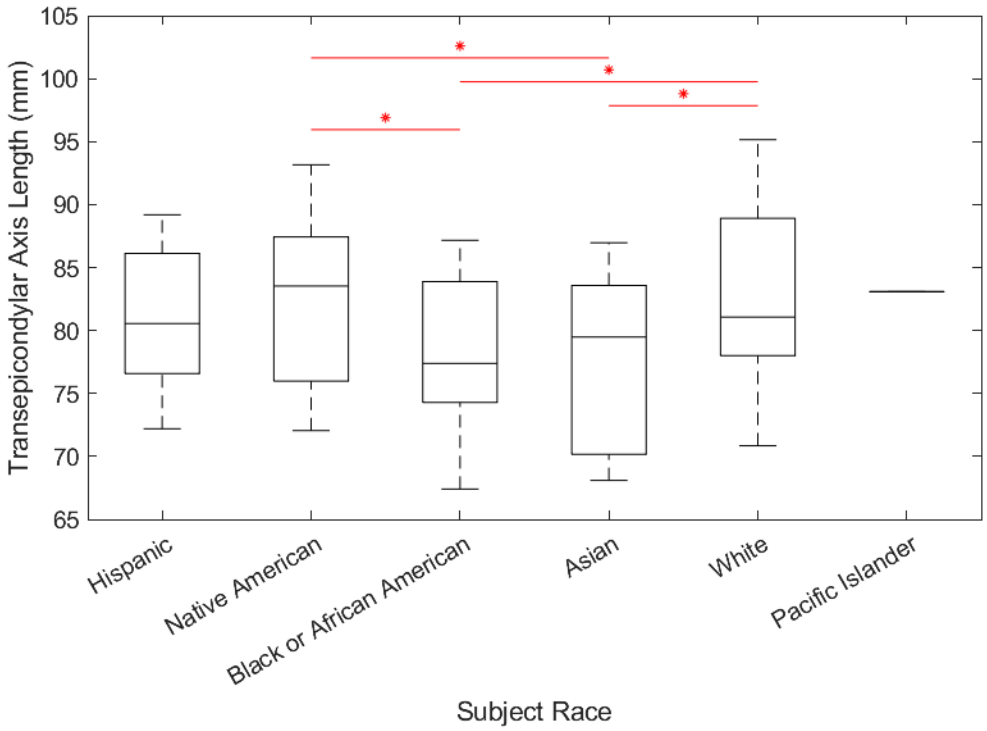

Figure 5-29: Transepicondylar axis length by subject race.

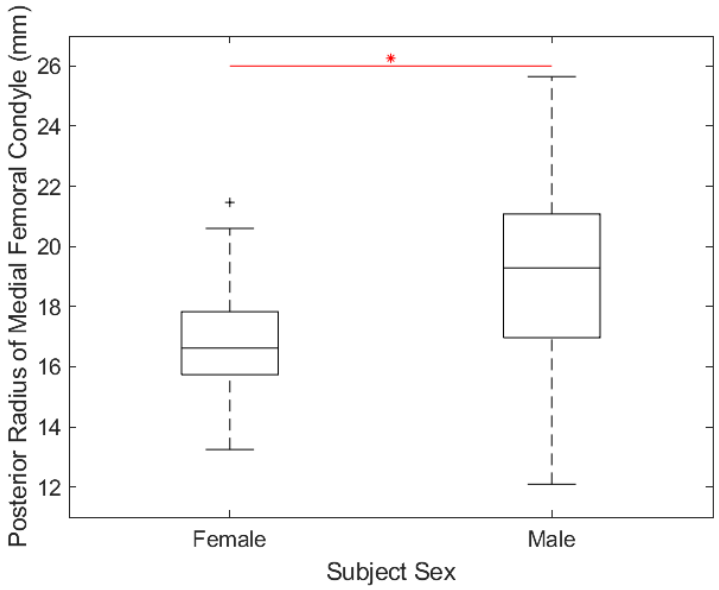

Figure 5-30: Posterior radius of the medial femoral condyle by subject sex.

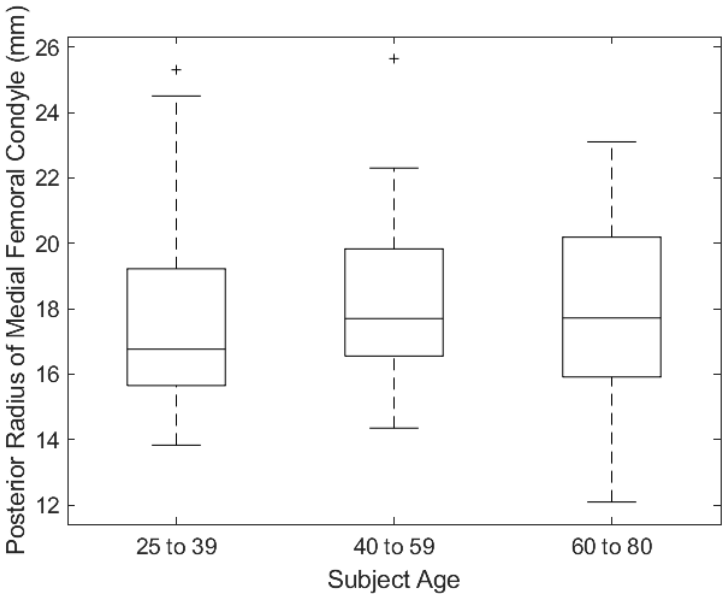

Figure 5-31: Posterior radius of the medial femoral condyle by subject age.

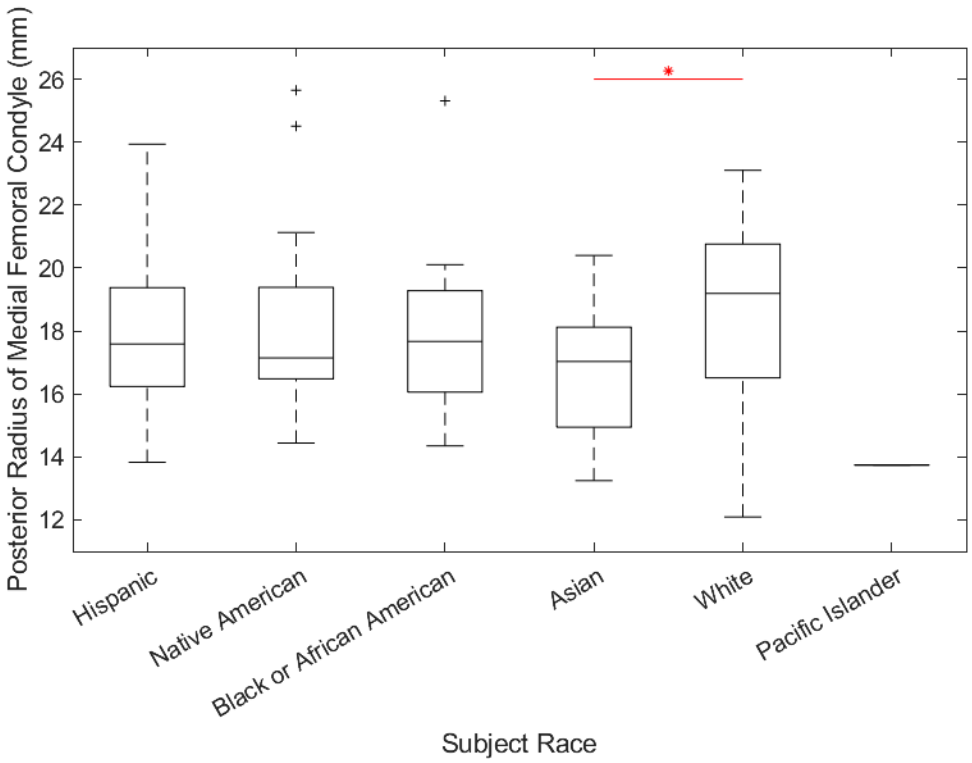

Figure 5-32: Posterior radius of the medial femoral condyle by subject race.

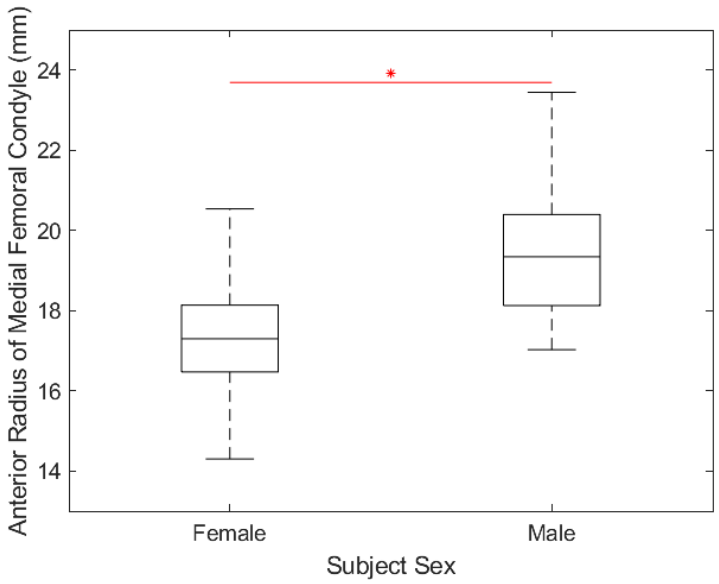

Figure 5-33: Anterior radius of the medial femoral condyle by subject sex.

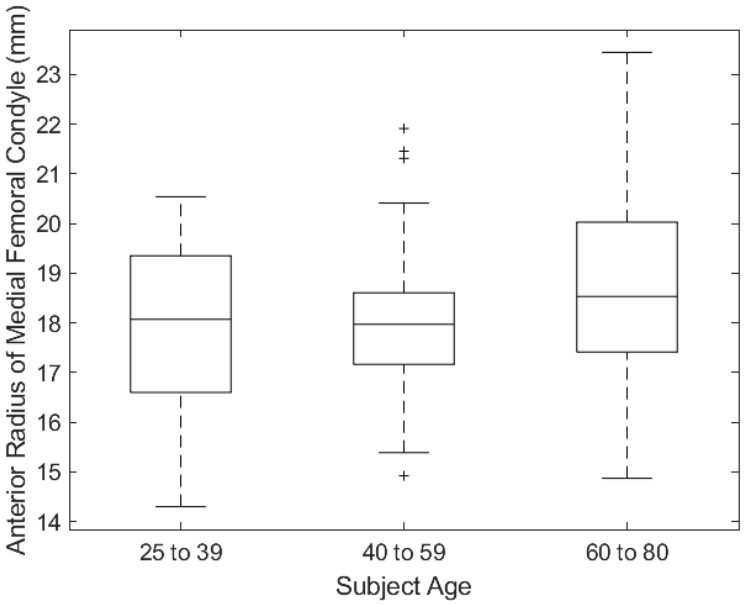

Figure 5-34: Anterior radius of the medial femoral condyle by subject race.

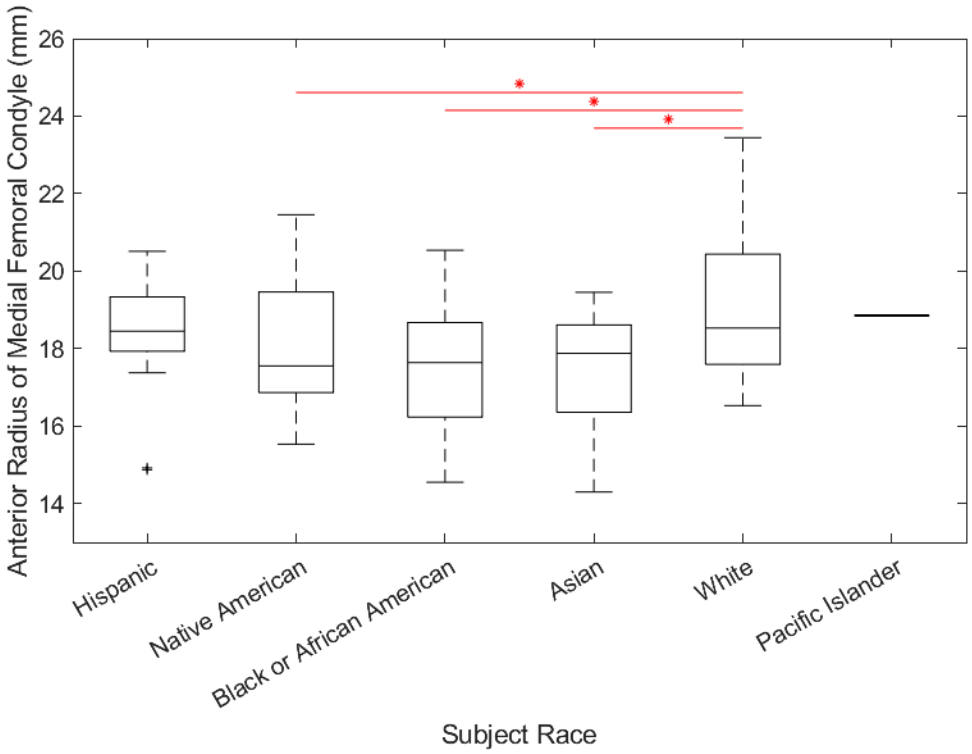

Figure 5-35: Anterior radius of the medial femoral condyle by subject race.

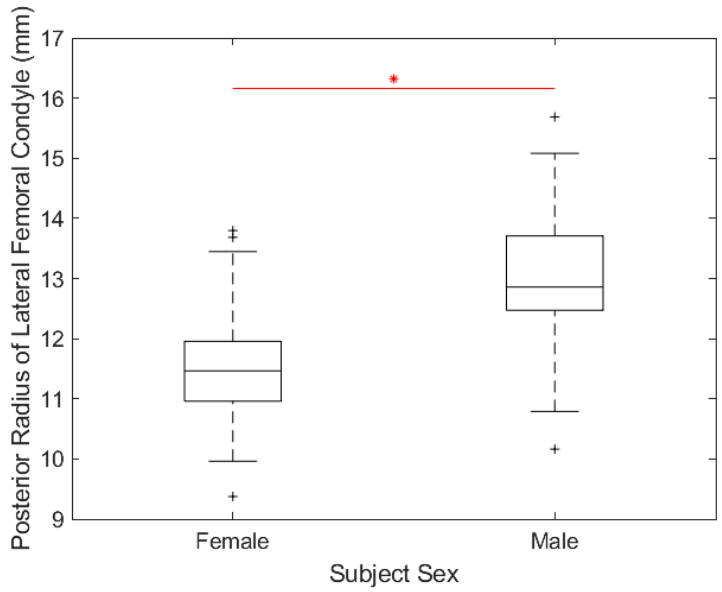

Figure 5-36: Posterior radius of the lateral femoral condyle by subject sex.

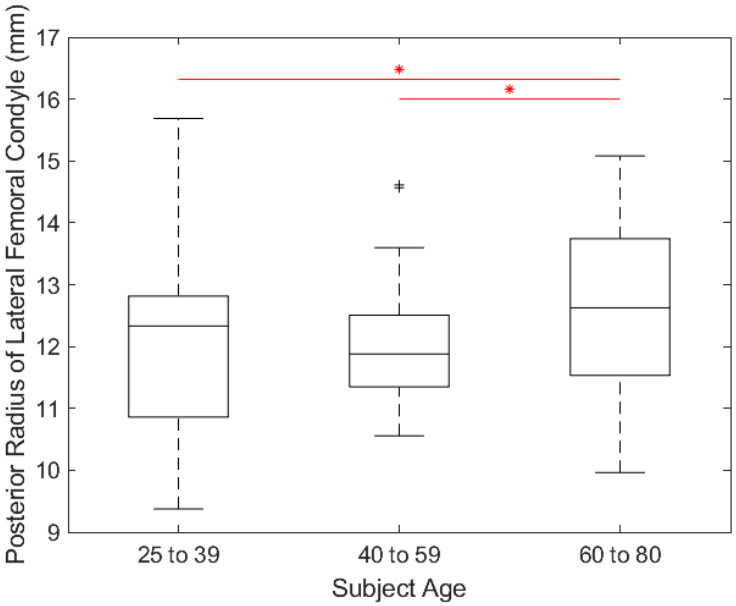

Figure 5-37: Posterior radius of the lateral femoral condyle by subject age.

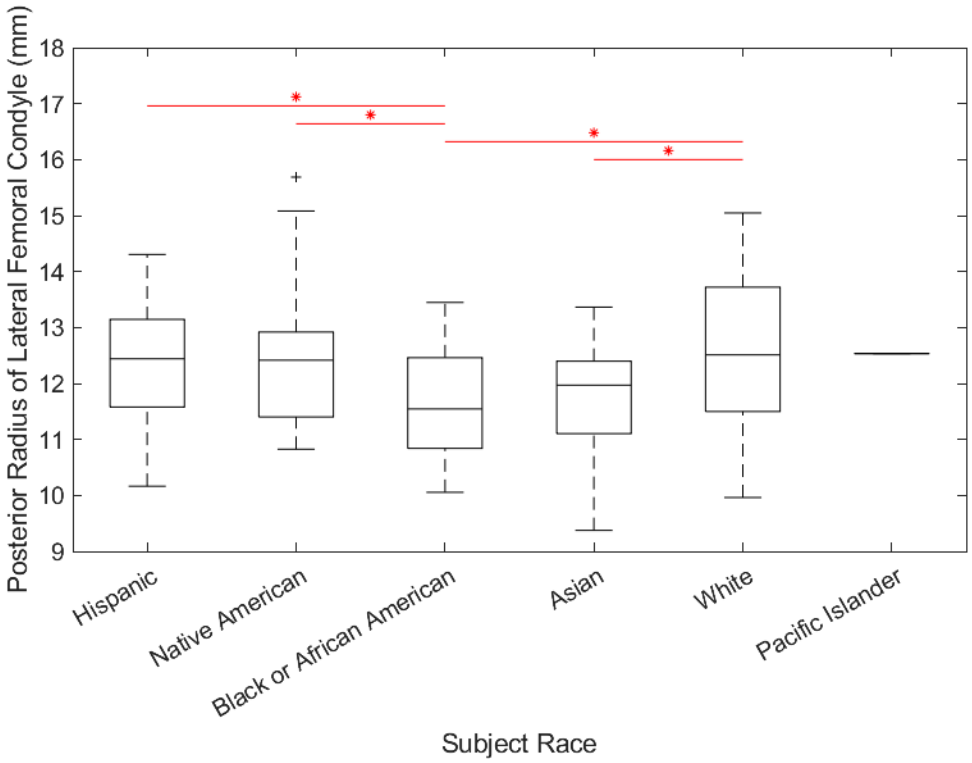

Figure 5-38: Posterior radius of the lateral femoral condyle by subject race.

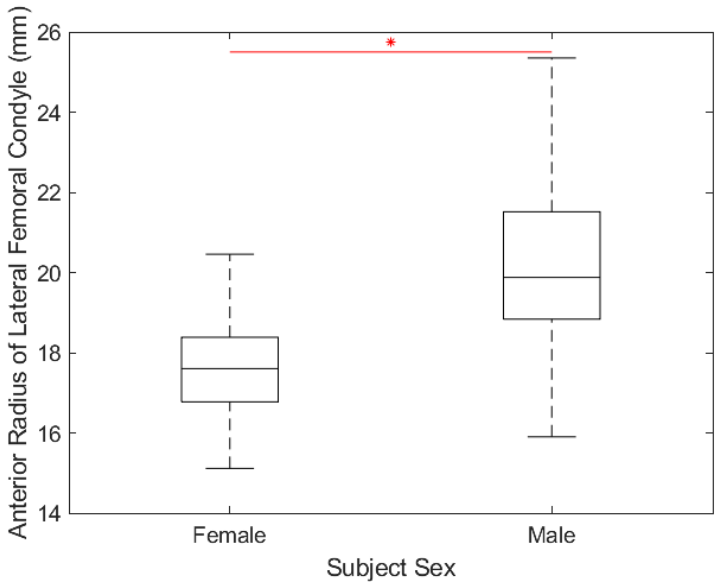

Figure 5-39: Anterior radius of the lateral femoral condyle by subject sex.

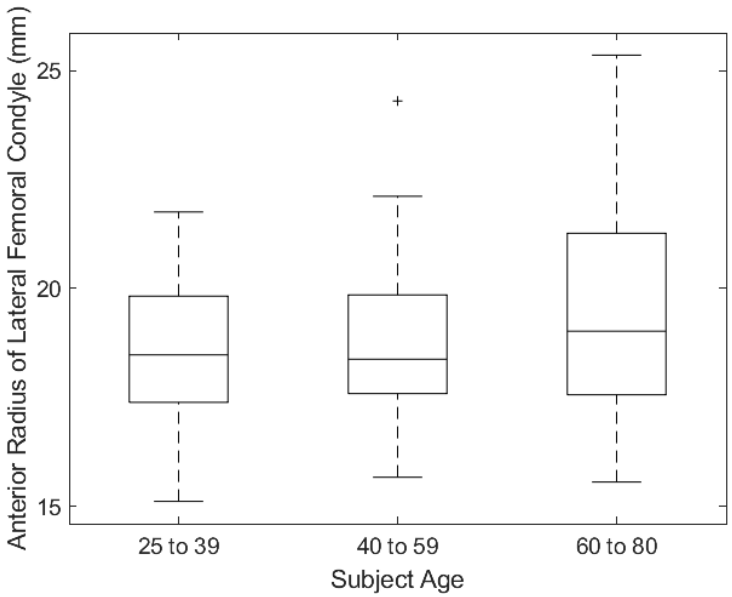

Figure 5-40: Aneterior radius of the lateral femoral condyle by subject age.

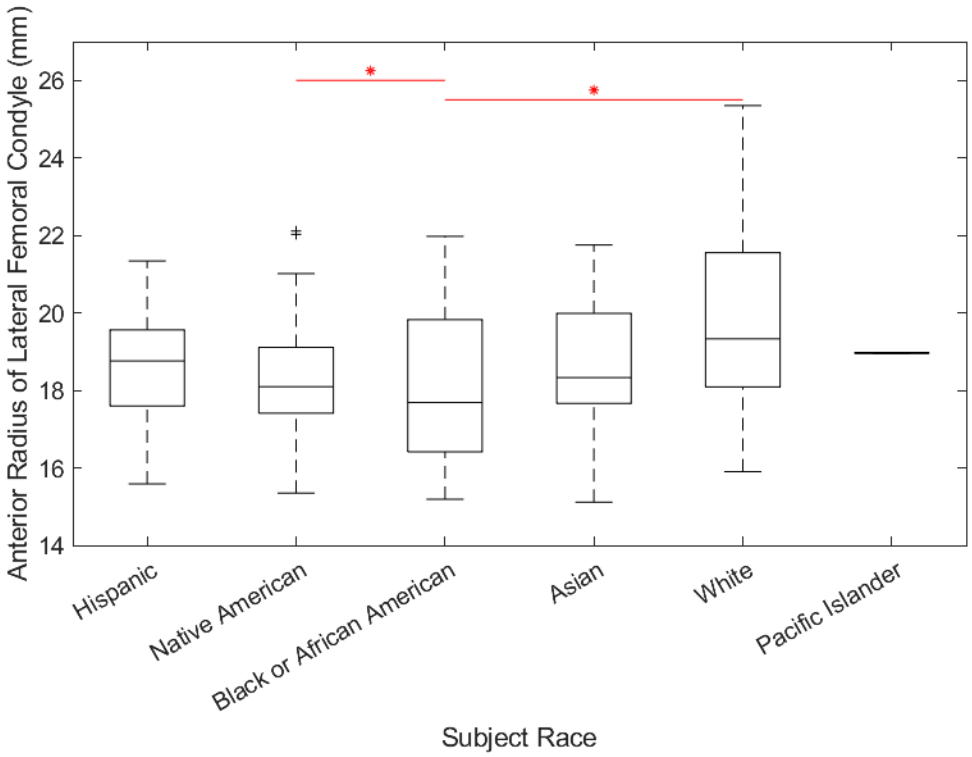

Figure 5-41: Anterior radius of the lateral femoral condyle by subject race.

Table 5-8: ANOVA-derived p-values indicating significant differences ($p < 0.05$) between populations. Highly significant differences ($p < 0.001$) are shown in bold.

	Sex	Age	Race	Height	BMI
PC1	**2.09E-11**	3.10E-02		**5.40E-06**	
PC2	**5.39E-05**	2.08E-03	2.89E-02	**2.22E-04**	
PC3					
PC4					
PC5					
PC6					
PC7		4.94E-02			
TEA	**9.82E-13**	3.02E-02		**5.98E-05**	
APH	**1.59E-04**			**4.53E-05**	
MAP	**1.02E-04**			**2.17E-06**	
LAP	**1.22E-04**			**1.26E-06**	
AML	**8.06E-06**			4.26E-03	
PML	**1.33E-09**			**2.12E-04**	
MCW				**1.89E-10**	
LCW	**9.94E-07**	1.00E-02		1.42E-02	
MACOAP	**1.69E-11**			9.14E-03	
LACOAP					
MACOML	3.19E-02			1.83E-02	
LACOML	2.94E-02			1.74E-02	
PCA	**1.22E-06**				
ACA					
NW					
PRMFC	1.25E-03				
ARMFC	**4.23E-05**			**7.38E-07**	
PARMFC	**6.12E-04**			1.29E-02	
HRMFC	**9.10E-06**		4.23E-03		
CRMFC	**4.27E-06**	4.48E-02		9.02E-03	
PRLFC	**2.61E-07**	4.91E-02			
ARLFC	**3.80E-05**			**1.24E-06**	
PARLFC					
HRLFC	**2.13E-03**			2.23E-02	
CRLFC	**5.11E-08**			2.73E-03	
PGRFC	**4.38E-09**				
ADMFC	1.37E-02		2.81E-02		
ADLFC					

5.3.2 Tibia

5.3.2.1 Subject Demographics

Initially, an equal number of scans were evaluated for each demographic category, ensuring balanced representation by sex, race, and a range of ages. However, the training set was unevenly reduced due to issues with scan quality, such as limited field of view or poor tissue differentiation for material property assignment, and the presence of severe osteoarthritis or osteophytes.

The resulting training data for the tibia represented 115 cases, 58 female and 57 male, with an average age of 52.7 (+/- 16.9) and 54.1 (+/- 18.5) years, respectively. Subjects obtained from the NMDID database (n = 72) came from Asian (n = 13), Black or African American (n = 17), Hispanic (n = 18), and Native American (n = 26) backgrounds. The remaining subjects (n = 42) included White (n = 41) and Hispanic (n = 1) subjects and were derived from a cadaveric training set previously developed the University of Denver. The distribution of the overall subject demographics is shown in Table 5-9. Subject age, height, and BMI data by demographic are shown in Figure 5-42, Figure 5-43, and Figure 5-44, respectively

Table 5-9: Subject demographics for the Tibia SSM training set.

	Asian	Black or African American	Hispanic	Native American	White	Total
Female	7	13	7	11	20	58
Age (Years)	45.7 +/- 19.3	46.6 +/- 15.0	41.0 +/- 14.3	46.5 +/- 9.9	66.6 +/- 12.7	52.7 +/- 16.9
Height (cm)	161.6 +/- 7.3	164.8 +/- 7.6	159.9 +/- 8.0	165.7 +/- 9.2	162.0 +/- 6.4	163.0 +/- 7.5
BMI (kg/m^2)	25.5 +/- 7.6	27.6 +/- 5.5	25.8 +/- 4.9	28.1 +/- 5.3	24.6 +/- 6.9	26.2 +/- 6.2
Male	6	4	11	15	21	57
Age (Years)	47 +/- 10.0	43.0 +/- 6.2	41.4 +/- 13.7	41.2 +/- 13.5	74.0 +/- 6.2	54.1 +/- 18.5
Height (cm)	171.3 +/- 8.7	174.0 +/- 3.3	172.5 +/- 4.1	175.2 +/- 9.3	174.8 +/- 8.0	174.0 +/- 7.5
BMI (kg/m2)	25.9 +/- 3.3	26.2 +/- 3.7	28.0 +/- 5.6	29.1 +/- 5.5	21.0 +/- 4.9	25.3 +/- 6.0
Total	13	17	18	26	41	115
Age (Years)	46.3 +/- 15.1	45.8 +/- 13.4	41.2 +/- 13.5	43.5 +/- 12.2	70.4 +/- 10.5	53.4 +/- 17.7
Height (cm)	166.1 +/- 9.2	167.0 +/- 7.8	167.6 +/- 8.5	171.2 +/- 10.2	168.5 +/- 9.6	168.5 +/- 9.3
BMI (kg/m2)	25.7 +/- 5.8	27.2 +/- 5.1	27.1 +/- 5.3	28.7 +/- 5.4	22.8 +/- 6.2	25.8 +/- 6.1

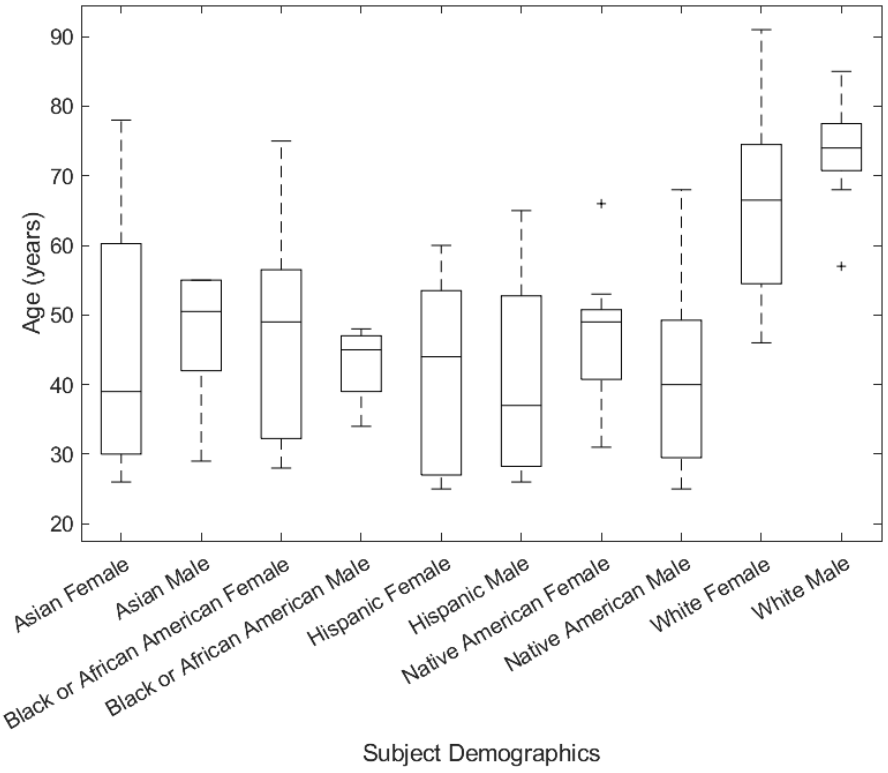

Figure 5-42: Distribution of subject age by demographic.

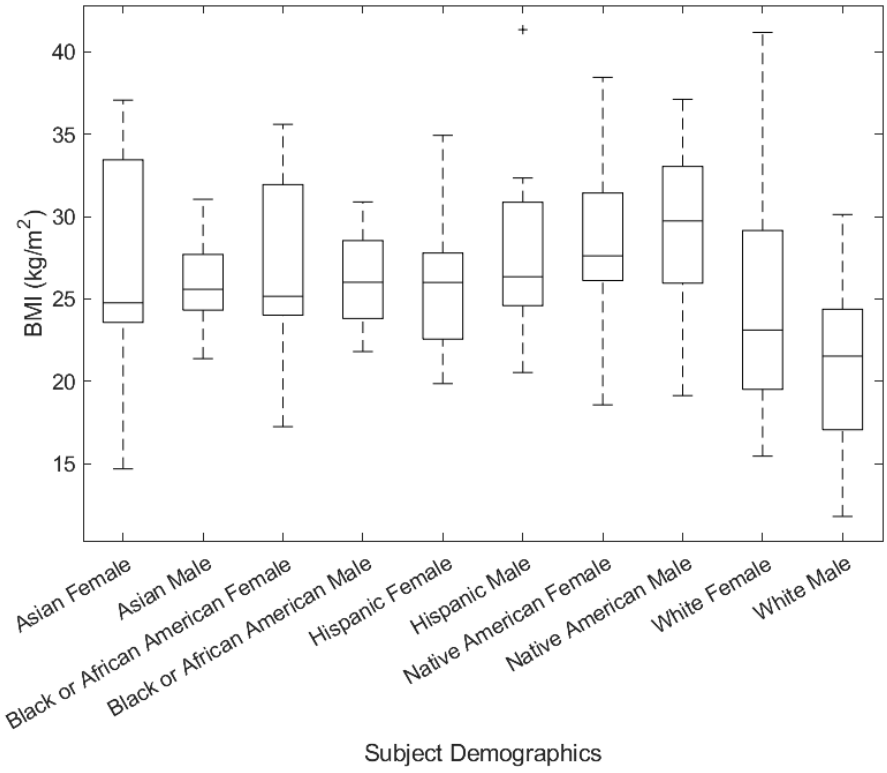

Figure 5-43: Distribution of subject BMI by demographic.

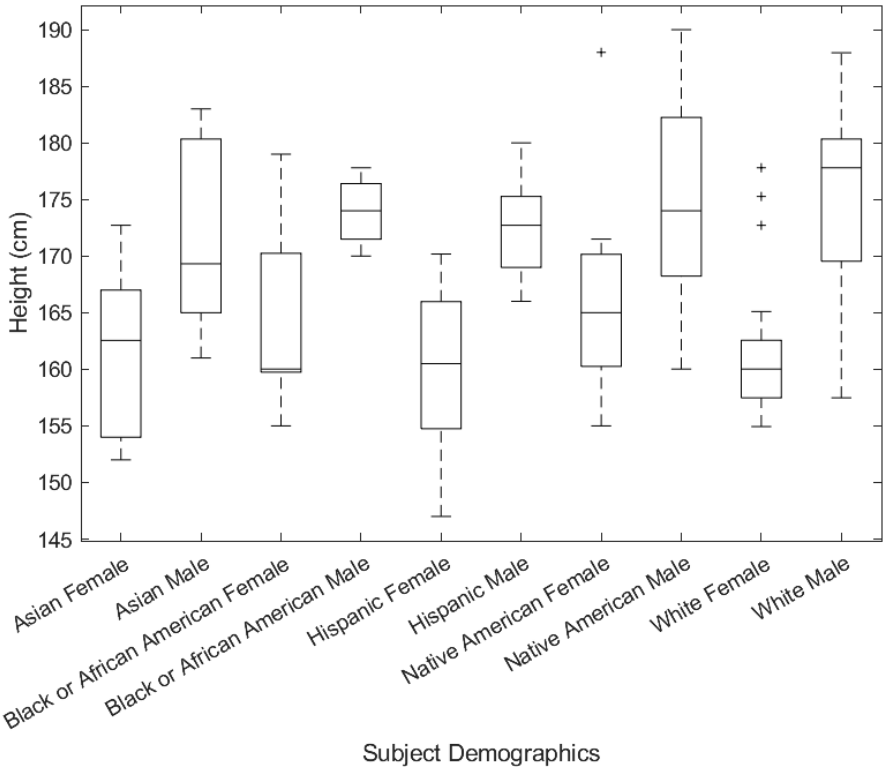

Figure 5-44: Distribution of subject height by demographic.

5.3.2.2 Model Quality

As with the Femur SSM, the accuracy of the Tibia SSM improved with the inclusion of a greater number of principal components (Figure 5-45). Achieving the clinically relevant submillimeter distance error of 0.6 mm required the incorporation of four principal components. Accuracy achieved by inclusion of the first ten principal components is outlined in Table 5-10. Figure 5-46 depicts the cumulative compactness of the Femur SSM, indicating that 14 principal components were necessary to encompass at least 95% of the model variance. The distribution of variance across the first ten principal

components is reported in Table 5-10. The generalizability of the model with increasing numbers of training set samples is shown in Figure 5-47. With 115 instances, the generalizability and specificity of the model were found to be 0.34 +/- 0.04 mm and 0.87 +/-0.15 mm, respectively (Table 5-11).

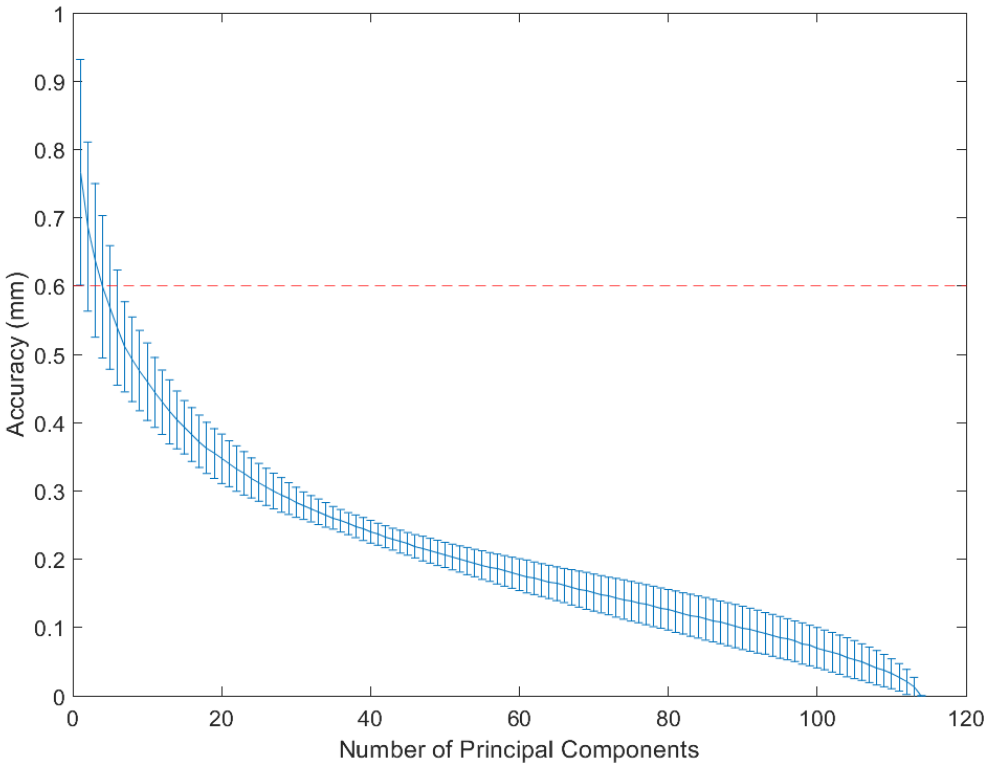

Figure 5-45: Accuracy of the Tibia SSM with the progressive addition of principal components to the model.

