Changes in Femoral Structure and Function Following Anterior Cruciate Ligament Injury and with Aging

by

Daniella M. Patton

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in the University of Michigan
2020

Doctoral Committee:

Professor Karl J. Jepsen, Chair
Professor Ellen M. Arruda
Research Professor James A. Ashton-Miller
Assistant Professor Todd L. Bredbenner, University of Colorado Colorado Springs
Professor Indika Rajapakse

Daniella M. Patton
pattondm@umich.edu
ORCID iD: 0000-0002-2693-7905

Dedication

To my parents, Jasson Patton and Carolina Lopez, who have been my relentless supporters and inspiration.

Thank you for being a constant reminder as to why I embarked on this academic endeavor. Without your encouragement and love the opportunity to write and complete this dissertation would not have been possible.

"The brick walls are there for a reason. The brick walls are not there to keep us out; the brick walls are there to give us a chance to show how badly we want something."

- Randy Pausche, The Last Lecture

Acknowledgements

The road to completing my Ph.D. was harder than I envisioned - a non-linear trajectory filled with dead-ends and do-overs. However, this has been the most positive and fruitful learning experience of my life. For this, I would like to thank many who have sacrificed their time and efforts to encourage, teach, and challenge me throughout my time at Michigan.

I would first like to express my sincerest gratitude to my academic advisor, Professor Karl Jepsen, who has by example shown me what a great scientist (and person) should be. Beyond the endless support, positivity, and academic training, I would also like to thank him for pushing me to think differently about problems – out of the box and through a big-picture lens. I truly feel capable of tackling any research question and am grateful for the tools he helped me develop to do so. I would like to also express my sincerest appreciation to my committee members, Professors Todd Bredbenner, James Ashton-Miller, Indika Rajapakse, and Ellen Arruda for graciously donating both time and expertise to the execution of this dissertation. Their insightful comments, constructive feedback, and thoughtful questions were valued greatly.

I am very grateful for Professor Steve Schlecht, who gave numerous informal 'lab lectures' on anatomy and histology and allowed me the opportunity to get involved with the ACL work written in this thesis. He has been both a mentor and a friend. I am hopeful that our paths will cross professionally in the future. Also, I would like to thank Professor Edward Wojtys, my clinical mentor through the translation research education certificate, who allowed me the opportunity to combine my interest in bone research and sports-related musculoskeletal injuries. His positivity and sharp clinical hypotheses greatly added to my experience here at Michigan.

This work would not have been possible without the many members of the ORL, who provided both professional and personal encouragement and support. In particular, I would like to thank Erin Bigelow, who was critical to helping me get set-up and was both my confidant and close friend while working in the lab. Further, I would like to express thanks to Rob Goulet, who has helped me in an innumerable number of ways throughout my Ph.D.: sending me articles of interest, answering endless questions, and working with me on code used in this dissertation. A special thank you must also be extended to Sharon Vaassen who helped me through all financial and funding questions/issues that I found here. To my friends in the ORL, thank you all for keeping me sane and bringing so much joy to my life with various Bone Crushers Trivia Nights, annual Tour De Fries events, and happy hours.

To my friends, I would like to express my sincerest gratitude for the endless laughs at our board game, wine and cheese, and movie nights. I am so lucky to be surrounded by so many kind, hard-working, and motivational individuals.

I will be forever thankful for my family. Mom and Dad, thank you for teaching me how to dream bring and for both being such a great example of the humble and hard-working people I strive to be. I would like to thank my sister Tara, who helped me get my mind off of research on our weekend adventures in Ann Arbor and Plymouth. Finally, I could not have finished this work without my fiancé Aaron. Thank you for being such a wonderful life partner and adventurer. You have been such a big driving force in completing this degree and I feel so fortunate that being reunited with you is something that I get to look forward to soon.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

The ACL, a ligament connected to the distal femur, has little regenerative capacity. In consequence, surgical intervention is required if a patient hopes to remain active following ACL injury. In addition to the long recovery time and associated morbidities (e.g., osteoarthritis) following surgery, up to 12% of the primary reconstructed ACL grafts will fail within 15 years. Revision reconstructions are inferior to primary ACL reconstructions, thus, understanding the mechanism of failure is critical to mitigating worst-case outcomes. Reasons for revision risk have largely focused on technical errors despite that biological factors may also be a cause. Bone, a biological factor, decreases in mass following ACL injury. However, how bone microstructure changes following injury has remained largely unexplored.

It was determined in this study that bone microstructure differs on a patient-by-patient basis undergoing ACL reconstructive surgery. Differences in microarchitecture could not be explained by time from injury to operation (i.e. time of disuse) or activity the patient was participating in at the moment of injury. Thus, differences in bone quality are due to variability present at baseline, in response to injury, and/or activity level following injury. Clinically, these findings are important because we are the first to show that bone quality varies across patient groups, pointing out that microstructure may be an important factor to consider in assessing ACL injury risk and surgical outcomes.

The second half of this thesis compared age-related and sex-specific differences in bone microstructure to whole bone strength in the proximal femur with the long term goal of improving diagnostic methods to assess osteoporotic hip fracture risk. Hip fragility fractures are costly, associated with a severe decrease in the quality of life, and nearly half of patients (>65 years) who suffer a hip fracture never regain normal function. Unfortunately, approximately fifty percent of patients that experience a hip fracture receive no prophylactic treatment prior to fragility fracture because they are not diagnosed as osteoporotic using current

clinical diagnostic methods. Both bone mass and microstructure change with age and the progression of osteoporosis. However, technical limitations have made it difficult to measure fracture risk from a biomechanical perspective - relating proximal femur bone strength and microstructure in synergy.

The second study determined that the magnitude of sex-specific differences in bone strength was greater than age-related strength loss endured throughout life. Further, there was no sex-specific difference in the rate of loss observed herein. Clinically, these findings demonstrate that if females could maximize bone quality early in life, they may be able to maintain the structural strength later on, even with bone loss, to mitigate fragility fractures altogether. Further, mechanical variables (i.e., stiffness and post-yield-displacement) and demographic data (i.e., age and sex) could not adequately explain variability in whole bone strength. Microstructural analysis in the femoral neck improved our ability to predict whole bone strength but demonstrated that sub-regional microstructural detail only modestly improved strength predictability in comparison to average measures across the femoral neck. Despite this, we found that increased levels of micro-architectural detail are needed to identify sex-specific differences in whole bone strength. Clinically, these findings demonstrate that regional analysis may be useful for identifying those at greatest risk of fracture earlier in life and in a sex-specific manner.

Chapter 1 Introduction

Bones in the human skeleton are extraordinary: developed to be both strong and light so an individual can experience substantial impact without fracture and move freely without onerous energy expenses. Beyond the well-known mechanical utility of bones - allowing humans to jump, run, and lift – they also play vital roles in protecting, maintaining mineral homeostasis, and forming blood (i.e., hematopoiesis) in the human body (1,2). For example, the ribcage protects the heart and the lungs and the skull protects the brain. The skeleton functions as a reservoir for minerals in the body (e.g., calcium, phosphorus, iron, and magnesium) essential for body organs to operate (1,2). Further, bone is blood-forming and acts as an endocrine organ to help to mediate phosphate and energy metabolism throughout life (1). To re-iterate, the human skeleton is extraordinary, playing roles in numerous vital functions in the human body. Unsurprisingly, if bone degrades as a result of a disease state or traumatic musculoskeletal injury the consequence can be catastrophic.

This thesis aims to characterize bone degeneration at either end of the femur (i.e., the distal and proximal metaphysis) - the strongest and longest bone in the body (Figure 1.1). The femur is the only bone in the upper mid-thigh in the human body and, as a result, bone atrophy here can severely impact an individual, potentially resulting in a prolonged decrease in functional capacity, increased risk of co-morbidities, and rise in all-cause mortality risk (3–5). This thesis will focus on bone changes following two different conditions: (1) following anterior cruciate ligament (ACL) injury in young females to better understand how bone health may impact ACL reconstructive surgery outcomes and (2) with age to better understand the type of degeneration that has the most deleterious effect on bone strength for both sexes. ACL injuries and osteoporosis affect different populations, have different associated co-morbidities and differing mortality risks (6–8). However, both result in the degeneration of bone, have similar sex-specific etiologies, and are serious public health burdens for both sexes (4,9).

More specifically, we will be studying bone microstructure acutely after ACL injury and regarding sex-specific and age-related changes in bone strength in cadaveric proximal femurs. Bone microstructure has, to the best of our knowledge, remained largely unexplored in the distal femur near the ACL following injury in patient studies. On the other hand, age-related changes of bone microstructure in the proximal femur have been studied extensively ex-vivo (10). However, despite known age-related changes and knowledge that bone microarchitecture plays an important role in whole bone strength in the proximal femur, technical limitations have made it difficult to measure bone strength and bone microstructure in synergy (10–12). Thus, this thesis will help to close two unique gaps in our current knowledge in research.

Bone degeneration (i.e., osteolysis) following ACL injury and reconstructive surgery can be seen in patients who experience tunnel expansion in the distal metaphysis (Figure 1.2, A). Unfortunately, tunnel expansion can severely impact patient outcomes and long term surgical success (13,14). Similarly, bone degeneration as a result of osteoporosis can also have serious consequences. For example, osteoporosis may result in a proximal femur so weak that it can fracture from a fall of standing height or less (i.e., fragility fracture) (Figure 1.2, B). Hip fractures are costly, associated with a severe decrease in the quality of life, and nearly 50% of patients (>65 years) who suffer a fracture to the hip never regain normal function (7,8). The aforementioned reasons demonstrate how bone degeneration in both the distal and proximal femur are both serious and costly.

It is well established that bone mass is lost in the first three months following ACL injury and with the progression of untreated osteoporosis (15–18). However, beyond bone mass, there are a variety of other important properties of bone quality that can also change following injury or disease. Bone microstructure, mineral make-up, and shape (i.e., properties of bone quality) can also directly impact bone's ability to function correctly (19,20). For example, from a load-bearing perspective, it is quite possible to have two trabecular structures with the same mass but different measures of strength (Figure 1.3). This demonstrates that while bone mass is important, mass alone leaves much unexplained in the context of bone functionality.

Factors of bone quality, particularly bone microstructure, need to be considered to understand bone disease, treatment, or injury. However, bone microstructure remains an elusive topic, partially due to the clinical problems such as increased radiation dosage, increased scan time, and technical limitations required to achieve such scans in practice. Changes in bone microstructure may result in decreased bone strength and increased risk in fragility fracture, beyond what is predictable from loss of bone mass in patients with osteoporosis (21). Further, while largely unexplored, degeneration of bone microstructure following ACL injury may impact the success of surgical repair and propensity of experiences concurrent co-morbidities (17,22). Thus, understanding bone changes at the microstructural level would be beneficial to inform on risk factors for those who experience ACL injury or hip fracture due to osteoporosis. Further, understanding bone microstructure following ACL injury or hip fracture may guide towards improved diagnostics and treatment methods in the future.

The work in this thesis relies on the use of a nano-computed tomography (Nano-CT) system (nanotom-s, phoenix|x-ray, GE Measurement & Control; Wunstorf, Germany) to visualize and quantify bone microstructure. Various techniques were used (i.e., CT scanning, image processing, statistical analysis, and mechanical testing), requiring that individuals with different expertise teach and train me to complete this work. The work presented throughout this thesis was completed by me but relied on a collaborative group. I. The remainder of this chapter will establish a foundation for bone biology to aid in understanding the material presented in the subsequent chapters of this dissertation. First, the composition, microarchitecture, and mechano-responsive behavior of bone will be described. Next, key measures of bone microarchitecture typically reported in the literature will be briefly presented. The final section contains an outline of the objectives of this dissertation.

Bone Composition

To play so many vital roles in the human body, bone must be both multiscale and hierarchical (1,2,23). At the nano-structural (10^{-9} m) level, bone is a composite material composed of minerals (65-70 %, e.g., calcium hydroxyapatite, phosphorous, and chloride), organic components (20-25%, e.g., type one collagen) and water (10%) (1,23). The chemical make-up

of bone allows it to be both strong and stiff (mineral component) and ductile (organic component) so that bones can withstand substantial amounts of energy before fracture. The collagen-mineral make-up can differ both in the distribution of material type and organization (i.e., in sheets, circumferential rings, concentrically around vascular channels, etc.) depending on the function and manner in which it was deposited (1,20).

Micro-Architecture of Bones: Cortical and Trabecular Bone

At the microstructural level (10^{-6} m), bone is arranged in space in a biologically efficient manner to provide vital functions for the body. Bone can be broadly categorized as either a cortical or trabecular structure (Figure 1.4). These categories of microstructure have both architectural and functional differences.

- Cortical bone is compact and dense, taking most of the role of load-bearing in the human body (1). This type of bone is the primary component of the shaft (i.e. diaphysis) and is present in the metaphysis of long bones.
- Trabecular bone or "spongy bone" is a porous structure (i.e., porosity of ~ 80%). Trabecular bone architecture is a lightweight scaffold that helps redirect stresses to the stronger cortical shell (1,11). Further, due to its architecture, trabecular bone has a large surface area allowing for quick mineral resorption. Trabecular bone is predominately found at the metaphysis of long bones and as a major component of the vertebrae (11).

Bone is a Mechanically Responsive/ Dynamic Structure

Bone is both a mechanically responsive organ optimized for loads endured daily and a dynamic structure that is constantly remodeling throughout life (24–28). For example, larger and heavier individuals tend to have bones adapted to be large and strong to withstand daily loads. However, beyond basic macroscale changes in bone structure (i.e., size and density), loading plays a critical role in the orientation, location, and density of bone microstructure (24–28). For example, cortical and trabecular regions in the proximal femur are highly heterogeneous based on loading conditions frequently endured (Figure 1.5).

To be an adaptive structure bone is in a continuous state of remodeling throughout life. Approximately 25% of trabecular and 3% of cortical bone is replaced through remodeling each

year in a healthy adult (29,30). Cells that play important roles in this remodeling process are osteoblasts to help lay down new bone, osteoclasts to resorb old bone, and osteocytes to signal when bone is exposed to mechanical stress or stimuli (29,30). The normal bone remodeling cycle is controlled by several endocrine and immunological factors (e.g., pro-inflammatory cytokines, glucocorticoids, and parathyroid hormone [PTH]) (30). Thus, there are a variety of ways in which coordinated osteoclast-osteoblast activity can be directly affected. Aging, immobility (i.e., long term bed rest), poor nutrition, and inflammation generally result in an uncoupling of osteoblast-osteoclast activity resulting in a net bone loss (29).

Despite many different mechanisms of bone loss, bone is generally lost systemically and intelligently, losing bone where lesser loads are experienced first and placing greater value on regions of primary stress. For example, in the proximal femur, trabecular bone is arranged in arcades to align with stressors experienced most frequently under normal loading (26). With age and progression of osteoporosis, trabecular arcades are resorbed in an orderly fashion, with arcades of lessor stressors resorbed first and arcades of primary stressors becoming more prominent as thinner trabecular get resorbed (25,26) (Figure 1.6). Thus, in healthy individuals, bone is an intelligently adapted structure that is largely dictated by (1) external forces (i.e., compressive, tensile, and shear loads) and (2) how effectively bone remodels (26,31).

Microstructure

A variety of morphometric indices derived from micro-CT scans are used to quantitatively describe bone microstructure in the literature. Generally speaking, bone is separated into cortical or trabecular volumes and morphometric variables are quantified separately to characterize bone architecture. Standard variables used to characterize trabecular bone are bone volume fraction (BV/TV or BVF), thickness (Tb.Th), spacing (Tb.Sp), and number (Tb.N) (32). Standard variables used to quantify cortical bone microstructure include total area (Tt.Ar), cortical area (Ct.Ar), cortical area fraction (Ct.Ar/Tt.Ar), and cortical thickness (Ct.Th) (32). Each of these variables is briefly described below (Table 1.1). There are other frequently used variables to quantify bone microstructure (e.g., structure model index, and degree of anisotropy) that will not be described here but are explained in detail elsewhere (32).

Chapter Objectives

The overarching goal of this thesis is to determine the extent to which bone microstructure in the distal femur metaphysis degenerates following ACL injury(Chapter 2) and how age-related and sex-specific differences in bone microstructure impact proximal femur bone strength (Chapters 3-5). If these goals are achieved, we may be able to guide clinicians towards improved diagnostics and treatment methods. In chapter two, whether or not bone microstructural changes occur following ACL injury will be assessed, working under the hypothesis that the longer the time from injury to operation (i.e., the longer the period of disuse) the more extreme the degeneration occurs in bone. If true, then the time from injury to operation may impact an individual's ability for the new ACL to effectively osseointegrate and may play a role in long term surgical success. In chapters three through five, a biomechanical approach will be used to establish how changes in whole bone microarchitecture in the proximal femur impact fracture risk. It is well appreciated that bone microarchitecture deteriorates with age, but there is currently little understanding of how such sex-specific changes in bone microstructure influences whole bone strength using a direct approach to assess bone mechanical properties (i.e., direct mechanical testing) (10).

Chapter two is a study that would not be possible without the support of Professors Edward Wojtys, Stephen Schlecht, and James Ashton-Miller. In this chapter bone micro-architecture at the ACL enthesis were quantified (i.e., the region where the ACL tendon connects to the distal femur) in young females who are undergoing reconstructive surgery. To our knowledge, we are the first to assess microstructural changes at the key location where bone is removed and the new ACL tendon will be placed during surgery. We will quantify the extent of change and location (e.g., cortical or trabecular regions) of bone degeneration to determine if factors, such as time from injury to operation, play a role in the quality of bone pre-surgical intervention. If large variations in bone micro-architecture exist, thisr finding may provide a new, biologically related, risk factor that plays a role in the success of graft osseointegration beyond the typically reported technical errors.

The remainder of this thesis will be focused on a rare collection of proximal femur cadaveric bones collected in the adult age range for both sexes (M: n=44, 18 - 89 years, F: n=40, 24 - 95

years). This rare collection was obtained and made available to me as a result of my work under Karl Jepsen, my mentor. Further, Karl Jepsen guided and supported me throughout all of the scientific chapters presented below. We have a large cadaveric study examining proximal femur bone strength using mechanical testing, and we are, to the best of our knowledge, the first to have the entire proximal femur scanned at a 27 μm resolution. Thus, we have the novel ability to overcome the field-of-view to resolution limitation so that bone microstructure of the entire proximal femur can be resolved using our Nano-CT. This will allow us to quantify microarchitecture at the key location where the femur fails and probe important questions regarding the contributions of whole bone microstructure in relation to whole bone strength.

Chapter three will establish the biomechanical properties of the proximal femurs used for all subsequent analyses of this dissertation. This work was completed with the support of Erin Bigelow. Proximal femurs will be loaded to failure in fall-to-side orientation using mechanical testing. From this, the relationship between stiffness and strength will be assessed in regard to sex and aging. Although men have stronger bones relative to body size compared to women (33), it is not known whether the relationship between stiffness and strength differs between sexes. We will examine this relationship and determine how demographic factors, in addition to other measurable mechanical properties such as brittleness (i.e., post-yield displacement) affect this relationship. Further, this relationship will be explored in the context of different anatomical regions using mechanical testing. Knowing whether the stiffness-strength relationship varies with age and sex at multiple whole bone sites in the human body is important for refining strength estimates, which will benefit efforts aimed at reducing fragility fractures (34).

The objective of chapter four will be to create a standard, automated, and accurate method to reliably segment proximal femur bones so that bone microarchitecture can be reliably calculated. In order to evaluate bone architecture, it is required that bone and the particular regions of interest are accurately identified via segmentation. Our collection of femurs widely varies in cortical and trabecular architecture within and across structures making it difficult to segment using a single global threshold (Figure 1.7). Thus, we will apply a machine learning approach using a fully convolutional neural network (FCNN) to segment bone. While there

have recently been unparalleled advances in the application of machine learning for segmentation of medical image data, to our knowledge, we are the first to demonstrate FCNN applicability in high-resolution ex-vivo CT scans; scans commonly used to assess disease progression and/or drug treatment response in musculoskeletal research (32,35). This chapter would not be possible without the help of key collaborators from Object Research Systems (ORS, Montreal, CA): Benjamin Provencher, Nicolas Piche, and Mike Marsh. Further, this work was completed with the collaborative efforts from key individuals at the University of Colorado Colorado Springs: Emilie Henning and Todd Bredbenner. Sean K. Carrol, a master's student in the Kinesiology department at UM, was another key contributor to work presented in this chapter. Finally, the work presented here could not be completed without Rob W. Goulet who was involved in every aspect of this study - helping to write the code used for analysis, order the computer parts to run machine learning methods, work with ORS developers to overcome technical problems, and help with the manual segmentation required in this study.

In chapter five microstructural traits (cortical and trabecular) in the femoral neck that best predict bone strength in males and females will be identified. This chapter builds off of chapter three where bone mechanical properties were defined and chapter four where a technique was created to accurately segment bone in our NanoCT scans. We will determine if adding more cortical and trabecular architectural details improves strength predictions. Our objective in this chapter is to close the gap in our current understanding of how bone microstructure relates to bone strength in an age- and sex-specific manner (36–40). We expect this work will provide an opportunity to re-map diagnostic metrics from DXA or QCT images in a sex-specific manner to improve fracture risk predictions. Rob W. Goulet was a key collaborator in helping to develop the custom script used to analyze the femurs in different levels of refinement in this study.

Finally, chapter six provides concluding remarks and suggests future directions of research. As will be demonstrated in the subsequent chapters, the collaborative network of clinicians, bone researchers (both faculty and staff), students, and software engineers made it possible to analyze bone microstructure in a novel and unique manner.

Figures

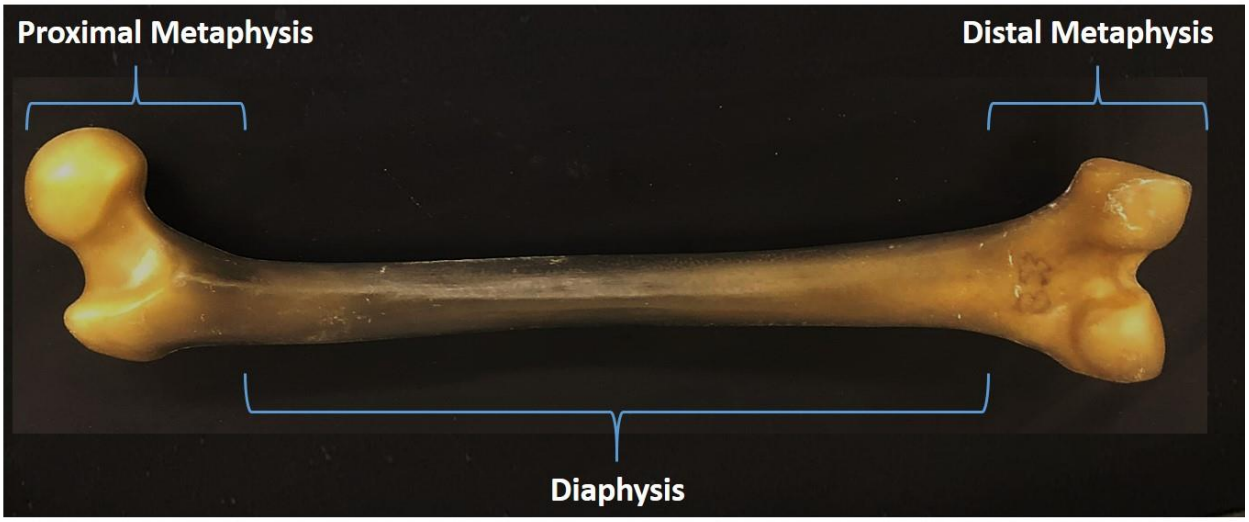


Figure 1.1 Human femur bone with the proximal metaphysis, distal metaphysis, and diaphysis labeled.

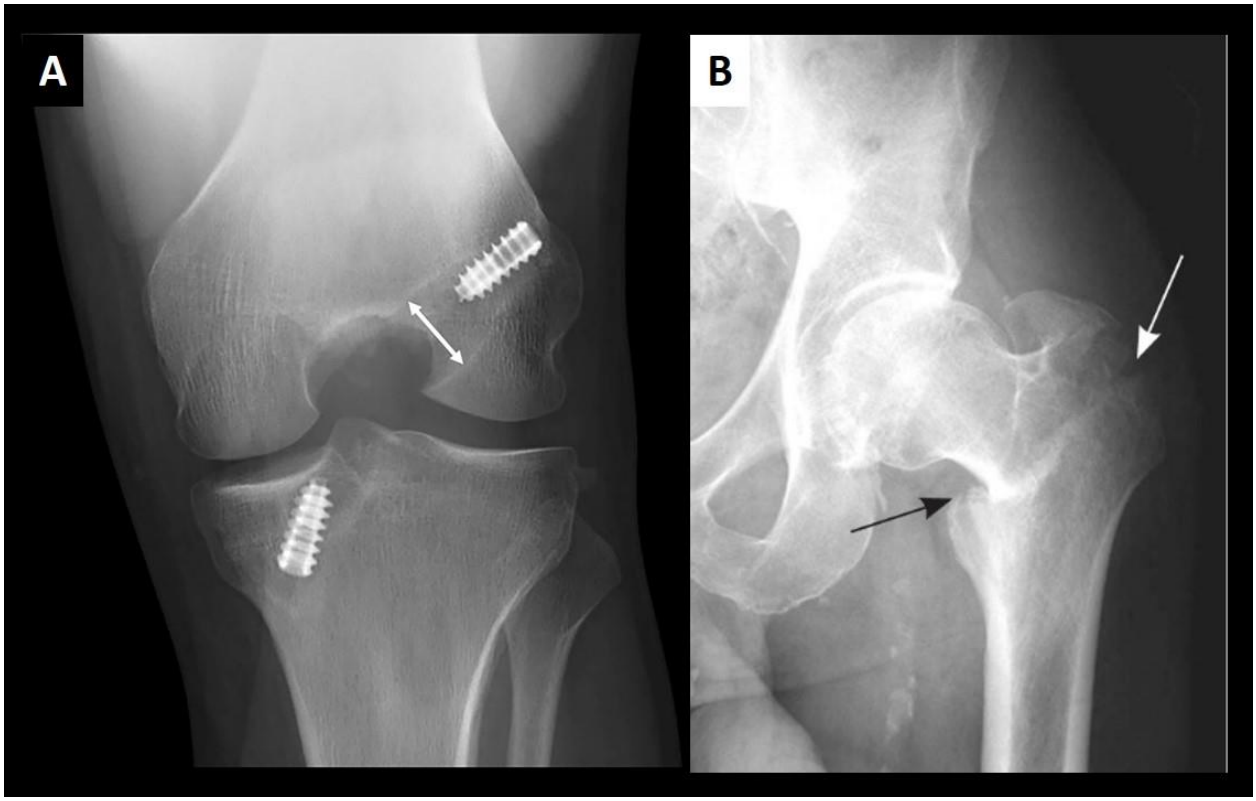


Figure 1.2 (A*) X-ray of patient (24-year-old male) with an unstable bone-patellar tendon-bone autograft. Note the significant expansion of the original graft tunnel on the femoral side (arrows) due to osteolytic activity. Courtesy of Professor Edward M. Wojtys. (B) X-Ray of intertrochanteric fragility fracture* (41).

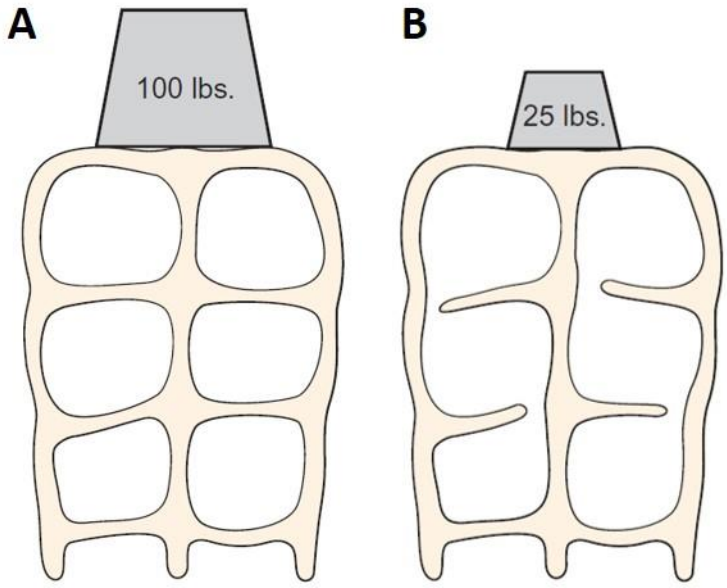


Figure 1.3 A figure demonstrating the maximum load two bone structures with similar amounts of material (i.e. mass) can hold. Two theoretical examples of cancellous bone with the same mass but different connectivity with a (B) poorly connected and fully connected structure (A) (1).

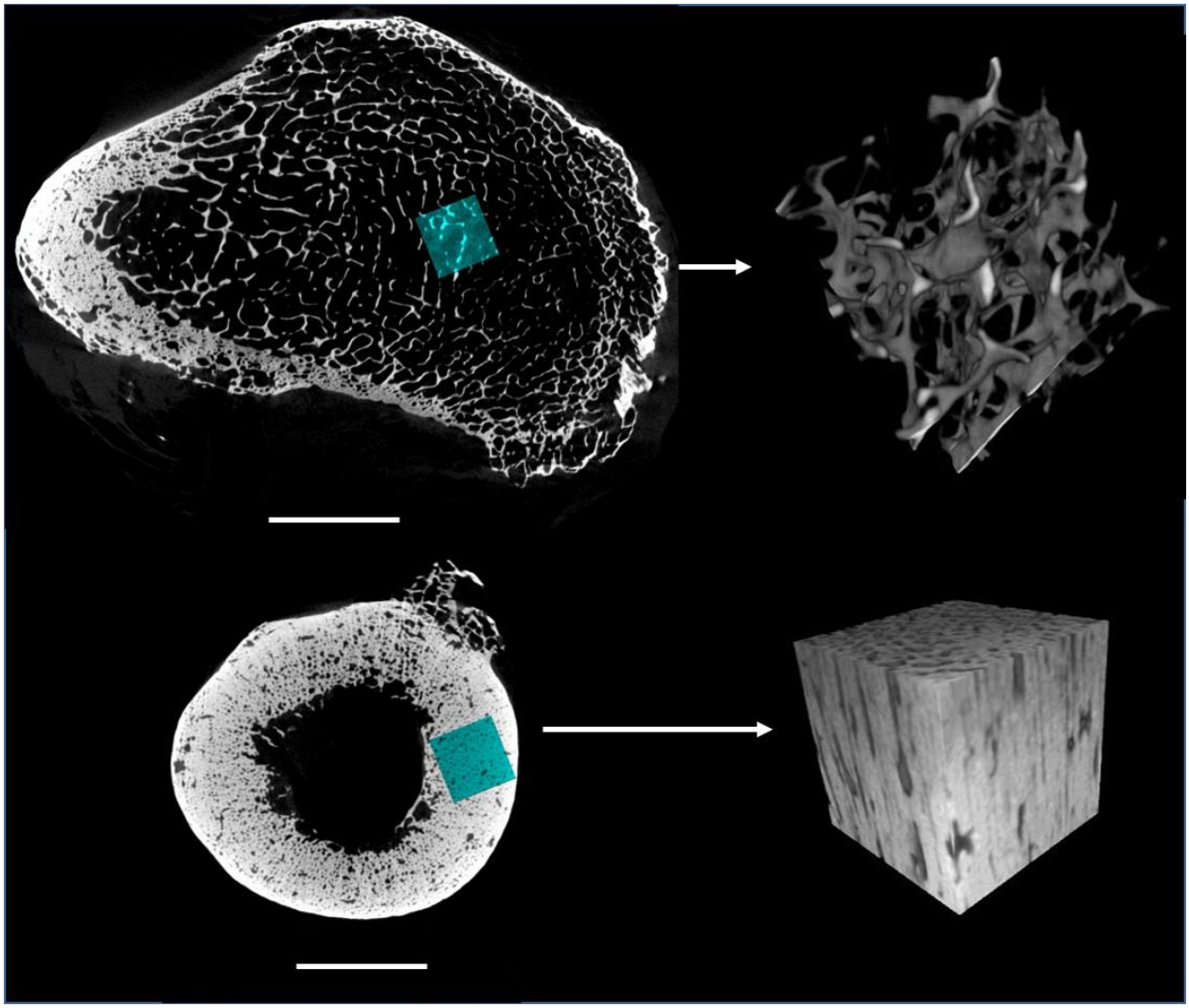


Figure 1.4 (Top) A volumetric cube (10 mm^3) of trabecular bone extracted from the metaphysis and (Bottom) cortical bone extracted from the diaphysis the femur (10 mm scale bar for both top and bottom images) (Nano-CT scan at 27 μm).

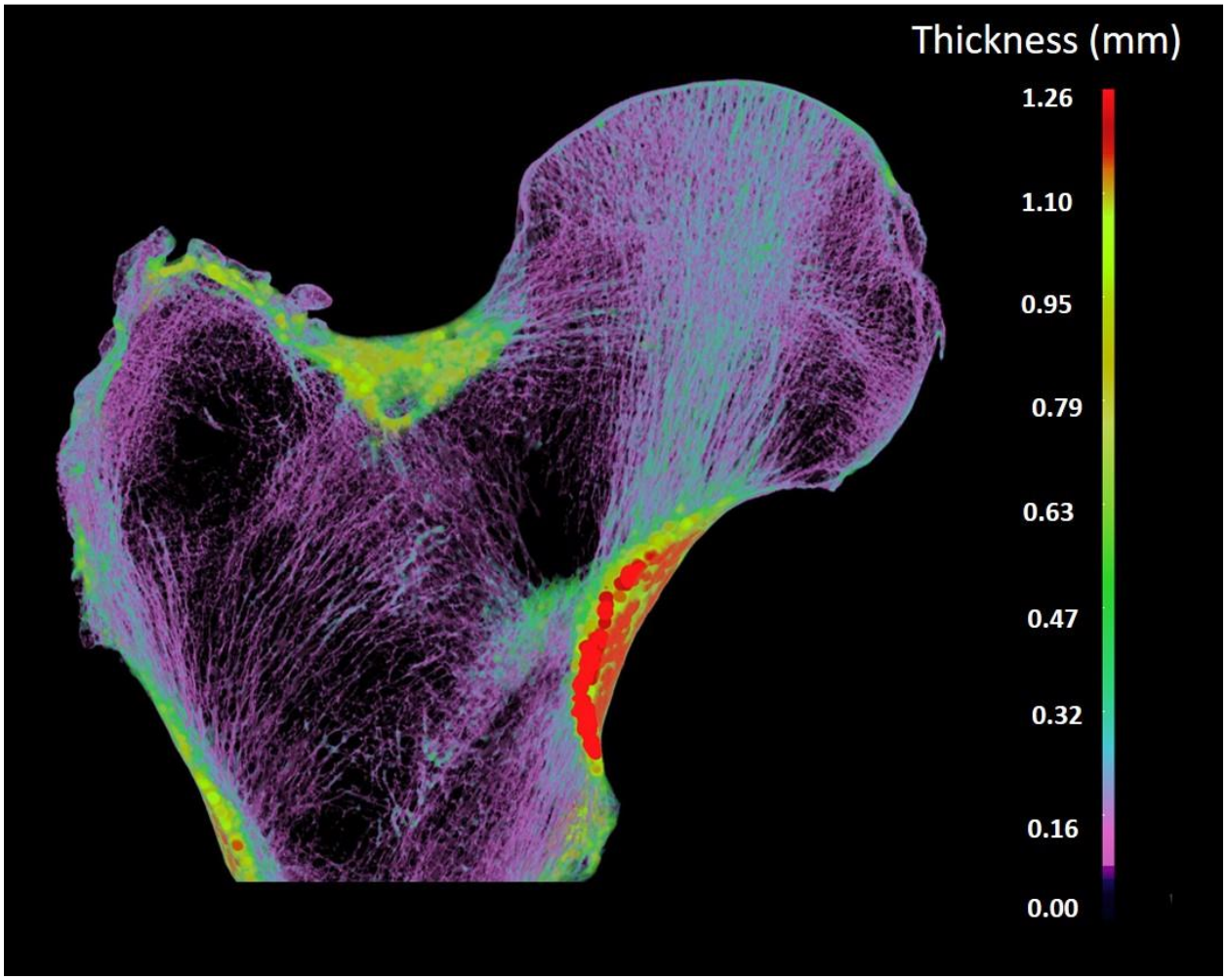


Figure 1.5 Volume thickness map of a proximal femur demonstrating the wide variability of cortical and trabecular microarchitecture in the proximal femur.

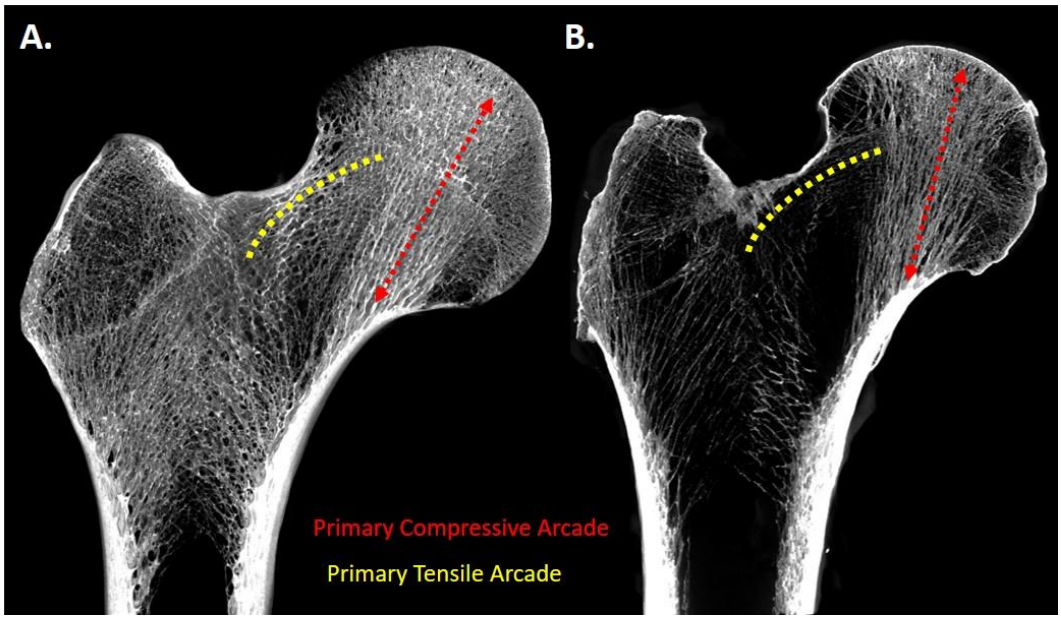


Figure 1.6 Nano-CT scans of a (A) 27-year-old and (B) 91-year-old female. With age the primary compressive arcade (red) remains intact while the primary tensile arcade (yellow) is resorbed.

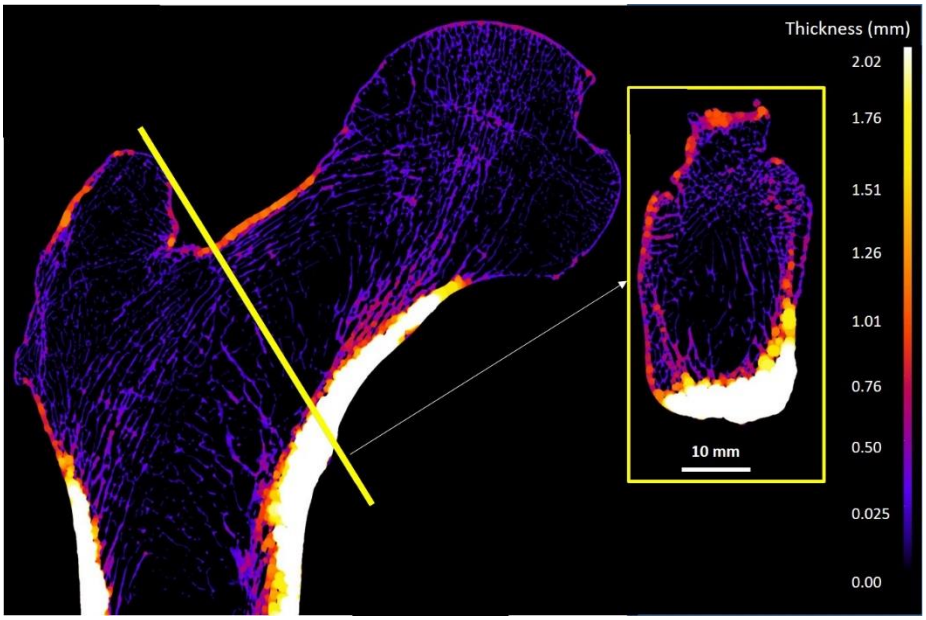


Figure 1.7 A volume thickness map of a femur with a coronal cross-section of the femoral neck. The volume thickness map demonstrates large variation in thickness, porosity, and bone density depending on the anatomical location.

Tables

Table 1.1 A brief description of variables commonly used to quantify cortical and trabecular bone microstructure.

Variables	**Definition**
BV/TV or BVF	The ratio of bone volume to the total volume of interest.
Tb.Th	The mean thickness of trabecular bone.
Tb.Sp	The mean distance between trabeculae.
Tb.N	A measure of the average number of trabeculae per unit length.
Tt.AR	Mean total cross-sectional area (calculated on a slice by slice basis).
Ct.Ar	Mean total cortical area (calculated on a slice by slice basis).
Ct.Ar/Tt.Ar	The ratio of Ct.Ar to Tt.Ar (described above).
Ct.Th	The mean cortical thickness.

Chapter 2 Bone Degeneration in Young Females Following Anterior Cruciate Ligament Injury

Anterior Cruciate Ligament Epidemiology

Injury of the anterior cruciate ligament (ACL) is a serious public health burden that is particularly prevalent amongst young individuals (<20 years) (42). The ACL, a ligament that connects the femur to the tibia, is a critical stabilizer of the knee and has little regenerative capacity (9). If a patient hopes to maintain an active lifestyle, surgical intervention is required to repair the ACL tissue (43). Despite 75-97% of patients reporting good outcomes following ACL reconstructive surgery (44–46), the long recovery time (6-9 months), costly hospital visits, and high level of associated morbidity (e.g. osteoarthritis) confirm such intervention is problematic and taxing to many individuals (6). In addition, as many as 5.8% and 12% of the primary reconstructed ACL grafts will fail within 5 and 15 years, respectively (47,48). Revision ACL reconstructions following graft failure exacerbate the aforementioned health concern (i.e., long recovery time and morbidity) since results of this procedure are inferior to the primary ACL reconstruction (49). Thus, understanding the mechanism of primary reconstruction failure is critical to help mitigate some of the worst-case outcomes following ACL injury.

The Role of Bone in the Reconstruction of an Injured ACL

ACL reconstructive surgery is technically complex and requires that an auto- or allo-graft (e.g., bone-patellar tendon-bone or hamstring tendons) be anchored in place to act as a scaffold for the tendon to form a new ACL. The anchoring sites of the graft are within the posteromedial lateral femoral condyle and the anteromedial tibial plateau of the knee joint.

A surgeon typically drills an anatomically oriented tunnel through the native ACL insertion sites with a trephine (Figure 2.1). The ACL graft can be threaded through the bone tunnels and anchored in tension using a bio-absorbable screw or hook. The bone anchoring sites are critically important in reconstruction since surgical success requires that adequate osteo- and

ligamentous integration take place for a new ACL to form. As evident from the phenomenon of tunnel expansion and bone osteolysis following both primary and revision ACL surgeries, insufficient biological integration between the auto- or allograft and native bone severely impacts patient outcomes and long term surgical success (13,14).

ACL Revisions: Contributory factors

Since results of revision ACL reconstruction are inferior to primary ACL reconstruction, considerable effort has focused on understanding the underlying mechanism of ACL revision failure (13,22,43,44,47,49,50). Reasons for ACL graft and revision failure can be categorized into technical errors, biologic factors and traumatic re-injury (13,22). A summary of each failure type is described in more detail below (Table 2.1).

Research on revision risk has largely focused on technical errors, which can account for up to 22 – 79% of error rates in acute graft failures (<6 months) and ~27% in late (>6 months) graft failures (22,49). More specifically, prior research has largely focused on errors that occur as a result of the tunnel position and the type of graft used in reconstruction (22,44,48,49,51). However, the mechanism of ACL graft failure is often multifactorial (22). While biological factors are a complex pathological entity, the biologic "health" of the ACL enthesis may also play a critical role in revision risk, even when it is not the primary cause of failure. For example, Harner reported that 14% and 25% of biological factors were either a direct or contributory cause of ACL graft failures (52). There is a need to better understand the biological factors that contribute to ACL revision risk. Despite the looming possibility of poor graft integration among patients who have received an ACL reconstruction, the condition of the mineralized matrices into which the ACL tunnel is drilled and the ACL graft is placed has remained largely unstudied.

