Improving the Reliability of HR-pQCT Measurements in Longitudinal Studies of Individuals with Osteogenesis Imperfecta (OI)

Submitted by

Mahdi Hosseinitabatabaei

August 2023

A THESIS SUBMITTED TO MCGILL UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF DOCTOR OF PHILOSOPHY

Department of Biomedical Engineering

McGill University

This thesis is dedicated to:

My lovely wife Mozhdeh for her continuous love and unwavering support. My dear son Nickan for bringing a new meaning to my life. My wonderful parents and sister for their endless encouragement and understanding over the years. My amazing in-laws for their encouragement and support.

Table of Contents

List of Figures:

Chapter 2:

List of Tables:

Abbreviations:

Abbreviation	Definition
3D	Three-dimensional
3D-DP	3D registration using different periosteal mask
3D-SP	3D registration using the same periosteal mask
3D-TB	3D registration using transformed boundary conditions
AIC	Akaike information criterion
App. Modulus	Apparent modulus
AUC	Area under the curve
AW	Adapted window
BMC	Bone mineral content
BMD	Bone mineral density
BMU	Basic multicellular unit
BP	Bisphosphonate
BV	Bone volume
CL	Minimum cluster size
CSA	Cross-sectional-area image registration (same as SM)
CT	Computed tomography
Ct.Ar	Cortical area
Ct.Po	Cortical porosity
Ct.Th	Cortical thickness
Ct.vBMC	Cortical volumetric bone mineral content
Ct.vBMD	Cortical volumetric bone mineral density
Dmov	Displacement boundary condition for moving image
Dref	Displacement boundary condition for reference image
DW	Disuse window
DXA	Dual-energy X-ray absorptiometry
EV	Eroded volume
EV/BV	Eroded volume normalized by bone volume
F load	Failure load
FE	Finite element
Fx	Fracture
GP	Growth plate

HA	Hydroxyapatite
HpSC	Hematopoietic stem cell
HR-pQCT	High-resolution peripheral quantitative computed tomography
LC	Lining cell
MA	Matched-angle
MA-DP	Matched-angle with different periosteal mask
MA-SP	Matched-angel with same periosteal mask
MC	Unspecified marrow cell
MES	Minimum effective strain
MESm	Minimum effective strain for modeling
MESp	Minimum effective strain for microdamage accumulation
MESr	Minimum effective strain for remodeling
microFE	micro-finite-element
MOW	Mild overuse window
Mr.Ar	Marrow cavity area
MRI	Magnetic resonance imaging
MV	Mineralized volume
MV/BV	Mineralized volume normalized by bone volume
Ob	Osteoblast
Oc	Osteoclast
ODN	Odanacatib
OI	Osteogenesis Imperfecta
OPG	Osteoprotegerin
Ot	Osteocyte
POW	Pathological overuse window
pQCT	Peripheral quantitative computed tomography
Q	Registration transformation matrix
QC1	Density quality control phantom provided by Scanco
QCT	Quantitative computed tomography
R^2	Regression coefficiet of determination
RANKL	Kappa-B ligand
RPP	Registration performance plot
Scl-Ab	Sclerostin neutralizing antibody

SD	Standard deviation
SEM	Standard error of mean
SM	Slice-match image registration (same as CSA)
StSC	Stromal stem cell
Tb.1/N.SD	Inhomogeneity of trabecular network
Tb.Ar	Trabecular area
Tb.BV/TV	Trabecular bone volume fraction
Tb.N	Trabecular number
Tb.Sp	Trabecular separation
Tb.Th	Trabecular thickness
Tb.vBMC	Trabecular volumetric bone mineral content
Tb.vBMD	Trabecular volumetric bone mineral density
Tt.vBMC	Total volumetric bone mineral content
Tt.vBMD	Total volumetric bone mineral density
XCT	First generation HR-pQCT scanner
XCT2	Second generation HR-pQCT scanner
XtremeCT	First generation HR-pQCT scanner (same as XCT)
XtremeCTII	Second generation HR-pQCT scanner (same as XCT2)

Abstract:

High-resolution peripheral-quantitative computed tomography (HR-pQCT) is an imaging tool for measuring bone microstructure, density, strength, and (re)modeling in the peripheral skeleton. There are challenges related to HR-pQCT application in longitudinal settings. Further, existing HR-pQCT data is limited for children with osteogenesis imperfecta (OI), a genetic disorder resulting in bone fragility.

Repositioning error in HR-pQCT imaging leads to different bone volumes assessed over time. Image registration is used to identify the same bone volume. While the commonly used cross-sectional area registration corrects for axial misalignment, 3D-registration additionally corrects for rotations. For 3D-registered micro finite-element analysis, other registration methods involving matched angle analysis or boundary transformations can be used to limit interpolation error. I investigated the effect of different registration methods on the short-term precision in adults with OI using same-day repeated scans. Image registration improved precision, with 3D-registration marginally outperforming cross-sectional area registration for trabecular and microFE outcomes.

Image registration is also used for timelapse HR-pQCT to quantify bone formation and resorption. However, there is no consensus on what image registration method or input image type to use. Further, proper definition of the periosteal mask, and proper methods for minimal noise and error are not well standardized. Finally, no validation is available in the literature for timelapse HR-pQCT. As part of a multicenter trial, I used the same-day repeated scans from 29 adults with OI to examine the influence of various parameters on HR-pQCT-derived bone formation and resorption. These parameters included the input image (binary or grayscale), registration method (3D or matched-angle), and segmentation mask (original or dilated mask). For grayscale images, I evaluated various values for the density difference between voxels to be considered formed or resorbed, the minimum size of formation/resorption clusters, and gaussian smoothing sigma. For both XCT and XCT2, a density threshold of 200 mg/cm^3, and a cluster size of 0 resulted in formation/resorption volumes approaching zero, with negligible effect of increasing the density threshold and cluster size, and negligible noise when combined with Gaussian noise reduction. I validated the selected method using a combination of repeated and longitudinal scans. Finally, using the selected and validated method, a positive dose-dependent effect of an anabolic drug (setrusumab) was observed on bone formation and resorption at the distal radius and tibia of adults with OI.

Pediatric HR-pQCT data is limited. Only one cross-sectional study has reported HR-pQCT data for children with OI at the metaphysis, although fractures in children with OI often occur in the diaphysis. I compared HR-pQCT measurements and their changes during 1 year between children with OI and age-

and sex-matched controls. The data showed more prominent differences between groups at the tibia metaphysis compared to the radius. The 1-year changes were comparable between the OI and control groups. At the diaphysis, differences between the two groups were mainly driven by area.

In conclusion, I discussed challenges related to longitudinal HR-pQCT studies in the OI population, and provided methodological solutions, and clinically relevant data.

Résumé:

La tomodensitométrie périphérique quantitative à haute résolution (HR-pQCT) est un outil d'imagerie permettant de mesurer la microstructure osseuse, la densité, la résistance et la (re)modélisation du squelette périphérique. Il existe des défis liés à l'application HR-pQCT dans les contextes longitudinaux. De plus, les données HR-pQCT existantes sont limitées pour les enfants atteints d'ostéogenèse imparfaite (OI), une maladie génétique entraînant une fragilité osseuse.

L'erreur de repositionnement dans l'imagerie HR-pQCT conduit à différents volumes osseux évalués au fil du temps. L'enregistrement d'image est utilisé pour identifier le même volume osseux. Alors que l'enregistrement de la section transversale couramment utilisé corrige le désalignement axial, l'enregistrement 3D corrige en outre les rotations. Pour la micro-analyse par éléments finis (microFE) d'images superposées en 3D, d'autres méthodes de superposition impliquant une analyse par angle apparié (MA) ou des transformations de limites (3D-TB) peuvent être utilisées pour limiter l'erreur d'interpolation. J'ai étudié l'effet de différentes méthodes d'enregistrement sur la précision à court terme chez les adultes atteints d'OI en utilisant des scans répétés le même jour. L'enregistrement d'image a amélioré la précision, l'enregistrement 3D surpassant légèrement l'enregistrement de la zone transversale pour les résultats trabéculaires et microFE.

La superposition d'images est également à la base de la quantification par HR-pQCT des changements de formation et résorption osseuses chez l'homme dans le temps. En outre, la définition correcte du masque périosté et les méthodes appropriées pour un minimum de bruit et d'erreur ne sont pas bien standardisées. Enfin, aucune validation n'est disponible dans la littérature pour le timelapse HR-pQCT. Dans le cadre d'un essai multicentrique, j'ai utilisé les scans répétés le même jour de 29 adultes atteints d'OI pour examiner l'influence de divers paramètres sur la formation et la résorption osseuse dérivées du HR-pQCT. Ces paramètres comprenaient l'image d'entrée (binaire ou en niveaux de gris), la méthode d'enregistrement (3D ou à MA) et le masque de segmentation (masque d'origine ou dilaté). Pour les images en niveaux de gris, j'ai évalué diverses valeurs pour la différence de densité entre les voxels à considérer comme formés ou résorbés, la taille minimale des amas de formation/résorption et le sigma de lissage gaussien. Pour XCT et XCT2, un seuil de densité de 200 mg/cm3 et une taille de cluster de 0 ont entraîné des volumes de formation/résorption proches de zéro, avec un effet négligeable d'augmentation du seuil de densité et de la taille des clusters, et un bruit négligeable lorsqu'il est combiné au bruit gaussien réduction. J'ai validé la méthode choisie en utilisant une combinaison de balayages répétés et longitudinaux. Enfin, en utilisant la méthode sélectionnée et validée, un effet dose-dépendant

positif d'un médicament anabolisant (setrusumab) a été observé sur la formation et la résorption osseuse au niveau du radius distal et du tibia d'adultes atteints d'OI.

Les données pédiatriques HR-pQCT sont limitées. Une seule étude transversale a rapporté des données HR-pQCT pour les enfants atteints d'OI au niveau de la métaphyse, bien que les fractures chez les enfants atteints d'OI se produisent souvent dans la diaphyse. J'ai comparé les mesures HR-pQCT et leurs changements pendant 1 an entre les enfants atteints d'OI et les témoins appariés selon l'âge et le sexe. Les données ont montré des différences plus importantes entre les groupes au niveau de la métaphyse du tibia par rapport au rayon. Les changements à 1 an étaient comparables entre les groupes IO et témoin. Au niveau de la diaphyse, les différences entre les deux groupes étaient principalement liées à la zone.

En conclusion, j'ai discuté des défis liés aux études longitudinales HR-pQCT dans la population IO, et j'ai fourni des solutions méthodologiques et des données cliniquement pertinentes.

Acknowledgements:

I would like to thank my supervisor, Dr. Bettina Willie, for giving me the opportunity to learn and thrive at McGill University and Shriners Hospital for Children. She had not only been a wonderful supervisor, but a real mentor who not only helped me in shaping my character and discovering my future ambitions and life goals, but also offered me opportunities that could help me reach my long-term goals. She has been very caring and supportive, open to ideas, easy to communicate with, and transparent, and I learned so much from her.

I am grateful to have the opportunity to work on innovative projects that can directly contribute to improve the quality of life of adults and children with bone related conditions, especially osteogenesis imperfecta. Working at Shriners Hospital for Children-Canada has been a privilege. Doing my PhD in this environment was truly inspiring for me, especially engaging with those who suffer from osteogenesis imperfecta, and researchers who work nonstop to improve the quality of life of those individuals.

Thank you to the members of our lab, Dr. Zimmermann, Dr. Rummler, Alice, Kyle, David, Isabela, Aren, Joseph, Taylor, Beatrice, Celine, Dr. Tauer, Samantha, Carolyn, Jack, Gabriella, Annie, Meghan, Leanne, and Meet. Special thanks to Catherine for being supportive is many critical occasions, and covering for scanning sessions, and helping with French translation on my abstracts.

I thank the members of my advisory committee, Dr. Svetlana Komarova and Dr. Robert Kearney for providing me with invaluable feedback that helped me learn more, and to better shape my research.

A would like to express my gratitude to the Department of Biomedical Engineering for their support, in particular Pina Sorrini for always taking the time and kindly helping me with my questions. I also thank Sabrina Teoli for assisting with administrative questions.

I would like to acknowledge the funding sources that have supported me during my doctoral studies: The Fonds de recherche du Quebec – Sante (doctoral training program), BBME department for the PhD excellence award, and International Differential Fee Waiver, Faculty of Medicine for Max Binz Doctoral Fellowship, The Network for Oral and Bone Health Research for Doctoral Fellowship, and several organizations and agencies for abstract and travel awards.

I would like to express my deepest love and gratitude to Mozhdeh for her endless love and support, being a friend, making my life full of joy, and for being there for me through all the ups and downs.

I thank my parents and my sister for their constant support paving the way for me to reach my goals. Last but not least, I thank my in-laws for being a source of encouragement and support, especially my mother-in-law who supported us during and after my son's birth, so I could submit my thesis in time!

Contribution to original knowledge:

In this thesis, I discussed several challenges related to longitudinal HR-pQCT studies in both adult and pediatric OI populations, and provided solutions for these challenges to improve the reliability of longitudinal HR-pQCT studies. In Chapter 4, I discussed repositioning error, which results in capturing slightly different bone regions overtime, hence obscuring the true biological changes occurring in bone. The goal was to identify the image registration method that could most effectively reduce the repositioning errors. I showed that image registration is indeed required to improve the precision of longitudinal images, and 3D registration is the preferred method compared to the typical registration based on matching the cross-sectional-area between scans. I also showed the importance of acquiring scans with minimal motion considering that image registration could only partially reduce the impact of motion. My approach to this problem had several unique aspects. First, our group presented the first set of HR-pQCT precision data for adults with osteogenesis imperfecta (OI). Second, I used data from a multicenter clinical trial that were representative of a real-world scenario. This is in contrast to other studies that collected scans from single imaging sites in a research setting. This difference can improve the translation of the findings of the study to clinical trials. Further, I used a systematic approach to compare several variations of 3D image registration, which was lacking in the literature. Studying these variations helped us to better understand the importance of subtle factors on the precision of HR-pQCT outcomes. Another important aspect of this study was our approach to interpret the precision errors. While the previous studies in the literature solely focused on the aggregate precision errors (i.e., mean or median) for different registration methods, we complimented the aggregate results with individual level results that could reveal subtle differences between the performance of image registration methods in terms of their performance to improve or even deteriorate precision errors. Using this extensive analysis, we were able to reveal the improvements in precision errors when using 3D image registration. Finally, we presented extensive details regarding our methodologies to improve reproducibility, and shared our scripts for public use on Github.