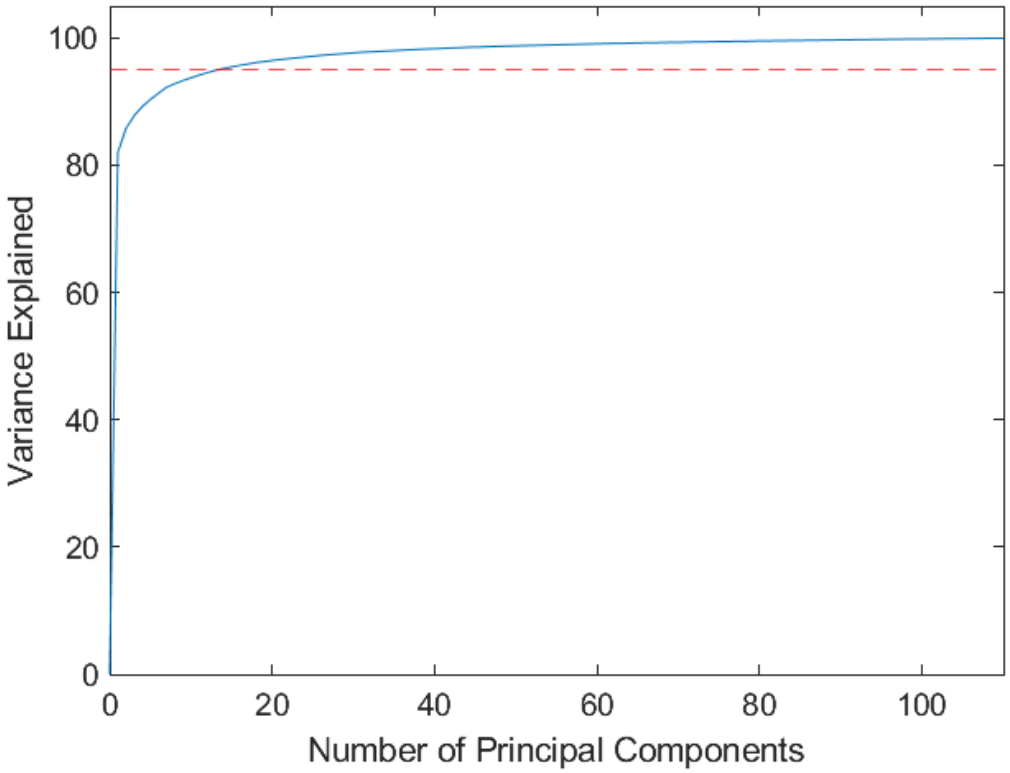

Figure 5-46: Cumulative compactness of the Tibia SSM for increasing number of PCs included.

Table 5-10: Accuracy, variance explained, and cumulative compactness for the first ten PCs of the Tibia SSM. Desired accuracy and compactness thresholds are highlighted.

Num PCs	Accuracy (mm)	Variance Explained (%)	Cumulative (%)
1	0.77 +/- 0.17	82.06	82.06
2	0.69 +/- 0.12	3.72	85.78
3	0.64 +/- 0.11	1.99	87.77
4	0.60 +/- 0.10	1.44	89.21
5	0.57 +/- 0.09	1.11	90.33
6	0.54 +/- 0.08	0.98	91.31
7	0.51 +/- 0.07	0.94	92.25
8	0.49 +/- 0.06	0.55	92.80
9	0.48 +/- 0.06	0.48	93.28
10	0.46 +/- 0.06	0.45	93.73
11	0.44 +/- 0.05	0.43	94.16
12	0.43 +/- 0.05	0.38	94.54
13	0.42 +/- 0.05	0.35	94.89
14	0.40 +/- 0.04	0.29	95.19
15	0.39 +/- 0.04	0.26	95.44

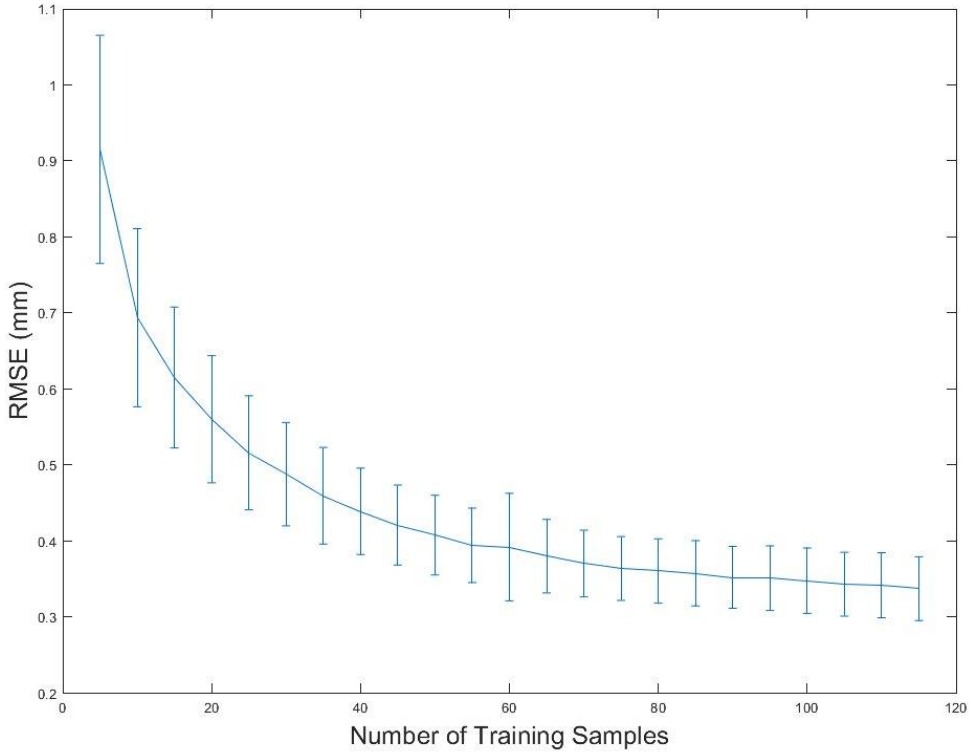

Figure 5-47: Generalizability of the Tibia SSM with the progressive addition of training samples to the register.

Table 5-11: Generalizability and specificity for the Tibia SSM.

Number of Instances	115
Generalizability (mm)	0.34 +/- 0.04
Specificity (mm)	0.87 +/- 0.15

5.3.2.3 Model Outcomes

The first four modes of shape variation of the Tibia SSM were visualized by perturbing the mean geometry at +/- 3.0 standard deviations. Correlation coefficients

between modes of variation and morphological parameters (Table 5-12) and between morphological parameters (Table 5-13) were calculated.

Mode 1 (Figure 5-48) depicted uniform scaling and exhibited strong correlations with various morphological parameters, including the size of the proximal tibia plateau, medial and lateral tibial posterior slopes, tubercle anteroposterior offset, tibial spine width, and medial plateau width (Table 5-12). Mode 2 (Figure 5-49) described a change in the “neck” of the tibia and a corresponding change in shaft size. While not highly correlated with any morphological parameters, Mode 2 was significantly correlated ($p < 0.05$) with posterior overhang and tubercle anteroposterior offset (Table 5-12). Mode 3 (Figure 5-50) described the sagittal plane "shelfiness" of the tibia and a lateralization of the tubercle. Although not strongly correlated with any morphological parameters, Mode 3 showed significant correlations ($p < 0.05$) with the anteroposterior length of the proximal tibia plateau, medial slope of the tibial plateau, posterior overhang, tubercle position, and the width of the medial and lateral plateaus (Table 5-12). Mode 4 (Figure 5-51) primarily described a medialization of the tubercle and was significantly correlated ($p < 0.05$) to the anteroposterior length of the proximal tibial plateau, the lateral tibial posterior slope, tibial asymmetry, anteroposterior and superior-inferior tubercle offset, and the lateral plateau width (Table 5-12).

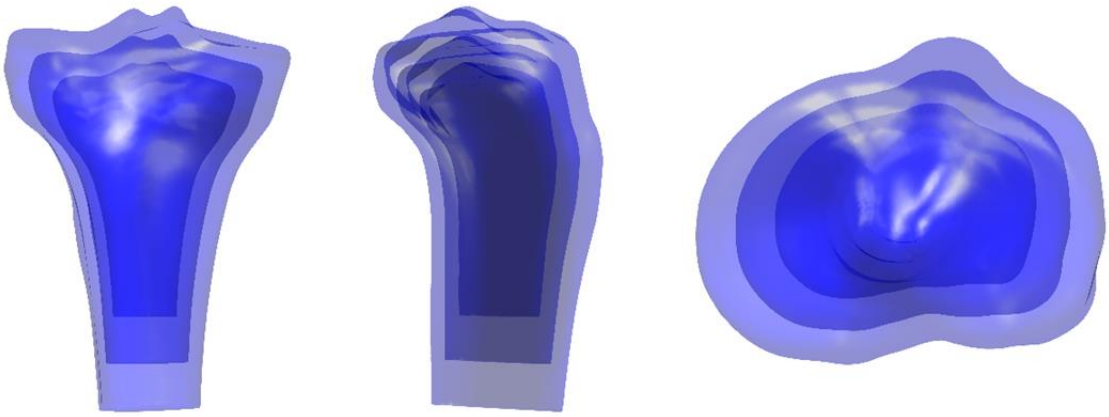

Figure 5-48: Shape variation in Mode 1 of the Tibia SSM. Variation is shown to +/- three standard deviations from the mean.

Figure 5-49: Shape variation in Mode 2 of the Tibia SSM. Variation is shown to +/- three standard deviations from the mean.

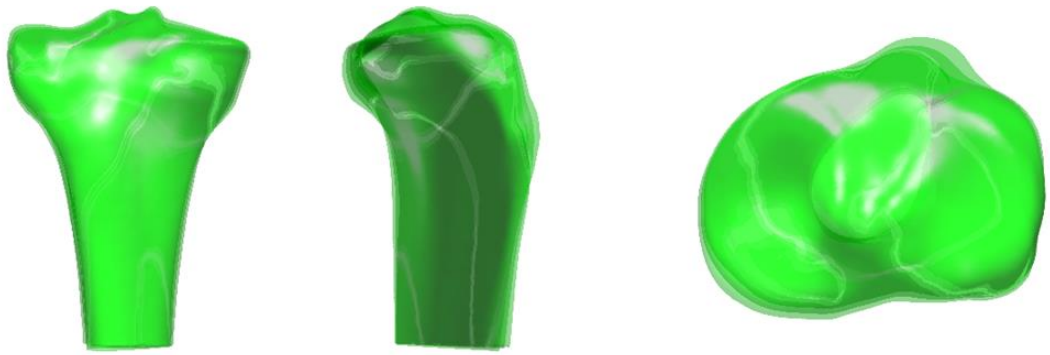

Figure 5-50: Shape variation in Mode 3 of the Tibia SSM. Variation is shown to +/- three standard deviations from the mean.

Figure 5-51: Shape variation in Mode 4 of the Tibia SSM. Variation is shown to +/- three standard deviations from the mean.

Table 5-12: Correlation coefficients (R) between PC scores and morphological parameters. Only statistically significant ($p < 0.05$) correlations are shown. High correlations ($R > 0.65$) are bolded.

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11	PC 12	PC 13	PC 14
tML	**-0.99**													
tAP	**-0.84**		-0.20	0.19							-0.18			
tMAP	**-0.90**													
tLAP	**-0.86**			0.21							-0.19			
MPTS	**0.35**		-0.21		0.22		-0.21	-0.37			0.25			
LPTS	**-0.19**			0.22				-0.27	-0.27					
CS					0.40	-0.29	-0.46							0.19
PO		0.28	-0.36		0.21	-0.24					0.19		0.25	-0.21
AS				-0.27					0.20		0.30	0.35		
TMLO	-0.27		-0.28							-0.20		0.26		
TAPO	**-0.85**	-0.23	-0.19	0.24										
TSIO			-0.40	0.19										
WS	-0.40					-0.32		0.19						0.27
WMTP	**-0.86**													-0.27
WLTP	**-0.64**		-0.18			0.26	0.20	-0.20						

Table 5-13: Correlation coefficients (R) between tibia morphological parameters. Only statistically significant ($p < 0.05$) correlations are shown. High correlations ($R > 0.65$) are bolded.

	tML	tAP	tMAP	tLAP	MPTS	LPTS	CS	PO	AS	TMLO	TAPO	TSIO	WS	WMTP	WLTP
tML															
tAP	**0.82**														
tMAP	**0.88**	**0.82**													
tLAP	**0.84**	**0.83**	**0.83**												
MPTS	-0.33	-0.40	-0.31	-0.32											
LPTS	0.20		0.19	0.23											
CS		-0.22													
PO						-0.23									
AS			0.24	-0.34				0.33							
TMLO	0.27	0.25	0.26	0.24		0.23									
TAPO	**0.82**	**0.81**	**0.77**	**0.78**	-0.22	0.24				0.27					
TSIO								0.22			0.19				
WS	0.40	0.24	0.29	0.24			0.41				0.35				
WMTP	**0.86**	**0.68**	**0.84**	**0.73**	-0.28					0.34	**0.68**		0.18		
WLTP	**0.66**	0.63	0.59	0.64	-0.29		-0.40			0.28	0.55		-0.36	0.55	

5.3.2.4 Interpopulation Comparisons

The sex and racial distribution of the scores pertaining to the first four principal components are depicted in Figure 5-52 (PC1 and PC2) and Figure 5-53 (PC3 and PC4).

To quantify interpopulation differences in PC scores and morphological parameters, unpaired two-sample t-tests and a series of ANOVA tests were conducted. Table 5-14 displays the p-values resulting from t-tests with significant differences ($p < 0.05$) between two populations. Boxplots illustrating the distribution of PCs one through four and select morphological populations across sex, race, and age categories are shown in Figures 5-49 to 5-75. Significant differences ($p < 0.05$) in these plots are indicated by a red bar. The sex and racial distribution of the anteroposterior and mediolateral lengths of the proximal tibial plateau are shown in Figure 5-81. Table 5-15 displays significant p values ($p < 0.05$) resulting from ANOVA tests for each PC or morphological parameter considering sex, race, age, height, and BMI.

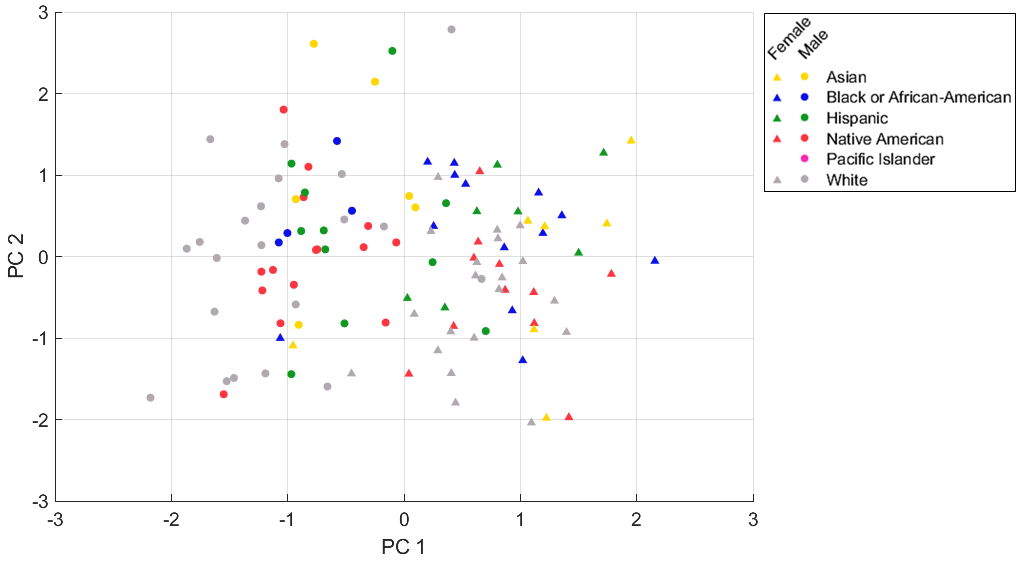

Figure 5-52: Sex and racial distribution of PC 1 and PC 2 scores of the Tibia SSM.

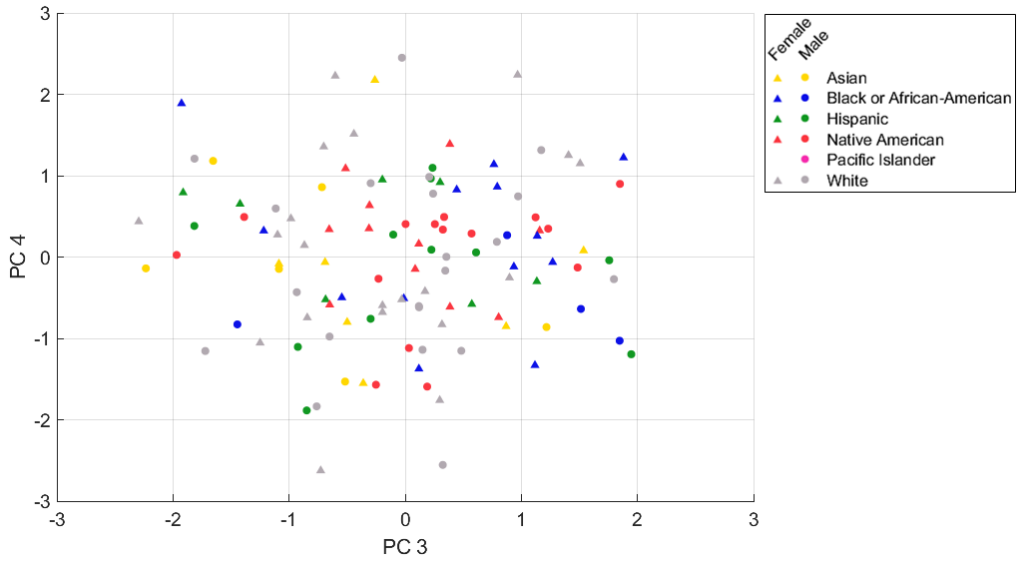

Figure 5-53: Sex and racial distribution of PC 3 and PC 4 scores of the Tibia SSM.

Table 5-14: P-values derived from unpaired two-sample t- tests indicating significant differences ($p < 0.05$) between populations. Highly significant differences ($p < 0.001$) are shown in bold.

	M/F	A/B	A/H	A/NA	A/W	B/H	B/NA	B/W	H/NA	H/W	NA/W	YA/MA	YA/OA	MA/OA
PC 1	**3.10E-25**							4.05E-02						
PC 2	3.08E-02						4.28E-02	4.04E-02						
PC 3		4.40E-02										2.60E-02		
PC 4														
PC 5		3.18E-02												
PC 6														
PC 7									4.04E-02					
PC 8			6.17E-03		2.33E-03				3.05E-02		2.81E-02		3.72E-02	
PC 9														
PC 10										3.50E-02	1.07E-02			
PC 11							5.37E-03	**4.10E-04**						
PC 12					8.19E-03			4.73E-03	2.83E-02		**2.12E-05**		2.07E-03	1.49E-03
PC 13											1.80E-02			
PC 14		3.20E-02					2.77E-02							
tML	**5.64E-26**							4.91E-02						
tAP	**1.32E-12**							3.94E-02						
tMAP	**1.80E-15**				4.43E-02		4.59E-02	3.15E-02						
tLAP	**1.00E-12**													
MPTS	**7.86E-05**													3.25E-02
LPTS				3.66E-02			2.89E-03				1.90E-03			
CS									4.02E-02					4.24E-02
PO						4.61E-02	3.52E-02							
AS						2.54E-02		3.84E-02						
TMLO													2.48E-02	
TAPO	**1.87E-12**				2.19E-02			6.98E-03						
TSIO		3.55E-02						**7.02E-05**		3.36E-02	3.15E-02		5.72E-03	
WS	**3.94E-04**				2.49E-02						4.30E-02			
WMTP	**1.38E-16**							4.58E-02						
WLTP	**8.93E-10**						4.07E-02							

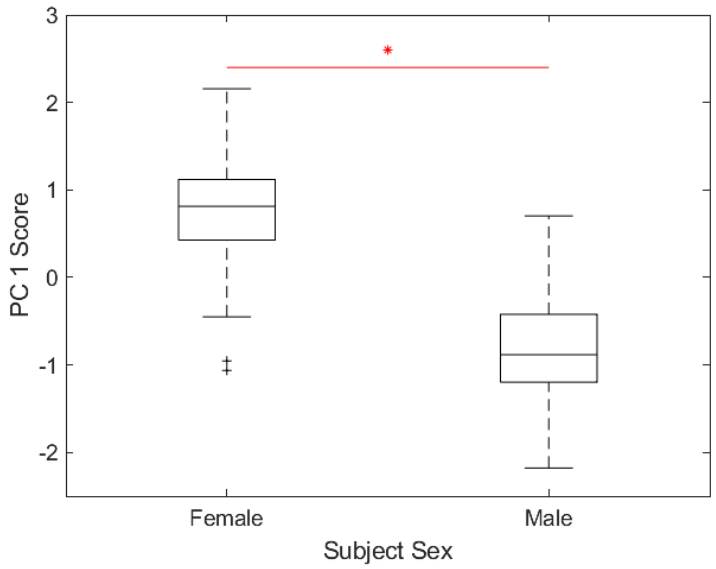

Figure 5-54: Tibia SSM PC 1 score by subject sex.

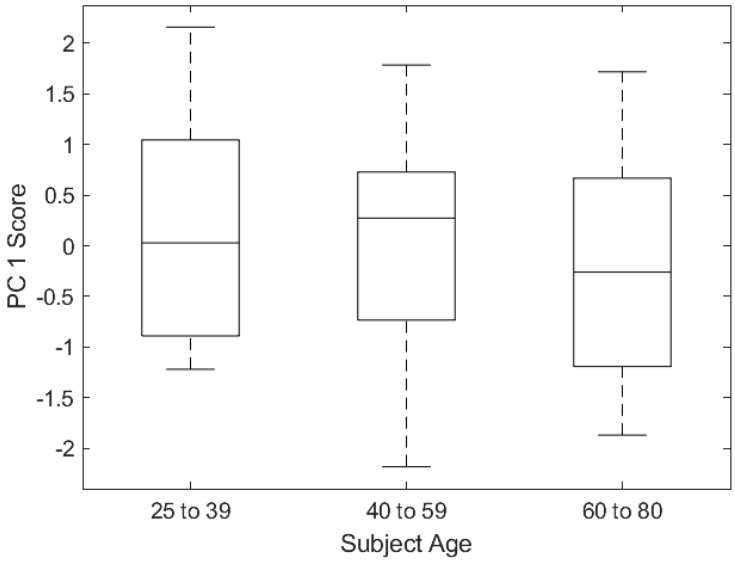

Figure 5-55: Tibia SSM PC 1 score by subject age.

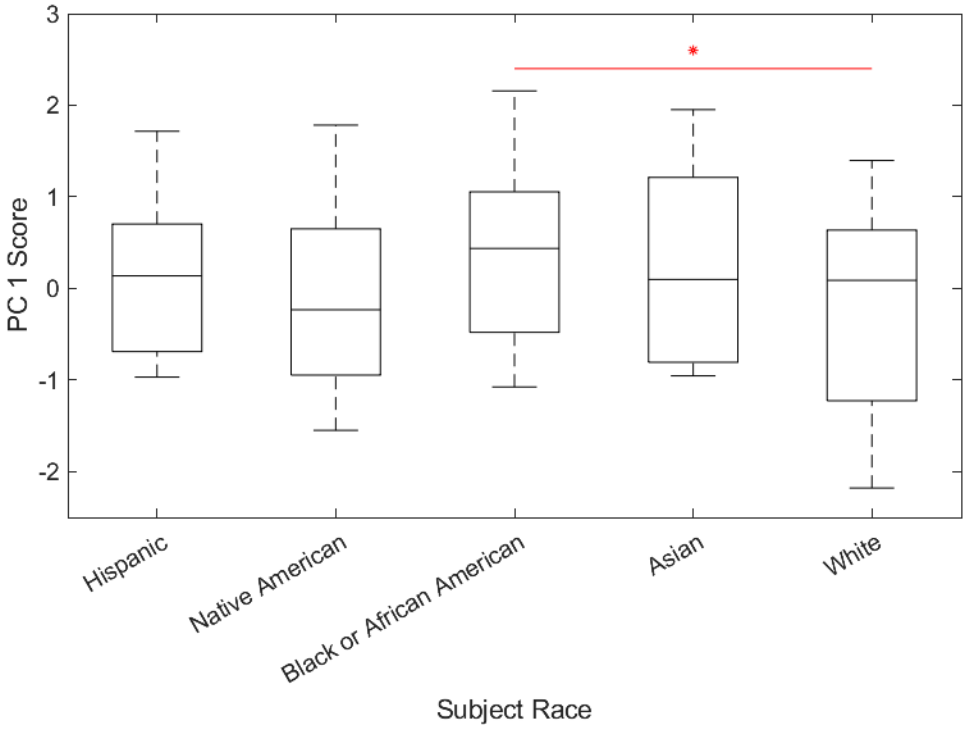

Figure 5-56: Tibia SSM PC 1 score by subject race.

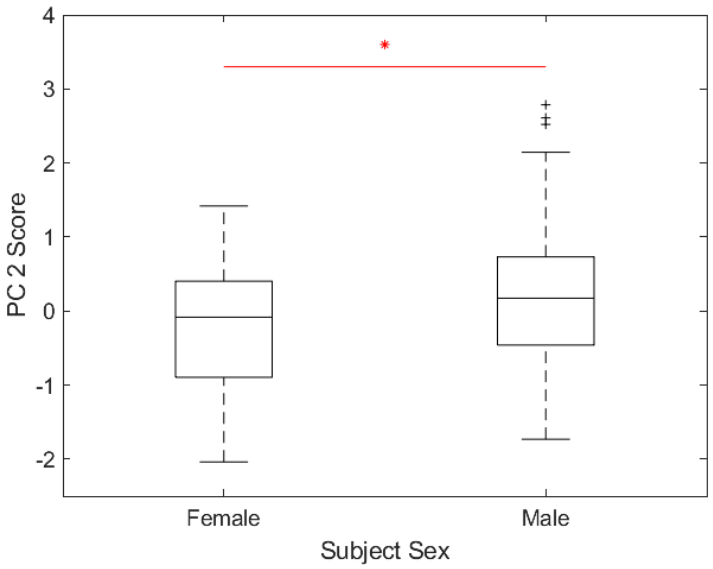

Figure 5-57: Tibia SSM PC 2 score by subject sex.

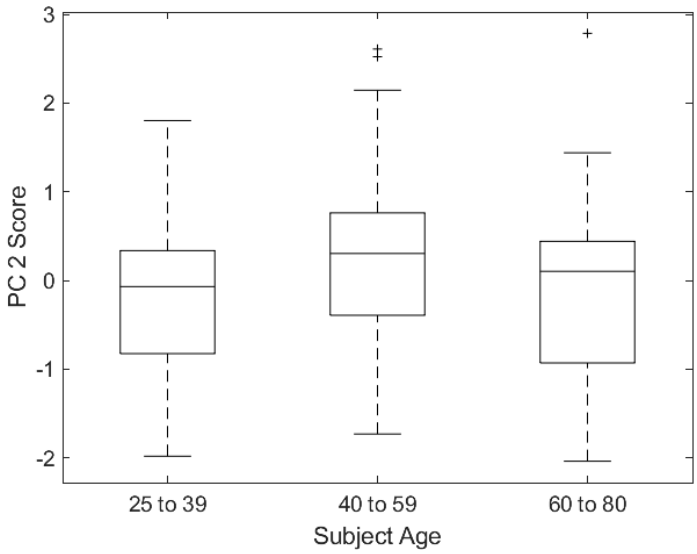

Figure 5-58: Tibia SSM PC 2 score by subject age.

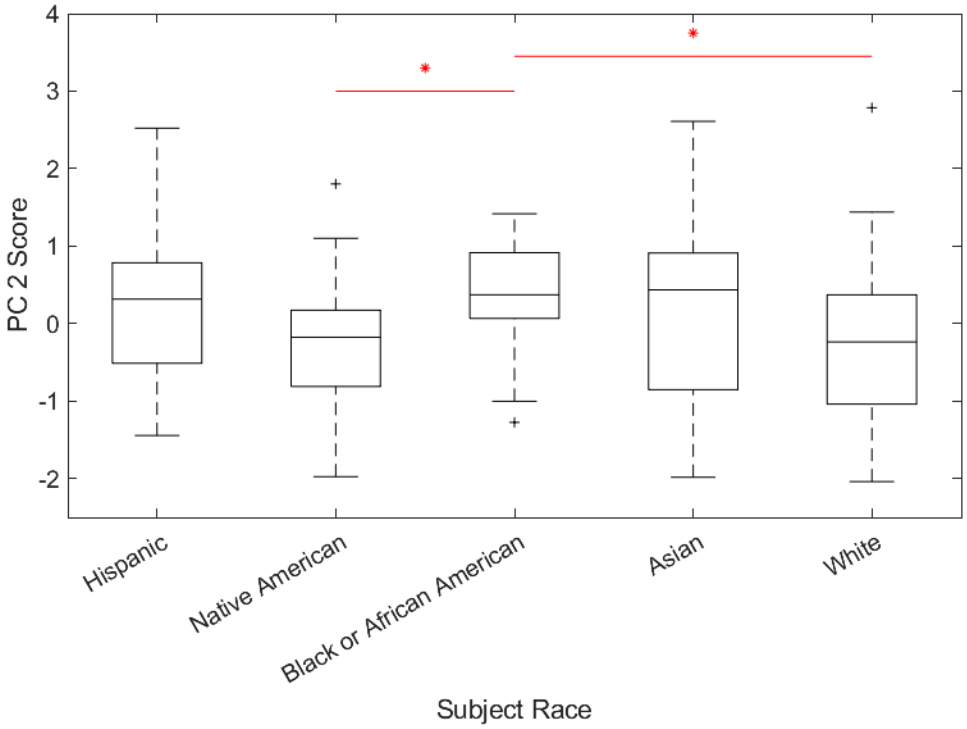

Figure 5-59: Tibia SSM PC 2 score by subject race.

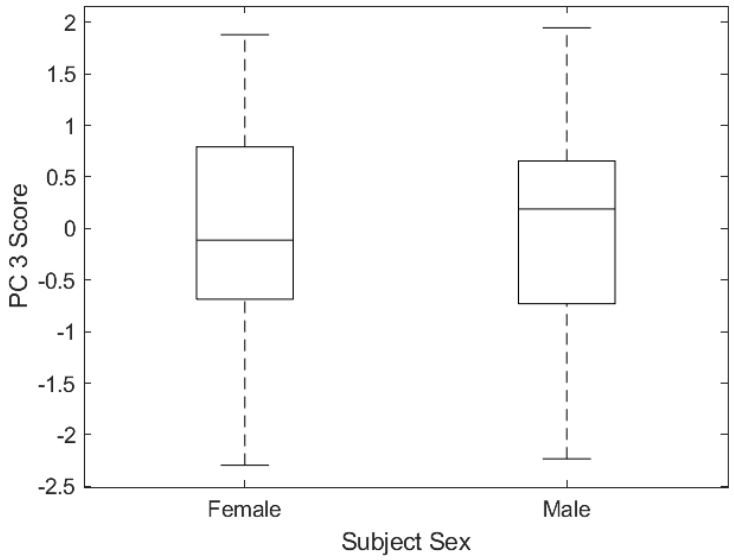

Figure 5-60: Tibia SSM PC 3 score by subject sex.

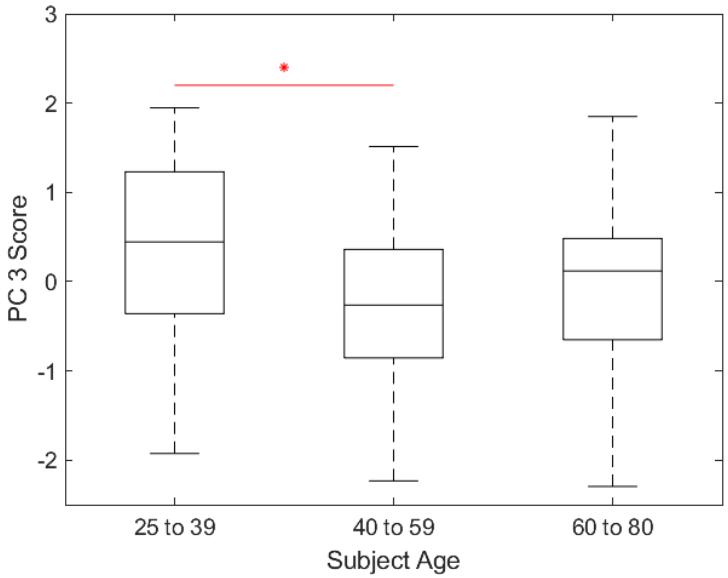

Figure 5-61: Tibia SSM PC 3 score by subject age.

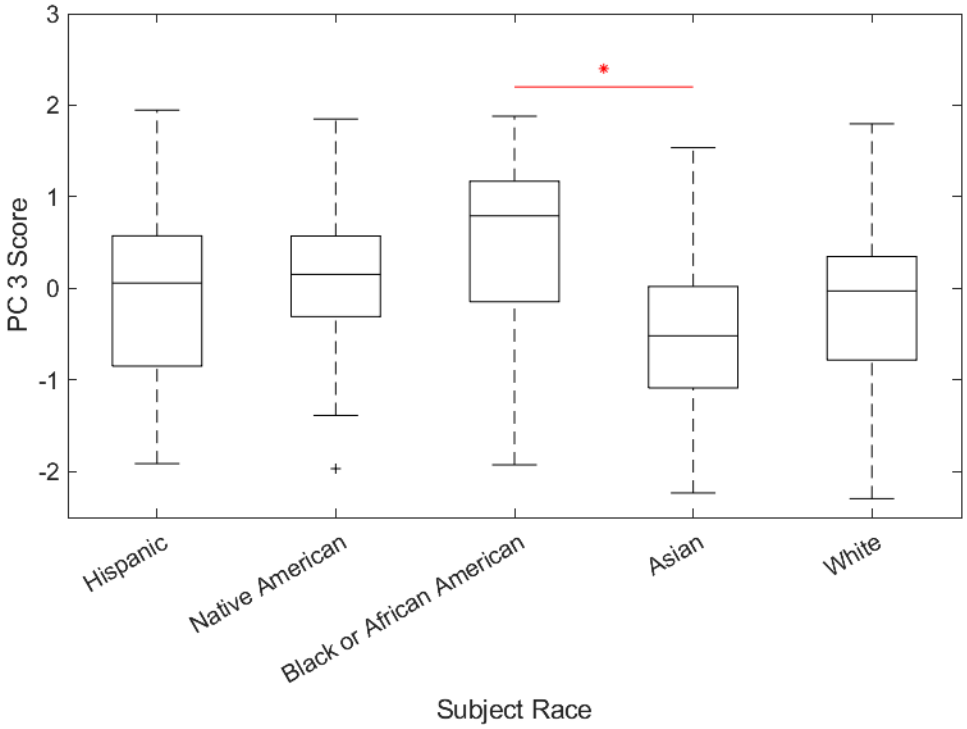

Figure 5-62: Tibia SSM PC 3 score by subject race.

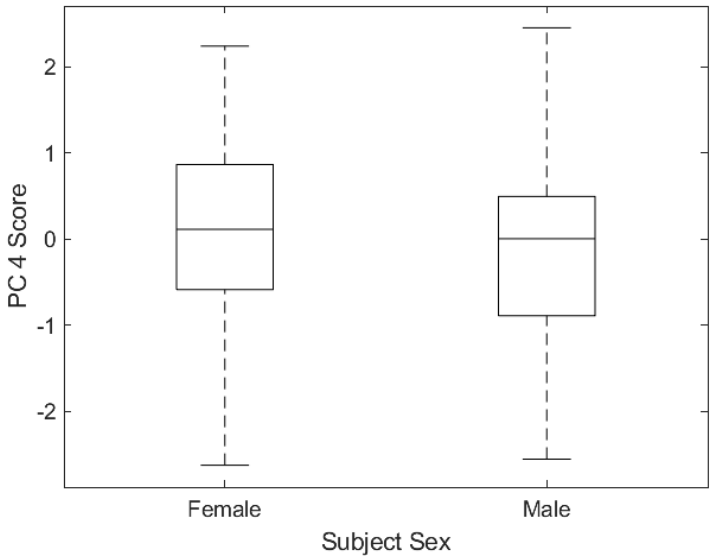

Figure 5-63: Tibia SSM PC 4 score by subject sex.

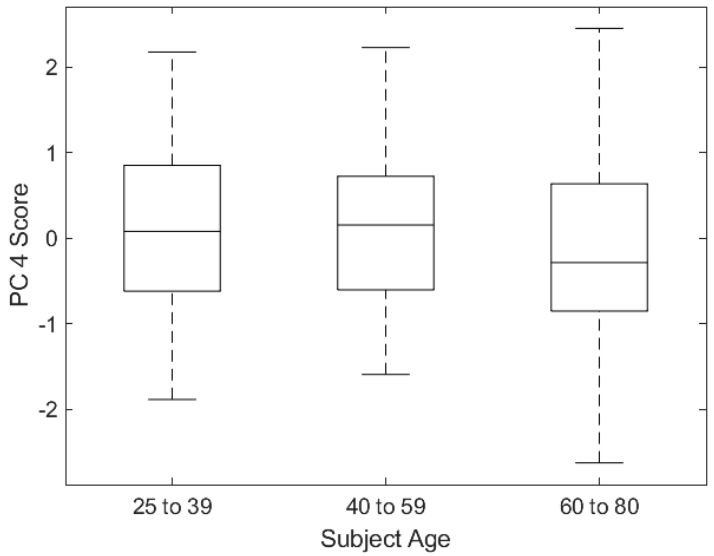

Figure 5-64: Tibia SSM PC 4 score by subject age.