Bone Loss Following ACL Injury

Likely due to altered loading kinematics, clinically measured bone mineral density (BMD) in the distal femur and proximal tibia decreases following an initial ACL injury (15–17). However, to determine the structural and compositional state of the mineralized matrices in

ACL injured patients, we need to understand changes in the underlying cortical and trabecular bone microstructure. Clinical BMD, while useful, cannot quantify microstructure or adequately inform on the mineralized changes occurring in the small region (e.g., 50 mm^3) adjacent to the femoral ACL enthesis, where ~95% of non-contact ACL ruptures occur (53). These details in architecture are needed to understand how the existing cortical and trabecular tissues within the femoral ACL enthesis affect the osseointegration of the repair graft. Understanding bone quality at the ACL enthesis and after ACL tears may help inform on decisions regarding the optimal timing for ACL reconstruction relative to the injury date. The objective of this study was to characterize architectural changes occurring within the femoral ACL enthesis of young females who suffered an ACL injury at a range of time intervals from ACL failure until reconstructive surgery (on average 8-12 weeks post-injury). We tested the hypothesis that injured femoral ACL entheses will show a significant decrease in cortical and trabecular bone mass compared to non-injured controls.

Methods

Sample population

Femoral ACL enthesis explants (10 mm in diameter) were collected from the injured knee of fifty-four female patients during ACL reconstructive surgery. Subjects ranged in age from 13 to 25 years. De-identified data recorded for most patients included age, sex, activity at moment of injury, time from injury to surgery and location of injury (Figure 2.2).

Patient explants were collected by one surgeon (EMW) to minimize variation in arthroscopic techniques. For explant extraction, a 10 mm diameter trephine was used in addition to standard 'outside-in' surgical practices (Figure 2.3). Upon extraction, femoral explants were stored at 4° C in 1x phosphate-buffered solution (PBS) and imaged within 72 hours. In addition to the patient tissue, twelve control femoral explants were acquired from paired knees of five female cadaveric donors and two additional unpaired knees from two donors ranging in age from 18 – 36 years from the Gift of Life Michigan and the University of Michigan Medical School. Donor knees were harvested within 48 hours following death and frozen at -20° C until ACL femoral explants could be extracted. Donor explants were extracted using equipment and techniques

identical to that used in the clinical setting. Control tissues were stored at 4° C in 1x PBS and imaged three-dimensionally within 72 hours. Both patient and cadaveric tissue use were approved by the University of Michigan Institutional Review Board and given exempt status.

Scanning preparation and acquisition

High resolution (14 um voxel size) scans of femoral ACL explants were acquired using a nanotom-s computed tomography system (phoenix|x-ray, GE Measurement & Control; Wunstorf, Germany) and consistent acquisition parameters (80 kV, 300 µA, 68 minutes, 1000 ms exposure time, 1000 images, 0.012'' aluminum filter). During the course of this study the nanotom-s was replaced with a nanotom-m (phoenix|x-ray, GE Inspection Technologies; Skaneateles, NY, USA), resulting in 22 of the femoral ACL explants included in this study being scanned on the newer system (70 kV, 300 µA, 34 minutes, 5000 ms exposure time, 1000 images, 0.012'' aluminum filter). The image acquisition parameters of the new system were adjusted so the two systems generated similar grey values (< 2% difference). Femoral explants were imaged in a 5 mL polypropylene scintillation vial, surrounded by polyurethane foam to prevent movement and saturated in 1x PBS to maintain tissue hydration. A calibration phantom containing air, water and a hydroxyapatite mimicker (1.69 mg/cc; Gammex, Middleton, WI, USA) was included in each scan. Image volumes were reconstructed using datos|x reconstruction software (phoenix|x-ray, GE Sensing and Inspection Technologies, GmbH, Wunstorf, Germany).

Volumetric analysis

Grey values from each reconstructed image were converted to Hounsfield units using the calibration phantom as described previously (54). Each explant was reoriented along the anteroposterior and mediolateral anatomical axes based on the curvature of the cortical shell using MicroView 2.0 software (Parallax Innovations, Inc., Ilderton, ON, Canada), and the cortical and trabecular matrices of each explant were manually segmented into two separate volumes of interest (VOI) (Figure 2.4). Each 10 mm diameter explant contained a 2.5 mm guide-pin hole due to the tissue extraction technique. This pin-hole was not included (i.e., digitally removed from all VOIs) in our analysis. Volumetric bone mineral density (vBMD),

relative bone volume fraction (BV/TV) and porosity (1 – BV/TV) were quantified for the cortical VOI. For the trabecular VOIs, analyses were standardized to 3.5 mm of trabecular bone adjacent to the most inferior aspect of the cortical matrix. Trabecular (Tb.) variables quantified included vBMD, BV/TV, trabecular thickness (Tb.Th) and trabecular spacing (Tb.Sp) in MicroView 2.0 software.

Regional Variation in Bone Microstructure Near the Femoral Enthesis

While explants extracted from patient ACL reconstructive surgery and control donors were completed exclusively by one individual (EMW and SHS, respectively), there is the possibility that the angle of the femoral explant tunnel relative to the enthesis may vary. To test how the angle of insertion of the trephine affects bone explant microstructural properties, one 18-year-old female cadaveric right knee was scanned in the nanotom-M at 60 μm resolution (110 kV, 200 μA, 95 minutes, 500 ms exposure time, 1900 images, 0.030'' aluminum filter) pre- and post- explant extraction. Image volumes were reconstructed using datos|x reconstruction software (phoenix|x-ray, GE Sensing and Inspection Technologies, GmbH, Wunstorf, Germany). Using Dragonfly software 4.0 (ORS, Montreal, Canada), image volumes were co-aligned to the femur pre- and post-extraction, the bone region was identified, and cortical and trabecular regions were manually segmented. Segmentation utilized a paintbrush tool able to highlight regions only within a certain threshold range and relied on manual correction within Dragonfly. Finally using the cylindrical tunnel as the reference ROI, the tunnel angle was digitally rotated by 10°, 20°, and 30° both medially and laterally relative to the initial tunnel location (Figure 2.5). Cortical and trabecular BV/TV were quantified ~3.5 mm adjacent to the inferior cortical matrix.

Statistical analysis

All data were analyzed using RStudio Team (2015) (RStudio, Inc., Boston, MA). A nonparametric Wilcoxon signed-rank test was conducted to determine if there were significant asymmetric differences between femoral explants harvested from the left and right knees of the control cadaveric donors (n=5). Results from control donors with explants from paired knees were averaged since there was no significant difference in cortical or trabecular results (See

Cortical and Trabecular Bone in Results Section). Individual data points acquired from each patient variable and each cadaveric control variable were combined to create injured (n=54) and non-injured (n=7, unpaired explants from 2 cadavers and 5 cadavers with bilateral explants) groupings. This was necessary to account for the inherent biological variation among individuals since the distribution of demographic data was not well powered (see Figure 2.2). On pooled data, an unpaired t-test with Welch's correction (i.e., to correct for unequal variances and/or sample sizes) was performed to test whether the cortical and trabecular parameters differed significantly between patient and control explants. Linear regression analysis as well as unpaired t-test with Welch's correction in groups of different time frames (1 - 7, 8 – 11, 12 – 16, and 17 + weeks) from injury to operation was used to test how time from injury to operation affected the architectural parameters derived from the cortical and trabecular VOIs. In order to assess the potential effect of age on bone microstructure, a multivariate regression including age, time from injury to operation, and the interaction of the two variables was used to identify significant independent predictors of bone microstructure. data. An alpha of 0.05 was used for all statistical analyses to identify significant differences between groups.

Results

Sample Population

Basketball (31%) and soccer (24%) were the most prevalent activities resulting in ACL injury in our cohort. The most frequent location of these injuries was adjacent to the femoral enthesis and the range of time from injury to operation varied from 4 to 78 weeks. The majority of primary reconstructive surgeries occurred within the first 16 weeks (85%). The median time in which a primary reconstructive surgery was completed post-ACL injury was 10 weeks.

Cortical Bone

The paired left and right femoral explants of the five control non-injured (NI) cadaveric donors showed no significant difference in the cortical measures (vBMD, $p = 0.313$; BV/TV, $p = 0.313$; porosity, $p = 0.438$). The injured (I) explants showed significant differences in vBMD (NI: 736.1 – 867.6 mg/cc; I: 451.2 – 891.9 mg/cc; $p < 0.001$), BV/TV (NI: 0.674 – 0.867; I:

0.401 – 0.792; $p = 0.001$), and porosity (NI: 0.133 – 0.326; I: 0.209 – 0.600; $p = 0.001$) compared to the non-injured femoral explants (Figure 2.6). F-tests comparing variances between non-injured and injured femoral explants was not significant for any cortical parameters (vBMD, $p = 0.112$; BV/TV, $p = 0.490$; porosity, $p = 0.489$). Large qualitative differences in cortical bone porosity between non-injured and injured femoral explants can be observed visually (Figure 2.7).

Linear regressions of patient explants showed no significant positive or negative association between any parameters and time between injury to surgery when control data was not included in the analysis (vBMD, $p = 0.284$; BV/TV, $p = 0.175$; porosity, $p = 0.177$). Explants were grouped by time from injury to operation (Control, 1-7, 8 – 11, 12 – 16, and 17 + weeks from injury to operation). The injured (I) explants in all groups showed significantly lower vBMD (1 - 7 weeks, $p < 0.001$; 8 - 11 weeks, $p = 0.007$; 12 - 16 weeks, $p = 0.002$; 17 + weeks, $p = 0.005$) and BV/TV (1 - 7 weeks, $p < 0.001$; 8 - 11 weeks, $p = 0.006$; 12 - 16 weeks, $p = 0.002$; 17 + weeks, $p = 0.008$) and higher porosity (1 - 7 weeks, $p < 0.001$; 8 - 11 weeks, $p = 0.007$; 12 - 16 weeks, $p = 0.002$; 17 + weeks, $p = 0.008$) compared to the non-injured femoral explants (Figure 2.8). No significant difference was found between patient groups for the cortical parameters.

Trabecular Bone

There were no significant differences in the trabecular measures between the paired left and right femoral explants removed from the five control non-injured (NI) cadaveric donors (vBMD, $p = 0.125$, BV/TV, $p = 0.313$; Tb.Th, $p = 0.313$; Tb.Sp, $p = 1.000$). The injured (I) vBMD was significantly lower (NI: 364.5 – 424.3 mg/cc; I: 246.7 – 529.6 mg/cc; $p = 0.013$) (Figure 2.9) compared to the non-injured femoral explants. Non-significant differences in BV/TV ($p = 0.314$), Tb.Th ($p = 0.412$), and Tb.Sp ($p = 0.828$) between non-injured and injured femoral explants were observed, suggesting that trabecular bone quality was not significantly affected by ACL injury in our patient cohort (Figure 2.9). F-tests comparing variances between non-injured and injured femoral explants were significant for vBMD ($p = 0.016$) but for no other trabecular variable (BV/TV, $p = 0.301$; Tb.Th, $p = 0.841$; Tb.Sp, $p = 0.543$). In addition, linear regressions showed no association between vBMD, BV/TV, Tb.Th, and Tb.Sp at the

time that had elapsed post-injury and prior to reconstructive surgery (vBMD, $p = 0.391$; BV/TV, $p = 0.284$; Tb.Th, $p = 0.157$; Tb.Sp, $p = 0.569$). Explants were grouped by time from injury to operation (Control, 1 - 7, 8 – 11, 12 – 16, and 17 + weeks from injury to operation) for vBMD. The injured (I) explants only showed significantly lower vBMD at 17 weeks or greater from injury to operation ($p = 0.018$) but for no other time frame (1 - 7 weeks, $p = 0.153$; 8 - 11 weeks, $p = 0.132$; 12 - 16 weeks, $p = 0.145$), compared to the non-injured femoral explants (Figure 2.10). No significant difference was found in vBMD between patients grouped by time from injury to operation.

Age and Time from Injury to Operation Effects on Bone Microstructure

For cortical variables, time from injury to operation (vBMD, $\beta = 6.7$, $p = 0.283$; BV/TV, $\beta = -4.8e\text{-}4$, $p = 0.932$; porosity, $\beta = 6.7$, $p = 0.283$), age (vBMD, $\beta = 12.1$, $p = 0.113$; BV/TV, $\beta = 3.8e\text{-}4$, $p = 0.956$; porosity, $\beta = 12.1$, $p = 0.113$), and the time from injury top operation-age interaction (vBMD, $\beta = -0.4$, $p = 0.212$; BV/TV, $\beta = -2.5e\text{-}5$, $p = 0.932$; porosity, $\beta = -0.4$, $p = 0.211$) were non-significant predictors of cortical bone microstructure (vBMD, R^2[adj.] = 0.02, $p = 0.295$; BV/TV, R^2[adj.] = -0.02, $p = 0.612$; porosity, R^2[adj.] = 0.02, $p = 0.295$). For trabecular variables, time from injury to operation (vBMD, $\beta = 0.64$, $p = 0.869$; BV/TV, $\beta = 3.2e\text{-}3$, $p = 301$; Tb.Th, $\beta = -1.3e\text{-}3$, $p = 0.509$; Tb.Sp, $\beta = 5.4e\text{-}3$, $p = 0.251$), age (vBMD, $\beta = 1.1$, $p = 0.818$; BV/TV, $\beta = 5.2e\text{-}3$, $p = 0.179$; Tb.Th, $\beta = -2.2e\text{-}3$, $p = 0.352$; Tb.Sp, $\beta = 1.8e\text{-}3$, $p = 0.756$), and the time from injury top operation-age interaction (vBMD, $\beta = 9.4e\text{-}3$, $p = 0.963$; BV/TV, $\beta = 1.5e\text{-}4$, $p = 0.349$; Tb.Th, $\beta = 5.3e\text{-}5$, $p = 0.609$; Tb.Sp, $\beta = -2.7e\text{-}4$, $p = 0.286$) were non-significant predictors of cortical bone microstructure (vBMD, R^2[adj.] = -0.04, $p = 0.832$; BV/TV, R^2[adj.] < 0.01, $p = 0.397$; Tb.Th, R^2[adj.] < 0.01, $p = 0.407$; Tb.Sp R^2[adj.] = - 0.02, $p = 0.585$).

Regional Variation in Bone Microstructure Near the Femoral Enthesis

Due to the small sample size (n=1), significant differences in bone microstructure relative to the orientation of the explant could not be assessed. However, for cortical results, it is clear that despite some variation in cortical BV/TV (0.88 - 0.94), all values in our control knee were higher than the patient cortical BV/TV in our explants (I: 0.401 – 0.792). Trabecular BV/TV

increased as the explant moved medially to laterally across the medial condyle of the distal femur (Table 2.2).

Discussion

Our findings support the hypothesis that there would be substantial bone loss within the femoral ACL enthesis following ACL injury. The femoral explants removed at the time of surgery revealed extensive and significant differences in cortical vBMD, BV/TV, and porosity between patient and control explants. Patient explants also showed a significant reduction in trabecular vBMD compared to controls. This trabecular reduction in vBMD could not be explained based on trabecular architectural changes. Further, our data support the hypothesis that increased time from injury to operation may result in greater mineralized tissue loss within the femoral ACL enthesis.

While changes in trabecular vBMD were observed, bone architectural changes following ACL injury were most evident at the cortical tissue comprising the femoral ACL enthesis. Compared to controls, patient explants had significantly lower BV/TV and vBMD and increased porosity, indicating that significant tissue mineralization loss and bone degeneration in the subchondral bone occurred post-injury. Bone loss following injury around the knee (both pre- and post-surgery) has been widely shown using clinical dual-energy X-ray absorptiometry (DXA) (55–60). However, prior studies focused on regional changes after extensive time had passed (~1 to 12 years post-surgery) and were largely focused on cancellous bone, the most affected tissue following prolonged immobilization, within the distal femoral and proximal tibial metaphyses (15,61). To our knowledge, the more variable subchondral bone that may be sclerotic, less homogenous, and is a part of the ACL enthesis, has not yet been explored in detail. In line with DXA-based studies, we believe altered weight-bearing and/or limb disuse could, in part, explain our observations.

When cortical morphology was sub-grouped by time from injury to operation, there was no significant difference in any patient sub-groups compared to each other. If altered weight-bearing and/or limb disuse were the only factors impacting bone quality, we would expect to

see continued bone loss over time (61). Instead, we observed a rapid significant decrease in cortical bone volume following ACL injury combined with an increase in bone remodeling activity (i.e., porosity) that is not recovered before ACL reconstructive surgery. Thus, a lack of weight-bearing activity and kinematic changes may not be the only mechanism driving degenerative bony changes following ACL injury. This idea is supported by Rittweger et al. who observed significant reductions in volumetric bone mineral content in patients five years after ACL reconstruction using a bone to bone (BTB) graft despite demonstrating a full recovery of knee extensor strength and patellar tendon stiffness (62).

We also observed a significant difference in trabecular vBMD but in no other trabecular measure (BV/TV, thickness, or spacing) in our patient group compared to controls. Our findings suggest that reduction in vBMD is not due to differences in microstructure and, instead, may be due to changes in the organic matrix (inorganic, organic, or water component) of the bone at the site of ACL injury (63). These findings were surprising because, as mentioned above, cancellous bone is the most affected tissue following prolonged immobilization (61). Kazakia et al. reported small but significant decreases in trabecular thickness and spacing but no change in trabecular BV/TV and BMD in patients who underwent a disuse period (six weeks) in the distal tibia compared to baseline (64). This is not fully in-line with our findings because we observed no significant difference in trabecular thickness, spacing, and BV/TV between injured and control explants. However, the bone at the ACL enthesis experiences tensile and shear forces (5) while the distal femur experiences mainly compressive forces (65). It is known that bone cells, such as osteoblasts, are mechanotransductory and respond to compressive and tensile forces differently (66). Perhaps, bones that are habitually loaded differently (i.e., compressive versus tensile and shear loading) are primed for a particular loading orientation and thus respond to changes in loads in a fundamentally different manner. However, such associations and speculations are complicated and would benefit from having access to a relevant animal model where the mineralized matrix and the underlying microstructure can be studied. Our patient population also demonstrated that there is considerable variation among patients (see Figure 2.7), despite that many fell well below (Tb.vBMD, Ct.vBMD, and Ct.BV/TV) or above (Ct.Porosity) the non-injured baseline. Time from injury to operation alone could not explain the difference observed in this study.

However, considering that the majority of the explants analyzed were from patients that were 8 to 12 weeks from injury, our results suggest that many patients have mineralized matrices that still may not be optimal for adequate osseointegration of an ACL auto-/allo-graft following reconstruction.

Our findings are novel in that these are the first assessments, using high resolution computed tomography imaging, of changes in cortical and trabecular architecture that occur at the exact location in which ACL graft fixation occurs at reconstruction. Further, we argue that our findings are directly related to ACL injury and cannot be simply related to cohort (e.g. same age, sex, and procurement) effects because we predominately observed architectural changes in cortical but not trabecular tissue. Less is known about how the mineralized matrices in a small focal area such as at the osseo-ligamentous interface of the ACL structurally and functionally responds to the application, or lack thereof, of mechanical forces. Despite the complex physiological kinematics of the knee, where the distal femur experiences both joint contact (reaction) forces and tibio-femoral (bone-on-bone contact) forces, loads directly at the ACL insertion site appear to largely shear and tensile forces from the ACL itself (5,67,68). Thus, based on our findings, in the context of what has been previously reported by others, we hypothesize that the bone loss observed in our patient femoral ACL explants is a highly localized example of the early disuse resulting from the loss of ACL loads (5). This conclusion is supported by the finding that there was a significant loss of localized trabecular vBMD and a significant increase in cortical porosity (an indication of remodeling activity) following ACL injury that is characteristic of rapid bone loss shortly (i.e., less than 2 months) after the matrix experiences disuse (69–73).

While none of the nanoCT scans of samples provided evidence of architectural disruption due to avulsion, we cannot rule out that some of the bone loss observed herein could also be the result of a microscopic subchondral avulsion of the calcified fibrocartilage in conjunction with the ACL rupture. Future studies need to characterize the viability of osteogenic cells in this region and characterize this site in animal models to better understand the early biological response to ACL tears in a setting where variables are better controlled (mechanism of tears, activity post-injury, sex, age, etc.). Nonetheless, clinically these findings may allude to an

important biological parameter that may affect ACL reconstruction success. Regardless of the mechanism responsible for these degenerative changes, our results indicate that the remodeling of mineralized tissue in the ACL enthesis occurs early following injury. Therefore, the timing of the re-establishment of tensile forces and weight-bearing activities on the ACL femoral enthesis appears to be a factor to consider in evaluating ACL reconstruction protocols. Supporting this conclusion, Tomita et al demonstrated that the timing of graft integration and the restoration of near-native mechanical properties differ between soft tissue and BTB grafts in dogs (74). Osseointegration may be jeopardized if the mineralized regions through which the ACL graft passed have degenerated, particularly if a soft tissue graft is used. Secondly, BTB grafts appear to fail less often than soft tissue grafts that lack a bony interface (75). For autograph reconstructions, additional research is needed to determine if the supply of viable bone cells (i.e., osteoblasts, osteocytes, chondro-fibroblasts) in the BTB constructs placed into the degenerated native bone of the femoral tunnel explain this divergent outcome between graft types.

There is additional concern for bone integrity loss after ACL injury when considering the phenomenon of tunnel expansion or osteolysis seen frequently with various ACL reconstruction techniques, both in primary and revision surgeries (14,59,76,77). This gradual tunnel expansion which has been seen with both auto- and allo-grafts may be facilitated by the localized loss of bony structural integrity in the enthesis when the ACL fails. ACL remnant preservation techniques have been utilized in an attempt to prevent tunnel osteolysis by preventing synovial fluid leaks from the knee joint (78). However, the root cause for bone degeneration may not be leaking synovial fluid but instead may be primarily from the prolonged loss of tensile forces on the ACL enthesis.

Other regions of the ACL-complex have demonstrated cell loss following an ACL injury that is time dependent with significant changes occurring within the first three months following injury (79). Cell apoptosis following injury has been documented in the ruptured ligament itself (80,81), but the cellular integrity of the mineralized matrices comprising the entheses remains unknown. The reality may be that the same process and timetable is occurring in bone. For those prescribing to ACL stump preservation during reconstructive surgery (82), the timing of

ACL surgery has become more important to minimize ligament cellular apoptosis. The same scenario may be in play within the mineralized entheseal zones of the ACL.

The primary limitation of this study was the small sample size in terms of the distribution of demographic data. Another limitation is that the 10 mm diameter trephine only provides a core sample of a portion of the ACL enthesis and the exact entheseal location of the sampling may vary (e.g. proximal vs. distal and medial vs. lateral) among patients and donors. However, based on our study testing the effect of extreme differences in angle insertion (medial vs. lateral) on resultant cortical and trabecular properties, angle related variability cannot account for the differences observed between control and patient explants.

The hypothesis that there would be substantial bone loss within the femoral ACL enthesis following ACL injury was supported in that our results suggest that the condition of the mineralized tissue, particularly that of the cortical matrix, into which a femoral tunnel is drilled and the ACL auto- or allo-graft is placed, may not be in a homeostatic remodeling state at the time most surgical interventions take place (~ 2 - 3 months following injury). Cortical bone volume and the density of the matrix were dramatically reduced 4 weeks out from injury, and showed little improvement over time compared to non-injured individuals. Trabecular bone vBMD decreased but was only significant 17+ weeks out from time of injury to operation, implying that the mechanism of trabecular bone loss may be a slow gradual decline. Moreover, the rapid pace at which bone remodeling occurred within this matrix may not provide a suitable population of bone-forming precursor cells within the first 12 weeks following injury, potentially jeopardizing the success of graft osseointegration, and thereby leading to graft failure and/or bone lysis within the tunnel.

Acknowledgements

I would like to gratefully acknowledge funding from the National Institutes of Health: NIH/NIAMS AR054821 (Edward M. Wojtys and James A. Ashton-Miller), AR070903 (Stephen H. Schlecht[SHS]) and AR069620 (Karl J. Jepsen [KJJ]) utilized to complete the work. Gratitude is also extended to Gift of Life Michigan and University of Michigan Medical

School for cadaveric and donor tissue used in this study. Lastly, I would like to thank Collin Martin, Michael Casden, Danielle Ochocki, Rebecca Falzon, and Lauren Battle for assistance with this project.

Figures

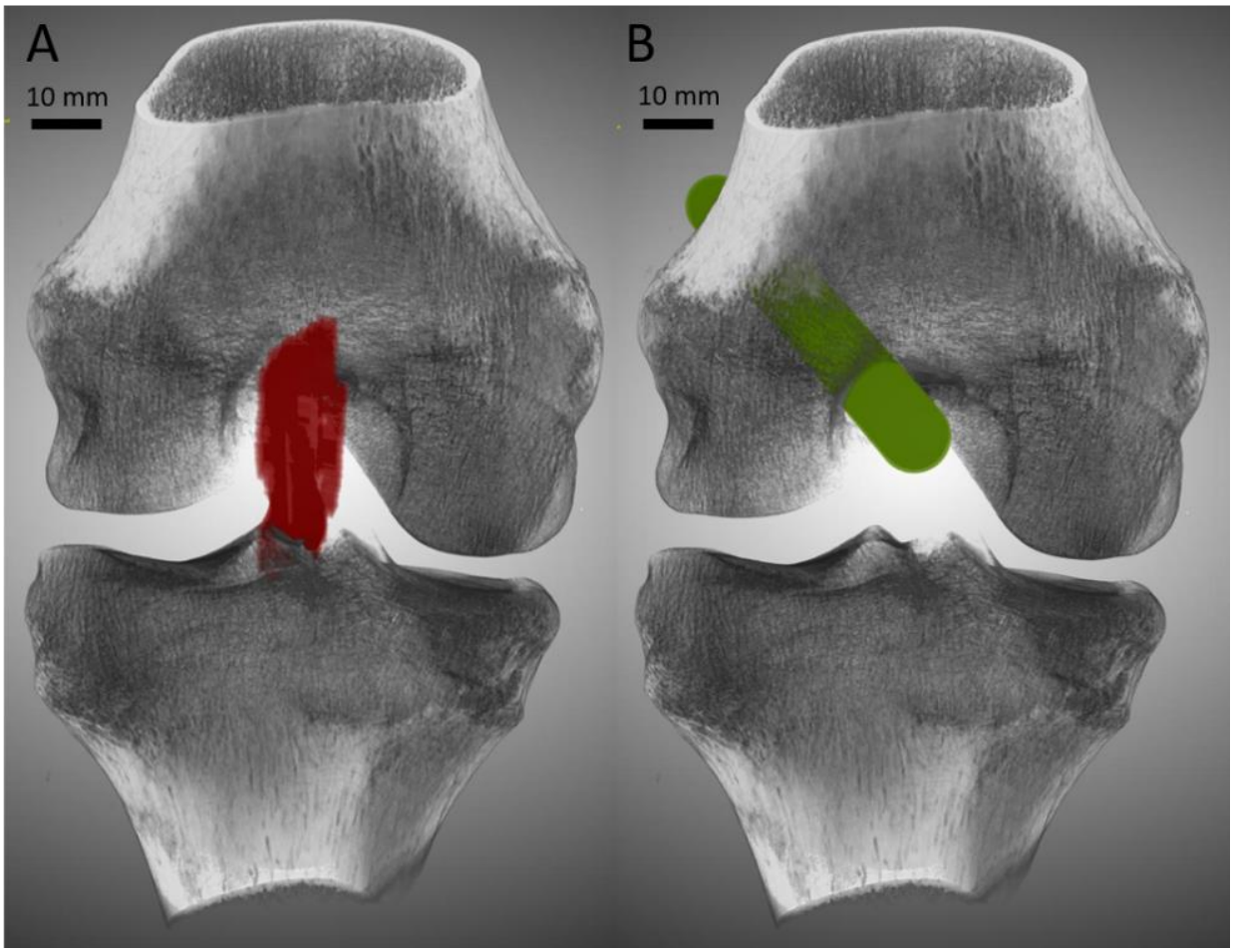


Figure 2.1 Female knee with an intact ACL (red) and (B) an example location of the femoral anchoring site created with a surgical trephine (green) (60 um, Nano-CT).

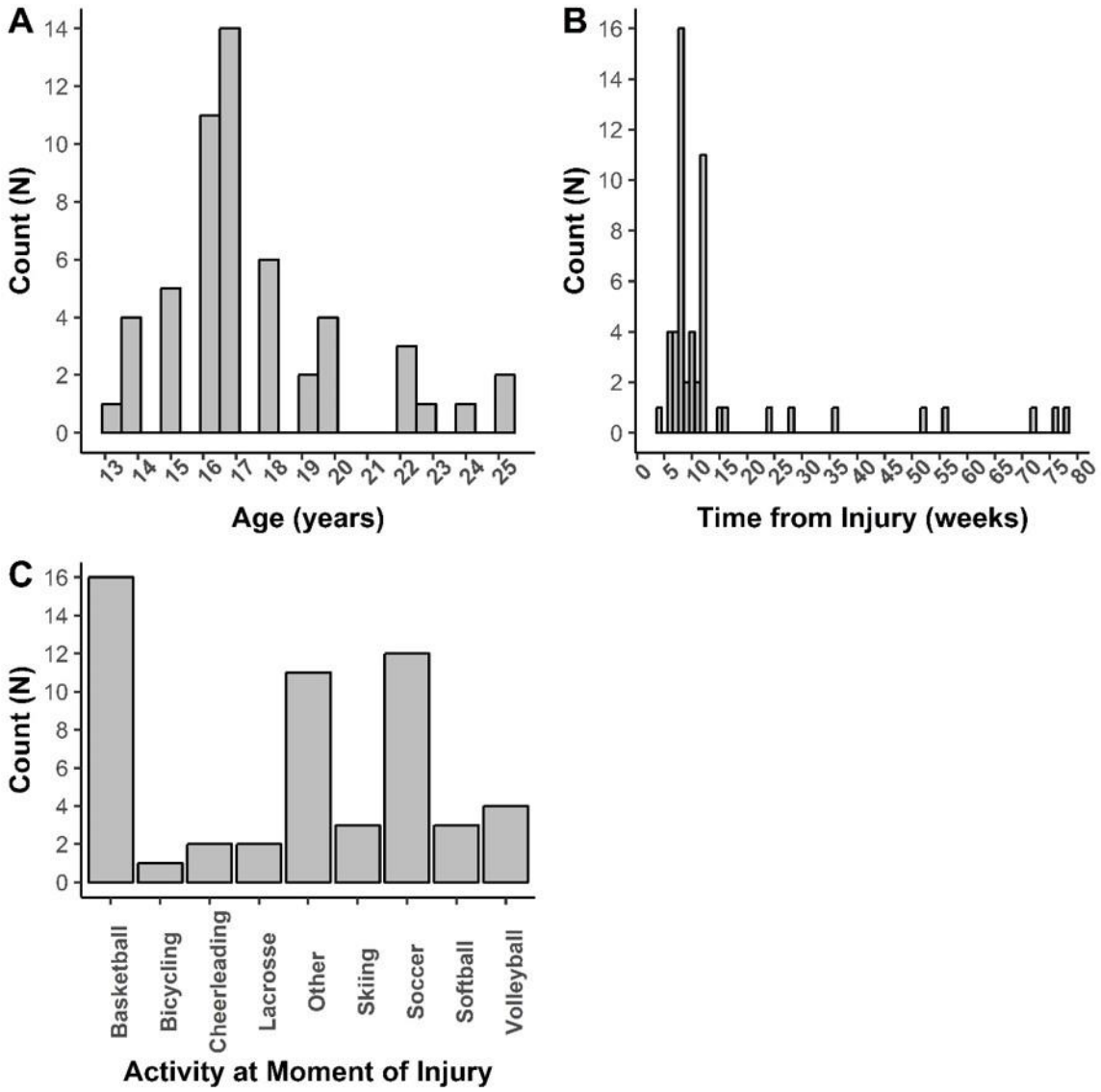


Figure 2.2 Histogram of patient (A) age, (B) time from injury to operation, and (C) activity at time of injury.

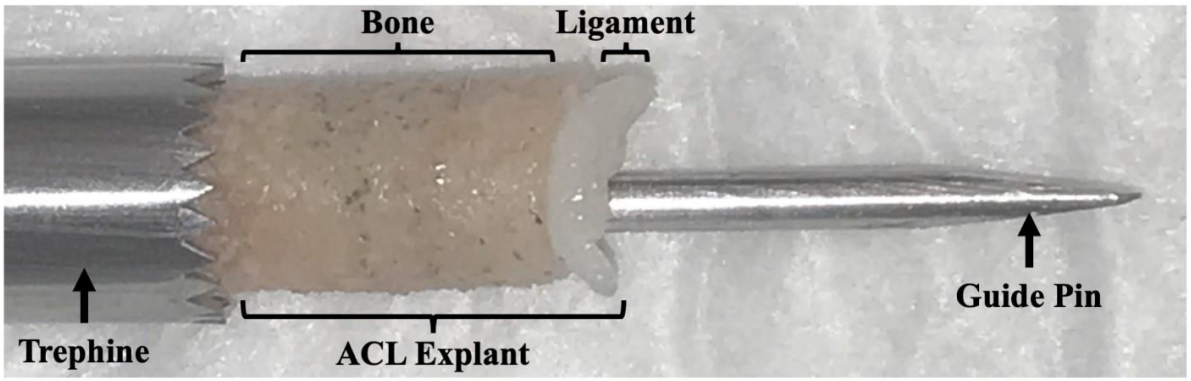


Figure 2.3 View of the 10mm trephine and guide pin used in the extraction procedure in both patients and cadavers. A sample femoral ACL explant is also shown. Reproduced with permission from Chen et al. 2019 (83).

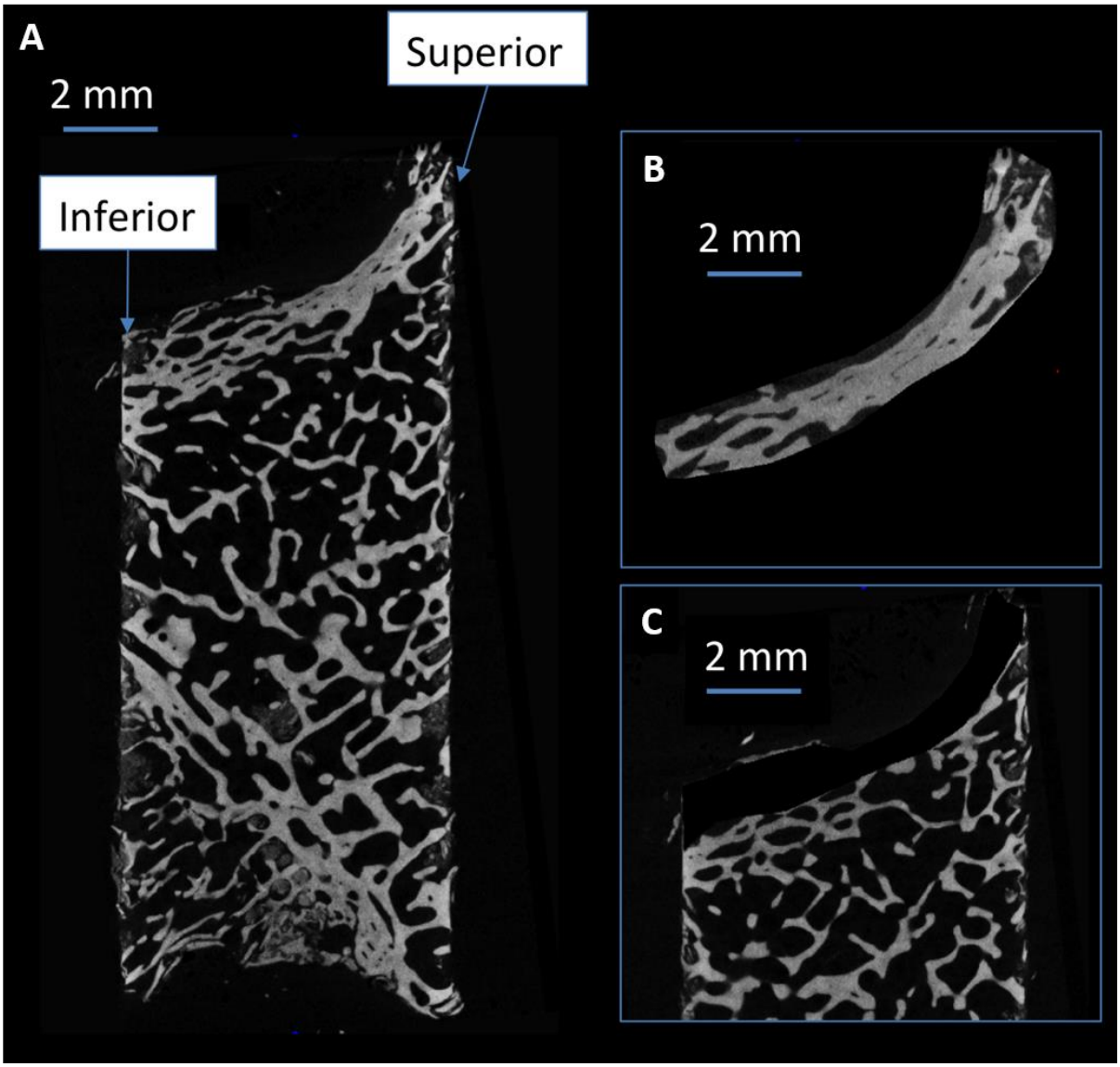


Figure 2.4 Representative explant: (A) reorientation of the explant in the y – z plane; (B) segmentation of the cortical VOI; and (C) segmentation of trabecular VOI.

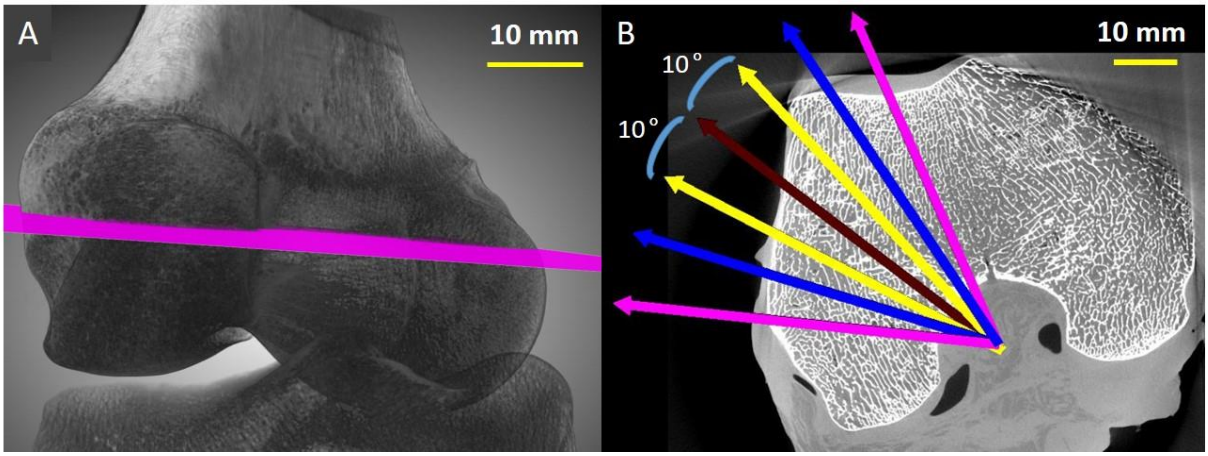


Figure 2.5 (A) The plane through which the explant VOI was rotated and (B) the angle in which the cylindrical VOI was rotated [10 °(yellow), 20 °(blue), and 30 ° (pink)] relative to the actual extraction point (mahogany) to measure bone morphology.

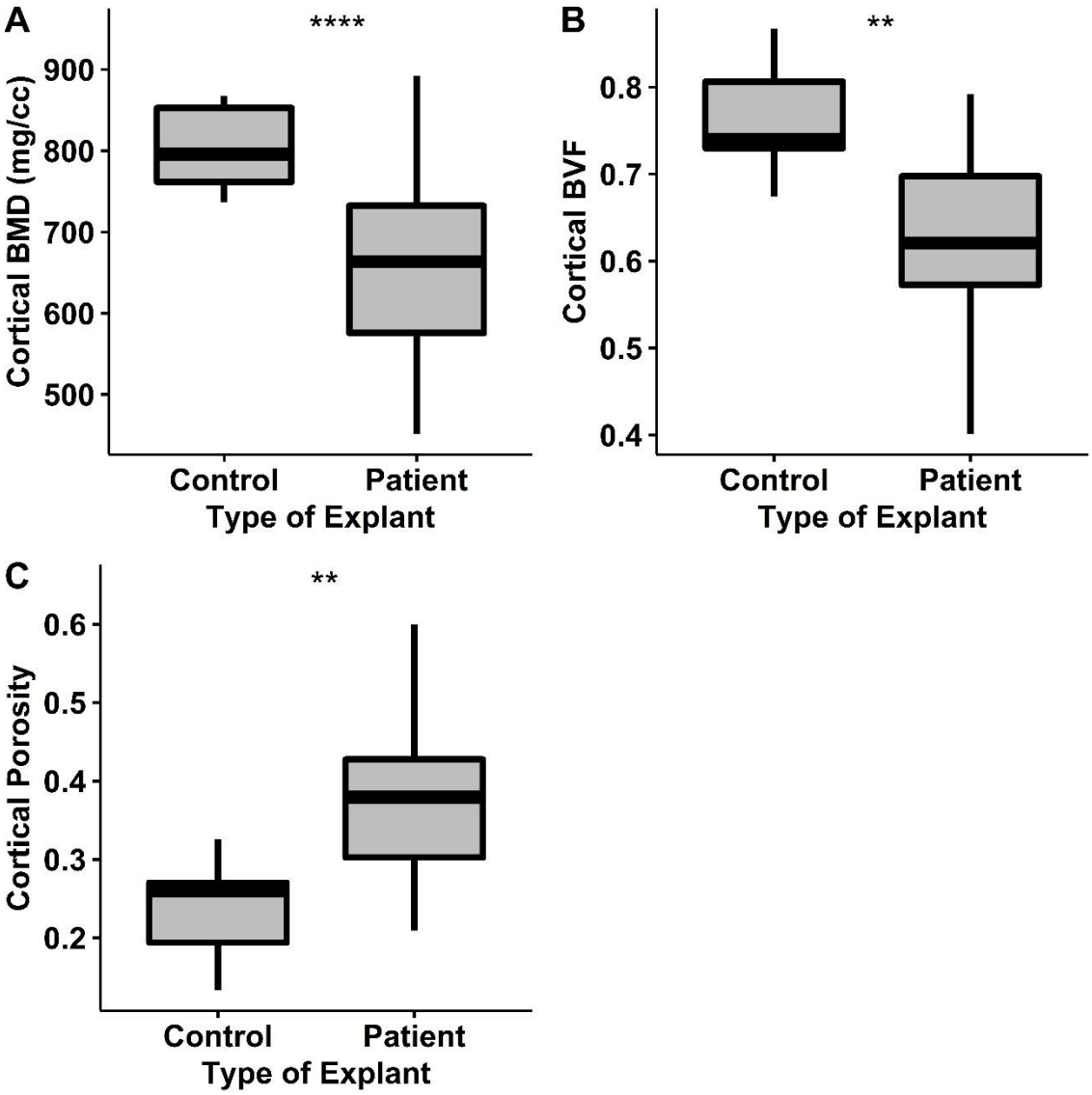


Figure 2.6 Unpaired t-test between patient and control explants for cortical (A) vBMD, (B) BV/TV, and (C) porosity.