In Chapter 5, I performed a comprehensive investigation on the combination of settings that would result in the most reliable quantification of bone formation and resorption using timelapse HR-pQCT. This study was also a part of a multicenter dose finding clinical trial in adults with OI. Timelapse HR-pQCT method is relatively recent, and only a few groups have investigated it. This study is however the first to investigate the effect of different settings to this extent using same-day repeated scans. The repeated scans were from the same dataset used in the study presented in Chapter 4. I first compared different image registration methods and found that 3D and matched angle image registration methods

result almost identical errors on the same-day repeated scans. I further compared the input image type for timelapse analysis. This comparison is important as some of the studies in the literature have opted to use binary input images, while some other studies used grayscale input images. I showed that the binary method results in high errors, which make it unreliable for longitudinal studies. In contrast, the grayscale method can be combined with proper settings to achieve low errors close to zero. One of such settings is noise reduction. While all of the studies in the literature have used post processing techniques to remove small clusters of formed and resorbed voxels as noise, for the first time I showed that reducing noise in the grayscale images prior to their subtraction is also possible and is preferable. Another important novelty in this work was to validate the timelapse analysis using a combination of repeated and longitudinal scans. Through this validation study, we showed that while there was negligible formation and resorption between the same-day repeated scans, matching either of the repeated scans with the same baseline scan result in similar regions identified as formation and resorption, confirming that the identified regions were not due to random noise, but rather were true biological changes. I also looked at two possible confounding factors for the first time, namely the effect of image rotation angle during registration and the drift in attenuation coefficients on the computed bone formation and resorption outcomes. From our analysis, we ruled out the notion that those factors could have driven the magnitude of bone formation and resorption. Finally, I implemented the selected and validated method, and found a positive dose-dependent effect of an anabolic drug (setrusumab) on bone formation and resorption of adults with OI. The results of these analysis gave us additional insight into the dynamics of changes in bone, rather than just the overall effect as conventionally shown by changes in bone density, morphology, and strength outcomes. All of the methodologies and scripts will be publicly available.

Finally in Chapter 6, I discussed some of the rather under-appreciated challenges specific to longitudinal HR-pQCT imaging in children. In this study, I investigated the feasibility of using rigid 3D image registration to align the longitudinal scans of growing children, and showed that it is indeed not feasible unless specific landmarks such as sclerotic line are present in the scans. I also acquired double-stack scans and investigated the misalignment between these stacks, and found that the misalignment between consecutive stacks is not scanner-related and is caused by limb movement. For average-based HR-pQCT outcomes including density, geometry, and microstructure, misalignment between the two stacks is not problematic. However, microFE is sensitive to misalignment. In those situations, either the double stacks can be analyzed separately or misalignment may be corrected. If alignment of consecutive stacks is necessary, leaving an overlap between the stacks is recommended. While a few studies have used multi-stack scans, the challenges related to the misalignment between then have never been

discussed. Another novelty in this study was including the diaphyseal region, which can provide unique insight considering that the bone undergoes a different mode of loading at this region compared to the metaphysis. Further, bone morphology in entirely different at the metaphysis and diaphysis. In addition to the technical novelties, this will also be the first study to report longitudinal HR-pQCT data in children with OI. This is especially important at the long bone diaphysis, which is a site of frequent fracture in these children. This study also included age- and sex-matched healthy controls to enable comparison between these two groups. The results at baseline showed that at the metaphysis, differences between the children with OI and healthy controls were more prominent at the tibia compared to the radius, mainly due to more deteriorated trabecular microstructure. At the diaphysis, I showed that the OI group had smaller cross-sectional area, whereas cortical density was similar. Overall, the results of this study suggest that the changes in bone measurements over the 1-year period are similar between the OI and control groups.

Contribution of authors:

Chapter 4. Evaluating the precision of HR-pQCT measurements using different registration methods in adults with OI: This study has been published in the Journal of Bone and Mineral Research (JBMR) in 2022 under the title

"3D image registration marginally improves the precision of HR-pQCT measurements compared to cross-sectional-area registration in adults with osteogenesis imperfecta"

Authors: **Seyedmahdi Hosseinitabatabaei**, Nicholas Mikolajewicz, Elizabeth A. Zimmermann, Maximilian Rummler, Beatrice Steyn, Catherine Julien, Frank Rauch, Bettina M. Willie

Contributions:

S. Hosseinitabatabaei – Conceptualization, data curation, formal analysis, investigation, methodology, software (developing scripts for image registration), visualization, writing of original draft

N. Mikolajewicz – Assistance with data curation and statistical analysis; partly involved in the writing of original draft

E. Zimmermann – Involved in study conceptualization and data curation, performing HR-pQCT scans, revision of the manuscript

M. Rummler – Assisted with performing HR-pQCT scans and revision of the manuscript

B. Steyn – Assisted with performing the 3D-registered analyses and revision of the manuscript

C. Julien – Assisted with performing HR-pQCT scans, data curation, and revision of the manuscript

F. Rauch – Assisted with interpretation, and revision of the manuscript

B. Willie – Conceptualization, formal analysis, funding acquisition, project administration, supervision, writing of original draft

Chapter 5. Spatiotemporal changes at the distal radius and tibia of adults with OI induced by setrusumab (anabolic drug): Currently in preparation for the Journal of Bone and Mineral Research (JBMR) under the title

"Non-invasive quantification of bone remodeling dynamics in adults with osteogenesis imperfecta using time-lapse HR-pQCT"

Authors: **Seyedmahdi Hosseinitabatabaei**, Isabela Vitienes, Maximillian Rummler, Annette Birkhold, Frank Rauch, Bettina M. Willie

Contributions:

S. Hosseinitabatabaei – Conceptualization, data curation, formal analysis, investigation, methodology, software (developing scripts for image registration and timelapse analyses), visualization, writing of original draft

I. Vitienes – Assistance with data curation and revision of the manuscript

M. Rummler – Assistance with data curation and revision of the manuscript

A. Birkhold – Assisted with interpretation, and revision of the manuscript

F. Rauch – Assisted with interpretation, and revision of the manuscript

B. Willie – Conceptualization, formal analysis, funding acquisition, project administration, supervision, writing of original draft

Chapter 6. Studying the natural history of the peripheral bones of children with OI and healthy controls using longitudinal HR-pQCT analysis: Currently in preparation for the Journal of Bone and Mineral Research (JBMR) under the title

"Natural history of the peripheral bones in children with osteogenesis imperfecta and age- and sex-matched healthy controls using longitudinal HR-pQCT analysis"

Authors: **Seyedmahdi Hosseinitabatabaei**, Samantha McCluskey, Elizabeth A. Zimmermann, Francis H. Glorieux, Fredrick Charbonneau, Frank Rauch, Bettina M. Willie

Contributions:

S. Hosseinitabatabaei – Conceptualization, data collection (HR-pQCT scanning) and curation, development of scanning protocol, formal analysis, investigation, methodology, software (developing scripts for image registration and stack alignment), visualization, writing of original draft

S. McCluskey – Assistance with data curation (analysis of scans) and revision of the manuscript

E. Zimmermann – Conceptualization and revision of the manuscript

F. Glorieux – Conceptualization and revision of the manuscript

F. Charbonneau – Recruitment of participants and coordination of the scanning sessions

F. Rauch – involved in the development of scanning protocol, interpretation of the results, and revision of the manuscript

B. Willie – Conceptualization, development of scanning protocol, formal analysis, funding acquisition, project administration, supervision, writing of original draft

Other (non-thesis) publications:

- M. Simon, M. Indermaur, D. Schenk, **S. Hosseinitabatabaei**, BM. Willie, P. Zysset, "Fabric-elasticity relationships of tibial trabecular bone are similar in osteogenesis imperfecta and healthy individuals", Bone, 2022
- N. Mikolajewicz*, EA. Zimmermann*, M. Rummler, **S. Hosseinitabatabaei**, C. Julien, F. Glorieux, F. Rauch, BM. Willie, "Multisite longitudinal calibration of HR-pQCT scanners and precision in osteogenesis imperfecta" (*shared authorship), Bone, 2021.
- FH. Glorieux, B. Langdahl, R. Chapurlat, S. Jan De Beur, VR. Sutton, K. Poole, K. Dahir, E. S Orwoll, BM. Willie, N. Mikolajewicz, E. Zimmermann, **S. Hosseinitabatabaei**, MS. Ominsky, C. Saville, J. Clancy, A. MacKinnon, A. Mistry, K. Javaid, "setrusumab for the Treatment of Osteogenesis Imperfecta: 12-Month Results from the Phase 2b ASTEROID Study"*, In preparation*
- I. Vitienes, N. Mikolajewicz, **S. Hosseinitabatabaei**, A. Bouchard, C. Julien, G. Graceffa, E. Ross, A. Rentsch, T. Widowski, R. Main, BM. Willie, "Pullet genetic strain and rearing housing influence the in vivo skeletal strain patterns during physical activity"*, Under revision in Bone*
- I. Vitienes, **S. Hosseinitabatabaei**, A. Bouchard, C. Julien, A. Rentsch, T. Widowski, R. Main, BM. Willie, "In vivo tibiotarsal axial stiffness is unaffected by genetic strain and rearing housing"*, In preparation*
- P. Asgharzadeh*, A. Birkhold*, EA. Zimmermann, **S. Hosseinitabatabaei**, F. Rauch, O. Rohle, BM. Willie, "Explainable osteogenesis imperfecta disease type classification using a gradient boosting machine learning model" (*shared authorship)*, In preparation*
- L. Gabel, K. Kent, S. Hosseinitabatabaei, AJ. Burghardt, MB. Leonard, F. Rauch, BM. Willie, "Assessing bone density, microarchitecture, and strength in pediatric populations using high-resolution peripheral quantitative computed tomography"*, In preparation*

Other on going (non-thesis) projects:

Building a cadaveric HR-pQCT phantom *(bone sections have been processed)***:**

In contrast to the *in-vivo* precision errors, inherent errors cannot be altered, and need to be quantified to realize the highest achievable precision. This is particularly important in multicenter studies, where cross-calibration between the measurements of different devices, each with their own errors is essential. *Ex-vivo* precision can be calculated by repeatedly scanning a static object such as a bone mimic phantom. These phantoms can also be used to evaluate long-term precision errors since their structure and density are stable over-time. However, bone mimic phantoms lack a realistic bone microstructure, which necessitates the availability of cadaveric bone phantoms. Cadaveric bone phantoms are embedded bone sections that are designed similar to the standard phantoms to avoid motion artifact and repositioning errors. They can be used to directly evaluate the *ex-vivo* precision of bone microstructural measurements. To my knowledge, only one other cadaveric bone phantom exists for HR-pQCT evaluation but it includes a limited range of bone densities from similarly aged donors. The phantom I have constructed contains a large range of bone densities. It will be used to assess the long-term precision of HR-pQCT measurements, as well as cross-calibration in future multicenter trials.

Effect of sclerotic lines on HR-pQCT measurements *(first version of the algorithm is developed and tested on several scans)***:**

Bisphosphonates are the common treatment for OI, especially in growing children. Since BPs inhibit bone resorption, growing OI subjects will develop sclerotic lines. Sclerotic lines are horizontal trabeculae containing some degree of cartilage thought to be caused by the temporary interruption of growth plate cartilage resorption, which can remain in bone even into adulthood. It is thought that the arrangement of cancellous bone in horizontal trabeculae is not optimal from a biomechanical perspective. Sclerotic lines appear as bright areas on images that resemble bone and falsely increase the measured bone mass. The presence of sclerotic lines can reduce the reliability of HR-pQCT scans because these artifacts result in measurements that may not be representative of true bone properties. These artifacts have not been examined in subjects with OI, which can impair correct conclusions from studies and clinical trials on bone properties using HR-pQCT. Therefore, it would be advantageous to determine the bone structure and density with sclerotic lines removed from HR-pQCT scans. This task cannot be accomplished manually because it is extremely time-consuming, subjective, inaccurate, and imprecise. Accordingly, there is a need to detect and remove sclerotic lines using objective and reliable computer algorithms. The

preliminary analysis on 5 scans indicated significant differences in trabecular parameters with changes up to 26% for trabecular bone volume fraction and density.

Correlation between pQCT and QCT strength measurements *(currently recruiting more participants, co-managed by Dr. Suzanne Morin)***:**
Obesity is associated with increased risks of falls and fractures. While individuals with obesity tend to have greater bone mineral density compared to non-obese, their bones are in fact weaker when normalized to their body mass. This indicates that bone density is not reliable for fracture risk prediction for individuals with obesity. Quantitative computed tomography (QCT) is a fast modality for scanning large regions of the axial and peripheral skeleton to estimate strength. However, the most relevant sites at the hip and spine typically expose the subject to high radiation dose. An alternative modality is peripheral QCT (pQCT) is an alternative modality. Although pQCT is limited to scanning small regions at the peripheral skeleton, it has an extremely low radiation dose. Our goal in this study is to investigate the correlation between bone parameters acquired from pQCT and QCT at the radius, tibia, and femur in adults with obesity. Using pQCT (XCT3000-Stratec) and QCT (Philips), we scanned the non-dominant radii, tibiae, and femurs (pQCT 4% and 25% from femoral condyles and QCT femoral neck) of 9 adults with obesity to study the correlations between strength outcomes between these modalities. Our preliminary results showed that bone strength outcomes in adults with obesity are highly correlated between pQCT and QCT, regardless of the scanning region.

Contribution to other projects:

The effect of 2-week bedrest on the distal radius and tibia using HR-pQCT *(managed by collaborators at the University of Saskatchewan. My contributions have been developing the imaging protocol, performing the scans, developing the scripts for image registration, and assistance with the interpretation of the findings):*
Bedrest studies enable us to study disuse-related bone changes (microgravity in space or bedrest due to illness) in a controlled fashion. Rate of bone loss during bedrest is typically larger than those of age related bone loss. Similarly, astronauts lose bone at a rapid rate. Bed rest also enables us to study the effectiveness of countermeasures against disuse-related bone loss. While a few HR-pQCT studies have investigated the effect of 60-day bedrest in healthy adults, understanding of how inactivity and countermeasures alter bone structure and density in older adults is lacking. This is important due to the

clinical relevance of older population, who are most affected by long-term sedentary bed rest, or possible participation of astronauts older than 50 years of age in space missions or space tourism. In this study, we investigated the effects of 2-week bedrest, with or without an exercise countermeasure, on bone microarchitecture and density at the distal radius and tibia in older adults.