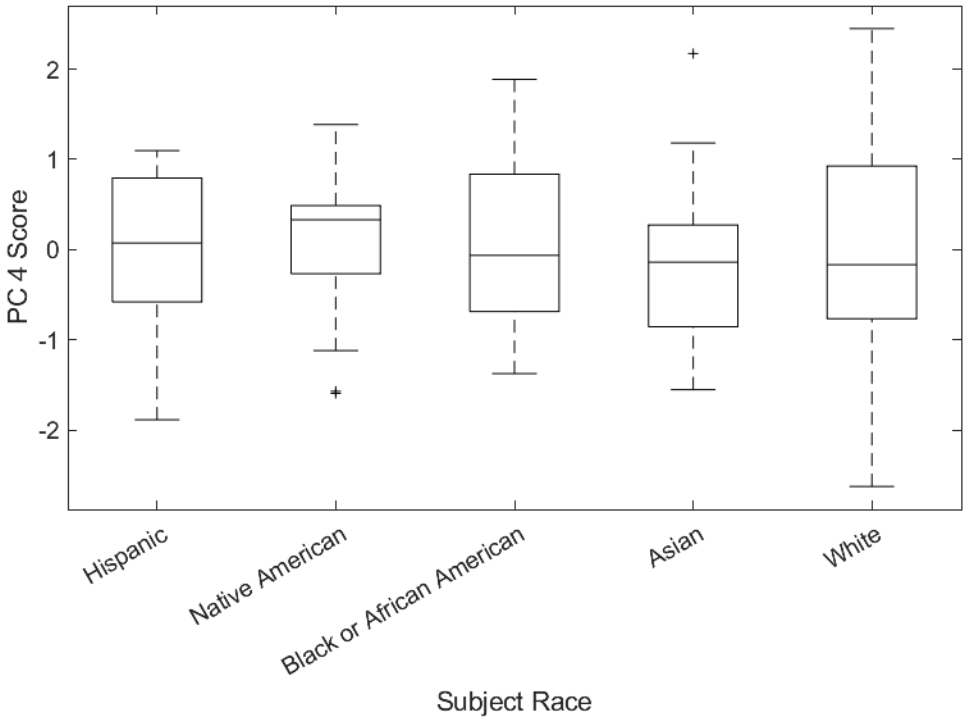

Figure 5-65: Tibia SSM PC 4 score by subject race.

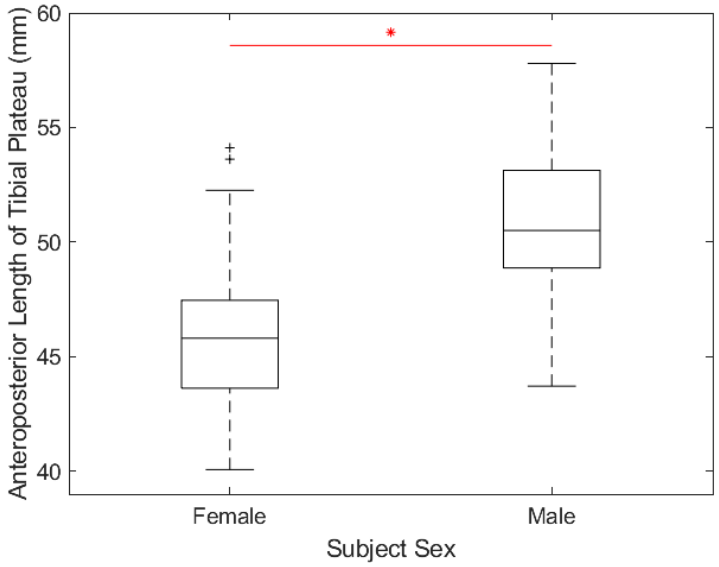

Figure 5-66: Anteroposterior length of the tibial plateau by subject sex.

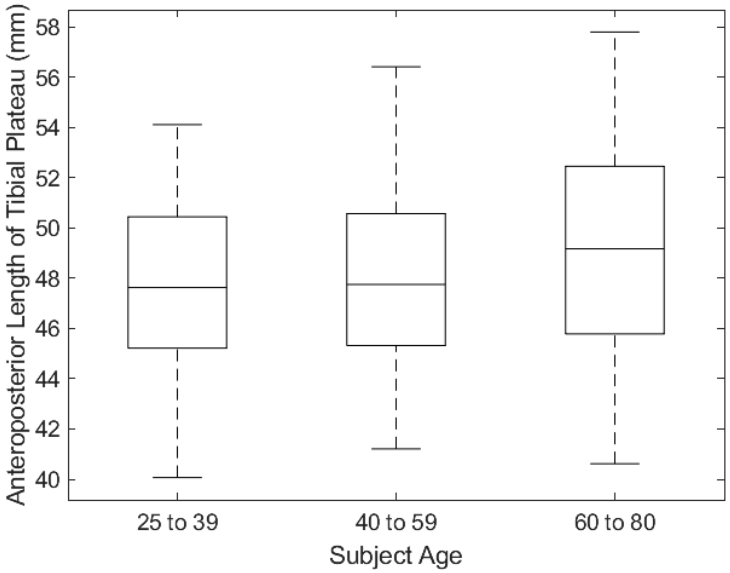

Figure 5-67: Anteroposterior length of the tibial plateau by subject age.

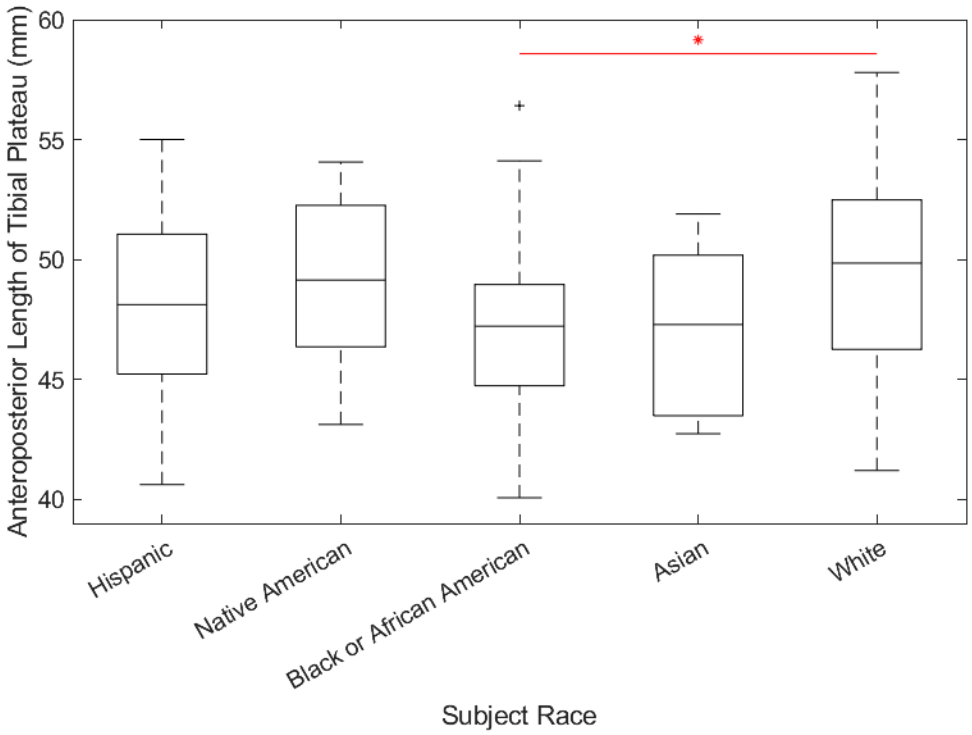

Figure 5-68: Anteroposterior length of the tibial plateau by subject race.

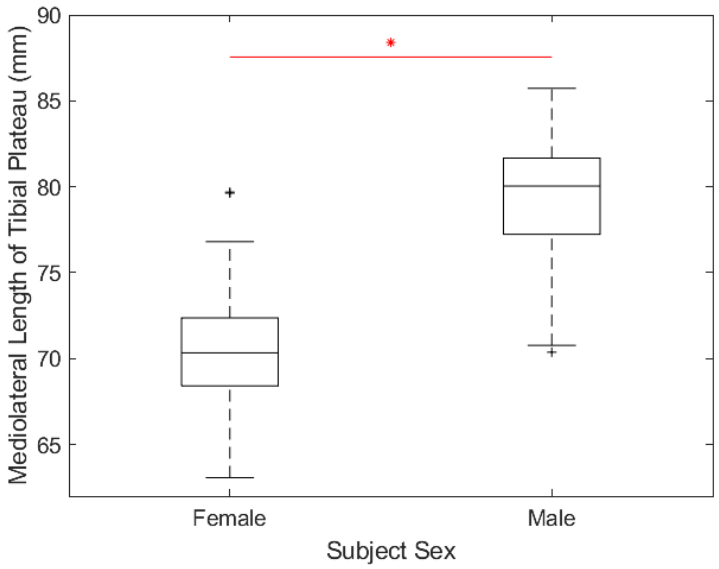

Figure 5-69: Mediolateral length of the tibial plateau by subject sex.

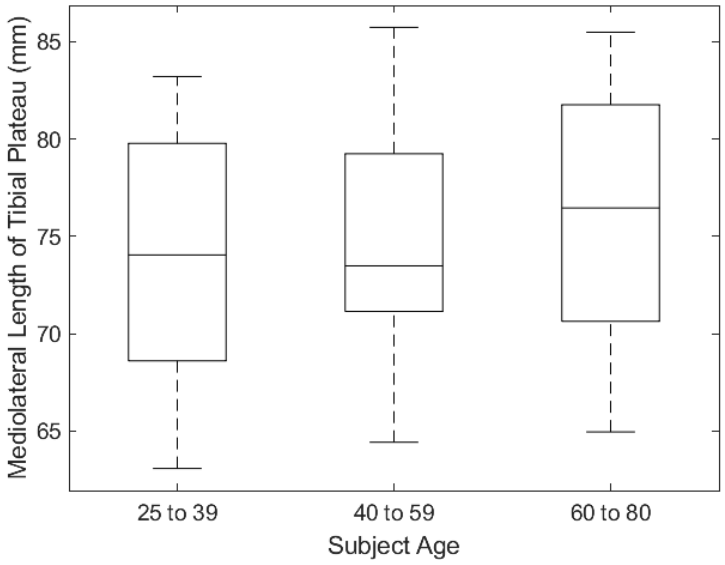

Figure 5-70: Mediolateral length of the tibial plateau by subject age.

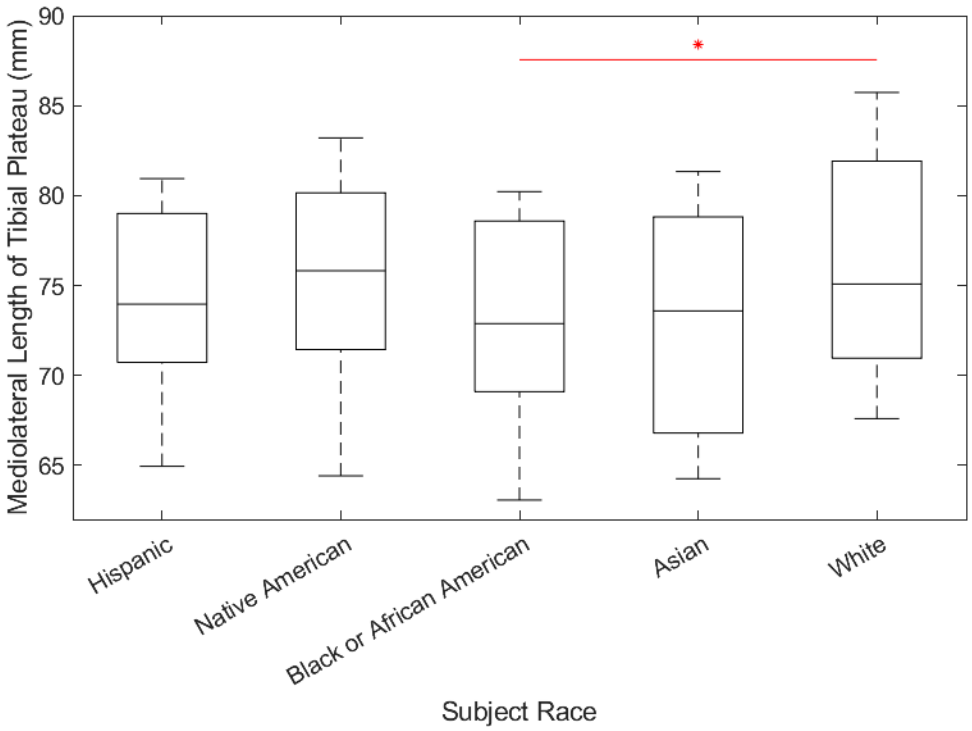

Figure 5-71: Mediolateral length of the tibial plateau by subject race.

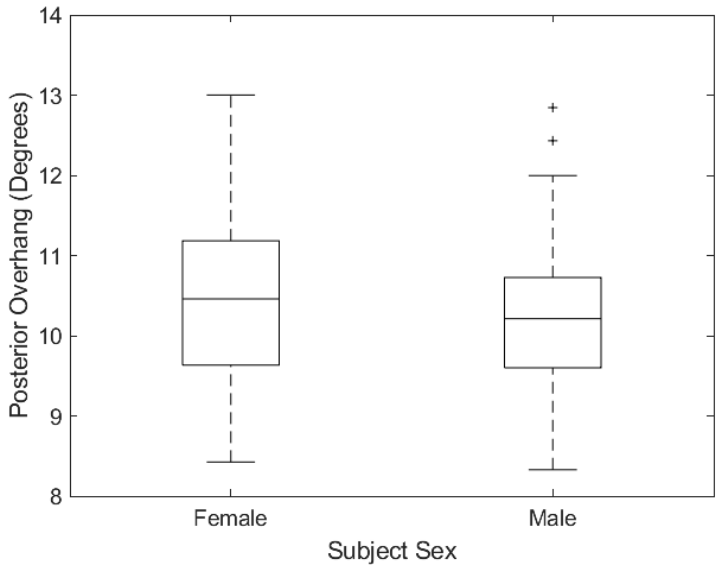

Figure 5-72: Posterior overhang (in degrees) by subject sex.

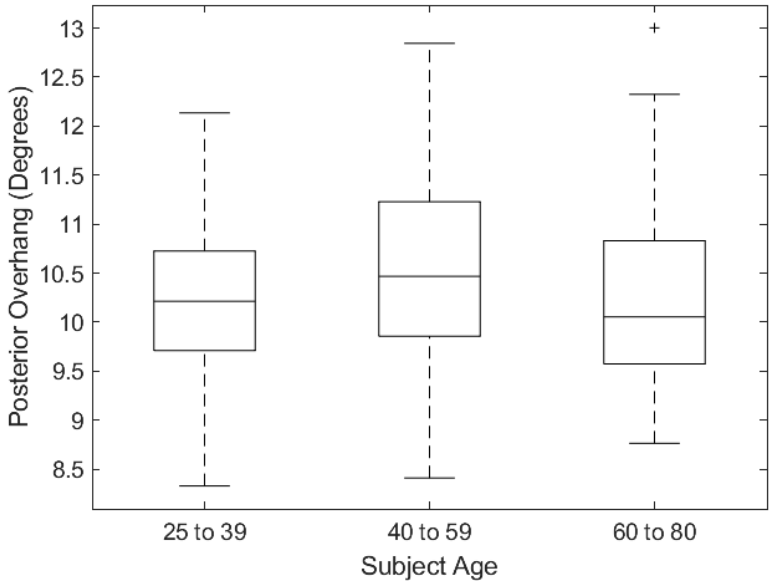

Figure 5-73: Posterior overhang (in degrees) by subject age.

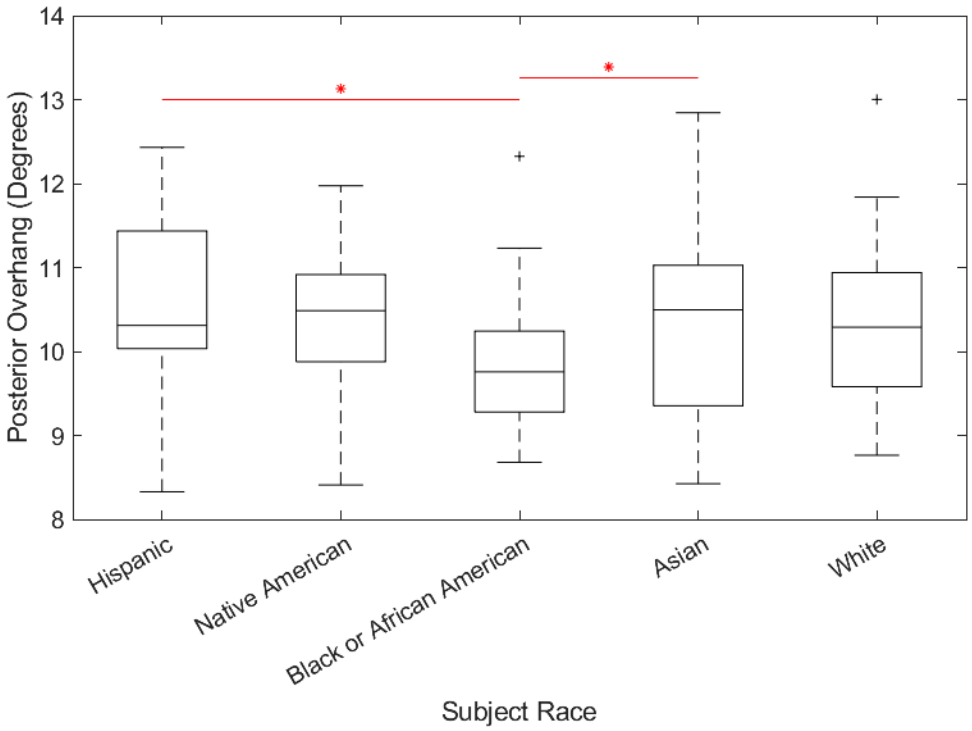

Figure 5-74: Posterior overhang (in degrees) by subject race.

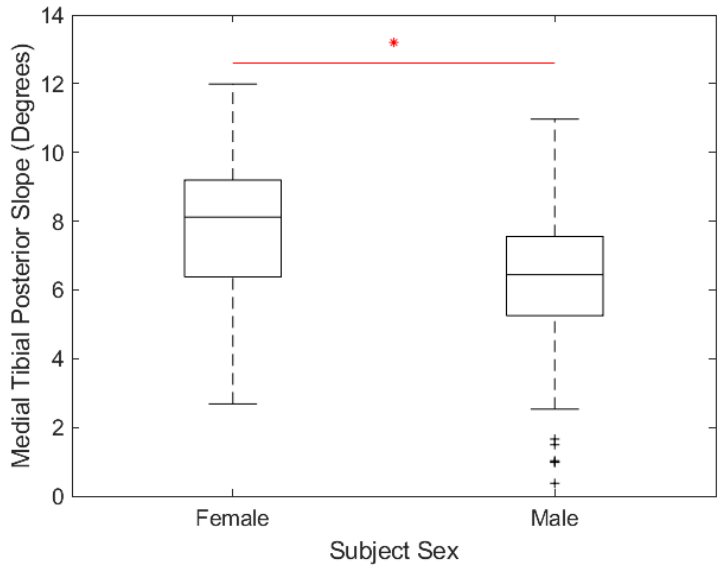

Figure 5-75: Medial tibial posterior slope by subject sex.

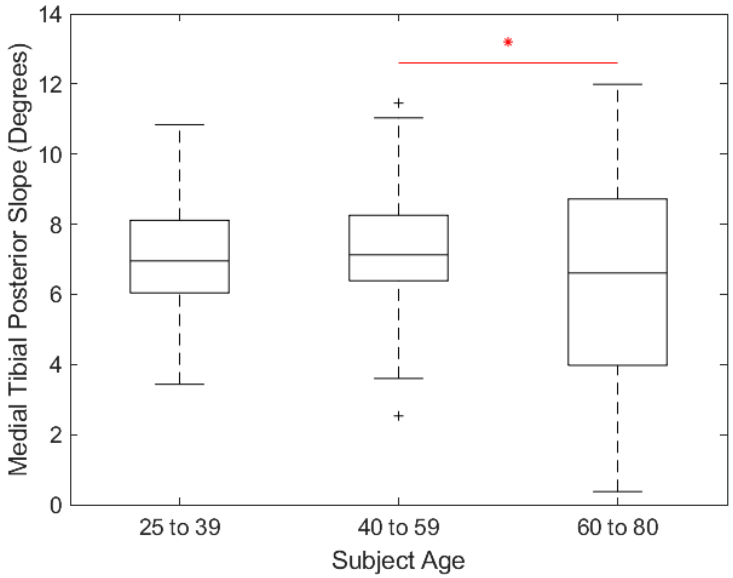

Figure 5-76: Medial tibial posterior slope by subject age.

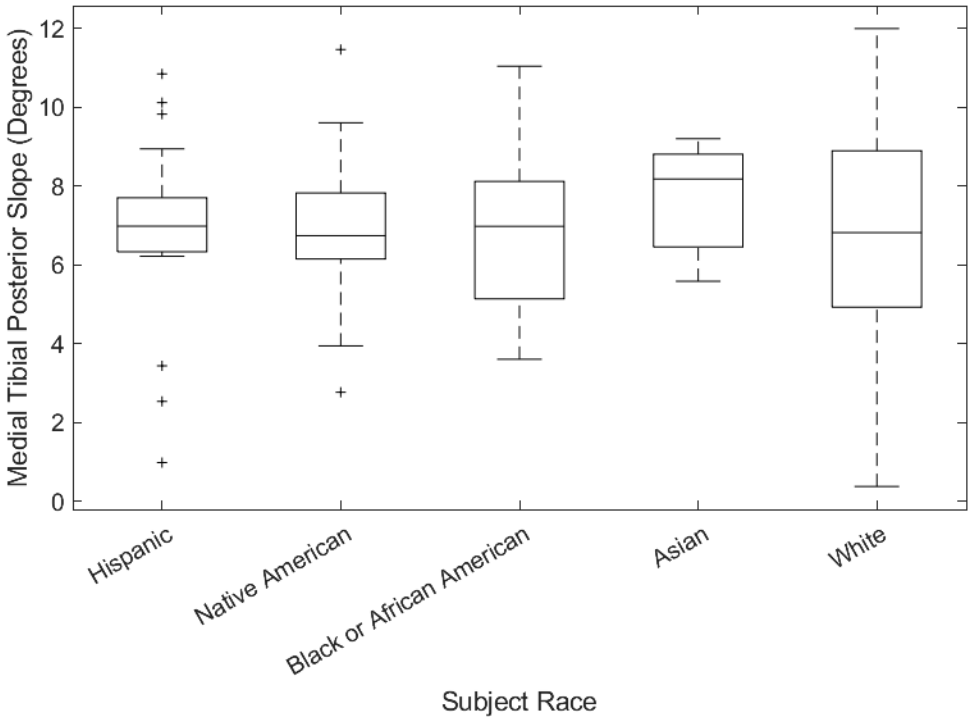

Figure 5-77: Medial tibial posterior slope by subject race.

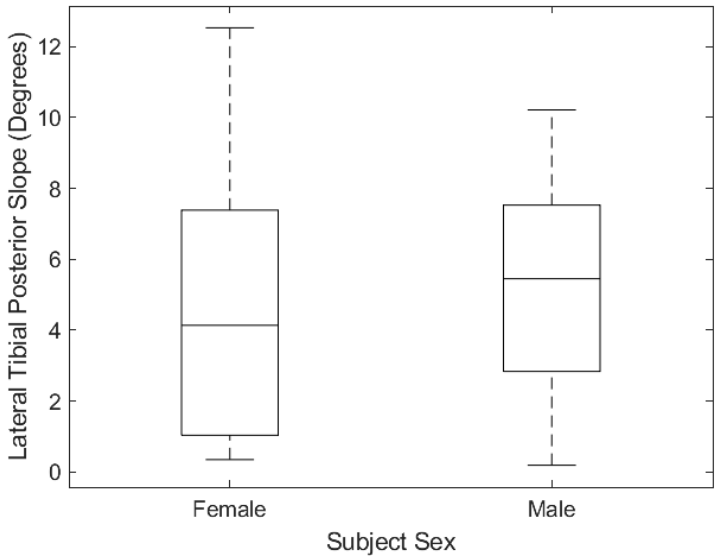

Figure 5-78: Lateral tibial posterior slope by subject sex.

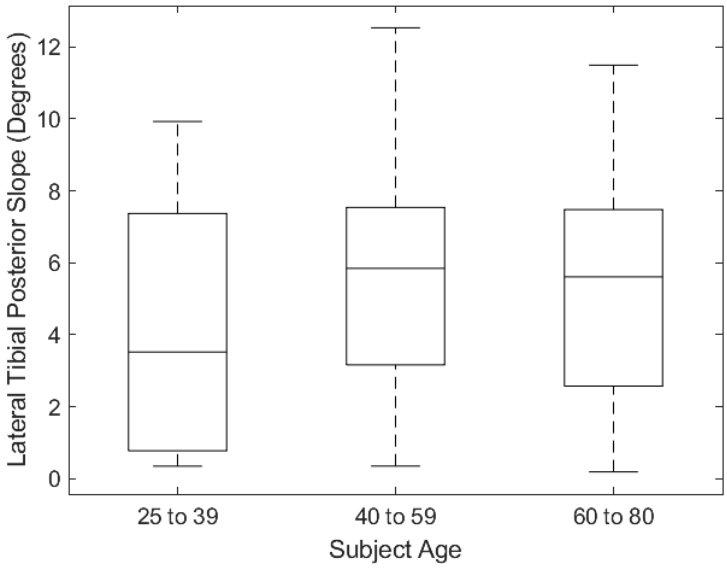

Figure 5-79: Lateral tibial posterior slope by subject age.

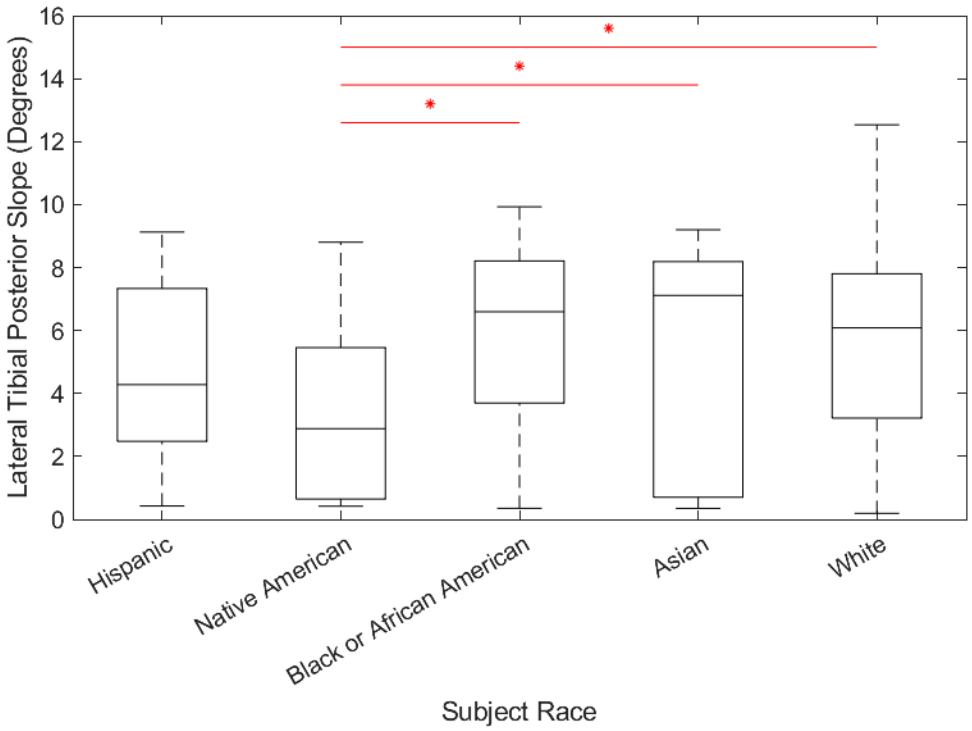

Figure 5-80: Lateral tibial posterior slope by subject race.

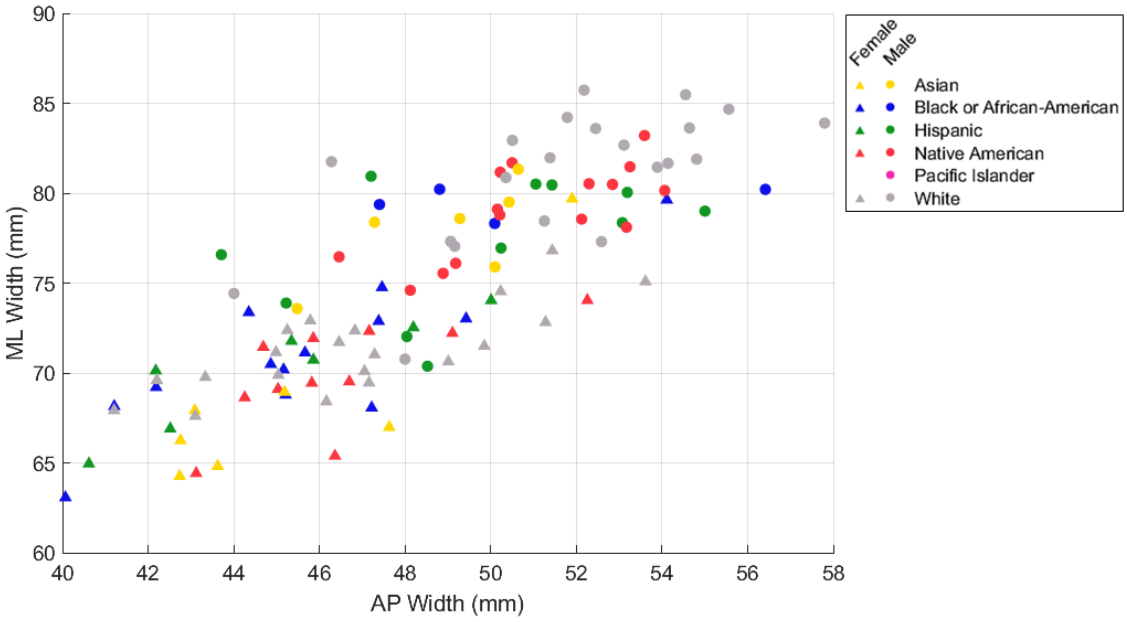

Figure 5-81: Sex and racial distribution of the anteroposterior and mediolaterial lengths of the proximal tibia plateau.

Table 5-15: ANOVA-derived p-values derived indicating significant differences ($p < 0.05$) between populations. Highly significant differences ($p < 0.001$) are shown in bold.

	Sex	Age	Race	Height	BMI
PC 1	**3.61E-14**			**3.27E-07**	
PC 2	**2.20E-06**		2.55E-02	**3.30E-05**	
PC 3					
PC 4					
PC 5					
PC 6					
PC 7					
PC 8			6.43E-03		
PC 9					
PC 10			2.55E-02		
PC 11			4.47E-03		
PC 12			3.08E-02		
PC 13					
PC 14					
tML	**2.33E-15**			**8.06E-06**	
tAP	**4.18E-04**			**1.17E-06**	
tMAP	**5.07E-06**			**1.26E-05**	
tLAP	**1.13E-04**			**1.77E-05**	
MPTS	3.83E-02				
LPTS			1.59E-02		
CS					
PO				3.56E-02	
AS					
TMLO		1.86E-02		4.02E-03	
TAPO	**8.67E-05**			**1.96E-04**	
TSIO					
WS	2.67E-03				
WMTP	**2.08E-07**			**1.65E-04**	
WLTP	**4.94E-04**			4.27E-03	

5.4 Discussion

Prior to discussing the outcomes of the model, it is important to first consider: "Is this model a valid representation of the training set population?" This question is addressed using four metrics: accuracy, compactness, generalizability, and specificity.

Both the femur and tibia models successfully recreated subjects with a clinically relevant submillimeter (image resolution size) distance error of 0.6 mm, using five (Table 5-3) and four (Table 6-3) principal components, respectively. Additionally, both models demonstrated compactness (Figure 5-7 and Figure 6-6), requiring seven (femur) and fourteen (tibia) principal components to encompass 95% of the variance in the data. For both models, the first and most dominant principal component (PC) represented size, explaining 86.10% (femur) and 82.06% (tibia) of the variance in the training set. This finding is in line with those of previous SSMs (Audenaert et al. 2019; Fugit, Bayoglu, and Laz (in review)). Subsequent PCs explained decreasing amounts of variance, with no single PC accounting for more than 4% of the variance in either the Femur (Table 5-3) or Tibia (Table 5-10) SSM.

Generalization accuracy was assessed through a leave-one-out analysis, which evaluated how well the model represented other unseen bones in the same class. With 111 femur and 115 tibia subjects, the generalization accuracy for the Femur SSM (0.33 ± 0.06 mm) and Tibia SSM (0.34 ± 0.04 mm) was below the 0.6 mm threshold. However, it should be noted that the number of subjects included in each model is below the 200 subjects required for convergence of the generalizability (Audenaert et al. 2019), and this is reflected in the lack of clear convergence of the reconstruction error even when the full training set was included (Figure 5-8 and Figure 5-47). This indicates that while our model is generalizable within clinically relevant accuracy, the inclusion of more subjects is necessary to sufficiently capture the true population variance, particularly given the

diverse populations included in our model. Both models also demonstrated the ability to generate realistic new bone instances, as evidenced by sub-millimeter specificity (

Table 5-4 and Table 5-11).

T-tests revealed significant size differences (PC 1) between the sexes in both the tibia ($p = 3.10E-25$) and femur ($p = 5.06E-21$). This agreed with the results of the ANOVA ($p = 1.08E-11$ and $p = 3.61E-14$ for the femur and tibia, respectively) and was consistent with the findings of Audenaert et al. (Audenaert et al. 2019). Size showed strong correlations with morphometric distance parameters in both the femur (Table 5-5) and tibia (Table 5-12). Consequently, it is unsurprising that significant sex differences were also found in these parameters using both t-tests (Table 5-7 and Table 5-14) and ANOVA (Table 5-8 and Table 5-15), such as TEA of the femur (Figure 5-27) and the AP (Figure 5-66) and ML (Figure 5-69) lengths of the tibial plateau.

Differences between racial and ethnic groups, while present, were less pronounced. Scatter plots for the PC scores, Figure 5-13 and Figure 5-14 for the femur and Figure 5-52 and Figure 5-53 for the tibia did not show strong trends with race. In the femur, t-tests revealed significant ($p < 0.05$) size differences between Asian subjects and White subjects, Blacks or African American subjects and White subjects, Asian subjects and Native American subjects, and Blacks or African American subjects and Native Americans subjects in the femur (Figure 5-17). These size differences translated into differences in morphometric parameters, with the TEA length differing significantly between these groups (Figure 5-29). In contrast to the t-tests, ANOVA did not reveal size-based differences between racial groups (Table 5-8 and Table 5-15); the differences

in the findings of the statistical tests are likely influenced by the small sample sizes and differences in sex composition in the subgroups. The ANOVA results agreed with the findings of Mahfouz et al., who found no significant difference in TEA length between Asian subjects and White subjects or Black or African American subjects and White subjects (Mahfouz et al. 2012). T-tests also identified differences between Black or African American and White subjects in APH, MAP, and PML (Table 5-7), which were not shown by ANOVA or by Mahfouz et al. (2012). However, differences in tibial metrics tML and tAP found by Mahfouz et al. were confirmed by t-tests (Table 6-6), as were femoral metric differences between Asians and Whites (APH, MAP, LAP, and PML) (Mahfouz et al. 2012). ANOVA found no racial differences between these or any other distance morphometric parameters in either the femur (Table 5-8) or tibia (Table 5-15).

T-tests showed statistically significant differences in parameters related to the radii of curvature of the femoral condyles across sex and racial and ethnic groups, particularly between Asian and White subjects and Black or African American and White subjects (

Table 5-6). These measures are important for implant development and sizing, suggesting significant implications for the femoral component of TKA. Mahfouz et al. also found curvature differences between these populations, although their method for defining curvature differed from ours, preventing a direct comparison (Mahfouz et al. 2012). While ANOVA did not identify differences in the same radii of curvature as the t-tests, significant inter-group racial differences were found in the horizontal radius of the

medial femur condyle and in the angle of divergence of the medial femur epicondyle (Table 5-8).

In the tibia, t-tests identified significant differences ($p < 0.05$) in posterior overhang were found between Black or African American subjects and Hispanic and Asian populations, respectively (Figure 5-74), which were not confirmed using ANOVA (Table 5-15). While the medial tibial posterior slope showed no differences between racial groups (Figure 5-77), Native American subjects exhibited significant differences from Black and African American, Asian, and White subjects in the angle of their lateral tibial posterior slope (Figure 5-80). Inter-race differences were also found in this metric using ANOVA (Table 5-15). Since the tibial plateau slope guides the angle of the resection cut for the tibial component in TKA, this finding may be particularly noteworthy.

While bone size changes with age are well documented (Riggs et al. 2004), t-tests found no significant age differences in PC 1 of either the femur (Figure 5-16) or the tibia (Figure 5-55). Conversely, ANOVA did find that PC 1 of the Femur varied significantly ($p < 0.05$) with age (Table 5-8). Age-related t-test differences were observed in Femur PC 2, the width of the femoral condyles, and the posterior radius of the femoral condyle, but not in the transepicondylar axis or the posterior condylar angle as found by Han et al. (Han et al. 2016). In contrast, ANOVA identified significant age-related differences with age in TEA, lateral condylar width, the coronal radius of the medial femoral condyle, and the posterior radius of the lateral femoral condyle (Table 5-8). In the tibia, t-test differences were found in the medial posterior tibial slope between middle age and old age, but not in the lateral posterior tibial slope (Table 5-14). Conversely, Han et al. found

differences between these same groups in the lateral posterior tibial slope rather than the medial side (Han et al. 2016). ANOVA identified only significant age-related differences in the tubercle medial lateral offset of the tibia (Table 5-15).

Our study has several limitations. Our training set includes a broader range of demographics than most studies of this type and is the only known model to the authors that includes Native American and Hispanic subjects. However, when considering the US population, it neglects Native Hawaiian/Pacific Islander and ethnically middle eastern populations. It is also crucial to recognize that racial identities are socially constructed and lack a biological basis (Cwalina et al. 2022). Without including detailed genetic data from rigorous genomic testing or accounting for potential confounding factors such as BMI or socioeconomic status, our results cannot definitively attribute the observed differences in our models to race. Nevertheless, race remains a significant determinant of access to societal resources and barriers to full inclusion, as with the presence of racial and ethnic disparities in TKA. The fact of inequality means that race and ethnicity continue to matter in important ways (Committee on Improving the Representation of Women and Underrepresented Minorities in Clinical Trials and Research 2022).

Furthermore, while we observed statistically significant differences in the size and morphology of the distal femur and proximal tibia, our intra-group sample sizes average about 20 subjects per racial background, which is insufficient to fully capture the variation within each demographic. In the future, we hope to investigate differences in the knee with a larger training set to comprehensively capture population variation. Nonetheless, this work represents a step forward in incorporating demographic diversity

into the development and validation of orthopedic devices, ensuring that treatments like TKA are effective across varied populations.