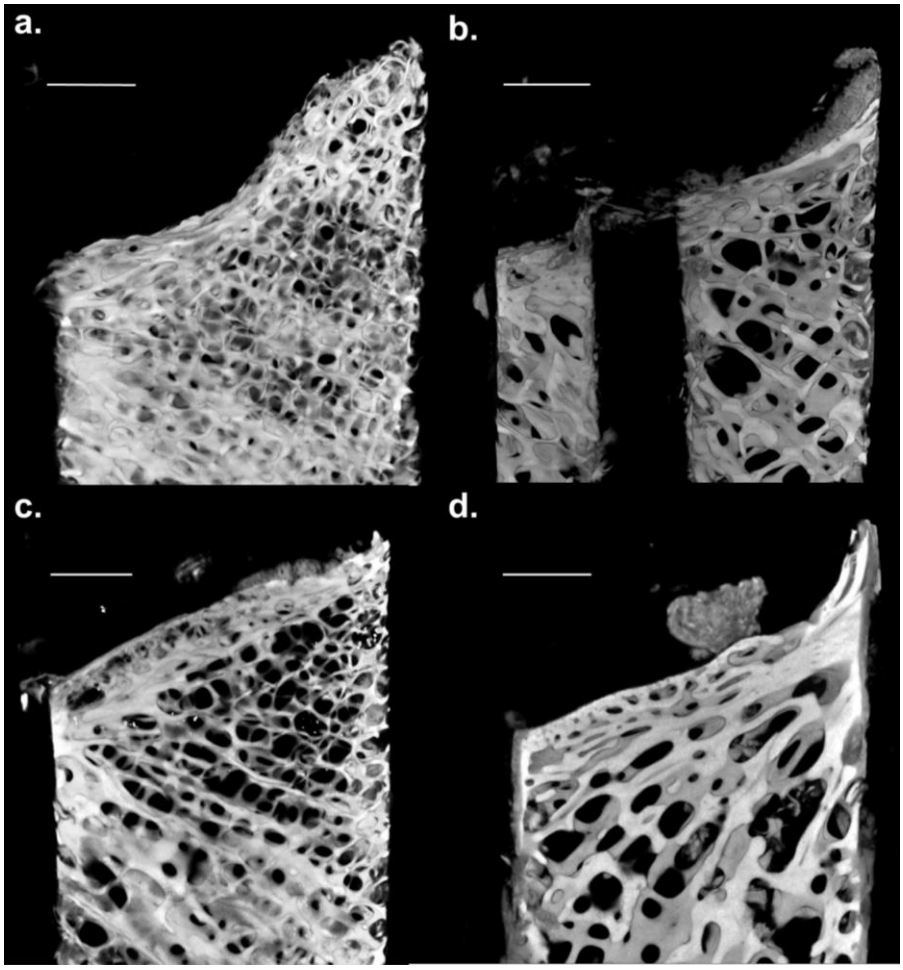


Figure 2.7 Three-dimensional scans of three patient explants with (a) little, (b) some, and a (c) significant amount of cortical porosity relative to the control (d). Note the presence of osteophytes (not included in analysis) at the entheseal boundary in both patient specimens (Scale bar = 2 mm).

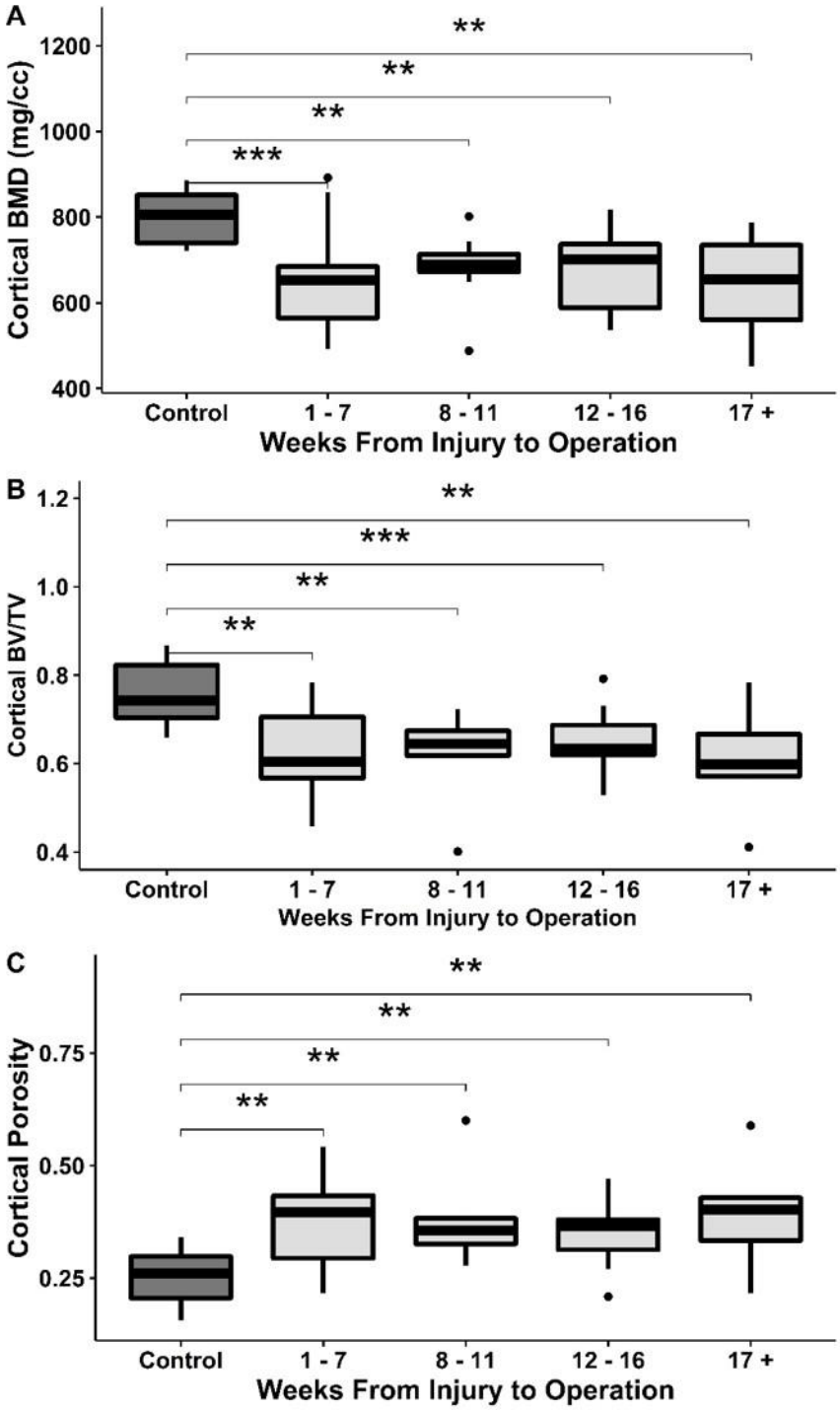


Figure 2.8 Unpaired t-tests between patient and control explants grouped by time from injury to operation for (A) vBMD, (B) BV/TV, and (C) porosity.

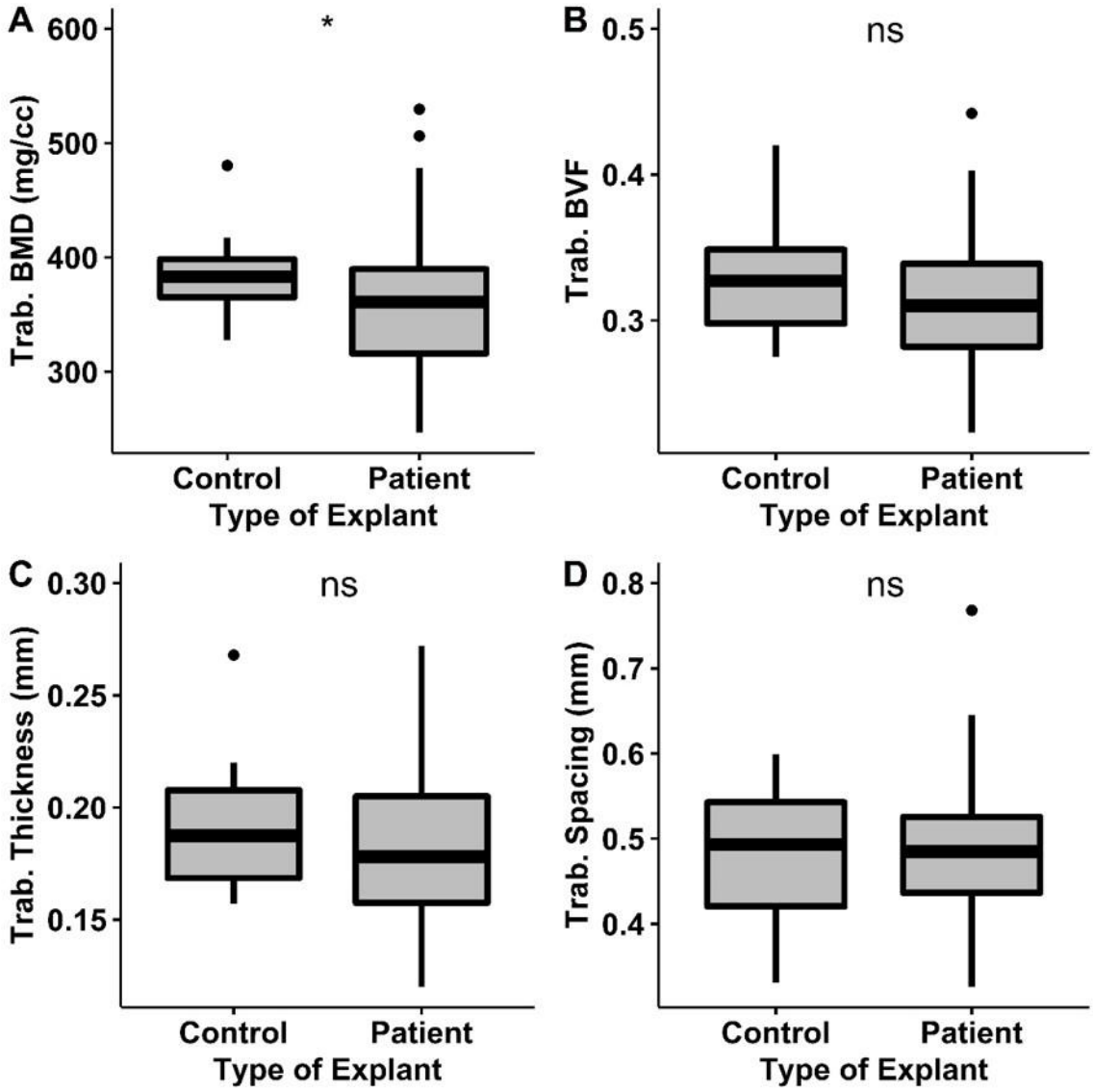


Figure 2.9 Unpaired t-tests between patient and control explants for trabecular (A) vBMD, (B), BV/TV, (C) thickness, and (D) spacing.

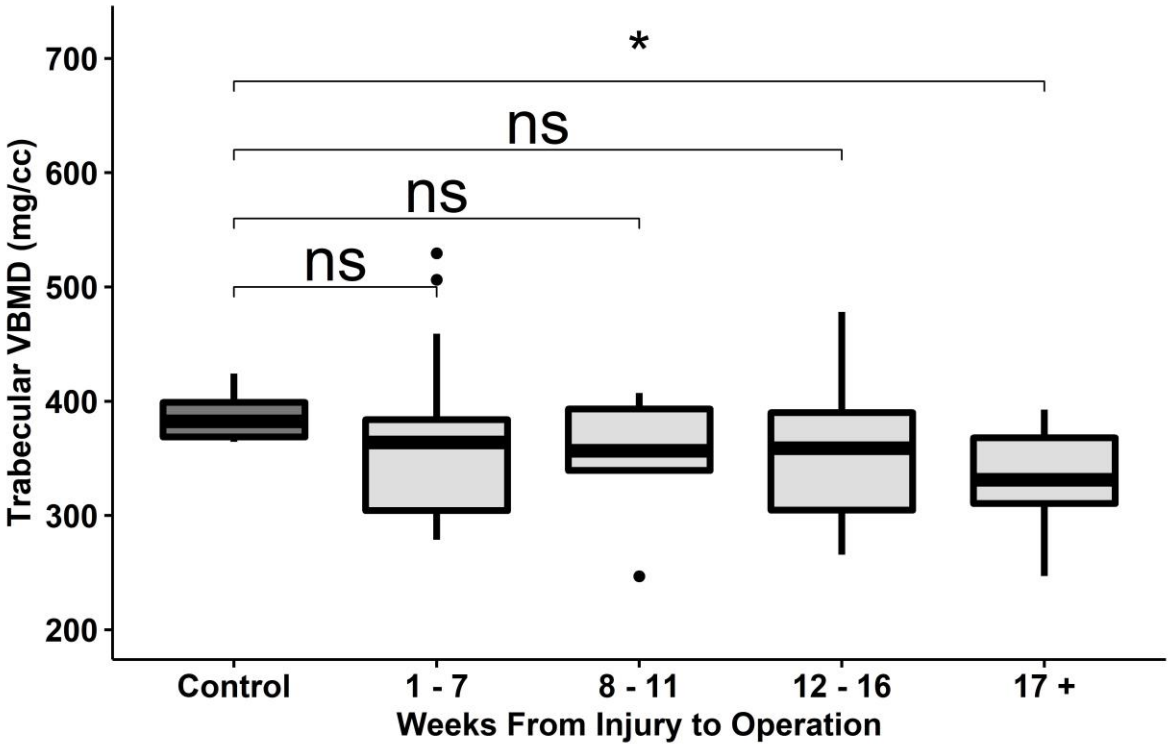


Figure 2.10 Unpaired t-tests between patient and control explants grouped by time from injury to operation for trabecular vBMD.

Tables

Table 2.1 A summary of the common categories of failure and specific reasons as to why a failure occurs within a specific category.

Failure Type	Reasons for ACL Failure	Citations
Technical Errors	Tunnel malposition (70-80% of surgeon error)	Wetzler et al., 1996; Bach Jr, 2003; Kamath et al., 2011; Spindler et al., 2004; Shaerf, 2014
	Inadequate ACL graft tissue or tension	
	Failure to recognize and treat surrounding injuries	
	Sup-optimal ACL surgical technique (varies by surgeon preference). Autografts (hamstring tendons [HS] and the "gold standard" bone patella tendon-bone [BPTB]), allografts, and synthetic grafts are all used for ACL reconstructive surgery	
Biologic Factors	Failure of graft incorporation	Graf, B.; Uhr, 1988; Shelbourne *et al.*, 1991; Harner *et al.*, 1992; Dye and Chew, 1993; Jaureguito and Paulos, 1996; Wetzler *et al.*, 1996; Bach Jr, 2003; Kamath *et al.*, 2011; Samitier *et al.*, 2015
	Biomechanical failure of the tissue (failure of "ligamentization")	
	Infection or rejection of the graft	
	Lack of bone plug healing	
	Osseous tunnel expansion or bone lysis	
Traumatic Re-injury	Direct blow to knee before complete graft incorporations and rehabilitation	Wetzler *et al.*, 1996; Bach Jr, 2003
	Direct blow to knee after resuming full activities	

Table 2.2 Differences in cortical and trabecular BV/TV when the angle in which the trephine drills out the explant it rotated about the femoral ACL enthesis.

Color	**Rotation Relative to Bone Analysis**	**Location**	**Ct.BV/TV**	**Tb.BV/TV**
Green	30	Medial	0.898	0.224
Blue	20	Medial	0.890	0.256
Yellow	10	Medial	0.882	0.286
Red	-	-	0.881	0.314
Yellow	10	Lateral	0.949	0.343
Blue	20	Lateral	0.893	0.357
Green	30	Lateral	0.879	0.347

Chapter 3 The Relationship between Whole Bone Stiffness and Strength is Age and Sex Dependent

Stiffness and Strength

Stiffness and strength are mechanical properties frequently used to define bone health (84). Stiffness is a measure of the load needed to induce a magnitude of deformation and, strength is a material's structural resistance to failure by fracture or excessive deformation (85). When loaded, bone is an elastic, visco-plastic material that exhibits both elastic (i.e., a linear relationship between stress and strain) and plastic behavior when loaded beyond (86–89). Plastic behavior occurs when a material deviates from linear proportionality such that load-induced microstructural rearrangements (i.e., deformation) cannot be reversed (85). In practice, bone stiffness and strength are quantified when bone is loaded in tension or compression using load-displacement or stress-strain curves (89,90). A load-deformation curve quantifies the amount of load needed to produce displacement (Figure 3.1) and is used to quantify stiffness, the slope of the linear portion of the curve, and strength, the maximum load a specimen can withstand before failure (89). The force required to fracture a specimen and the manner in which ultimate failure occurs depends many factors: the physical size of the specimen (i.e., length, height, width, and shape), the collagen and mineral make-up of the bone, and the stress experienced (i.e., tensile, compressive, and/or shear) (85).

For bone, generally speaking, it is believed that stiffness and strength reflect different properties: strength reflects structure and stiffness reflects mineralization despite that both stiffness and strength are related to a combination of bone mass, geometric distribution, and material properties (i.e., tissue composition) (84). As a result, the biomechanical status of bone may be poorly described if only stiffness or strength is reported. For instance, ex-vivo tests have shown that osteoporotic bones tend to be stiffer than healthy bones, but also more brittle, resulting in an increased risk of fracture (84).

Assumptions in bone on the stiffness strength relationship

Despite that stiffness and strength are different mechanical properties, researchers have shown that the relationship between bone stiffness and strength is linear and highly correlated (Table 3.1). Some have interpreted this to mean that bone stiffness, which is computationally easier to model (i.e., post-yield behavior modeling requires linear iterative methods to convert complicated nonlinear numerical problems into a sequence of linear problems), is a surrogate measure of strength (91,92). As a result, computational models such as finite element analysis (FEA) developed to estimate strength often model failure only to the elastic limit using the Pistoia criterion (i.e., the force where 2 - 4 % of the tissue exceeds a predetermined strain limit) (34,93,94). How additional demographic factors known to impact mechanical properties (e.g., sex and age) impacts the stiffness-strength relationship remains largely unexplored.

The relationship between whole bone stiffness and strength is age and sex-dependent (95)

A fragility fracture is a mechanical event that occurs when a low-energy force applied to the bone, such as during a fall from a standing height, exceeds bone strength and results in structural failure (89). Fractures occur through a process involving nonlinear material and structural behavior which leads to the accumulation of submicroscopic damage merging into a macroscopic crack (96). Since strength cannot be measured in situ, assessments of fracture risk rely on correlations between bone strength and surrogate indices, such as morphological traits (90) or results from engineering-based finite element analyses (FEA) (97). Noninvasive linear-elastic estimates of strength depend on a strong association between the in situ stiffness (linear, elastic deformation) and strength (non-linear, plastic deformation and failure) (34,93). While some FEA models use linear and non-linear estimates to predict bone strength (98), those that rely on linear computational techniques may not accurately predict strength due to assumptions that ignore nonlinearities in structural behavior (i.e., post-yield displacement and post-yield load) (34,93).

For tubular structures, like long bone diaphyses, stiffness is expected to correlate strongly with strength, because both measures depend on similar morphological and material properties (99). Whether a similar correlation between stiffness and strength holds for fracture-prone cortical-

cancellous structures (e.g., proximal femur) is not well understood. Correlations between stiffness and strength have been limited to studies conducted at the tissue-level, often for a single sex, and at the whole bone level but only for diaphyseal structures (86,92,100). Thus, the stiffness-strength relationship has been established at the tissue level but is not well understood at the whole bone level (101,102). To our knowledge, no published studies tested how the relationship between whole bone stiffness and strength varies between sexes and with age at different anatomical sites in fresh-frozen cadavers.

Overarching goal

The goal of this study was to test whether the relationship between stiffness and strength varies with sex and age. We tested the weight-bearing femoral diaphysis and proximal femur and the non-weight bearing radial diaphysis. Although men have stronger bones relative to body size compared to women (33), it is not known whether the relationship between stiffness and strength differs between sexes. With aging, bones become more brittle, thereby affecting crack toughening mechanisms (20). Nawathe et al. reported that changing tissue-level material properties from fully ductile to brittle (i.e., no post-yield displacement) using finite element modeling resulted in a ~40% decrease in the estimated whole bone strength in the proximal femur when loaded in fall-to-side fracture (103). We postulated that this age-related increase in brittleness, which we define as a decrease in post-yield displacement, would lead to premature propagation of the fatal crack and thus reduce whole bone strength in older bones beyond that which is predictable from stiffness. Thus, we hypothesize that the relationship between stiffness and strength will depend on age and post-yield displacement. Knowing whether the stiffness-strength relationship varies with age and sex at multiple whole bone sites is important for refining strength estimates, which may benefit efforts aimed at reducing fragility fractures by using accurate surrogate measures of bone strength as an alternate metric to diagnose and treat patients (34).

Material and Methods

Sample population

Table 3.2 shows a summary of the age distribution and the number of bones in all test groups. Bodyweight and height, measured at the time of death, were provided when medical history was present. Following procurement, bones were wrapped in PBS-soaked gauze and stored frozen at -40º C.

Mechanical testing of the long bone diaphysis

The proximal and distal metaphyses were embedded in square molds filled with acrylic resin (Ortho-Jet BCA, Lang Dental, Wheeling, IL, USA) using a custom alignment fixture (Figure 3.2). Specimens were aligned so the anterior-posterior-medial-lateral quadrants coincided with the flat sides of the acrylic blocks. The acrylic blocks interfaced with parallel aluminum guide walls to prevent specimen rotation during testing. The diaphyses were loaded to failure in four-point bending using an Instron 8511 materials testing system (Instron, Inc., Norwood, MA, USA) (104). Lower loading points were positioned at 25% and 75% of the bone length and upper loading points were positioned at one-third and two-thirds of the lower span length. Each sample was subjected to three pre-yield load-unload conditioning cycles before being loaded to failure at a displacement rate of 0.1 mm/s. The loading protocol was validated by testing aluminum cylinders and confirming that the derived material modulus was within 1% of textbook values. Femurs were loaded in the posteroanterior (PA) direction (anterior surface in tension) and radii were loaded in the medial-lateral direction (lateral surface in tension. Loading directions were chosen to coincide with the natural curvature of the bones. Both the radial and femoral diaphyses lack symmetry so results may differ for other loading orientations. Load-displacement curves were adjusted for test fixture geometry and used to determine bending stiffness (Nm^2), yield load (bending moment, Nm), post-yield deflection (1/m), post-yield load (Nm), and maximum bending moment (Nm) (104). Displacement was measured as the deflection of the upper loading points. The yield point was defined as the intersection between lines describing a 10% stiffness loss from the initial tangent stiffness with the load-displacement curve. Post-yield load (PYL) was calculated by subtracting the bending moment at yield from the maximum bending moment. Post-yield displacement (PYD) was calculated as the amount of deflection between the yield point and failure. For simplicity, bending stiffness, yield load, post-yield deflection, post-yield

load, and yield load will be referred to as stiffness, yield load, post-yield deflection, post-yield load, and strength, respectively, for the remainder of this study.

Mechanical testing of proximal femurs

Proximal femurs were cut 16.5 cm from the superior aspect of the femoral head. The femoral shaft was embedded in a 5 cm square aluminum channel filled with acrylic resin (Ortho-Jet BCA, Lang Dental, Wheeling, IL, USA) using a custom alignment fixture (91). Before mechanical testing, the proximal femurs were imaged using a nano-computed tomography system (nanotom-s, phoenix|x-ray, GE Measurement & Control; Wunstorf, Germany) (27 μm voxel size, 110 kV, 200 μA, 546 minutes). Morphological analyses of the proximal femurs is considered in later work (Chapter 5). To achieve a simulated fall-to-the-side loading configuration, proximal femurs were oriented with the shaft at 10° of inclination with respect to the horizontal surface and the femoral neck in 15° of internal rotation (Figure 3.3) (37,105). Custom Bondo (3M, Maplewood, MN, USA) molds were used to distribute the load applied to the greater trochanter during testing. Proximal femurs were subjected to a 100 N pre-load then loaded to failure at 100 mm/s through a metal acetabular cup that was best fit to the femoral head size. Stiffness (N/mm), yield load (N), PYD (mm), PYL (N), and maximum load (N) were calculated from load-displacement curves. A validation study, which involved indenting a hemispherical steel platen at 100 mm/sec into the Bondo pads, determined that the deflection attributable to the load cell and Bondo pads was 0.04 mm (0.02 mm – 0.1 mm), which accounted for 0.96% (0.56% - 2.2%) of the total displacement of the fractured femurs (Appendix A).

Statistical analysis

Whole bone strength for the diaphyses refers to the maximum bending moment, and whole bone strength for the proximal femur refers to maximum load. The relationship between stiffness and strength was assessed using least-squares linear regression. Sex-specific differences in the slopes and y-intercepts of the linear regressions were determined by ANCOVA for each bone site. The degree to which strength varied for a given stiffness was determined by calculating the 90% prediction bands and measuring strength at the average stiffness value (Minitab 16 e-academy, Inc., State College, Pennsylvania USA). Multiple linear regression analysis was conducted to test

if stiffness, age, PYD, and PYL were independent predictors of strength. Variance inflation factors (VIF) assessed if independent predictors of strength exhibited severe multicollinearity within the model. While there is no well-defined critical value to indicate severe multicollinearity, it is generally accepted that VIFs ranging from 5-10 signify a problem (106). Whole bone strength values were compared across sites using least-squares linear regression. A regression analysis was also conducted between the residuals of the stiffness-strength regressions at different bone sites to test whether a bone that tended to be weak (or strong) for a given stiffness at one site was also weak (or strong) for a given stiffness at other sites. Sex-specific differences in the slopes and y-intercepts of these regressions were determined by ANCOVA.

Results

Relationship between stiffness and strength

Bone strength correlated significantly with stiffness for males and females at all three bone sites (Figure 3.4). A comparison of the stiffness-strength regressions between males and females showed a significant difference in y-intercepts but not slopes for the radial diaphysis (Slope: $p=0.100$, Intercept: $p=0.015$) and the proximal femur (Slope: $p=0.613$, Intercept $p<0.001$), indicating that male radial diaphysis and proximal femurs were significantly stronger for a given stiffness compared to females. Using regression equations (Figure 3.4), it was determined that male bones were 158% (~1850 N) and 119% (~6 Nm) stronger than female bones for the proximal femur and radial diaphysis, respectively, when compared at the mean stiffness value for females. Although significant sex-specific differences were found for the radial diaphysis, there was limited overlap in stiffness values between male and female bones (15.12 – 21.02 Nm^2), suggesting that this particular sex-specific comparison was not appropriate. In contrast, the y-intercept and slope did not differ between males and females for the femoral diaphysis, even when the analysis was limited to the range of overlapping stiffness values (186.5 – 354.6 Nm^2). Thus, male and female bones showed a similar stiffness-strength relationship for the femoral diaphysis.

The degree to which strength varied for a given stiffness was determined for each bone site by calculating strength at the 90% prediction bands at the average stiffness (Table 3.3, Example

Calculation: Figure 3.5). The percent difference between the lowest and highest values of strength for males and females were 37% and 38% for the radial diaphysis, 42% and70% for the femoral diaphysis, and 97 % and 108% for the proximal femur, respectively. Qualitatively, high-resolution nanoCT images of proximal femurs obtained from young and elderly male and female donors showed large differences in bone morphology and internal microstructure (e.g., femoral neck length, proximal femur head size, amount and location of trabecular bone, etc.) for proximal femurs with similar stiffness but with a 50-100% difference in strength (Figure 3.6).

The stiffness-strength relationship: age, sex, and brittleness effects

The relative contributions of stiffness, age, PYD, and PYL to bone strength were determined by conducting a multivariate regression analysis (Table 3.4). Stiffness remained a significant predictor of strength in all regressions, even when age, PYL, and PYD were included in the model. Age was a significant ($p<0.05$) or borderline significant ($p<0.1$) independent predictor of strength at all bone sites for both sexes except for the female radial diaphysis. PYD was a significant ($p<0.05$) or borderline significant ($p<0.1$) independent predictor of strength for both sexes at both femur sites and female radial diaphyses. Post-yield load was a significant independent predictor for male femoral diaphyses and proximal femurs and female radial diaphyses. Adjusted R^2 values determining how well the variance of bone strength fit the tested multiple linear regression models ranged from 46.9% to 91.5% among the test groups. There was only one case in which VIF values suggested a potential multi-collinearity problem (Female radial diaphysis: PYL VIF = 5.647). However, this variable was not a significant independent predictor of bone strength and the impact on adjusted R-squared values was not further explored. Generalized linear models revealed that there were no significant 2-way, 3-way, and 4-way interactions among independent variables at any bone site or for either sex (data not shown). Thus, no strong interactions among the independent variables exist.

Comparing bone strength and residuals of the stiffness-strength relationship across anatomical sites

Strength correlated significantly across bone sites for male donors (Figure 3.7). These regressions were borderline significant for female donors. Regressions across bone sites were also conducted for using the stiffness-strength residuals (Figure 3.8). Significant positive correlations were observed only when comparing the male radial diaphysis and femoral diaphysis (R^2=0.18, p=0.04). Thus, perhaps due to differences in bone type (proportion of cortical and trabecular bone) and differences in normal in-vivo loading experienced across bone sites, donors that tended to have low (or high) strength for a given stiffness at one site did not tend to show low (or high) strength values at other sites.

Discussion

Cadaveric femoral diaphyses, radial diaphyses, and proximal femurs were loaded to failure to test how sex, age, and brittleness affected the relationship between stiffness and strength. For the diaphyses, strength correlated well with stiffness, as expected for a tubular structure, with strength values varying by as much as 37 – 70% for a given stiffness. In comparison to the diaphyses, the relationship between stiffness and strength was weaker for the proximal femur, as evidenced by a lower proportion of the variance in bone strength that can be explained by stiffness (i.e., R-squared values), with strength varying as much as 97 – 108% (i.e., 2-fold) for a given stiffness. PYD (brittleness) and age had independent effects on the stiffness-strength relationship for all three bone sites, indicating that older and more brittle bones had a lower strength than would be predicted from stiffness alone. Finally, the relationship between stiffness and strength varied with sex for the proximal femur with males showing 158% (~1850 N) greater strength than females at matched stiffness values. To our knowledge, this is the first study to report significant sex, age, and brittleness effects on the stiffness-strength relationship of whole bones. Our findings suggest that surrogate indices of strength that rely on stiffness may be improved by adjusting for sex, age, and brittleness effects.

The sex-specific nature of the stiffness-strength relationship was observed for the proximal femur where male bones were approximately 158% stronger than female bones at matched stiffness

values, respectively. Prior work reported sex-specific differences in bone stiffness and strength individually (33,107,108), and it has been estimated that male bones are stronger relative to body size compared to female bones (33,109,110). However, no studies have reported the sex-specific nature of the stiffness-strength relationship observed herein. Although it was not our goal to identify a mechanism that would explain the sex-specific differences in the stiffness-strength relationship, our results suggest that the physical bone traits that define bone stiffness may differ from those that define strength and that structure-function associations differ between men and women. Changes in porosity and external morphology may partially explain the decoupling of stiffness and strength. Likewise, sex-specific changes in collagen cross-links may contribute to the decoupling of stiffness and strength (111,112). Variation in proximal femur strength between and within sexes may also be attributed in part to trabecular microstructural redundancy. Less microstructurally redundant bones require a smaller proportion of bone to fail, are less able to effectively transmit load, and thus are weaker (21,113). If females are less microstructurally redundant than males, then this may explain why strength but not stiffness declines across the age-range examined. With increasing age, women have an increased propensity to fracture compared to men (114) and prior work has identified bone traits (external size, geometry, BMD, etc.) that may contribute to the increased strength indices of men compared to women (115). Our study suggested that the cumulative effect of these bone traits resulted in a stiffness-strength relationship that varied with sex and age for proximal femur and the male diaphysis.

Age, PYD, and PYL, in addition to stiffness, were significant independent predictors of strength for the proximal femur, and these variables were mostly significant at the femoral and radial diaphysis for both sexes. A large proportion of the variance in bone strength (Adjusted R^2 = 46.9% – 91.5%) that can be explained by multiple linear regressions, including measures of elastic and plastic mechanical behavior and no measures of bone morphology or tissue-level mechanical properties. This outcome indicated that older, more brittle bones tend to sustain a lower post-yield load, and thus a lower strength relative to stiffness. Previous studies examining age-related changes in bone mechanics were typically conducted at the tissue level (116–118). Few cadaveric studies have reported age-changes in whole bone mechanical properties. Our results are consistent with prior work reporting that post-yield properties (e.g., strength, fracture toughness, post-yield strain) degrade with age (37,101,102). This decrease in post-yield

displacement, (i.e. increase in brittleness) reflects changes in crack tolerance of cortical and trabecular bone which may lead to premature failure and thus a proportionally lower strength of older bones (102,116). Although most studies agree that bones tend to exhibit less PYD with age, stiffness has been shown to decrease (105,118), not change (37,101,119), or even increase with age (117). Discrepancies in how stiffness changes with age among studies likely arise from differences in scale (whole bone level versus tissue-level), testing mode (compression, tension, torsion, 4-point bending), anatomic site, and/or tissue handling. Our results are in line with findings by Nawathe et al., who showed a ~ 40% decrease in simulated whole bone strength for the proximal femur loaded to failure in fall-to-side fracture when tissue level post-yield behavior of bone was computationally changed from a fully ductile to brittle behavior (2015). Our study utilized the natural variation in whole bone post-yield displacement across the adult age range to study how brittleness affected bone strength. The current study is unique because we tested a large number of samples to assess bone mechanical behavior at three different sites using consistent tissue handling methods. The clinical implication of finding that the age-related increase in brittleness may contribute to the age-related decrease in strength is that post-yield properties depend on material behavior, which is difficult to measure non-invasively. This outcome would suggest that the degree to which morphological traits can be used to predict strength becomes progressively limited with aging. Thus, determining how the relative contributions of material and morphological traits to whole bone strength change with age, site, and sex may benefit efforts to improve strength estimates and fracture risk.

Proximal femur strength varied as much as 97% and 108% at the mean stiffness for males and females, respectively, suggesting that bone strength may not be accurately predicted based solely on information arising within the linear-elastic range of loading (i.e., stiffness). Whole bone strength correlated across bone sites on an absolute basis but not relative to stiffness (Figures 5 and 6), consistent with prior work (120–122). The variance in bone strength being poorly described by bone strength at different bone sites for the female bones may be partially attributed to the lower number of paired samples in this cohort. Intra-skeletal elements (cortical TMD and cortical area) are less highly correlated in females compared to males at the radial and femoral diaphysis (122). If bone material properties are less uniform across the female skeleton, differences in mechanical properties may be further accentuated when comparing across

diaphyseal and the cortical-cancellous proximal femur. Intra-skeletal comparisons were studied to begin understanding whether peripheral sites can predict strength changes in the proximal femur. Herein, we tested whether strength correlated across sites (Figure 5), leaving mechanistic details to follow up research. We believe that strength correlated across sites for three primary reasons: (1) body size effects (i.e., bigger people tend to have bigger, stronger bones), (2) bone morphology (i.e., bone robustness correlates across skeletal sites (122)), and (3) age-related changes in bone structure and material properties. For the latter factor, if bone structure and material properties change similarly across sites, then we would expect that bone strength would also show similar age-related declines across sites, and thereby contributing to the correlations in strength across sites. These factors remain to be teased out in future work to better understand whether peripheral bones provide a meaningful site to monitor the age-declines in the strength of the fracture-prone proximal femur.

Similar correlations of whole bone strength across anatomical sites have been observed in formalin-fixed cadavers (123); these outcomes are limited because formalin affects been shown to significantly affect Young's modulus, yield strain, and ultimate strain of bone mechanical properties loaded in compression (87,124). Although we did not investigate the biomechanical mechanisms that would explain the correlation of strength across sites, similarities in stiffness measures at central and peripheral cortico-cancellous sites may be due to similarities in areal BMD, volumetric BMD, geometry, and microstructure (121). The lack of correlation of the residuals from the stiffness-strength regressions across bone sites could be attributed to the fairly narrow range of residual values for the radial and femoral diaphyses. The lack of strong correlations across bone sites between the stiffness-strength relationship suggests that age-related changes in strength and stiffness may arise through different rates of structural and material changes. Future work needs to tease out the material and geometrical contributions to whole bone stiffness and strength for both sexes to better explain the outcomes observed in this study. Clinically, this outcome would mean that site-specific strength estimates may be needed to predict fracture risk for women and that the sum of factors that affect the stiffness-strength relationship at one site may not be observed at another, despite the similarity in strength on an absolute value.

Directly measuring whole bone mechanical properties for a large cohort of cadaveric specimens is a strength of this study. However, some limitations need to be addressed. Because the cadaveric bones had no known musculoskeletal disease or injury, our donors may represent a stronger subgroup within the elderly population and thus may underestimate the declines in bone stiffness and strength with aging. Bodyweight and height were not available for all donors, which limited our ability to adjust for body size effects and investigate temporal trends. The proximal femur testing protocol (39,94,105) was limited to a constant loading rate and direction (i.e., sideways fall). Although limitations such as loading condition, orientation, and rate exist for all ex vivo mechanical tests, the outcomes should provide a reasonable approximation of the in situ whole bone strength. Proximal femurs were loaded to failure at a rate that was three orders of magnitude greater than the diaphyseal sites. It is unclear how the stiffness-strength relationships would change with different loading modes. However, bones become more brittle at higher loading rates, which may partially explain the greater variation among the proximal femur mechanical properties compared to the diaphysis (88). The in-vivo initial impact velocity from a fall to the side may be up to 35 times higher than the load in which bone was fractured in this study (125). It is well established that bone sustains higher forces at higher loading rate (105). However, McElhaney observed only a 12% change in compressive strength when there was a 300-fold increase in load, suggesting that the mechanical test results observed here may not significantly differ from what would be observed in a clinical fracture (126).. Finally, relationships among material and geometrical properties were not explored, but are needed to provide insight into the decoupling between whole bone stiffness and strength.

In conclusion, whole bone strength was impacted by stiffness and age-related declines in ductility and other age-related factors. Thus, bones appear to become weaker relative to stiffness with aging. Finally, the relationship between stiffness and strength varied between sexes for the proximal femur where males were twice as strong as stiffness-matched females. Both PYD and age affected the stiffness-strength relationship, to varying degrees, indicating that including these variables in addition to stiffness may improve estimates of whole bone strength. Future work will determine why the relationship between stiffness and strength changes with aging.

Acknowledgments

The research reported in this chapter was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health **(**KJJ: AR065424, AR069620; SHS: DE007057, David H. Kohn: T32DE007057, DE00705; Todd L. Bredbenner: AR06424**).** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Figures

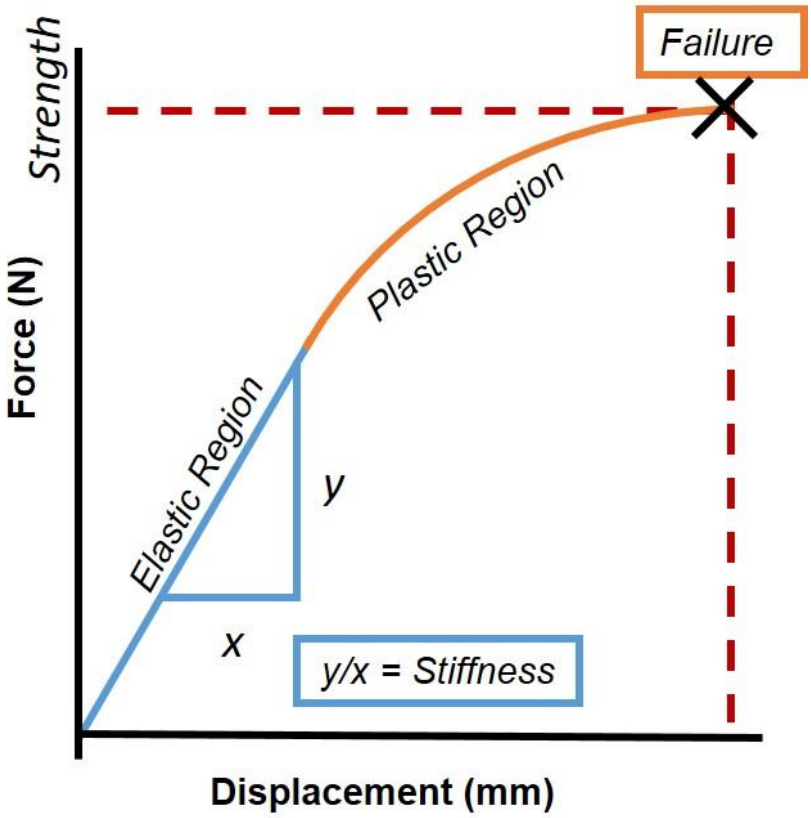


Figure 3.1 Example load-deformation curve of a material that exhibits both elastic and plastic behavior.

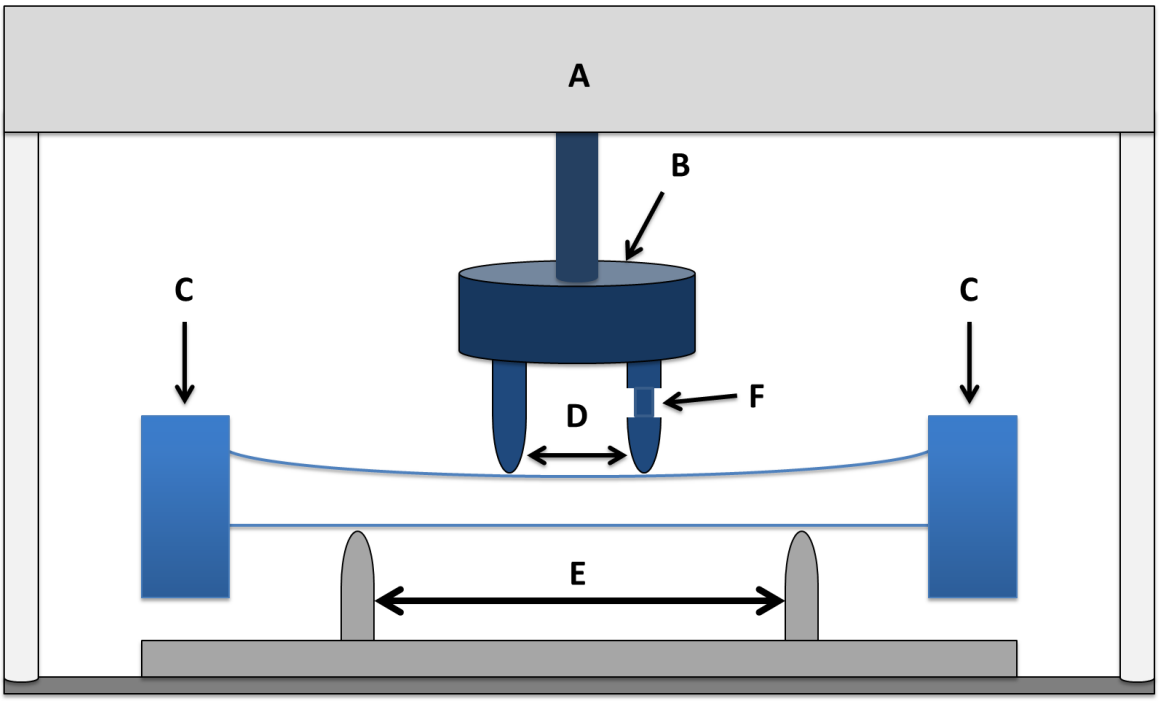


Figure 3.2 Schematic of the 4-point bending testing fixture used to assess whole bone mechanical properties of the femoral and radial diaphysis. Elements of the system include the (A) Instron 8511 material test frame, (B) 10 kN load cell, (C) square molds of acrylic resin used to prevent sample rotation during loading, (D) upper loading points at 33% lower span length and centered around the lower span length, (E) the lower span length at 25% and 75% total bone length, and (F) an adjustable upper loading point used to ensure contact at all 4 loading points along the non-uniform bone geometry. The (D) upper and (E) lower loading points are dictated by the physical length of every femur and, in result, were adjusted for every diaphysis loaded to failure.