Effect of genetic strain and housing condition on the mechanical strain and stiffness of tibiotarsus in egg-laying hens *(Another PhD student's project. My contributions included performing surgeries to attach strain gauges on tibiotarsus of chickens, collecting strain data, in-vivo loading of chickens, and interpretation of some of the results)*:

Mammalian bone volume and structure is primarily dictated by mechanical stimuli. On the other hand, oviparous vertebrates rely on bone as a source of calcium for the mineralization of their eggshells. Egg-laying hens suffer from osteoporosis that could be due to disuse and the calcium demands of laying eggs. In this study, we measured in vivo mechanical strains at the tibiotarsus midshaft during habitual activities in two commercially-relevant genetic strains of chickens that were reared in housing conditions allowing for different levels of physical activity to investigate the effect of genetic strain and level of physical activity on in vivo strains. We further used the strains identified during habitual loading to estimate anabolic loading levels at the tibiotarsus. We then applied daily doses of the estimated anabolic load to the tibiotarsus in a controlled loading condition to study bone remolding due to this loading for different genetic strains and housing conditions.

Chapter 1. Introduction:

High-resolution peripheral-quantitative computed tomography (HR-pQCT) is a promising non-invasive, low radiation imaging tool for measuring bone microstructure and density in the peripheral skeleton. Combined with micro-finite element modeling, HR-pQCT can also be used to estimate bone strength. Studies have shown that HR-pQCT derived bone density, microstructure and strength are predictors of fracture risk. The high resolution of HR-pQCT also offers the opportunity to quantify bone formation and resorption.

There are several technical challenges related to HR-pQCT applications that are not well appreciated and standardized. In order to improve the reliability of HR-pQCT measurements and studies, it is important to acknowledge and address these challenges. Further, considering that HR-pQCT is a relatively recent technology, sufficient data is lacking for many population groups. Some of the challenges and complexities associated with HR-pQCT are general, while some are more specific to particular populations. In my thesis, I will discuss and address some of the general and population-specific challenges associated with HR-pQCT. The population of interest in this thesis is those with osteogenesis imperfecta (OI), which is a devastating collagen-related genetic disorder resulting in bone fragility.

In chapter 2, I will discuss the concepts and literature relevant to my research, and explain the gaps that are motivations for the studies included in this thesis. In chapter 3, the objectives of this thesis are given. Chapters 4, 5, and 6 provide three manuscripts that cover the objectives of my thesis, followed by a general discussion and concluding remarks in chapter 7.

In longitudinal HR-pQCT studies, inaccuracies arise in successive HR-pQCT scans due to repositioning error (shifts along the longitudinal axis or rotational misalignment) that make it challenging to precisely image the exactly same bony region over time. To identify the same bone volumes at each time point, image registration is used. Two main classes of image registration are cross-sectional-area (CSA) and three-dimensional (3D). Each of these methods has its own pros and cons. A major difference between these methods is that the CSA method can only correct for translational misalignments, while the 3D method can also correct for rotational misalignments. Variations of 3D registration exist that may resolve some of its limitations. In chapter 4, a study on investigating the effect of different image registration methods on short-term in vivo precision in adults with OI is presented.

Image registration is also the basis for time-lapse HR-pQCT, an emerging method to quantify bone formation and resorption in humans. However, it is not clear if 3D or matched-angle (MA) registration methods is superior for timelapse analysis. Aside from image registration method, there is no consensus on what input image type to use, as some studies have used grayscale images, while some other studies used binary input images. Further, proper definition of the periosteal mask, and proper methods for minimal noise and error are not well standardized. Finally, there is no proper validation available in the literature. In chapter 5, I present a study as part of a phase-2b multicenter clinical trial, where we used the same-day repeated scans from 29 adults OI to examine the influence of various parameters on HR-pQCT-derived bone formation and resorption to identify the preferred methodology. We then validated the selected method using a combination of repeated and longitudinal scans. We finally used the selected and validated method on clinical data to assess the effect of different doses of an anabolic drug (setrusumab) on bone formation and resorption, as well as net changes in bone at the distal radius and tibia of adults with OI.

While much HR-pQCT data exist for adults, such data is limited for children, especially longitudinal studies. Many of the techniques used for adults to improve reliability, such as image registration, may not be readily applicable to children due to bone growth. Accordingly, longitudinal HR-pQCT studies in children are not well standardized, and the challenges and opportunities are not well understood. Additionally, no longitudinal HR-pQCT data are available for children with OI. Therefore, in chapter 6, I present a study in which longitudinal HR-pQCT data are presented for children with OI, and age- and sex-matched healthy controls. We also performed a feasibility study on using image registration for growing children to clarify its limitations and potentials.

Chapter 2. Background and relevant literature:

2.1. Bone as a hierarchical composite:

Bone is a highly hierarchical structure, with properties that can be analyzed from nano to macro length scales (Figure 1)(Fratzl & Weinkamer, 2007; Rho et al., 1998). At the nano-scale, the building blocks of bone can be divided into organic and inorganic compartments, which constitute nearly 30% and 60% of total bone tissue weight, respectively. The remaining 10% of bone tissue weight is water. The organic phase consists of ~90% type I collagen, while the remaining 10% consists of non-collagenous proteins and lipid (Wagermaier et al., 2015). Type I collagen molecules are created from three polypeptide chains that assemble into a triple helix form. After the cleavage of the non-helical ends of the polypeptides, each molecule is 300 nm in length and 1.23-1.5 nm in thickness. The collagen molecules then self-assemble into a staggered and parallel twist to form a fibril. Because of the 67 nm space between the ends of the collagen molecules and an offset from row to row, gap and overlap zones exist within the fibril, which produces an oscillating surface topography with a characteristic axial repeat pattern called the D-periodicity (

Figure 2). The inorganic or mineral phase is an impure form of calcium phosphate known as hydroxyapatite (Boskey, 2013; Fratzl & Weinkamer, 2007).

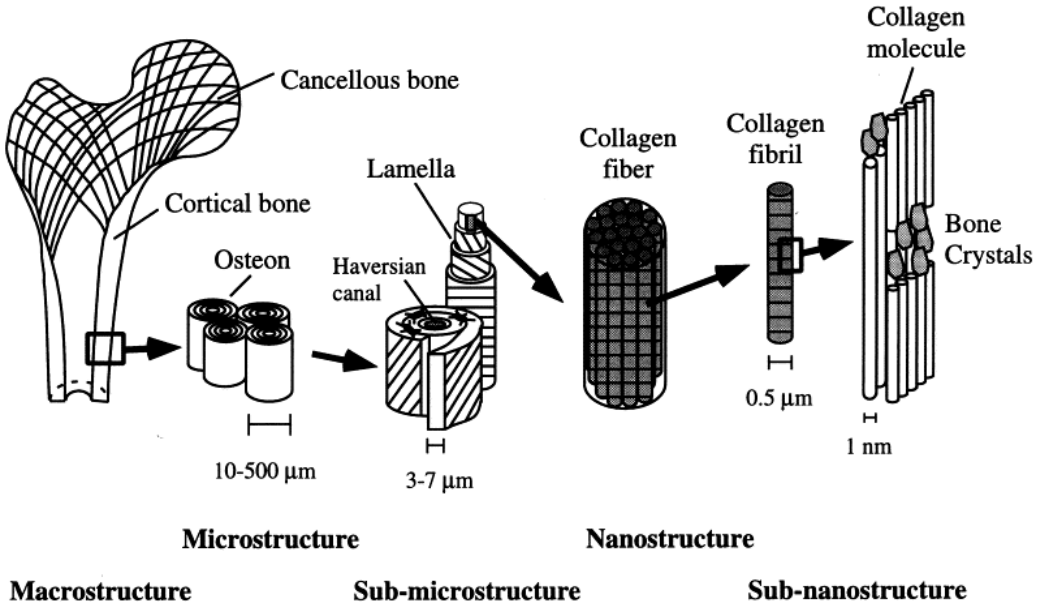


Figure 1. Hierarchical structural organization of bone. At the highest scale, cortical and cancellous (trabecular) bones are found that constitute bone structure. At the micro-scale, osteons are found with Haversian systems and lamellae. At the nano-scale, collagen fibers are found, which are made up of collagen fibrils. Finally, collagen fibrils are composed of bone mineral crystals, collagen molecules, and non-collagenous proteins. Taken with permission from Mechanical properties and the hierarchical structure of bone, Rho et. al., Medical Engineering & Physics, 1998

Bone mineral nucleates within the gap regions between the ends of the collagen fibrils, and along the spaces that run longitudinally between the fibrils, as well as the suggestion that nucleation can start in the overlap zone as well (

Figure 2)(Landis et al., 1993; Rho et al., 1998). The mineral particles deposit on the collagen array and mature into thin plate-like structures which have a space of 67 nm between them. The mineral platelets are mainly parallel to the longitudinal axis of the collagen fibril. The thickness of these mineral particles is in the range of 1.5 to 4 nm, and the lengths varies over a wide range (Fratzl & Weinkamer, 2007; Rho et al., 1998). The mineral particles are not entirely made up of calcium and phosphate, and usually contain different impurities that substitute constituent ions in the lattice or are absorbed onto the crystal surface.

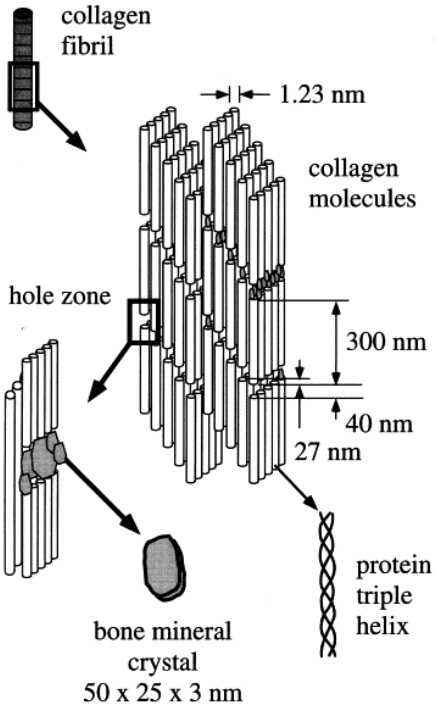


Figure 2. A schematic illustration of the assembly of collagen molecules and bone mineral crystals. The 67 nm periodic pattern results from the presence of adjacent hole (40 nm) and overlap (27 nm) regions of the assembled molecules. Taken with permission from Mechanical properties and the hierarchical structure of bone, Rho et. al., Medical Engineering & Physics, 1998

The mineralized collagen fibrils are between 0.1 and 3 μm thick and constitute the universal building block of bone (Rho et al., 1998). At the next level, the mineralized fibrils form arrays with diameters in the range of 1 to 10 μm (Fratzl & Weinkamer, 2007). These arrays can be in a variety of patterns such as parallel fibril arrays, woven fiber structures, plywood-like structures and radial fibril arrays. The organization of the arrays is characteristic for different structures of bone, including lamellar bone, woven bone and fibrolamellar bone (Figure 3)(Wagermaier et al., 2015).

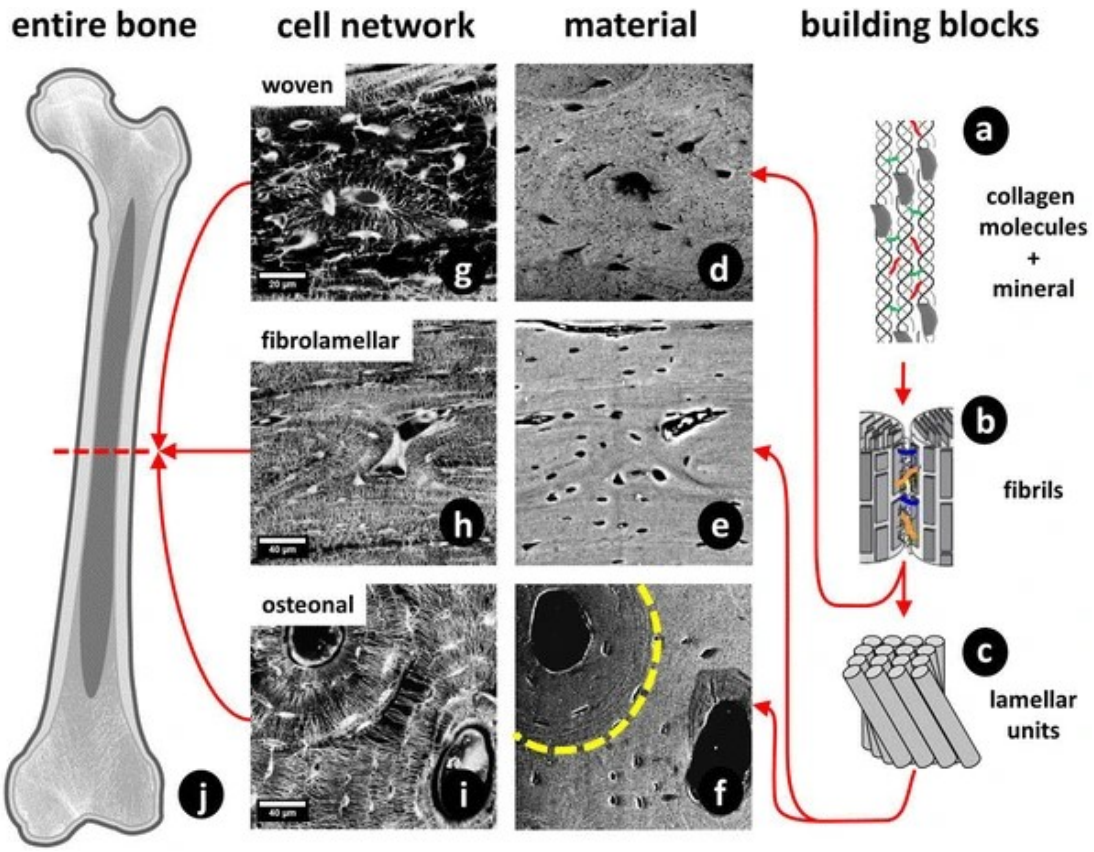


Figure 3. Hierarchical organization of bone. At the smallest scale, collagen molecules, connected by crosslinks, and embedded mineral particles are found (a). Collagen molecules assemble to create fibrils (b). The fibrils can assemble into lamellar units (c) or directly into woven bone (d, g). The lamella unit can either create fibrolamellar bone (e,h) or osteonal (lamellar) bone (f, i). Taken with permission from Fragility of Bone Material Controlled by Internal Interfaces, Wagermaier et. al., Calcif Tissue Int, 2015

Woven bone can be defined as randomly oriented fibril arrays, and is typically found in rapidly formed bone such as during fracture healing to act as a transient basis for subsequent formation of organized bone (Reznikov et al., 2014). The more organized bone is typically lamellar bone that replaces woven bone or old lamellar bone through remodeling processes. Lamellar bone is commonly formed from layers of fibril arrays stacked in a rotated plywood design. The fibrils are alternating in orientation around an axis perpendicular to the layers. The lamellar bone can be found in two main forms. In the first form, lamellar bone is concentrically arranged around a canal that contains blood vessels and nerves. This structure is known as an osteon, and runs along the long axis of bone (Figure 4). Osteons are categorized as primary formed during initial bone

formation and as secondary osteons, also called Haversian osteons, formed during remodelling of existing bone. Osteons are usually found in cortical bone. The second form of lamella bone is typically found in trabecular bone, where lamellae are organized parallel to a common direction (Figure 4)(Reznikov et al., 2014).