6. Material Property Variation in the Knee Considering Diverse Populations

6.1 Introduction

Total Knee Arthroplasty (TKA) is a widely performed surgical procedure aimed at alleviating pain and restoring function in patients with severe knee osteoarthritis (Varacallo, Luo, and Johanson, n.d.). Despite its general success, disparities in postoperative outcomes have been observed across different racial and ethnic groups, with minority populations often experiencing less favorable results (Shahid and Singh 2016). A major cause for TKA revision is mechanical loosening or loss of implant fixation (Bozic et al. 2010). Early migration of the implant is a strong predictor of clinical failure (Albrektsson et al. 1981), often leading to revisions due to mechanical instability (Pijls et al. 2012). One crucial factor linked to this early migration is lower bone quality (Anderson et al. 2020).

Emerging evidence points to differences in bone mineral density (BMD) across racial and ethnic groups; for instance, non-Hispanic Black adults typically exhibit higher BMD and lower rates of bone loss compared to other ethnicities (Noel, Santos, and Wright 2021). Conversely, US Asian adults may have lower BMD than non-Hispanic White adults, although the data on this remains limited and sometimes inconclusive. The directionality of these BMD differences also appears to vary by skeletal site. For example, Mexican American adults show higher mean BMD of the proximal femur

compared to non-Hispanic White adults, despite having a lower mean BMD for the total body and other sites (Noel, Santos, and Wright 2021).

Although age-, site-, and sex-specific high-resolution peripheral quantitative computed tomography (HR-pQCT) reference data for bone density in various populations has been developed (Koy et al. 2022; Alvarenga et al. 2022; Yu et al. 2020; Hung et al. 2015), significant population gaps remain. Further, there has been limited work leveraging statistical models to explore material property variation in the bones of the knee across ethnic and racial groups, despite the noted suitability of these methods for the task (Bah et al. 2015). To the author's knowledge, no statistical intensity models of the knee have been created from a racially and ethnically diverse training set, nor has any such model been leveraged to investigate the material properties of the tibia and femur across these groups.

Accordingly, the objective of this work is to develop a statistical intensity model of the knee using a diverse training set. This model aims to investigate differences in tibia and femur bone quality across sex, racial and ethnic, and age groups. A secondary aim is to explore the relationship between bone quality and shape, providing a comprehensive understanding of how these factors interplay in the bones of the knee.

6.2 Methods

6.2.1 Subject Selection

The New Mexico Decedent Image Database (NMDID) is a comprehensive dataset containing whole-body CT scans and associated metadata for over 15,000 individuals who died in New Mexico between 2010 and 2017 (Edgar et al. 2020). This dataset offers

an invaluable resource for accessing a diverse population of scans. For this work, an effort was made to gather a sample representative of the United States population (Barton et al., n.d.), including equal numbers of males and females of various ages from Black or African American, Native American, Asian, and Hispanic backgrounds. A sample of White subjects were obtained from the training set of an SSIM previously developed at the University of Denver (Bayoglu et al. 2020). These race and ethnicity metrics were self-reported by the decedents in the 2010 census. It is important to note that for the purposes of this work, "Hispanic" refers to decedents who identified as racially Hispanic, reflecting the usage in the 2010 census where many New Mexicans selected Hispanic as an "Other" race option (Edgar et al. 2020). The inclusion criteria for this study required subjects to be aged between 25 and 80 years, with a BMI of less than 40, based on cadaver height and weight. Subjects were excluded if they had a recorded history of cancers, tumors, or malignancies, any recorded injuries to the lower limbs, a manner of death that could result in lower limb injuries, cadaver decomposition, or severe osteoarthritis.

6.2.2 Bone Model Generation

Each subject file downloaded from the NMDID database contained 26 CT scans. Among these, three scans held importance for this work: the “thin bone torso” and “thin bone lower extremity” scans, utilized to acquire subject bony geometry, and the “thin soft-tissue lower extremity” scans, used to obtain the subject bony material properties. Before processing, a preliminary assessment of each subject's respective scans was conducted to check for factors that may render the subject unsuitable for our model’s

purposes, including the presence of moderate to severe osteoarthritis in the knee joints, intraosseous cannulae (typically in the proximal tibia) or orthopedic implants in the lower extremity under consideration, or scan artifacts within regions of interest. Both of the subject's lower extremities were considered for use. Only one leg was included per subject. However, if neither leg proved appropriate for use, the subject was subsequently excluded from further analysis.

Pre-processing started by calibrating the "thin soft-tissue lower extremity" scan. Since phantoms are not included in scans from the NMDID database, we utilized a phantomless air-fat-muscle calibration method, similar to the one described by Eggermont et al. (Eggermont et al. 2019). The calibration process was automated using a MATLAB (MathWorks, Natick, MA) script. Subjects for which phantomless calibration failed due to homogeneity in Hounsfield Units of their fat and muscle were excluded from further analysis.

After evaluating and successfully calibrating the scan, the subject's bony geometry was then segmented from the "thin bone torso" (proximal femur) and "thin bone lower extremity" (distal femur, tibia, distal fibula) scans using ScanIP software (Synopsys, Mountain View, CA) and the location of nine key landmarks on the femur (hip center, femur medial and lateral epicondyles) and tibia (anterior medial plateau, posterior medial plateau, anterior lateral plateau, posterior lateral plateau, ankle center) were identified in HyperMesh (Altair, Troy, MI). Left instances were mirrored across the sagittal plane. Local anatomical coordinate systems were created using the previously identified landmarks for each femur and tibia in an automated MATLAB script. The nodes

describing the surfaces of the tibia and femur were transformed from the global CT coordinate system into their respective anatomical coordinate system along with their corresponding landmarks. The transformation matrices defining the transformation between the CT and anatomical coordinate systems were retained for future use.

6.2.3 Template Selection and Meshing

Using a workflow similar to Fugit et al. (in review), median right tibia and femur were selected as the templates and aligned in their local coordinate systems as described above. Tibial template meshes were constructed for surfaces (triangular (tri), 0.95 ± 0.16 mm edge length, 20,296 nodes and 40,588 tri elements) and solids (tetrahedral (tet), 0.96 ± 0.14 mm edge length, 239,903 nodes and 1,364,343 tetrahedral elements). The diaphysis of the tibial template geometry was resected at a ratio of 1.38 between SI length and ML condylar width. Femoral template meshes were constructed for surfaces (2D, 0.82 ± 0.19 mm edge length, 49,767 nodes and 99,530 tri elements) and solids (3D, 1.03 ± 0.16 mm edge length, 418,569 nodes and 2,337,654 tetrahedral elements). The diaphysis of the femoral template geometry was resected at a ratio of 1.74 between SI length and ML condylar width.

6.2.4 Mesh Registration and Morphing

Registration of the instances was fully automated in a MATLAB script. First, since the model is concerned only with the proximal and distal portions of the tibia and femur, respectively, the template resection ratio was applied to trim the shaft of the instance. While the entire tibia was segmented in the NMDID subjects, there were instances where femurs lacked "sufficient shaft" to meet the required aspect ratio due to segmentation

being done across two scans. These cases were subsequently excluded from further analysis.

After adjusting the instance to the correct height, an ICP-based algorithm (Kroon 2024) was used to first rigidly align the instance to the template and then affinely deform the template onto the instance. Following the methods outlined by Andreassen et al. (Andreassen et al. 2024), a Generalized Regression Neural Network (GRNN)-based algorithm completed the morphing of the tri template onto the tri instance. The nodal coordinates of the final deformed template mesh were recorded as the registered instance tri mesh. Morphing error was quantified as the root mean square (RMS) error of the distances between the deformed template points and their respective closest point in the instance mesh. The template tet mesh was subsequently morphed onto the tri instance mesh using GRNN trained on the displacements of the template tri mesh. The nodal coordinates of the final morphed tetrahedral mesh were recorded as the registered instance tetrahedral mesh. Morphing error was quantified as the root mean square (RMS) error of the distances between the surface points of the deformed template tetrahedral mesh and their respective closest point in the instance tri mesh. Tetrahedral mesh quality was evaluated using MATLAB's meshQuality function.

6.2.5 Material Property Assignment

Material properties were derived from the CT scan of each instance and were assigned to the respective registered tetrahedral meshes in MATLAB. Registered meshes were first transformed back into global CT space using the established transformations. Material properties for each element were reported as apparent bone mineral density. To

calculate this, the intensity of the voxel nearest to each node was identified and averaged for the nodes in each element. The average intensity was then converted to BMD using the subject-specific conversion of HU to BMD determined during the scan calibration process and then further converted to apparent density using the established relationships shown in the Equations 6-1 to 6-3 below:

$$BMD = a + b \cdot HU \qquad \text{Equation (6-1)}$$

$$\rho_{ash} = k_{ash} \cdot BMD + l_{ash} \qquad \text{Equation (6-2)}$$

$$\rho_{app} = \frac{\rho_{ash}}{m_{app}} \qquad \text{Equation (6-3)}$$

where *HU* is Hounsfield unit, *BMD* is bone mineral density, a and b are calibration constants, *ρ_ash* and *ρ_app* denote the ash and apparent densities, respectively. Conversion factors *k_ash*, *l_ash*, and *m_app* were adapted from existing literature (J. H. Keyak, Lee, and Skinner 1994; Schileo et al. 2007; Joyce H Keyak et al. 2005) and are shown in Table 6-1. As with the nodal coordinate information, the element material properties for each instance were combined into a matrix: the Material Property Register. Figure 6-1 visualizes the entire registration process from segmentation to Material Property Register creation.

Table 6-1: Conversion factors used to convert HU to BMD in the Tibia and Femur.

Conversion Factors				
	Tibia		Femur	
	Value	Source	Value	Source
k_{ash}	0.953	Keyak 1994	0.887	Keyak 2005
l_{ash}	45.7	Keyak 1994	63.3	Keyak 2005
m_{app}	0.55	Keyak 1994	0.593	Schileo 2008

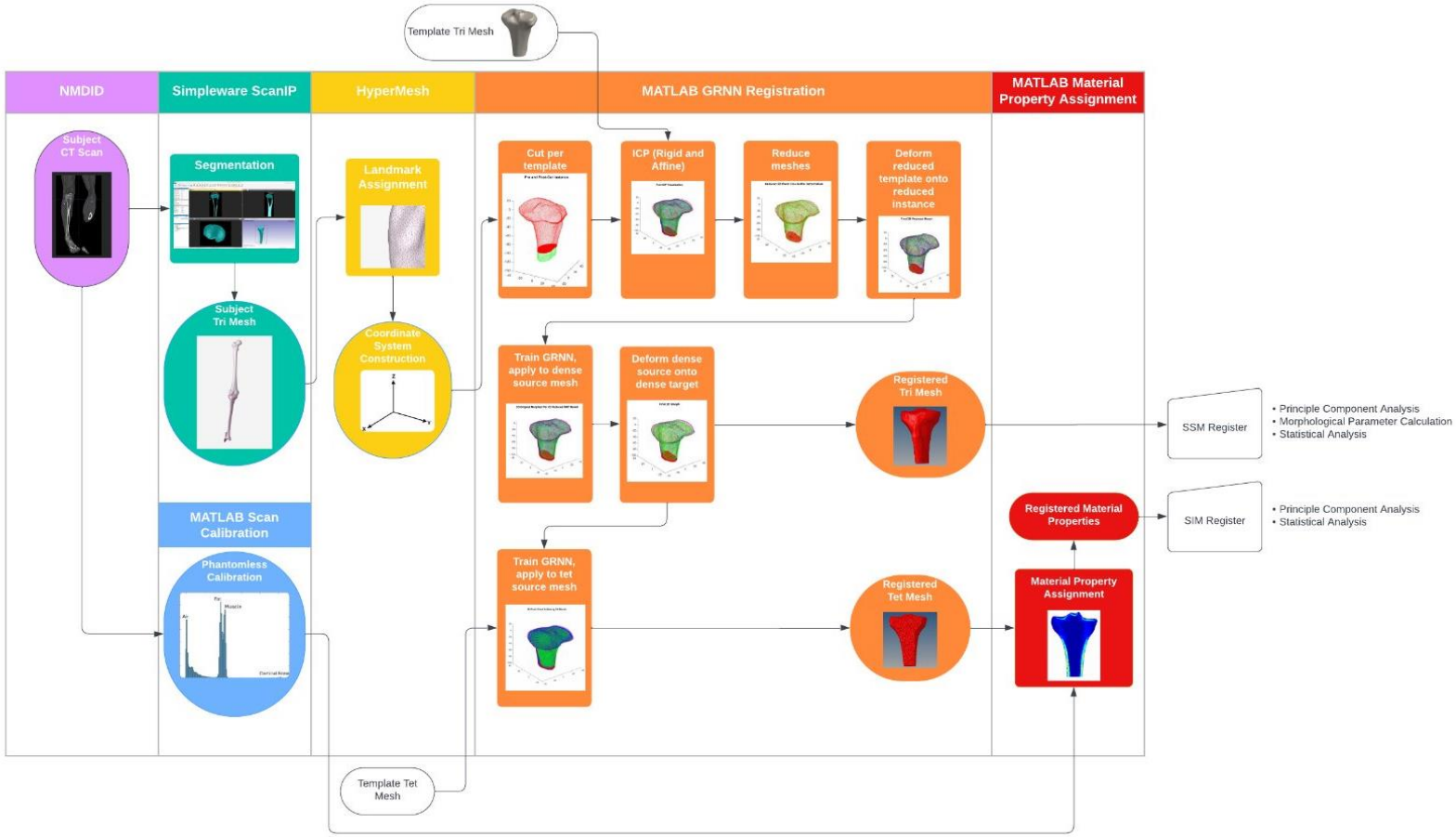

Figure 6-1: Visualization of the registration process from subject segmentation (left) to register creation (right).

6.2.6 Model Generation

Using the registered shape and material property data, a combined Statistical Shape and Intensity Model can be developed. Prior studies have shown that the variability in shape and material property data were not strongly correlated (Burton et al. 2019; Taghizadeh et al. 2017). As such, it is possible to decouple the models, which allows the independent consideration of variation in shape and material property maps.

Principal component analysis was therefore performed individually on both the tri surface registers and material property registers, yielding Statistical Shape Models and Statistical Intensity Models of the Proximal Tibia and Distal Femur, respectively. Metrics of accuracy, compactness, generalizability, and specificity were computed across various levels of SIM principal component (PC) inclusion to evaluate SIM model quality (Audenaert et al. 2019).

6.2.7 Statistical Analysis

Visualization of the first four modes of intensity variation in each model was achieved by adjusting the average material properties within +/- 3.0 standard deviations and superimposing them onto a mean geometry generated from the SSM. Correlation coefficients between the first ten shape and intensity modes were calculated to investigate the relation between shape and material properties.

To quantify interpopulation variances in SIM PC scores, unpaired two-sample t-tests were conducted across different sex and race groups, including Male/Female (M/F), Asian/Black or African American (A/B), Asian/Hispanic (A/H), Asian/Native American (A/NA), Asian/White (A/W), Black or African American/Hispanic (B/H), Black or

African American/Native American (B/NA), Black or African American/White (B/W), Hispanic/Native American (H/NA), Hispanic/White (H/W), and Native American/White (NA/W). Subjects were further categorized into age groups: "young" (ages 25-39), "middle" (ages 40-59), and "old" (ages 60-80), with subsequent two-sample t-tests performed to compare young age/middle age (YA/MA), young age/old age (YA/OA), and middle age/old age (MA/OA). Box plots were employed to visually represent distributions of the first four PC scores and select morphological parameters.

Further, a series of Analysis of Variance (ANOVA) tests were conducted to investigate inter-group differences while accounting for imbalances in population sizes. The independent variables included scan source, sex, race, age, height, and BMI, with the latter three being continuous variables. Although the results of this statistical test are preliminary, this analysis is important when considering multiple independent variables that could potentially impact bone material properties.

6.3 Results

6.3.1 Femur

6.3.1.1 Subject Demographics

Initially, an equal number of scans were evaluated for each demographic category, ensuring balanced representation by sex, race, and a range of ages. However, the training set was unevenly reduced due to issues with scan quality, such as limited field of view or poor tissue differentiation for material property assignment, and the presence of severe osteoarthritis or osteophytes.

The resulting training data for the femur represents 111 cases, 55 female and 56 male, with an average age of 52.8 (+/- 17.3) and 54.7 (+/- 19.1) years, respectively. Subjects obtained from the NMDID database (n = 70) came from Asian (n = 13), Black or African American (16), Hispanic (18), and Native American (n = 10) backgrounds. The remaining subjects (n = 41) included White (n = 38), Hispanic (n = 1), and Pacific Islander (n = 1) subjects and came from the training set of an SSIM previously developed at the University of Denver. The distribution of the overall subject demographics is shown in Table 6-2. Subject age, height, and BMI data by demographic are shown in Figure 6-2, Figure 6-3, and Figure 6-4, respective

Table 6-2: Subject demographics for the Femur SSIM training set.

	Asian	Black or African American	Hispanic	Native American	Pacific Islander	White	Total
Female	7	12	8	10	0	18	55
Age (Years)	45.7 +/- 19.3	44.3 +/- 12.9	45.1 +/- 17.6	46.5 +/- 10.4		68.2 +/- 12.5	52.8 +/- 17.3
Height (cm)	161.6 +/- 7.3	164.4 +/- 7.9	160.7 +/- 7.7	166.2 +/- 9.6		161.7 +/- 5.6	163.0 +/- 7. 4
BMI (kg/m^2)	25.5 +/- 7.6	27.8 +/- 5.7	27.0 +/- 5.6	27.8 +/- 5.5		23.7 +/- 5.9	26.0 +/- 6.0
Male	6	4	11	14	1	20	56
Age (Years)	47 +/- 10.0	50.5 +/- 16.8	41.4 +/- 13.7	39.3 +/- 11.7	77 +/- 0	74.9 +/- 6.3	54.7 +/- 19.1
Height (cm)	171.3 +/- 8.7	174.6 +/- 4.4	172.5 +/- 4.1	175.7 +/- 9.4	165.1 +/- 0	174.4 +/- 8.1	173.9 +/- 7.6
BMI (kg/m2)	25.9 +/- 3.3	27.7 +/- 4.7	28.0 +/- 5.6	29.3 +/- 5.7	29.3 +/- 0	20.5 +/- 4.6	25.4 +/- 6.1
Total	13	16	19	24	1	38	111
Age (Years)	46.3 +/- 15.1	45.8 +/- 13.7	42.9 +/- 15.1	42.3 +/- 11.5	77 +/- 0	71.7 +/- 10.2	53.8 +/- 18.2
Height (cm)	166.1 +/- 9.2	167.0 +/- 8.3	167.5 +/- 8.3	171.8 +/- 10.4	165.1 +/- 0	168.4 +/- 9.4	168.5 +/- 9.3
BMI (kg/m2)	25.7 +/- 5.8	27.8 +/- 5.3	27.5 +/- 5.5	28.6 +/- 5.5	29.3 +/- 0	22.0 +/- 5.4	25.7 +/- 6.1

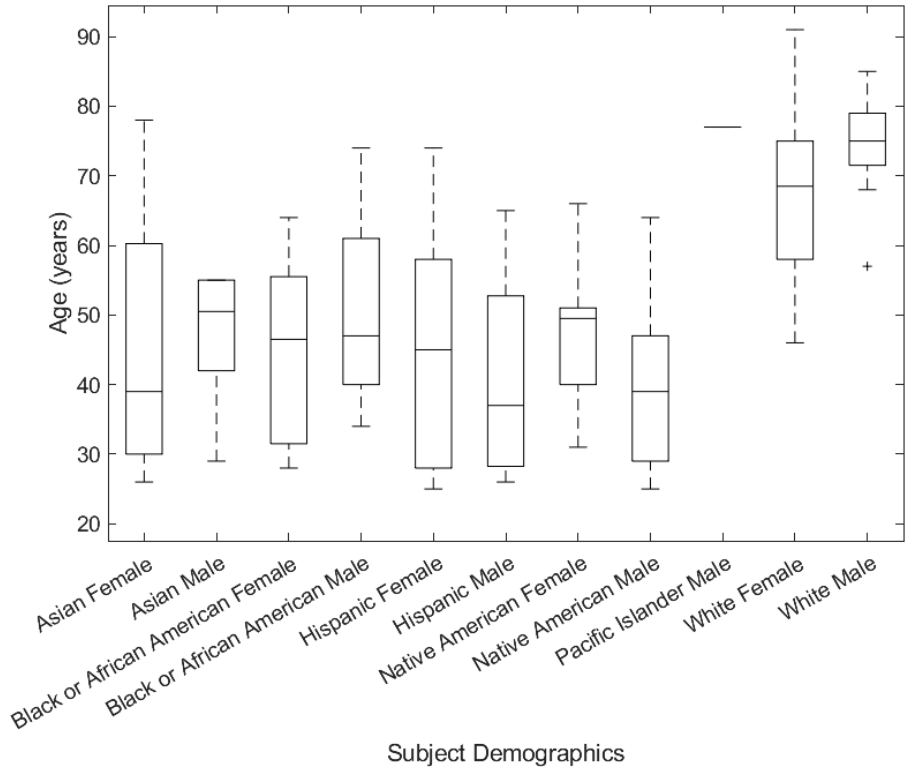

Figure 6-2: Distribution of subject age by demographic.

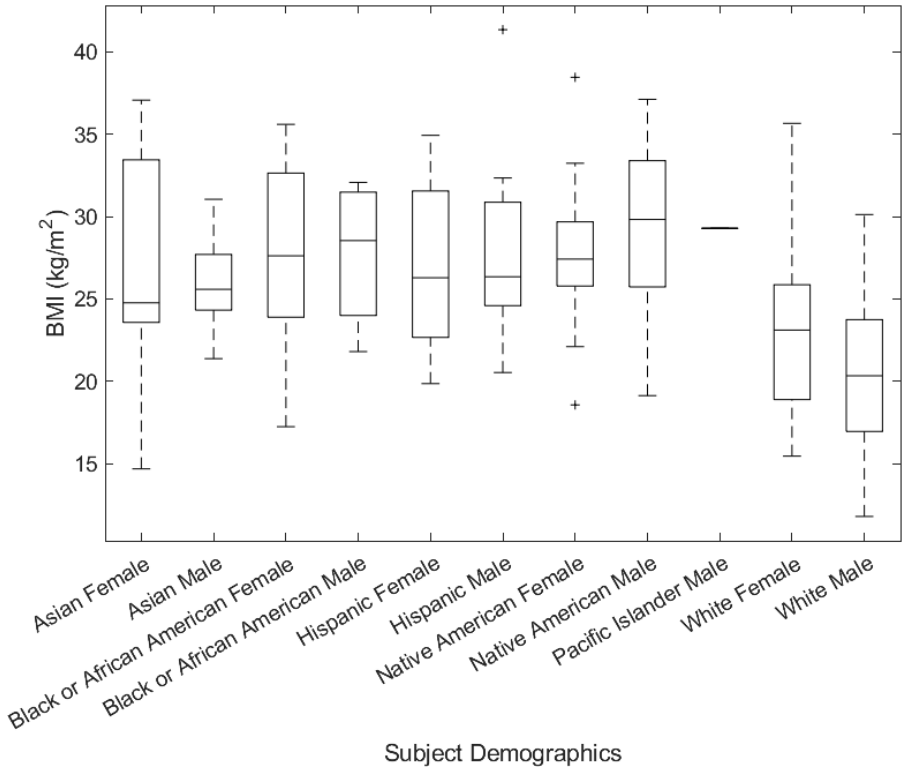

Figure 6-3: Distribution of subject BMI by demographic.

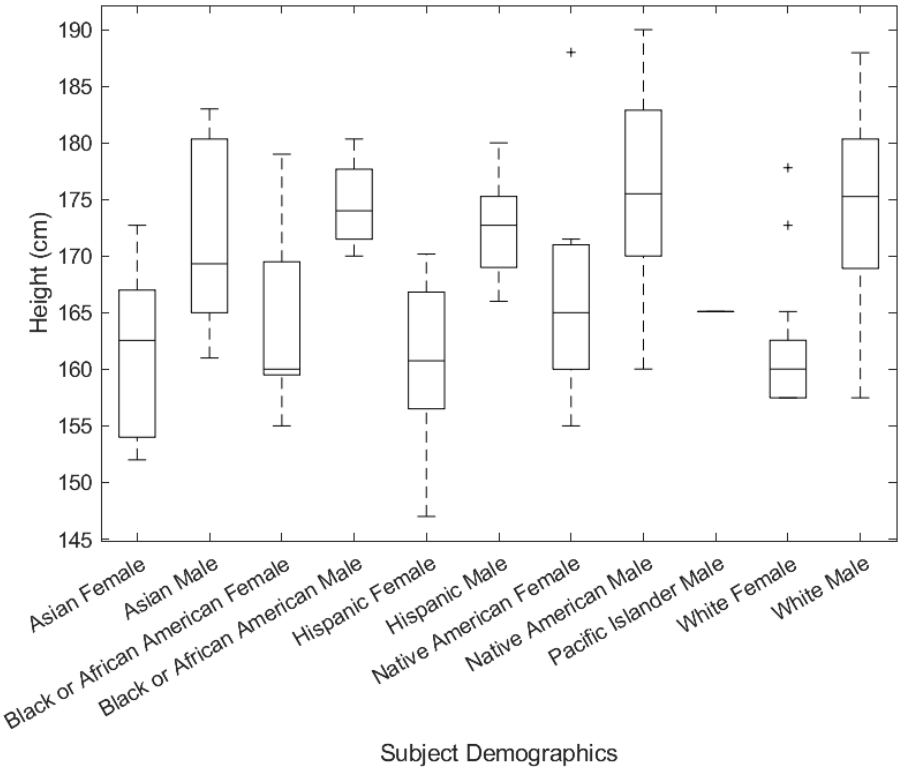

Figure 6-4: Distribution of subject height by demographic.

6.3.1.2 Model Quality

Measures of accuracy, compactness, generalizability, and specificity were computed to assess model quality. The accuracy of the Femur SIM improved with the inclusion of a greater number of principal components (Figure 6-5). Accuracies achieved by inclusion of selected numbers of principal components are outlined in Table 6-3. Figure 6-6 depicts the cumulative compactness of the Femur SIM, indicating that 79 principal components were necessary to encompass at least 95% of the model variance. The distribution of variance across selected principal components is reported in Table 6-3. The generalizability of the model with increasing numbers of training set samples is shown in

Figure 6-7. With 111 instances, the generalizability and specificity of the model were found to be 101.81 +/- 16.71 kg/m^3 and 180.49 +/- 19.69 kg/m^3, respectively (Table 6-4).

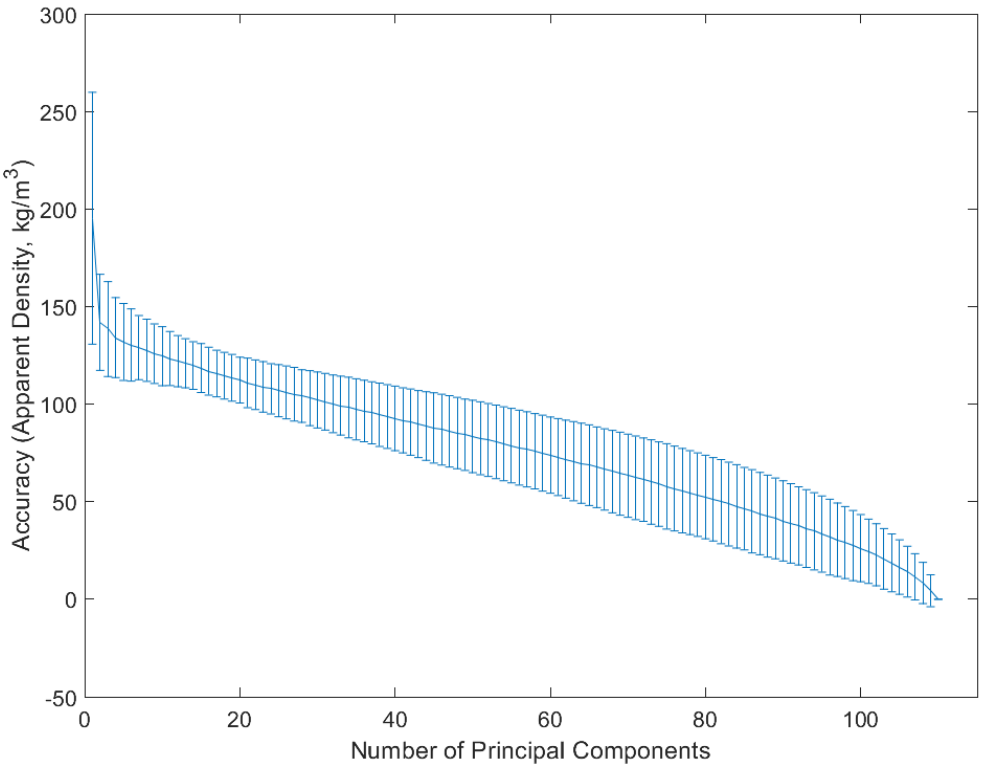

Figure 6-5: Accuracy of the Femur SIM with the progressive addition of principal components to the model.

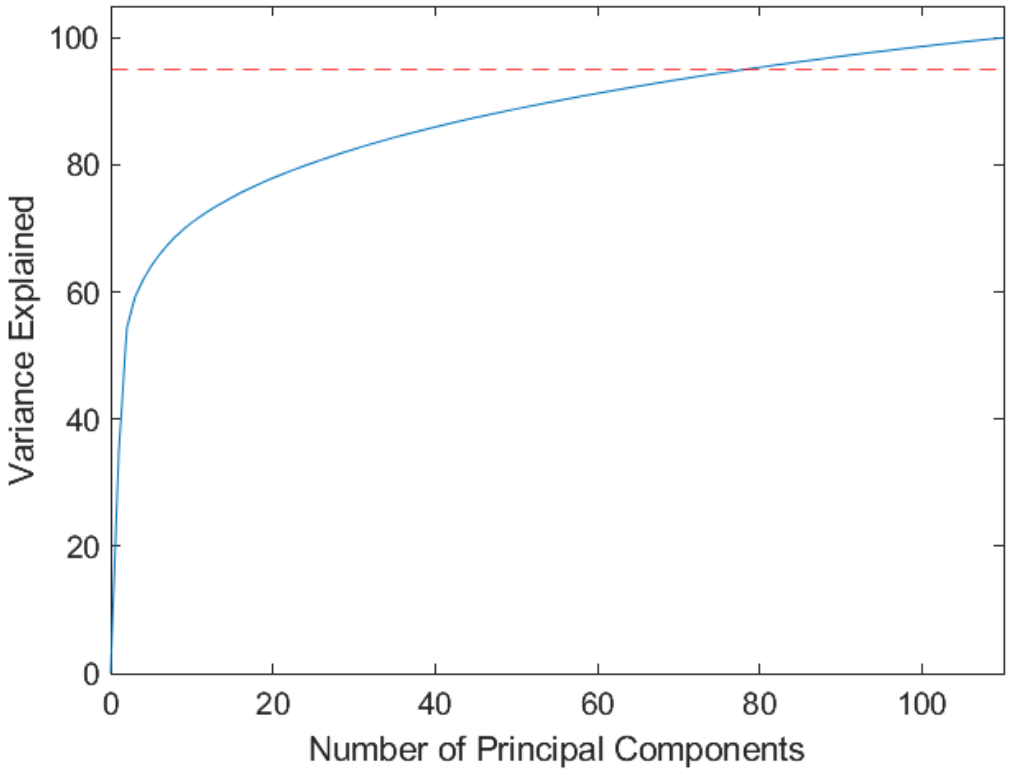

Figure 6-6: Cumulative compactness of the Femur SIM for increasing number of PCs included.

Table 6-3: Accuracy, variance explained, and cumulative compactness for select PCs of the Femur SIM. The PC at which the desired compactness threshold is reached is highlighted.

Num PCs	Accuracy (Apparent Density, kg/m^3)	Variance Explained (%)	Cumulative (%)
1	195.24 +/- 64.54	34.60	34.60
2	141.91 +/- 24.65	19.76	54.36
3	138.45 +/- 24.33	4.87	59.23
4	134.09 +/- 20.49	2.71	61.94
5	131.85 +/- 19.71	2.17	64.11
6	130.30 +/- 18.49	1.75	65.86
7	128.94 +/- 16.44	1.55	67.40
8	127.58 +/- 15.96	1.35	68.75
9	125.82 +/- 15.26	1.15	69.90
10	124.55 +/- 15.14	1.04	70.94
20	112.31 +/- 11.72	0.54	77.93
30	102.14 +/- 14.35	0.40	82.48
40	92.66 +/- 16.55	0.31	85.97
50	83.41 +/- 18.55	0.26	88.81
60	73.88 +/- 19.54	0.23	91.26
70	63.31 +/- 21.29	0.20	93.43
79	53.59 +/- 21.39	0.19	95.18
110	1.34E-12 +/- 5.88E-13	0.121	100

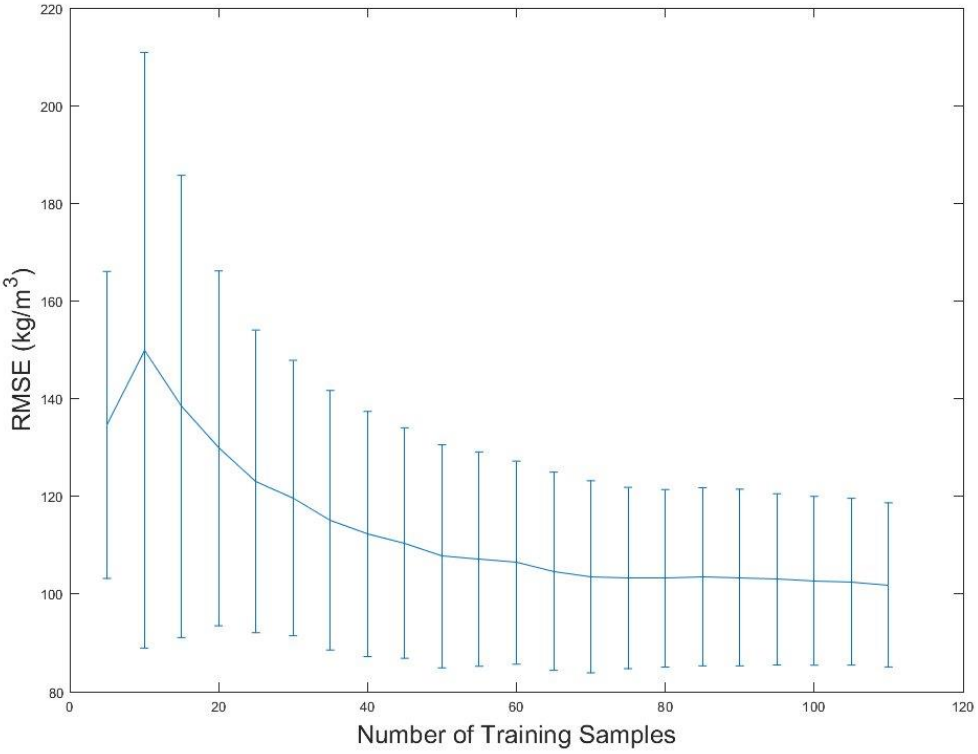

Figure 6-7: Generalizability of the Femur SIM with the progressive addition of training samples to the register.

Table 6-4: Generalizability and specificity for the Femur SIM.

Number of Instances	110
Generalizability (Apparent Density, kg/m^3)	101.81 +/- 16.71
Specificity (Apparent Density, kg/m^3)	180.49 +/- 19.69

6.3.1.3 Model Outcomes

The first four modes of variation in the Femur SIM are visualized by perturbing the mean geometry at +/- 3.0 standard deviations and displaying the resulting material properties in select cut planes, shown in Figure 6-8. Mode 1 (Figure 6-9) described an overall change in bone quality. Mode 2 (Figure 6-10) appeared to capture changes in

thickness of the cortical wall, while also potentially capturing the edge artifact due to air surrounding the bone in material property assignment, which can be influenced by the segmentation process. Mode 3 (Figure 6-11) described primarily a change in trabecular bone density. Finally, Mode 4 (Figure 6-12) captures variation in cortical thickness and a small complementary change in trabecular density.

Correlation coefficients between the first ten SSM (from Chapter 5) and SIM modes were calculated, and significant correlations were reported (Table 6-5). Notably, none of the correlations between the shape and intensity modes of the femur were greater than 0.5 or strong, thereby supporting the approach to decouple the SSM and SIM.

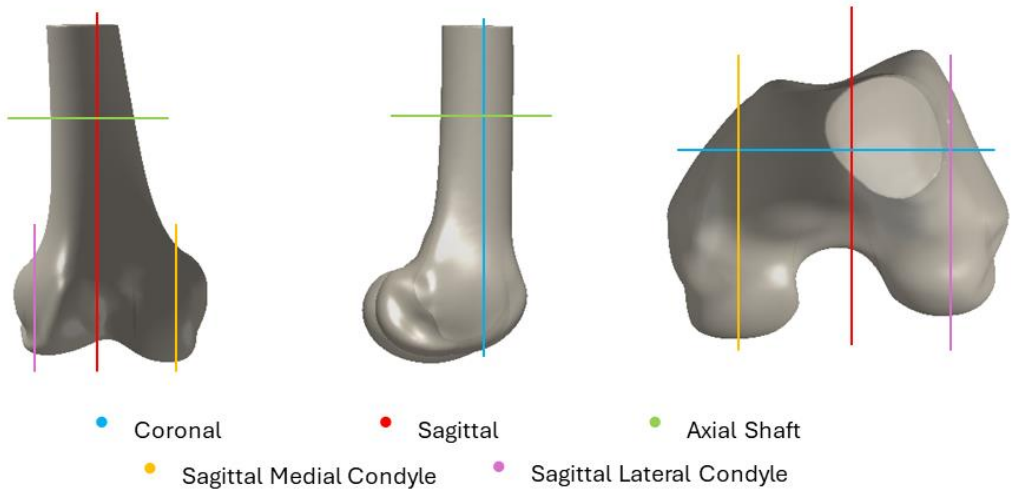

Figure 6-8: Cut planes used to visualize material properties in the femur.