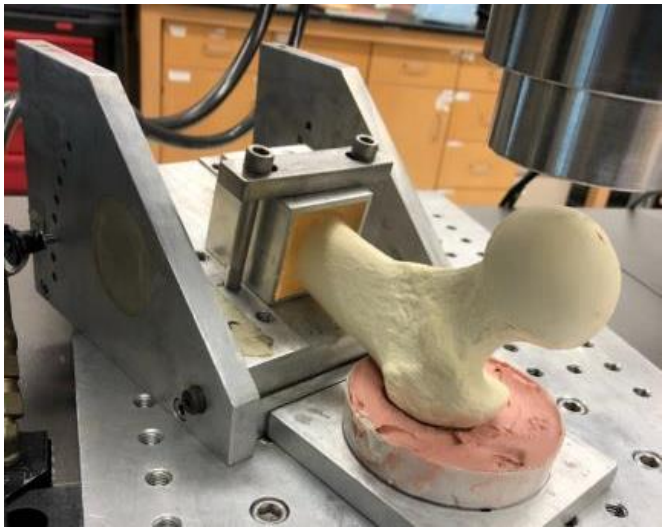
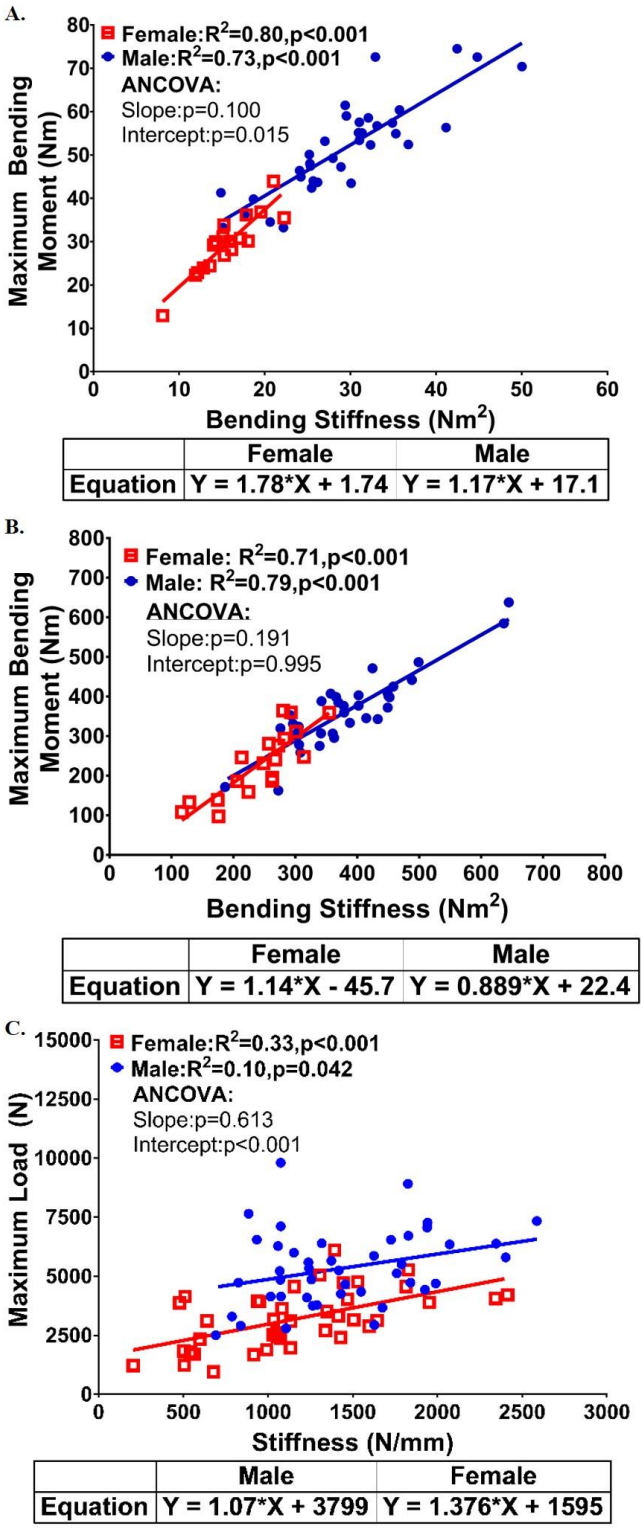


Figure 3.3 Example set-up of a proximal femur loaded in fall-to-side testing. The proximal femurs were oriented with the shaft at 10 ° of inclination with respect to the horizontal surface and the neck with 15° of internal rotation.



	Female	Male
Equation	Y = 1.78*X + 1.74	Y = 1.17*X + 17.1

	Female	Male
Equation	Y = 1.14*X - 45.7	Y = 0.889*X + 22.4

	Male	Female
Equation	Y = 1.07*X + 3799	Y = 1.376*X + 1595

Figure 3.4 Linear regressions between whole bone stiffness and strength for the (A) radial diaphysis (B) femoral diaphysis, and (C) proximal femur.

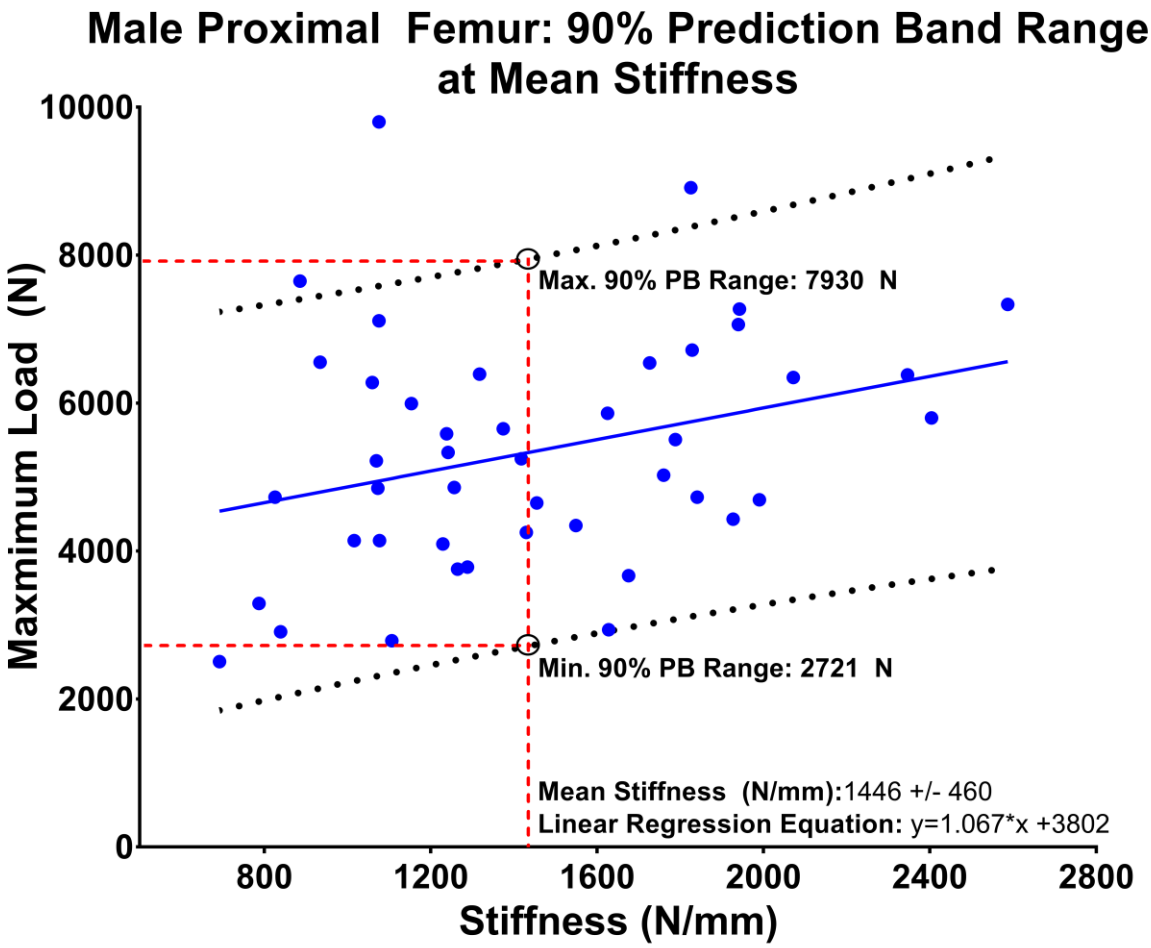


Figure 3.5 Example (male proximal femurs) showing how the range in whole bone strength for a given stiffness was calculated from the 90% prediction bands (PB) and expressed relative to the mean stiffness.

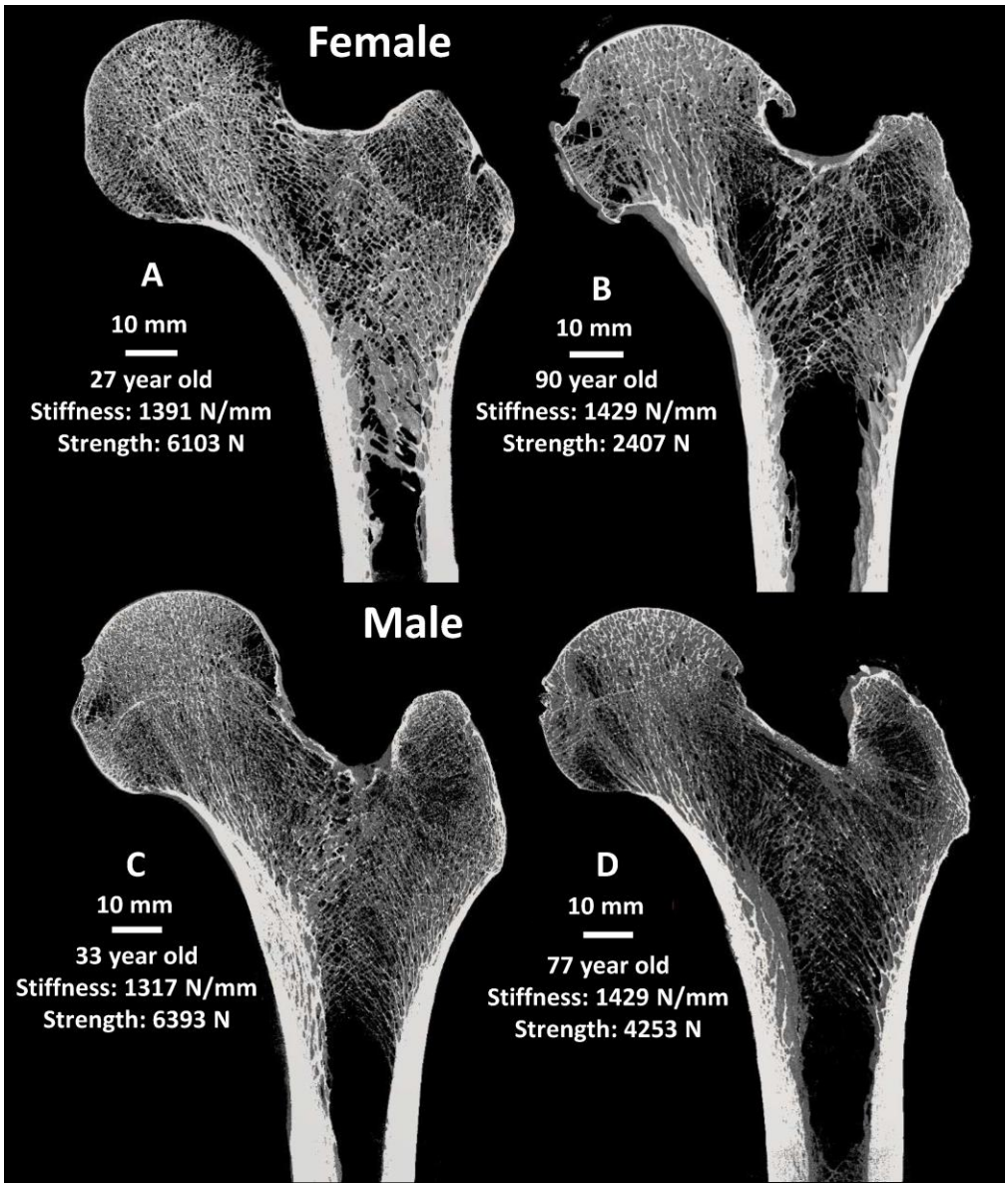
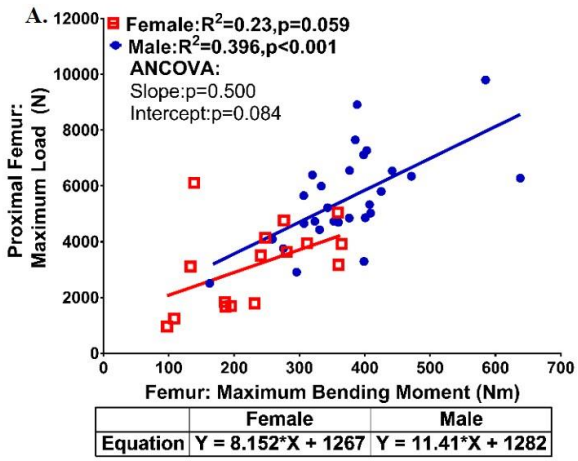
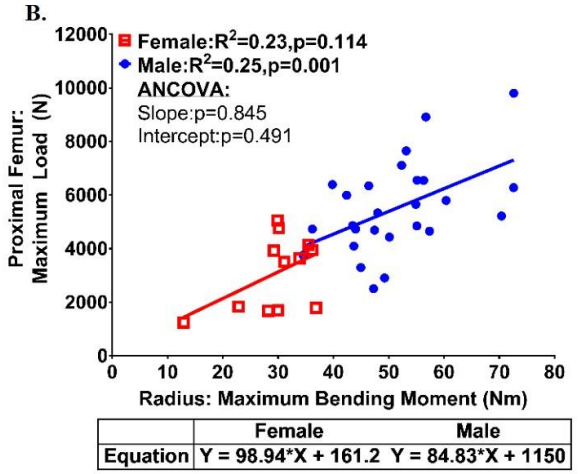


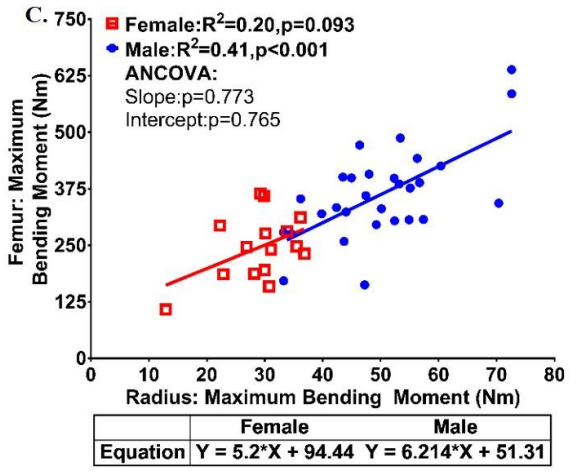
Figure 3.6 NanoCT images of proximal femurs showing similar whole bone stiffness by sex but different strength for a (A) 27-year-old female [Stiffness: 1391 N/mm, Strength: 6103 N], (B) 90-year-old female [Stiffness: 1429 N/mm, Strength: 2407 N], (C) 33-year-old [Stiffness: 129 N/mm, Strength: 6393 N], and (D) 77-year-old male [Stiffness: 1429 N/mm, Strength: 4253 N].



	Female	Male
Equation	Y = 8.152*X + 1267	Y = 11.41*X + 1282



	Female	Male
Equation	Y = 98.94*X + 161.2	Y = 84.83*X + 1150



	Female	Male
Equation	Y = 5.2*X + 94.44	Y = 6.214*X + 51.31

Figure 3.7 Comparison of whole bone strength between the (A) femoral diaphysis and the proximal femur, (B) radial diaphysis and the proximal femur, (C) and the radial diaphysis and the femoral diaphysis.

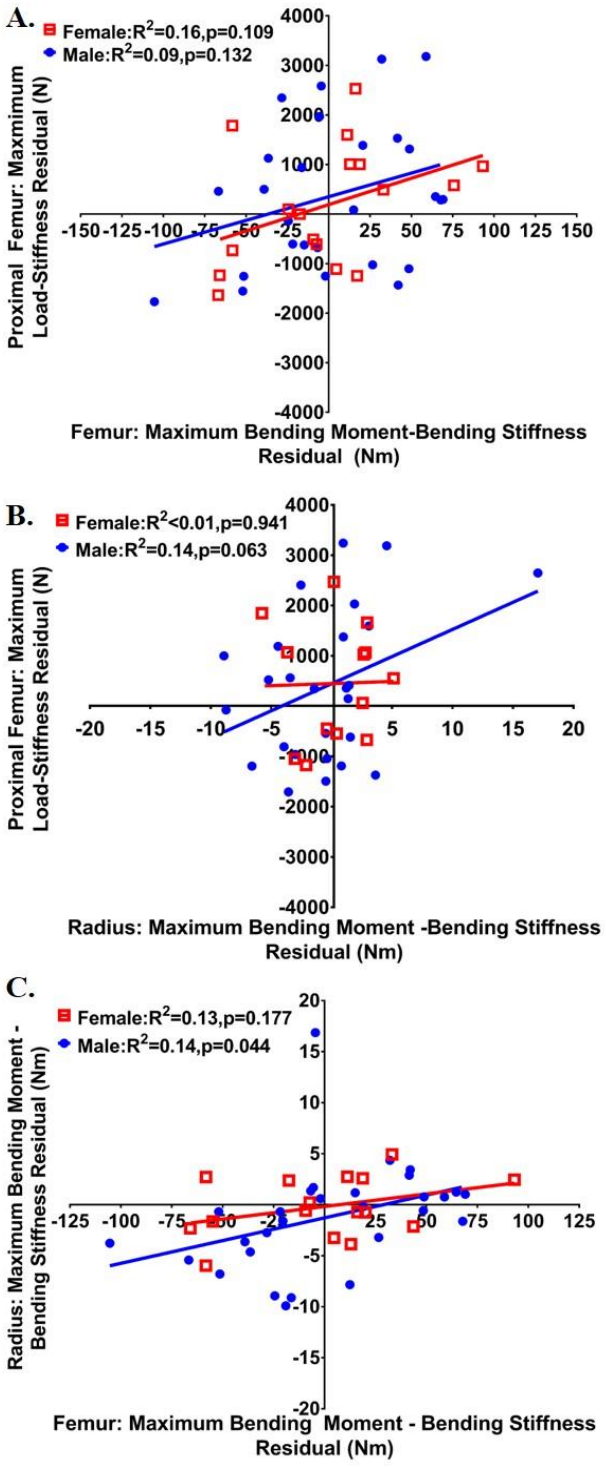


Figure 3.8 Comparison of residuals calculated from the stiffness-strength regressions between the (A) femoral diaphysis and the proximal femur, (B) radial diaphysis and the proximal femur, and (C) the radial diaphysis and the femoral diaphysis.

Tables

Table 3.1 Summary of various studies demonstrating a linear relationship between bone stiffness and strength.

Bone Specimen	**Sample Size**	**Linear Stiffness-Strength Relationship Slope[Intercept] (R^2)**	**Mechanical Test**	**Citation**
Cadaveric Ulnar	n=45	NA[NA] (0.92)	Three point bending with the posterior surface was in tension.	(86)
Bovine (bilateral tibia and femur) and Cadaveric Vertebral Cancellous	Cadaver: n=5 (all F) Bovine: N/A	Pooled analysis 0.0038[0] (0.91)	Bovine cortical bone: Tension; Human Cancellous Bone: Compression	(92)
Cadaveric Vertebral Cancellous	n=28 (F[16]; M[12])	30.93[1.758] (0.88)	Uniaxial compression in the infero—superior direction	(127)
Bovine proximal tibia	n=48	Compression: 9.4e-3[0] (0.78) Tension: 5.1e-3[0] (0.91)	Uniaxial Compression (n = 30) and tension (n = 29)	(128)

Table 3.2 Distribution of bone samples relative to age, sex, and site.

Sex	**Female**			**Male**		
Bones	**Number of Samples**	**Mean Age +/-SD (years)**	**Age Range (years)**	**Number of Samples**	**Mean Age +/- SD (years)**	**Age Range (years)**
Radius	19	59 +/- 22	23 - 95	36	54 +/- 23	18 - 89
Femur	19	57 +/- 21	24 - 95	34	59 +/- 20	18 - 89
Proximal Femur	40	63 +/- 21	24 - 95	44	58 +/- 19	18 - 89

Table 3.3 Comparison of the maximum load for the proximal femur, femoral diaphysis, and radial diaphysis.

Bone Site	Sex	Mean Stiffness	Strength at Mean Stiffness	Minimum 90% PB	Maximum 90% PB	90% Prediction Band Strength Range	% Range Compare to Mean Strength
Proximal Femur	F	1163 +/- 509	3195	1451	4922	3471	**108%**
	M	1446 +/- 460	5344	2721	7930	5209	**97%**
Femur	F	244 +/- 62	232	149	313	164	**70%**
	M	386 +/- 95	366	289	443	154	**42%**
Radius	F	16 +/-3	29	24	35	11	**38%**
	M	29 +/- 3	52	42	61	19	**37%**

* Stiffness values are given in Nm^2 for the femoral and radial diaphyses and N/mm for the proximal femur. Strength values are given in Nm for the diaphyses and in N for the proximal femur.

Table 3.4 Multiple linear regression analysis between whole bone strength and stiffness [Diaphysis: Nm2, Proximal Femur: N/mm], age [years], PYD [Diaphysis: 1/m, Proximal Femur: mm], and post-yield load (PYL) [Diaphysis: Nm, Proximal Femur: N] (bold font, p<0.05; italic font, p<0.10).

Site	Sex	Predictive Variable	Slope of Coefficient (B)	Standardized Slope of Coefficient (Normalized B)	SE	t	P	VIF
Femur Diaphysis	Male	Constant	-20.94		49.23	-0.43	0.674	
		Stiffness	1.03	1.03	0.12	8.37	**<0.001**	3.788
		Age	-0.78	-0.17	0.30	-2.60	**0.014**	1.087
		PYD	57.58	0.25	17.74	3.25	**0.003**	1.501
		Post-Yield Load	-0.25	-0.23	0.15	-1.73	*0.094*	*4.307*
	Adjusted R^2: 86.7%							
	Female	Constant	-10.30		43.61	-0.24	0.817	
		Stiffness	0.66	0.48	0.20	3.29	**0.005**	2.991
		Age	-0.63	-0.17	0.36	-1.78	*0.097*	1.220
		PYD	37.01	0.26	20.36	1.82	*0.091*	*2.947*
		Post-Yield Load	0.54	0.32	0.40	1.58	0.137	5.647
	Adjusted R^2: 87.3%							
Radial Diaphysis	Male	**Constant**	15.46		6.11	2.53	**0.017**	
		Stiffness	1.26	0.92	0.18	7.02	**<0.001**	2.767
		Age	-0.09	-0.19	0.04	-1.99	*0.055*	*1.428*
		PYD	0.93	0.15	0.65	1.42	0.167	1.729
		Post-Yield Load	-0.015	-0.31	0.13	-0.12	0.909	2.669
	Adjusted R^2: 78.40%							
	Female	Constant	3.30		3.45	0.96	0.356	
		Stiffness	1.13	0.57	0.20	5.66	**<0.001**	2.125
		Age	-0.024	-0.80	0.02	-1.04	0.318	1.251
		PYD	0.79	0.19	0.31	2.52	**0.024**	1.175
		Post-Yield Load	0.42	0.35	0.12	3.49	**0.004**	2.154
	Adjusted R^2: 91.50%							
Proximal Femur	Male	**Constant**	5371		1077	4.99	**<0.001**	
		Stiffness	1.44	0.41	0.4098	3.5	**0.001**	1.088
		Age	-32.30	-0.37	10.87	-2.97	**0.005**	1.244
		PYD	-229.14	-0.32	95.66	-2.4	**0.022**	1.411
		Post-Yield Load	0.79	0.56	0.1869	4.25	**<0.001**	1.378
	Adjusted R^2: 46.9%							
	Female	**Constant**	3775.2		568.9	6.64	**<0.001**	
		Stiffness	1.29	0.54	0.26	4.97	**<0.001**	1.070
		Age	-28.67	-0.49	6.16	-4.65	**<0.001**	1.024
		PYD	-82.36	-0.26	35.23	-2.34	**0.025**	1.116
		Post-Yield Load	0.32	0.15	0.23	1.4	0.171	1.034
	Adjusted R^2: 57.5%							

Chapter 4 Machine Learning Algorithms Improve Bone Segmentation in nanoComputed Tomography Images

Introduction

Background

Micro-computed tomography (Micro-CT) imaging is the "gold standard" method to assess three-dimensional (3D) bone morphology and microstructure when studying bone disease and treatment (32,129). With the rise of high resolution peripheral quantitative computed tomography (HR-pQCT) and additional imaging modalities capable of resolving bone microstructure, robust tools must exist to evaluate bone architecture (130,131). To evaluate bone architecture, it is required that bone and the particular regions of interest (i.e., cortical, trabecular, metaphyseal, or diaphyseal volumes) are accurately identified via segmentation. Segmentation, sometimes based on intensity thresholding, locates voxels of the interior of an object or an object border for quantitative analysis. Two particular types of segmentation, bone/background and cortical/trabecular, are critical and challenging steps in bone analysis.

Segmenting Bone from Background

To quantify bone microstructure, one must choose a method (often a threshold value) to segment bone from non-bone (i.e., background) voxels (32). Traditional morphometric outcomes (e.g., bone volume fraction [BVF], trabecular thickness [Tb.Th] and spacing [Tb.Sp]), are sensitive to variations in global threshold values (132–135). For example, a variation in a grey-level threshold of 7.1% can constitute up to a 35 μm difference in Tb.Th, 306 μm in Tb.Sp, and 0.07 in BVF (i.e., percent bias range of 13% Tb.Th, 29.2% Tb.Sp, amd 22.0% BVF) (135). Consequently, the threshold value chosen has a direct and profound effect on study reproducibility and the biological interpretation of the data. Despite this sensitivity, there is no standard method to segment bone from non-bone voxels (32,134).

Common Thresholding Methods used in Bone Research

Both global and local thresholds are used to identify bone in grey-scale CT-scans.

- A global threshold is frequently used and requires that a single greyscale intensity value be chosen to separate bone from background; working under the assumption that there is a bimodal histogram (Figure 4.1) (136,137). The threshold value can be a fixed operator-selected grey-scale CT value, percentage of the grey-level spectrum, or set using an automatic approach such as the Otsu method (135,138,139). However, bones are challenging to simply threshold because there may be significant overlap in grey-scale values between bone and the surrounding tissue, bone structures may not always have a uniform density, and system scanning artifacts may be present. Thus, while frequently used, there are a variety of circumstances in which global thresholds perform poorly in bone research.
- Local thresholding (e.g., region growing, mean weighted average, and edge-based detection) allows for a threshold value to change dynamically, determining a threshold for each voxel separately by considering only nearby voxels of an arbitrary radius. These more sophisticated approaches solve some of the common global threshold challenges but come at the cost that they are often not publicly/easily available to use without advanced programming knowledge, are computationally intensive, and/or require setting several experiment-specific parameters (64,140,141).

Bone research would benefit from new, easily accessible, automatic methods to segment bone from background. A new method that could segment bone efficiently, consistently, and to a wide variety of complex bone structures would be of great benefit to the field.

Segmenting Cortical from Trabecular Bone.

To quantify cortical and trabecular bone microstructure in any CT scan, it is required that each region is digitally identified. The "gold standard" approach for cortical/trabecular segmentation is a semi-automated slice-by-slice hand contouring approach (142–146). While a manual approach is often used, it is sensitive to operator error and as a result, one operator must perform all contouring in a study (32,144,147). Unfortunately, manual contouring is tedious and arduous. For example, manual cortical/trabecular segmentation of the proximal femur femoral neck can surpass four hours per bone. Snake algorithms, which "snap" the contours

into place, and interpolation functions can help quicken the process (148) but perform poorly in regions with high cortical porosity, in subjects with low BMD compared to the background, or when large changes in bone structure exist. In conclusion, the "gold standard" procedure for cortical/trabecular separation is inefficient, non-trivial, and prone to operator error.

Automatic Methods to Segment Cortical from Trabecular Bone

Many semi-automated approaches have been proposed to help standardize bone segmentation (142,144,145). However, they all face certain limitations in that they are complicated (e.g. based on region growing, energy minimizations spline curves, and deformable models), depend on specimen orientation, are affected by the scan-dependent signal to noise ratio, and/or require optimizing algorithm-specific parameters on a trial and error basis (145,149). The most popular automated technique proposed by Buie et al. is applicable to different species (e.g., human radii and mice tibiae), is easy to implement, and requires minimal input (144). However, Buie's method cannot handle cortical surface gaps such as Volkmann canals (requires digitally filling surface gaps) and misidentifies cortical as trabecular bone when the cortical region is highly porous (144). To further complicate matters, some have even proposed segmenting a third region, the transitional zone, which is the "trabecularized" inner cortex (143,150). While the interesting transitional zone is a site of rigorous intracortical remodeling, introducing a third region further complicates the segmentation process.

Given all existing approaches to segment cortical from trabecular bone, it is surprising that there has been no convergence on a single segmentation approach in orthopedic research. Instead, the method of choice varies from study to study, likely impacting reported results, and contributing to the 'crisis' of reproducibility that is observed in research today. More nuanced, standardized approaches that are easily applicable and reproducible and require minimal user input are needed for the segmentation of bone in micro-CT scans.

Proximal Femur Femoral Neck Nano-CT Scans: A Challenging Segmentation Problem

We have a collection of high resolution (27 μm voxel size) scans of cadaveric proximal femurs and are interested in quantifying femoral neck microarchitecture. The femoral neck is a

complicated structure that varies in cortical (e.g., thickness, porosity, and vBMD) and trabecular (e.g., anatomical location and density distribution, vBMD, and thickness) architecture within and across structures (Figure 4.2). Moreover, based on visual observation, scan quality substantially varied in our dataset. Scanning parameters in any protocol (e.g. resolution, voxel size, and step position) should be dictated by the physical size and density of the sample (151). The proximal femurs used in this dissertation substantially varied in size, shape, and density because donors were collected from diverse demographic backgrounds (Chapter 3, pg. 45). Thus, while the scanning protocol was optimized to perform well on most femurs, scans of the smallest and/or least dense and largest and/or densest proximal femurs were sub-optimal. As a result, noise (graininess) and the presence or absence of cone beam artifacts (i.e., artifacts at the edges of materials near edges of the detector due magnification to resolution ratio) varied from scan-to-scan. Spatial density variation in the morphologically complex femur also resulted in the presence of beam hardening (insufficient penetration of the sample) and scattering on the rare occasion in which metal particles were present in scans (151). The presence and extremity of scanning artifacts varied from scan-to-scan so existing artifact minimizing post-processing algorithms were not included in our workflow (151).

Deep Learning and neural networks: A New Tool for Segmentation

In recent years, there have been unparalleled advances in the application of deep learning, a type of machine learning, for classification, object detection, segmentation, and registration of medical image data (35). Machine learning is the science of building algorithms to solve practical problems (outcome) based on a quantifiable phenomenon (dataset) (152,153). Deep learning is a mathematical algorithm where model parameters are not modeled directly from the features in the training examples (e.g., linear regression analysis), but from the outputs of the proceeding layers, commonly referred to as hidden layers (49,152,154). To better explain this, the most well-known of these models, namely, convolutional neural networks (CNNs), will be described in more detail below.

Convolutional Neural Networks (CNNs)

Neural networks are mathematical models that are loosely inspired by biological neural networks present in the brain. A neural network consists of a basic unit called a node (inspired by neurons) that transmits information to other nodes via connections (inspired by dendrites and synapses). Many nodes are connected and arranged in hidden layers. The network is programmed (i.e. trained) to take greyscale input data (e.g. images) to automatically generate an output (e.g. region of interest identification).

CNNs are neural networks that rely on convolutional layers (i.e., mathematical operations that act as a filter in hidden layers) whose values are adjusted using a backpropagation algorithm during the training of the network to optimize the classification (i.e., segmentation) output (please see Chapter Four Methods for more detail) (153,154). Convolutional steps allow for multi-class predictions of multiple pixels with fewer computations (i.e., improving the segmentation model efficiency and decreasing computational requirements) (154). The weights/values of the filters in the hidden layers are the key components of the network underlying the architecture that are changed during training to optimize segmentation/classification performance (153,154).

However, in CNNs, there are a series of operations (i.e., convolutional and pooling layers) used in the underlying architecture that result in a smaller segmentation output than input image (155). To overcome this limitation, some researchers have relied on deconvolutions operations to up-sample the reduced size feature maps (i.e., backward-strides convolution) to restore the initial image size (155). This type of network is a CNN without fully connected layers and is referred to as a fully convolutional neural network (FCNN) (154,155). While the details and underlying architecture of complex deep neural networks, CNNs, and FCNNs are beyond the scope of this study, an FCNN with U-net architecture will be used for this study and is briefly described below (153,156).

The U-net is one of the most well-known FCNNs for medical image segmentation (153,154,157). The U-net relies on the deconvolutions used in FCNNs and applies skip connections between contracting (convolution) and expanding (deconvolution) paths so that

global features can propagate to higher resolution layers in the hidden layers (158). This architecture has been shown to repeatedly shown to outperform other biomedical image segmentation method with few training images (ISBI cell tracking challenge winner of 2015) (158).

CNNs in the Musculoskeletal Field

The use of CNNs has proven to be the most successful type of artificial neural network for image analysis problems, effectively permeating the entire medical imaging community and repeatedly outperforming other approaches (35). In the musculoskeletal field, promising results have been shown in automatic vertebral identification and segmentation of the vertebrae, whole body, and for the proximal femur in both magnetic resonance (MR) and computed tomography (CT) scans (159–163). One limitation of CNNs is that they historically required a large amount of training data - a requirement that may be difficult to meet in research where data-sets are limited in size and manual annotation is time-consuming. However, one FCNN with U-net architecture (ISBI 2015 Cell Tracking Challenge winner) overcomes such a limitation and performs well with minimal training data (158). Using a FCNN with U-net architecture provides promise for automatic segmentation tasks of high-resolution scans of the cadaveric proximal femur. The remainder of this chapter will use FCNN and FCNN with U-net architecture interchangeably.

Objectives

Although most bones can be visually identified in CT scans without difficulties, developing a segmentation method that is both precise and automated is still a challenge. Thus, our first objective is to create a FCNN with U-net architecture to threshold bone from background in our femoral neck nano-CT scans. We hypothesize that our FCNN will outperform the commonly used automatic Otsu threshold method. The trained FCNN will be compared to ground-truth (i.e., manually segmented bone volumes) and Otsu segmentation of extracted coronal cross-sections of the femoral neck. Bone architectural results will also be quantified to measure the impact of the segmentation technique on architectural results. Our second objective is to create a FCNN that can automatically delineate cortical from trabecular bone. We plan to create a

segmentation method that is accurate, and completely reproducible, and performs well despite biological and scanning artifact variability. If successful, introducing such a technique would save time and effort at the hand of researchers while also creating a method that allows for improved quality and reproducibility.

Methods

The sample population, preparation, and scanning protocol were described in detail previously (Chapter 3, pg. 45). Briefly, male and female (Female[n=40],24-95yrs, Male[n=40]:18-89 yrs) proximal femurs were cut 16.5 cm from the superior aspect of the femoral head, the shaft was embedded in acrylic resin using a custom alignment fixture, and each femur was imaged using a nano-computed tomography system (nanotom-s, phoenix|x-ray, GE Measurement & Control; Wunstorf, Germany) (27 μm voxel size, 110 kV, 200 μA, 546 minutes).

Image Processing

Image volumes were reconstructed using datos|x reconstruction software as 32-bit float volume files (phoenix|x-ray, GE Sensing and Inspection Technologies, GmbH, Wunstorf, Germany). Grey values from each reconstructed image were converted to Hounsfield units using the calibration phantom as described previously (54). Due to computer memory limitations, scans were down-sampled to 16-bit signed integer volumes and then a 3-D median filter (radius = 3) in Matlab was applied (The MathWorks Inc., Natick, Massachusetts, United States)**.** The femoral neck region of interest was extracted using two parallel proximal and distal planes. Briefly, the proximal femur was oriented so that the femoral shaft was parallel to the z-axis, and the line between the most medial aspect of the femoral head and lateral aspect of the proximal femur (i.e., the greater trochanter) were parallel with the transverse plane (x-y plane) in Dragonfly). Next, the inferior aspect of the lesser trochanter and the base of the greater trochanter were identified in the y-z plane (i.e., used as anatomical landmarks to generate the distal plane). The proximal femur was rotated about the point of the lesser trochanter in the y-z plane until the line between the base of the greater trochanter and the inferior aspect of the lesser trochanter were parallel to the z-axis. The y-z plane in Dragonfly that intersects through these two points and orthogonal to this view is considered the first plane of extraction of the femoral neck. A parallel plane that contains the point of intersection between the sphere fit to

femur head and the superior aspect of the femoral neck were used to identify the proximal extraction plane (Dragonfly 4.0, Object Research Systems; Montreal, QC, Canada). This resulted in a 16.9 mm slab on average from the femoral neck region being extracted for analysis. The extraction procedure was repeated three separate times on proximal femur scans (n=3) and percent differences in bone volume (BV), total volume (TV), and BVF were 1.38%, 1.86%, and 1.05% respectively, indicating high reproducibility.

Creating a FCNN:

All machine learning models (FCNN with U-net architecture) were trained within the framework of Dragonfly software 4.0 (ORS, Montreal, Canada) on an HP Z820 Workstation with the following specifications:

- Windows 7 Ultimate, 64-bit operating system
- 192 GB of RAM
- NVIDIA GeForce RTX 2080 Ti graphics card
- Intel XEON CPU E5-2650 version 2 processor

A FCNN with U-net architecture is trained via supervised learning and thus requires a ground-truth (GT) dataset: a set of input objects (cross-sections) and the desired output (manually segmented ROI). The ground truth dataset is typically divided into three subsections: training, validation, and test set (153,154).

- A training set is the set of ground truth data that is used to build the model. While there is no standard requirement, the general rule of thumb is that ~70% of GT images are used for training the model.
- The validation set (i.e. a hold-out set) is used to assess the model weights and determine the best hyper-parameters during training (described below). The validation set is generally much smaller than the training set (~15% of GT images).
- The test set (~15% of GT images) is used to assess the quality of the final model.

Once the training, validation, and test set are defined, the hyper-parameters, or the variables which determined the network structure and/or how the network is trained, need to be selected (described in Table 4.1) (153,154,157). Hyper-parameters used in the model are set before starting the training process. The metrics were selected by hand via systematic testing of model

output against the test set. Please refer to the original paper by Ronneberger et al. for additional information regarding the architecture (158).

Model Assessment via the use of Similarity metrics of overlap

To assess the model performance against the ground-truth test set, the following metrics were quantified (164):

- True positive (TP): The number of voxels in which the segmentation method correctly labeled voxels (e.g., bone) in our dataset.
- True Negative (TN): The number of voxels in which the segmentation method correctly predicted unlabeled voxels (e.g. background).
- False Positive (FP): The number of voxels in which the segmentation method incorrectly predicts labeled voxels in the dataset.
- False Negative (FN): The number of voxels in which a voxel should be labeled, and segmentation method incorrectly predicts an unlabeled voxel.
- Dice Coefficient (DICE): DICE, frequently called the overlap index, is the most widely used metric for testing the quality of a segmentation method. This metric relies on four separate cardinalities (TP, TN, FP, and FN).

$$DICE = \frac{2 * TP}{2 * TP + FP + FN}$$

- Volumetric Similarity (VS): A measure that considers the volumes of segments to indicate similarity.

$$VS = 1 - \frac{|FN - FP|}{2 * TP + FP + FN}$$

- Kappa-Coefficient (KAP): A measure of agreement between two samples which takes into account agreement caused by chance, making the algorithm more robust.

$$KAP = \frac{f_a - f_c}{N - f_c}$$

$$where:$$

$$f_a = TP + TN,$$

$$f_c = \frac{(TP + TN)(TN + TP) + (FP + TP)(TN + TP)}{N},$$

$$and\ N = number\ of\ voxels$$

Objective 1: Selecting the Ground-Truth Data for the FCNN

To create a FCNN, a ground truth (GT) dataset must be created. For this study, a ground truth image is defined as a femoral neck cross-section with a corresponding region of interest (i.e. bone) that is manually identified on a voxel by voxel basis. The bone ground truth region (ROI) was manually identified with a paintbrush tool able to highlight regions only within a certain threshold range and then manually corrected within Dragonfly. Manual segmentation is time-consuming, averaging ~3.5 hours per single slice in the dataset. Given the time required to segment bone from background, including slices from all bones in our dataset (n=94) or even from all slices in a single dataset was not practical (Femoral Neck Mean Coronal Cross-sectional Slices = 608 +/- 135). However, the FCNN must be trained on GT data that adequately represents the variability of the dataset to perform well (153). Thus, the network was initially trained on a very small set of GT data that was divided exclusively into the training (90% of the GT data) and validation (10% of the GT data randomly removed) set. Model performance was initially based on qualitative assessment and went through several iterations.

- Iteration 1: Three slices were extracted from the most superior, middle, and inferior regions from five randomly selected femoral neck volumes. The FCNN was trained with the following hyper-parameters: input patch size = 112, epoch number = 10, stride-to-input ratio = 1, batch size = 2. Model performance was qualitatively assessed on a subset of the femoral neck cross-sections (Figure 4.3). Femoral neck cross-sections where the network performed most poorly (12 femoral neck volumes from 30 volumes tested) were extracted, manually corrected, and then added to the GT data.
- Iteration 2: Seventeen bone volumes with three slices extracted/bone were used to train a FCNN (hyper-parameters: input patch size = 176, epoch number = 50, stride-to-input ratio =1, batch size = 4). Model performance was assessed by testing the FCNN on 30 additional femoral neck cross-sections that were not used in the training. Model performance was ranked on a scale of 1 (perfect performance) – 5 (worst performance) based on visual assessment. While the segmentation performed well in the majority of cases (~63%), under-identification of bone occurred in scans of low contrast and the

FCNN classified background as bone in scans where the histogram had been shifted to a brighter greyscale range due to metal artifacts (Figure 4.4).

The four volumes of femoral neck cross-sections where the FCNN performed most poorly were identified as cross-sections to add to the final training set. The final GT dataset (i.e. training, validation, and test set) was created based on finding the critical femoral neck cross-sections needed to accurately represent the variability in the density/distribution of bone and scan quality in our cohort. The final GT data used for training and testing our FCNN are described below:

- GT Data for the Training and Validation set: The ground truth data set was created from 21 femoral necks volumes varying in size and bone volume/distribution (male n=12 [27-87 years], female n=9 [29-91 years]). Six 2-D coronal cross-sections were extracted at various locations across each femoral neck (Mean length 18 mm; SD: 3 mm). The validation set consisted of one slice randomly selected from 12 different bones (12 coronal slices) in the ground truth data-set. All remaining coronal slices (114 slices) were used for the training of the network model.
- Test Set: The ground truth test set was created from 10 slices (1 slice/bone) from the 10 bones not used for the training or validation of the FCNN. Selected scancs represented the demographic diversity in the dataset (M [n=5] 24 – 95 years, F[n=5] 29-89 years). The ground truth ROIs of the test set were created by manual segmentation with the caveat that three individuals (DMP, MK, RG) manually segmented each slice and the average of three ROIs was considered the GT test data (i.e., if two or more individuals defined a voxel as bone, that voxel was considered a bone voxel).

Objective 1: Selecting the FCNN hyper-parameters

Briefly, hyper-parameters were selected based on systematic testing of model performance under the guidance of Benjamin Provencher (Dragonfly developer for ORS; Montreal, CA). FCNNs were iteratively trained using different hyper-parameters (e.g., model depth, input patch size, batch size) and the resultant network quality was assessed by quantifying overlap with the test set (153) for results of systematic testing completed by Benjamin Provencher. The following hyper-parameters were selected for the final model: input patch size = 64, stride-to-

input ratio = 0.8, batch size = 64, epoch number = 125, and a categorical cross-entropy loss function. In addition, the reduce learning rate on plateau function, a function that decreases the size of the step and optimizer takes to minimize the loss function when training a FCNN,was selected.

Objective 1: Comparing the final FCNNs to the Otsu Method

Two FCNNs were created and compared to the Otsu threshold method. The Otsu method, named after the inventor Nobuyuki Otsu, is an image thresholding method that automatically binarizes greyscale images into the foreground and background. The Otsu method is an algorithm that iterates through all threshold values in a scan and selects the value that minimizes the within-class variance (i.e. the addition of the variance of the foreground and background multiplied by the weights) (138). Both FCNNs were trained using the training data and hyper-parameters described above. One FCNN, the original FCNN, was trained as previously described. A second FCNN, the augmented FCNN, was trained in the same manner except that images in the training data-set were artificially augmented with additional images created by flipping, rotating, shearing, and elastically stretching original images to add more training to the GT dataset without requiring manual labelling. Resultant network performances were evaluated by quantifying TP, TN, FP, FN, DICE, RI, and VS metrics on the test data.

The FCNN segmentation quality was compared to various iterations of the Otsu method on the test dataset (Figure 4.5). (A) The Otsu was first tested on a single cross-section with no manual correction (group labeled as “Otsu”). (B) Next, Otsu was run on the same cross-section but where air pockets in the femoral neck were identified and digitally replaced with grey-scale values representative of the marrow region (“Otsu Air”). (C and D) Otsu was run again on the original cross-section but for cortical and marrow area separately (“Otsu CM”). Finally, the Otsu was run on the cortical and marrow region separately but on the cross-sections where the air pockets were digitally replaced (“CM Air”). To test the impact of the segmentation technique on the architectural results, Ct.BVF, Tb.BVF, and Tb.Th were calculated for both FCNNs (Original and Augmented) and the Otsu (Otsu, Otsu Air, Ostu CM, CM Air) segmentations on the test set.