At the macro scale, human bone tissue is generally classified as cortical and trabecular (cancellous) (Figure 4). Cortical bone is dense with a porosity of nearly 6% (Fratzl & Weinkamer, 2007), and is found primarily in the shaft of long bones and the outer shell around trabecular bone at the proximal and distal ends of bones and the vertebrae. Trabecular bone has a spongy architecture with a porosity of nearly 80% (Fratzl & Weinkamer, 2007), and is located within cortical bone, in medullary cavities at the ends of long bones, and in the interior of short bones, such as spinal vertebrae.

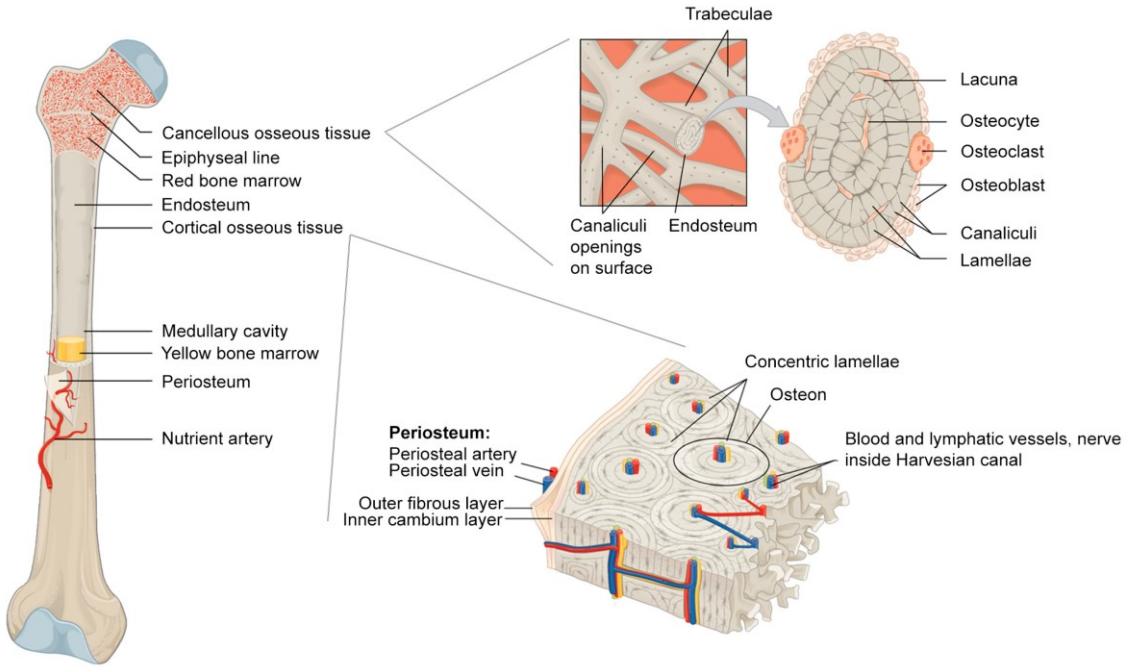


Figure 4. Schematic view of cortical and trabecular bone in human long bone. Cortical bone is consists o osteons parallel to the bones long axis, while trabecular bone has a network within the medullary cavity and follows the direction of strain. Adapted from The Components of Bone and What They Can Teach Us about Regeneration, Le et. al., Materials, 2018

Bone material typically combines sufficient stiffness and strength with high toughness. Stiffness is related to the ability of bone to resist bending, while toughness is the capacity of bone to absorb energy during impact to prevent bone fracture. Stiffness and toughness are contradictory

(i.e., high value of one sacrifices the value of the other) (Ritchie, 2011). In materials such as ceramics, high stiffness and strength (i.e., resistance to failure) are often accompanied by low toughness, whereas materials with high toughness usually have low stiffness and are deformable. The mineral phase of bone provides stiffness, while the organic matrix reduces the inherent brittleness of the mineral and provides toughness (Burr, 2002).

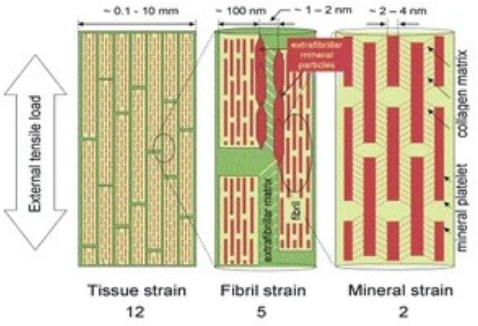
In bone, the overall stiffness can be estimated from the average of the local values of the stiffness. On the other hand, toughness and strength cannot be calculated just based on the corresponding values of the constituents, but rather depend on the nucleation (i.e., formation) and crack propagation (Wagermaier et al., 2015). These two factors depend on tiny defects in the materials such as pores, microdamage, interfaces, and fiber directions. Hence, to reduce fragility, bone has mechanisms at different scales to dissipate energy, such that less energy will be available for crack nucleation and propagation to minimize the risk of fracture (Figure 5)(Wagermaier et al., 2015).

At the smallest scale (levels 1 and 2 on Figure 5), shear deformation between mineral and collagen as well as between collagen fibrils is shown to dissipate energy and contribute to toughness (Himadri S. Gupta et al., 2005, 2006). At level 3, the formation of additional micro-cracks ahead of crack tips reduce the energy available to propagate the crack (H. S. Gupta & Zioupos, 2008; Zioupos & Currey, 1998). At level 4, the periodic variation of elastic modulus reduces the crack driving force by trapping the crack in repeatedly decreasing and increasing moduli (Fratzl et al., 2007; O. Kolednik et al., 2014; Otmar Kolednik et al., 2011). At level 5, the energy of crack is dissipated by crack deviation between the lamellae and the formation of microcracks. The deviation of cracks occur at relatively weak interfaces, such as between lamellae or at cement lines (Koester et al., 2008; Nalla et al., 2006; Peterlik et al., 2006). Finally at level 6, toughness is also increased by fibers bridging the crack by nucleating new cracks that can dissipate energy (Nalla et al., 2003).

Figure 5. Six identified bone toughening mechanisms from smallest to largest scales. Levels 1 and 2 show that shear deformation between mineral and collagen and between collagen fibrils as a toughening mechanism. At the bottom of each graph, the magnitude of relative magnitude of the stains are shown. Level 3 shows the penetration of rhodamine stain (black arrows) into micro-cracks that were formed ahead of crack tip as energy dissipating mechanism. The yellow line shows the splitting and deviating crack. Level 4 indicates the periodic changes of the indentation modulus within successive lamellae. Level 5 shows crack deviation across lamellar bone. Level 6 shows crack bridging by uncracked ligaments (black arrows). Taken with permission from Fragility of Bone Material Controlled by Internal Interfaces, Wagermaier et. al., Calcif Tissue Int, 2015

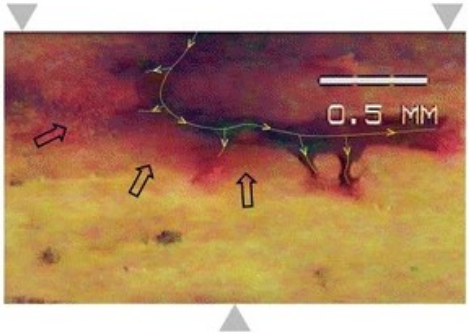
Level 1 and 2:

Plasticity of fibril – mineral and fibril – fibril interfaces



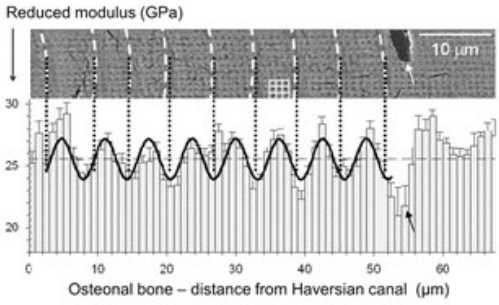
Level 3:

Microcracking ahead of crack tip



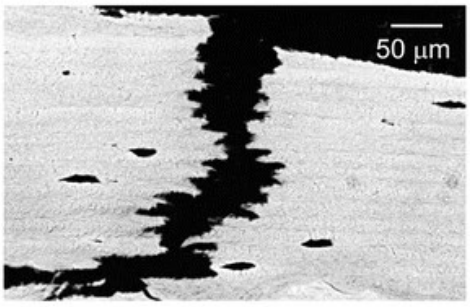
Level 4:

Periodic modulation of material properties



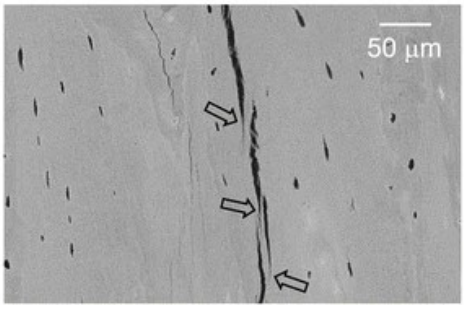
Level 5:

Crack deviation



Level 6:

Crack bridging



2.2. Bone as a living organ:

Aside from the mineral particles, collagen matrix, and water, several cell types can be found in bone as a living organ. These cells enable the bone to undergo continuous changes, regeneration, repairing microdamage and adaptation to changing environmental stimuli. Three types of bone cells can be identified: osteoblasts, osteoclasts and osteocytes (Bellido, T., Plotkin, L. I., & Bruzzaniti, 2019). Osteoblasts are responsible for bone formation by secreting a collagenous bone matrix (known as osteoid), followed by mineralization with hydroxyapatite particles (Bellido, T., Plotkin, L. I., & Bruzzaniti, 2019). Some of the osteoblasts become embedded into the mineralized bone and become osteocytes (Bellido, T., Plotkin, L. I., & Bruzzaniti, 2019). Osteocytes are the most populous bone cells and are found within the bone in lacunae and form a network with neighbouring cells via their dendrites which pass through small channels, called canaliculi (Bellido, T., Plotkin, L. I., & Bruzzaniti, 2019). Osteoclasts are the third type of cells and are responsible for bone resorption. Together, the three cell types work together to perform bone modeling, changes the shape of bone during growth and remodeling, to maintain skeletal integrity throughout the life (Figure 6). Osteocytes are known as the cells that coordinate the activity of osteoblasts and osteoclasts in response to environmental cues such as mechanical stimuli or hormonal changes by secreting factors such as sclerostin, receptor activator of nuclear factor kappa-B ligand (RANKL) or osteoprotegerin (OPG)(Bellido, T., Plotkin, L. I., & Bruzzaniti, 2019; Bonewald, 2011; Robling et al., 2008).

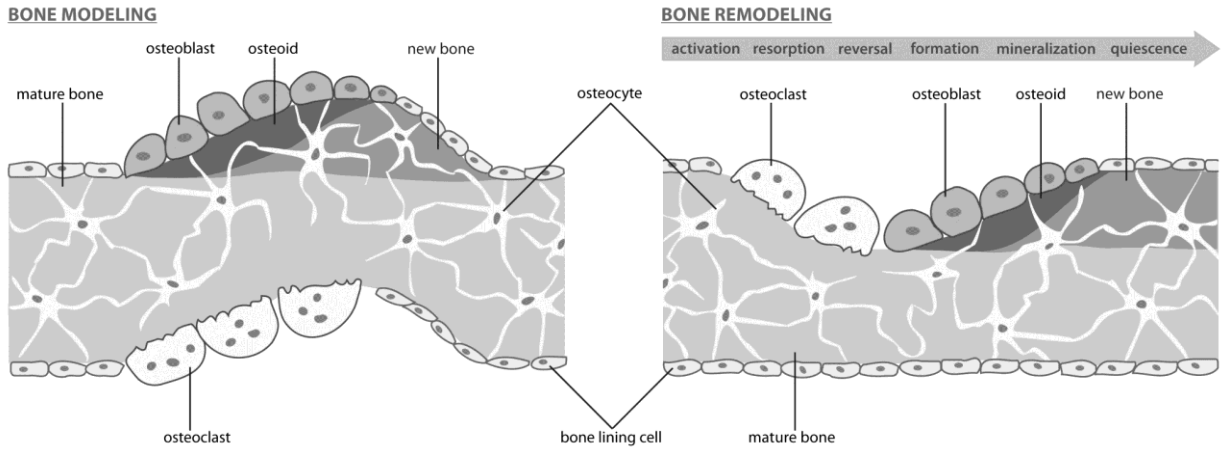


Figure 6. Schematic illustration of bone modeling (left image) and bone remodeling (right image). In bone modeling, bone formation and resorption are uncoupled. While osteoblasts are laying

down bone on the top of the bone, osteoclasts resorbing it on the other side. In bone remodeling, formation and resorption are spatially and temporally coupled. After an activation signal, osteoclasts resorb the old bone, followed by osteoblasts laying down osteoid, which is later mineralizes into new bone. Taken with permission from Chapter 25: Registered Micro-Computed Tomography Data as a Four-Dimensional Imaging Biomarker of Bone Formation and Resorption, Annette I. Birkhold and Bettina M. Willie, Biomarkers in Bone Disease, Vol. 1, 2016

The adaptation of bone in response to different stimuli was first proposed in the 19th century by Julius Wolff, who observed that the trabecular patterns followed the trajectory lines of principal stress in weight-bearing bones of the lower-limbs (Brand, 2011). Later, Harold Frost proposed the mechanostat theory for the adaptation of bone (Figure 7)(Frost, 1987, 1994). Based on Frost's theory, bone operates in a negative feedback system, bone mass and architecture are locally controlled by bone cells in order to maintain the mechanical strain in bone at a target set point. Accordingly, bone is formed when the actual strain exceeds the set point, and resorbed when the actual strain is below the set point, bringing the strain to the target set point. Frost's theory defined an equilibrium range (adapted window in Figure 7) of strain magnitudes that would not results in bone response, such that certain minimum effective strains above and below equilibrium range must be surpassed to elicit a formation or resorption response in the form of bone modeling or remodeling. In a similar model proposed by Beaupre et. al., (Beaupré et al., 1990) this adapted window was named lazy or dead zone. Modeling is defined as spatially and temporally uncoupled and independent bone formation and resorption that is mainly responsible for bone growth (Robling, A. G., Daly, R., Fuchs, R. K., & Burr, 2019; Willie et al., 2020). In contrast, remodeling describes a process, in which resorption followed by formation are temporally and spatially coupled. Remodeling occurs throughout ones lifetime, and is mainly responsible for maintaining bone mass and replacing damaged tissue (Robling, A. G., Daly, R., Fuchs, R. K., & Burr, 2019; Willie et al., 2020).