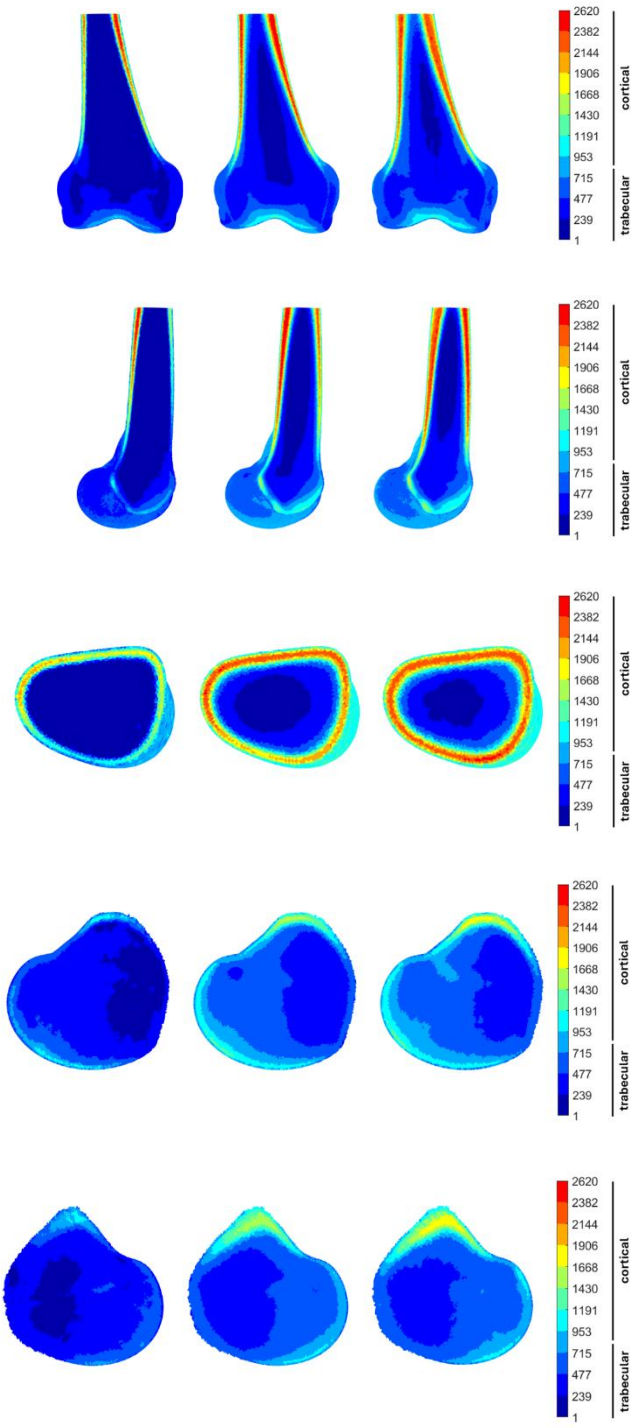

Figure 6-9: Material property variation in Mode 1 of the Femur SIM. Variations extending to +/- three standard deviations from the mean are displayed along key anatomical planes, including the coronal midline, sagittal midline, transverse shaft, sagittal medial condyle, and sagittal lateral condyle.

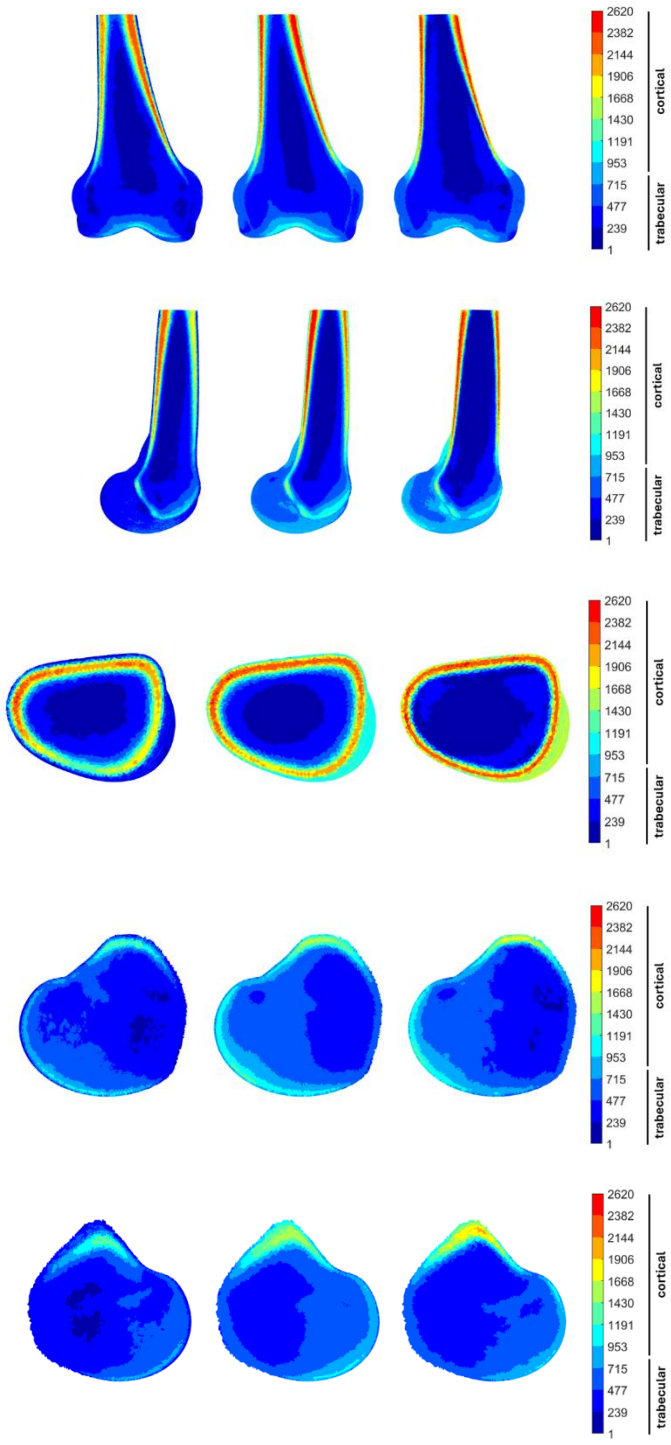

Figure 6-10: Material property variation in Mode 2 of the Femur SIM. Variations extending to +/- three standard deviations from the mean are displayed along key anatomical planes, including the coronal midline, sagittal midline, transverse shaft, sagittal medial condyle, and sagittal lateral condyle.

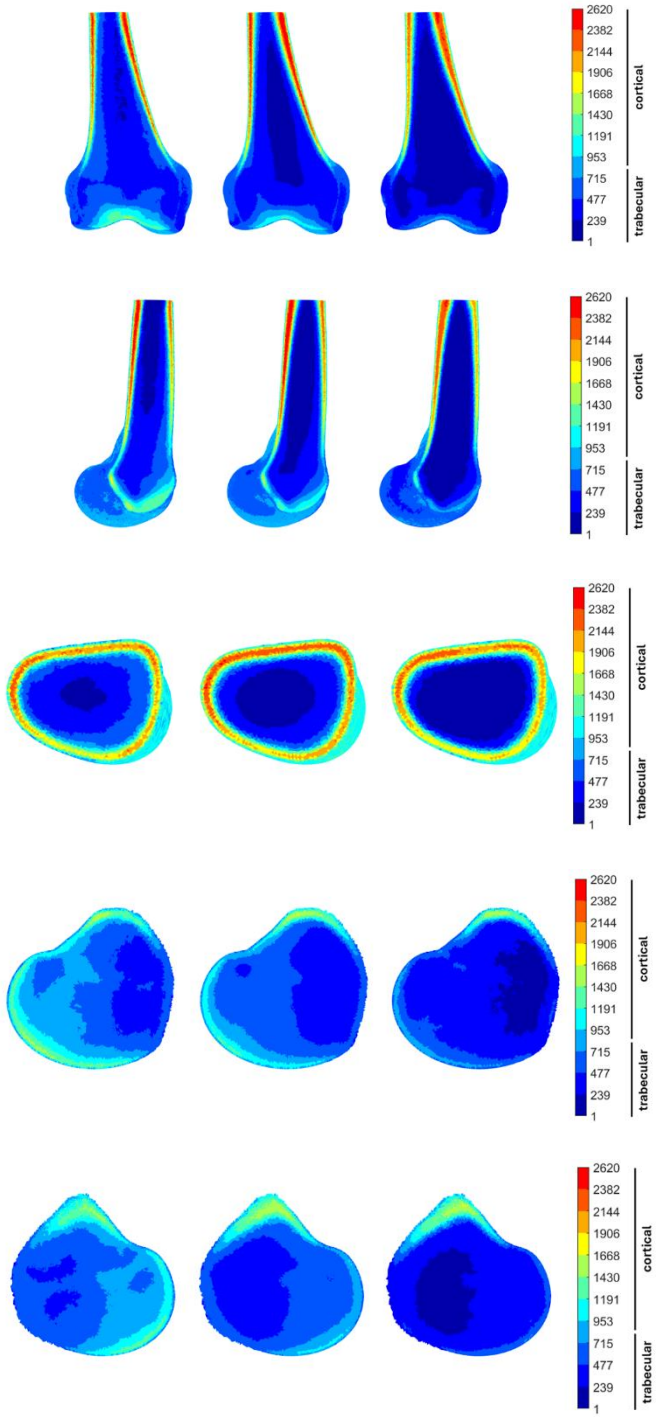

Figure 6-11: Material property variation in Mode 3 of the Femur SIM. Variations extending to +/- three standard deviations from the mean are displayed along key anatomical planes, including the coronal midline, sagittal midline, transverse shaft, sagittal medial condy

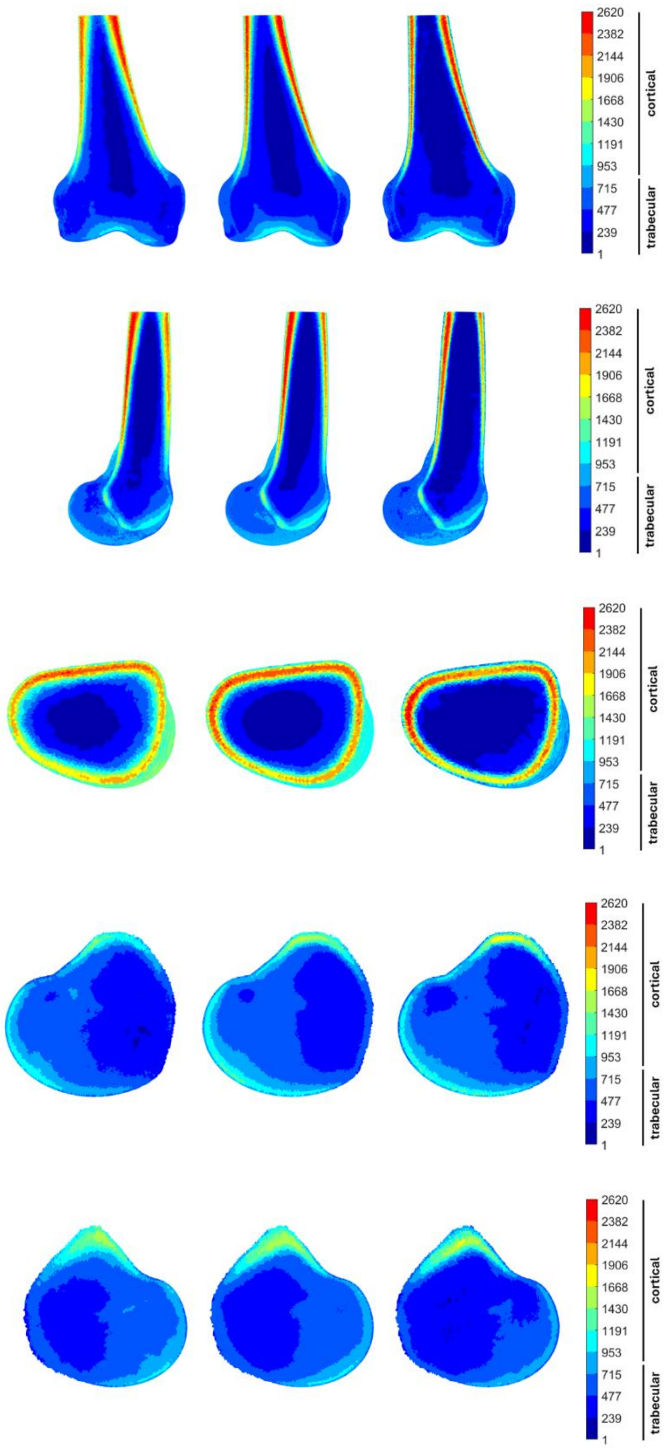

Figure 6-12: Material property variation in Mode 4 of the Femur SIM. Variations extending to +/- three standard deviations from the mean are displayed along key anatomical planes, including the coronal midline, sagittal midline, transverse shaft, sagittal medial condyle.

Table 6-5: Correlation coefficients (R) between Femur SSM and SIM PC scores. Only statistically significant (p<0.05) correlations are shown. High correlations (R> 0.65) are bolded.

	SSM PC 1	SSM PC 2	SSM PC 3	SSM PC 4	SSM PC 5	SSM PC 6	SSM PC 7	SSM PC 8	SSM PC 9	SSM PC 10	SIM PC 1	SIM PC 2	SIM PC 3	SIM PC 4	SIM PC 5	SIM PC 6	SIM PC 7	SIM PC 8	SIM PC 9	SIM PC 10
SSM PC 1														-0.26						
SSM PC 2																				
SSM PC 3																				
SSM PC 4											-0.20				-0.27					
SSM PC 5																	-0.42			
SSM PC 6											-0.28									
SSM PC 7																-0.19	-0.25	-0.27		-0.23
SSM PC 8													0.22						0.28	
SSM PC 9												-0.23				0.26				
SSM PC 10													-0.20		-0.20			0.20		
SIM PC 1				-0.20		-0.28														
SIM PC 2									-0.23											
SIM PC 3								0.22		-0.20										
SIM PC 4	-0.26																			
SIM PC 5				-0.27						-0.20										
SIM PC 6							-0.19		0.26											
SIM PC 7					-0.42		-0.25													
SIM PC 8							-0.27			0.20										
SIM PC 9								0.28												
SIM PC 10							-0.23													

6.3.1.4 Interpopulation Comparisons

The sex and racial distribution of the scores pertaining to the first four principal components are depicted in Figure 6-13 (PC1 and PC2) and Figure 6-14 (PC3 and PC4). In Figure 6-13, differences were observed in the scatter of PC 2 for the White population; this population was derived from a cadaveric dataset at DU instead of from the NMDID database and the observed differences may be due to the scan source and segmentation process.

To quantify interpopulation differences in PC scores and morphological parameters, unpaired two-sample t-tests and a series of ANOVA tests were conducted. Table 6-6

displays the p-values resulting from t-tests with significant differences ($p < 0.05$) between two populations. Boxplots illustrating the distribution of PCs one through four across sex, race, and age categories are shown in Figures 6-10 to 6-21. Significant differences ($p < 0.05$) in these plots are indicated by a red bar. Table 6-7 displays significant p values ($p < 0.05$) resulting from ANOVA tests for the first ten PCs considering factors of scan source, sex, race, age, height, and BMI.

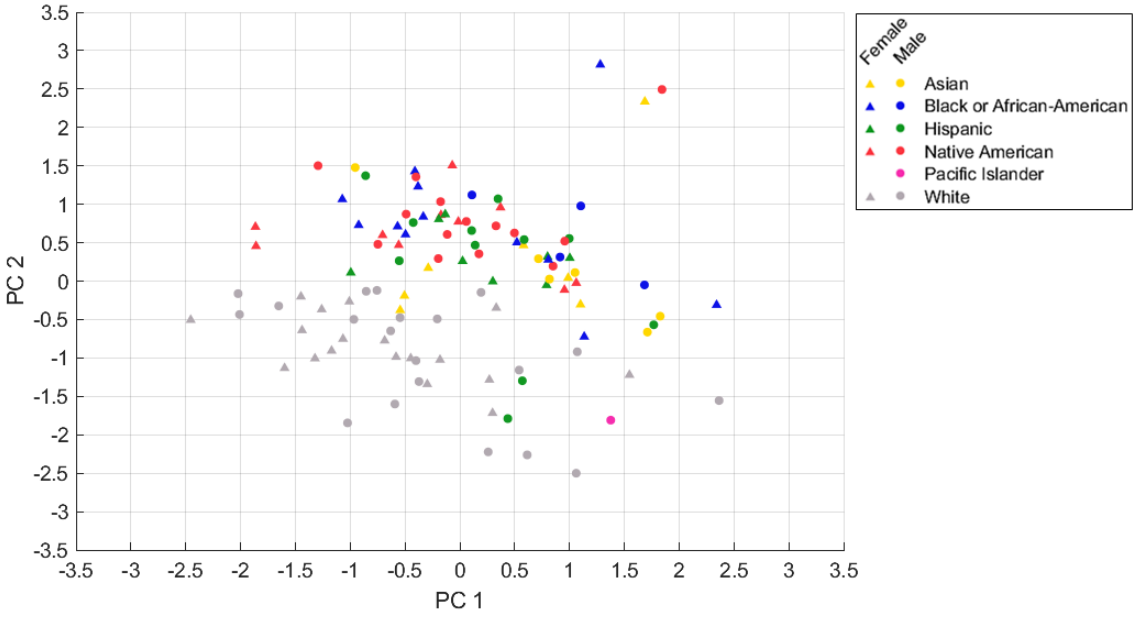

Figure 6-13: Sex and racial distribution of PC 1 and PC 2 scores of the Femur SIM.

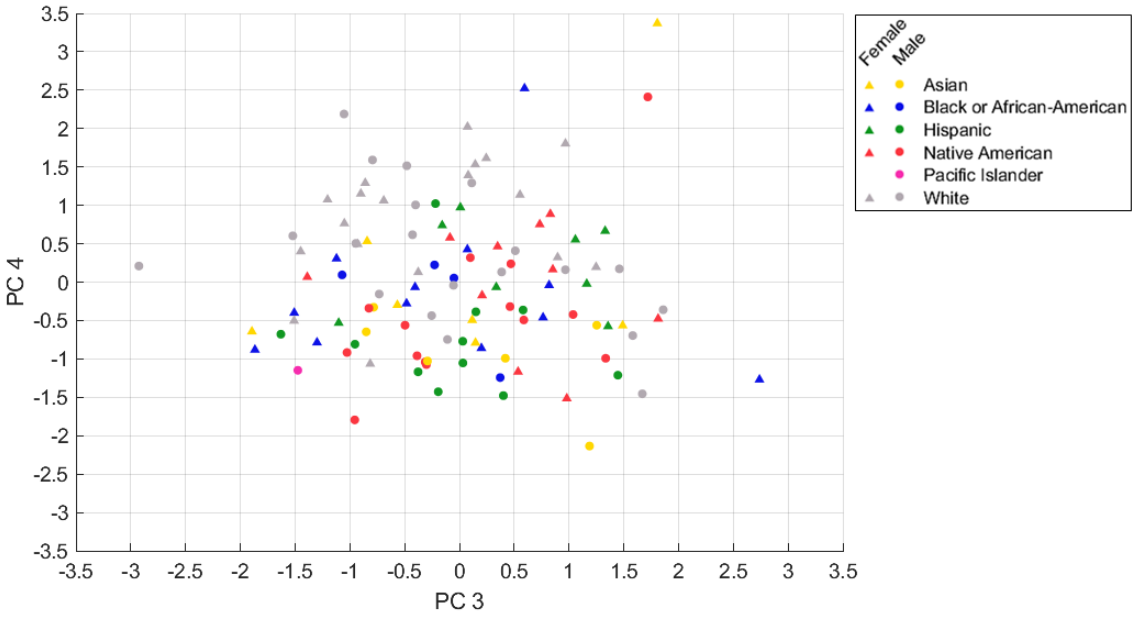

Figure 6-14: Sex and racial distribution of PC 3 and PC 4 scores of the Femur SIM.

Table 6-6: P-values derived from unpaired two-sample t-tests indicating significant differences in PC scores one to ten ($p < 0.05$) between populations. Highly significant differences ($p < 0.001$) are shown in bold.

	M/F	A/B	A/H	A/NA	A/W	B/H	B/NA	B/W	H/NA	H/W	NA/W	YA/MA	YA/OA	MA/OA
PC 1				3.13E-02	1.09E-03			7.58E-03		6.70E-03			**1.73E-04**	1.81E-02
PC 2				2.70E-02	**3.10E-06**			**8.60E-11**	1.62E-02	**8.73E-08**	**1.39E-15**		**1.40E-10**	**1.28E-07**
PC 3												1.24E-03		
PC 4	2.95E-03				5.80E-03			7.79E-03		**3.90E-04**	**7.48E-04**			
PC 5	**4.25E-05**													
PC 6	4.18E-03	8.10E-03				4.70E-02	1.17E-02	1.37E-02				1.32E-02		**9.11E-04**
PC 7														
PC 8	1.68E-03		3.80E-02	1.16E-02	3.85E-02									
PC 9	1.81E-03													
PC 10														

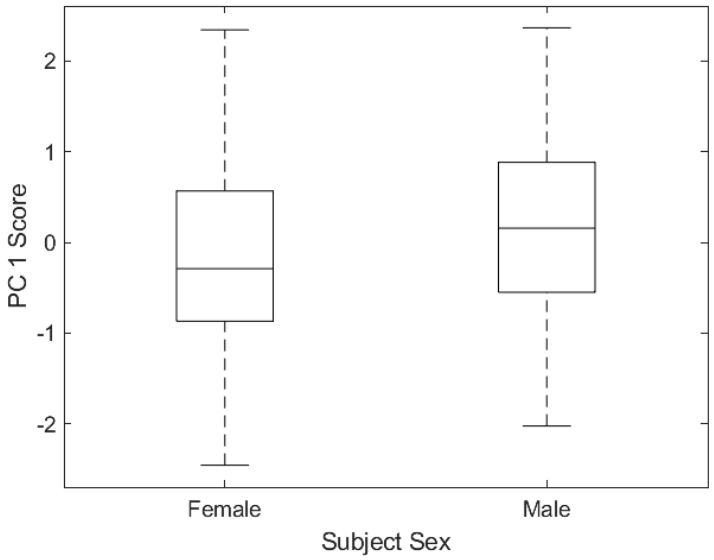

Figure 6-15: Femur SIM PC 1 score by subject sex.

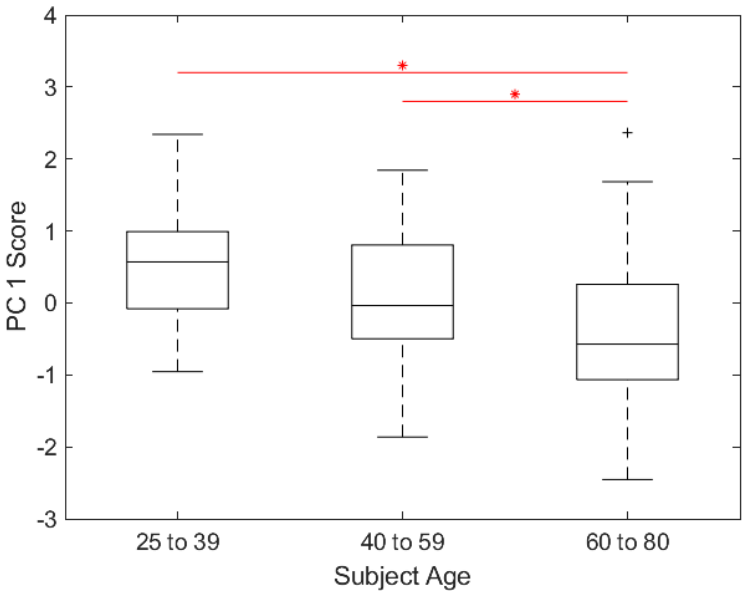

Figure 6-16: Femur SIM PC 1 score by subject age.

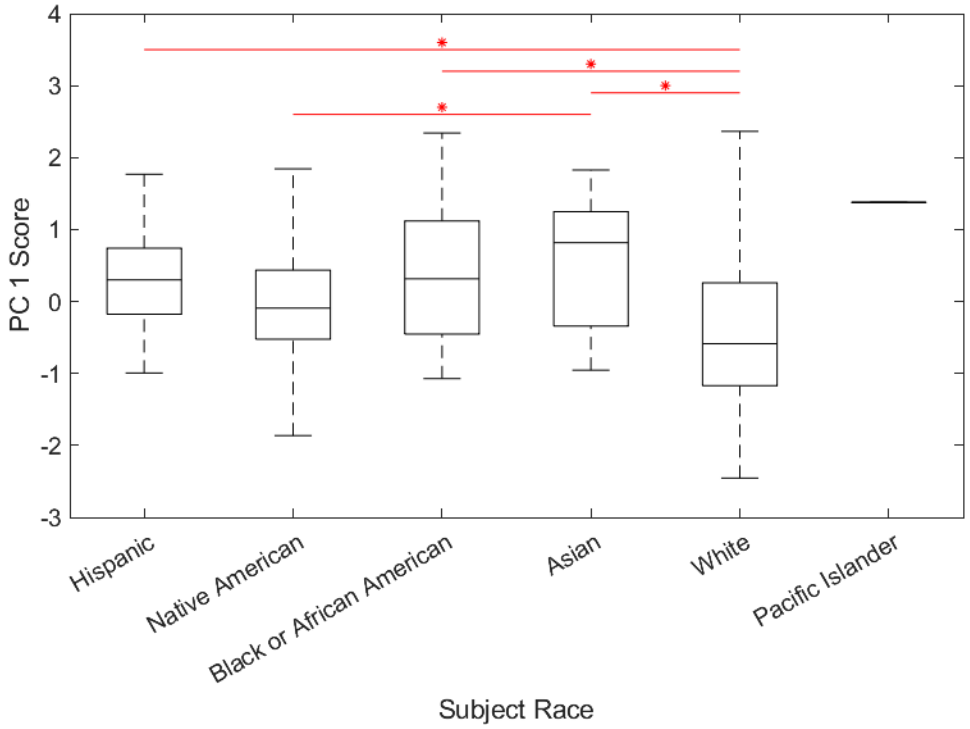

Figure 6-17: Femur SIM PC 1 score by subject race.

Figure 6-18: Femur SIM PC 2 score by subject sex.

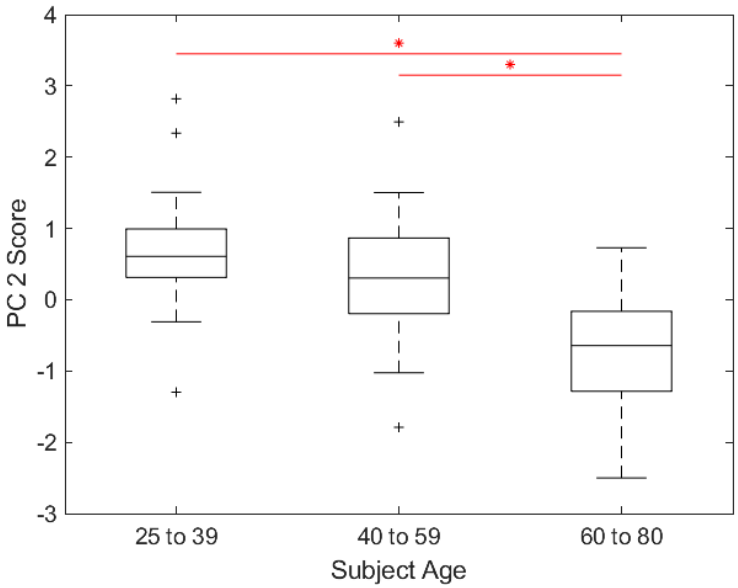

Figure 6-19: Femur SIM PC 2 score by subject age.

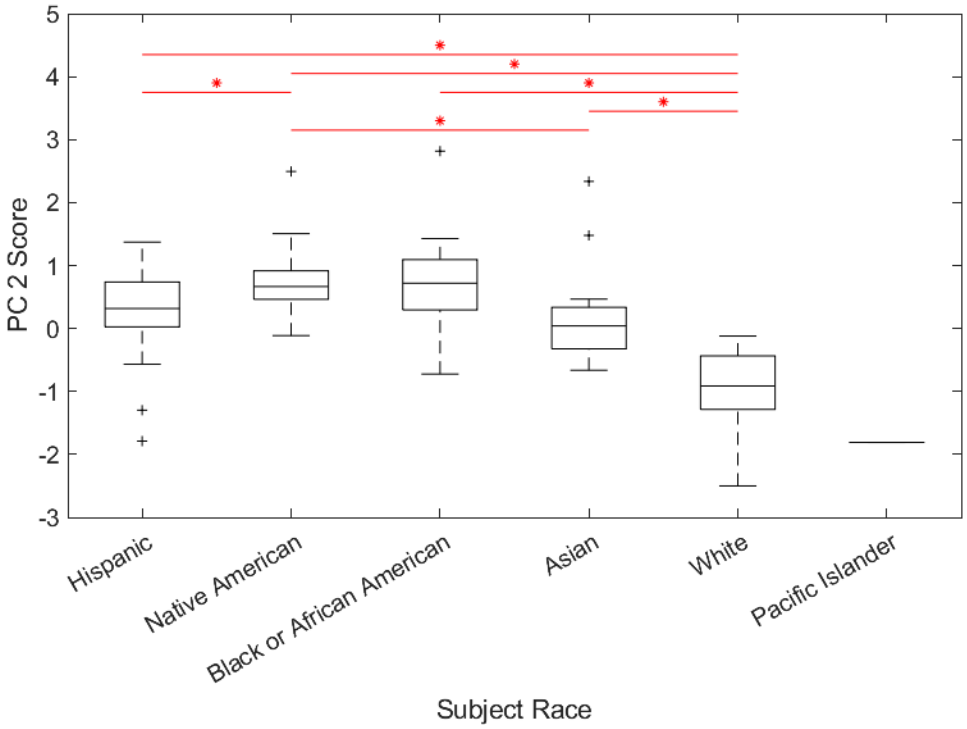

Figure 6-20: Femur SIM PC 2 score by subject race.

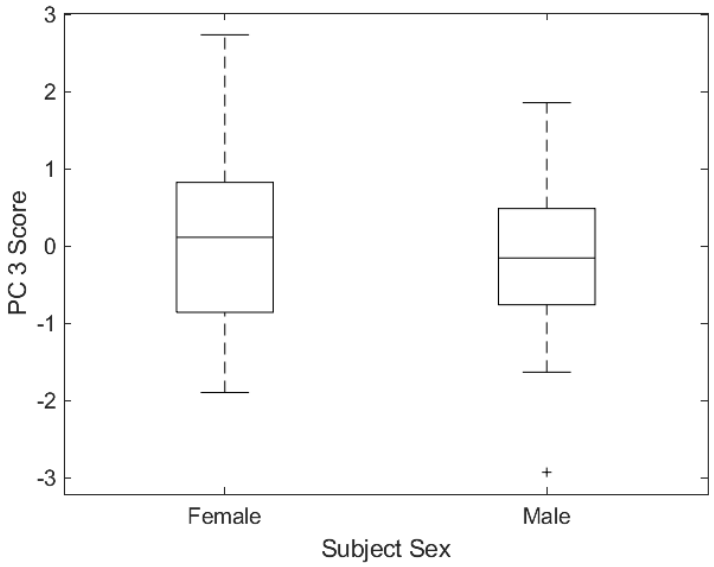

Figure 6-21: Femur SIM PC 3 score by subject sex.

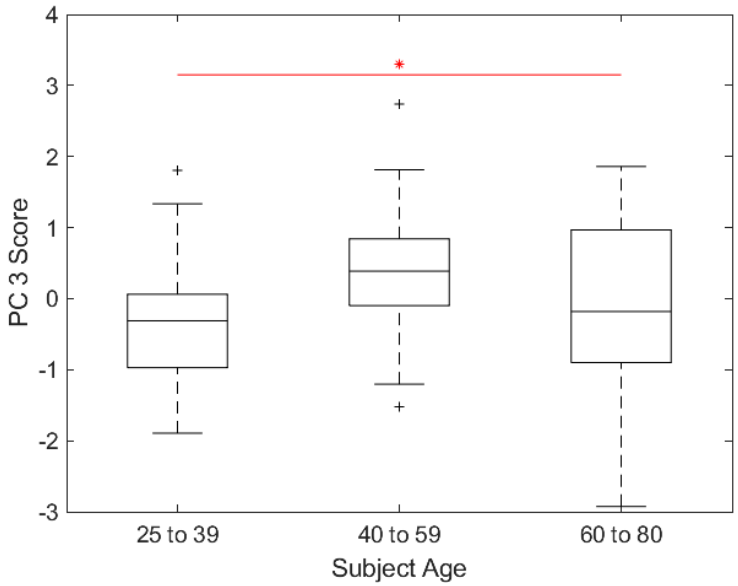

Figure 6-22: Femur SIM PC 3 score by subject age.

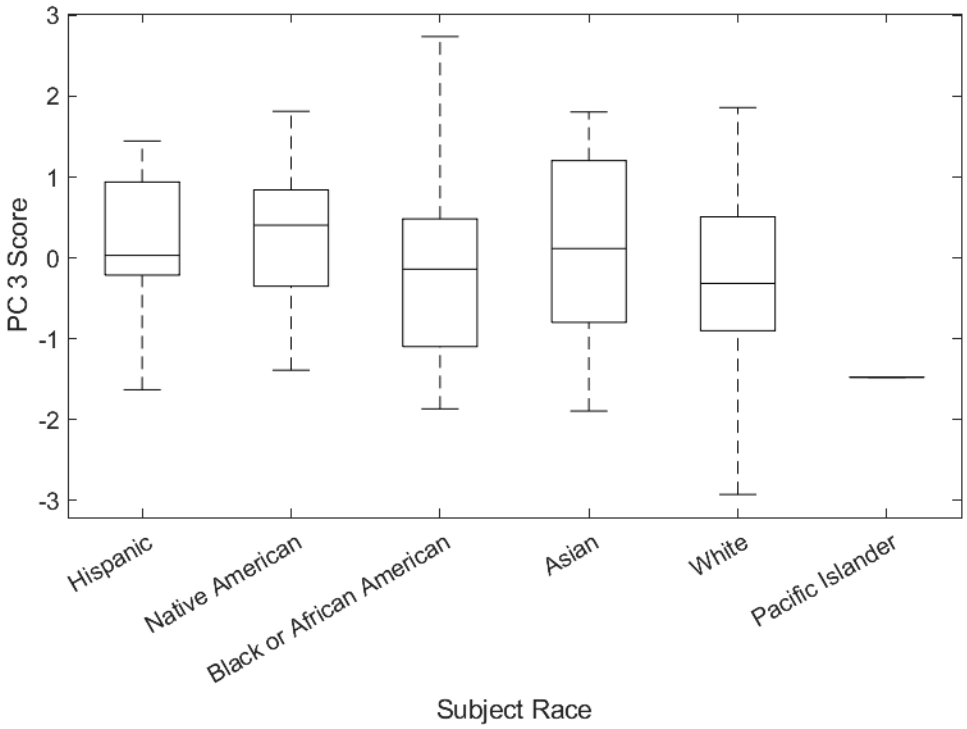

Figure 6-23: Femur SIM PC 3 score by subject race.

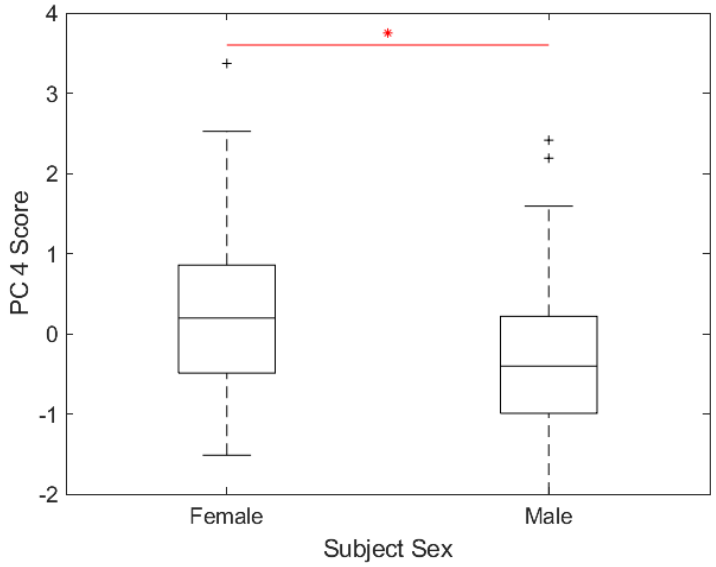

Figure 6-24: Femur SIM PC 4 score by subject sex.

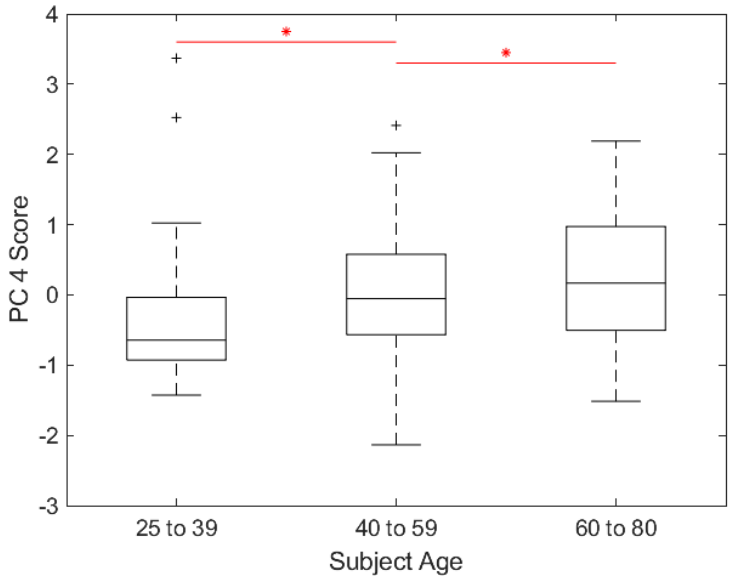

Figure 6-25: Femur SIM PC 4 score by subject age.

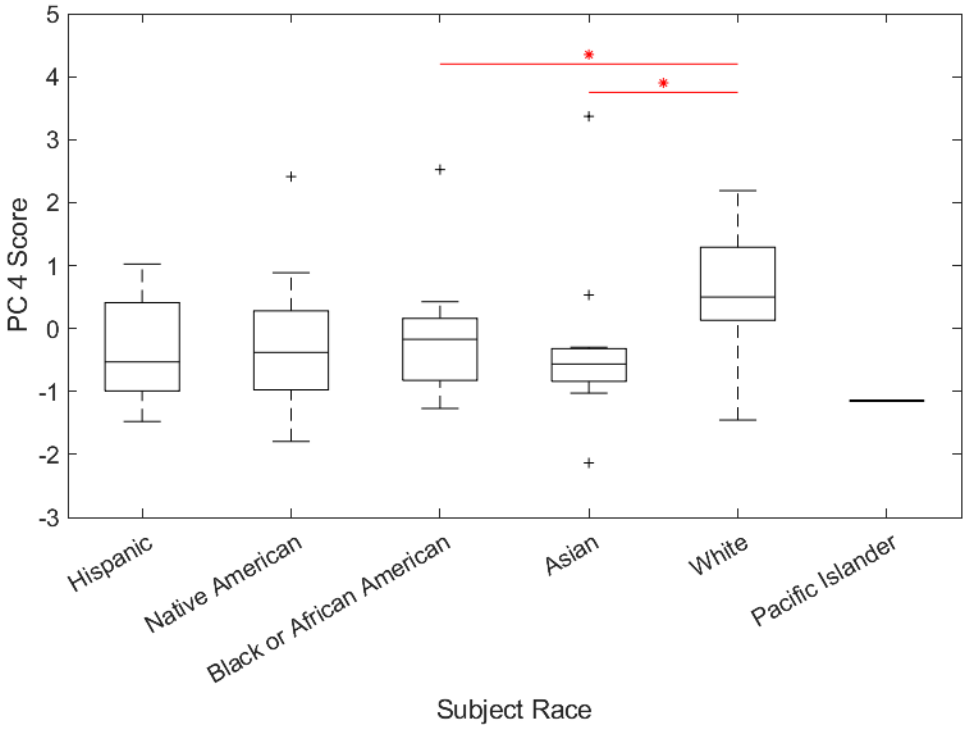

Figure 6-26: Femur SIM PC 4 score by subject race.