Testing FCNN results on Test Set Bone Cubes

For each proximal femur scan in the GT test set (n=10), two three-dimensional bone cubes (8 mm^3) were extracted from the center of the femoral neck and femoral head. Cubes were extracted because the Otsu method is typically performed on bone biopsies and a comparison of these volumes may better represent differences in Otsu and FCNN performance (135). The femoral neck is the region where the network was trained and the femoral head (i.e. representative of the compressive arcade), is a volume never seen in the training network. Qualitative assessment between segmentation methods (FCNN Original, FCNN Augmentation, and Otsu) as well as quantitative assessment of trabecular BVF (Tb.BVF), thickness (Tb.Th), spacing (Tb.Sp), number (Tb.N), and connectivity density (Conn.D) across the different segmentation methods were conducted.

Objective 2: Create a single neural network that can threshold cortical from trabecular area in the sample set.

In Objective 1, the critical femoral neck cross-sections needed to correctly capture the variability in our data set were identified for accurate bone/background segmentation. Therefore, to train a new FCNN for cortical bone segmentation, the same GT cross-sections (male n=12 [27-87 years], female n=9 [29-91 years]) were used for the training (114 GT slices) and validation (12 GT slice) set. Also, the same test set (10 additional GT slices not used for training or validation) for bone background identification was used to test the FCNN quality for cortical bone identification. Please see example coronal GT cross-sections with the cortical bone highlighted in red, observed in Figure 4.6.

Similar to Objective 1, the FCNN was initially trained using hyper-parameters based on systematic testing of training results (data not shown). The Iteration 1 model (hyper-parameters: input patch size = 256, stride-to-input ratio = 1, epoch number = 150, loss function = categorical cross-entropy) was run on all femoral neck volumes not used in training the model and qualitatively evaluated. Of the 94 bones in the dataset, the bones in which the

segmentation algorithm performed most poorly (n=8) were added to the training data with the caveat that only five slices were extracted instead of six. The final model had a GT training set of 144 slices and a validation set of 12 slices (hyper-parameters: input patch size = 256, stride to input ratio = 1, epoch number = 150, loss function = categorical cross-entropy, optimization). DICE coefficients were quantified using the GT truth test set to assess model performance. Two additional metrics, the Rand index (RI), a measure of similarity between clustering, and the true positive rate (TPR), the portion of positive voxels in the ground truth that are also identified as positive by the FCNN, were quantified (164).

Experiments and Statistical Results

All statistics were conducted with R 3.1.2 (165) using RStudio (166). For Objective 1, the final training accuracy of FCNNs relative to the validation set were reported. Means and standard deviations of all similarity measures of overlap were reported for FCNN and Otsu methods relative to the test set. Significant differences in model performances were evaluated using paired t-test with a Welch approximation to adjust for unequal variances for all similarity measures of overlap. Mean and standard deviations of Ct.BVF, Tb.BVF, and Tb.Th were reported for the single slice coronal cross-sections using the GT, FCNNs, and Otsu segmentation methods. A paired t-test with Welch approximation and Bonferri correction for multiple comparisons was used to test for differences in architectural results between the GT and various segmentation approaches. Percent differences in the results between the augmented FCNN and the GT and the Otsu CM-Air and GT were also reported. Linear regression analysis of the trabecular architectural results (Tb.BVF, Tb.Th, Tb.Sp, Tb.N, and Conn.D) was conducted to compare the Augmented FCNN to the Otsu and Original FCNN for both the femoral head and neck bone cubes. Finally, means and standard deviations of similarity measure of overlap (DICE, VS, KAP, RI, and TPR) were reported for the cortical bone segmentation FCNN compared to the GT test set.

Results

Objective 1a: Single Slice Results Comparing the final FCNNs to the Otsu Threshold

The original and augmented FCNN had a training accuracy of 0.994 and 0.995 compared to the validation set, respectively. All DICE, VS, and KAP coefficients (Mean[SD]) were reported for FCNN and Otsu methods (Table 4.2). There was no significant difference between the two FCNNs in terms of DICE ($p = 0.983$), VS ($p = 0.982$), or KAP ($p = 0.277$). However, the augmented neural network had significantly more TP and TN voxels identified and significantly less FP voxels identified ($p<0.001$ in all cases). As a result, the Augmented FCNN was used as the reference dataset when comparing the FCNN performance to all Otsu results (paired t-test). For all similarity metrics of overlap, the FCNN had a significantly higher (i.e. better) coefficient measure compared to the ground truth test set ($p < 0.01$) (Table 4.3). As is visible in Figure 4.7, when additional manual intervention steps are added to the Otsu threshold (blue), the DICE coefficient improves but still significantly underperforms compared to augmented FCNN (red).

Objective 1b: Impact of Segmentation on Bone Architectural Results

Ct.BVF, Tb.BVF, and Tb.Th results for the ground truth (GT) and all FCNN and Otsu methods are reported below (Table 4.4). The percent difference between the GT and augmented FCNN on architectural results were on average 1.4 % lower, 1.2% higher, and 3.1% higher for Ct.BVF, Tb.BVF, and Tb.Th compared to the ground truth results, respectively. The percent difference between the GT and Otsu CM-AP on architectural results were on average 2.4 %, 39.0 %, and 21.7 % higher for Ct.BVF, Tb.BVF, and Tb.Th compared to the ground truth results, respectively. Cortical and trabecular results quantified using every Otsu threshold significantly differed from the GT segmentation results except for Tb.th using the All-AP and CM-AP and for Tb.BVF using the CM-AP threshold (Table 4.5). Surprisingly, there was a small but significant difference in Ct.BVF for the Augmentation FCNNs compared to the GT result. Ct.BVF, Tb.BVF, and Tb.Th results are visually presented below (Figure 4.8).

Objective 1c: Trabecular Bone Architectural Results in Extracted Bone Cubes

One bone cube extracted from the femoral neck had very little trabecular bone (BVF ≈ 0.0016) and the Otsu method was unable to threshold bone from background. This outlier was removed from the following analysis. Minimal differences in architectural results (Table 4.6) and segmentation quality (Figure 4.9) were observed between all segmentation methods in the femoral neck and compressive arcade bone cubes. Linear regression analysis of trabecular results between both the Original and Augmented FCNNs was tightly linearly correlated (Mean: $p<0.001$; R^2[adj] = 0.998). Linear regression analysis of the augmented FCNs and the Otsu method was completed for trabecular BVF, thickness, spacing, number, and connectivity density for bone cubes extracted from the femoral head and neck (Figure 4.10). Linear regression analysis was highly correlated for some (BVF, Tb.Sp, and Tb.N) but not all (Tb.Th, and Conn.D) results. Our findings demonstrate that trained FCNNs perform equally well in small extracted cubes of the proximal femur compared to the Otsu and that a trained FCNN can segment bone from background in regions never seen in training.

Objective 2: FCNN to Automatically Segment Cortical Area

The FCNN trained for cortical bone segmentation had an accuracy of 0.9800 compared to the validation set. Compared to the GT segmentation on the test set, the trained FCNN had high DICE (0.956 +/- 0.021), VS (0.991 +/- 0.006), KAP (0.954 +/- 0.022), RI (0.991 +/- 0.004), and TPR (0.962 +/- 0.019) (Figure 4.11)

Discussion

In this study we trained two FCNNs to segment bone from background and delineate cortical from marrow area in nano-CT scans of the femoral neck. While visually possible to separate bone in CT scans, there are still difficulties with automated segmentation methods in practice. Our FCNNs overcame such challenges despite variability in bone structure and scan quality in isolated bone specimens in the dataset. These findings support the hypothesis that a FCNN trained to segment bone from background outperforms the commonly used Otsu threshold. Both networks are publicly available and free for academic use (available for download on the infinite toolbox in Dragonfly 4.0). Thus, this method is currently an immediate and alternative

method for bone segmentation in CT scans. Finally, both networks were created with the exact same underlying FCNN architecture, demonstrating the wide range applicability of using FCNNs to segment structures in high-resolution CT scans.

FCNNs have proven to be a promising segmentation tool in the musculoskeletal field, particularly for clinical magnetic resonance (MR) and CT scans for cartilage and bone identification (e.g., proximal femur, vertebral body, and skull) (154,159–161,163). Despite widespread research on clinical applications, we are, to the best of our knowledge, the first to demonstrate FCNN applicability in high-resolution ex-vivo CT scans; scans commonly used to assess disease progression and/or drug treatment response in musculoskeletal research (32). Bone segmentation methods still often rely on a global threshold (11,32,167). While global thresholds often perform well on scans of bone biopsies, we showed that a global Otsu threshold can overestimate Tb.BVF by 39% on highly heterogeneous structures. Since segmentation may have a profound and critical effect on the findings and reproducibility of any study, the method should be selected with great care. Our FCNN outperformed other approaches with minimal differences in architectural results (1.2% for Ct.BVF, 3.5 % for Tb.BVF, and 3.5% Tb.Th) compared to GT controls. Our study demonstrates how effective a well-trained FCNN can be for segmentation of CT scans.

A second substantial finding was that our FCNN, trained to segment entire cross-sections, performed well on bone cubes extracted from both the femoral neck and the femoral head. Due to the field-of-view to resolution limitations with micro-CT, cadaveric bone architectural analysis is typically quantified on bones on biopsies (129). We showed that our FCNN is able to accurately segment structures that are commonly used for architectural analysis (32). In fact, we demonstrated that the input image does not need to contain the same basic structure (i.e., a cross-section with cortical bone surrounding trabecular area), to perform well. Further, our FCNN, based on qualitative assessment, performed as well as the Otsu method on femoral head volumes, volumes not seen in training or testing of the network. The advantages of using a deep learning approach for segmentation are obvious, and our findings prove that our network will perform well, at the very least, on different regions of cadaveric bones scanned with the same acquisition parameters. Future work will focus on testing the versatility of these FCNNs on

wide-ranging datasets (e.g., anatomically different bones, scanned with different a system, and varied scanning acquisition parameters).

In this study, we found no difference in segmentation quality between the original and augmented FCNN. However, others have shown that adding augmentation to training data improved FCNN segmentation accuracy (154,158). We believe this discrepancy occurred due to two main factors. (1) The augmentation metrics applied to our training data (i.e., rotation, shear, and flipping) did not meaningfully address any variability observed in our dataset. For example, we rotated our training data with augmentation but the femoral neck cross-sections were aligned in the same direction prior to applying the FCNN. (2) Our GT training data was pre-selected to represent the variability in our femoral neck dataset. We believe that the network accuracy was saturated pre-augmentation. However, there was also no negative impact on the quality of the models when augmentation was applied. Thus, while results may not portray this, augmentation still has the potential to add tremendous value to FCNN segmentation quality, particularly if the training data is limited and the augmentation meaningfully addresses the variability of observed cases in the dataset. Future work should test training a FCNN with a smaller dataset and systematically evaluate augmenting training data in a way that represents dataset variability to improve FCNN accuracy.

We developed an accurate and novel FCNN to automatically delineate cortical from marrow area in scans of femoral neck cross-sections. To our knowledge, we are the first to apply deep learning for automatic segmentation of this difficult, time consuming, and arduous task. As previously mentioned, many alternative semi-automated approaches have been proposed to standardize and quicken this segmentation process (142,144,145). However, I am the first to propose a method that requires no preparatory manual intervention/correction or setting dataset optimized algorithm-specific parameters. Instead, our FCNN is an out-of-box approach that was trained to perform well despite variations in signal to noise ratio. Interestingly, the same training set used for the bone background segmentation was not adequate to accurately train a network for cortical/marrow segmentation (Iteration #1). Cortical-trabecular segmentation in the femoral neck is complicated due to the fact that cortical bone has regional variability in thickness and porosity, and requires a user to have an understanding of the underlying

structural biology. Perhaps the larger the requirement of a FCNN to "learn" biologically driven segmentation, the larger the requirement for training data.

Despite success with training FCNNs for segmentation, some limitations need to be addressed. Substantial time and effort was spent on manually creating the GT dataset for the bone/background (3 hours/slice) and cortical/marrow (30 min/slice) FCNNs. Given the scan-to-scan variability in our data, I felt there was no existing alternative segmentation method that would perform accurately. However, efficiency may have been improved if a smaller GT dataset was used and augmentation was added to more effectively represent the scan variability (e.g., add noise and shifting the histogram). A second limitation of this study was that this network was a trained 2-D FCNN which was applied on a slice-by-slice basis to segment a 3-D structure. Theoretically, there are a number of advantages to using a network algorithm able to take input from all dimensions to determine a voxel type (i.e. accurately assessing partial volume effects). In fact, multiple studies have demonstrated that a 3-D FCNN outperforms a 2-D FCNN in terms of segmentation accuracy (163,168). However, given the nature and size of the scans in our study, it was not feasible to train a network with 3-D dimensional GT data. Thirdly, our hyper-parameters were selected via manual selection with systematic testing. Instead, a number of automated methods exist (e.g., grid search and random search optimization algorithms) and could be applied to concisely and accurately determine the best hyper-parameters for model performance (169). Finally, while the basic architecture is known for any FCNN, neural networks are notorious for being considered a black-box approach for segmentation (153,170). It is important to understand the underlying mechanics/mathematics of any segmentation approach to trust the algorithm and identify potential pitfalls. Some methods, such as Local Interpretable Model-agnostic Explanations (LIME), can help explain the underlying mathematical structure and can provide an explanation for individual predictions (170). Future work should focus on applying such approaches to better explain why our FCNNs performed well on the test data. Alternatively, other tensor based approaches for segmentation provide promise, but the feasibility on CT scans has yet to be assessed (171,172). Despite limitations, it is clear that we have conclusively created two FCNNs to accurately segment bone from background and cortical from marrow area in the femoral neck.

In conclusion, we have created two FCNNs that have been trained to accurately segment cortical from trabecular bone and bone from background in nano-CT scans of the femoral neck that have both biological and scanning artifact variability. FCNNs developed in this study are free and available for use on Dragonfly 4.0 and provide a reproducible, completely automated, approach for segmentation. Our FCNNs accurately quantify architectural results and provides a novel solution to overcome the 'crisis' of reproducibility that may occur due to threshold errors in musculoskeletal research today.

Acknowledgments

The research reported in this chapter would not have been possible without the help of many. First off, I would like to acknowledge that this work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health **(**KJJ: AR065424, AR069620; TLB: AR06424). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Further, I would like to graciously acknowledge the employees from Orthopedic Research Systems (ORS), in particular, Mike Marsh, Benjamin Provencher, and Nicolas Piche, who were critical in getting this work started and helped with model and hyper-parameter selection using Dragonfly Software 4.0. In addition, I would like to thank Sean Carroll who helped reorient and extract the bone cubes used for architectural analysis. Further, I would like to thank Todd L. Bredbenner, who donated a significant amount of time to provide mentorship regarding testing the model accuracy. Finally, as mentioned earlier, this work would not have been possible without Rob Goulet, who was involved in every aspect of this project.

Figures

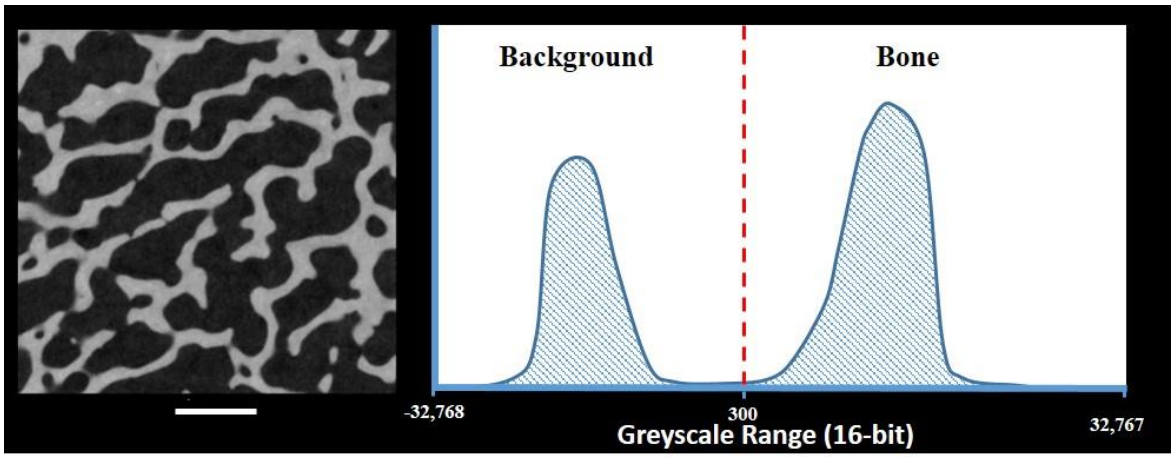


Figure 4.1 (left) Cross-section of a nano-CT scan (14 μm, GE, Germany) where bone and background can be easily delineated (1 mm scale bar) and (right) an example figure of a bimodal histogram where bone and background have no overlapping greyscale values.

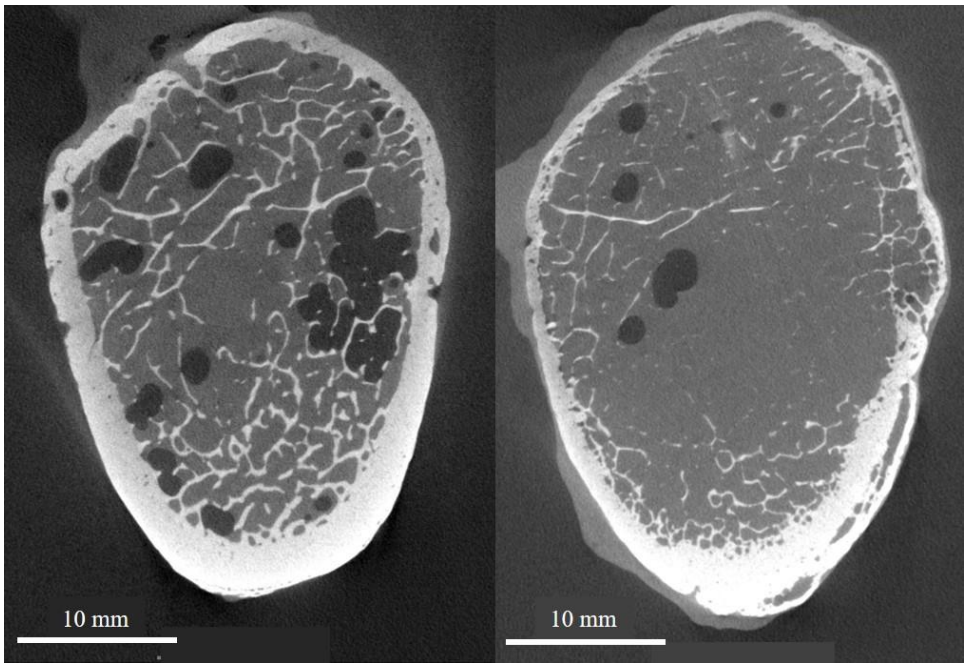


Figure 4.2 Two femoral neck cross-sections of 28 y.o. (left) and 98 y.o. (right) female cadaveric specimens demonstrating large differences in BVF, thickness, and grayscale intensity within and across scans.

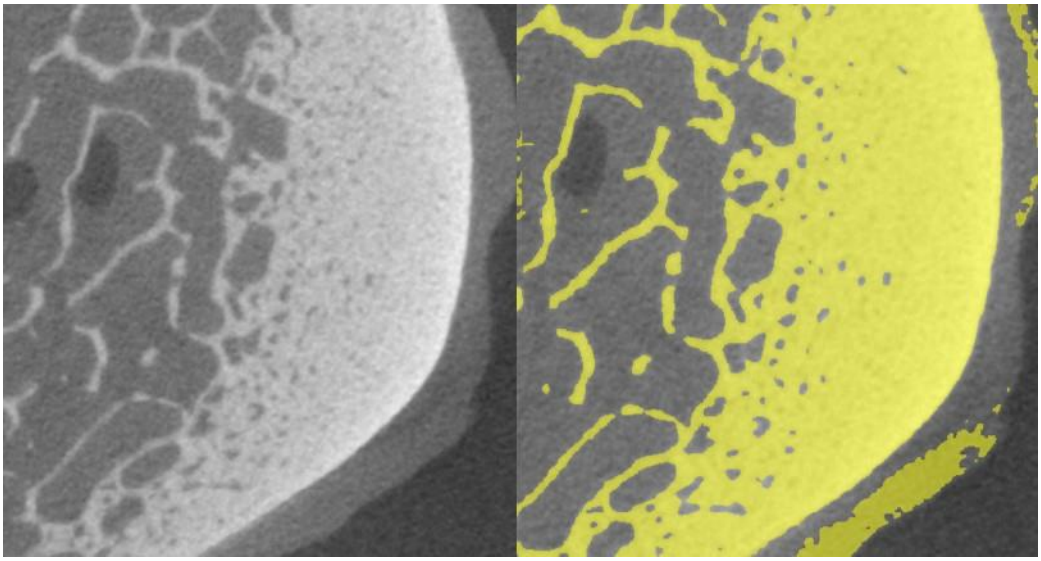


Figure 4.3 Example output of a FCNN post-training (iteration 1) on a femoral neck cross-section (left) and output (right) where the segmentation method is misidentifying soft tissue as bone

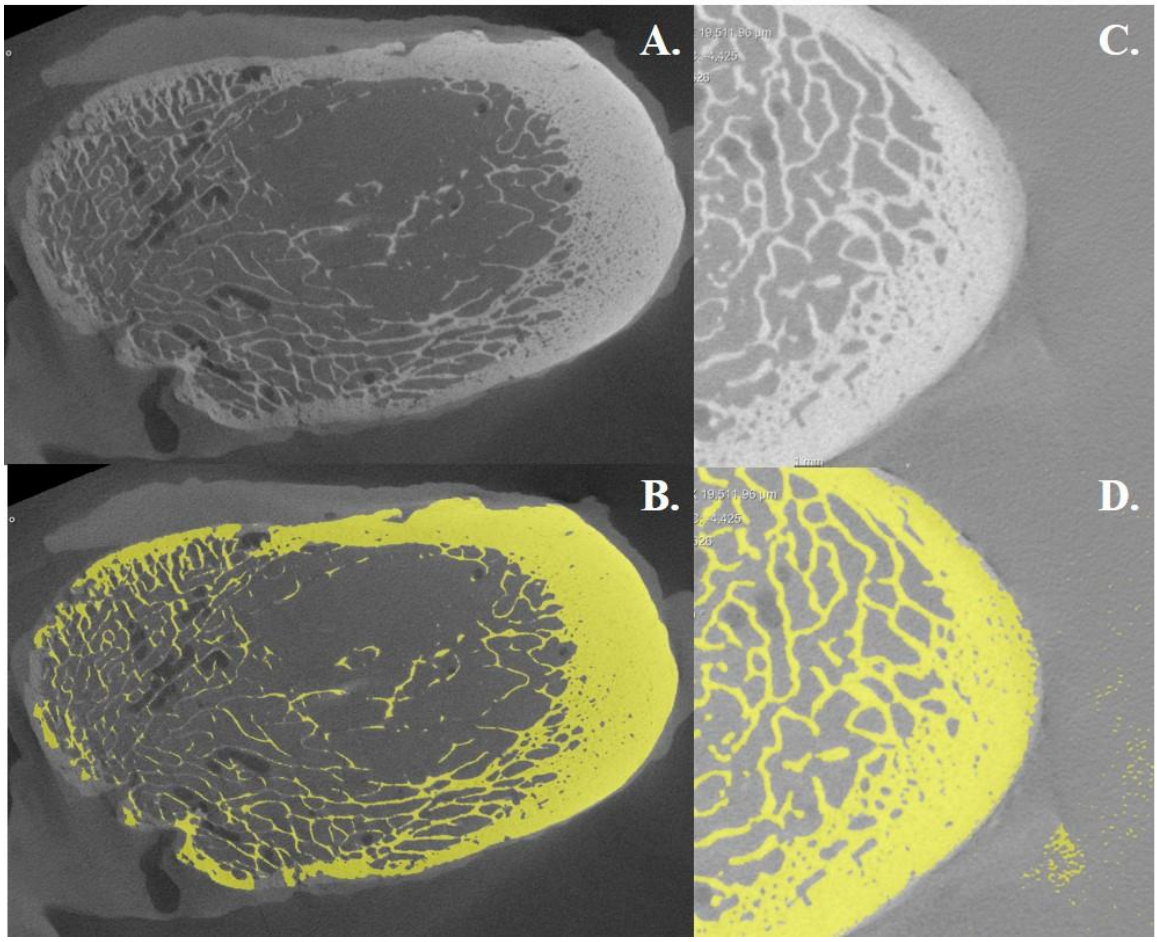


Figure 4.4 A scan where there is relatively poor contrast between bone and background (A) and a shifted histogram (C). These outlier scans are representative of worst case segmentation output (B and D) of the FCNN trained in iteration two. Cross-sections from these scans were used for training the final model.

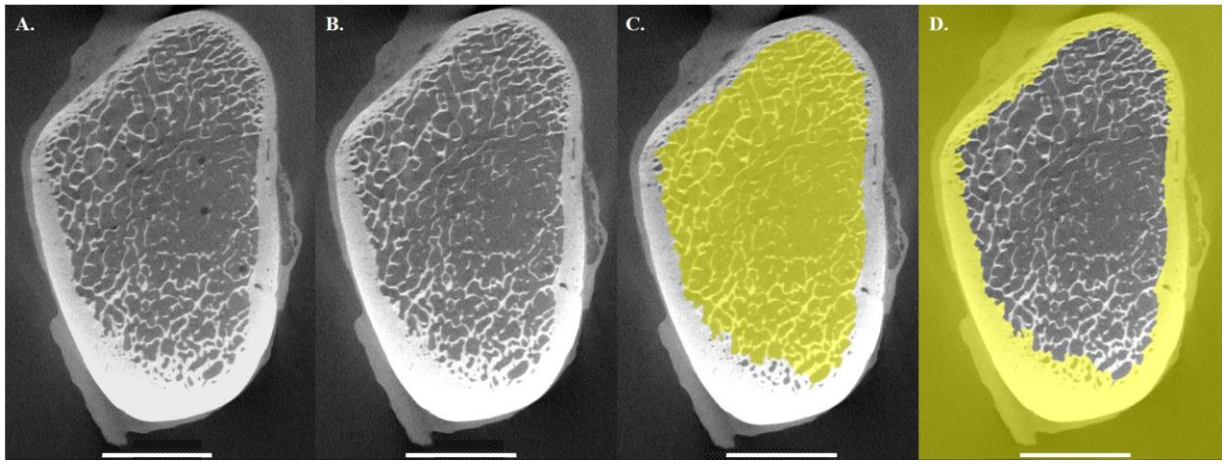


Figure 4.5 Example cross-sections for Otsu segmentation on; (A)Otsu: the entire image, (B) Otsu Air: the entire image with air pockets digitally replaced, and Otsu CM: where the Otsu method was run on the (C) marrow, and (D) cortical regions separately (scale bar = 10 mm).

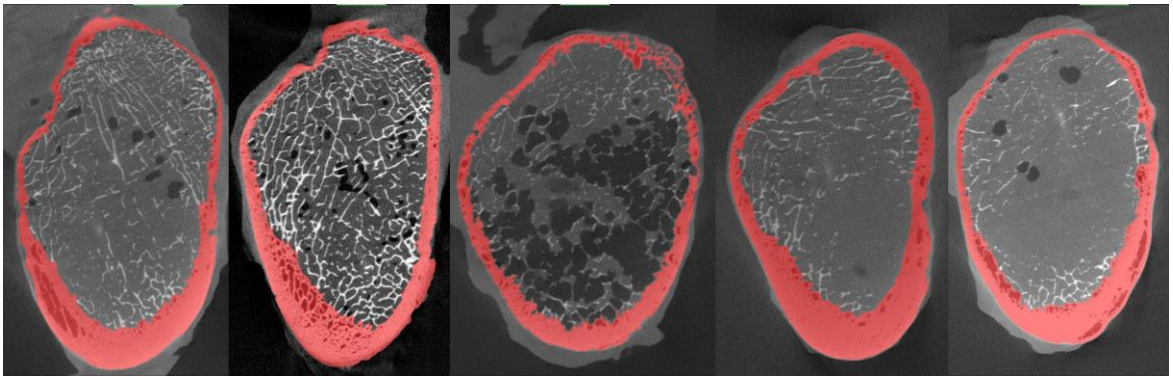


Figure 4.6 A panel of coronal femoral neck cross-sections that were used as GT data for the training a FCNN to segment cortical bone (red) from the background.

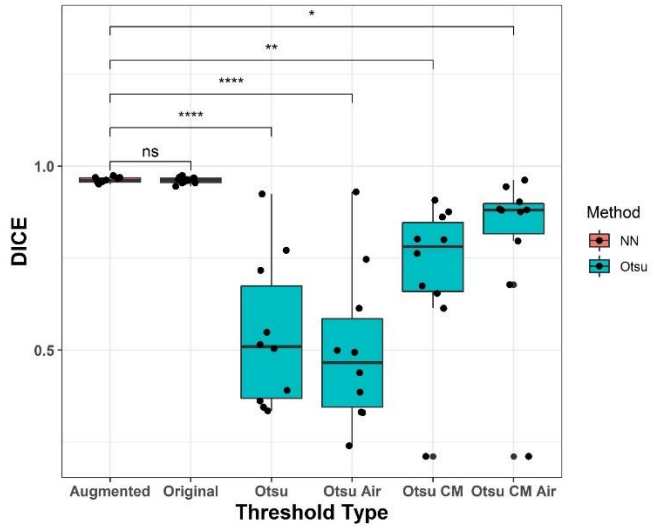


Figure 4.7 DICE values of both FCNNs (Red) and Otsu methods (Blue). The FCNNs were not significantly different from each other but the augmented FCNN significantly outperformed all Otsu methods (Blue).

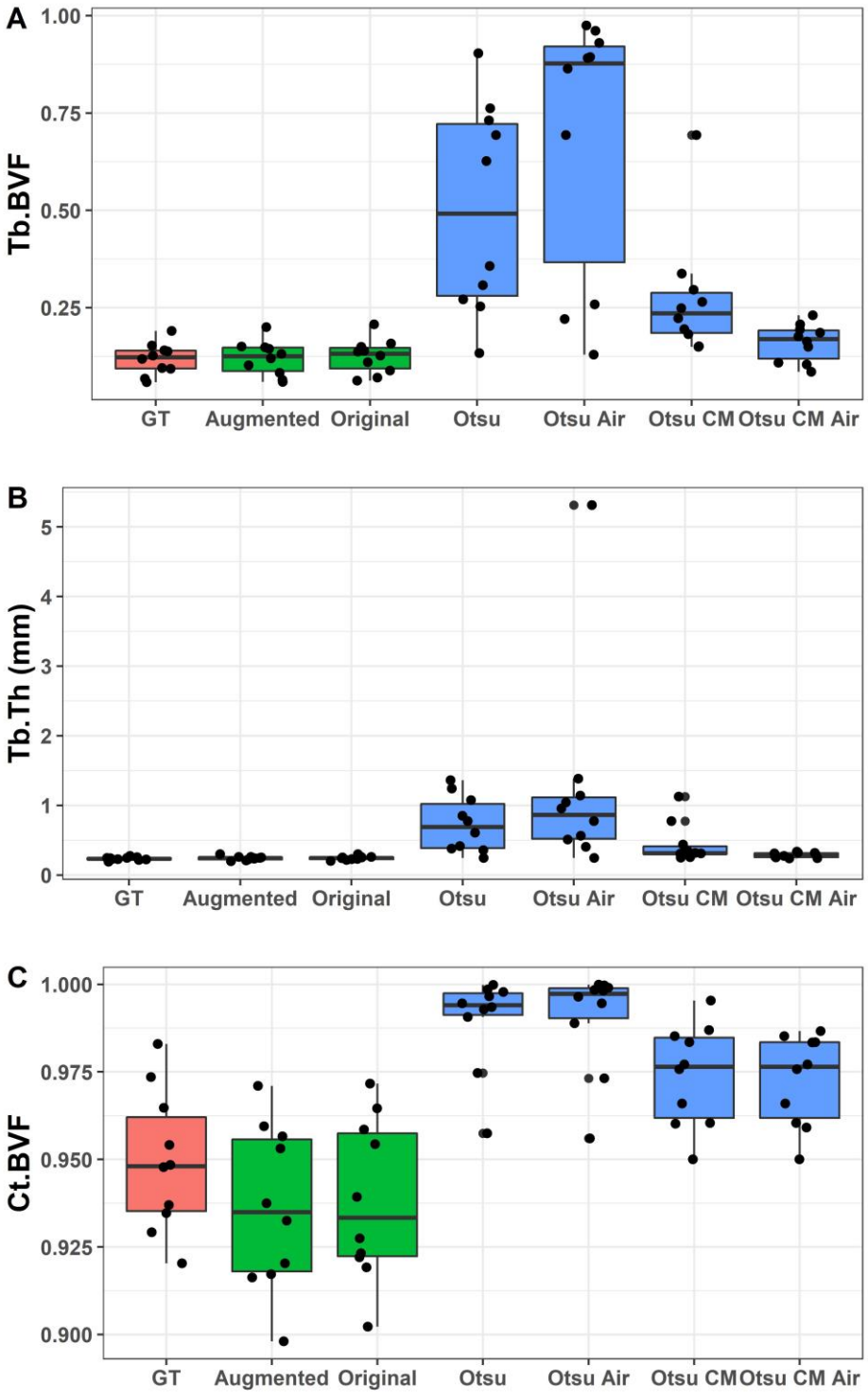


Figure 4.8 A box and whiskers plot of (A) Tb.BVF, (B) Tb.Th, and (C) Ct.BVF results for the GT (red), FCNNs (Green) and Otsu (blue) methods to segment bone from background.

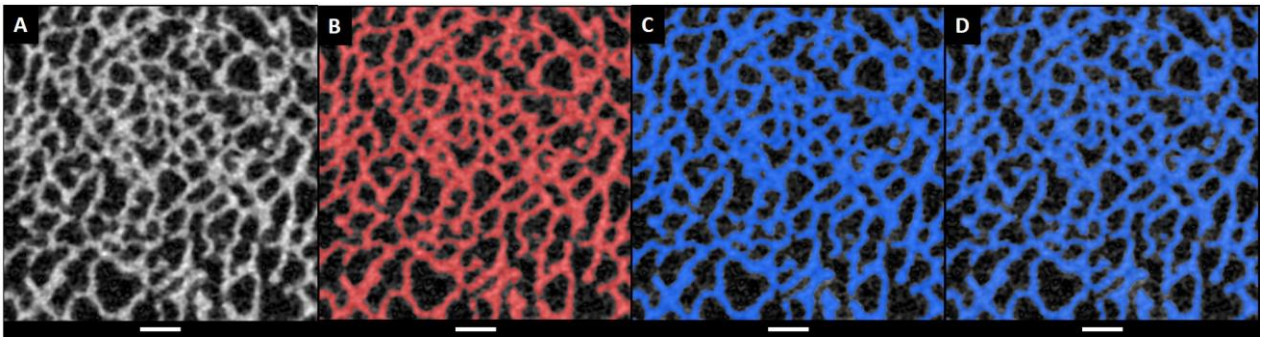


Figure 4.9 Panel Comparison of a (A) cross-section of a trabecular bone cube extracted from the femoral head, (B) the Otsu segmentation highlighted in red, (C) the original FCNN segmentation highlighted in blue, and (D) the augmented FCNN segmentation indicated in blue.

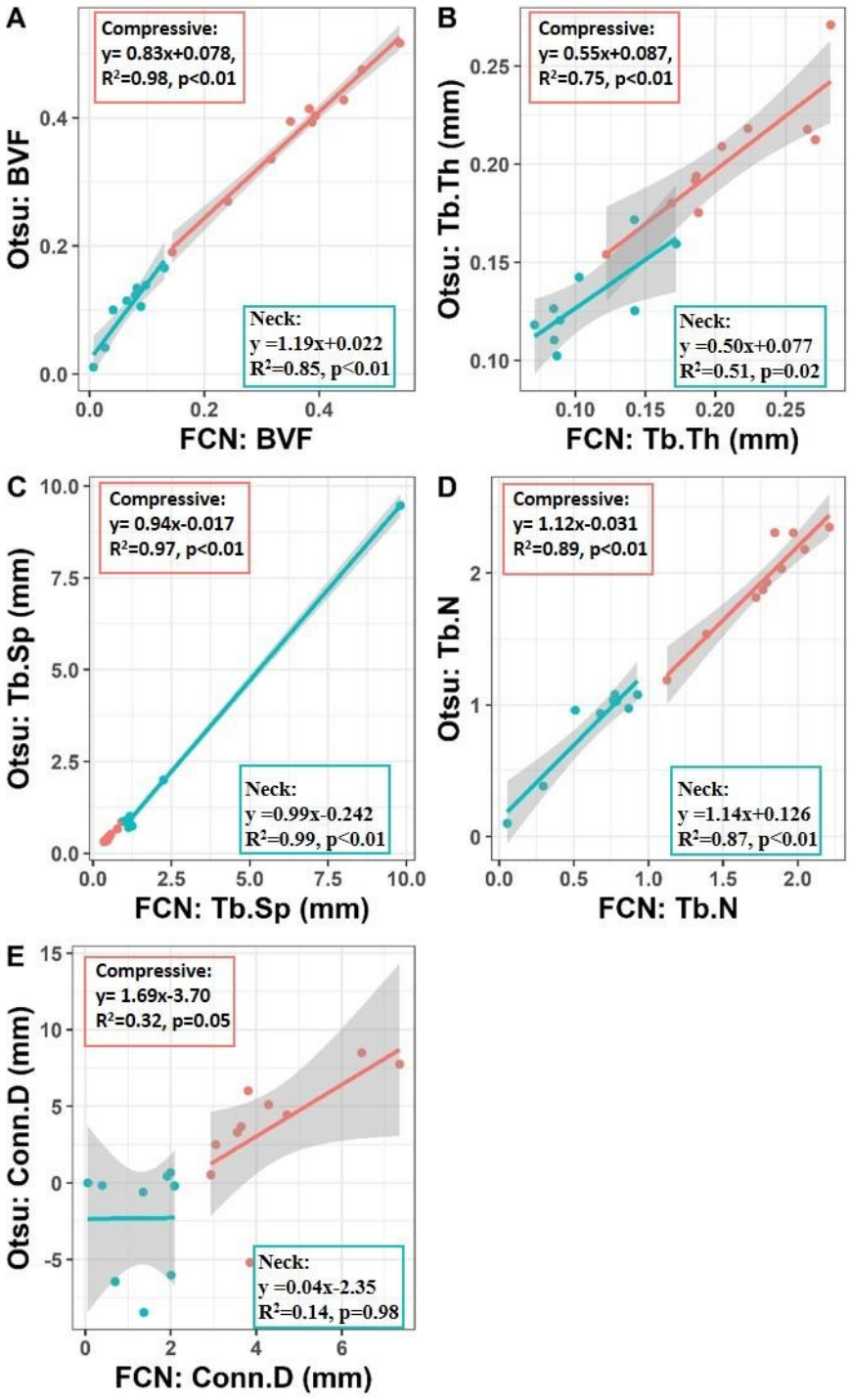


Figure 4.10 Linear regression analysis of bone cubes extracted from the femoral neck (red) and head (blue) for BVF, Tb.Th, Tb.Sp, Tb.N, and Conn.D for the augmented FCNN (labeled FCN) and Otsu segmentation methods.

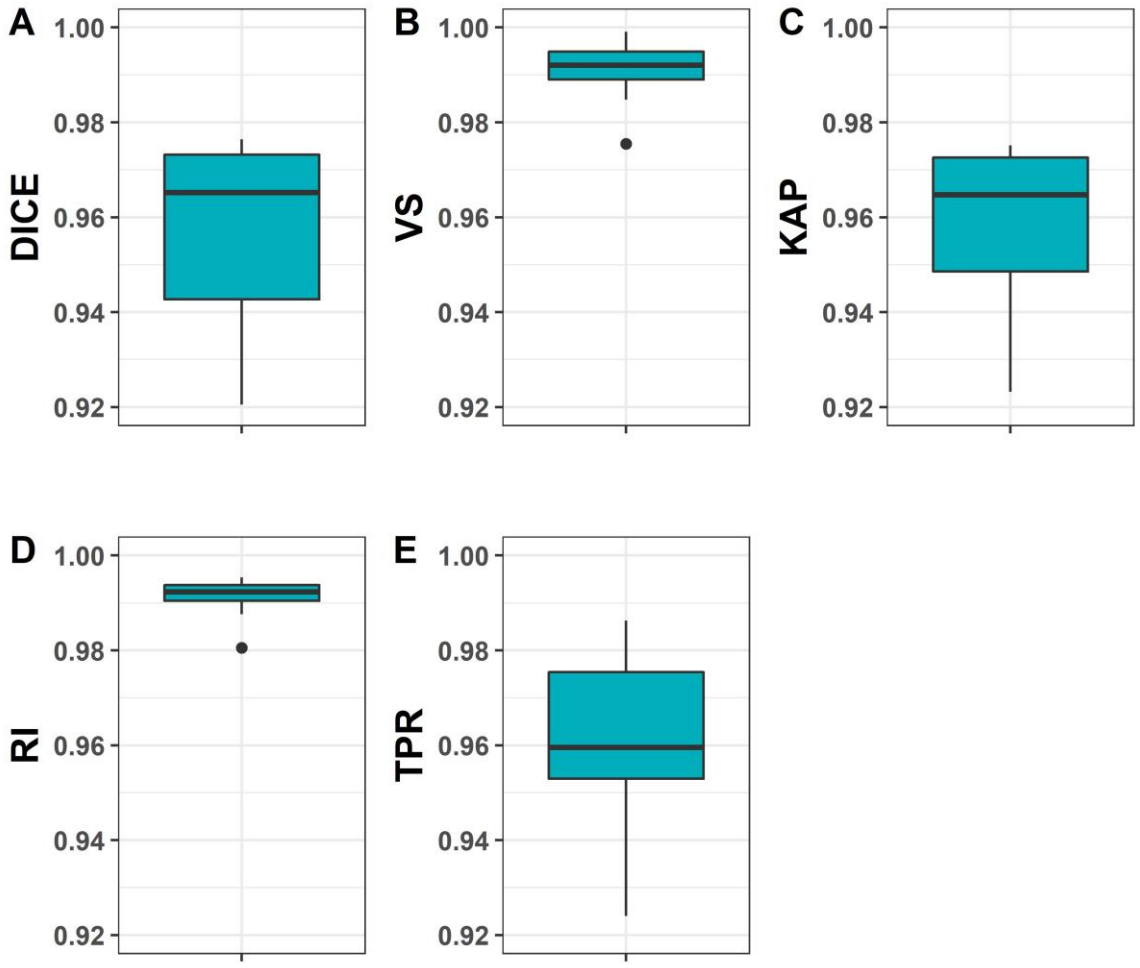


Figure 4.11 Box and whiskers plot for (A) DICE, (B) VS, (C) KAP, (D) RI, and (E) TPR of a FCNN for cortical segmentation compared to the GT test set data.

Tables

Table 4.1 Definition of hyper-parameters that require manual selection for a FCNN with U-net architecture.

Hyper-parameters	**Definition**
Patch Size	During training, 2-D data is split into a user defined input patch size that is smaller than the data-set dimensions. Patches are moved around the data-set using a sliding window to predict the label of each pixel in a patch.
Batch Size	The batch size defines the number of patches to work through before updating the internal model parameters in training.
Epoch Number	Defines the number of times a learning algorithm will pass through (i.e. run the model, assess the errors, and update the internal parameters) while training the model.
Loss Function	A function used in training a network to compute the error, or distance between a model outcome and the desired solution (target).
Optimization Algorithm	Objective function: A heuristic algorithm used for searching for an optimal solution. It is used to determine which weights and the particular magnitude of weights are updated.
Stride-to-Input Ratio	Objective function: How far the patch moves in the sliding window from one position to the next relative to the patch size.

Table 4.2 DICE, VS, and KAP (Mean [SD]) measures were reported for the two FCNNs (Original and Augmented) and all Otsu methods (Otsu on the entire cross-section [All], Otsu on the entire cross-section with the air pockets replaced [All-AP], cortical/marrow regions separate[CM], and CM-AP).