Frost's theory is based on several strain thresholds (Figure 7). If the effective strain reaches an upper or a lower threshold, bone adapts its mass and architecture (Figure 7)(Frost, 1987; Robling, A. G., Daly, R., Fuchs, R. K., & Burr, 2019; Willie et al., 2020). For effective strain below the lower threshold, bone is resorbed to reduce stiffness. If the mechanical strain due to loading exceeds a larger upper threshold, bone is formed via modeling process (Frost, 1987; Robling, A. G., Daly, R., Fuchs, R. K., & Burr, 2019; Willie et al., 2020). For effective strains that are below the modeling threshold, bone undergoes continuous remodeling to repair microdamage

(Figure 7)(Frost, 1987; Robling, A. G., Daly, R., Fuchs, R. K., & Burr, 2019; Willie et al., 2020). For very high effective strains, bone will develop microdamage, diffuse damage, or eventually fail.

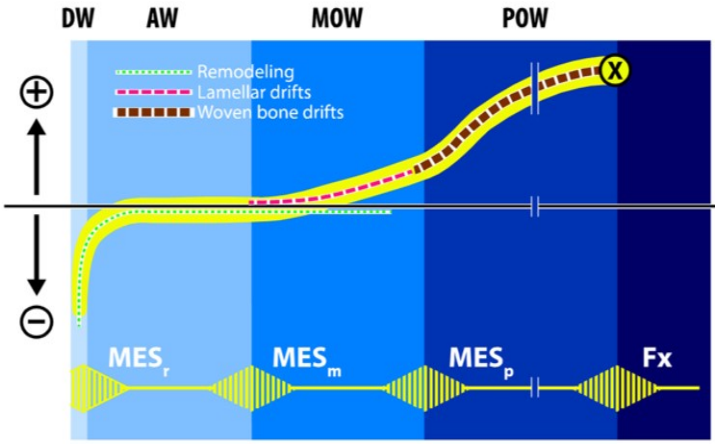


Figure 7. The mechanostat theory proposed by Frost. The horizontal axis depicts peak bone strain and the vertical axis either net loss (-) or gain (+) of bone mass. The yellow pulsed line highlights the threshold values of minimum effective strain (MES) for remodeling (MES_r), modeling (MES_m), microdamage accumulation (MES_p) and fracture strain (Fx). Depending on the strain levels, bone loss occurs if disused (DW), bone maintenance during normal use (adapted window, AW), bone gain while increasing the use (mild overuse window, MOW) and bone damage if overused (pathological overuse window, POW). This figure was adapted with permission from: The murine axial compression tibial loading model to study bone mechanobiology: Implementing the model and reporting results, Russell P. Main, et. al., Journal of Orthopaedic Research, 2019

Bone remodeling is fundamentally the same in cortical and trabecular bone compartments, which is executed by bone basic multicellular unit (BMU)(Parfitt, 1994). BMU comprises active osteoclasts and active osteoblasts, and is a transient functional groupings of these cells that progress through the bone, remove old bone and replace it with new bone. In both the cortical and trabecular bones, a BMU maintains its size, shape, and individual identity as it moves through or across the bone, by continuously recruiting new cells (Parfitt, 1994). While the activity of BMU is the same between cortical and trabecular compartments, the overall remodeling process is geometrically different between the two (Parfitt, 1994). Cortical bone remodeling time is typically in the form of intracortical remodeling, where the process originates on the wall of a Haversian or Volkmann canal, and the necessary cells travel to the selected site via the circulation. However, if the site of origin is adjacent to hematopoietic bone marrow, the necessary cells could migrate to the bone surface directly without entering the circulation. This scenario is similar to the remodeling occurring in trabecular bone, as it is surrounded by bone marrow (Parfitt, 1994).

The remodeling process for cortical bone is named osteonal remodeling (Figure 8), which corresponds with the process in which BMU create tunnels along the long axis of the bone (i.e., along the osteons). The tip of the BMU tunnel are osteoclasts resorbing bone, and the other end of the tunnel are osteoblasts laying bone around a circular cross-section (Parfitt, 1994).

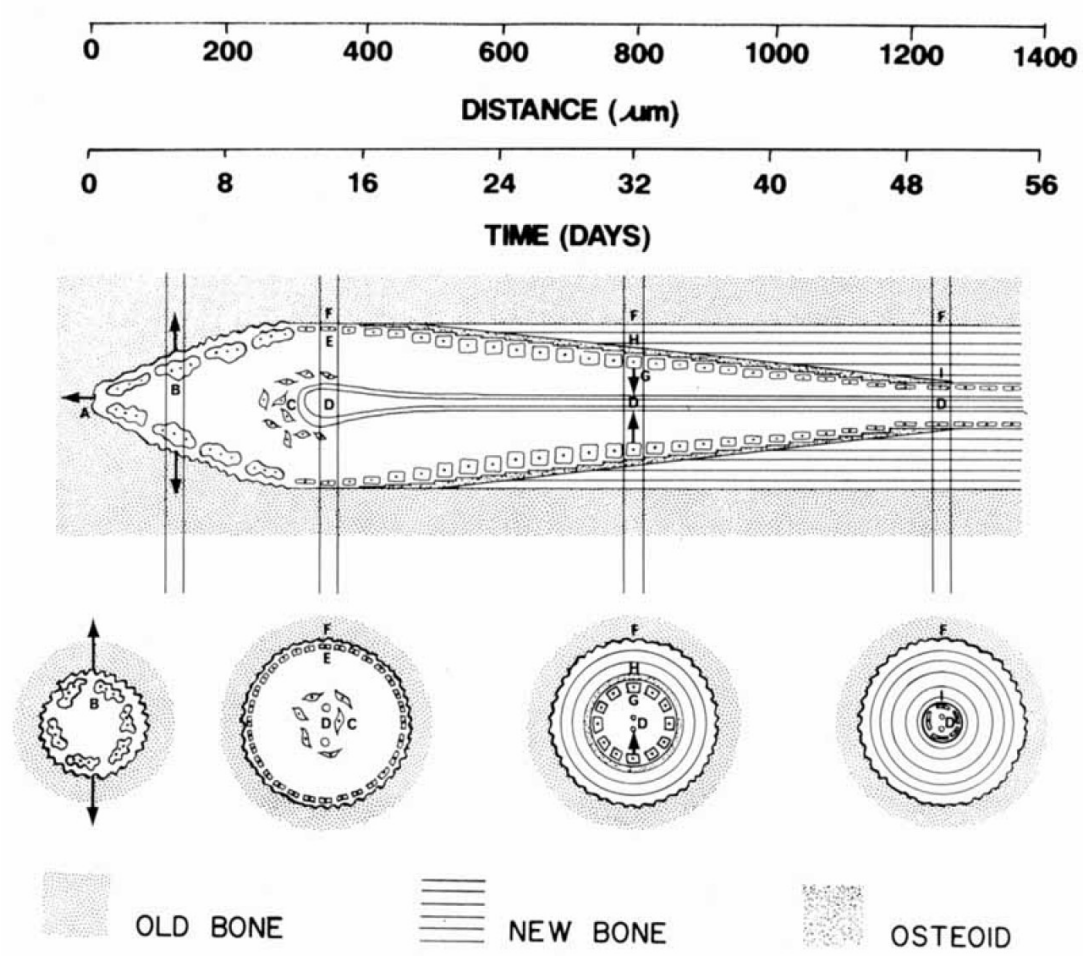


Figure 8. Schematic diagram of cortical BMU. The top figure indicates a side view of the 3D BMU, while the circular cross-sections correspond to the different stages of remodeling as shown by parallel lines. The two top axes indicate distance and time scales based on advance of 25 μm/day. A) apex of cutting cone moving from right to left; B) osteoclasts resorbing bone in a centrifugal manner, although forward erosion is just as likely; C) location of dividing precursors of preosteoclasts and preosteoblasts; D) capillary loop; E) the reversal zone; F) cement line between the new and old bone; G) osteoblasts forming bone toward center; H) osteoid; I) lining cells at periphery of canal of completed Haversian system. Adapted from Osteonal and Hemi-Osteonal Remodeling: The Spatial and Temporal Framework for Signal Traffic in Adult Human Bone, Parfitt, Journal of Cellular Biochemistry, 1994

The remodeling for trabecular bone is named hemi-osteonal remodeling (Figure 9), which suggests that its shape is similar to a half of the cortical BMU. Although visualization of remodeling in trabecular is more difficult due to varying orientation of single trabeculae, each BMU can be considered a BMU tunnel that is cut in half, and has a semi-circle cross-section, while the process is the same as cortical BMU. Accordingly, the resorbed regions appear as trenches, rather than tunnels (Parfitt, 1994).

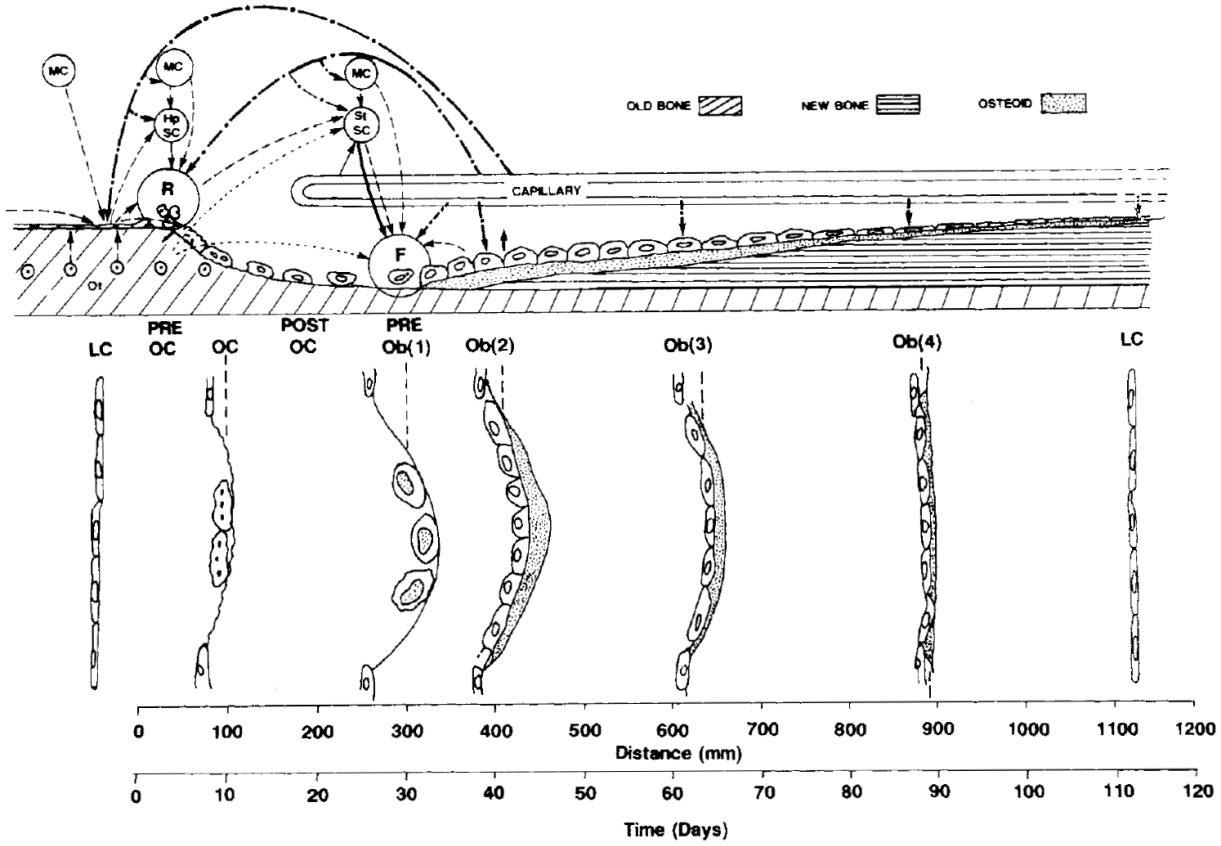


Figure 9. A diagram of hemi-osteonal remodeling. The top part of the figures indicates a cancellous BMU is and the lower part shows selected transverse sections. The moving direction is from right to left, digging a trench across the surface. The distance and time scales assume longitudinal advance at 10 μm/day, which is based on an estimated total BMU length of 1,200 μm and an estimated time for completion at a full remodeling cycle of 120 days. The solid arrows indicate the pathways of cell recruitment and movement. R = region of preosteoclast recruitment as the initiation of resorption; F = region of preosteoblast recruitment and the initiation of formation; LC = lining cell; Ot = osteocyte; Oc = osteoclast; Ob. = osteoblast; HpSC = hematopoietic stem cell; StSC = stromal stem cell; MC = unspecified marrow cell. Adapted from Osteonal and Hemi-Osteonal Remodeling: The Spatial and Temporal Framework for Signal Traffic in Adult Human Bone, Parfitt, Journal of Cellular Biochemistry, 1994

2.3. Osteogenesis Imperfecta (OI):

Osteogenesis imperfecta (OI), or brittle bone disease, is a collagen-related genetic disorder resulting in bone fragility (Marini, 2018). OI could occur in approximately 1 in 15,000-20,000 births (Oakley & Reece, 2010). At least 18 genetic mutations can lead to an OI phenotype (Tauer et al., 2019). However, in most cases (~85%), OI is caused by dominant mutations in either the COL1A1 or COL1A2 gene, which code for collagen type I. The 2019 Nosology and Classification of Genetic Skeletal Disorders distinguishes four phenotypical OI types caused by these mutations: type I - mild, type II - perinatally lethal, type III - severe and type IV – moderate (Mortier et al., 2019). These mutations occur during the production of collagen in osteoblasts and can result in quantitative defects (i.e., reduced amount of normal type I collagen production) or qualitative defects (i.e., production of collagen molecules with altered structure). Quantitative defects result in lower bone mass with normal quality and are associated with milder osteogenesis imperfecta such as type I, while qualitative defects can cause the more severe types of OI such as type II, III, and IV (Hald et al., 2016; Marini, 2018).