Table 6-7: ANOVA-derived p-values indicating significant differences ($p < 0.05$) between populations. Highly significant differences ($p < 0.001$) are shown in bold.

	Sex	Age	Race	Height	BMI	Source
PC 1	8.77E-03	7.94E-03	3.47E-02		2.41E-02	
PC 2				1.63E-02		9.33E-03
PC 3		1.49E-02		**2.03E-04**		
PC 4	1.55E-02					
PC 5	**1.37E-04**	4.75E-03				
PC 6	2.71E-02					
PC 7						
PC 8				1.59E-02	3.57E-02	
PC 9		4.55E-02		1.50E-02		
PC 10		2.03E-02				

6.3.2 Tibia

6.3.2.1 Subject Demographics

Initially, an equal number of scans were evaluated for each demographic category, ensuring balanced representation by sex, race, and a range of ages. However, the training set was unevenly reduced due to issues with scan quality, such as limited field of view or poor tissue differentiation for material property assignment, and the presence of severe osteoarthritis or osteophytes.

The resulting training data for the tibia represents 115 cases, 58 female and 57 male, with an average age of 52.7 (+/- 16.9) and 54.1 (+/- 18.5) years, respectively. Subjects obtained from the NMDID database (n = 72) came from Asian (n = 13), Black or African American (n = 17), Hispanic (n = 18), and Native American (n = 26) backgrounds. The remaining subjects (n = 42) included White (n = 41) and Hispanic (n = 1) subjects and came from the training set of an SSIM previously developed at the University of Denver. The distribution of the overall subject demographics is shown in Table 6-8. Subject age, height, and BMI data by demographic are shown in Figure 6-27, Figure 6-28, and Figure 6-29, respectively.

Table 6-8: Subject demographics for the Tibia SSIM training set.

	Asian	Black or African American	Hispanic	Native American	White	Total
Female	7	13	7	11	20	58
Age (Years)	45.7 +/- 19.3	46.6 +/- 15.0	41.0 +/- 14.3	46.5 +/- 9.9	66.6 +/- 12.7	52.7 +/- 16.9
Height (cm)	161.6 +/- 7.3	164.8 +/- 7.6	159.9 +/- 8.0	165.7 +/- 9.2	162.0 +/- 6.4	163.0 +/- 7.5
BMI (kg/m^2)	25.5 +/- 7.6	27.6 +/- 5.5	25.8 +/- 4.9	28.1 +/- 5.3	24.6 +/- 6.9	26.2 +/- 6.2
Male	6	4	11	15	21	57
Age (Years)	47 +/- 10.0	43.0 +/- 6.2	41.4 +/- 13.7	41.2 +/- 13.5	74.0 +/- 6.2	54.1 +/- 18.5
Height (cm)	171.3 +/- 8.7	174.0 +/- 3.3	172.5 +/- 4.1	175.2 +/- 9.3	174.8 +/- 8.0	174.0 +/- 7.5
BMI (kg/m2)	25.9 +/- 3.3	26.2 +/- 3.7	28.0 +/- 5.6	29.1 +/- 5.5	21.0 +/- 4.9	25.3 +/- 6.0
Total	13	17	18	26	41	115
Age (Years)	46.3 +/- 15.1	45.8 +/- 13.4	41.2 +/- 13.5	43.5 +/- 12.2	70.4 +/- 10.5	53.4 +/- 17.7
Height (cm)	166.1 +/- 9.2	167.0 +/- 7.8	167.6 +/- 8.5	171.2 +/- 10.2	168.5 +/- 9.6	168.5 +/- 9.3
BMI (kg/m2)	25.7 +/- 5.8	27.2 +/- 5.1	27.1 +/- 5.3	28.7 +/- 5.4	22.8 +/- 6.2	25.8 +/- 6.1

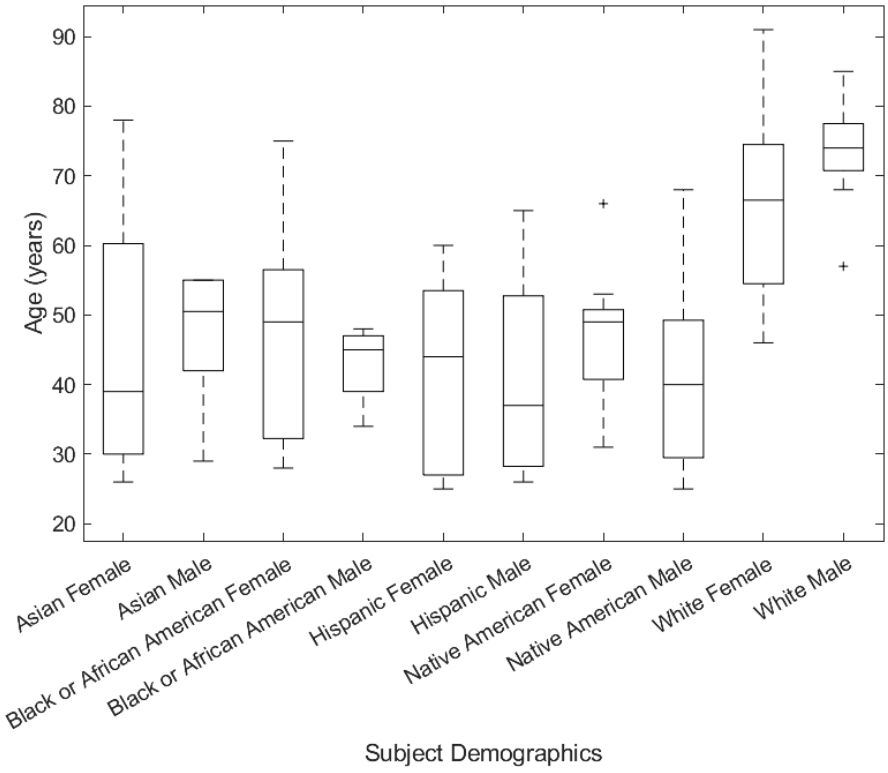

Figure 6-27: Distribution of subject age by demographic.

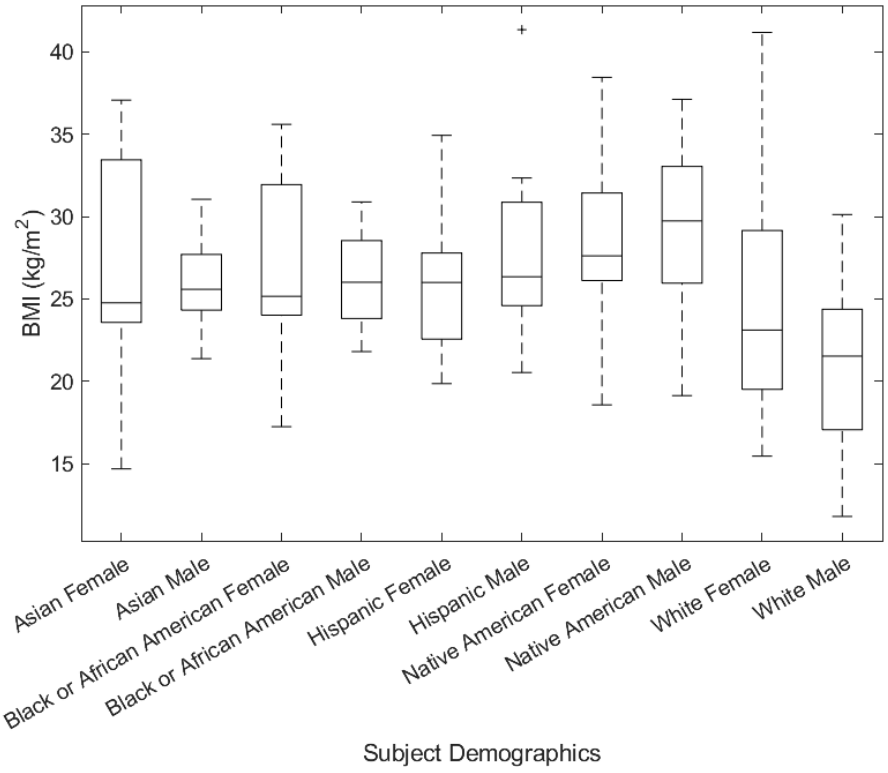

Figure 6-28: Distribution of subject BMI by demographic.

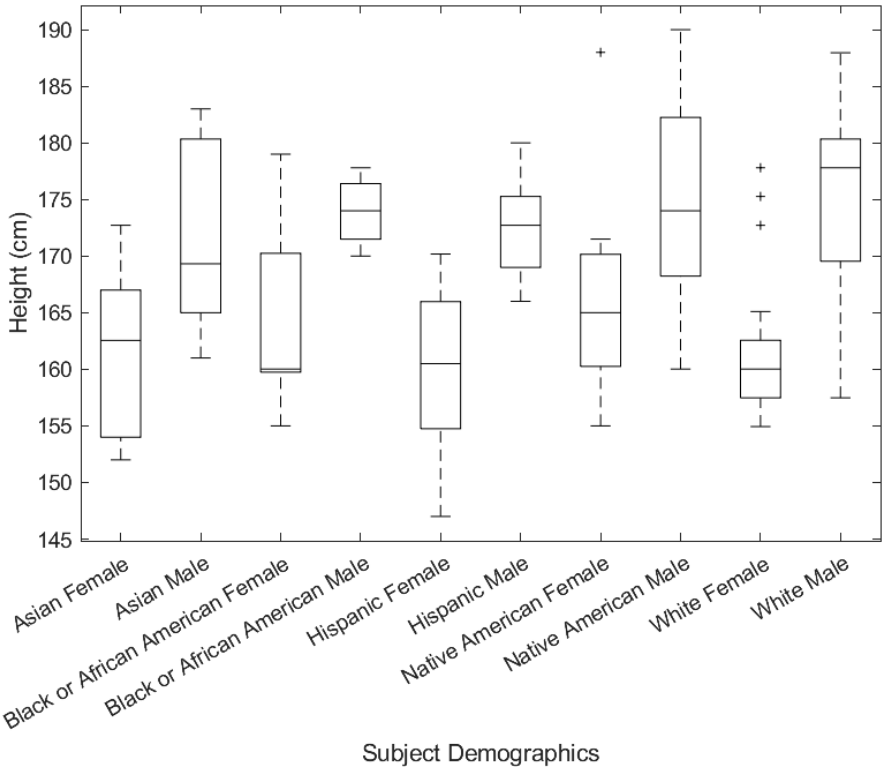

Figure 6-29: Distribution of subject height by demographic.

6.3.2.2 Model Quality

Measures of accuracy, compactness, generalizability, and specificity were computed to assess model quality. The accuracy of the Tibia SIM improved with the inclusion of a greater number of principal components (Figure 6-30). Accuracies achieved by inclusion of selected numbers of principal components are outlined in Table 6-9. Figure 6-31 depicts the cumulative compactness of the Femur SIM, indicating that 82 principal components were necessary to encompass at least 95% of the model variance. The distribution of variance across selected principal components is reported in Table 6-9. The generalizability of the model with increasing numbers of training set samples is shown in Figure 6-32. With 115 instances, the generalizability and specificity of the

model were found to be 101.81 +/- 16.71 kg/m^3 and 180.49 +/- 19.69 kg/m^3, respectively (Table 6-4).

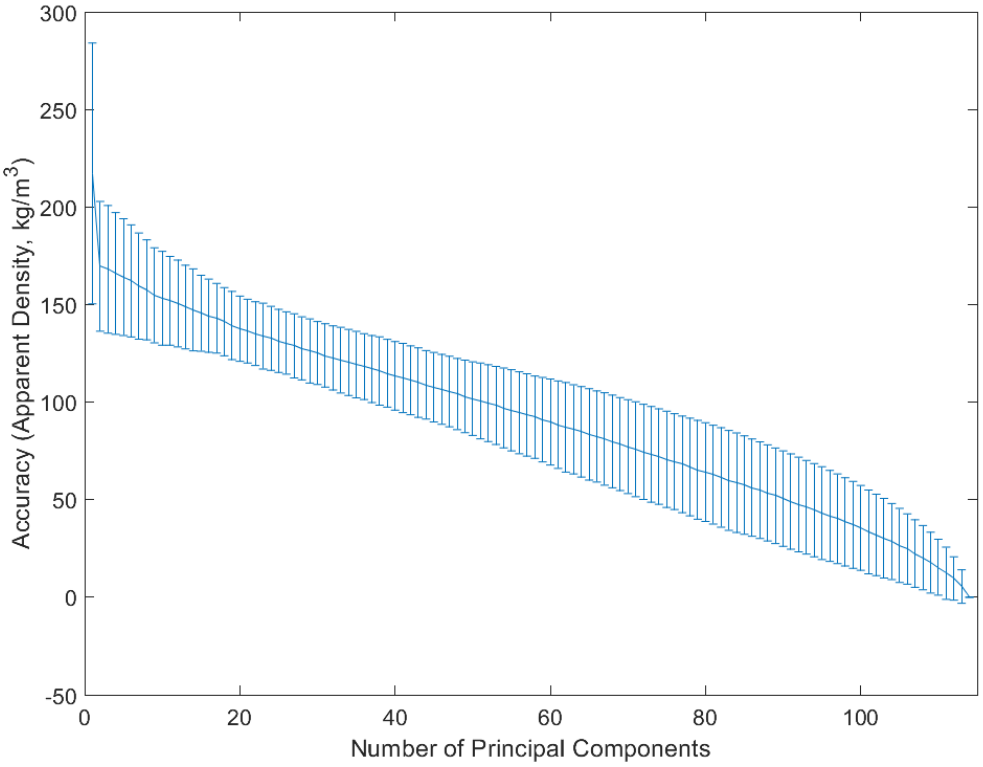

Figure 6-30: Accuracy of the Tibia SIM with the progressive addition of principal components to the model.

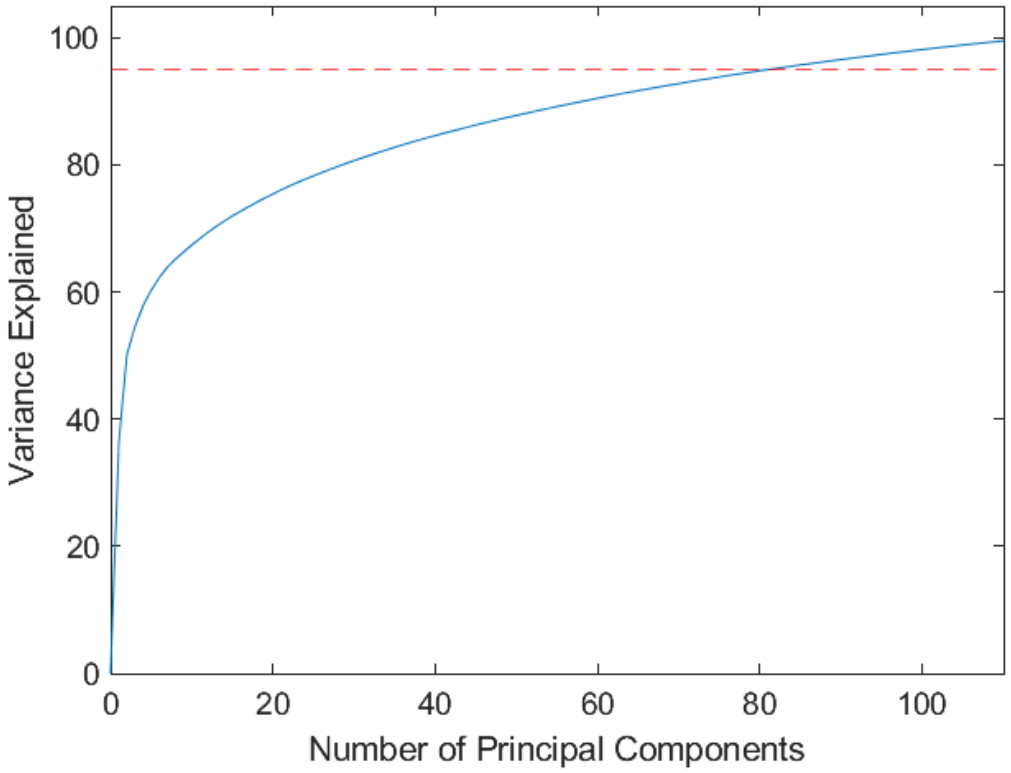

Figure 6-31: Cumulative compactness of the Tibia SIM for increasing number of PCs included.

Table 6-9: Accuracy, variance explained, and cumulative compactness for select PCs of the Tibia SIM. The PC at which the desired compactness threshold is reached is highlighted.

Num PCs	Accuracy (Apparent Density, kg/m^3)	Variance Explained (%)	Cumulative (%)
1	217.20 +/- 66.80	36.10	36.10
2	169.60 +/- 33.23	14.02	50.12
3	168.03 +/- 32.71	4.45	54.58
4	165.90 +/- 31.20	3.29	57.86
5	164.05 +/- 29.91	2.43	60.29
6	162.10 +/- 28.70	2.01	62.30
7	159.45 +/- 27.20	1.63	63.92
8	157.54 +/- 25.61	1.28	65.20
9	154.72 +/- 24.34	1.12	66.32
10	153.23 +/- 24.06	1.10	67.42
20	137.61 +/- 16.71	0.65	75.44
30	125.14 +/- 16.14	0.45	80.68
40	113.48 +/- 17.56	0.35	84.62
50	101.77 +/- 18.85	0.29	87.79
60	89.80 +/- 22.05	0.25	90.48
70	77.21 +/- 24.06	0.22	92.79
80	64.11 +/- 25.29	0.19	94.81
82	61.39 +/- 25.50	0.18	95.18
114	1.40E-12 +/- 6.65E-13	0.11	100.00

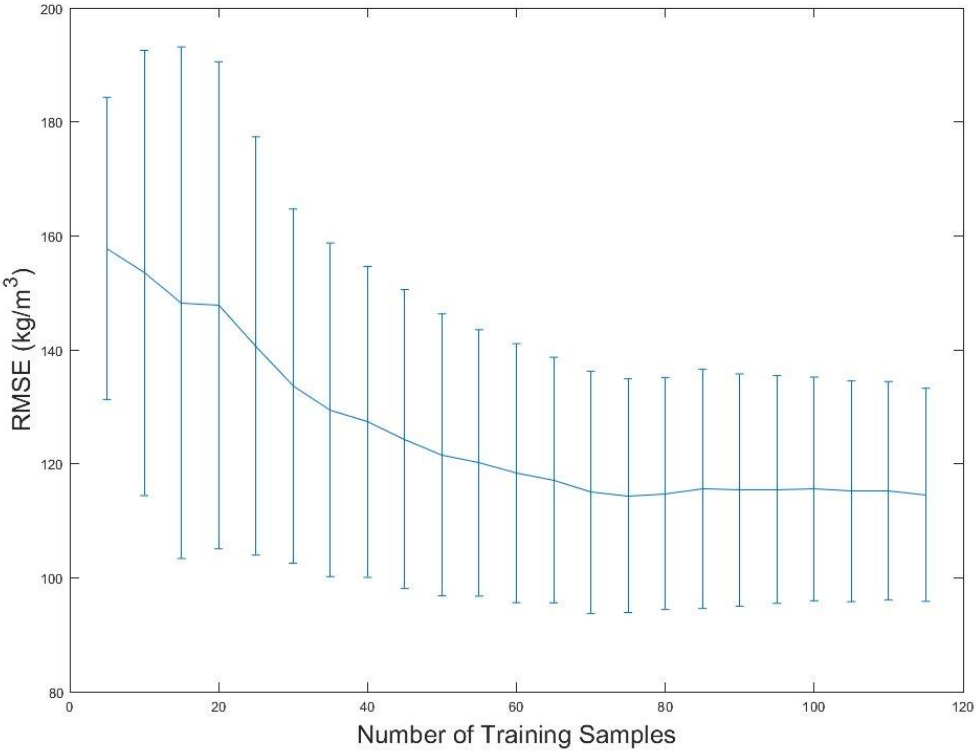

Figure 6-32: Generalizability of the Tibia SIM with the progressive addition of training samples to the register.

Table 6-10: Generalizability and specificity for the Tibia SIM.

Number of Instances	114
Generalizability (Apparent Density, kg/m^3)	114.55 +/- 1870
Specificity (Apparent Density, kg/m^3)	218.18 +/- 20.71

6.3.2.3 Model Outcomes

The first four modes of variation of the Tibia SIM were visualized by perturbing the mean geometry at +/- 3.0 standard deviations and displaying the resulting material properties in select cut planes, shown in Figure 6-33. Mode 1 (Figure 6-34) described an

overall change in bone quality. Mode 2 (Figure 6-35) appeared to capture changes in thickness of the cortical wall, while also potentially capturing the edge artifact due to air surrounding the bone in material property assignment, which can be influenced by the segmentation process. Mode 3 (Figure 6-36) described a change in trabecular density and in the cortical thickness at the tubercle. Finally, Mode 4 (Figure 6-37) captured variation in overall anterior and tubercle cortical density alongside a complementary trabecular variation.

Correlation coefficients between the first ten SSM (from Chapter 5) and SIM modes were calculated, and significant correlations were reported (Table 6-11). As similarly observed in the Femur, none of the correlations between the shape and intensity modes of the tibia were greater than 0.5 or strong, thereby supporting the approach to decouple the SSM and SIM.

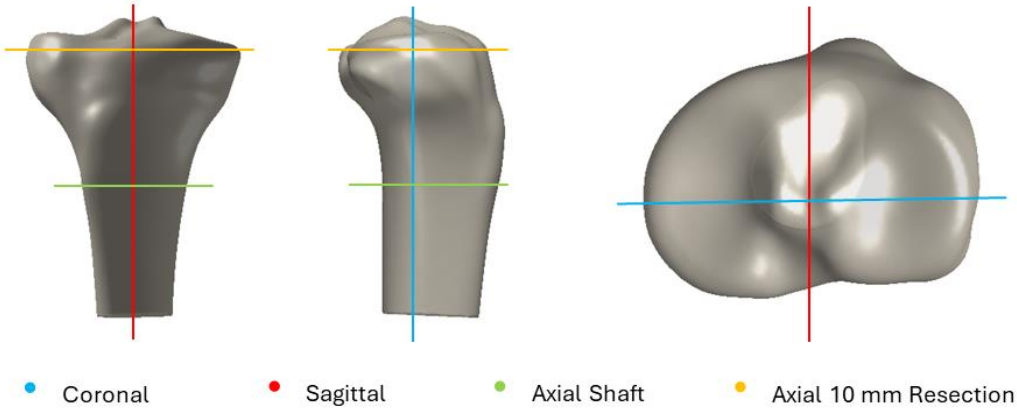

Figure 6-33: Cut planes used to visualize material property variation in the tibia.

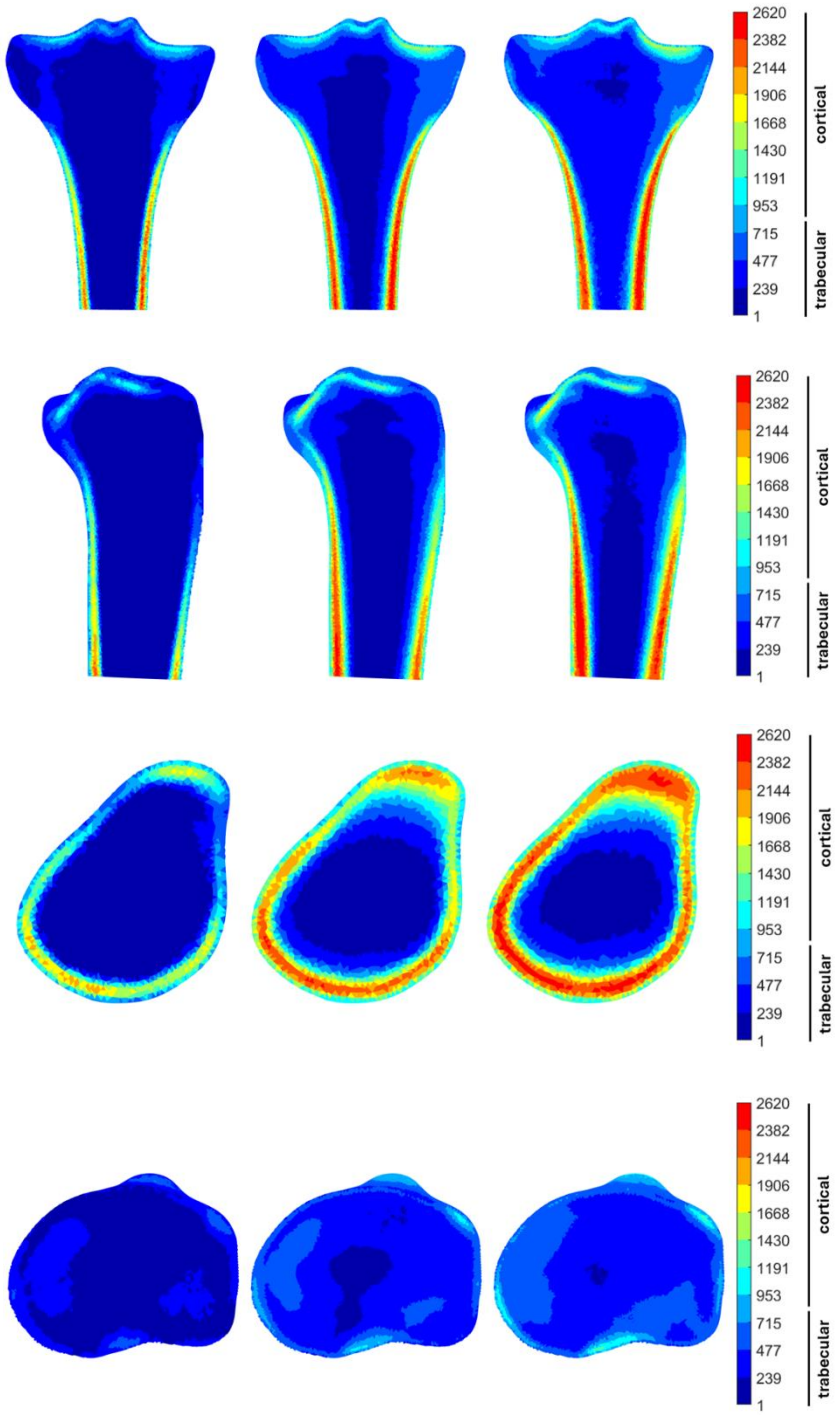

Figure 6-34: Material property variation in Mode 1 of the Tibia SIM. Variations extending to +/- three standard deviations from the mean are displayed along key anatomical planes, including the coronal midline, sagittal midline, transverse shaft, and a clinically relevant transverse resection plane.

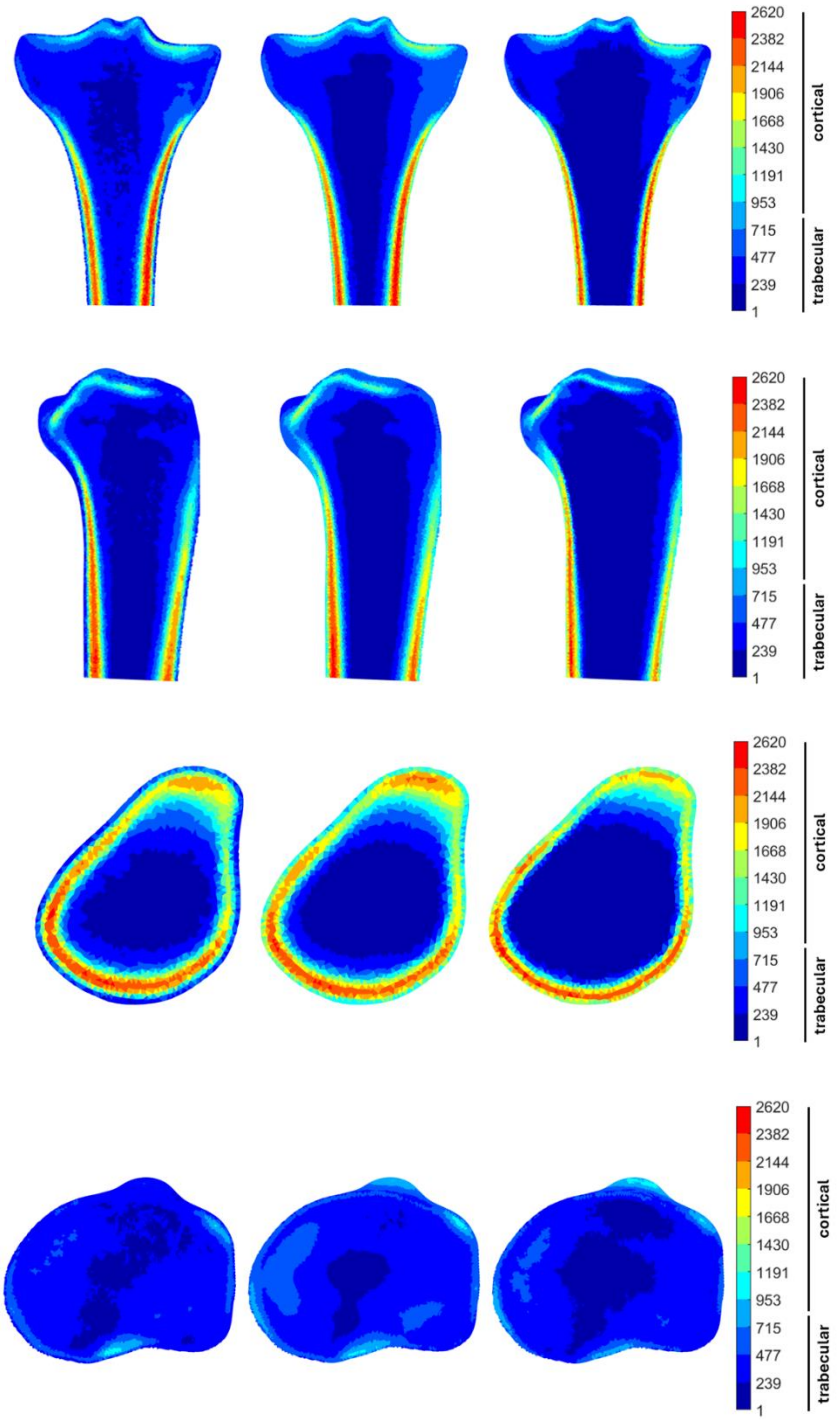

Figure 6-35: Material property variation in Mode 2 of the Tibia SIM. Variations extending to +/- three standard deviations from the mean are displayed along key anatomical planes, including the coronal midline, sagittal midline, transverse shaft, and a clinically relevant transverse resection plane.

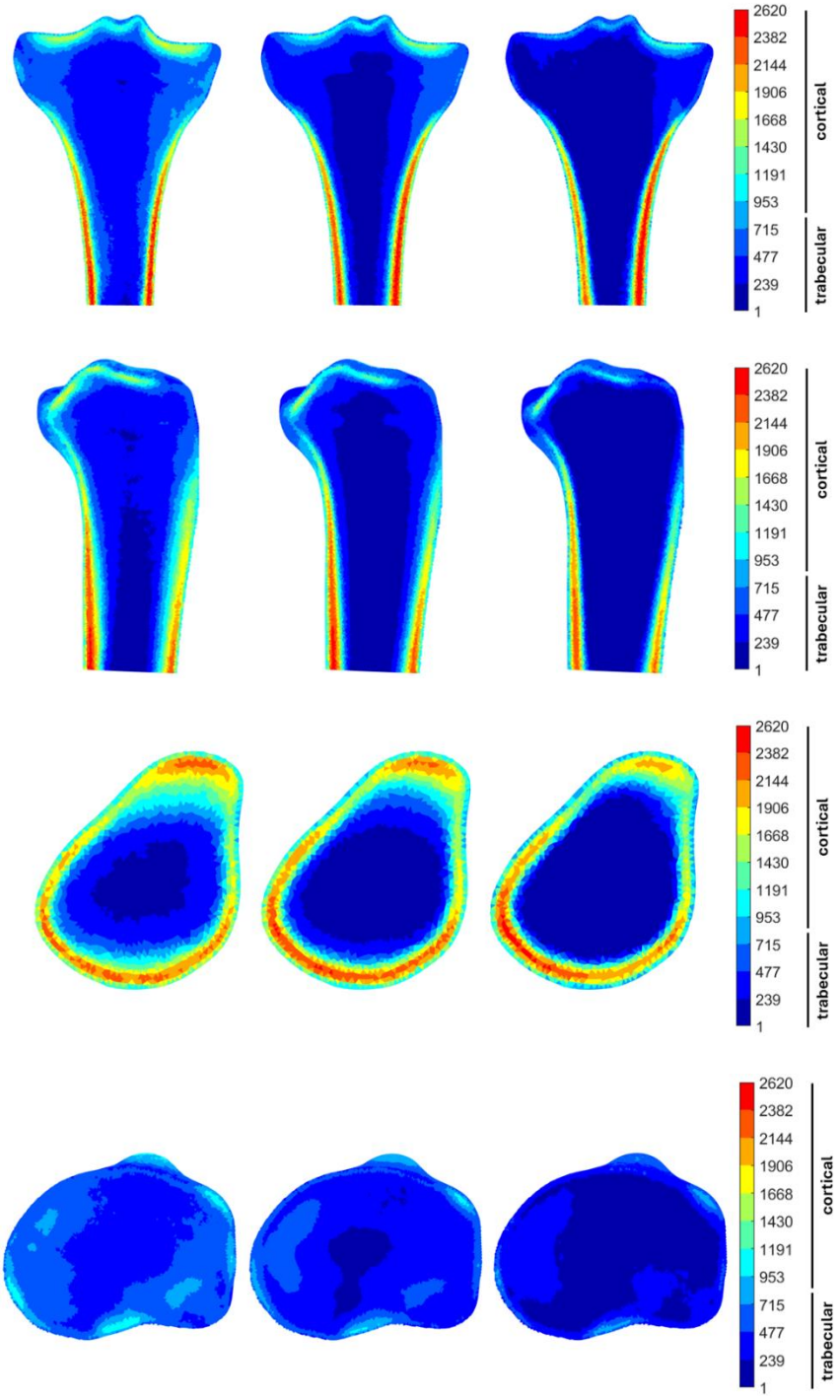

Figure 6-36: Material property variation in Mode 3 of the Tibia SIM. Variations extending to +/- three standard deviations from the mean are displayed along key anatomical planes, including the coronal midline, sagittal midline, transverse shaft, and a clinically relevant resection plane.

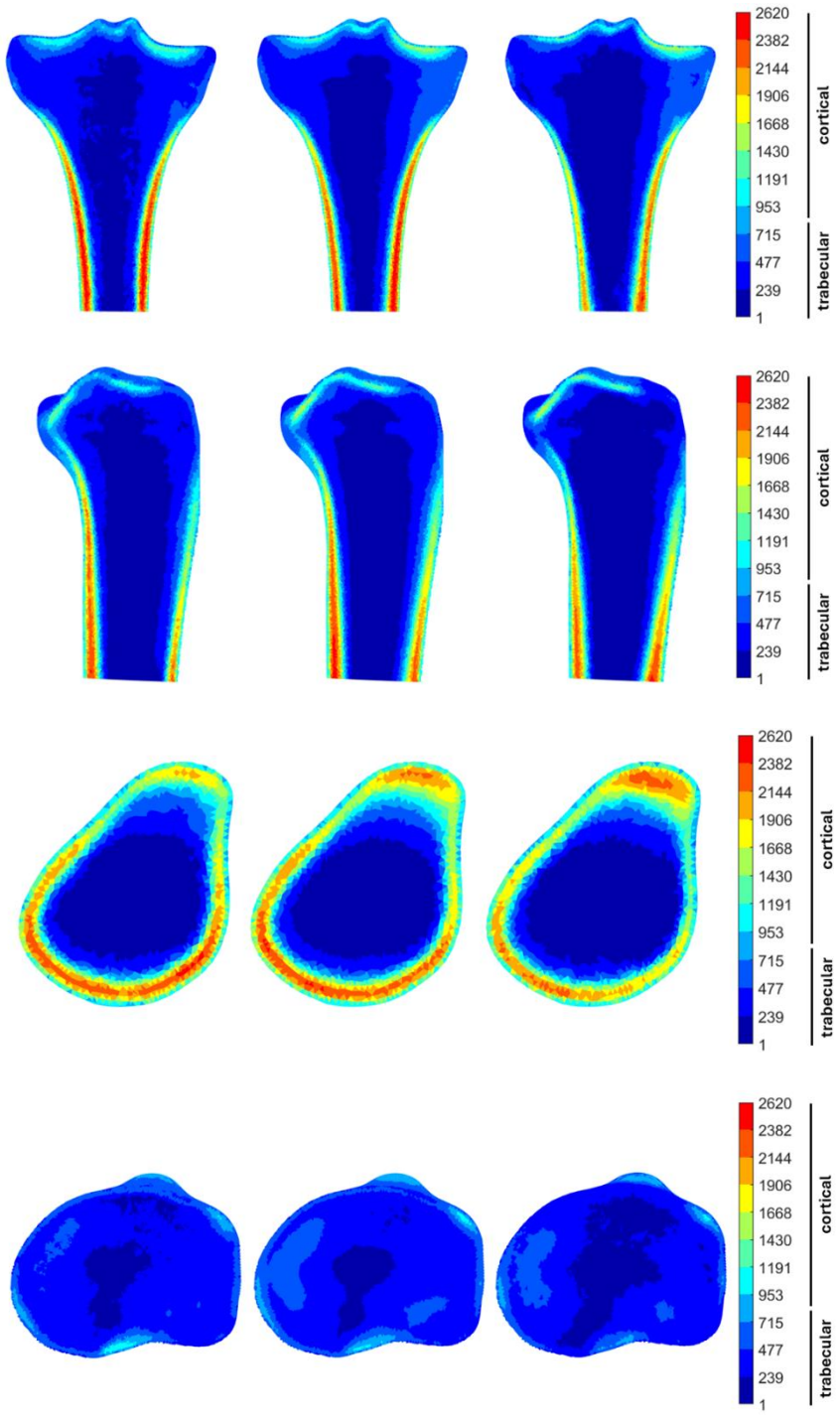

Figure 6-37: Material property variation in Mode 4 of the Tibia SIM. Variations extending to +/- three standard deviations from the mean are displayed along key anatomical planes, including the coronal midline, sagittal midline, transverse shaft, and a clinically relevant transverse resection plane.