Type	**Labels**	**DICE**	**VS**	**KAP**
FCNN	Original	0.961 [0.01]	0.986 [0.01]	0.958 [0.01]
FCNN	Augmentation	0.962 [0.01]	0.986 [0.01]	0.959 [0.01]
Otsu	All	0.541 [0.20]	0.580 [0.20]	0.484 [0.23]
Otsu	All-AP	0.501 [0.21]	0.539 [0.21]	0.435 [0.24]
Otsu	CM-AP	0.716 [0.20]	0.761 [0.12]	0.685 [0.23]
Otsu	CM-AP	0.801 [0.22]	0.851 [0.12]	0.778 [0.26]

Table 4.3 P-values from paired t-test testing for significant differences between Augmented FCNN to the Original FCNN and all Otsu methods (p-values <0.05 are in bold).

Type	**Labels**	**DICE**	**VS**	**KAP**	**TP**	**TN**	**FP**	**FN**
FCNN	Original	0.982	0.982	0.277	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Otsu	All	**<0.001**	**<0.001**	**<0.001**	0.888	**<0.001**	**0.001**	0.424
Otsu	All-AP	**<0.001**	**<0.001**	**<0.001**	0.246	**<0.001**	**<0.001**	0.423
Otsu	CM-AP	**<0.001**	**<0.001**	**0.005**	0.677	**0.015**	**<0.001**	0.394
Otsu	CM-AP	**0.007**	**0.007**	**0.049**	0.589	0.077	**0.009**	0.382

Table 4.4 Mean and standard deviations calculated for all GT, FCNN, and Otsu bone segmentation methods for Ct.BVF, Tb.BVF, and Tb.Th..

Type	**Labels**	**Ct.BVF**	**Tb.BVF**	**Tb.Th (mm)**
GT	-	0.949 [0.020]	0.118 [0.040]	0.235 [0.023]
FCNN	Original	0.936 [0.023]	0.120 [0.044]	0.242 [0.029]
FCNN	Augmentation	0.938 [0.023]	0.125 [0.044]	0.245 [0.030]
Otsu	All	0.990 [0.013]	0.504 [0.267]	0.731 [0.395]
Otsu	All-AP	0.990 [0.015]	0.682 [0.341]	1.233 [1.48]
Otsu	CM-AP	0.974 [0.014]	0.274 [0.160]	0.443 [0.285]
Otsu	CM-AP	0.973 [0.013]	0.160 [0.048]	0.285 [0.035]

Table 4.5 All p-values of paired t-test with Bonferri correction comparing the difference between the ground truth and Ct.BVF, Tb.BVF, and Tb.Th for all FCNNs and Otsu bone segmentation methods ($p < 0.05$ values in bold).

Type	**Labels**	**Ct.BVF**	**Tb.BVF**	**Tb.Th (mm)**
FCNN	Original	0.053	0.568	0.071
FCNN	Augmentation	**0.030**	1.000	0.380
Otsu	All	**0.001**	**0.004**	**0.019**
Otsu	All-AP	**0.001**	**0.002**	0.364
Otsu	CM-AP	**0.005**	0.068	0.253
Otsu	CM-AP	**0.005**	**0.001**	**0.001**

Table 4.6 Representation of average differences in segmentation using either the original FCNN, augmented FCNN, or Otsu methods are presented below.

Bone Architecture	Compressive Arcade Box			Tensile Arcade Box		
	Original	**Augmentation**	**Otsu**	**Original**	**Augmentation**	**Otsu**
BVF	0.371 [0.11]	0.367 [0.11]	0.382 [0.10]	0.065 [0.04]	0.062 [0.04]	0.167 [0.20]
Tb.th (mm)	0.203 [0.04]	0.210 [0.05]	0.202 [0.03]	0.106 [0.04]	0.117 [0.04]	0.188 [0.18]
Tb.Sp (mm)	0.560 [0.19]	0.533 [0.18]	0.480 [0.17]	2.741 [3.19]	2.758 [3.20]	1.766 [2.75]
Conn.D	4.212 [1.52]	4.363 [1.54]	3.67 [3.92]	1.247 [0.96]	1.169 [0.86]	-0.396 [6.90]
Tb.N	1.735 [0.33]	1.776 [0.32]	1.95 [0.37]	0.583 [0.34]	0.583 [0.34]	1.073 [0.80]

Chapter 5 Associations Between Regional Variation in bone microstructure and hip strength for men and women

Introduction

Osteoporosis, Hip Fractures, and Sex-Specific Etiology

Osteoporosis is a skeletal disorder characterized by deterioration of bone tissue, resulting in decreased bone strength and predisposition to fractures (173,174). Osteoporosis, a slow and insidious disease that occurs with advancing age, often leaves a patient asymptomatic until a fragility fracture occurs (175). Fragility fractures are fractures that occur with minimal trauma (e.g., fall from standing height or less) and are associated with several adverse outcomes: increased risk of subsequent fracture, associated morbidities, increased mortality rate, etc. (176). Hip fracture, one particular type of fragility fracture, are so common globally that one occurs every 20 seconds (177). Hip fractures are costly, associated with a severe decrease in the quality of life, and nearly 50% of patients (>65 years) who suffer a fracture to the hip never regain normal function (7,8). Thus, as the proportion of elderly individuals increases globally, diagnostic and treatment methods must be improved to mitigate osteoporosis-related hip fragility fractures.

Hip fractures are a serious public health problem for both sexes (4,7,173,178). However, there are distinct gender-related differences in both the presence and associated outcomes. Approximately 75% of hip fractures occur in females, but males have up to a 2-fold increase in mortality rate compared to age-matched females (4,8,179). Differences in hip fracture risk between males and females are largely attributed to two main factors:

(1) On average, males are larger in height and weight compared to females and thus have bones that are physiologically developed to be larger and stronger.
(2) While bone loss occurs with age for both sexes, females lose bone at a faster rate and earlier in life compared to males. Sex-specific differences in bone loss largely occur

> because men do not experience the rapid loss of bone mass that women do following menopause. Menopause results in rapid hormonal changes (i.e., estrogen deficiency) which impairs the carefully orchestrated normal bone remodeling process. This deficiency, at the most basic level, results in an increase in osteoclastic bone resorption activity without an increase in bone formation, leading to a net loss of bone (180).

These two factors play an important role in the sex-specific nature of hip fragility fracture risk because bone, as mentioned previously (Chapter 1, pg. 4), is a mechanically responsive organ that is optimized at both the micro- and macro scale based on habitual loading. As a result, sex-specific differences in loading due to anatomical differences (e.g., pelvis shape and forces from muscle attachments) may also play a role. Sex-specific differences in the surrounding anatomies are briefly described in the following section.

Anatomical differences between males and females

It is well documented that females have a significantly wider pelvis then body-size matched males (181–183). However, beyond this basic difference, there are a variety of biomechanically relevant, sexually dimorphic, differences at the joint, where the femur head (ball) and the acetabulum (socket) of the pelvis are in direct contact, and in surrounding tissues (Table 5.1). The ball and socket joint acts as the site of primary loading during normal activities (e.g., walking and running) in the proximal femur and, as a result, is a critically important site for macro- and micro- structural development of bone. Known sex-specific anatomical differences surrounding the proximal femur are described (Table 5.1).

Surprisingly, known sex-specific anatomical differences surrounding the proximal femur have remained largely unexplored in the context of osteoporosis. The magnitude and local orientation of loading likely differ in a sex-specific manner and these differences add to the argument that how bone is lost and optimized varies differentially relative to sex.

Current Diagnostic Method to Assess Fragility Fracture Risk and Limitations

The gold standard method to clinically assess hip fracture risk uses a two-dimensional bone mineral density (aBMD) T-score of the femoral neck (18). A T-score uses BMD of a young, healthy, gender-specific population as a reference (BMD_{ref}) and is defined using the following equation:

$$T - score = \frac{BMD_{patient} - BMD_{ref}}{SD_{BMDref}} \ (Standard\ Deviation = SD)$$

A patient is diagnosed as osteoporotic if they have a T-score that is below -2.5 (18). Despite being the clinical gold standard method to assess fracture risk, aBMD T-scores are limited because patients are frequently left underdiagnosed and undertreated. Staggeringly, it has been shown that 90% of men and 70% of women who suffered a fragility hip fracture received no prior treatment (US Medicare population from 2001–2011) (184,185). Reasons for under-diagnosis of individuals who fracture are multi-factorial; however, BMD scores are inherently limited because nearly half of the individuals that fracture have BMD T-scores greater than -2.5 (186). Thus, we move beyond the existing DXA-derived BMD paradigm of diagnosis to improve fracture risk predictions.

A Hip Fracture is a Biomechanical Event that Depends on Underlying Bone Microstructure

A hip fracture occurs when the proximal femur is loaded beyond its structural strength. Therefore, a biomechanical approach that considers bone strength provides a more direct approach to evaluating sex-specific differences in fracture susceptibility (187). Bone strength is determined by a combination of bone size, shape, microstructure, and material properties (28,188,189). BMD can inform on the quantity of bone, but currently gives no insight into bone quality, which, by definition, are all bone traits (i.e., microstructure, morphology, material properties, and shape) that affect bone strength but are not accounted for by bone mass or quantity (Figure 5.1) (Hernandez and Keaveny, 2006). A fundamental issue that hinders our ability to improve fracture risk is that there is a lack of understanding of the influence of bone quality on sex-specific differences in bone strength (19,190). Microstructural properties are a measure of bone quality, that likely play an important role in bone strength because they vary substantially across individuals and with aging, diseases, and treatments (20,101,191–194).

Age-Related Changes in Bone Microstructure

With age and progression of osteoporosis, trabecular arcades aligned with compressive and tensile stresses in the proximal are resorbed in an orderly fashion, with arcades of lessor stressors resorbed first and arcades of primary stressors becoming more prominent as thinner trabecular get resorbed (25,26) (Figure 1.6).

Three-dimensional quantification of age-related changes in trabecular architecture has historically required that trabecular bone biopsies be extracted and scanned via Micro-CT. From these studies, much has been learned in terms of the sex-specific differences in the manner and timing in which bone is lost. Generally speaking, at any age, men have higher measures of bone volume fraction and trabecular thickness and lower measures of trabecular separation compared to females in the femoral neck and greater trochanter (195). For females, there is a severe loss in bone volume fraction (BVF) that begins mid-life and continues throughout life in the femoral neck (12,195). For men, it has also been shown in most (10,11,24,196), but not all (12) studies, that trabecular bone microstructure deteriorates with age. However, changes are less extreme, occur later in life, and loss occurs predominately through trabecular thinning rather than loss of connectivity compared to females (10,11,24,196). Loss of trabecular connectivity negatively impacts bone strength to a greater extent than trabecular thinning and is likely another sex-specific difference in bone structure that results in an increased risk of fragility fractures in females (197).

Cortical bone, similar to trabecular bone, experiences bone degeneration with age for both sexes. Normal healthy cortical bone is highly heterogeneous, with large variations both in bone thickness and porosity depending on the relative distal-proximal and superior-inferior location of the femoral neck (Figure 5.3) (198,199). Bones in men tend to be larger in cross-sectional area and have thicker cortices compared to those in women (200). With age, cortical bone thins as porosity increases, with changes most pronounced in the superior region of the proximal femur (10,201). However, age-related cortical bone degeneration is more extreme in females compared to males (10,201). Regardless of sex, cortical bone degeneration is associated with increased hip fracture risk and is an important measure of bone quality that should be assessed to better understand whole bone strength (201–203).

Changes in bone microarchitecture coupled with the loss of bone mass are thought to be the primary reasons for decreases in femoral strength with age. As shown previously in both a cadaveric [CAD] and computational [COMP] model study, a severe decrease in bone strength occurs between the ages of 20 and 90 years (CAD: Loss of 70% [F] and 47% [M]; COMP: Loss of 55% [F] and 39% [M]) (39,98,204,205). [CAD] While BMD was a strong predictor of strength, strength decreased 40% faster than BMD and BMD could not account for sex-specific differences in bone strength, even when corrected for size (39). [COMP] Similarly, BMD could not fully predict age-related changes in bone strength because BMD only decreased by 26% in women and 21% in men over 20 to 90 years (98). Since BMD fails to fully explain age and sex-specific variations in the femoral strength, it is quite evident that changes in bone microstructure lead to a greater increase in hip fragility than what is predicted by BMD alone.

Proximal Femur Bone Microstructure and Whole Bone Strength

Despite known age-related changes in bone microarchitecture in the proximal femur, how bone microarchitecture directly impacts whole bone strength remains an elusive topic. This gap in knowledge is partially due to micro-CT technical limitations (i.e., small field size) which restricts the analysis of proximal femur microarchitecture to extracted sub-volumes (10–12). Nonetheless, computational studies have evaluated the influence of micromechanics on whole bone strength (206,207). Despite limitations (e.g., small sample size, lack of focus on sex-specific differences, and finite element based assumptions), computationally based findings are an important step in understanding the relationship between microarchitecture and whole bone strength. Findings are briefly described below.

- Both cortical and trabecular bone in the femoral neck play critical roles in the load transfer process when resisting fall-to-side failure (103,206,207).
- With age and bone loss, there is a marked decrease in trabecular micro-architectural "structural redundancy" which impacts whole bone strength (206).
- Only a small proportion of bone tissue (1.5% - 6.4%) needs to fail for structural failure to occur in the proximal femur (21,207).

To improve our understanding of sex-specific and age-related changes in whole bone strength, bone micro-architecture and its relationship to whole bone strength must be evaluated. Bone

micro-architecture cannot be directly measured in the clinic. However, understanding how bone architecture differs between sexes and impacts whole bone strength may help inform on decisions to improve diagnostic methods to identify those at greatest risk of hip fracture earlier in life.

Objectives

The objective of this study is to identify the microstructural traits (cortical and trabecular) in the femoral neck that best predict bone strength in males and females. We will determine if adding more cortical and trabecular architectural details improves strength predictions. To address this objective, analyses will be completed in three levels of refinement from least to most detail. As described above, the femoral neck plays a critical role in resisting structural failure in the proximal femur when loaded in a fall-to-side orientation. While BMD alone is a predictor of strength, BMD cannot fully account for age and sex-related variations in bone strength. By repeating analysis in three levels of refinement from least to most architectural detail, we aim to identify the level of architectural detail necessary to explain the age and sex-specific differences in bone strength. In the lowest refinement level, femoral neck total bone volume fraction and basic measures of external morphology (i.e., total area) will be quantified to capture information similar to what one would obtain from BMD. In level of refinement two, average cortical and trabecular architectural measures in the femoral neck will be quantified. In level of refinement three, cortical and trabecular microarchitecture will be quantified in specific sub-regions (i.e., superior, proximal, inferior, and distal) of the femoral neck.

We hypothesize that cortical and trabecular parameters both greatly contribute to the whole bone strength and that localized regions of the femoral neck, such as the inferior distal and superior proximal regions loosely aligned with the compressive and tensile arcades, better explain bone mechanical strength than average measures across the femoral neck. We will test for sex-specific differences in microstructure that contribute to bone strength in level of refinement three that are not resolvable at lower levels of refinement. Our objective is to close the gap in our current understanding of how bone microstructure relates to bone strength in an age- and sex-specific manner (36–40). We expect this work will provide an opportunity to re-

map diagnostic metrics from DXA or QCT images in a sex-specific manner to improve fracture risk predictions.

Methods

Sample population, Scanning, and Image Processing

Cadaveric femurs were collected for both sexes in the adult age range (Male [n=44], 18-95 years; Female [n=40], 24-95 years). Bones were scanned, loaded to failure in fall-to-the-side loading configuration, and bone strength was recorded as previously described (Chapter 3, pg. 47). Photographic images were taken of proximal femurs post-failure and two Orthopedic surgeons (F.F., M.H., Chapter 5 Acknowledgments) classified fracture type based on both the basic (femoral neck and trochanteric) and Müller AO (A1-A1/B1-B3) classification (208). Inter-observer agreement rates for basic and AO classifications were 1.00 and 0.96, respectively. Femoral neck volumes were extracted (previously described in Chapter 4, pg. 82) and bone background and cortical marrow separation were completed with the final trained FCNNs described in Chapter 4. Femoral neck bone/background and cortical/marrow segmentation were visually assessed in Dragonfly software 4.0 (ORS, Montreal, Canada). The FCNN bone/background performed poorly on one extracted femoral neck volume. As a result, six coronal 2-D slices were extracted from the volume, ground-truth images were manually created, and a new network was trained and applied to the volume in which the FCNN performed poorly. The cortical/marrow FCNN segmentation required minimal manual correction near the inferior aspect of the femoral neck where the cortical shell was most porous.

Cortical and Trabecular Microstructure Quantification

Measures of cortical and trabecular architectural traits were quantified using a custom written plug-in that is publicly available through the Infinite Toolbox in Dragonfly software 4.0 (ORS, Montreal, Canada, https://infinitetoolbox.theobjects.com/category/Plugins). Volumes in interest were defined and Bone volume fraction was determined by counting the number of bone voxels and normalizing the total number of voxel with these volumes. Average thickness measures were determined by averaging a 3-D volume thickness map which labeled each voxel of the VOI as

the diameter of the largest sphere that can fit in the VOI at that location. Moments of inertia were quantified relative to the cortical shell (Appendix A). Each femoral neck volume was padded by 200 slices on the proximal and distal boundary before analysis to remove the possibility of boundary condition errors for cortical bone (i.e., underestimation of bone thickness).

Cortical and Trabecular Microstructure Quantification

To assess how increased detail improves whole bone strength predictability, the analysis was completed in three different levels of refinement from least (1) to most (3) detail (Figure 5.4). The first analysis takes information only on the external size and the volume fraction of bone (i.e., few details); the second adds information on cortical and trabecular traits separately; the final adds regional information for cortical and trabecular volumes separately. Architectural and morphometric results used in each level of refinement to predict whole bone strength are listed below:

- Level 1: Bone volume fraction (BVF) of the entire neck, total cross-sectional area (Tt.Ar) (i.e., mean, minimum [min], and maximum [max]).
- Level 2: Tt.Ar [min], Cortical thickness (Ct.Th), cortical bone volume fraction (Ct.BVF) trabecular thickness (Tb.Th), trabecular bone volume fraction (Tb.BVF), cortical area (Ct.Ar), marrow area (Ma.Ar), and principal moments of inertia (I_{min}, I_{max}).
- Level 3: Cortical and marrow regions were segmented into superior-proximal [SP], superior-distal [SD], inferior-proximal [IP], inferior-distal [ID]) regions for a total of 8 sub-regions (4 cortical and 4 trabecular VOIs). The traits quantified for the cortical regions included Ct.Th and Ct.BVF. The traits quantified for the trabecular regions included Tb.Th and Tb.BVF. Tt.Ar [min] was included so that the impact of external morphometry was considered in the analysis.

Statistical Analysis

All analyses were completed in RStudio (166). Strength-age linear regressions were calculated for both sexes. Further, cadaveric femurs were sub-divided by fracture type to determine if there were significant differences in frequency between sexes (Chi-Square), in age, or the age-normalized strength (linear regression method (189)) for a given fracture type. Plots were created using the *ggplot2* package (209).

The analysis was completed in each level of refinement using the least absolute shrinkage and operator (LASSO) method from the *glmnet* package in RStudio (210). LASSO regression provides (1) high prediction accuracy, (2) can shrink and remove coefficients without a substantial increase of the bias, (3) can perform well on datasets with few observations and many features, and (4) can increase the model interpretability by eliminating irrelevant variables (211,212). The LASSO method was used to objectively identify the variables of greatest importance in predicting strength for males, females, and combined data-sets for the level of refinements one and two. For the level of refinement three, sex was included as a covariate for all models. Considering interaction with sex allowed us to identify regional volumes that predict bone strength for one sex but not the other.

LASSO minimizes the sum of squared errors by taking into account a tuning parameter λ. The tuning parameter is determined based on test data (a hold-out set) and is selected to minimize the sum of squared errors in the linear regression model. Technically, the tuning parameter can be any value from 0 to infinity. When λ is set to 0, no parameters are eliminated and the regression behaves as a linear regression. If λ equals ∞, all coefficients are eliminated. When λ is increased there is an increase in bias and when decreased there is an increase in variance, thus presenting a trade-off between bias and variance. Thus, the optimal λ varies by model and requires systematic testing, such as cross-validation, to ensure reproducibility.

Five-fold cross-validation was used to determine the λ value in all models in this study. Briefly, two-hundred λ values were tested (Range: 0.001- 100000, Sequence: 10^{ji}, $j_i = \sum_{-3}^{5} j_{i-1} + 0.04$) for each model to determine the λ value that resulted in the minimum mean cross-validated error (i.e., λ_{min}). In addition, the lambda value that results in the most regularized model (i.e., λ_{1se}) was determined. By definition, λ_{1se} is the λ value that lies within one standard error of the optimal value (i.e., λ_{min}). (Figure 5.6). In the context of this analysis, the λ_{min} model is the model with the smallest number of coefficients that are highly accurate (213). Thus, the final LASSO model of interest will be based on the λ_{1se}.To ensure reproducibility and convergence to consistent λ_{min} and λ_{1se} values, the five-fold cross-validation was repeated 1000 times and the median λ_{min} and λ_{1se} values were recorded. Beta (β) values for variables that were significant predictors of strength,

the percent (null) deviance of bone strength the model explained (% Dev) (i.e., a measure 1 – deviance ratio of the model divided by the null deviance, which is how well the response variable is predicted by the model the includes only the intercept), the sum squared of the residuals of the final model (SSR), and the R^2 values were reported for each model. As mentioned above, for practical purposes, the λ_{1se} model was considered the final reported model. It is generally recommended that the "one-standard-error rule" is used when selecting a model because it is the simplest model with accuracy comparable to the model that resulted in a minimum mean cross-validated error (213). Linear regression analysis against age was run on variables that were identified as significant predictors of strength in our final LASSO models. Sex-specific differences in the intercept and slope were also considered (ANCOVA).

Results

Basic Results: Strength, Age, and Fracture Type

Bone strength (i.e., maximum load) correlated negatively with age for females (R^2=0.170, p=0.004, Eq. Max Load[N]=-25.5*Age[years]+4843) and males (R^2=0.102, p=0.021, Eq. Max Load[N]=-29.7*Age[years] +7070) (Figure 5.5). When age was considered as an additional factor, age (Intercept: p< 0.001) but not the age-sex interaction (Slope: p=0.780) was significant. Male proximal femurs were approximately 70% (2227 N) stronger for a given age compared to females (Figure 5.5). Only a subset of our cadaveric cohort had height and weight data available in their medical history (F: n=20, M: n=35). However, based on this subset males in this sample were, on average, 25% heavier and 8% taller than females in this sample ([F] Weight: 150+/-45 lbs., Height: 65+/- 3 in.; [M] Weight: 187+/-49 lbs., Height: 70+/-3 in.; Unpaired t-test: Weight p<0.001, Height p=0.010). When categorized by fracture type (Basic Classification), female femurs experienced significantly more trochanteric (T) compared to femoral neck/cervical (C) fractures than male femurs (Chi-Square: p=0.013). Further, male cadaveric femurs that failed with a C-fracture were significantly older than those that failed with a T-fracture (C: 64+/-17 years, T: 53+/-20 years, Unpaired t-test: p=0.053). There was no difference in age between fracture types for females (C: 67+/-23, T: 64+/-21, Unpaired t-test: p=0.697). Strength values normalized for age did not differ between fracture groups for either sex (Unpaired t-test: [M] p=0.968, [F] p=0.093).

Level of Refinement One

An example plot of the λ_{min} and λ_{1se} chosen for our combined (male and female) model is shown in Figure 5.6. The number of variables, percent deviance the model explains, sum squared of the residuals, and the multiple R^2 for male, female, and combined data λ_{min} and λ_{1se} models are described (Table 5.2). With this level of structural refinement, only 45–58% of the percent deviance of strength was explained by BVF and measures of outer bone size (TtAr [min], TtAr [max]) in the final models. When separated by sex, three variables were identified as important predictors of strength for males (mean BVF, Tt.Ar [min] and Tt.Ar [max]). Unlike males, only one variable, mean BVF, was included in the final LASSO model for females. For the combined model with λ_{1se}, three variables, BVF, Tt.Ar [min], and sex, were best predictors of whole bone strength and the model explained 60.2% of the deviance in bone strength (Figure 5.7). Final unadjusted β values in the LASSO regression were also recorded (Table 5.3). Of all significant predictors of strength, only BVF significantly changed with age for males and females (M: R^2[adj.]=0.10, p=0.025, Eq. BVF=-1.04E-2*Age+0.394; F: R^2[adj.]=0.68, p=0.005, Eq. BVF=-1.29E-2*Age+0.401). There was no significant difference for the age-related BVF loss in the femoral neck between sexes (Slope: p=0.731; Intercept: p<0.001). Based on our findings in this cross-sectional study we would expect a male and female to lose 19% and 24% of femoral neck BVF, respectively, between the ages of 20-90 years.

Level of Refinement Two

The number of variables, the percent deviance of the model explained, sum squared of the residuals, and the multiple R^2 for male, female, and combined data λ_{min} and λ_{1se} models were described for the level of refinement two (Table 5.4). With this level of structural refinement, only 45–59% of the deviance in strength was explained by the model including trabecular and cortical BVF and measures of outer bone size (TtAr [max]). Briefly, when examining the simplified models (λ_{1se}), two variables were considered significant predictors of strength for both males and females (Ct.Ar, Tb.BVF). For the combined model, five variables, (Ct.Ar, Tb.BVF, Imax, Tt.Ar [min], and Sex), were best predictors of whole bone strength and explained 61.8 % of the deviance of bone strength was predicted by the model. Final unadjusted β values in the

LASSO regression were also quantified (Table 5.5). Of all significant predictors of strength, only Tb.BVF significantly decreased with age for both sexes (M: R^2[adj.]=0.19, p=0.002, Eq. Tb.BVF=-1.15*Age+0.212; F: R^2[adj.]=0.45, p<0.001, Eq. Tb.BVF=-1.01*Age+0.173). When including sex as a covariate and interaction term with age, sex (Intercept: p<0.001) but not the sex-age interaction (Slope: p=0.731) were significant predictors of strength. Based on our findings we would expect a male and female to lose 43% and 46% of Tb.BVF, respectively between the ages of 20-90 years. For the final variables included in our LASSO regression predicting bone strength, males had significantly higher measure of I_{max} ([M] 22717+/-7443, [F] 13648+/-3922, p<0.001), Ct.Ar ([M] 238+/-47, [F] 186+/-43, p<0.001), Tb.BVF ([M] 0.14+/-0.04, [F] 0.11+/-0.03, p<0.001), and Tt.Ar (see level of Refinement 1).

Level of Refinement Three

The number of variables, the percent deviance of the model explained, sum squared of the residuals, and the multiple R^2 for the λ_{min} and λ_{1se} models on the combined data including sex as an interaction term are described (Table 5.6). Only variables that are included in the final models for λ_{min} and λ_{1se} were presented (Table 5.7). Seven variables were considered significant predictors of strength in our combined data-set. The final model explained 64.4 % of the deviance in whole bone strength. Of the seven variables, four were local regional variables ([IP] Tb.BVF, [ID] Tb.BVF, [SP] Ct.Th, [SD] Tb.BVF), two were local regional variables with sex interactions ([IP] Ct.Th, [ID] Tb.BVF), and one was directly related to external morphometry (Tt.Ar [min]) (Table 5.7). Many variables included in our final model that were significant predictors of strength for males were also significant predictors of strength for females. However, the magnitude of the effect of Tb.BVF in the ID region and Ct.Th in IP region was a significant predictor of strength for males but not females, and thus allude to the need for more sex-specific models (Figure 5.8). When separated by sex, males had a significantly higher measure of Tb.BVF in the ID region (M: 0.123+/-0.054, F: 0.079+/-0.036; Unpaired t-test: p<0.001) but not for Ct.Th in IP region (M: 1.98+/-1.11, F: 1.68+/-0.70; Unpaired t-test: p=0.105). Finally, linear regression analysis comparing age to all cortical and trabecular sub-volumes was examined. All variables that significantly changed with age are indicated in bold (Table 5.8). Tb.BVF significantly changed with age in the ID, SP, and SD regions with age for both sexes. When sex was considered as an additional factor in age-related changes in Tb.BVF, sex (Intercept: [ID] p< 0.001, [SP]

p=0.009), [SD] p=0.002) but not the age-sex interaction (Slope: [ID] p=0 0.363, [SP] p=0.967, [SD] p=0.883) were significant. These findings indicate that at any given age males have a higher measure of Tb.BVF, but there was no sex-specific difference in the rate at which trabecular bone waslost.

Discussion

The objective of this study was to identify the microstructural traits in the femoral neck that best predict bone strength in males and females. We hypothesized that cortical and trabecular parameters both greatly contribute to whole bone strength and that localized regions better explain bone strength than average measures across the femoral neck. Our findings support this hypothesis because we found that both cortical and trabecular architectural measures together were the best predictors of whole bone strength in our combined models. In addition, th percent deviance in bone strength that the final models explained improved modestly (level 1: 60.2%, 2: 61.8%, and 3: 64.4%) as more details were added to each level of refinement. In the most detailed level of refinement, two regional variables with sex-interactions were included in the final model, demonstrating that additional variables were necessary to predict bone strength for males but not females. These findings show how age-related and sex-specific local differences in micro-architecture impact bone strength and provide insight into new ways to improve fracture risk diagnostics.

Age, Bone Strength, and Sex

We found significant negative correlations between whole bone strength and age for both sexes. There was no sex-specific difference in the rate of whole bone strength loss with age, but males were 2227 N (~70%) stronger than age-matched females. In a subset of our data, we found that males were only, 25% heavier and 8% taller than females Thus, differences in stature could not fully explain differences in bone strength which were 70% higher in the male compared to female femurs. On the other hand, females lose approximately 1785 N (~55 N) in bone strength with age (20-90 years). Meaning, the magnitude of the difference in bone strength at any given age for females compared to males is ~ 500 N greater than the projected strength loss throughout life from normal aging. Determining the structural (e.g., size) and

architectural (e.g., local trabecular micro-architecture, cortical thickness, etc.) properties of bone that makes males stronger than females at baseline may provide new treatment methods to decrease fracture risk. Although our study was cross-sectional in study design, if the magnitude of strength loss is the same, regardless of sex, then baseline measures of bone strength may play a more critical role than what has been previously established. Our findings are in line with a recent cadaveric study but differ from computational studies who found that there was a sex-specific difference in the rate of strength-loss with age (98,214,215). Contrasting results may be due to differences in the analysis (direct mechanical testing vs. strength estimates), study type (cross-sectional vs. longitudinal), and/or population (cadaveric vs. patient). Regardless of differences in the study designs, these findings elucidate and confirm a substantial sex-specific difference in bone composition as related to bone strength, and suggest that there should be a shift in research focus to identify the factors that allow an individual to reach optimal bone mass early in life, as opposed to identifying how to regain mass once appreciable bone loss has already occurred.

Fracture Type

When categorized by fracture type, female proximal femurs failed more often in the intertrochanteric region compared to male femurs, whereas, cervical fractures were more common in the male proximal femurs. Age was not significantly associated with fracture type in the female femur. Clinical studies have shown that women who suffer trochanteric fractures have more generalized bone loss and are older compared to those who suffer a cervical fracture (216,217). While we did not find the fracture type age-association reported previously for females, results showing trochanteric fractures are more common in females (i.e., weaker and more highly associated with the microstructural bone loss with age) than males is in line with prior interpretations of fracture prevalence in female-specific studies (216,217). Differences in this study compared to clinical findings may be due to the nature of cadaveric studies which limits analysis to smaller sample sizes and controlled loading conditions.

However, our findings agree with prior clinical studies showing an increased prevalence of cervical fractures in males with age (218). It has been suggested that the mechanism of cervical

and trochanteric fractures is inherently different and patients should be treated based on the risk of fracture type (219). While we agree that patient-specific diagnostic and treatment methods are necessary to improve fracture risk, we also qualitatively observed that cervical and trochanteric fractures' initial point of failure often occurs at the same location (Figure 5.9). Perhaps, instead of labeling failure based on clinical fracture type, future studies should label failure type based on the initial points of failure by Nano-CT scanning the femur post-failure and investigate bone microstructure at the fracture site.

Level of Refinement One

The final LASSO regression models included three variables for males and one variable, femoral neck BVF, for females. These models explained 33.3% and ~50% of the proportion of deviance of bone strength data for males and females, respectively. While males had a significantly higher femoral neck bone volume fraction at any given age compared to females, the rate of bone loss with age did not differ between sexes. Besides significant sex-specific differences in the physical size and baseline levels of bone volume fraction, there is no obvious explanation as to why one variable explains so much more of the deviance in the data for female but not male proximal femurs data (110). Regardless, femoral neck BVF alone predicts a substantial proportion of variability for females only, and more detailed analysis (e.g., quantification of architectural traits, more localized analysis) is necessary to improve predictability, particularly for males.

We determined that the percent strength loss (44% males, 70% females) was much greater than percent femoral neck BVF loss (19% males, 24% females) that occurs with age, implying that loss in BVF alone cannot explain observed drops in whole bone strength in this study. These findings are in line with previous BMD based studies showing a loss in bone strength throughout life was much larger than observed BMD changes (98,215). While not quantified, the aim of this level of refinement was to present results "similar" to what would be obtained from BMD (i.e., average measures across the femoral neck that did not require knowledge in the underlying microstructure). Our findings, at the very least, are similar to what was reported previously, suggesting that more information is required to assess fracture risk in males, in

addition to external size. However, studies focused on sex-specific differences in fracture risk based on BMD generally show male and female fracture risk is equivalent at the same absolute BMD value (220–223).

Despite finding sex-specific differences in all parameters in our final model (Tt.Ar[min]) and BVF), sex was still considered a significant predictor of bone strength. Thus, we are unable to fully tease out sex-specific differences in bone strength and more detailed analysis is required. However, the results of this study suggest that rate of femoral neck bone loss (measured using BVF) in high-resolution scans cannot explain sex-specific differences in bone strength or strength changes with age throughout life, further demonstrating the need for consideration of bone architecture.

Level of Refinement Two

We observed a modest increase in percent deviance in whole bone strength that our final sex-combined model (60.2% vs. 61.8%) explained. Such a modest increase in percent deviance explained was surprising given that additional important predictors of bone strength (i.e., trabecular and cortical thickness) were included in our analysis (197,198). However, it has been previously shown that structural failure results from a small proportion of bone tissue (1.5% - 6.4%) in the proximal femur (21,207). Results reported here were average cortical and trabecular results of the entire femoral neck and thus may not fully represent the critical fracture point or local region, that best predicts whole bone strength. Another potential reason as to why such minimal improvements in fracture risk were observed may be due to the fact that some individuals have different capacities to maintain bone throughout life (i.e., “fast losers” and “slow losers”) (224). If different mechanisms of bone loss exist, then bone parameters most highly predictive of strength may differ by sub-group. As this is a cross-sectional cadaveric study, it is not possible to group our data by rate of bone loss. However, we have shown in a prior study that external bone size may be one factor that gives insight into diverging mechanisms into which bone is maintained and lost throughout life (37,225). Thus, while our final model explains over 60% of the deviance of whole bone strength, much work is still needed to determine if, what, and why different mechanisms of bone loss exist (i.e.,

accurately quantify architectural details in longitudinal studies with accurate bone strength estimates).

Our final model included five variables Ct.Ar, Tb.BVF, I_{max}, external morphometry (Tt.Ar [min]), and sex. Similar to level of refinement one, males had significantly higher measures of bone architecture compared to females (110). However sex-specific differences in bone strength cannot be attributed solely to the included variables as sex was also selected for inclusion. Compared to the 70% loss in bone strength for females and 44% strength loss in males, we observed a loss of 46% and 42% in trabecular BVF for females and males, respectively from 20-90 years. The rate of bone loss was similar to the rate of the strength lost throughout life for males but not females in our linear regression analysis. Thus, while Tb.BVF is not the only important variable to consider for males, changes in this compartment are of the same order of magnitude of the strength lost throughout life. If it were possible to separate cortical from trabecular bone in clinical DEXA scans, BMD measures based exclusively on trabecular area may allow for improved fracture risk predictability for males. However, for females, trabecular BVF is lost at a slower rate than strength throughout life implying that a sub-analysis needs to be completed to tease out which variables change throughout life and impact whole bone strength.

For both male and female-specific models, the same two variables, Ct.Ar and Tb.BVF were predictors of whole bone strength. Including two variables, one cortical and one trabecular bone variable, is important and in line with prior findings, revealing both cortical and trabecular bone share load when resisting fall-to-side failure (103). When comparing sex-specific λ_{1se} models, we found that there was a drop in the percent deviance the model explains of strength from level of refinement one (F: 50%, M: 44%) to two (F:49%, M:29%). However, there was an increase in percent deviance explained for sex-specific λ_{min} models from level of refinement 1 to 2. By definition, λ_{1se} is the λ value that lies within one standard error of the optimal value (i.e., λ_{min}) Clearly, the nature of selecting the λ_{min} value in this case resulted in over-conservative sex-specific final models in terms of the number of variables included (212). Other variables not assessed here (e.g., collagen cross-linking and bone mineralization, external bone size, local bone microstructure, etc.) impact bone strength (20,111,112). While not

assessed, it is possible that a few bones in our cohort have abnormally high/low mineralization levels or differences in local bone microarchitecture which largely impacted bone strength. Regardless of the strength predictability, our findings show a need for region-specific analysis in order to better predict bone strength.

Level of refinement Three

In level of refinement three, we found that the final model explained about 64.4% of the deviance in whole bone strength. This final model had modest improvements in bone strength predictability compared to prior combined models (Level of Refinement 1: 60.2% vs. Level of Refinement 2: 61.8%). Such a modest increase in the percent deviance of bone strength explained, was again, surprising given that local measures of bone microarchitecture, loosely aligned with the compressive and tensile arcade, were considered. However, while we did not see large changes in deviance the model explained, we found differences in the set of variables that best predicted bone strength for males and females. Sex-specific differences included in the variables observed here demonstrate the need for localized measures of architectural analysis to accurately identify sex-specific differences in bone strength.

Five variables with no gender-interaction were used in the final LASSO regression model to predict bone strength. Independent predictors included three trabecular BVF (IP, ID, SD) variables, one cortical thickness (SP) variable, and total area (Tt.Ar [min]). Two of three trabecular regions (IP and ID), loosely aligned with the primary compressive arcade, were important predictors of bone strength. These regions are predominately loaded in tension, a loading condition where bone is considerably weaker than in compression (226). Unsurprisingly, this is also the region where we qualitatively observed most initial fracture failures to occur (21,207). These findings elucidate the importance of the compressive arcade, an arcade that has been shown to lose less bone than the tensile arcade throughout life, as a key region needed to understand whole bone strength. Trabecular BVF in the SD region of the femoral neck is also an important predictor of bone strength. The SD region is loosely aligned with the tensile arcade, a region of trabecular bone that experiences more prominent loss earlier in life and with the progression of osteoporosis (25,26). Historically, this arcade has been

suggested as a key location for assessing osteoporosis progression (25,26). While our study is in agreement with prior findings showing that there is substantial bone loss throughout life in the tensile arcade, our findings also point to the need for further analysis of the compressive arcade to better understand why bones are strong or weak.

Interestingly, there was no observed loss in bone volume fraction with age in either the IP or ID trabecular regions of the femoral neck. However, there was a significant decrease in the SD region for both sexes. These findings demonstrate that many of the properties important to predict whole bone strength may be an inherent property of an individual (or a property that does not change with age) that naturally makes some better adapted to resist loads than others. In all trabecular regions, males had a significantly higher BVF value with age compared to females, but there was no sex-specific difference in the rate of bone loss in any region. Clearly, regardless of sex, trabecular bone, particularly in the compressive arcade, plays an important role in bone strength. These findings warrant future arcade-focused studies in order to better identify how bone resists failure when loaded in the fall-to-side orientation.

Cortical thickness in the SP region of the femoral neck was included in our final model. While cortical thickness did not change with age, cortical bone in the SP region of the femoral neck is a key location of compressive force when loaded to failure in fall-to-side orientation (207). It has been shown that the more prominent the bone loss, the more extreme and localized the compressive force on the superior aspect of the femoral neck (207). These findings indicate that the regions with the most extreme tensile and compressive loading in the femoral neck are also the most important variables included in our model to predict whole bone strength. Finally, Tt.Ar [min] was also included in this model, confirming that overall size of the bone plays an important role in predicting whole bone strength (33). These findings confirm that many of the important variables used to predict bone strength were properties that did not change with age and naturally made some bones better adapted to resist loads than others.

We determined that two variables had a sex-specific interaction in predicting bone strength for males. These two regions for analysis included cortical mean thickness (IP) and trabecular BVF (ID) for males. Trabecular BVF was also a significant independent predictor of strength

without interaction effects, demonstrating that not only is this an important site for both sexes, but the magnitude of effect on bone strength differs in a sex-specific manner. Again, as previously mentioned, this is a region where there is extreme tensile loading and is the location where many fractures occurred in this study. Future studies while focus on quantifying architectural properties in this particular region. The cortical thickness variable in the IP region is a location where tensile loading occurs. While not assessed, we hypothesize that due to the fact that males have significantly more bone than females at a given age, the load is being more effectively distributed across the male femur compared to the female femur. This would, in part, explain why more male-specific variables are included in our model and why the cortical IP region is included for males but not females in this study. Others have proposed that structural redundancy is a key factor to consider when predicting whole bone strength (21). Future studies will focus on the ID region of the compressive arcade and how this particular region plays such an important role in predicting bone strength in this study.

Limitations and future work

This study has many advantages: a large sample set of cadaveric bones for both sexes, mechanical testing results measured directly, and scans of the entire proximal femur at a resolution able to quantify bone microarchitecture. To our knowledge, there are no other datasets of cadaveric proximal femurs scanned at the high resolution (27 um) used here. Despite this, there are a number of limitations that need to be addressed. First off, cadaveric studies are limited in that bones can only be loaded to failure in one loading condition/orientation and medical history is limited, as discussed previously (Chapter 3, pg. 50). Further, the in-vivo initial impact velocity from a fall to the side may be up to 35 times higher than the load in which bone was fractured in this study (125). It is well established that bone sustains higher forces at higher loading rate (105). However, McElhaney observed only a 12% change in compressive strength when there was a 300-fold increase in load, suggesting that the mechanical test results observed here may not significantly differ from what would be observed in a clinical fracture (126). Also, while our bone segmentation is consistent and accurate in this study (DICE=0.96), there may be inherent bias because all ground-truth images

in the FCNN training and validation, set were created by a single person, as previously described (Chapter 4, pg. 77).

In addition, the femoral neck was manually extracted using anatomical landmarks. While shown to be a reproducible protocol, a preferable method for femur femoral neck extraction would be through the use of statistical shape modeling, where regions could be extracted in an entirely automated manner and not subject to operator error (227). In addition, generally speaking, statistical shape modeling may help elucidate sex-specific differences in the entire proximal femur structure, beyond the femoral neck region, and explain bone strength variability in our cohort (227). Further, the femoral neck region was extracted as a representative volume of the femoral neck BMD-based DEXA measures. Due to the nature of the extraction protocol, femoral neck volume varied from sample to sample and was generally larger than the ~1.5 cm wide region used for DEXA based BMD measures. However, we believe that our volume better captured the regions where proximal femurs in our study typically broke (the inferior portion of the lesser trochanter) than the DEXA based BMD volume. Finally, no measures of bone mineral make-up or bone orientation (i.e., anisotropy) were quantified, despite the fact that the material composition plays a large role in bone mechanical properties (e.g., collagen, mineralization) (20,111,112). Despite these limitations, we feel that this body of work is unique and adds a major contribution to the literature. Future work will focus on using statistical shape modeling, three-dimensional volumetric mapping, and trabecular spacing to better evaluate strength variability observed herein.

Conclusions

The objective of this study was to identify the microstructural traits in the femoral neck that best predict bone strength in males and females. Surprisingly, we found that the most basic measure in our analysis (level of refinement 1) predicted whole bone strength, as assessed using the percent deviance the model explaines, nearly as well as all other more detailed iterations. However, at the most detailed level of refinement, we found novel sex-specific differences in regions used to predict whole bone strength for males. Further, our study determined that regions highly associated with the compressive arcade were important

variables to include for predicting whole bone strength for both sexes. These studies allude to the need for more region-specific analysis, to better understand whole bone strength in the future.

This work required the use of a Nano-CT system, something that due to the small scan volume, radiation dosage, and time required for scanning, will never be used in a clinical environment. Despite this, these findings may provide important insight that may lead to new and improved diagnostic techniques (e.g. femoral neck BMD determined on a local, rather than global, basis) or machine learning techniques to extract high-level features from lower resolution DEXA scans. The feasibility of this has yet to be fully explored in the literature. However, we believe that the findings described here will allow us to explore age and strength related changes in a new and novel way, providing new insight into ways to improve fracture risk diagnostics.