Individuals with OI usually experience many fractures during their life, especially in severe types. The low bone strength is complex and can be studied in multiple length scales. The brittle and weak structure of bone in OI subjects originates from brittle bone tissue properties (tissue level properties) and deteriorated structural and geometrical properties (macro-level properties) (Fratzl-Zelman et al., 2014; Joan C Marini, Antonella Forlino, Hans Peter Bächinger, Nick J Bishop, Peter H Byers, Anne De Paepe, Francois Fassier, Nadja Fratzl-Zelman, Kenneth M Kozloff, Deborah Krakow, Kathleen Montpetit, 2015; Nijhuis et al., 2019). OI bone tissue commonly exhibits increased mineralization with smaller and thinner platelets that are densely packed (Fratzl-Zelman et al., 2014). One theory related to lower strength in OI bones is that some of the toughening mechanisms are not optimal due to increased mineralization and collagen defects. More specifically, it is likely that less pronounced lamellar bone compromises toughening mechanisms based on periodic modulation of the indentation modulus within successive lamellae (Level 4 in Figure 5)(Wagermaier et al., 2015), as well as crack undulating deviation across lamellar bone (Level 5 Figure 5)(Wagermaier et al., 2015). Hence, the putative inability of OI bone to remodel primary woven bone into a high-quality lamellar structure may be a major source of lower strength,

rather than collagen defect itself. This has been reported in several studies (Carriero et al., 2014; Fratzl-Zelman et al., 2014).

Interestingly, two recent studies have shown that micro-compressive and micro-tensile properties of OI bone are not inferior compared to those of healthy bone. Based on the results of micro-compressive testing, Indermaur et. al.,(Indermaur et al., 2021) report similar elastic properties but higher ultimate stress in OI bone compared to healthy bone. One possible explanation for this observation was that most bone fractures are initiated in tensile loading. Therefore, a followup study was conducted to assess micro-tensile properties of OI bone. Interestingly, tensile properties of OI bone were not inferior to those of healthy controls. However, a consistent relationship was observed between tensile properties and the degree of tissue mineralization and collagen fiber orientation, which can in fact be in line with the different levels of toughening mechanisms. Together with the same observation for compressive properties in OI and control bones, these studies suggested that the observed bone fragility in OI bones may be mostly driven by lower bone mass and deteriorated architecture. However, as the authors pointed out, their study had a limited sample size, and the test was different from real in vivo conditions.

In summary, the fragility of OI bones is complex and multifactorial. Overall, the altered geometry and inadequate physical properties of the bone matrix work together to decrease load bearing capacity of OI bone.

2.4. Pharmacological therapy for OI:

There are currently no curative treatments for OI. OI treatment is focused on decreasing pain and fractures and increasing bone mass and mobility. Most of the pharmacological approaches tested on people with OI were originally developed to treat osteoporosis by increasing bone mass. The pharmacological treatments include: bisphosphonates, Denosumab, growth hormone, Teriparatide, TGF-β inhibitor, sclerostin-inhibitory antibodies, and combined therapy (Figure 10) (Marom et al., 2016). While some of these drugs are used off label for OI, there currently exists no FDA/EMA/Health Canada approved drug therapy for this population. Bone treatments can increase bone mass either by decreasing bone resorption (anti-catabolic) or increasing bone formation (anabolic). Combined therapy is another strategy that benefit from the synergistic effect of anabolic and anti-catabolic treatments. Bisphosphonates are anti-catabolic treatments that are

commonly used for individuals with OI. Bisphosphonates are reported to reduce fractures and pain in children with OI (Land et al., 2006; Plotkin et al., 2000), while they are less effective in adults (Marom et al., 2016). Because bisphosphonates stop bone resorption, they can result in sclerotic lines in growing children, which are horizontal trabeculae containing some degree of cartilage thought to be caused by the temporary interruption of growth plate cartilage resorption (Rauch et al., 2004). One concern with sclerotic lines is reduced reliability of bone computed tomography (CT) measurements, as sclerotic lines resemble dense bone and result in over-estimated bone mass measurements.

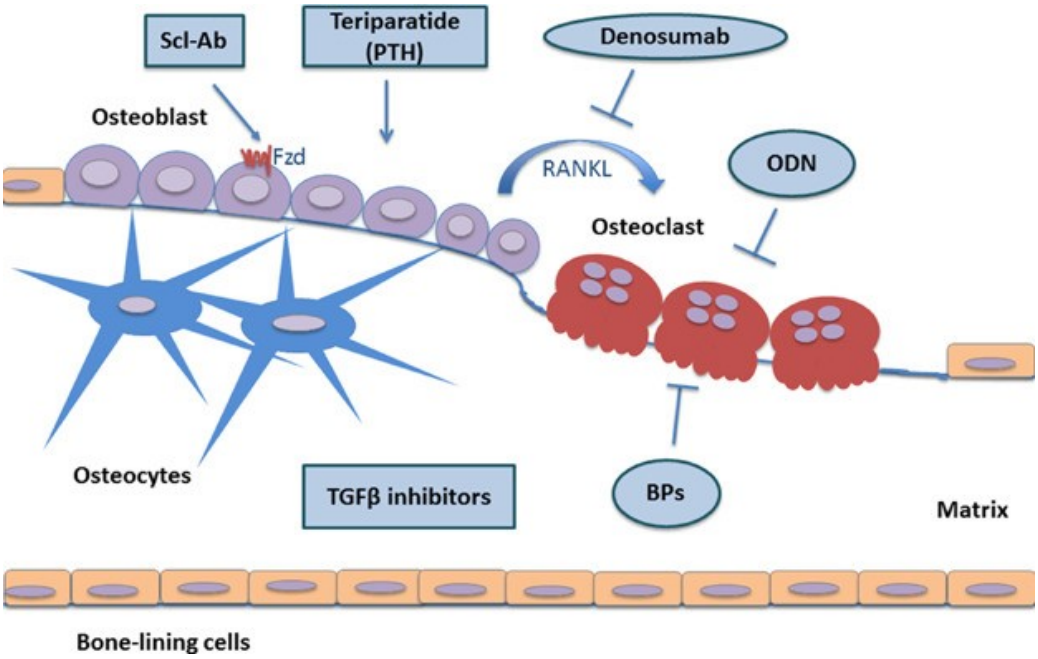


Figure 10. Schematic illustration of the target of therapeutic interventions. Bisphosphonates (BPs) reside in the bone matrix and inhibit osteoclast activity. Cathepsin K inhibitors such as Odanacatib (ODN) also inhibit the resorptive activity of osteoclasts. Denosumab targets the receptor activator of nuclear factor kappa-B ligand (RANKL), to inhibit the formation and activation of osteoclasts. Teriparatide promotes osteoblast differentiation and activity. Sclerostin inhibitory antibodies (Scl-Ab) increase bone formation via Wnt signaling. Adapted from Pharmacological and biological therapeutic strategies for osteogenesis imperfecta, Marom et. al., Medical Genetics, 2016

A recent group of anabolic treatments developed are sclerostin-inhibitory antibodies (Scl-Ab). Sclerostin is a protein that reduces bone formation by inhibiting WNT/b-catenin signaling in osteoblasts and is predominantly secreted by osteocytes. Several preclinical studies on OI mice models have shown strong anabolic effect of sclerostin neutralizing antibody. These studies collectively showed improved trabecular and cortical bone mass, and increased stiffness and

strength, suggesting Scl-Ab is a promising treatment for OI (Grafe et al., 2016; Lee et al., 2017; Nijhuis et al., 2019; Roschger et al., 2014; B. P. Sinder et al., 2014; Benjamin P. Sinder et al., 2013, 2015). Scl-Abs (Romosozumab) have been approved by FDA to treat osteoporosis (McClung et al., 2014; Recker et al., 2015). In an open-label, phase 2a trial, pharmacodynamics and safety of a sclerostin neutralizing antibody (setrusumab) was investigated on 14 adults with moderate OI (Glorieux et al., 2017). That study showed that setrusumab stimulates bone formation, reduces bone resorption, and increases lumbar spine aBMD in adults with moderate OI, with no treatment-related adverse events and fractures. This led to a recently completed phase 2b, multicenter, multinational, double-blind, dose finding clinical trial on setrusumab. Some data used throughout this thesis are from this clinical trial.

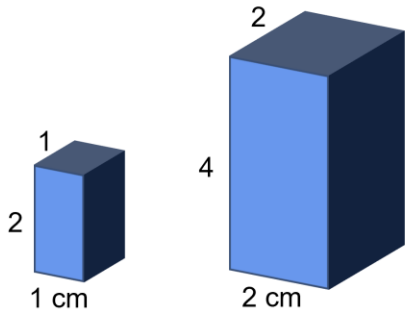
2.5. Non-invasive assessment of bone:

The 'gold standard' method for measuring bone strength is mechanical loading until failure, which is not feasible clinically. Therefore, different non-invasive imaging modalities have been developed to predict bone strength. Radiography is the most common method used to non-invasively evaluate bone. Radiography involving x-rays is based on the attenuation of photon beams based on the electron density (atomic number which in turn depends on mass density) of the body tissues. Bone mineral highly attenuates x-rays due to its high electron density, whereas soft tissues (e.g. fat and muscle) have low electron densities and are low attenuating tissues. The differences in the attenuation of different tissues is the key to separating bone from other tissues. A disadvantage of x-rays is the radiation exposure to the subject, while other methods such as magnetic resonance imaging (MRI) and ultrasound do not irradiate the participants. However, x-ray methods are more informative and interpretable and they provide high contrast for bone, and can be calibrated against bone density. The key x-ray modalities are summarised below.

2.5.1. Dual-energy x-ray absorptiometry (DXA):

Dual-energy x-ray absorptiometry (DXA) is the current standard method to estimate bone fragility through measures of areal bone mineral density (aBMD), which was first introduced in 1987 by Hologic. As the name suggests, two x-ray beams with different energy levels are focused on the target site on body. A single x-ray beam is attenuated by both soft tissue and bone, and it is not possible to determine, how much attenuation was attributable to the bone. On the other hand, the attenuation coefficients and their ratio change with the energy of the x-ray beam. DXA uses the

difference in total absorption between the two beams with different energy levels to subtract out the absorption by soft tissue, leaving just the absorption by bone, which is related to bone density. Since the DXA projection is a 2D image, BMD measured is usually referred to as areal BMD (aBMD). DXA is highly accessible at most hospitals, has low effective radiation dose, and is capable of measuring BMD at the axial and peripheral skeleton for estimation of fracture risk and monitoring of aBMD changes over time. Some non-BMD measurements can be acquired using DXA images, such as vertebral fracture assessment, hip structural analysis, trabecular bone score and the measurement of body composition. However, DXA has several limitations. First, it cannot separate cortical and trabecular compartments because the image indicate a projection of 3D bones in 2D, and due to low resolution. Second, DXA aBMD measurements and their precision are dependent on soft-tissue thickness (Caksa et al., 2019). Another limitation is related to the two-dimensional nature of DXA measurements, as bone depth is not accounted for, which can underestimate and overestimate aBMD in small and large bones, respectively (Figure 11).



Volumetric bone mineral density (g/cm^3)	1.0	1.0
Projected Area (cm^2)	2.0	8.0
Volume (cm^3)	2.0	16.0
BMC (g)	2.0	16.0
Areal bone mineral density (g/cm^2)	1.0	2.0

Figure 11. Effect of size on measured bone mineral parameters. Using DXA, the region of interest is the projected area, which is equal to the area of the front face of the sample. The bone mineral content (BMC) is the total amount of bone mineral (g) in the sample. An areal density (g/cm^2), bone mineral density (BMD), is calculated as BMC over projected area. While both samples have the same volumetric bone density, the areal BMD measured by DXA is larger for the larger sample. Reproduced from New Approaches for Interpreting Projected Bone Densitometry Data, Carter et. al., JBMR, 1992

2.5.2. Quantitative computed tomography (QCT):

Quantitative computed tomography is an approach that enables quantitative analysis of bone density by calibrating Hounsfield units with bone density using a standard phantom. Unlike DXA, quantitative computed tomography (QCT) can obtain volumetric (3D) measurements of bone, such as volumetric BMD (vBMD) and bone geometry. QCT can also obtain some measurements separately for the trabecular and cortical compartments; although the accuracy of these measurements is dependent on the slice thickness and pixel size (Zysset et al., 2015). Scanning times for QCT are quite short, on the order of a few seconds. The QCT measurements of femoral neck vBMD have been used for hip fracture risk predictions (e.g., FRAX score)(Yang et al., 2012). A major limitation with QCT is the high radiation dose due to the greater amount of soft tissue at the central sites. Further, despite being able to separate trabecular and cortical compartments, spatial resolution is not high enough to fully resolve bone microstructure. Considering the mentioned limitations and higher prices for QCT devices, this modality is not as widely used as DXA, despite being developed earlier than DXA in 1970s (Zysset et al., 2015).

2.5.3. Peripheral QCT (pQCT):

Peripheral QCT (pQCT, Stratec) is a compact QCT device dedicated to scanning peripheral bones. The typical scanning protocol for pQCT is to acquire single slices of 2 mm thickness with an in-plane voxel size of 200-800 μm at different regions of the limb (A. Wong, 2016). These regions are selected such that they represent metaphyseal and diaphyseal bone, as well as muscle. Similar to QCT, pQCT can separate cortical and trabecular compartments. An advantage of pQCT is the low effective radiation dose due to the low amount of soft-tissue (~1 μSV per slice). On the other hand, pQCT has several limitations. First, the spatial resolution is not high enough to fully resolve bone microstructure. Further, pQCT averages bone measurements through the thickness of the slices,. Therefore, it can eliminate some of the variations, especially at the transitioning zone at the distal radius (A. Wong, 2016). Finally, variations in participant repositioning can cause measurement imprecision, as different scans may capture slightly different bone regions (Rinaldi et al., 2011; Swinford & Warden, 2010; A. K. O. Wong et al., 2015).