Table 6-11: Correlation Coefficients (R) between Tibia SSM and SIM PC scores. Only statistically significant ($p<0.05$) correlations are shown. High correlations ($R> 0.65$) are bolded.

	SSM PC 1	SSM PC 2	SSM PC 3	SSM PC 4	SSM PC 5	SSM PC 6	SSM PC 7	SSM PC 8	SSM PC 9	SSM PC 10	SIM PC 1	SIM PC 2	SIM PC 3	SIM PC 4	SIM PC 5	SIM PC 6	SIM PC 7	SIM PC 8	SIM PC 9	SIM PC 10
SSM PC 1																				
SSM PC 2																-0.19				
SSM PC 3																				
SSM PC 4													0.30		0.26	-0.18				
SSM PC 5																	0.31			
SSM PC 6															0.20					-0.24
SSM PC 7																			-0.20	
SSM PC 8																				
SSM PC 9																0.20				
SSM PC 10											-0.25			0.19	-0.25				0.24	
SIM PC 1										-0.25										
SIM PC 2																				
SIM PC 3				0.30																
SIM PC 4										0.19										
SIM PC 5				0.26		0.20				-0.25										
SIM PC 6		-0.19		-0.18					0.20											
SIM PC 7					0.31															
SIM PC 8																				
SIM PC 9							-0.20			0.24										
SIM PC 10						-0.24														

6.3.2.4 Interpopulation Comparisons

The sex and racial distribution of the scores pertaining to the first four principal components are depicted in Figure 6-38 (PC1 and PC2) and Figure 6-39 (PC3 and PC4). Similar to the femur results, differences were observed in the scatter of PC 2 (Figure 6-38) for the White population; this population was derived from a cadaveric dataset at DU instead of from the NMDID database and the observed differences may be due to the scan source and segmentation process.

To quantify interpopulation differences in PC scores and morphological parameters, unpaired two-sample t-tests and a series of ANOVA tests were conducted. Table 6-12

displays the p-values from t-tests with significant differences ($p < 0.05$) between two populations. Boxplots illustrating the distribution of PCs one through four across sex, race, and age categories are shown in Figures 6-30 to 6-41. Significant differences ($p < 0.05$) in these plots are indicated by a red bar. Table 6-13 displays significant p values ($p < 0.05$) resulting from a series of ANOVA tests for the first ten PCs considering factors of scan source, sex, race, age, height, and BMI.

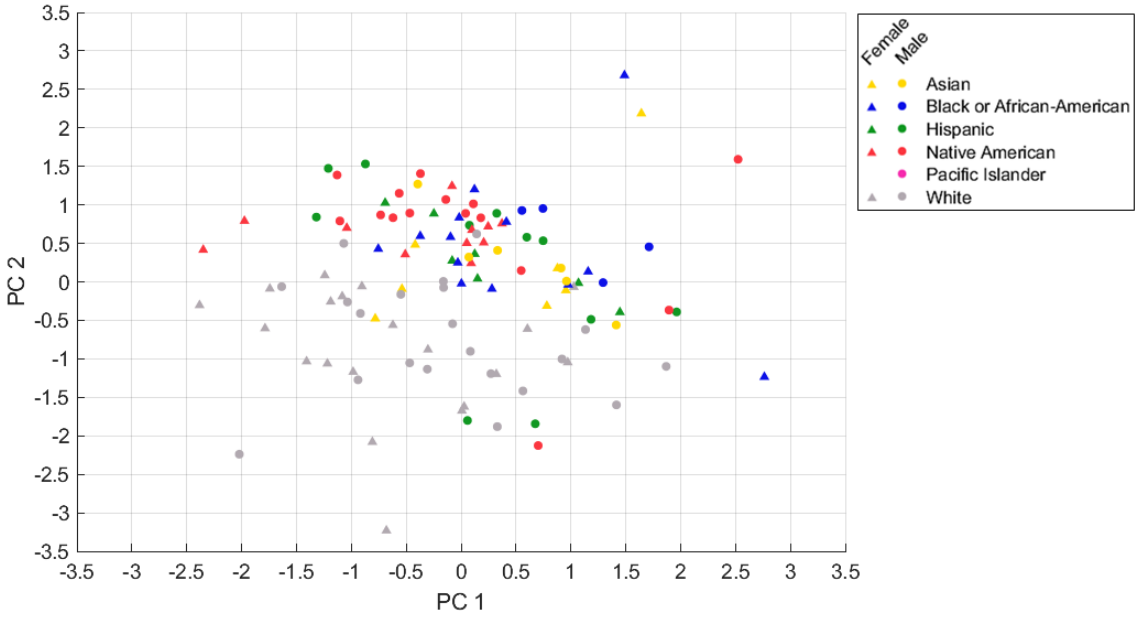

Figure 6-38: Sex and racial distribution of PC 1 and PC 2 scores of the Tibia SIM.

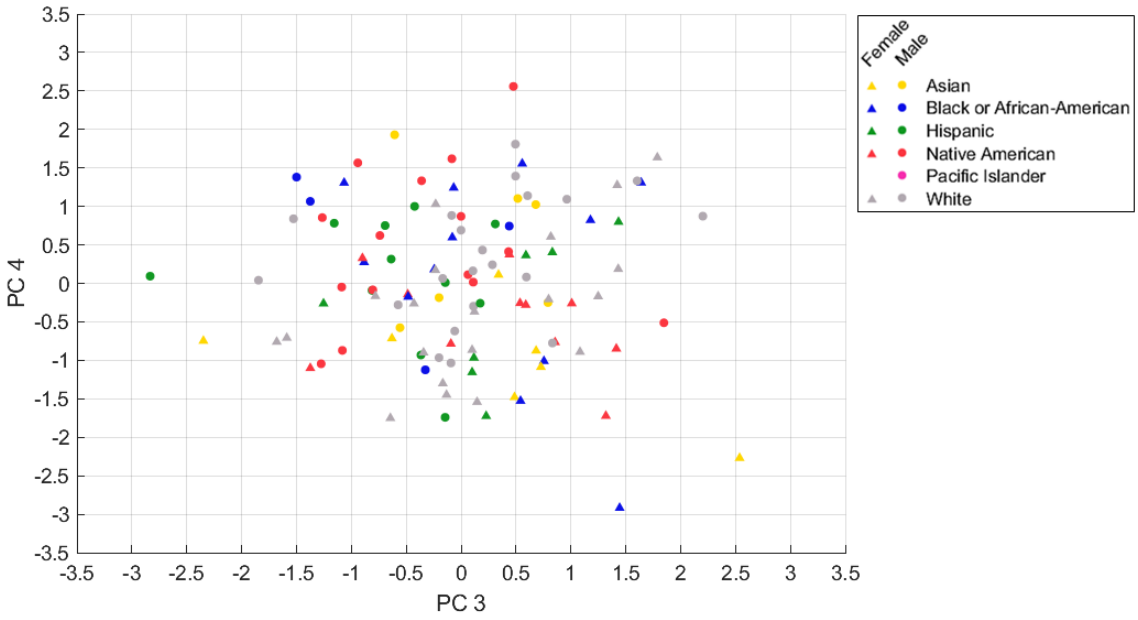

Figure 6-39: Sex and racial distribution of PC 3 and PC 4 scores of the Tibia SIM.

Table 6-12: P-values derived from unpaired two-sample t-tests indicating significant differences in PC scores one to ten ($p < 0.05$) between populations. Highly significant differences ($p < 0.001$) are shown in bold.

	M/F	A/B	A/H	A/NA	A/W	B/H	B/NA	B/W	H/NA	H/W	NA/W	YA/MA	YA/OA	MA/OA
PC 1					7.30E-03		1.54E-02	**6.67E-04**		2.75E-02		2.07E-02	**4.73E-06**	2.33E-03
PC 2					**5.11E-05**			**3.26E-07**		**3.89E-05**	**5.26E-11**		**2.35E-06**	**2.74E-06**
PC 3												**6.77E-04**	1.12E-02	
PC 4	**9.78E-05**											1.36E-03		1.73E-02
PC 5	4.76E-02						4.61E-02							
PC 6					2.69E-02		2.60E-02				1.11E-02			
PC 7	**2.27E-04**							3.23E-03			1.77E-02			1.08E-02
PC 8	1.24E-03											1.09E-02		2.05E-02
PC 9												1.44E-02	5.49E-03	
PC 10														

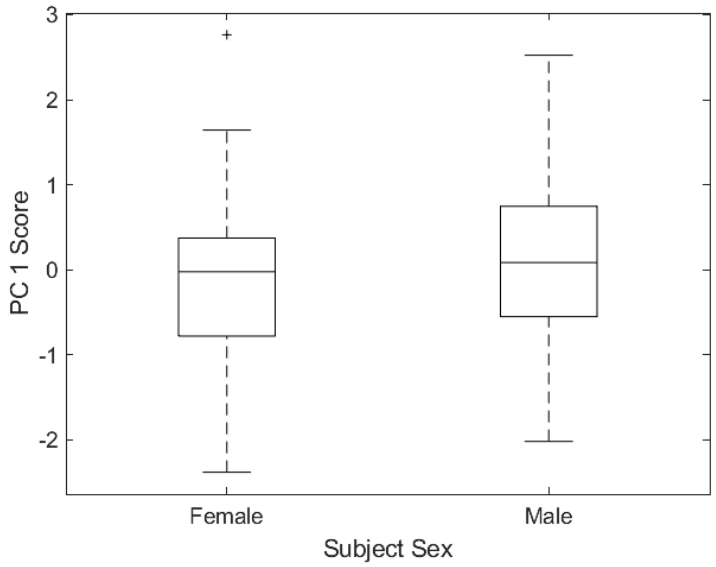

Figure 6-40: Tibia SIM PC 1 score by subject sex.

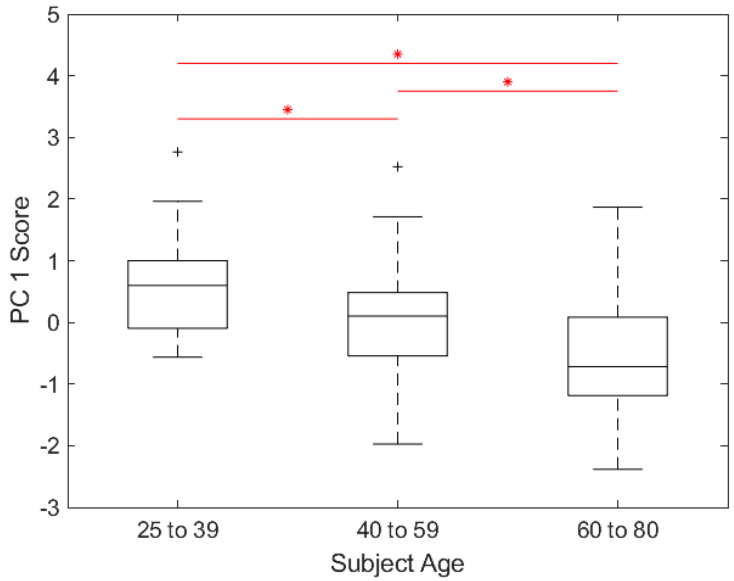

Figure 6-41: Tibia SIM PC 1 score by subject age.

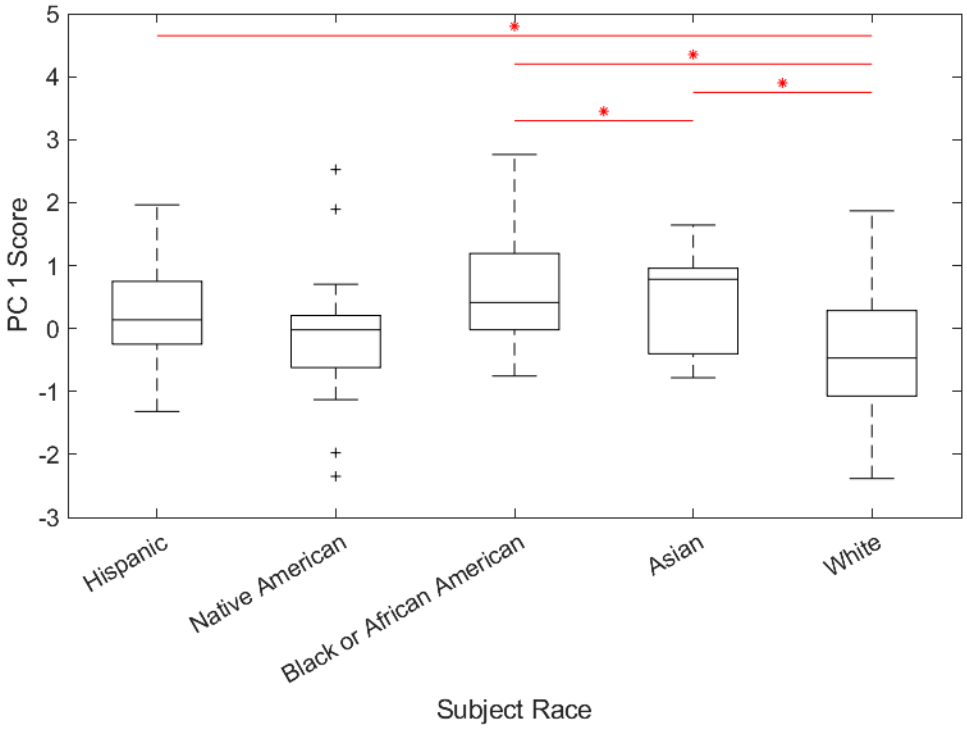

Figure 6-42: Tibia SIM PC 1 score by subject race.

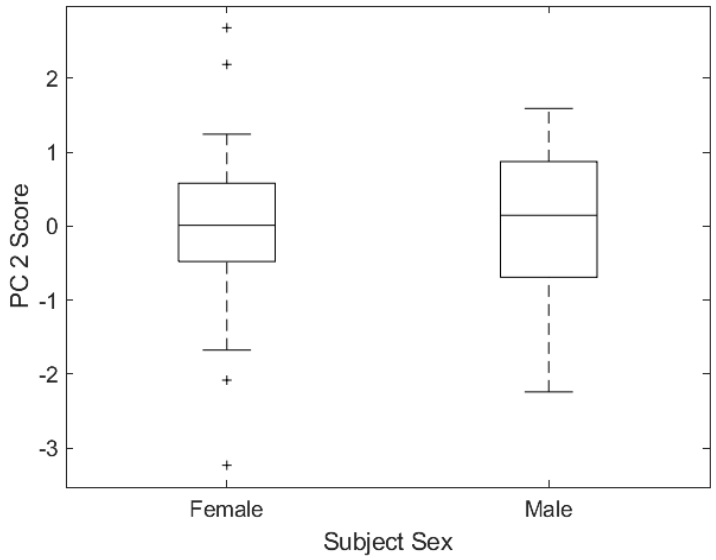

Figure 6-43: Tibia SIM PC 2 score by subject sex.

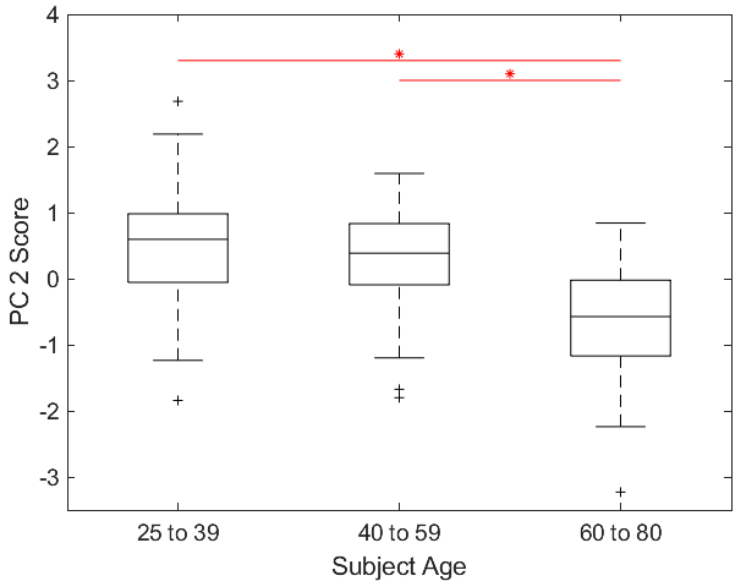

Figure 6-44: Tibia SIM PC 2 score by subject age.

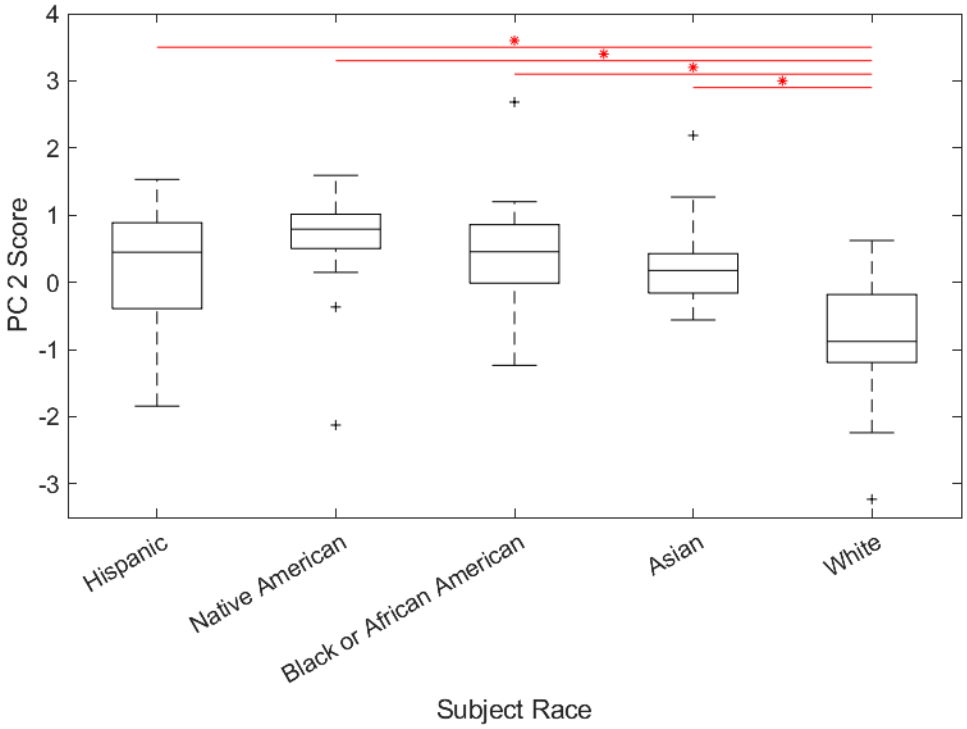

Figure 6-45: Tibia SIM PC 2 score by subject race.

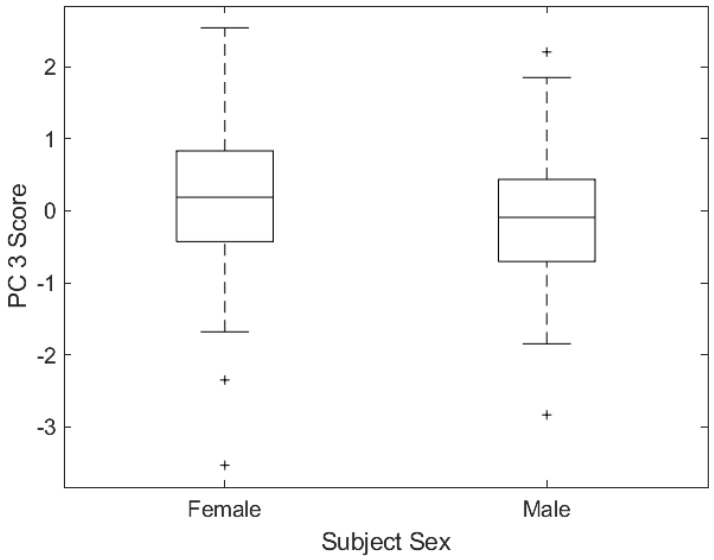

Figure 6-46: Tibia SIM PC 3 score by subject sex.

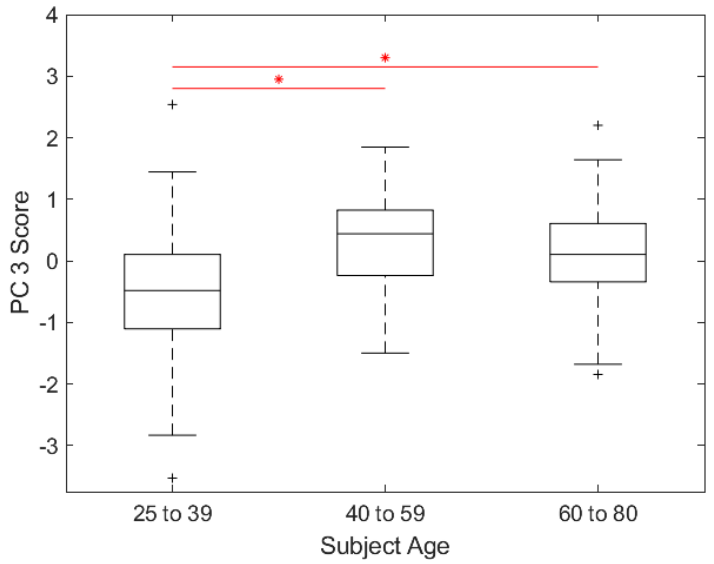

Figure 6-47: Tibia SIM PC 3 score by subject age.

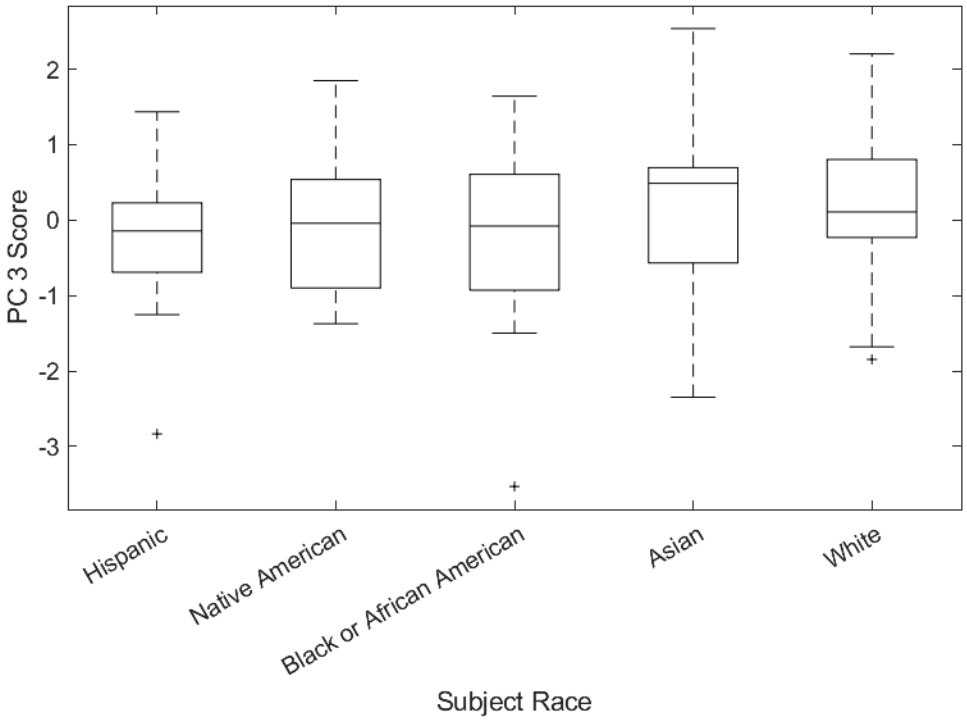

Figure 6-48: Tibia SIM PC 3 score by subject race.

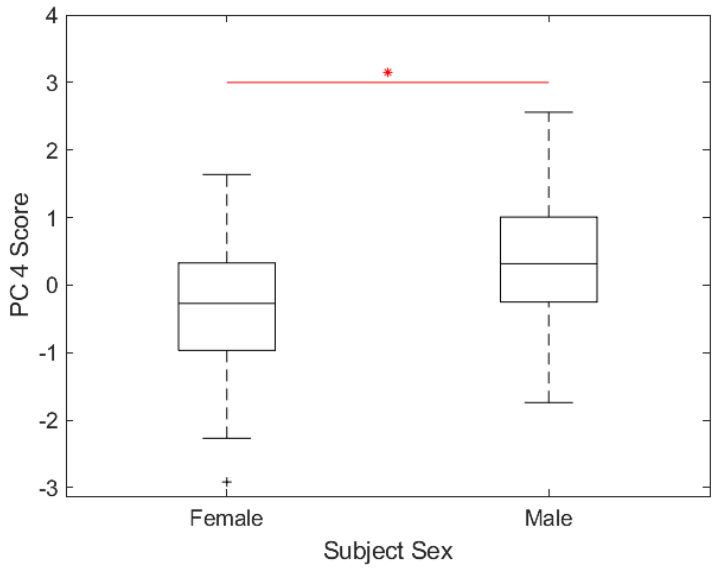

Figure 6-49: Tibia SIM PC 4 score by subject sex.

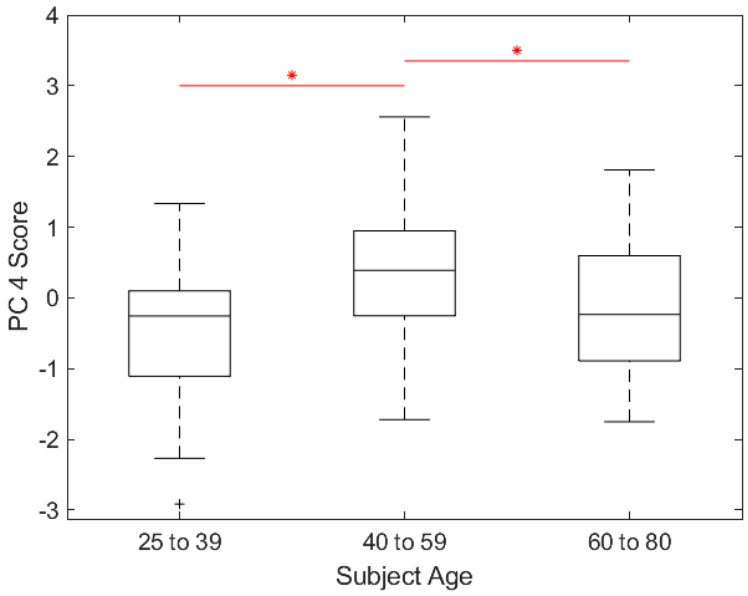

Figure 6-50: Tibia SIM PC 4 score by subject age.

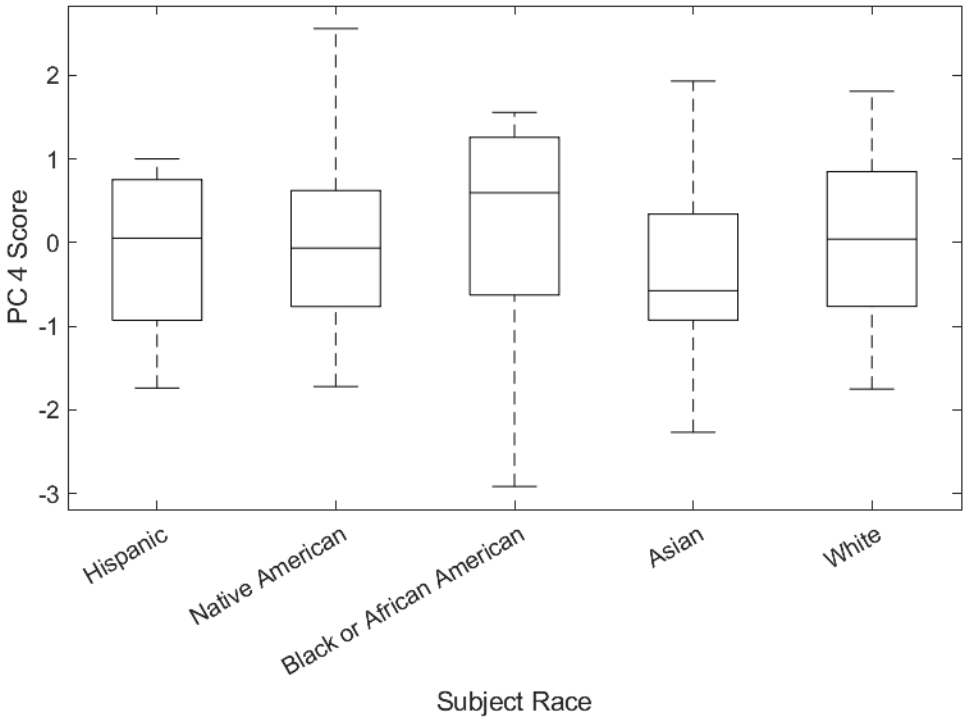

Figure 6-51: Tibia SIM PC 4 score by subject race.

Table 6-13: ANOVA-derived p-values indicating significant differences ($p < 0.05$) between populations. Highly significant differences ($p < 0.001$) are shown in bold.

	Sex	Age	Race	Height	BMI	Source
PC 1	4.24E-03	**1.64E-04**	4.44E-03			
PC 2				7.46E-03		**3.47E-05**
PC 3				4.37E-03		
PC 4	7.42E-03	1.13E-02			1.83E-02	1.36E-02
PC 5	1.39E-02	2.80E-02				
PC 6				4.11E-02		
PC 7						
PC 8				3.56E-02		
PC 9						
PC 10						

6.4 Discussion

This work importantly quantifies the variability in bone property distribution for a population considering diversity in gender, race and ethnicity, and age. The quality of the statistical model can be evaluated in the context of four metrics: accuracy, compactness, generalizability, and specificity. Material property values assigned to the bones were physiologic and comparable to commonly reported values for the apparent density of cortical and cancellous bone (Szabo and Rimnac 2022; B. Li and Aspden 1997). With all PCs included, the model recreated register subjects with a sub-unit accuracy of 1.34E-12 +/- 5.88E-13 kg/m^3 (Femur) and 1.40E-12 +/- 6.65E-13 kg/m^3 (tibia). Unlike the SSM models describing shape, neither of the SIM models was compact (Figure 6-6 and Figure 6-31), requiring 79 (femur) and 82 (tibia) PCs to account for 95% of the variance in the model. This could indicate the model "overfitted" the data, creating intensity modes based on noise. Generalization accuracy was assessed through a leave-one-out analysis, which evaluated how well the model represented other unseen bone maps in the same class. With 111 femur and 115 tibia subjects, the generalization accuracy was 101.81 +/- 16.71 kg/m^3 for the Femur SIM and 114.55 +/- 18.70 kg/m^3 for the Tibia SIM. Given that the reconstruction error did not converge with the full set of training samples in either the Femur (Figure 6-7) or Tibia (Figure 6-32) SIM, and that the number of subjects included in each model is below the 200 subjects required for the convergence of shape generalizability (Audenaert et al. 2019), it is likely that the addition of more subjects to the training set would improve the generalizability of these models. Both models demonstrated the ability to generate realistic new material property maps with a

specificity of 180.49 +/-19.69 kg/m^3 for the Femur SIM and 218.18 +/-20.71 kg/m^3 for the Tibia SIM.

The first principal component encapsulates an overall change in bone quality across both the femur (Figure 6-9) and tibia (Figure 6-34). Interestingly, neither the t-tests or ANOVA found significant disparities in PC 1 scores between males and females in the Tibia SIM (Figure 6-40). This observation supports the idea that gender-related distinctions in bone mass diminish after adjusting for body size (Nieves et al. 2005). While no t-test differences were found in PC 1 scores between the sexes in the Femur SIM (Figure 6-15), ANOVA did identify significant differences in this mode of variation between males and females (Table 6-7). Significant differences ($p < 0.05$) were detected in PC1 between White subjects and those of Hispanic, Black or African American, and Asian descent in both the Femur and Tibia SIMs, with White subjects having a lower average overall bone quality (Figure 6-17 and Figure 6-42). Additionally, differences in PC 1 were identified between Black or African American and Asian subjects in the Tibia SIM (Figure 6-42), as well as between Native American and Asian subjects in the Femur SIM (Figure 6-17). Inter-race differences in PC 1 were also identified by ANOVA in both the Femur (Table 6-7) and Tibia (Table 6-13) SIMs. Both t-tests and ANOVA identified significant differences in Mode 1 between age groups in the Femur (t-test, young age/old age, middle age/old age) and the Tibia (t-test, all age groups). This may be related to the overall decrease in bone density expected as we age (Riggs et al. 2004). Similarly, t-test analysis of PC 3, primarily associated with trabecular bone density in both Femur (Figure 6-11) and Tibia (Figure 6-36) SIMs, revealed notable differences

between young and old cohorts (Figure 6-22 and Figure 6-47). These findings were confirmed by ANOVA in the femur (Table 6-7), but not in tibia (Table 6-13). This is in line with the continual loss of trabecular volumetric bone density observed throughout life (Riggs et al. 2004). Lastly, PC 4, related to overall cortical thickness in the Femur SIM (Figure 6-12) and anterior cortical thickness in the Tibia SIM (Figure 6-37), exhibited significant disparities between males and females, with females having greater overall cortical thickness in the femur (Figure 6-24) and males having greater anterior cortical thickness in the tibia (Figure 6-49). ANOVA confirmed the presence of sex related differences in PC 4 of both models (Table 5-8 and Table 5-15). T-test distinctions were also apparent between White cohorts and both Black or African American and Asian cohorts in this mode for the Femur SIM (Figure 6-26). Race-related PC 4 differences were not confirmed found using ANOVA (Table 5-8 and Table 5-15).

It is important to note that the training set of these models draws from two distinct data sources, namely the NMDID and a prior DU SSM dataset. While consistency is maintained in the methods within each group, differences exist between them. Notably, subjects from the previous SSM dataset underwent calibration using a phantom, while NMDID subjects were calibrated using a phantomless method. Studies comparing phantom calibration with phantomless methods have demonstrated comparable results (Eggermont et al. 2019), so it is expected that this should not have a significant effect on the results. Accordingly, the distribution of the subjects in the PC1 (overall bone quality) scatter plots, where one would expect calibration-related differences to manifest, does not indicate a strong difference between the phantom and phantomless calibration subjects

(Figure 6-10 and Figure 6-35), although significant ($p<0.05$) differences were found between White (DU) subjects and subjects from other racial or ethnic backgrounds (NMDID). The differences in PC 1 could also be due to a notable age discrepancy between the two subject sets, as subjects from the previous SSM dataset tended to be older on average compared to those from the NMDID (Table 6-2 and Table 6-8).

Moreover, segmentation quality may vary between these groups due to potential human error. This discrepancy is likely evident in the distribution of PC 2 scores, which described cortical wall thickness but may be impacted by the edge artifact when differentiating bone and surrounding air (Figure 6-10 and Figure 6-35). Previous SSM subjects tend to exhibit more air in their material property maps compared to NMDID subjects, leading to an uneven distribution of PC2 scores (Figure 6-13 and Figure 6-38). In addition to segmentation quality, these inconsistencies could also be due to differences in scan parameters, although the slice thickness and pixel size in both groups of scans were comparable. Since the White subject group exclusively comprises individuals from the previous SSM dataset, the calibration method, age, and segmentation quality represent confounding factors with White racial identity.

Our training set includes a broader range of demographics than most studies of this type and is the only known model to the authors that includes Native American and Hispanic subjects. However, when considering the US population, the lack of Native Hawaiian/Pacific Islander and ethnically middle eastern populations is a limitation. It is also crucial to recognize that racial identities are socially constructed and lack a biological basis (Cwalina et al. 2022). Without including detailed genetic data from

rigorous genomic testing or accounting for potential confounding factors such as socioeconomic status, our results cannot definitively attribute the observed differences in our models to race. Nevertheless, race remains a significant determinant of access to societal resources and barriers to full inclusion, as with the presence of racial and ethnic disparities in TKA. The fact of inequality means that race and ethnicity continue to matter in important ways (Committee on Improving the Representation of Women and Underrepresented Minorities in Clinical Trials and Research 2022).

Furthermore, while we observed statistically significant differences in the bone material properties of the distal femur and proximal tibia, our intra-group sample sizes average about 20 subjects per racial background, which is insufficient to fully capture the variation within each demographic. In the future, a more robust training set will be needed to investigate differences in the knee with a larger training set to comprehensively capture population variation.

In closing, significant differences in bone material property modes of variation were identified between sex, racial and ethnic, and age groups. Further work with a larger training set is needed to improve the quality of the model and the robustness of the results.

7. Application: SSIM Graphical User Interface

Quantifying bone shape is essential for assessing the fit, coverage, and sizing of implant designs. Likewise, accurate representations of bone material properties are crucial for computational evaluations of initial fixation and micromotion in bone-implant constructs. Given the ability of SSIMs to do both, medical device developers are keenly interested in tools that seamlessly integrate the outcomes of these models into their development workflows.

Previously, a Graphical User Interface (GUI)-based model interface was developed to streamline the creation of FEA-ready tibia and femur meshes using SSM outcomes. This tool effectively addressed questions concerning bone shape, enabling the generation of bones representative of various population subsets and of "worst-case" scenarios through user-defined mode adjustments. However, this GUI left crucial inquiries about bone quality variation and its impact on bone-implant systems unanswered. Accordingly, a new GUI-based model interface was created to apply material property maps to shape model instances, building on the previous interface to support consideration of anatomic variation in implant design evaluation.

The SIM portion of the GUI utilized the GRNN-based morphing algorithm developed by Andreassen et al. to register a tetrahedral mesh to user-defined shape instances. Pre-defined material properties could then be assigned to these registered tetrahedral meshes

and exported as ready-to-use files for FEA. The main panel of the SSIM GUI is shown in Figure 7-1.