Acknowledgments

I would like to gratefully acknowledge funding from the National Institutes of Health: NIH/NIAMS - AR070903 [SHS], AR065424 [KJJ], AR069620 [KJJ], and AR06424 [TLB]. The content presented in this chapter does not necessarily represent the official views of the National Institutes of Health. In addition, I would like to thank the two orthopedic surgeons, Fred T. Finney and Mark E. Hake, collaborators through the University of Michigan Department of Orthopedic Surgery, who donated their time and expertise to label our fractured femurs by clinical fracture type. Further, I would like to thank Todd L. Bredbenner, who donated a significant amount of time to provide mentorship regarding femoral neck extraction for analysis and the statistical analysis used in this chapter. Finally, I would like to thank Rob Goulet, who was critical in helping to create the code used for the sub-regional analysis used in this chapter.

Figures

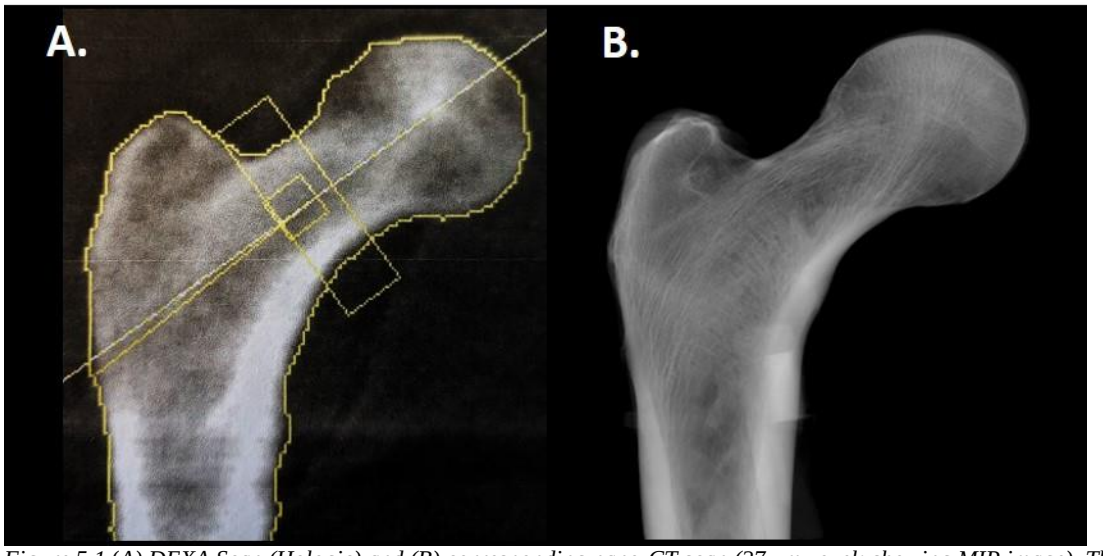


Figure 5.1 (A) DEXA Scan (Hologic) and (B) corresponding nano-CT scan (27 um voxel; showing MIP image). The DEXA scan, while insightful, cannot account for 3-D trabecular orientation and thickness variability visible in the Nano-CT scan (27 μm). Courtesy of Karl Jepsen and Jacob Applegate.

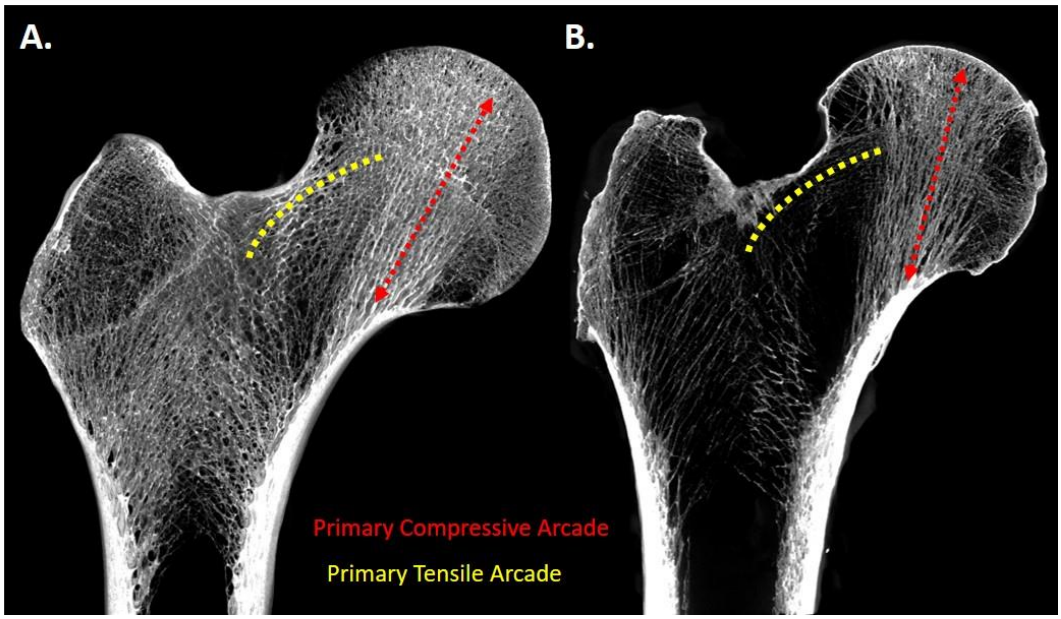


Figure 5.2 Nano-CT scans of a (A) 27-year-old and (B) a 91-year-old female. With age the primary compressive arcade (red) remains intact while the primary tensile arcade (yellow) is almost entirely resorbed.

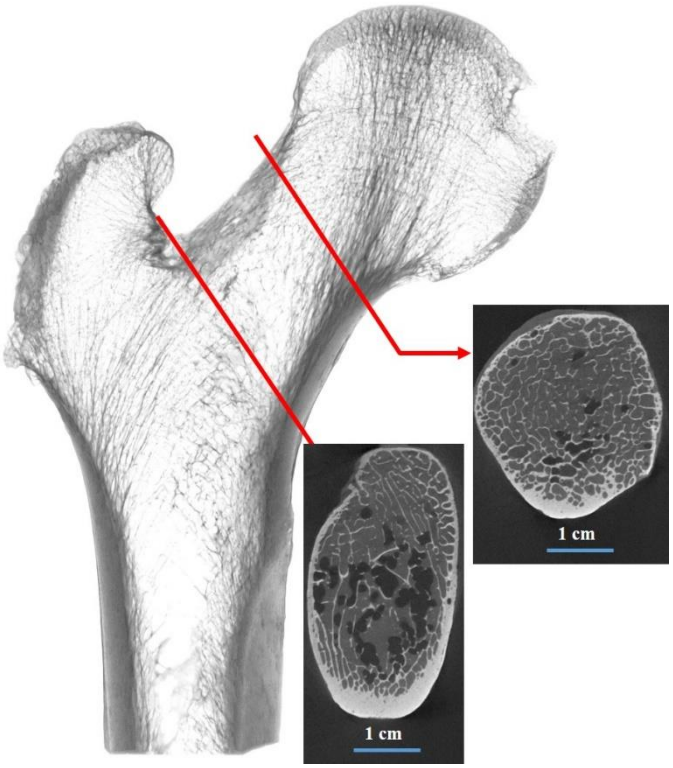


Figure 5.3 Two coronal cross-sections of the distal (left) and proximal (right) region of the femoral neck, demonstrating the high level of cortical heterogeneity in a 77-year-old female.

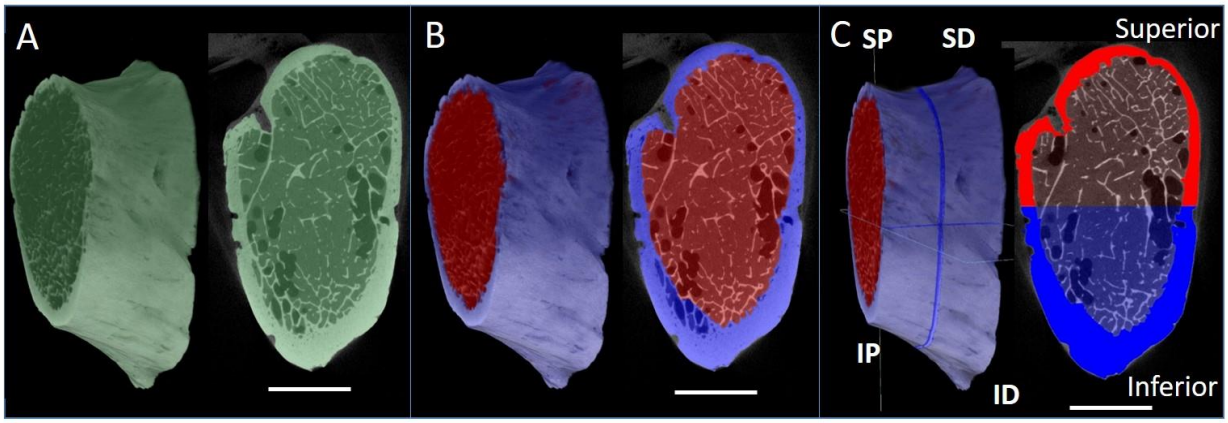


Figure 5.4 Visualization of the three levels of refinement from one (left) to three (right). Results were quantified for the (A) entire femoral neck VOI (green) in level one, for the (B) average cortical (blue) and trabecular (red) volumes in the entire femoral neck volume in level two, and for (C) cortical and trabecular SP, SD, IP, and ID volume quadrants in the femoral neck in level three.

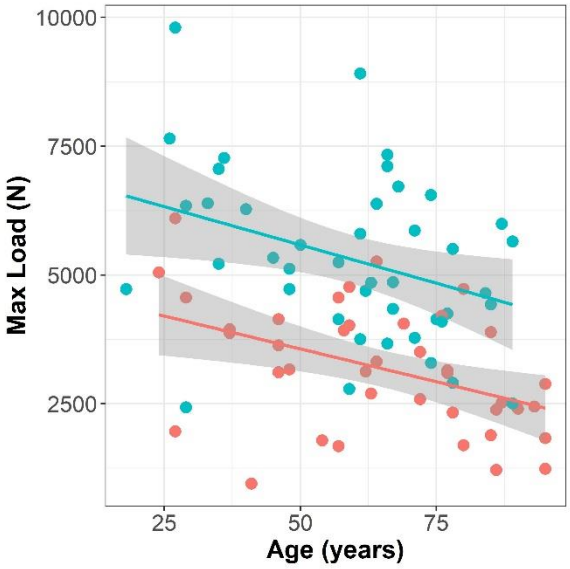


Figure 5.5 Linear regressions of significant age-related changes in maximum load for males (blue) and females (red).

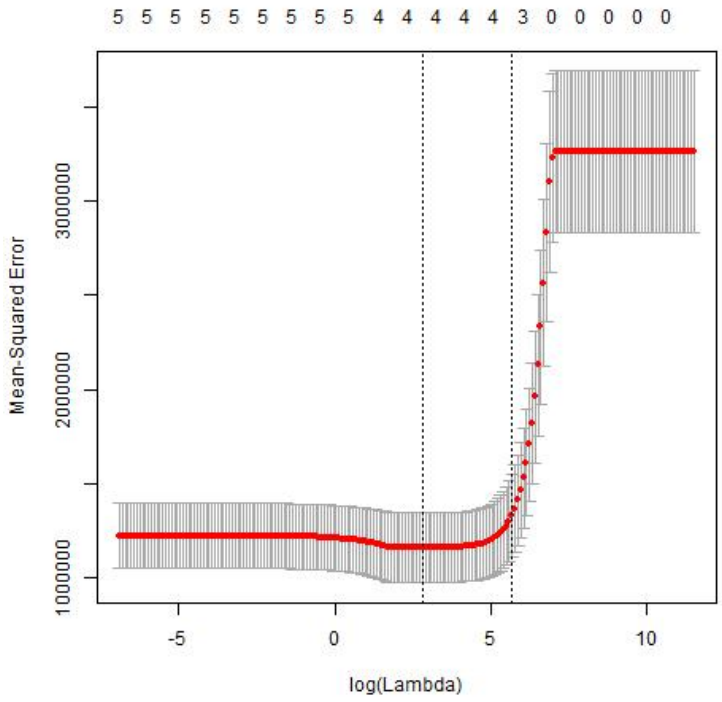


Figure 5.6 Example plot of the log(λ) versus the model mean square error. The λmin is indicated by the dotted vertical line on the left and λ1se value is indicated by the right-most vertical dotted line. The number of variables included in the final model for a corresponding λ value are listed above the plot.

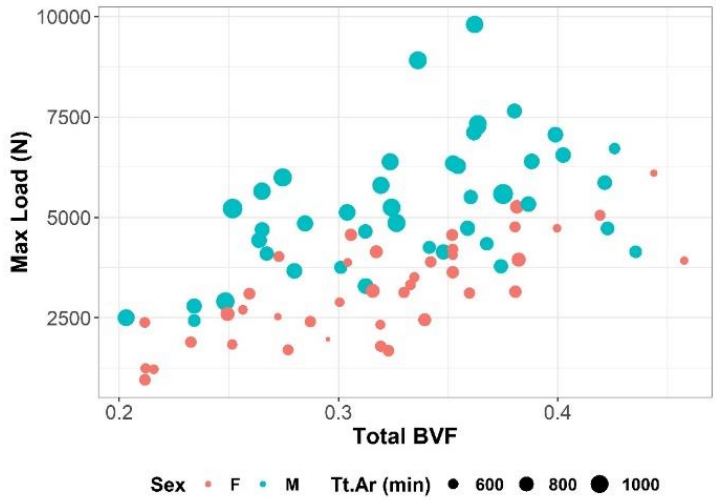


Figure 5.7 A plot including the three variables (sex [color], min. Tt.Ar [point size], and mean BVF [X-axis]), relative to maximum load. Variables presented were selected as the most important variables in the combined model based on λ_{1se} as described in the methods

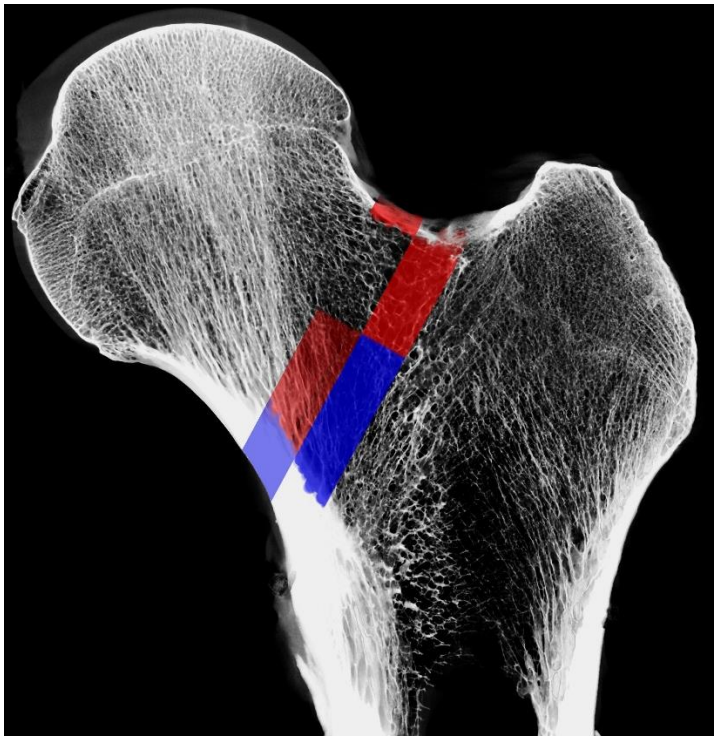


Figure 5.8 Volumes included in the final level of refinement for both sexes (yellow) and for significant interactons with males (blue) for the cortical [IP] volume and the [ID] region, which is significant independent predictor in the model.

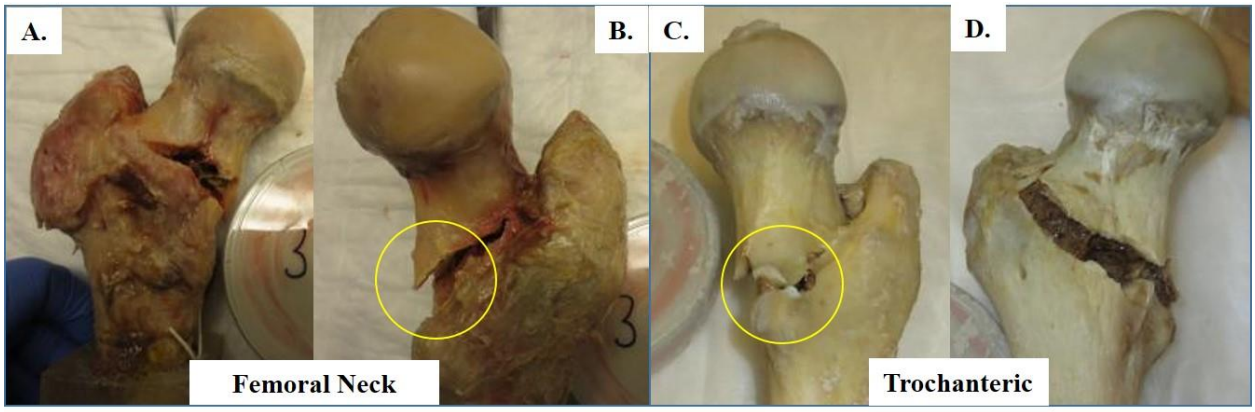


Figure 5.9 Views of a femoral neck (A -B) and trochanteric (C-D) fracture with the same initial point of failure (circled in yellow).

Tables

Table 5.1 Summary of significant differences in anatomy surrounding the proximal femur for males and females.

Property	Definition	Sex-Specific Difference	Citations
Proximal femur head size	The radius size of the femoral head.	Males have a significantly larger femoral head than females which results in ~30% larger surface area for males.	(181,228)
Proximal femur Valgus or oblique angle (bicondylar)	A measure of how the femur angles medially from the hip to the knee. Defined as the angle between an axis through the shaft of the femur and a line perpendicular to the intracondylar plane.	Males have a significantly lower bicondylar angle.	(182,229)
Femoral Version (angle of torsion)	A measure of the twist between the proximal and distal ends of the femoral diaphysis.	Overall, females display a larger femoral angle of version than males.	(182)
Surrounding muscles that are essential for hip function	Muscle Attachments: The gluteus medius, tensor fasciae latae and gluteus maximus	Males have significantly larger muscle volumes compared to women when normalized by weight	(230)
Acetabulum - Abnormal Pathology	When the socket for the proximal femur head has either under-coverage (acetabular dysplasia) or over-coverage (pincer femoro-acetabular impingement [FAI])	While a wide range of this variability is considered normal, both under- and over-coverage is more common in females than males	(181)

Table 5.2 The number of variables (DF), the median λ_{min} and λ_{1se} values, the percent deviance the model explains (%Dev), the sum of squared residuals (SSR), and the multiple R^2 are listed for all of the models (males, females, and combined) in level of refinement one.

Type	Male		Female		Combined	
	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}
DF	3	3	1	1	4	3
λ	50.526	352.971	116.232	352.971	28.994	321.764
%Dev	48.6%	33.3%	57.1%	49.5%	66.9%	60.2%
SSR	5.69E+07	7.39E+07	2.51E+07	2.96E+07	8.65E+07	1.04E+08
RSQ	0.489	0.445	0.580	0.580	0.670	0.669

Table 5.3 Significant variables included in the final LASSO model are listed for male, female, and for combined datasets in level of refinement one.

Type	Male		Female		Combined	
	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}
(Intercept)	-6554	219	-949	270	-5571	-2506
BVF	17959	9541	13050	9213	15692	10855
Tt.Ar [mean]	-	0.24	-	-	-	-
Tt.Ar [max]	2.13	-	-	-	0.39	-
Tt.Ar [min]	3.83	1.95	-	-	4.60	3.42
Sex	N/A	N/A	N/A	N/A	588	465

Table 5.4 The number of variables (DF), the median λ_{min} and λ_{1se} values, the percent deviance the model explains (%Dev), the sum of squared residuals (SSR), and the multiple R^2 are listed for all of the models (males, females, and combined) in level of refinement two.

Type	Male		Female		Combined	
	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}
DF	7	2	6	2	6	5
Lambda	88	511	73	322	61	387
%Dev	57.6%	28.9%	62.7%	49.1%	70.6%	61.8%
SSR	4.69E+07	7.87E+07	2.18E+07	2.98E+07	7.68E+07	9.99E+07
RSQ	0.588	0.446	0.634	0.593	0.709	0.696

Table 5.5 Unadjusted β values for variables included in the final LASSO model for male, female, and for combined datasets in level of refinement two.

Type	Male		Female		Combined	
	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}
(Intercept)	836	3329	-355	860	-1053	18
Ct.BVF	-217	-	419	-	-	-
Ct.Th	-	-	47	-	-	-
Ct.Ar	6.73	3.50	8.78	5.15	7.85	6.00
Tb.BVF	22277	8037	16753	12730	20220	13483
Tb.Th	-11011	-	3141	-	-4606	-
Ma.Ar	-	-	-1.95	-	-	-
I_{min}	-	-	-	-	-	-
I_{max}	7.94E-03	-	0.00E+00	-	1.92E-02	1.29E-02
J_{polar}	4.64E-05	-	-	-	-	-
Tt.Ar [min]	2.48	-	-	-	1.59	0.95
Sex	N/A	N/A	N/A	N/A	383	228

Table 5.6 The number of variables (DF), the median λ_{min} and λ_{1se} values, the percent deviance in bone strength the model explains (%Dev), the sum of squared residuals (SSR), and the multiple R^2 are listed for the combined models with sex as an interaction in level of refinement three.

Type	Combined	
	λ_{min}	λ_{1se}
DF	12	7
Lambda	61	293
%Dev	74.1%	64.4%
SSR	6.78E+07	9.31E+07
RSQ	0.745	0.693

Table 5.7 Beta values for variables included in the final LASSO model for male, female, and for combined datasets in level of refinement three.

Region	Bone Type	Type	Combined	
			ʎ min	**ʎ 1se**
		(Intercept)	1183	576
IP	Cortical	BVF	1444	-
		Thickness	196	-
		Thickness (interaction)	-	87
	Trabecular	BVF	6426	2745
ID	Trabecular	BVF	3719	567
		BVF (interaction)	1279	4085
		Thickness	-4636	-
SP	Cortical	Thickness	757	474
	Trabecular	Thickness	-3067	-
SD	Cortical	BVF	-2276	-
	Trabecular	BVF	9019	7898
Outer Area		Tt.Ar [min]	2.22	2.12
		Tt.Ar [min] (Interaction)	0.639	-

Table 5.8 Linear regression analysis for all sub-volumes in which a variable significantly changed with age for at least one sex. Variables that significantly changed with age are indicated in bold.

Region	Variable	Sex	Age (Slope)	Intercept	R^2 [Adj.]	p
IP	Ct.BVF	**F**	**-1.41E-03**	**1.02E+00**	**0.43**	**< 0.001**
		M	3.50E-04	9.14E-01	0.02	0.172
ID	Ct.BVF	**F**	**-9.14E-04**	**9.64E-01**	**0.10**	**0.026**
		M	2.05E-04	9.12E-01	-0.01	0.511
	Tb.BVF	**F**	**-8.36E-04**	**1.34E-01**	**0.21**	**0.002**
		M	**-1.26E-03**	**1.98E-01**	**0.18**	**0.003**
SP	Ct.BVF	**F**	**-1.04E-03**	**9.36E-01**	**0.16**	**0.006**
		M	-4.60E-05	8.63E-01	-0.02	0.909
	Tb.BVF	**F**	**-1.08E-03**	**1.79E-01**	**0.37**	**< 0.001**
		M	**-1.10E-03**	**2.03E-01**	**0.16**	**0.005**
SD	Ct.BVF	**F**	**-8.47E-04**	**9.54E-01**	**0.17**	**0.005**
		M	2.92E-04	8.64E-01	-0.01	0.497
	Tb.BVF	**F**	**-1.09E-03**	**1.72E-01**	**0.35**	**< 0.001**
		M	**-1.15E-03**	**2.04E-01**	**0.17**	**0.004**
	Tb.Th	F	-3.01E-04	2.12E-01	0.04	0.111
		M	**-5.89E-04**	**2.41E-01**	**0.09**	**0.028**

Chapter 6 Discussion Closing Remarks

This thesis aimed to characterize bone degeneration at either end (i.e., the distal and proximal metaphysis) of the femur. Our findings were novel, demonstrating that bone changes occur immediately following ACL injury in young females at the exact location where the bone explant is removed and new ACL is placed during reconstructive surgery. Also, we found sex-specific differences in regional microstructural variables in the proximal femur are needed to best predict whole bone strength. Clinically, these results demonstrate that bone microarchitecture should be an important factor to consider in ACL reconstructive surgery and that regional analysis in the femoral neck may allow us to better identify individuals at the greatest risk of fracture in an improved, sex-specific manner. The remainder of this chapter will briefly summarize our findings and recommend future directions of study.

In chapter two, we determined that bone quality greatly differed on a patient-by-patient basis undergoing ACL reconstructive surgery. Differences in microarchitecture could not be explained by time from injury to operation (i.e. time of disuse), age, or activity the patient was participating in when ACL injury occurred. Thus, differences in bone quality may be due to variability present at baseline, in response to injury, or activity an individual participates in after injury. Clinically, these findings are important because we are the first to show that bone quality varies across patient groups and points out that this may be an important factor to consider in the context of ACL injury risk and long term surgical outcomes. For example, the reason behind bone tunnel enlargement following ACL surgical repair is still largely unknown, but we established that differences in bone quality may be a plausible explanation.

Many questions arose from the findings in this study. How much did physical activity vary in patients post ACL injury and before surgery? Is the same variability observed in males? Is bone loss driven by inflammation, disuse, or some combination of both? Future work should consider including both a questionnaire and the clinical MRI scan confirming ACL injury in

addition to the patient explant. A questionnaire with details regarding self-reported pain, immobility, and daily activities before surgery would be helpful to address questions in the context of our current findings. A clinical MRI would be useful to identify if macroscale, in addition to micro-scale, changes in bone morphology are evident. Repeating our analysis with male patient explants would be helpful to determine if sex-specific differences in bone quality exist. Microstructural changes occurred predominately in the cortical shell, which was surprising since trabecular bone generally has a higher remodeling rate. Thus, the bone loss observed may be a result of an acute inflammatory response. Future work should quantify osteophyte mass in explant scans and measure inflammatory markers in the blood serum, if possible, to address this hypothesis. However, some of our questions (i.e., genetic differences, baseline levels of bone quality, mechanism of ACL failure, etc.) will not be possible to tease out from our patient cohort. Future work will require animal models wherein age, genetic make-up, and ACL failure mechanism are controlled to fully understand the changes that occur in bone microstructure following injury.

Findings from our cadaveric proximal femur studies (Ch. 3 - 5) provide novel insight for future research. We first determined that the magnitude of sex-specific differences in bone strength was greater than age-related strength loss endured throughout life. Further, there was no sex-specific difference in the rate of strength or bone loss. Clinically, these findings demonstrate that if females could maximize bone quality early in life, they may be able to maintain the structural strength later on, even with bone loss, to mitigate fragility fractures altogether. Further, mechanical variables (i.e., stiffness and post-yield-displacement) and demographic data (i.e., age and sex) did not fully explain variability in whole bone strength. Femoral neck analysis improved our ability to predict whole bone strength but demonstrated that increased level of regional microstructural detail only modestly improved strength predictability. Despite this, our findings demonstrate that increased regional level of micro-architectural detail is needed to identify sex-specific differences in whole bone strength that could not be identified in lower levels of refinement. Clinically, these findings demonstrate that regional analysis on DEXA scans may be useful for identifying those at greatest risk of fracture earlier in life and in a sex-specific manner.

Based on these findings, there are a few, immediate, studies that would largely increase our current understanding of bone biology and whole bone strength using our cadaveric dataset. We found that BVF in the trabecular inferior-distal region was included in all of our strength-predictive models and proximal femur failure qualitatively began and coalesced from this same region. Thus, future work should focus on quantifying bone microstructure (loosely aligned with the compressive arcade) to identify why this is a point where failure typically occurs. Further, we observed large variations in bone macrostructure in terms of neck length, shaft angle, and femoral head size. All of the aforementioned variables play a role in the normal loading on the femur and likely result in small but important differences in the alignment of the compressive and tensile arcades. As the arcades make-up a large proportion of trabecular bone mass, it would be helpful to use statistical shape modeling to identify the magnitude of such differences and how they play a role in how the load is transmitted across the femur structure. Finally, the connectivity of trabecular bone is the key property that makes trabecular architecture strong and able to transmit loads. Since only a small proportion of tissue needs to fail for entire structural failure to occur, it would be useful if cortical-trabecular connectivity was assessed using a network theory approach. Network theory would allow us to identify the weakest links in the trabecular structures, which may be more informative than average measures of trabecular thickness and connectivity in a bone volume.

Long term, to identify those who are most susceptible to fracturing we need longitudinal studies following those who lose bone strength through a combination of decreased mass and microstructure with age. It is not possible to resolve bone microstructure using clinically available tools, but databases with longitudinal DEXA scans and fracture outcomes already exist. If key regional measures of bone microstructure could be accurately predicted using machine learning tools, we would immediately have access to a large cohort of data to bridge this gap in knowledge. Thus future work should focus on establishing what three-dimensional microstructural variables may be derivable from 2-D DEXA scans using a variety of machine learning models on a cohort of femurs with both nano-CT and DEXA scans.

In closure, this research improves our current understanding of bone microstructural degeneration that occurs in the distal femur following ACL injury and provides new insight

into regional sex-specific contributors of bone microstructure that predict bone strength. The latter studies provided insight into the etiology of sex-specific differences in hip fracture risk. Chapter one establishes how cortical and trabecular bone at the ACL enthesis changes following injury and sheds light on the variability in bone quality that cannot be simply explained by factors considered herein. Chapter three re-established the relationship between stiffness-strength in the proximal femur, revealing that this relationship also differs in a sex-specific manner. Chapter four presented a novel FCNN segmentation method that significantly outperformed the Otsu approach, the most common segmentation method used in bone research. Chapter five suggests that regional variability is needed to tease out sex-specific differences in bone microstructure that best predict bone strength. This chapter also demonstrated that increased levels of architectural detail only modestly improved our ability to predict bone strength. Finally, this dissertation suggests future directions needed to better understand microstructural changes observed following ACL injury and describes future studies needed to fully understand how bone microstructure is related to bone strength in a sex-specific manner in the proximal femur.

Appendix A

- 4-Point Bending Validation: Briefly, for the diaphysis, our protocol was validated by testing aluminum cylinders under the same loading conditions and confirmed that the derived material modulus was within 1% of textbook values. This validation test was completed previously by Karl Jepsen and Erin Bigelow.
- Proximal Femur Fall-to-Side Validation: Additional sites of deflection not attributable to the proximal femur in our testing configuration include the load cell and the custom bondo pad (Figure A.6.1).

The amount of deflection not attributed to the bone was quantified by indenting a steel platen at 100 mm/sec (same rate as the femoral tests) into the bondo pad so that loads were greater than what was observed in proximal femur fall-to-side testing (1 mm displacement). The steel platen was designed to have a rounded base and a size similar to that of the greater trochanter (Figure A.6.2). Stiffness was quantified three separate times from load-deflection curves (Recording Frequency: 1000 Hz) in Labview 2012. The calculated mean stiffness was 77490 N/mm. The amount of deflection that may have been attributable to the load cell and bondo pad was calculated on twenty bones (F[n=10]: 24-57 years, M[n=10]: 34 -78 years) analyzed in our dataset. The tests showed that the mean displacement potentially attributable to the test fixture design was 0.04 mm (0.02 mm – 0.1 mm), which accounted for 0.96% (0.56% - 2.2%) of the total displacement of the fractured femurs. Thus, the displacement of the proximal femur measured during testing could be reliably attributed to that of the bone, with minimal contributions of the test fixture. As such, the amount of variation in stiffness should be representative of the complex deformation that occurs during a fall to the side test. This will likely include deformations due to bending, torsion, compression and possibly local crushing – all of which likely occur during a fall to the side.

Figures

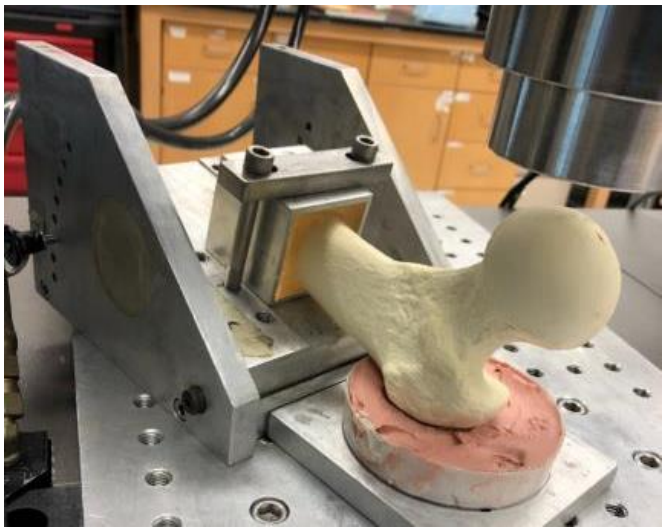


Figure A.6.1 Example set-up of a proximal femur loaded in fall-to-side testing. The proximal femurs were oriented with the shaft at 10 ° of inclination with respect to the horizontal surface and the neck with 15° of internal rotation.

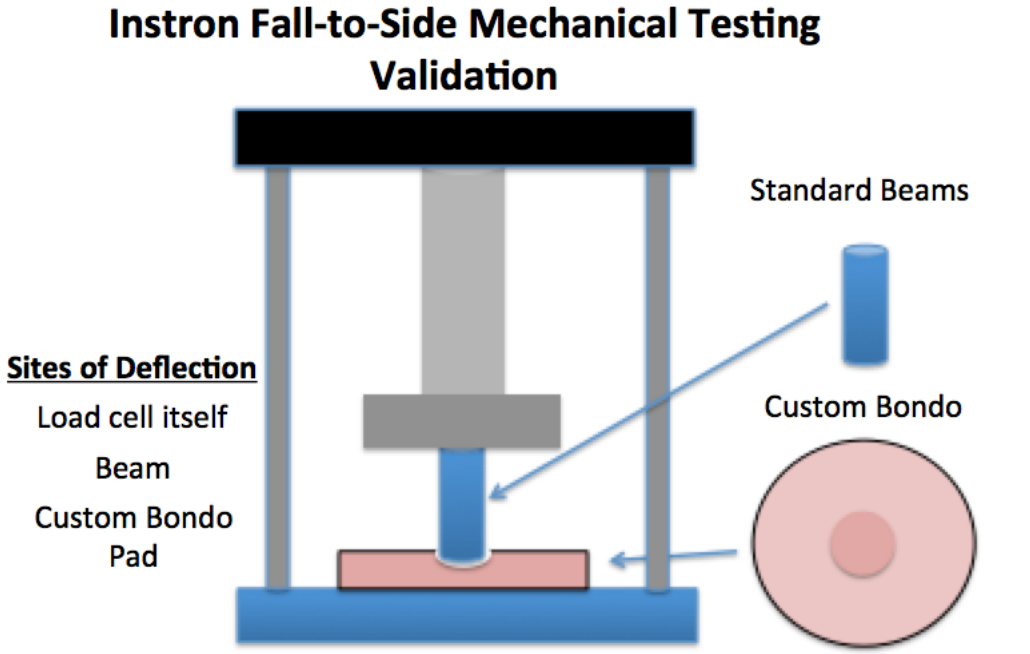


Figure A.6.2 Schematic of fall-to-side validation test with a steal platen loaded into a custom made bondo pad.

Appendix B

A summary of the analysis completed by Benjamin Provencher is briefly described. The objective of this analysis was to select the best hyperparameters for the U-net architecture to optimize bone segmentation. The same training, validation, and test set used in Chapter 4 were also used here (pg. 78).The depth of a U-Net model/architecture was defined here as the number of max-pooling layers. If the depth of a model was increased but the filter count decreased, the total number of learnable parameters can be the same. The models were tested with training parameters either in the 22 or 88 million range (Table B.1). Segmentation quality, measured in terms of DICE, was recorded for the various models tested (Table B.2). Unexpectedly, deeper models did not necessarily perform better. Also, adding more parameters also did not result in better segmentation. The second part of this experiment was to test the same models with different patch sizes (64, 128, 144, 176, 192, 208, 256) (Table B.3). The best result was obtained with smaller patch sizes. These findings helped guide us towards selecting the optimal hyper-parameters for the models used in Chapter 4.

Tables

Table B.1 Small and large models averaging 22 million and 88 million parameters.

Model Name	Parameter Count	Depth	First Layer Filter Count	PatchSize
U-Net_Bones_d_3_fc_128_org	21755714	3	128	64
U-Net_Bones_d_4_fc_64_org	21958050	4	64	64
U-Net_Bones_d_5_fc_32_org	22008530	5	32	64
U-Net_Bones_d_6_fc_16_org	22021098	6	16	64
U-Net_Bones_d_7_fc_8_org	22024214	7	8	64
U-Net_Bones_d_4_fc_128_org	87821122	4	128	64
U-Net_Bones_d_5_fc_64_org	88023458	5	64	64
U-Net_Bones_d_6_fc_32_org	88073938	6	32	64
U-Net_Bones_d_7_fc_16_org	88086506	7	8	64

Table B.2 Results of all small and large models in terms of DICE, VS, and KAP.

Model	DICE	VS	KAP	(DICE + VS + KAP)/3
U-Net_Bones_d_4_fc_128_org	0.964	0.986	0.961	0.970
U-Net_Bones_d_4_fc_64_org	0.964	0.985	0.961	0.970
U-Net_Bones_d_5_fc_32_org	0.964	0.987	0.961	0.971
U-Net_Bones_d_6_fc_16_org	0.964	0.986	0.961	0.970
U-Net_Bones_d_7_fc_16_org	0.964	0.985	0.961	0.970
U-Net_Bones_d_5_fc_64_org	0.963	0.986	0.961	0.970
U-Net_Bones_d_6_fc_32_org	0.963	0.986	0.961	0.970
U-Net_Bones_d_3_fc_128_org	0.963	0.985	0.960	0.970
U-Net_Bones_d_7_fc_8_org	0.962	0.985	0.960	0.969

Table B.3 Results of models tested with different patch size.

Model Name	Patch Size	DICE (mean)	DICE (SD)
U-Net_Bones_org_d_4_fc_64_ps_64	64	0.964	0.007
U-Net_Bones_org_d_4_fc_64_ps_128	128	0.961	0.010
U-Net_Bones_org_d_4_fc_64_ps_144	144	0.963	0.009
U-Net_Bones_org_d_4_fc_64_ps_176	176	0.962	0.010
U-Net_Bones_org_d_4_fc_64_ps_192	192	0.962	0.010
U-Net_Bones_org_d_4_fc_64_ps_208	208	0.963	0.008
U-Net_Bones_org_d_4_fc_64_ps_256	256	0.962	0.010
U-Net_Bones_org_d_4_fc_64_ps_256	256	0.962	0.010

Appendix C

All equations used to calculate the polar moments of inertia are described in the figure below (Figure C.). This figure was adapted from a slide created by Karl Jepsen.

Figure

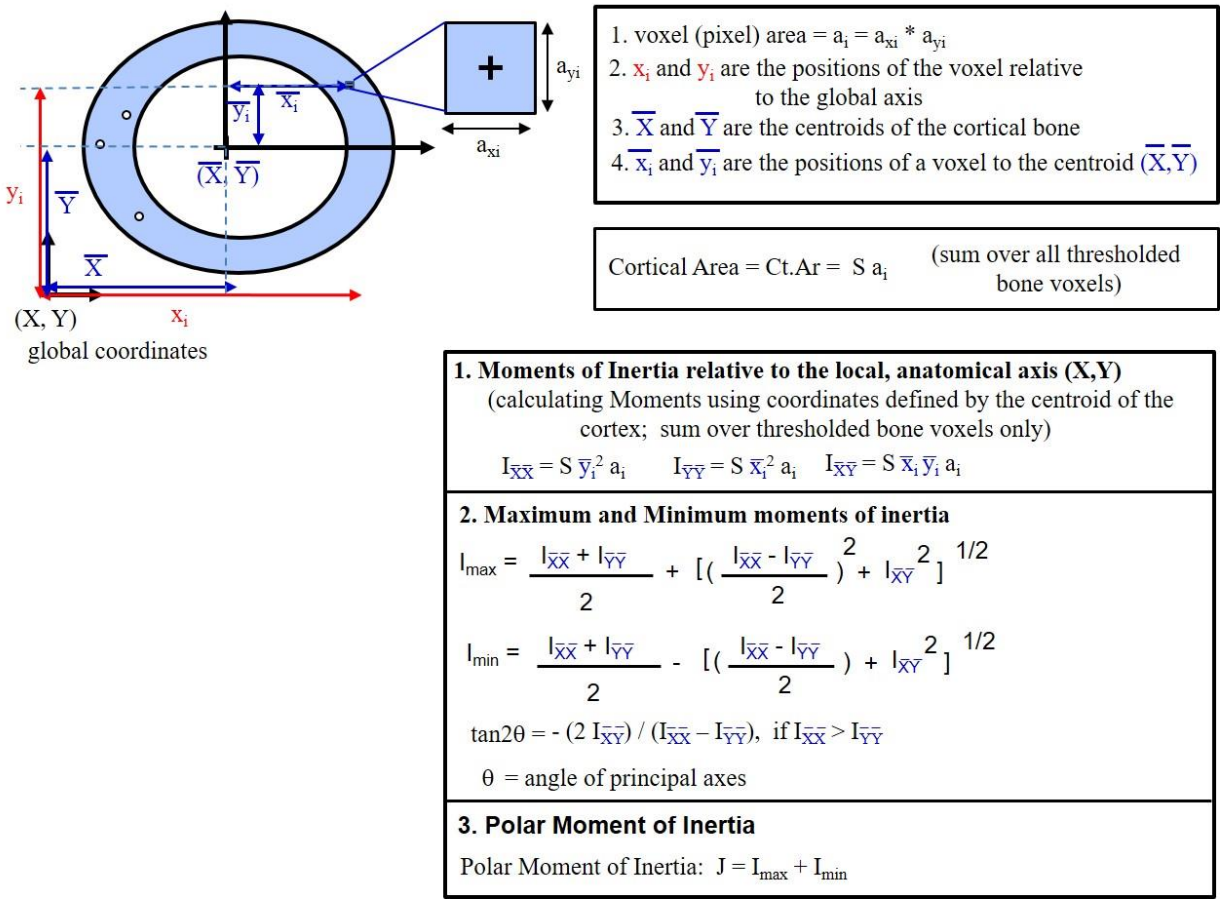


Figure C.1 A figure describing how polar moments of inertia were calculated on a slice by slice basis. Figure courtesy of Karl Jepsen.