2.5.4. High-resolution peripheral Quantitative Computed Tomography:

In contrast to pQCT, high resolution peripheral quantitative computed tomography (HR-pQCT, Scanco) is emerging as a powerful non-invasive bone imaging modality capable of assessing volumetric BMD, microarchitecture and strength, and distinguishing cancellous and cortical bone in-vivo at a voxel size of 82 μm (first generation XtremeCT scanner) or 61 μm (second generation XtremeCTII scanner). Typical scanning regions have a length of ~10 mm and are located at distal radius and distal tibia. The scanning time is about 3 minutes, and the radiation dose is approximately 3 μSv per scan, which is low compared to the annual radiation dose of 2.4 mSv for individuals (Nishiyama & Shane, 2013). HR-pQCT resolves most of the limitations associated with pQCT. The high resolution of HR-pQCT can resolve bone microarchitecture. The slice thickness in HR-pQCT is 61 (XtremeCTII) or 82 (XtremeCT) μm, which is much finer than 2 mm slice thickness of pQCT. Further, since the scanning volume is ~10 mm, and is formed by a stack of 110 (XtremeCT) or 168 (XtremeCTII) slices, it allows for image registration to alleviate repositioning errors in longitudinal scans.

2.6. Evaluation of bone structure, strength, and changes using HR-pQCT:

The development of HR-pQCT has improved the *in vivo* assessment of bone microarchitecture, density, and strength in the peripheral skeleton, most commonly the distal radius and tibia. HR-pQCT has been widely used in multiple cross-sectional and longitudinal studies on different populations to understand the effect of aging, diseases, growth, physical activity and micro-gravity, and to study fracture healing (Gabel et al., 2021; Geusens et al., 2014; Whittier et al., 2020). Owing to the resolution of HR-pQCT, bone microstructural parameters such as cortical and trabecular thickness, cortical porosity, trabecular number, and trabecular separation can be measured (Figure 12). The first-generation HR-pQCT scanner has a nominal isotropic resolution of 82 μm, which is at the limit for measuring the thickness of individual human trabeculae (Laib & Rüegsegger, 1999; Joshua A MacNeil & Boyd, 2007), so a method was developed to derive trabecular microarchitectural measurements indirectly, termed the 'derived approach'. This method works based on trabecular BMD and ridge extraction to measure trabecular number. The second-generation HR-pQCT scanner (XtremeCTII) can assess bone microarchitecture at peripheral limbs at a 61 μm nominal isotropic voxel size, which enables the direct extraction of

structural outcomes using an approach based on the distance transformation. An advantage of the direct assessment of trabecular microarchitecture is that the morphological measurements are independent of measured density. Studies have shown that the major difference between the scanners is for trabecular thickness (Manske et al., 2015). When combined with micro-finite element (microFE) method, HR-pQCT can be used to non-invasively estimate bone strength *in vivo* (Figure 12) (B. van Rietbergen & Ito, 2015; Varga et al., 2020). Studies have shown that HR-pQCT derived bone density, microstructure and strength are better predictors of fracture risk beyond DXA aBMD (Mikolajewicz et al., 2019; Samelson et al., 2019). FE-derived bone strength predictions can be used to identify individuals at risk of fracture and select the ideal treatment option that reduces this risk.

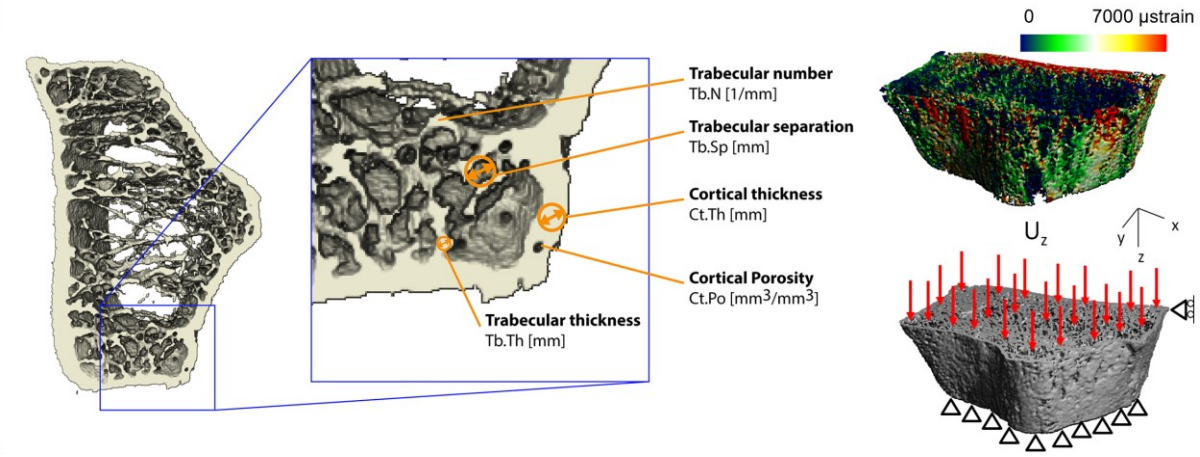


Figure 12. (left) A 3D render of a HR-pQCT scan showing some of the architectural outcomes, (right) and another 3D render showing the compressive boundary conditions and loads for microFE analysis, and the resulting strain distributions

Further, HR-pQCT allows for following temporal changes in cortical and trabecular bone by comparing registered (i.e., aligned) longitudinal scans. The registration of the scans enables us to measure changes in bone densitometric, structural, and biomechanical properties over a common region. This method has been used in multiple longitudinal studies on the effect of treatments and diseases such as osteoporosis (Burghardt, Kazakia, et al., 2010; L. A. Burt et al., 2017; Lauren A. Burt et al., 2017, 2018; Manske et al., 2015; Nishiyama et al., 2015; Peters et al., 2019; Tsai et al., 2015, 2016, 2017).

2.7. Precision of HR-pQCT measurements:

To detect and monitor temporal changes in bone density, microstructure, and strength, the short-term and long-term precision errors (or repeatability) of HR-pQCT measurements must be known (Glüer et al., 1995). Precision can be measured from repeated scans of subjects (*in vivo*) or bone mimic phantoms (*ex vivo*), so that any difference between the scans indicates error. Bone mimic phantoms have known densities that relate the intensity of voxels to bone density. Long-term precision can only be evaluated using the repeated scans of bone mimic phantoms because bones of living subjects change with time. The *in vivo* precision of HR-pQCT measurements is affected by different sources such as inherent error in the system, movement artifacts, physical positioning, and image processing (J A MacNeil & Boyd, 2008). Movement artifacts can be reduced by proper training and immobilization of the subject, although they cannot be completely removed. The *in vivo* precision errors of HR-pQCT measurements at the radius are reported to be below 1% for volumetric bone mineral density, 2.5% to 6.3% for structural parameters, and <3.5% for finite element outcomes (Boutroy et al., 2005; Burghardt, Buie, et al., 2010; Engelke et al., 2012; Kawalilak et al., 2016; J A MacNeil & Boyd, 2008; Mueller et al., 2009; Paggiosi et al., 2014). Precision errors are typically lower at the tibia than at the radius (Burghardt, Buie, et al., 2010; Engelke et al., 2012; Kawalilak et al., 2016).

A challenge in longitudinal HR-pQCT measurements is patient repositioning error, which results in different bone volumes being assessed at follow-up time-points. Image registration maps scans from one time point to another to identify the same bone volume. Image registration methods can be categorized into two-dimensional (2D) and three-dimensional (3D) (Whittier et al., 2020). The 2D method is in fact based on matching the cross-sectional area of the two scans, hence the name cross-sectional-area (CSA) registration is more accurate. The CSA method is also known as slice-match (SM) method, which aligns two images along the longitudinal axis (more proximal or distal) until the cross-sectional areas of the stacks are best aligned. Accordingly, the CSA registration can only correct for translation misalignments (Figure 13). Further, since the CSA method matches the area between two scans, it may be inaccurate when changes in bone area are expected, such as in response to treatment or diseases (Whittier et al., 2020). To correct for rotational misalignments along the x, y, and z axes, 3D-registration is required, which aligns the

images through a combination of translations and rotations to maximize the cross-correlation of voxel intensities (Figure 13).

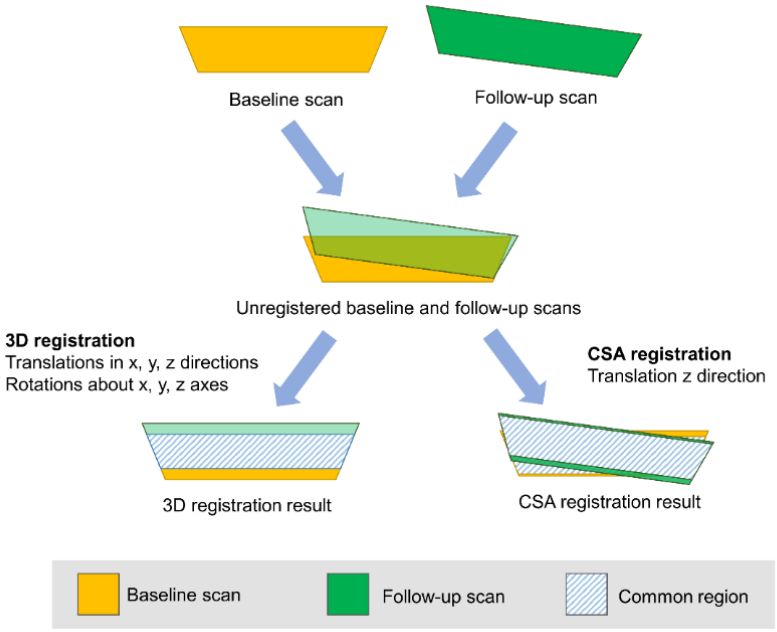


Figure 13. Schematic illustration of the differences between cross-sectional-area (CSA) and three-dimensional (3D) registration methods. The hashed regions in the bottom row indicate the common region between the two scans after image registration. In the CSA method, the followup image is translated iteratively in the longitudinal direction only until the best similarity between the total area of the bones in each image is obtained. 3D registration rotates and translates follow-up image to maximize the similarity between the intensity of voxels between the images. Modified and adapted with permission from Longitudinal bone microarchitectural changes are best detected using image registration, Kemp et. al., Osteoporosis International, 2020

3D registration can correct for both translational and rotational misalignments. However, 3D-registered microFE using Scanco's built-in solver requires two scans to have the same orientation with respect to the applied compressive loading, consequently introducing interpolation error. One approach to reduce interpolation error is to use matched-angle (MA) registration, which involves counteracting interpolation error by reducing its bias through rotating both the moving and reference images by the same angles but in opposite directions. This method was applied to tibial microCT images of rats where it was demonstrated to improve precision of trabecular microstructural parameters (de Bakker et al., 2016), and it was recently applied to HR-pQCT images for evaluating bone remodeling (Bert van Rietbergen et al., 2021). Another approach

to eliminate interpolation error is to transform the boundary condition instead of the images since the loading vectors can be transformed without error (Plett et al., 2021).

To date, only five studies have investigated the precision of 3D-registered HR-pQCT measurements, but none have been in individuals with OI. None of the previous studies reported results from multiple scanning sites and both scanner generations, thus suggesting that the scans were performed by highly trained personnel in a research setting rather than "real world" clinical trial scenario. In addition, none of the five studies examined MA-registered scans, and only two of the studies (Ellouz et al., 2014)(Plett et al., 2021) reported 3D-registered precisions for microFE, while they used different types of 3D-registered microFE. Finally, due to the lack of 3D-registered analyses in a multicenter study design, there is a need for a standardized evaluation of registration methods to inform best practice.

2.8. Evaluation of bone formation and resorption using HR-pQCT:

Registered longitudinal scans can also be used to monitor bone formation and resorption by subtracting images from two timepoints in a voxel-by-voxel fashion (Figure 14) (Christen & Müller, 2017).

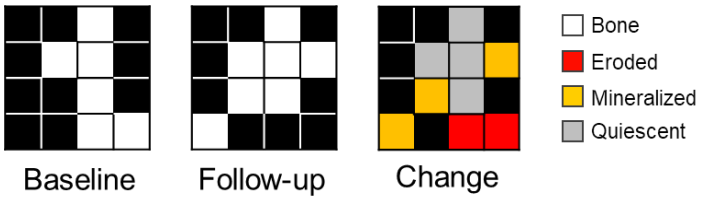


Figure 14. Schematic diagram of voxel-by-voxel comparison of two images using timelapse HR-pQCT. Each square indicates a voxel. White voxels are bone, while black voxels are background. If a voxel is present at followup, but not at baseline, it is considered mineralized (formed). If a voxel is present at baseline but not at followup, it is considered eroded (resorbed). If a voxel is present in both images, it is categorized as quiescent

Timelapse in vivo morphometry methods have been developed and validated in preclinical microCT models to allow the monitoring of cortical and cancellous bone formation and resorption (Birkhold et al., 2014a; Checa et al., 2014; Schulte et al., 2011; Waarsing et al., 2004). The qualitative approach was first developed using microCT scans of rats (Waarsing et al., 2004) and then expanded by other groups to quantify bone formation and resorption (Figure 15)(Birkhold et al., 2014b, 2014a; Schulte et al., 2011).

To date, six research groups have implemented timelapse HR-pQCT (Figure 15)(Atkins et al., 2021; Brunet et al., 2020; Christen et al., 2014, 2018; Collins et al., 2022; Du et al., 2020; Mancuso & Troy, 2020; Zhang et al., 2020). Contrary to the animal studies where binarized images were subtracted to identify bone formation and resorption, timelapse HR-pQCT studies have used binary or grayscale images as inputs (Figure 15). Four of the six studies used grayscale images as inputs. The first study assessed the long-term bone remodeling with aging in postmenopausal women, as well as computing errors of timelapse HR-pQCT (Christen et al., 2014, 2018). Another group (Mancuso & Troy, 2020) used grayscale images to relate bone strain to local changes in radius microstructure following forearm loading in postmenopausal women. A third group (Zhang et al., 2020) used the grayscale method to show higher bone resorption at the distal radius in progressive adolescent idiopathic scoliosis compared to the non-progressive form. The last group (Brunet et al., 2020) studied bone (re)modeling in metacarpophalangeal joints of participants with rheumatoid arthritis using grayscale images. In contrast, the remaining two studies used binary images as inputs. One group (Du et al., 2020) quantified bone (re)modeling in the tibia of postmenopausal women induced by 6 months of high impact exercise, while the other assessed bone formation and resorption at the distal radius of adults during fracture healing (Atkins et al., 2021; Collins et al., 2022).