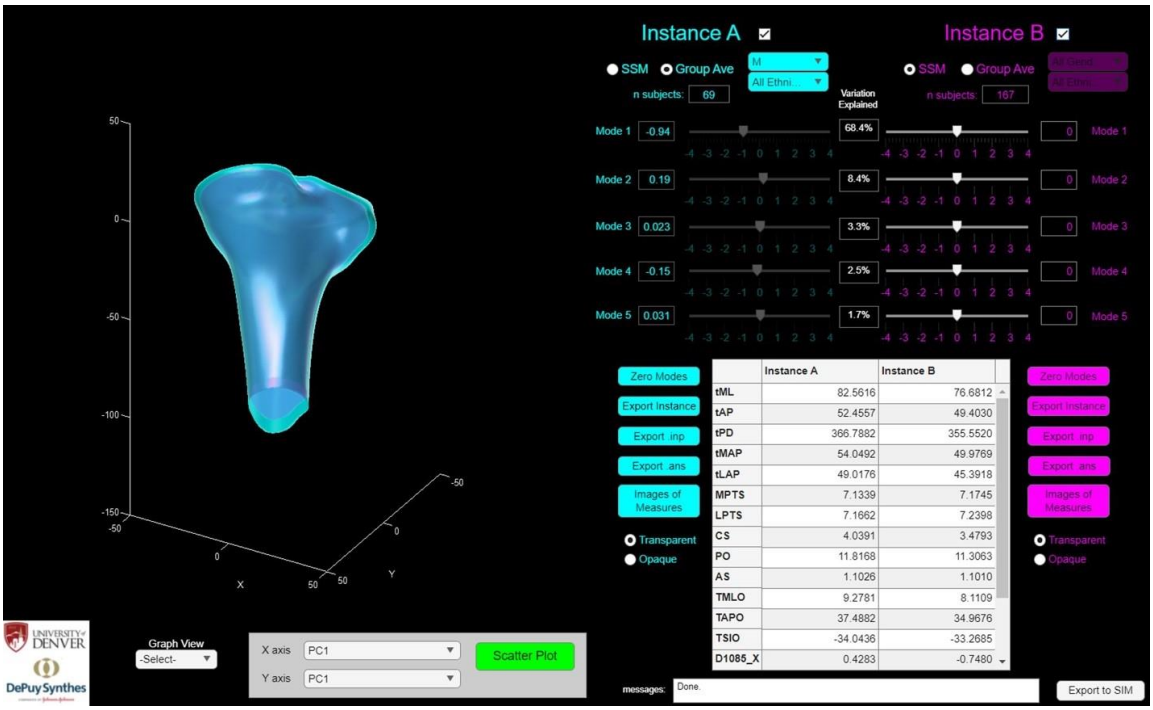

Figure 7-1: Main panel of the SSIM GUI.

From this panel, users create new instances by manually adjusting SSM modes or parsing populations based on gender and ethnicity. These new instances can be exported as STLs or as job files for Abaqus (.inp) or ANSYS (.ans). The new SIM subpanel is accessible through the "Export to SIM" button in the bottom right of the main panel of the DU SSIM. Upon pressing the button, the SIM subpanel appears, shown in Figure 7-2.

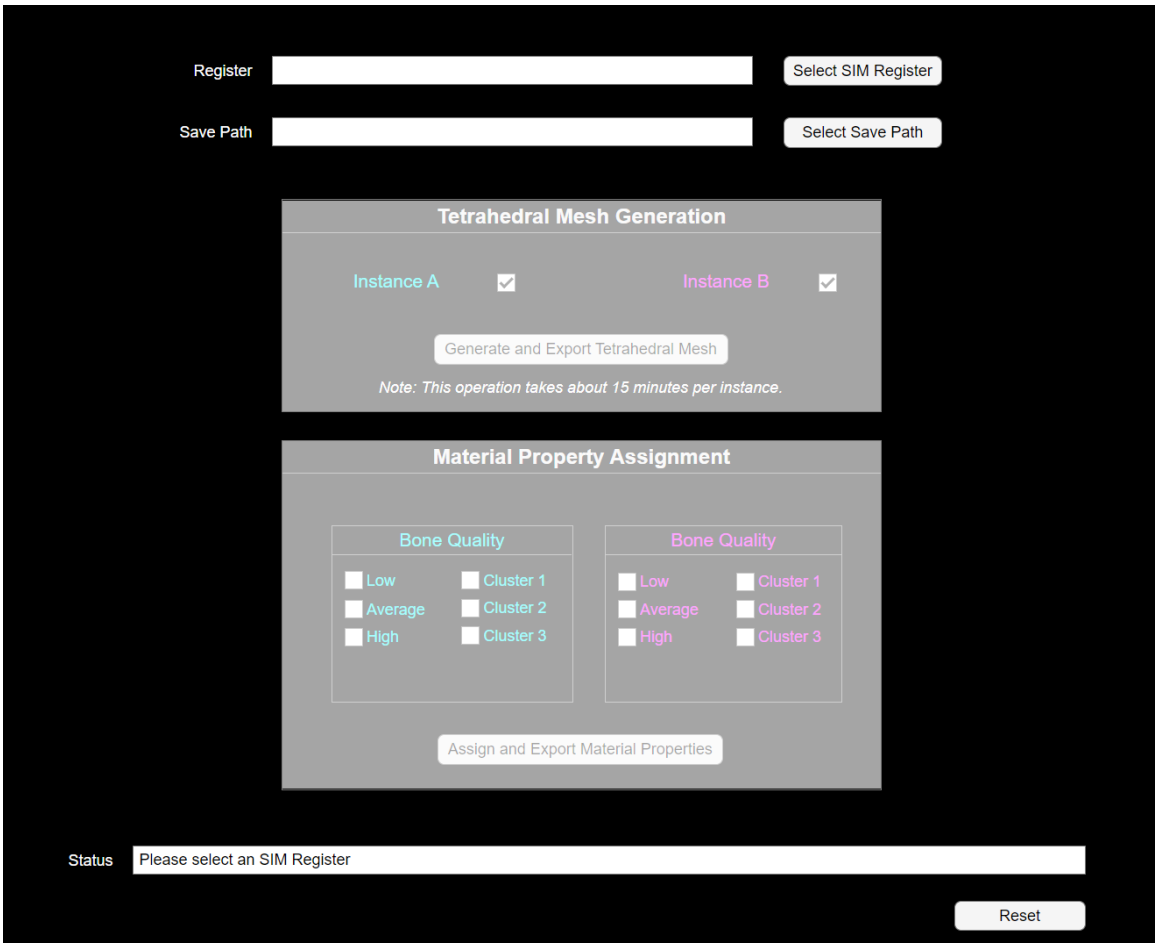

Figure 7-2: The SIM subpanel of the SSIM GUI, which is used to assign material properties. Portions of the panel are grayed out and become accessible as the user progresses through the process.

The SIM subpanel simplifies the complex process of generating a tetrahedral mesh and assigning material properties by guiding users through the process via the "Status" bar. The GUI assists users with SIM File Selection, Save Path Selection, Tetrahedral Mesh Generation, and Material Property Assignment. Since SIM modes are less intuitive than SSM modes, users are provided with six and seven material property maps for the Tibia and Femur, respectively, categorized into three groups:

1) AVERAGE: Represents the average bone quality from the SIM training set.

2) LOW and HIGH: Represents instances with the lowest and highest bone quality in the SIM training set.
3) CLUSTERS: Based on a k-means cluster analysis of the first three SIM modes. While changes in SSM modes are relatively intuitive, pre-defined material maps provided an intuitive alternative to the not-so-intuitive process of toggling SIM modes.

This tool was developed in partnership with DePuy Synthes, a leading orthopaedic implant manufacturer. It exemplifies the practical application and relevance of SSIMs in orthopedic implant design, underscoring the role these models can play in enhancing implant design, and ultimately, the outcomes of TKA and UKA for a diverse patient population.

8. Concluding Remarks

8.1 Future Work

While the SSIM developed in this project represents a step towards inclusion in orthopaedic devices, the work is far from complete. First, the models are limited by the size of their training set, with an average of only 20 subjects per racial or ethnic population included. Additionally, the model included only one subject from Pacific Islander/Native Hawaiian ancestry and none from ethnically middle eastern backgrounds. These factors affect the generalizability of the models, meaning that they cannot accurately represent population variation. Ideally, as per metrics set by Audenaert et al., the training set should include 200 instances per population. Achieving this is challenging due to the time-consuming nature of the registration pipeline. However, the NMDID database plans to develop a database of segmented subjects, which could streamline the database-to-register workflow.

In the current work, statistical analyses were performed using a series of t-test comparisons. While this approach is likely appropriate for the comparisons of gender with two groups, an Analysis of Variation (ANOVA)-based approach (Kim 2014) would be better when evaluating the statistical differences between the multiple racial and ethnic and age groups. Therefore, further exploration of the ANOVA results is needed to draw credible conclusions from either the SSM or SIM work.

Additionally, unexpected issues arose when incorporating data from the previously developed SSIM. Several subjects had segmentation issues, such as oversmoothing or undersmoothing, while others showed moderate to severe osteoarthritis, including the presence of osteophytes. While edge artifacts are a known challenge when creating material property maps, the quality of segmentation of these prior subjects could be improved. This problem could be addressed by re-segmenting a subset of the models or incorporating White subjects from the NMDID database if the current models are too osteoarthritic to segment successfully.

Once robustly constructed, it would be interesting to evaluate whether the anthropometric differences identified in the models are meaningful for the fit of modern TKA systems. For example, virtual implantations could be conducted to assess overhang and underhang, helping to assess the sufficiency of the implant sizing range. The SSM models could also be used to investigate new or personalized implant designs by predicting the knee kinematics associated with the different radii of curvature identified by the SSM to see if they yield any significant differences in the knee mechanics TKA systems aim to replicate.

In the current work, we developed individual SSM modes of the femur and tibia. We recognize the development of joint-level models that consider the relative alignment of multiple structures could also be valuable. The as-scanned position of the limb in the NMDID dataset was not well controlled, making it difficult to identify the lower limb alignment, e.g., the hip-knee-ankle (HKA) angle. However, aligning the joint based on

anatomical landmarks (Wang et al. 2023; Gielis et al., n.d.) could offer additional insights, particularly for surgical planning applications.

In addition to the HKA angle, other morphological parameters relevant to Total Knee Arthroplasty (TKA) pre-planning and implantation, such as the angle formed by a line connecting the most distal points on the femur condyles to the mechanical axis in the coronal plane or a metric for the angle of the femoral shaft in the coronal or sagittal planes, could be considered. Incorporating these parameters may contribute to a more comprehensive understanding of the modes of variation in the SSMs.

While the new GRNN-based registration workflow has been validated in the construction of these models, the selected parameters tuned to register some bones were not robust enough to morph all instances in the training set. Parameters were selected considering morphing accuracy and also computation time. Although a user familiar with the program can mitigate this issue, it would be preferable for the algorithm to function without intervention for all bones. This may require further refinement of the parameters, the algorithm, or a hybrid approach where CPD deforms the triangular mesh and a GRNN morphs the tetrahedral mesh based on the CPD morph.

Finally, the GUI based on the new GRNN morphing process successfully streamlines the process of creating FEA-ready bone models, its lengthy processing time makes it less attractive to users. This limitation could be addressed by tweaking the GRNN morphing parameters or adopting an alternative tetrahedral shape statistical model approach. In this approach the files imported into the app would contain a register of tetrahedral meshes matching that of the tri register. Constructing a new tetrahedral mesh for material

property assignment would no longer involve morphing a tetrahedral template onto a triangular instance. Rather, the tetrahedral mesh could be constructed as a new instance of the tetrahedral shape model, a rapid and computationally inexpensive process.

8.2 Concluding Remarks

The SSIM developed in this project represents a significant advancement in personalized care and inclusivity for orthopedic medical devices. Notably, it is the first model to include Native American and Hispanic subjects and compare these groups with three other racial and ethnic backgrounds. Although the current training set is not robust enough to draw definitive conclusions about differences between these groups, or among sex and age groups, this project successfully validated an NMDID database-to-register workflow that can be used to construct a more comprehensive model in the future

References

Aitchison, G.A, D.W.L Hukins, J.J Parry, D.E.T Shepherd, and S.G Trotman. 2009. "A Review of the Design Process for Implantable Orthopedic Medical Devices." *The Open Biomedical Engineering Journal* 3 (July):21–27. https://doi.org/10.2174/1874120700903010021.

Albrektsson, T., P.-I. Brånemark, H.-A. Hansson, and J. Lindström. 1981. "Osseointegrated Titanium Implants: *Requirements for Ensuring a Long-Lasting, Direct Bone-to-Implant Anchorage in Man*." *Acta Orthopaedica Scandinavica* 52 (2): 155–70. https://doi.org/10.3109/17453678108991776.

Alvarenga, J.C., V.F. Caparbo, D.S. Domiciano, and R.M.R. Pereira. 2022. "Age-Related Reference Data of Bone Microarchitecture, Volumetric Bone Density, and Bone Strength Parameters in a Population of Healthy Brazilian Men: An HR-pQCT Study." *Osteoporosis International* 33 (6): 1309–21. https://doi.org/10.1007/s00198-021-06288-5.

Anderson, Kyle D., Frank C. Ko, Amarjit S. Virdi, D. Rick Sumner, and Ryan D. Ross. 2020. "Biomechanics of Implant Fixation in Osteoporotic Bone." *Current Osteoporosis Reports* 18 (5): 577–86. https://doi.org/10.1007/s11914-020-00614-2.

Andreassen, Thor E., Donald R. Hume, Landon D. Hamilton, Sean E. Higinbotham, and Kevin B. Shelburne. 2024. "Automated 2D and 3D Finite Element Overclosure Adjustment and Mesh Morphing Using Generalized Regression Neural Networks." *Medical Engineering & Physics* 126 (April):104136. https://doi.org/10.1016/j.medengphy.2024.104136.

Asseln, Malte, Christoph Hänisch, Fabian Schick, and Klaus Radermacher. 2018. "Gender Differences in Knee Morphology and the Prospects for Implant Design in Total Knee Replacement." *The Knee* 25 (4): 545–58. https://doi.org/10.1016/j.knee.2018.04.005.

Atik, O. Şahap, Laszlo Rudolf Hangody, and Sualp Turan. 2023. "Total versus Unicompartmental Knee Arthroplasty." *Joint Diseases and Related Surgery* 34 (2): 235–36. https://doi.org/10.52312/jdrs.2023.57913.

Audenaert, E. A., C. Pattyn, G. Steenackers, J. De Roeck, D. Vandermeulen, and P. Claes. 2019. “Statistical Shape Modeling of Skeletal Anatomy for Sex Discrimination: Their Training Size, Sexual Dimorphism, and Asymmetry.” *Frontiers in Bioengineering and Biotechnology* 7 (November):302. https://doi.org/10.3389/fbioe.2019.00302.

Bah, Mamadou T., Junfen Shi, Martin Browne, Yanneck Suchier, Fabien Lefebvre, Philippe Young, Leonard King, Doug G. Dunlop, and Markus O. Heller. 2015. “Exploring Inter-Subject Anatomic Variability Using a Population of Patient-Specific Femurs and a Statistical Shape and Intensity Model.” *Medical Engineering & Physics* 37 (10): 995–1007. https://doi.org/10.1016/j.medengphy.2015.08.004.

Barton, Barbara, Romsai Tony Boonyasai, Cecilia Hahn, Doreen Bonnett, Joann Fitzell, Elizabeth Goldstein, Irene Hall, et al. n.d. “2023 National Healthcare Quality and Disparities Report.”

Bayoglu, Riza, William Fugit, Paul J. Rullkoetter, Chadd W. Clary, and Peter J. Laz. 2020. “Development of Statistical Models to Investigate Shape and Material Property Variation in the Knee: Proximal Tibia and Distal Femur.” Final Report for Statement of Work. KNEE SOW 3.1 – Development of a Statistical Shape, Alignment, and Intensity Model of the Lower Extremity.

Besl, P.J., and Neil D. McKay. 1992. “A Method for Registration of 3-D Shapes.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (2): 239–56. https://doi.org/10.1109/34.121791.

Biham, Sivan. 2022. “Point Cloud Registration — Classic Approaches.” Medium. August 5, 2022. https://towardsdatascience.com/point-cloud-registration-classic-approaches-d6191302b0b2.

Bourne, Robert B., Bert M. Chesworth, Aileen M. Davis, Nizar N. Mahomed, and Kory D. J. Charron. 2010. “Patient Satisfaction after Total Knee Arthroplasty: Who Is Satisfied and Who Is Not?” *Clinical Orthopaedics and Related Research* 468 (1): 57–63. https://doi.org/10.1007/s11999-009-1119-9.

Bozic, Kevin J., Steven M. Kurtz, Edmund Lau, Kevin Ong, Vanessa Chiu, Thomas P. Vail, Harry E. Rubash, and Daniel J. Berry. 2010. "The Epidemiology of Revision Total Knee Arthroplasty in the United States." *Clinical Orthopaedics & Related Research* 468 (1): 45–51. https://doi.org/10.1007/s11999-009-0945-0.

Bruce, Olivia L., Michael Baggaley, Arash Khassetarash, Ifaz T. Haider, and W. Brent Edwards. 2022. "Tibial-Fibular Geometry and Density Variations Associated with Elevated Bone Strain and Sex Disparities in Young Active Adults." *Bone* 161 (August):116443. https://doi.org/10.1016/j.bone.2022.116443.

Bruce, Olivia L., and W. Brent Edwards. 2023. "Sex Disparities in Tibia-Fibula Geometry and Density Are Associated with Elevated Bone Strain in Females: A Cross-Validation Study." *Bone* 173 (August):116803. https://doi.org/10.1016/j.bone.2023.116803.

Burton, William S., Irene Sintini, Jason M. Chavarria, James R. Brownhill, and Peter J. Laz. 2019. "Assessment of Scapular Morphology and Bone Quality with Statistical Models." *Computer Methods in Biomechanics and Biomedical Engineering* 22 (4): 341–51. https://doi.org/10.1080/10255842.2018.1556260.

Cavaignac, Etienne, Ke Li, Marie Faruch, Frederic Savall, Philippe Chiron, W. Huang, and Norbert Telmon. 2017. "Three-Dimensional Geometric Morphometric Analysis Reveals Ethnic Dimorphism in the Shape of the Femur." *Journal of Experimental Orthopaedics* 4 (1): 13. https://doi.org/10.1186/s40634-017-0088-2.

ChatGPT. https://chatgpt.com.

Committee on Improving the Representation of Women and Underrepresented Minorities in Clinical Trials and Research, Committee on Women in Science, Engineering, and Medicine, Policy and Global Affairs, and National Academies of Sciences, Engineering, and Medicine. 2022. *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups*. Edited by Kirsten Bibbins-Domingo and Alex Helman. Washington, D.C.: National Academies Press. https://doi.org/10.17226/26479.

Cwalina, Thomas B., Tarun K. Jella, Grigory A. Manyak, Andy Kuo, and Atul F. Kamath. 2022. "Is Our Science Representative? A Systematic Review of Racial and Ethnic Diversity in Orthopaedic Clinical Trials from 2000 to 2020." *Clinical*

Orthopaedics and Related Research 480 (5): 848–58. https://doi.org/10.1097/CORR.0000000000002050.

Cyrill Stachniss, dir. 2020. *Iterative Closest Point (ICP) - 5 Minutes with Cyrill*. https://www.youtube.com/watch?v=QWDM4cFdKrE.

Dryden, Ian L., and Kanti V. Mardia. 2016. *Statistical Shape Analysis: With Applications in R*. New York, UNITED KINGDOM: John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/du/detail.action?docID=4573538.

Edgar, Heather, Shamsi Daneshvari Berry, Emily Moes, Natalie Adolphi, Patrick Bridges, and KB Nolte. 2020. "New Mexico Decedent Image Database (NMDID)." https://doi.org/10.25827/5S8C-N515.

Eggermont, Florieke, Nico Verdonschot, Yvette Van Der Linden, and Esther Tanck. 2019. "Calibration with or without Phantom for Fracture Risk Prediction in Cancer Patients with Femoral Bone Metastases Using CT-Based Finite Element Models." Edited by María Angeles Pérez. *PLOS ONE* 14 (7): e0220564. https://doi.org/10.1371/journal.pone.0220564.

Fugit, William, Riza Bayoglu, and Peter J. Laz. (in review). "Accuracy Tradeoffs between Individual and Joint-Level Statistical Shape Models of Knee Morphology." *Medical Imaging and Physics*.

Galloway, Francis, Max Kahnt, Heiko Ramm, Peter Worsley, Stefan Zachow, Prasanth Nair, and Mark Taylor. 2013. "A Large Scale Finite Element Study of a Cementless Osseointegrated Tibial Tray." *Journal of Biomechanics* 46 (11): 1900. https://doi.org/10.1016/j.jbiomech.2013.04.021.

Gielis, Willem Paul, Hassan Rayegan, Vahid Arbabi, Seyed Y Ahmadi Brooghani, Claudia Lindner, Tim F Cootes, Pim A de Jong, H Weinans, and Roel J H Custers. n.d. "Predicting the Mechanical Hip–Knee–Ankle Angle Accurately from Standard Knee Radiographs: A Cross-Validation Experiment in 100 Patients." *Acta Orthopaedica* 91 (6): 732–37. https://doi.org/10.1080/17453674.2020.1779516.

Golyanik, Vladislav. 2020. "Introduction." In *Robust Methods for Dense Monocular Non-Rigid 3D Reconstruction and Alignment of Point Clouds*, edited by Vladislav

Golyanik, 1–11. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-30567-3_1.

Grassi, Lorenzo, Sami P. Väänänen, and Hanna Isaksson. 2021. “Statistical Shape and Appearance Models: Development Towards Improved Osteoporosis Care.” *Current Osteoporosis Reports* 19 (6): 676–87. https://doi.org/10.1007/s11914-021-00711-w.

Han, Hyuksoo, Sohee Oh, Chong Bum Chang, and Seung-baik Kang. 2016. “Anthropometric Difference of the Knee on MRI According to Gender and Age Groups.” *Surgical and Radiologic Anatomy: SRA* 38 (2): 203–11. https://doi.org/10.1007/s00276-015-1536-2.

Ho, Wei-Pin, Cheng-Kung Cheng, and Jiann-Jong Liau. 2006. “Morphometrical Measurements of Resected Surface of Femurs in Chinese Knees: Correlation to the Sizing of Current Femoral Implants.” *The Knee* 13 (1): 12–14. https://doi.org/10.1016/j.knee.2005.05.002.

Hu, Daniel A., James B. Hu, Ariel Lee, William J. Rubenstein, Kevin M. Hwang, Said A. Ibrahim, and Alfred C. Kuo. 2022. “What Factors Lead to Racial Disparities in Outcomes After Total Knee Arthroplasty?” *Journal of Racial and Ethnic Health Disparities* 9 (6): 2317–22. https://doi.org/10.1007/s40615-021-01168-4.

Hung, V. W. Y., T. Y. Zhu, W.-H. Cheung, T.-N. Fong, F. W. P. Yu, L.-K. Hung, K.-S. Leung, J. C. Y. Cheng, T.-P. Lam, and L. Qin. 2015. “Age-Related Differences in Volumetric Bone Mineral Density, Microarchitecture, and Bone Strength of Distal Radius and Tibia in Chinese Women: A High-Resolution pQCT Reference Database Study.” *Osteoporosis International* 26 (6): 1691–1703. https://doi.org/10.1007/s00198-015-3045-x.

Ibrahim, Said A., Roslyn A. Stone, Xiaoyan Han, Peter Cohen, Michael J. Fine, William G. Henderson, Shukri F. Khuri, and C. Kent Kwoh. 2005. “Racial/Ethnic Differences in Surgical Outcomes in Veterans Following Knee or Hip Arthroplasty.” *Arthritis & Rheumatism* 52 (10): 3143–51. https://doi.org/10.1002/art.21304.

Issa, Tariq Ziad, Mark J. Lambrechts, Jasmine S. Lin, Parker L. Brush, Jose A. Canseco, Alan S. Hilibrand, Christopher K. Kepler, Gregory D. Schroeder, and Alexander R. Vaccaro. 2023. “Diversity in Orthopaedic Surgery Medical Device Clinical Trials: An Analysis of the Food and Drug Administration Safety and Innovation Act.” *The Journal of the American Academy of Orthopaedic Surgeons* 31 (3): 155–65. https://doi.org/10.5435/JAAOS-D-22-00704.

Jangseob Kim, dir. 2020. *Scan Matching Algorithm Using ICP (Iterative Closest Points)*. https://www.youtube.com/watch?v=6qVZvKIkJMo.

Jolliffe, Ian T. 1990. “PRINCIPAL COMPONENT ANALYSIS: A BEGINNER’S GUIDE - I. Introduction and Application.” *Weather* 45 (10): 375–82. https://doi.org/10.1002/j.1477-8696.1990.tb05558.x.

Kaick, Oliver van, Hao Zhang, Ghassan Hamarneh, and Daniel Cohen-Or. 2011. “A Survey on Shape Correspondence.” *Computer Graphics Forum* 30 (6): 1681–1707. https://doi.org/10.1111/j.1467-8659.2011.01884.x.

Keyak, J. H., I. Y. Lee, and H. B. Skinner. 1994. “Correlations between Orthogonal Mechanical Properties and Density of Trabecular Bone: Use of Different Densitometric Measures.” *Journal of Biomedical Materials Research* 28 (11): 1329–36. https://doi.org/10.1002/jbm.820281111.

Keyak, Joyce H, Tadashi S Kaneko, Jamshid Tehranzadeh, and Harry B Skinner. 2005. “Predicting Proximal Femoral Strength Using Structural Engineering Models:” *Clinical Orthopaedics and Related Research* NA; (437): 219–28. https://doi.org/10.1097/01.blo.0000164400.37905.22.

Kim, Hae-Young. 2014. “Analysis of Variance (ANOVA) Comparing Means of More than Two Groups.” *Restorative Dentistry & Endodontics* 39 (1): 74. https://doi.org/10.5395/rde.2014.39.1.74.

Kim, T. K., Mark Phillips, Mohit Bhandari, John Watson, and Rajesh Malhotra. 2017. “What Differences in Morphologic Features of the Knee Exist Among Patients of Various Races? A Systematic Review.” *Clinical Orthopaedics and Related Research* 475 (1): 170–82. https://doi.org/10.1007/s11999-016-5097-4.

Koy, E. How Shing, A. Amouzougan, E. Biver, R. Chapurlat, T. Chevalley, S. L. Ferrari, A. Fouilloux, et al. 2022. “Reference Microarchitectural Values Measured by HR-pQCT in a Franco-Swiss Cohort of Young Adult Women.” *Osteoporosis International* 33 (3): 703–9. https://doi.org/10.1007/s00198-021-06193-x.

Kroon, Dirk-Jan. 2024. “Finite Iterative Closest Point.” https://www.mathworks.com/matlabcentral/fileexchange/24301-finite-iterative-closest-point.

Kurtz, Steven, Kevin Ong, Edmund Lau, Fionna Mowat, and Michael Halpern. 2007. “Projections of Primary and Revision Hip and Knee Arthroplasty in the United States from 2005 to 2030:” *The Journal of Bone & Joint Surgery* 89 (4): 780–85. https://doi.org/10.2106/JBJS.F.00222.

Kwak, Dai Soon, Quang Bang Tao, Mitsugu Todo, and Insu Jeon. 2012. “Determination of Representative Dimension Parameter Values of Korean Knee Joints for Knee Joint Implant Design.” *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 226 (5): 368–76. https://doi.org/10.1177/0954411912442025.

Li, B., and R.M. Aspden. 1997. “Material Properties of Bone from the Femoral Neck and Calcar Femorale of Patients with Osteoporosis or Osteoarthritis.” *Osteoporosis International* 7 (5): 450–56. https://doi.org/10.1007/s001980050032.

Li, Ke, Etienne Cavaignac, Wei Xu, Qiang Cheng, Nobert Telmon, and Wei Huang. 2018a. “Morphometric Evaluation of the Knee in Chinese Population Reveals Sexual Dimorphism and Age-Related Differences.” *International Orthopaedics* 42 (10): 2349–56. https://doi.org/10.1007/s00264-018-3826-x.

———. 2018b. “Morphometric Evaluation of the Knee in Chinese Population Reveals Sexual Dimorphism and Age-Related Differences.” *International Orthopaedics* 42 (10): 2349–56. https://doi.org/10.1007/s00264-018-3826-x.

Long, Huibin, Qiang Liu, Heyong Yin, Kai Wang, Naicheng Diao, Yuqing Zhang, Jianhao Lin, and Ai Guo. 2022. “Prevalence Trends of Site-Specific Osteoarthritis From 1990 to 2019: Findings From the Global Burden of Disease Study 2019.” *Arthritis & Rheumatology* 74 (7): 1172–83. https://doi.org/10.1002/art.42089.

Ma, Qian-Li, Joseph D. Lipman, Cheng-Kung Cheng, Xiao-Nan Wang, Yi-Yuan Zhang, and Bin You. 2017. "A Comparison Between Chinese and Caucasian 3-Dimensional Bony Morphometry in Presimulated and Postsimulated Osteotomy for Total Knee Arthroplasty." *The Journal of Arthroplasty* 32 (9): 2878–86. https://doi.org/10.1016/j.arth.2017.03.069.

Mahfouz, Mohamed, Emam ElHak Abdel Fatah, Lyndsay Smith Bowers, and Giles Scuderi. 2012. "Three-Dimensional Morphology of the Knee Reveals Ethnic Differences." *Clinical Orthopaedics & Related Research* 470 (1): 172–85. https://doi.org/10.1007/s11999-011-2089-2.

Mahfouz, Mohamed R., Emam ElHak Abdel Fatah, Lyndsay Bowers, and Giles Scuderi. 2015. "A New Method for Calculating Femoral Anterior Cortex Point Location and Its Effect on Component Sizing and Placement." *Clinical Orthopaedics & Related Research* 473 (1): 126–32. https://doi.org/10.1007/s11999-014-3930-1.

Martel-Pelletier, Johanne, Andrew J. Barr, Flavia M. Cicuttini, Philip G. Conaghan, Cyrus Cooper, Mary B. Goldring, Steven R. Goldring, Graeme Jones, Andrew J. Teichtahl, and Jean-Pierre Pelletier. 2016. "Osteoarthritis (Primer)." *Nature Reviews: Disease Primers* 2 (1). https://doi.org/10.1038/nrdp.2016.72.

Montaseri, Atousa, M Reza Alinaghizadeh, and Seied Mahdavi. 2012. "Physical Properties of Ethyl Methacrylate as a Bolus in Radiotherapy." *Iranian Journal of Medical Physics* 9 (July):127–34. https://doi.org/10.22038/ijmp.2012.318.

Myronenko, Andriy and Xubo Song. 2010. "Point Set Registration: Coherent Point Drift." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (12): 2262–75. https://doi.org/10.1109/TPAMI.2010.46.

Nieves, Jeri W, Carmelo Formica, Jamie Ruffing, Marsha Zion, Patricia Garrett, Robert Lindsay, and Felicia Cosman. 2005. "Males Have Larger Skeletal Size and Bone Mass Than Females, Despite Comparable Body Size." *Journal of Bone and Mineral Research* 20 (3): 529–35. https://doi.org/10.1359/JBMR.041005.

"NIH Consensus Statement on Total Knee Replacement." 2003. *NIH Consensus and State-of-the-Science Statements* 20 (1): 1–34.

Noel, Sabrina E, Michelly P Santos, and Nicole C Wright. 2021. “Racial and Ethnic Disparities in Bone Health and Outcomes in the United States.” *Journal of Bone and Mineral Research : The Official Journal of the American Society for Bone and Mineral Research* 36 (10): 1881–1905. https://doi.org/10.1002/jbmr.4417.

Perrot, Serge. 2015. “Osteoarthritis Pain.” *Best Practice & Research Clinical Rheumatology*, Pain in systemic rheumatic diseases, 29 (1): 90–97. https://doi.org/10.1016/j.berh.2015.04.017.

Pijls, Bart G, Edward R Valstar, Klaas-Auke Nouta, Josepha Wm Plevier, Marta Fiocco, Saskia Middeldorp, and Rob Ghh Nelissen. 2012. “Early Migration of Tibial Components Is Associated with Late Revision: A Systematic Review and Meta-Analysis of 21,000 Knee Arthroplasties.” *Acta Orthopaedica* 83 (6): 614–24. https://doi.org/10.3109/17453674.2012.747052.

Pouli, Nektaria, Roshan Das Nair, Nadina Berrice Lincoln, and David Walsh. 2014. “The Experience of Living with Knee Osteoarthritis: Exploring Illness and Treatment Beliefs through Thematic Analysis.” *Disability and Rehabilitation* 36 (7): 600–607. https://doi.org/10.3109/09638288.2013.805257.

Qingqiong Deng, Mingquan Zhou, and Zhongke Wu. 2010. “An Automatic Non-Rigid Registration Method for Dense Surface Models.” In *2010 IEEE International Conference on Intelligent Computing and Intelligent Systems*, 888–92. Xiamen, China: IEEE. https://doi.org/10.1109/ICICISYS.2010.5658814.

Riggs, B Lawrence, L Joseph Melton, Richard A Robb, Jon J Camp, Elizabeth J Atkinson, James M Peterson, Peggy A Rouleau, Cynthia H McCollough, Mary L Bouxsein, and Sundeep Khosla. 2004. “Population-Based Study of Age and Sex Differences in Bone Volumetric Density, Size, Geometry, and Structure at Different Skeletal Sites.” *Journal of Bone and Mineral Research* 19 (12): 1945–54. https://doi.org/10.1359/jbmr.040916.

Ritter, Merrill A., Jennifer T. Wing, Michael E. Berend, Kenneth E. Davis, and John B. Meding. 2008. “The Clinical Effect of Gender on Outcome of Total Knee Arthroplasty.” *The Journal of Arthroplasty* 23 (3): 331–36. https://doi.org/10.1016/j.arth.2007.10.031.

Rudisill, Samuel S., Nathan H. Varady, Aseal Birir, Susan M. Goodman, Michael L. Parks, and Troy B. Amen. 2023. “Racial and Ethnic Disparities in Total Joint Arthroplasty Care: A Contemporary Systematic Review and Meta-Analysis.” *The Journal of Arthroplasty* 38 (1): 171-187.e18. https://doi.org/10.1016/j.arth.2022.08.006.

Schileo, Enrico, Fulvia Taddei, Andrea Malandrino, Luca Cristofolini, and Marco Viceconti. 2007. “Subject-Specific Finite Element Models Can Accurately Predict Strain Levels in Long Bones.” *Journal of Biomechanics* 40 (13): 2982–89. https://doi.org/10.1016/j.jbiomech.2007.02.010.

Shahid, Hania, and Jasvinder A. Singh. 2016. “Racial/Ethnic Disparity in Rates and Outcomes of Total Joint Arthroplasty.” *Current Rheumatology Reports* 18 (4): 20. https://doi.org/10.1007/s11926-016-0570-3.

Smedley, Audrey, and Brian D. Smedley. 2005. “Race as Biology Is Fiction, Racism as a Social Problem Is Real: Anthropological and Historical Perspectives on the Social Construction of Race: American Psychologist.” *American Psychologist*, Genes, Race, and Psychology in the Genome Era, 60 (1): 16–26. https://doi.org/10.1037/0003-066X.60.1.16.

Stanbrook, Matthew B., and Bukola Salami. 2023. “CMAJ’s New Guidance on the Reporting of Race and Ethnicity in Research Articles.” *CMAJ : Canadian Medical Association Journal* 195 (6): E236–38. https://doi.org/10.1503/cmaj.230144.

Stegmann, Mikkel B, and David Delgado Gomez. n.d. “A Brief Introduction to Statistical Shape Analysis,” 15.

Szabo, Emily, and Clare Rimnac. 2022. “Biomechanics of Immature Human Cortical Bone: A Systematic Review.” *Journal of the Mechanical Behavior of Biomedical Materials* 125 (January):104889. https://doi.org/10.1016/j.jmbbm.2021.104889.

Taghizadeh, Elham, Vimal Chandran, Mauricio Reyes, Philippe Zysset, and Philippe Büchler. 2017. “Statistical Analysis of the Inter-Individual Variations of the Bone Shape, Volume Fraction and Fabric and Their Correlations in the Proximal Femur.” *Bone* 103 (October):252–61. https://doi.org/10.1016/j.bone.2017.07.012.

Taylor, Mark, and Patrick J. Prendergast. 2015. “Four Decades of Finite Element Analysis of Orthopaedic Devices: Where Are We Now and What Are the Opportunities?” *Journal of Biomechanics* 48 (5): 767–78. https://doi.org/10.1016/j.jbiomech.2014.12.019.

Temerinac-Ott, Maja. 2012. “Multiview Reconstruction for 3D Images from Light Sheet Based Fluorescence Microscopy.”

“The Stanford 3D Scanning Repository.” n.d. Accessed April 23, 2024. https://graphics.stanford.edu/data/3Dscanrep/.

Vaidya, Shrinand V., Chitranjan S. Ranawat, Alaric Aroojis, and N.S. Laud. 2000. “Anthropometric Measurements to Design Total Knee Prostheses for the Indian Population.” *The Journal of Arthroplasty* 15 (1): 79–85. https://doi.org/10.1016/S0883-5403(00)91285-3.

Varacallo, Matthew, T David Luo, and Norman A Johanson. n.d. “Total Knee Arthroplasty Techniques.”

Wang, Jinhong, Thomas A. G. Hall, Omar Musbahi, Gareth G. Jones, and Richard J. van Arkel. 2023. “Predicting Hip-Knee-Ankle and Femorotibial Angles from Knee Radiographs with Deep Learning.” *The Knee* 42 (June):281–88. https://doi.org/10.1016/j.knee.2023.03.010.

Wise, Barton L., Felix Liu, Lisa Kritikos, John A. Lynch, Neeta Parimi, Yuqing Zhang, and Nancy E. Lane. 2016. “The Association of Distal Femur and Proximal Tibia Shape with Sex: The Osteoarthritis Initiative.” *Seminars in Arthritis and Rheumatism* 46 (1): 20–26. https://doi.org/10.1016/j.semarthrit.2016.02.006.

Yu, Fan, Yuping Xu, Yanfang Hou, Yuanyuan Lin, Ruizhi Jiajue, Yan Jiang, Ou Wang, et al. 2020. “Age-, Site-, and Sex-Specific Normative Centile Curves for HR-pQCT-Derived Microarchitectural and Bone Strength Parameters in a Chinese Mainland Population.” *Journal of Bone and Mineral Research* 35 (11): 2159–70. https://doi.org/10.1002/jbmr.4116.

Zhang, Ju, David Ackland, and Justin Fernandez. 2018. “Point-Cloud Registration Using Adaptive Radial Basis Functions.” *Computer Methods in Biomechanics and Biomedical Engineering* 21 (7): 498–502. https://doi.org/10.1080/10255842.2018.1484914.