Bibliography

1. Burr DB, Akkus O. Bone Morphology and Organization [Internet]. Basic and Applied Bone Biology. Elsevier Inc.; 2013. 3–25 p. Available from: http://dx.doi.org/10.1016/B978-0-12-416015-6.00001-0
2. U.S. Dept. of HHS. Bone health and osteoporosis: a report of the Surgeon General. US Heal Hum Serv [Internet]. 2004;437. Available from: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Bone+Health+and+Osteoporosis+A+Report+of+the+Surgeon+General#0
3. Hebel JR, Zimmerman SI, Wehren LE, Magaziner J, Hawkes WG, Orwig DL. Gender Differences in Mortality After Hip Fracture: The Role of Infection. J Bone Miner Res. 2006;18(12):2231–7.
4. Panula J, Pihlajamäki H, Mattila VM, Jaatinen P, Vahlberg T, Aarnio P, et al. Mortality and cause of death in hip fracture patients aged 65 or older: A population-based study. BMC Musculoskelet Disord. 2011;12(105):2–6.
5. Dargel J, Gotter M, Mader K, Pennig D, Koebke J, Schmidt-Wiethoff R. Biomechanics of the anterior cruciate ligament and implications for surgical reconstruction. Strateg Trauma Limb Reconstr. 2007;2(1):1–12.
6. Gianotti SM, Marshall SW, Hume PA, Bunt L. Incidence of anterior cruciate ligament injury and other knee ligament injuries: A national population-based study. J Sci Med Sport. 2009;12(6):622–7.
7. Mossey JM, Mutran E, Knott K, Craik R. Determinants of recovery 12 months after hip fracture: The importance of psychosocial factors. Am J Public Health. 1989;79(3):279–86.
8. Braithwaite RS, Col NF, Wong JB. Estimating hip fracture morbidity, mortality and costs. J Am Geriatr Soc. 2003;51(3):364–70.
9. Cimino F, Volk BS, Setter D. Anterior cruciate ligament injury: Diagnosis, management, and prevention. Am Fam Physician. 2010;82(8):917–22.
10. Chen H, Zhou X, Fujita H, Onozuka M, Kubo K-Y. Age-Related Changes in Trabecular and Cortical Bone Microstructure. Int J Endocrinol. 2013;2013:1–9.
11. Parkinson IH, Fazzalari NL. Characterisation of Trabecular Bone Structure. In: Skeletal aging and osteoporosis. Berlin, Heidelberg: Springer; 2013. p. 31–51.
12. Lochmüller EM, Matsuura M, Bauer J, Hitzl W, Link TM, Müller R, et al. Site-specific deterioration of trabecular bone architecture in men and women with advancing age. J Bone Miner Res. 2008 Dec;23(12):1964–73.
13. Samitier G, Marcano AI, Alentorn-Geli E, Cugat R, Farmer KW, Moser MW. Failure of Anterior Cruciate Ligament Reconstruction. Arch bone Jt Surg. 2015 Oct;3(4):220–40.
14. Ahn JH, Ha KI, Chung YS, Kim JH. Bone tunnel enlargement after anterior cruciate ligament reconstruction. J Korean Orthop Assoc. 2018;36(3):239.
15. Van Meer BL, Waarsing JH, van Eijsden WA, Meuffels DE, van Arkel ERA, Verhaar JAN, et al. Bone mineral density changes in the knee following anterior cruciate

ligament rupture. Osteoarthr Cartil. 2014;22(1):154–61.
16. Nyland J, Fisher B, Brand E, Krupp R, Caborn DNM. Osseous deficits after anterior cruciate ligament injury and reconstruction: A systematic literature review with suggestions to improve osseous homeostasis. Arthrosc J Arthrosc Relat Surg. 2010;26(9):1248–57.
17. Boyd SK, Matyas JR, Wohl GR, Kantzas A, Steven K, Matyas JR, et al. Early regional adaptation of periarticular bone mineral density after anterior cruciate ligament injury. J Appl Physiol. 2000;89(6):2359–64.
18. Kanis JA. Diagnosis of osteoporosis and assessment of fracture risk. Lancet. 2002 Jun 1;359(9321):1929–36.
19. Hernandez C, Keaveny T. A biomechanical perspective on bone quality. Bone. 2006;23(1):1–7.
20. Zimmermann EA, Schaible E, Bale H, Barth HD, Tang SY, Reichert P, et al. Age-related changes in the plasticity and toughness of human cortical bone at multiple length scales. Proc Natl Acad Sci. 2011;108(35):14416–21.
21. Nawathe S, Akhlaghpour H, Bouxsein ML, Keaveny TM. Microstructural failure mechanisms in the human proximal femur for sideways fall loading. J Bone Miner Res. 2014;29(2):507–15.
22. Wetzler MJ, Bartolozzi AB, Gillespie M, Rubenstein DL, Ciccotti MG, Miller LS. Revision anterior cruciate ligament reconstruction. Arthrosc (Print)Objective Tec Orthopeadics. 1996;5(3):181–9.
23. Vaz MF, Canhão H, Fonseca JE. Bone : A Composite Natural Material. In: Command B-B on, editor. Advances in Composite Materials—Analysis of Natural and Man-Made Materials. Rijeka, Croatia: InTech; 2011. p. 195–229.
24. Turunen MJ, Prantner V, Jurvelin JS, Kröger H, Isaksson H. Composition and microarchitecture of human trabecular bone change with age and differ between anatomical locations. Bone. 2013 May;54(1):118–25.
25. Frost HM. Wolff's Law and bone's structural adaptations to mechanical usage: an overview for clinicians. Angle Orthod. 1994;64(3):175–88.
26. Manmoha Singh, A.R. Nagrath PSM. Changes in Trabecular of the Femur as an Pattern Index of the Upper of Osteoporosis. J Bone Jt Surg. 1970;52(3):457–67.
27. Nazarian A, Muller J, Zurakowski D, Müller R, Snyder BD. Densitometric, morphometric and mechanical distributions in the human proximal femur. J Biomech. 2007;40(11):2573–9.
28. Choksi P, Jepsen KJ, Clines GA. The challenges of diagnosing osteoporosis and the limitations of currently available tools. Clin Diabetes Endocrinol. 2018;4(1):12.
29. Hardy R, Cooper MS. Bone loss in inflammatory disorders. J Endocrinol. 2009;201(3):309–20.
30. Clarke B. Normal bone anatomy and physiology. Clin J Am Soc Nephrol. 2008;3(Supplement 3):S131–9.
31. Rudman KE, Aspden RM, Meakin JR. Compression or tension? The stress distribution in the proximal femur. Biomed Eng Online. 2006;5:1–7.
32. Bouxsein ML, Boyd SK, Christiansen BA, Guldberg RE, Jepsen KJ, Müller R. Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. J Bone Miner Res. 2010;25(7):1468–86.
33. Schlecht SH, Bigelow EMR, Jepsen KJ. How Does Bone Strength Compare Across Sex,

Site, and Ethnicity? Clin Orthop Relat Res. 2015;473(8):2540–7.
34. Zysset P, Qin L, Lang T, Khosla S, Leslie WD, Shepherd JA, et al. Clinical Use of Quantitative Computed Tomography-Based Finite Element Analysis of the Hip and Spine in the Management of Osteoporosis in Adults: The 2015 ISCD Official Positions-Part II. J Clin Densitom. 2015;18(3):359–92.
35. Litjens G, Kooi T, Ehteshami BB, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88.
36. Assessment of the strength of proximal femur in vitro: Relationship to femoral bone mineral density and femoral geometry. Bone. 1997;20(3):213–8.
37. Epelboym Y, Gendron RN, Mayer J, Fusco J, Nasser P, Gross G, et al. The interindividual variation in femoral neck width is associated with the acquisition of predictable sets of morphological and tissue-quality traits and differential bone loss patterns. J Bone Miner Res. 2012;27(7):1501–10.
38. Lotz JC, Hayes WC. The use of quantitative computed tomography to estimate risk of fracture of the hip from falls Use of Quantitative to Estimate Computed Risk Tomography of Fracture of the Hip from Falls. J Bone Joint Surg Am. 2006;72(5):689–700.
39. Rezaei A, Dragomir-Daescu D. Femoral Strength Changes Faster with Age Than BMD in Both Women and Men: A Biomechanical Study. J Bone Miner Res. 2015;30(12):2200–6.
40. Courtney AC, Wachtel EF, Myers ER, Hayes WC. Age-related reductions in the strength of the femur tested in a fall-loading configuration Age-Related Tested in the Strength of the Femur in a Fall-Loading Configuration. J Bone Jt Surg Am. 1995;77(3):387–95.
41. Leblanc KIME, Jr HLM, Leblanc LL, State L, Orleans N. Hip Fracture: Diagnosis, Treatment, and Secondary Prevention. Am Fam Physician. 2014;89(12):945–51.
42. Sanders TL, Maradit Kremers H, Bryan AJ, Larson DR, Dahm DL, Levy BA, et al. Incidence of anterior cruciate ligament tears and reconstruction: A 21-year population-based study. Am J Sports Med. 2016 Jun 1;44(6):1502–7.
43. Jaureguito JW, Paulos LE. Why grafts fail. Clin Orthop Relat Res. 1996;325:25–41.
44. Bach Jr BR. Revision posterior cruciate ligament surgery. Sports Med Arthrosc. 2003;19(10 (Suppl1)):14–29.
45. Biau DJ, Tournoux C, Katsahian S, Schranz P, Nizard R. ACL reconstruction: A meta-analysis of functional scores. Clin Orthop Relat Res. 2007;(458):180–7.
46. Spindler KP, Kuhn JE, Freedman KB, Matthews CE, Dittus RS, Harrell FE. Anterior cruciate ligament reconstruction autograft choice: Bone-tendon-bone versus hamstring. Does it really matter? A systematic review. Am J Sports Med. 2004;32(8):1986–95.
47. Bourke HE, Salmon LJ, Waller A, Patterson V, Pinczewski LA. Survival of the anterior cruciate ligament graft and the contralateral ACL at a minimum of 15 years. Am J Sports Med. 2012 Sep;40(9):1985–92.
48. Wright RW, Magnussen RA, Dunn WR, Spindler KP. Ipsilateral graft and contralateral ACL rupture at five years or more following ACL reconstruction: A systematic review. J Bone Joint Surg Am. 2011;93(12):1159.
49. Kamath G V., Redfern JC, Greis PE, Burks RT. Revision anterior cruciate ligament reconstruction. Am J Sports Med. 2011;39(1):199–217.
50. Schlumberger M, Schuster P, Schulz M, Immendörfer M, Mayer P, Bartholomä J, et al. Traumatic graft rupture after primary and revision anterior cruciate ligament

reconstruction: retrospective analysis of incidence and risk factors in 2915 cases. Knee Surgery, Sport Traumatol Arthrosc. 2017 May 1;25(5):1535–41.
51. Shaerf DA. Anterior cruciate ligament reconstruction best practice: A review of graft choice. World J Orthop. 2014;5(1):23.
52. Harner CD. Revision anterior cruciate ligament reconstruction using fresh-frozen allograft tissue. In: Instructional Course 64th Annual American Academy of Orthopedic Surgeon Meeting. San Francisco; 1997. p. 13–7.
53. Wojtys EM, Beaulieu ML, Ashton-Miller JA. New perspectives on ACL injury: On the role of repetitive sub-maximal knee loading in causing ACL fatigue failure. J Orthop Res. 2016;34(12):2059–68.
54. Smith L, Bigelow EMR, Jepsen KJ. Systematic Evaluation of Skeletal Mechanical Function. Curr Protoc Mouse Biol. 2013;3(2):39–67.
55. Bayar A, Sar S, Keser S, Özdolap Ş, Tuncay İ, Ege A. Regional bone density changes in anterior cruciate ligament deficient knees: A DEXA study. Knee. 2008;15(5):373–7.
56. Leppälä J, Kannus P, Natri A, Pasanen M, Sievänen H, Vuori I, et al. Effect of anterior cruciate ligament injury of the knee on bone mineral density of the spine and affected lower extremity: A prospective one-year follow-up study. Calcif Tissue Int. 1999;64(4):357–63.
57. Kannus P, Sievänen H, Järvinen M, Heinonen A, Oja P, Vuori I. A cruciate ligament injury produces considerable, permanent osteoporosis in the affected knee. J Bone Miner Res. 1992;7(12):1429–34.
58. Reiman MP, Rogers ME, Manske RC. Interlimb Differences in Lower Extremity Bone Mineral Density Following Anterior Cruciate Ligament Reconstruction. J Orthop Sport Phys Ther. 2006;36(11):837–44.
59. Takata S, Abbaspour A, Kashihara M, Nakao S, Yasui N. Unilateral chronic insufficiency of anterior cruciate ligament decreases bone mineral content and lean mass of the injured lower extremity. J Med Investig. 2007;54(3,4):316–21.
60. Sievänen H, Kannus P, Heinonen A, Oja P, Vuori I. Bone mineral density and muscle strength of lower extremities after long-term strength training, subsequent knee ligament injury and rehabilitation: A unique 2-year follow-up of a 26-year-old female student. Bone. 1994;15(1):85–90.
61. LeBlanc AD, Spector ER, Evans HJ, Sibonga JD. Skeletal responses to space flight and the bed rest analog: A review. J Musculoskelet Neuronal Interact. 2007;7(1):33.
62. Rittweger J, Reeves ND, Narici M V., Belavỳ DL, Maganaris CN, Maffulli N. Persisting side-to-side differences in bone mineral content, but not in muscle strength and tendon stiffness after anterior cruciate ligament reconstruction. Clin Physiol Funct Imaging. 2011;31(1):73–9.
63. Feng X. Chemical and Biochemical Basis of Bone Cell. Nih Public Acess. 2010;3(2):975–90.
64. Burghardt AJ, Kazakia GJ, Majumdar S. A local adaptive threshold strategy for high resolution peripheral quantitative computed tomography of trabecular bone. Ann Biomed Eng. 2007;35(10):1678–86.
65. Sasimontonkul S, Bay BK, Pavol MJ. Bone contact forces on the distal tibia during the stance phase of running. J Biomech. 2007;40(15):3503–9.
66. Zhong Z, Zeng XL, Ni JH, Huang XF. Comparison of the biological response of osteoblasts after tension and compression. Eur J Orthod. 2013;35(1):59–65.

67. Thambyah A, Pereira BP, Wyss U. Estimation of bone-on-bone contact forces in the tibiofemoral joint during walking. Knee. 2005;12(5):383–8.
68. Madeti BK, Chalamalasetti SR, Bolla Pragada SKS siva rao. Biomechanics of knee joint — A review. Front Mech Eng. 2015;10(2):176–86.
69. Bloomfield SA. Disuse osteopenia. Curr Osteoporos Rep. 2010;8(2):91–7.
70. Qin W, Bauman WA, Cardozo C. Bone and muscle loss after spinal cord injury: Organ interactions. Ann N Y Acad Sci. 2010;1211:66–84.
71. Schlecht SH, Pinto DC, Agnew AM, Stout SD. Brief communication: The effects of disuse on the mechanical properties of bone: What unloading tells us about the adaptive nature of skeletal tissue. Am J Phys Anthropol. 2012;149(4):599–605.
72. Uhthoff H, Jaworski Z. Bone loss in response to long-term immobilisation. J Bone Joint Surg Br. 2018;60-B(3):420–9.
73. Frost HM. Bone's Mechanostat: A 2003 Update. Anat Rec - Part A Discov Mol Cell Evol Biol. 2003;275(2):1081–101.
74. Tomita F, Yasuda K, Mikami S, Sakai T, Yamazaki S, Tohyama H. Comparisons of intraosseous graft healing between the doubled flexor tendon graft and the bone-patellar tendon-bone graft in anterior cruciate ligament reconstruction. Arthroscopy. 2001;17(5):461–76.
75. Mohtadi NG, Chan DS. A Randomized Trial Comparing Patellar, Hamstring and Double-bundle ACL Reconstruction at 5-yrs. Orthop J Sport Med. 2017;5(7_suppl6):2325967117S0024.
76. Sauer S, Lind M. Bone Tunnel Enlargement after ACL Reconstruction with Hamstring Autograft Is Dependent on Original Bone Tunnel Diameter. Surg J. 2017;03(02):e96–100.
77. Wilson TC, Kantaras A, Atay A, Johnson DL. Tunnel Enlargement after Anterior Cruciate Ligament Surgery. Am J Sports Med. 2004;32(2):543–9.
78. Tanabe Y, Yasuda K, Kondo E, Kitamura N. Clinical results of anterior cruciate ligament reconstruction with ligament remnant tissue preservation: A systematic review. Asia-Pacific J Sport Med Arthrosc Rehabil Technol. 2016;4:1–8.
79. Inokuchi T, Matsumoto T, Takayama K, Nakano N, Zhang S, Araki D, et al. Influence of the Injury-to-Surgery Interval on the Healing Potential of Human Anterior Cruciate Ligament-Derived Cells. Am J Sports Med. 2017;45(6):1359–69.
80. Nohmi S, Yamamoto Y, Mizukami H, Ishibashi Y, Tsuda E, Maniwa K, et al. Post injury changes in the properties of mesenchymal stem cells derived from human anterior cruciate ligaments. Int Orthop. 2012;36(7):1515–22.
81. Zhang S, Matsumoto T, Uefuji A, Matsushita T, Takayama K, Araki D, et al. Anterior cruciate ligament remnant tissue harvested within 3-months after injury predicts higher healing potential. BMC Musculoskelet Disord [Internet]. 2015;16(1):1–9. Available from: http://dx.doi.org/10.1186/s12891-015-0855-0
82. Tanabe Y, Yasuda K, Kondo E, Kitamura N. Clinical results of anterior cruciate ligament reconstruction with ligament remnant tissue preservation: A systematic review. Asia-Pacific J Sport Med Arthrosc Rehabil Technol [Internet]. 2016;4:1–8. Available from: http://dx.doi.org/10.1016/j.asmart.2016.04.001
83. Chen J, Kim J, Shao W, Schlecht SH, Baek SY, Jones AK, et al. An Anterior Cruciate Ligament Failure Mechanism. Am J Sports Med. 2019 Jul 1;47(9):2067–76.
84. Turner CH. Bone strength: Current concepts. Ann N Y Acad Sci. 2006;1068(1):429–46.

85. Roylance D. Mechanical Properties of Materials [Internet]. 1st ed. Cambridge, MA: Wiley; 2000. 1–128 p. Available from: http://www.rivm.nl/dsresource?objectid=08d9708d-a363-47d7-bff7-e9d116b08477&type=pdf&disposition=inline
86. Jurist JM, Foltz AS. Human ulnar bending stiffness, mineral content, geometry and strength. J Biomech. 1977;10(8):455–9.
87. Öhman C, Dall'Ara E, Baleani M, Jan SVS, Viceconti M. The effects of embalming using a 4% formalin solution on the compressive mechanical properties of human cortical bone. Clin Biomech. 2008;23(10):1294–8.
88. Yu B, Zhao GF, Lim JI, Lee YK. Compressive mechanical properties of bovine cortical bone under varied loading rates. Proc Inst Mech Eng Part H J Eng Med. 2011;225(10):941–7.
89. Bouxsein ML, Morgan EF. Biomechanics of bone and fractures. In: Osteoporosis. 2nd ed. San Diego: Acadmeic Press; 2008. p. 509–31.
90. Ammann P, Rizzoli R. Bone strength and its determinants. Osteoporos Int. 2003;14(S3):13–8.
91. Cody DD, Gross GJ, Hou FJ, Spencer HJ, Goldstein SA, Fyhrie DP. Femoral strength is better predicted by finite element models than QCT and DXA. J Biomech. 1999;32(10):1013–20.
92. Fyhrie DP, Vashishth D. Bone stiffness predicts strength similarly for human vertebral cancellous bone in compression and for cortical bone in tension. Bone. 2000;26(2):169–73.
93. Engelke K, van Rietbergen B, Zysset P. FEA to Measure Bone Strength: A Review. Clin Rev Bone Miner Metab. 2016;14(1):26–37.
94. Dall'Ara E, Eastell R, Viceconti M, Pahr D, Yang L. Experimental validation of DXA-based finite element models for prediction of femoral strength. J Mech Behav Biomed Mater. 2016;63:17–25.
95. Patton DM, Bigelow EMR, Schlecht SH, Kohn DH, Bredbenner TL, Jepsen KJ. The relationship between whole bone stiffness and strength is age and sex dependent. J Biomech. 2019;83:125–33.
96. Thurner PJ, Erickson B, Jungmann R, Schriock Z, Weaver JC, Fantner GE, et al. High-speed photography of compressed human trabecular bone correlates whitening to microscopic damage. Eng Fract Mech. 2007;74(12):1928–41.
97. MacNeil JA, Boyd SK. Bone strength at the distal radius can be estimated from high-resolution peripheral quantitative computed tomography and the finite element method. Bone. 2008;42(6):1203–13.
98. Keaveny TM, Kopperdahl DL, Melton LJ, Hoffmann PF, Amin S, Riggs BL, et al. Age-dependence of femoral strength in white women and men. J Bone Miner Res. 2010;25(5):994–1001.
99. Kontulainen SA, Johnston JD, Liu D, Leung C, Oxland TR, Mckay HA. Strength indices from pQCT imaging predict up to 85% of variance in bone failure properties at tibial epiphysis and diaphysis. J Musculoskelet Neuronal Interact. 2008;8(4):401–9.
100. van Rietbergen B, Ito K. A survey of micro-finite element analysis for clinical assessment of bone strength: The first decade. J Biomech. 2015;48(5):832–41.
101. McCalden RW, McGlough JA, Barker MB, Court-Brown CM. Age-related changes in the tensile properties of cortical bone. The relative importance of changes in porosity,

mineralization and microstructure. J Bone Jt Surg - Ser A. 1993;75(8):1193–205.
102. Zioupos P, Currey JD. Changes in the stiffness, strength, and toughness of human cortical bone with age. Bone. 1998;22(1):57–66.
103. Nawathe S, Nguyen BP, Barzanian N, Akhlaghpour H, Bouxsein ML, Keaveny TM. Cortical and trabecular load sharing in the human femoral neck. J Biomech. 2015 Mar 18;48(5):816–22.
104. Jepsen KJ, Centi A, Duarte GF, Galloway K, Goldman H, Hampson N, et al. Biological constraints that limit compensation of a common skeletal trait variant lead to inequivalence of tibial function among healthy young adults. J Bone Miner Res. 2011;26(12):2872–85.
105. Courtney AC, Wachtel EF, Myers ER, Hayes WC. Effects of loading rate on strength of the proximal femur. Calcif Tissue Int. 1994;55(1):53–8.
106. Stine RA. Graphical interpretation of variance inflation factors. Am Stat. 1995;49(1):53–6.
107. Carpenter RD, Sigurdsson S, Zhao S, Lu Y, Eiriksdottir G, Sigurdsson G, et al. Effects of Age and Sex on the Strength and Cortical Thickness of the Femoral Neck. Bone. 2011;48(4):741–7.
108. Sherk VD, Bemben DA. Age and sex differences in estimated tibia strength: influence of measurement site. J Clin Densitom. 2013;16(2):196–203.
109. Looker AC, Beck TJ, Orwoll ES. Does body size account for gender differences in femur bone density and geometry? J Bone Miner Res. 2001;16(7):1291–9.
110. Nieves JW, Formica C, Ruffing J, Zion M, Garrett P, Lindsay R, et al. Males have larger skeletal size and bone mass than females, despite comparable body size. J Bone Miner Res. 2005;20(3):529–35.
111. McNerny EMB, Gong B, Morris MD, Kohn DH. Bone fracture toughness and strength correlate with collagen cross-link maturity in a dose-controlled lathyrism mouse model. J Bone Miner Res. 2015;30(3):446–55.
112. Nyman JS, Roy A, Acuna RL, Gayle HJ, Reyes MJ, Tyler JH, et al. Age-related effect on the concentration of collagen crosslinks in human osteonal and interstitial bone tissue. Bone. 2006;39(6):1210–7.
113. Fields AJ, Nawathe S, Eswaran SK, Jekir MG, Adams MF, Papadopoulos P, et al. Vertebral fragility and structural redundancy. J Bone Miner Res. 2012;27(10):2152–8.
114. Wentz L, Liu P-Y, Haymes E, Ilich JZ. Females Have a Greater Incidence of Stress Fractures Than Males in Both Military and Athletic Populations: A Systemic Review. Mil Med. 2011;176(4):420–30.
115. Cawthon PM. Gender differences in osteoporosis and fractures. Clin Orthop Relat Res. 2011;469(7):1900–5.
116. Nalla RK, Kruzic JJ, Kinney JH, Balooch M, Ager JW, Ritchie RO. Role of microstructure in the aging-related deterioration of the toughness of human cortical bone. Mater Sci Eng. 2006;26(8):1251–60.
117. Burr DB, Martin RB. The effects of composition, structure and age on the torsional properties of the human radius. J Biomech. 1983;16(8):603–8.
118. Ding M, Dalstra M, Danielsen C, Kabel J, Hvid I, Linde F. Age Variations in the Properties of Human Tibial and trabeuclar Bone. J Bone Jt Surg. 1997;79(6):995–1002.
119. Burstein AH, Reilly DT, Martens M. Aging of bone tissue: mechanical properties. J Bone Jt Surg. 1976;58:82–6.

120. Vico L, Zouch M, Amirouche A, Frère D, Laroche N, Koller B, et al. High-resolution pQCT analysis at the distal radius and tibia discriminates patients with recent wrist and femoral neck fractures. J Bone Miner Res. 2008;23(11):1741–50.
121. Liu XS, Cohen A, Shane E, Yin PT, Stein EM, Rogers H, et al. Bone density, geometry, microstructure, and stiffness: Relationships between peripheral and central skeletal sites assessed by DXA, HR-pQCT, and cQCT in premenopausal women. J Bone Miner Res. 2010;25(10):2229–38.
122. Schlecht SH, Bigelow EMR, Jepsen KJ. Mapping the natural variation in whole bone stiffness and strength across skeletal sites. Bone. 2014;67(Supplement C):15–22.
123. Eckstein F, Lochmüller E-M, Lill CA, Kuhn V, Schneider E, Delling G, et al. Bone Strength at Clinically Relevant Sites Displays Substantial Heterogeneity and Is Best Predicted From Site-Specific Bone Densitometry. J Bone Miner Res. 2002;17(1):162–71.
124. Currey JD, Brear K, Zioupos P, Reilly GC. Effect of formaldehyde fixation on some mechanical properties of bovine bone. Biomaterials. 1995;16(16):1267–71.
125. Feldman F, Robinovitch SN. Reducing hip fracture risk during sideways falls: Evidence in young adults of the protective effects of impact to the hands and stepping. J Biomech. 2007;40(12):2612–8.
126. Mcelhaney JH. Dynamic response of bone and muscle tissue. J Appl Physiol. 1966;21(4):1231–6.
127. Hou FJ, Lang SM, Hoshaw SJ, Reimann DA, Fyhrie DP. Human vertebral body apparent and hard tissue stiffness. J Biomech. 1998;31(11):1009–15.
128. Keaveny TM, Wachtel EF, Ford CM, Hayes WC. Differences between the tensile and compressive strengths of bovine tibial trabecular bone depend on modulus. J Biomech. 1994;27(9):1137–46.
129. Feldkamp LA, Goldstein SA, Parfitt MA, Jesion G, Kleerekoper M. The direct examination of three dimensional bone architecture in vitro by computed tomography. J Bone Miner Res. 1989;4(1):3–11.
130. Nishiyama KK, Shane E. Clinical Imaging of Bone Microarchitecture with HR-pQCT. 2014;11(2):147–55.
131. Singh S, Bray TJP, Hall-Craggs MA. Quantifying bone structure, micro-architecture, and pathophysiology with MRI. Clin Radiol. 2018;73(3):221–30.
132. Ruegsegger P, Koller B, Muller R. A Microtomographic System for the Nondestructive Evaluation of Bone Architecture. Calficied Tissue Int. 1995;58:24–9.
133. Ito M, Nakamura T, Matsumoto T, Tsurusaki K, Hayashi K. Analysis of trabecular microarchitecture of human iliac bone using microcomputed tomography in patients with hip arthrosis with or without vertebral fracture. Bone. 1998;23(2):163–9.
134. Hara T, Tanck E, Homminga J, Huiskes R. The influence of microcomputed tomography threshold variations on the assessment of structural and mechanical trabecular bone properties. Bone. 2002;31(1):107–9.
135. Parkinson IH, Badiei A, Fazzalari NL. Variation in segmentation of bone from micro-CT imaging: Implications for quantitative morphometric analysis. Australas Phys Eng Sci Med. 2008;31(2):160–4.
136. Nawathe S, Akhlaghpour H, Bouxsein ML, Keaveny TM. Microstructural Failure Mechanisms in the Human Proximal Femur for Sideways Fall Loading. J Bone Miner Res. 2014;29(2):507–15.

137. Whitmarsh T, Otake Y, Uemura K, Takao M, Sugano N, Sato Y. A cross-sectional study on the age-related cortical and trabecular bone changes at the femoral head in elderly female hip fracture patients. Sci Rep. 2019 Jan;9(1):1–8.
138. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans Syst Man Cybern. 1979;9(1):62–6.
139. Stauber M, Rapillard L, van Lenthe GH, Zysset P, Müller R. Importance of Individual Rods and Plates in the Assessment of Bone Quality and Their Contribution to Bone Stiffness. J Bone Miner Res. 2006;21(4):586–95.
140. Johari N, Singh N, Magid A, Rotman SR, Weiss AM, Meinel L, et al. Age-Related Changes in Trabecular Architecture Differ in Female and Male C57BL/6J Mice. J Bone Miner Res. 1990;22(5):1238–9.
141. Glatt V, Canalis E, Stadmeyer L, Bouxsein ML. Age-Related Changes in Trabecular Architecture Differ in Female and Male C57BL/6J Mice. J Bone Miner Res. 2007;22(8):1197–207.
142. Li C, Jin D, Chen C, Letuchy EM, Janz KF, Burns TL, et al. Automated cortical bone segmentation for multirow-detector CT imaging with validation and application to human studies. Med Phys. 2015;42(8):4553–65.
143. Zebaze R, Seeman E. Cortical Bone: A Challenging Geography. J Bone Miner Res. 2015 Jan 1;30(1):24–9.
144. Buie HR, Campbell GM, Klinck RJ, MacNeil JA, Boyd SK. Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-CT bone analysis. Bone. 2007;41(4):505–15.
145. Ang I, Fox M, Polk JD, Kersh ME. A structure-based algorithm for automated separation of subchondral bone in micro-computed tomography data. engrXiv. 2018;
146. Laib A, Hauselmann HJ, Ruegsegger P. In vivo high resolution 3D-QCT of the human forearm. Technol Health Care. 1998 Dec;6(5–6):329–37.
147. Lublinsky S, Ozcivici E, Judex S. An automated algorithm to detect the trabecular-cortical bone interface in micro-computed tomographic images. Calcif Tissue Int. 2007;81(4):285–93.
148. Kass M, Witkin A, Terzopoulos D. Snakes : Active Contour Models. Int J Comput Vis. 1988;1(4):321–31.
149. Gelaude F, Sloten J Vander, Lauwers B. Semi-automated segmentation and visualisation of outer bone cortex from medical images. Comput Methods Biomech Biomed Engin. 2006;9(1):65–77.
150. Gabel L, Macdonald HM, McKay HA. Reply to: Challenges in the Acquisition and Analysis of Bone Microstructure During Growth. J Bone Miner Res. 2016;31(12):2242–3.
151. du Plessis A, Broeckhoven C, Guelpa A, le Roux SG. Laboratory x-ray micro-computed tomography: A user guideline for biological samples. Gigascience. 2017;6(6):1–11.
152. Mishra C, Gupta DL. Deep Machine Learning and Neural Networks : An Overview. 2017;6(2):66–74.
153. Burkov A. The Hundred-Page Machine Learning Book. Illustrated, editor. Expert Systems. Andriy Burkov, 2019; 2019. 1–160 p.
154. Campanini R, Wenzel DM, Rosa F La. A deep learning approach to bone segmentation in CT scans. Università di Bologna; 2017.
155. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic

segmentation. In: IEEE conference on computer vision and pattern recognition. 2015. p. 3431–40.
156. Goodfellow I, Bengio Y, Courville A. Regularization for Deep Learning [Internet]. Deep Learning. MIT Press; 2016. 216–261 p. Available from: http://www.deeplearningbook.org
157. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018;9(4):611–29.
158. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham; 2015. p. 234–41.
159. Forsberg D, Sjöblom E, Sunshine JL. Detection and Labeling of Vertebrae in MR Images Using Deep Learning with Clinical Annotations as Training Data. J Digit Imaging. 2017;30(4):406–12.
160. Deniz CM, Xiang S, Hallyburton RS, Welbeck A, Babb JS, Honig S, et al. Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks. Sci Rep. 2018;8(1):1–14.
161. Klein A, Warszawski J, Hillengaß J, Maier-Hein KH. Towards whole-body CT bone segmentation. In: Vieweg S, editor. Bildverarbeitung für die Medizin. Berlin, Heidelberg; 2018. p. 204–9.
162. Lessmann N, van Ginneken B, de Jong PA, Išgum I. Iterative fully convolutional neural networks for automatic vertebra segmentation. 2018;
163. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. Magn Reson Med. 2018;79(4):2379–91.
164. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Med Imaging. 2015;15(1):29.
165. R Development Core Team R. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2011.
166. Allaire JJ. RStudio: Integrated development environment for R. J Wildl Manage. 2015;
167. Tassani S, Korfiatis V, Matsopoulos GK. Influence of segmentation on micro-CT images of trabecular bone. J Microsc. 2014;256(2):75–81.
168. Krčah M, Székely G, Blanc R. Fully automatic and fast segmentation of the femur bone from 3D-CT images with no shape prior. In: IEEE international symposium on biomedical imaging: from nano to macro. IEEE; 2011. p. 2087–90.
169. Feurer M, Hutter F. Hyperparameter Optimization. In: Automated Machine Learning. The Springer Series; 2019. p. 3–33.
170. Ribeiro MT, Guestrin C. "Why Should I Trust You ?" Explaining the Predictions of Any Classifier. In: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2016. p. 1135–44.
171. Chen CAN, Surana A, Bloch A, Rajapakse I. Multilinear time invariant systems theory. In: Control and its Applications. Society for Industrial and Applied Mathematics; 2019. p. 118–25.
172. Malcolm J, Rathi Y, Tannenbaum A. A Graph Cut Approach to Image Segmentation in Tensor Space. In: IEEE Conference on Computer Vision and Pattern Recognition. 2007. p. 1–8.
173. Kanis JA, Glüer CC. An update on the diagnosis and assessment of osteoporosis with

densitometry. Osteoporos Int. 2000;11(3):192–202.
174. Kanis, J. A. GC. Consensus development conference: Diagnosis, prophylaxis, and treatment of osteoporosis. Am J Med. 1993 Jun;94(6):646–50.
175. Vondracek SF, Linnebur SA. Diagnosis and management of osteoporosis in the older senior. Clin Interv Aging. 2009;4(1):121–36.
176. Nazrun AS, Tzar MN, Mokhtar SA, Mohamed IN. A systematic review of the outcomes of osteoporotic fracture patients after hospital discharge: Morbidity, subsequent fractures, and mortality. Ther Clin Risk Manag. 2014;10:937–48.
177. Cooper C, Campion G, Melton LJ. Hip fractures in the elderly: A world-wide projection. Osteoporos Int. 1992;2(6):285–9.
178. Kannus P, Pakkaria J, Sievanen H, Heinonen A, Vuori I, Jarvinen M. Epidemiology of Hip Fractures. Bone. 1996;18(3):S57–63.
179. Sullivan KJ, Husak LE, Altebarmakian M, Brox WT. Demographic factors in hip fracture incidence and mortality rates in California, 2000-2011. J Orthop Surg Res. 2016;11(1):1–10.
180. Gallagher CJ, Tella SH. The prevention and treatment of osteoporosis. Arthritis Rheum. 1998;11(2):124–34.
181. Lewis CL, Laudicina NM, Khuu A, Loverro KL. The Human Pelvis: Variation in Structure and Function During Gait. Anat Rec. 2017;300(4):633–42.
182. Purcell M. Sex Differences In The Femur and Acetabulum: Biomechanical Analysis With Forensic Significance. Texas State University; 2013.
183. Kurki HK. Pelvic dimorphism in relation to body size and body size dimorphism in humans. J Hum Evol. 2011;61(6):631–43.
184. Binkley N, Adler R, Bilezikian JP. Osteoporosis Diagnosis in Men: The T-score Controversy Revisited. Bone. 2011;23(1):1–7.
185. Solomon D, Johnston S, Boytsov N, McMorrow D, Lane J, Krohn K. Osteoporosis Medication Use After Hip Fracture in U.S. Patients Between 2002 and 2011. J Bone Miner Res. 2014;29(9):1926–8.
186. Siris ES, Chen YT, Abbott TA, Barrett-Connor E, Miller PD, Wehren LE, et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. Arch Intern Med. 2004;164(10):1108–12.
187. Roberts BJ, Thrall E, Muller JA, Bouxsein ML. Comparison of hip fracture risk prediction by femoral aBMD to experimentally measured factor of risk. Bone. 2009;46(3):742–6.
188. van der Meulen MC, Jepsen KJ, Mikić B. Understanding bone strength: size isn't everything. Bone. 2008;29(2):101–4.
189. Jepsen KJ, Silva MJ, Vashishth D, Guo XE, Van Der Meulen MCH. Establishing biomechanical mechanisms in mouse models: Practical guidelines for systematically evaluating phenotypic changes in the diaphyses of long bones. J Bone Miner Res. 2015;30(6):951–66.
190. Bouxsein ML. Bone quality: where do we go from here? Osteoporos Int. 2003;14(5):118–27.
191. Keaveny TM, Morgan EF, Niebur GL, Yeh OC. Biomechanics of Trabecular Bone. Annu Rev Biomed Eng. 2001;3(1):307–33.
192. Turner CH. Determinants of skeletal fragility and bone quality. J Musculoskelet Neuronal Interact. 2002;2(6):527–8.

193. Macdonald HM, Nishiyama KK, Kang J, Hanley DA, Boyd SK. Age-related patterns of trabecular and cortical bone loss differ between sexes and skeletal sites: A population-based HR-pQCT study. J Bone Miner Res. 2011;26(1):50–62.
194. Carretta R, Stüssi E, Müller R, Lorenzetti S. Within subject heterogeneity in tissue-level post-yield mechanical and material properties in human trabecular bone. J Mech Behav Biomed Mater. 2013;24:64–73.
195. Eckstein F, Matsuura M, Kuhn V, Priemel M, Müller R, Link TM, et al. Sex differences of human trabecular bone microstructure in aging are site-dependent. J Bone Miner Res. 2007;22(6):817–24.
196. Chiba K, Burghardt AJ, Osaki M, Majumdar S. Heterogeneity of bone microstructure in the femoral head in patients with osteoporosis: An ex vivo HR-pQCT study. Bone. 2013 Sep;56(1):139–46.
197. Van Der Linden JC, Homminga J, Verhaar JAN, Weinans H. Mechanical consequences of bone loss in cancellous bone. J Bone Miner Res. 2001;16(3):457–65.
198. Kersh ME, Pandy MG, Bui QM, Jones AC, Arns CH, Knackstedt MA, et al. The heterogeneity in femoral neck structure and strength. J Bone Miner Res. 2013 May;28(5):1022–8.
199. Bell KL, Loveridge N, Power J, Garrahan N, Stanton M, Lunt M, et al. Structure of the femoral neck in hip fracture: Cortical bone loss in the inferoanterior to superoposterior axis. J Bone Miner Res. 1999;14(1):111–9.
200. Wang L, Cheng XG, Su YB, Brown K, Xu L, Li K, et al. Sex-related variations in cortical and trabecular bone of the femoral neck in an elderly Chinese population. Osteoporos Int. 2017 Aug;28(8):2391–9.
201. Malo MKH, Rohrbach D, Isaksson H, Töyräs J, Jurvelin JS, Tamminen IS, et al. Longitudinal elastic properties and porosity of cortical bone tissue vary with age in human proximal femur. Bone. 2013 Apr;53(2):451–8.
202. Bell KL, Loveridge N, Power J, Garrahan N, Meggitt BF, Reeve J. Regional differences in cortical porosity in the fractured femoral neck. Bone. 1999 Jan;24(1):57–64.
203. Crabtree N, Loveridge N, Parker M, Rushton N, Power J, Bell KL, et al. Intracapsular hip fracture and the region-specific loss of cortical bone: Analysis by peripheral quantitative computed tomography. J Bone Miner Res. 2001;16(7):1318–28.
204. Cheng XG, Lowet G, Boonen S, Nicholson PHF, Brys P, Nijs J, et al. Assessment of the strength of proximal femur in vitro: Relationship to femoral bone mineral density and femoral geometry. Bone. 1997 Mar;20(3):213–8.
205. Langton CM, Pisharody S, Keyak JH. Comparison of 3D finite element analysis derived stiffness and BMD to determine the failure load of the excised proximal femur. Med Eng Phys. 2009 Jul;31(6):668–72.
206. Nawathe S. Micromechanics of the Human Proximal Femur: Role of Microstructure and Tissue-Level Ductility on Femoral Strength. UC Berkeley Electronic Theses and Dissertations. UC Berkeley; 2014.
207. Verhulp E, van Rietbergen B, Huiskes R. Load distribution in the healthy and osteoporotic human proximal femur during a fall to the side. Bone. 2008;42(1):30–5.
208. AO Foundation. Müller AO Classification of Fractures—Long Bones [Internet]. AO Foundation. 2010. Available from: www.aofoundation.org
209. Wickham H. ggplot2. Wiley Interdiscip Rev Comput Stat. 2011;
210. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models

via coordinate descent. J Stat Softw. 2010;33(1):1.
211. Fonti V. Feature Selection using LASSO. VU Amsterdam. VU Amsterdam; 2017.
212. Tibshirani R. Regression shrinkage and selection via the lasso : a retrospective. R Stat Soc. 2011;73(3):273–82.
213. Friedman THRTJ. The Elements of Statistical Learning Data Mining, Inference, and Prediction. Vol. 26, Springer. Springer; 2011. 505–516 p.
214. Lang TF, Sigurdsson S, Karlsdottir G, Oskarsdottir D, Sigmarsdottir A, Chengshi J, et al. Age-related loss of proximal femoral strength in elderly men and women: The Age Gene/Environment Susceptibility Study - Reykjavik. Bone. 2012 Mar;50(3):743–8.
215. Rezaei A, Dragomir-Daescu D. Femoral Strength Changes Faster with Age Than BMD in Both Women and Men: A Biomechanical Study. J Bone Miner Res. 2015;30(12):2200–6.
216. Schott AM, Cormier C, Hans D, Favier F, Hausherr E, Dargent-Molina P, et al. How hip and whole-body bone mineral density predict hip fracture in elderly women: The EPIDOS prospective study. Osteoporos Int. 1998;8:247–54.
217. Mautalen CA, Vega EM, Einhorn TA. Are the etiologies of cervical and trochanteric hip fractures different? Bone. 1996;18(3):133S-137S.
218. Baudoin C, Fardellone P, Sebert JL. Effect of sex and age on the ratio of cervical to trochanteric hip fracture: A meta-analysis of 16 reports on 36,451 cases. Acta Orthop. 1993;64(6):647–53.
219. Tanner D, Kloseck M, Crilly R, Chesworth B, Gilliland J. Hip fracture types in men and women change differently with age. BMD Geriatr. 2010;10(12):1–4.
220. Cauley JA, Zmuda JM, Wisniewski SR, Krishnaswami S, Palermo L, Stone KL, et al. Bone mineral density and prevalent vertebral fractures in men and women. Osteoporos Int. 2004;15(1):32–7.
221. Selby PL, Davies M, Adams JE. Do men and women fracture bones at similar bone densities? Osteoporos Int. 2000;11(2):153–7.
222. Langsetmo L, Leslie WD, Zhou W, Goltzman D, Kovacs CS, Prior JC, et al. Using the same bone density reference database for men and women provides a simpler estimation of fracture risk. J Bone Miner Res. 2010 Oct;25(10):2108–14.
223. O'Neill TW, Lunt M, Silman AJ, Felsenberg D, Benevolenskaya LI, Bhalla AK, et al. The relationship between bone density and incident vertebral fracture in men and women. J Bone Miner Res. 2002 Dec 1;17(12):2214–21.
224. Brown JP, Josse RG, Canada SAC of the OS of. 2002 clinical practice guidelines for the diagnosis and management of osteoporosis in Canada. CMAJ. 2002 Nov 12;167(10 Suppl):S1–34.
225. Bigelow EMR, Patton DM, Ward FS, Ciarelli A, Casden M, Clark A, et al. External Bone Size Is a Key Determinant of Strength-Decline Trajectories of Aging Male Radii. J Bone Miner Res. 2019;
226. Reilly DT, Burstein AH. The elastic and ultimate properties of compact bone tissue. J Biomech. 1975;8(6).
227. Bredbenner TL, Eliason TD, Potter RS, Mason RL, Havill LM, Nicolella DP. Statistical shape modeling describes variation in tibia and femur surface geometry between Control and Incidence groups from the Osteoarthritis Initiative database. J Biomech [Internet]. 2010;43(9):1780–6. Available from: http://dx.doi.org/10.1016/j.jbiomech.2010.02.015
228. Brinckmann P, Hoefert H, Jongen HT. Sex differences in the skeletal geometry of the

human pelvis and hip joint. J Biomech. 1981;14(6):427–30.
229. Holliday TW, Hutchinson VT, Morrow MMB, Livesay GA. Geometric morphometric analyses of hominid proximal femora: Taxonomic and phylogenetic considerations. HOMO- J Comp Hum Biol [Internet]. 2010;61(1):3–15. Available from: http://dx.doi.org/10.1016/j.jchb.2010.01.001
230. Preininger B, Schmorl K, Von Roth P, Winkler T, Matziolis G, Perka C, et al. The sex specificity of hip-joint muscles offers an explanation for better results in men after total hip arthroplasty. Int Orthop. 2012;36(6):1143–8.