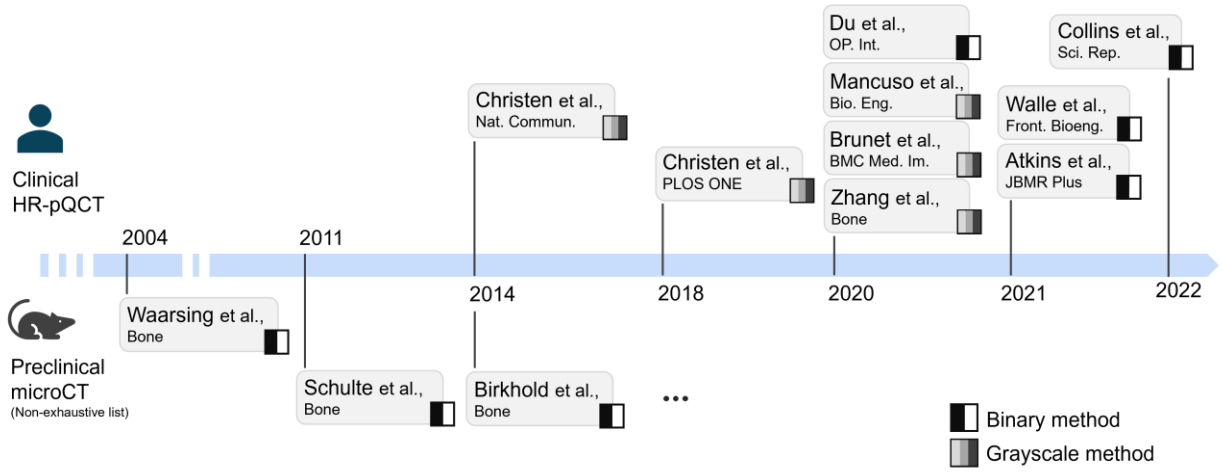


Figure 15. Timeline of the development and implementation of timelapse imaging in preclinical (bottom row) and clinical (top row) studies. The first implementation of timelapse analysis was done by Waarsing et. al., using timelapse microCT scans of rats tibia, which was only qualitative (i.e., visualize formed and reabsorbed sites). The list of preclinical studies is not complete, and

only shows a few highlight studies. The small squares at the bottom-right of each panel indicates whether binary or grayscale input images were used

Although a few studies have reported the errors associated with timelapse HR-pQCT, no one has yet performed a thorough parametric analysis and any validation to determine a preferred method. All of the mentioned studies used the standard 3D-registration method, which is prone to interpolation error (i.e., due to image rotation)(Hosseinitabatabaei et al., 2022; Bert van Rietbergen et al., 2021). It is not known if matched-angle registration (MA)(de Bakker et al., 2016; Hosseinitabatabaei et al., 2022; Bert van Rietbergen et al., 2021), can reduce the impact of interpolation error on timelapse HR-pQCT.

To date, no study has reported timelapse HR-pQCT in an OI population. Importantly, all of the studies on timelapse HR-pQCT were performed in a research setting at a single center with highly trained personnel. In a realistic multicenter clinical trial, which is common for OI, each participant is longitudinally scanned at a single center and radiology technicians at the different clinical centers have varying levels of training using HR-pQCT.

2.9. Available HR-pQCT data from OI population:

Limited HR-pQCT data is available from individuals with OI. Only four studies have reported bone geometry, density, and microarchitecture in the distal radius and tibia in adults with OI using HR-pQCT (Folkestad et al., 2012; Hald et al., 2016; Kocijan et al., 2015; Rolvien et al., 2018), aside from a case-report (Plachel et al., 2015). The first study reported lower total bone area at the radius, decreased trabecular number, increased trabecular separation, higher trabecular inhomogeneity, and decreased vBMD for 39 adults with type I OI compared to age and sex matched healthy controls (Folkestad et al., 2012). The same group (Hald et al., 2016) later reported significant differences between the HR-pQCT measurements of 85 adults with OI type I and IV. They reported lower vBMD, thinner cortices, and reduced trabecular number for type I OI compared to type IV. As suggested by the authors of the second study, the results of these two studies also implied similar HR-pQCT measurements between type IV OI and healthy individuals, while individuals with type IV OI suffer from low bone strength. Another study of 30 adults with OI types I, III, and IV, and 30 age and sex matched controls showed that OI types III and IV have

lower trabecular BMD and more deteriorated structure than type I in radius and tibia. They reported the similar trend for trabecular and cortical structural parameters but higher total BMD in tibia and cortical BMD at radius and tibia (Kocijan et al., 2015). The latest study on 21 adults with OI types I and IV, and age and sex matched controls indicated deteriorated trabecular and cortical structure and lower trabecular BMD for OI subjects, while cortical and total BMDs were comparable (Rolvien et al., 2018).

So far, only one study has reported HR-pQCT data for pediatric OI participants (Fennimore et al., 2020). Their participants consisted of 9 children with mild (*N=7*) and severe (*N=2*) OI, aged between 9-15 years old. The goal of this cross-sectional study was to investigate the feasibility of HR-pQCT imaging for pediatric OI participants considering their deformed and fragile bones. The scans were acquired using first generation H-pQCT scanner (XtremeCT) using the standard imaging and analysis protocols. However, this study provides no longitudinal information on HR-pQCT imaging of pediatric OI population, despite its importance due to growth.

Chapter 3. Objectives:

HR-pQCT is a relatively recent imaging technology that offers many opportunities for a variety of analysis on bone, while there are numerous technical subtleties related to HR-pQCT applications that are not well appreciated and standardized. Addressing these complexities are crucial to maximize the benefit from HR-pQCT, and to improve the reliability of HR-pQCT measurements. Some of the challenges and complexities associated with HR-pQCT are more general, while some are more specific to certain populations. Some of the general complexities are proper image registration for longitudinal studies, or proper approach for time-lapse HR-pQCT analysis to assess bone formation and resorption. An example of population specific complexities is that limited precision data is available for the osteogenesis imperfecta population. Another important example is the difficulties and confusions regarding pediatric HR-pQCT imaging and image analysis, such as proper scanning region or proper image registration method. To address some of the biggest challenges related to HR-pQCT reliability and standardization, my thesis is divided into three parts. Together, these three parts provide novel methodological approaches for reliable HR-pQCT image analysis, as well as clinically relevant data for the OI population:

1. **Evaluating the precision of HR-pQCT measurements using different registration methods in adults with OI:** Precision of HR-pQCT measurements is one of the main determinants of its reliability in longitudinal studies. Image registration is a promising method to improve the reliability of HR-pQCT measurements by ensuring that the same bone region is being analyzed over time. However, several variations of image registration exist, which require a systematic analysis to identify the best approach. Image registration is also required for more advanced HR-pQCT analysis such as timelapse to quantify bone formation and resorption. Last but not least, HR-pQCT precision data from OI population is lacking.
2. **Quantifying spatiotemporal changes at the distal radius and tibia of adults with OI induced by setrusumab (anabolic drug):** A more advanced analysis using timelapse HR-pQCT imaging can be used to quantify bone formation and resorption. There is however a need for standardizing timelapse HR-pQCT analysis based on a systematic approach. One area that requires standardization is the proper image registration technique. Other factors are the proper input image type, and proper

definition of bone masks. Further, to facilitate the clinical use of timelapse analysis, it is important to implement it in a realistic clinical setting. Using the time-lapse method, this study shows the effect of anabolic medication on the distal radius and tibia of adults with OI as part of a phase 2b clinical trial.

3. **Studying the natural history of the peripheral bones of children with OI and healthy controls using longitudinal HR-pQCT analysis:** While plenty of HR-pQCT data exist for adults, such data is limited for children, and no longitudinal data are available. Further, many of the techniques used for adults for improved reliability may not be readily applicable to children due to bone growth. Combining these two factors, longitudinal HR-pQCT studies in children are poorly standardized, and the challenges and opportunities are not well understood. Additionally, HR-pQCT data are rare for pediatric OI population, and in fact no longitudinal data exists for this population. To tackle these gaps, we performed a feasibility study on using image registration for growing children to clarify its limitations and potentials, and produced the first set of longitudinal HR-pQCT data for children with OI, and age and sex-matched controls. By improving our understanding of the challenges with longitudinal studies in growing children, and providing recommendations, this study can improve the reliability of HR-pQCT for the pediatric population.

The specific objectives of each study are:

Part 1 - Evaluating the precision of HR-pQCT measurements using different registration methods in adults with OI:

1. To investigate whether different image registration methods improve the short-term in vivo precision of HR-pQCT outcomes at the distal radius and tibia in adults with OI
2. To evaluate the influence of subject- and scanner-related factors on the extent to which registration improved precision errors in both scanner generations (XCT versus XCT2) at the radius and tibia

Part 2 – Quantifying spatiotemporal changes at the distal radius and tibia of adults with OI induced by setrusumab (anabolic drug):

1. To identify the errors associated with timelapse analysis using different settings on the in-vivo repeated scans from the participants in the clinical trial, and to select the preferred approach based on the resulting errors
2. To provide more rigorous validations for the selected method for timelapse analysis using the repeated scans and longitudinal data
3. To use the selected and validated method to evaluate bone formation and resorption induced by different doses of setrusumab in participants with OI in a clinical trial.

***Part 3 - Studying the natural history of the peripheral bones of children with OI and healthy controls using* longitudinal HR-pQCT analysis:**

1. To acquire one-year natural history data from children with OI and age-matched healthy controls using HR-pQCT and DXA
2. To assess the feasibility of image registration on longitudinal pediatric HR-pQCT scans

Chapter 4. Evaluating the precision of HR-pQCT measurements using different registration methods in adults with OI:

Status: *Published, Journal of Bone and Mineral Research (JBMR), 2022*

Publication title:

3D image registration marginally improves the precision of HR-pQCT measurements compared to cross-sectional-area registration in adults with osteogenesis imperfecta

Seyedmahdi Hosseinitabatabaei[a,b,c], Nicholas Mikolajewicz[a], Elizabeth A. Zimmermann[a,c,d], Maximilian Rummler[a,c], Beatrice Steyn[a,b,c], Catherine Julien[a,c], Frank Rauch[a], Bettina M. Willie[a,b,c,†]

[a]*Research Centre, Shriners Hospital for Children-Canada, Montreal, Canada*
[b]*Department of Biomedical Engineering, McGill University, Montreal, Canada*
[c]*Department of Pediatric Surgery, McGill University, Montreal, Canada*
[d]*Faculty of Dentistry, McGill University, Montreal, Canada*

Abstract:
Repositioning error in longitudinal high-resolution peripheral-quantitative computed tomography (HR-pQCT) imaging can lead to different bone volumes being assessed over time. To identify the same bone volumes at each time point, image registration is used. While cross-sectional area image registration corrects axial misalignment, 3D-registration additionally corrects rotations. Other registration methods involving matched angle analysis (MA) or boundary transformations (3D-TB) can be used to limit interpolation error in 3D-registering micro finite-element data. We investigated the effect of different image registration methods on short-term *in vivo* precision in adults with osteogenesis imperfecta, a collagen-related genetic disorder resulting in low bone mass, impaired quality, and increased fragility. The radii and tibiae of 29 participants were imaged twice on the same day with full repositioning. We compared the precision error of different image registration methods for density, microstructural and micro finite-element outcomes with data stratified based on anatomical site, motion status, and scanner generation. Regardless of the stratification, we found that image registration improved precision for total and trabecular bone mineral densities, trabecular and cortical bone mineral contents, area measurements, trabecular bone volume fraction, separation, and heterogeneity, as well as cortical thickness and perimeter. 3D-registration marginally outperformed cross-sectional area registration for some outcomes, such as trabecular bone volume fraction and separation. Similarly, precision of micro finite-element outcomes was improved after image registration, with 3D-TB and MA methods providing greatest improvements. Our regression model confirmed the beneficial effect of image registration on HR-pQCT precision errors, while motion had a detrimental effect on precision even after image registration. Collectively, our results indicate that 3D-registration is recommended for longitudinal HR-pQCT imaging in adults with osteogenesis imperfecta. Since our precision errors are similar to those of healthy adults, these results can likely be extended to other populations, although future studies are needed to confirm this.

4.1. Introduction:

High-resolution peripheral-quantitative computed tomography (HR-pQCT) is a promising non-invasive imaging tool for measuring bone density, microstructure, and strength at the distal peripheral skeleton.[1,2] A challenge in longitudinal HR-pQCT measurements is patient repositioning error, which can result in different bone volumes being assessed at follow-up time-points. To identify the same bone volumes at each time point, cross-sectional-area (CSA) or three-dimensional (3D) image registration methods can be used.[1]

The default CSA method currently used in HR-pQCT scanners (XtremeCT: XCT and XtremeCT II: XCT2, Scanco Medical, Bruttisellen, Switzerland) only corrects translational misalignment (i.e., positioning the limb more proximal or distal) and is insufficient to identify the common volume of interest (VOI) for rotational misalignment. Further, since the CSA method matches the area between two scans, it may be inaccurate when changes in bone area are expected, such as periosteal bone apposition or resorption in response to treatment or disease.[1,3] 3D registration methods can correct for both translational and rotational misalignments. Although, 3D registration is a superb alternative for microstructural and density parameters, it has disadvantages when used to register micro-finite-element (microFE) outcomes. The default approach used for microFE analysis in Scanco Medical's built-in solver simulates compression in a direction aligned with the scanner axis, not the bone axis. Consequently, to correct for rotational misalignments using 3D-registration, the images must be rotated to align them with respect to the scanner axis, thereby introducing interpolation error. An alternative approach to reduce interpolation error is to use matched-angle (MA) image registration. Image interpolation acts as a low pass filter and blurs (widens) the image. Using MA registration, the moving and reference images are both rotated by the same angle in opposite directions, and the error due to image rotation is propagated to both images, hence reducing the bias in interpolation error. MA registration was previously applied to tibial microCT images of rats, where it was demonstrated to improve precision of trabecular microstructural parameters,[4] and recently applied to HR-pQCT images for evaluating cortical retraction.[3] Another approach to eliminate interpolation error is to transform the microFE boundary conditions (i.e., compressive loading vector) instead of the images after finding the common volume using 3D-registration, since the loading vectors can be transformed without error.[5]

Although to date, five studies have investigated the precision of 3D-registered HR-pQCT measurements, a lack of consensus remains on which image registration method to use, especially for microFE data. MacNeil and Boyd showed that the majority of 3D-registered XCT density and microstructural outcomes tended to be more precise than CSA-registered using one-week apart repeated scans of 30 healthy adults (ages 20-40 years).[6] Recently, using second generation XCT2 scanner, Kemp and colleagues reported that compared to unregistered scans, both CSA and 3D registrations similarly improved short-term *in vivo* precision in 60 healthy adults (64.7 ± 5.1 years) for most radial parameters, and only tibial trabecular BMD, bone volume fraction, and cortical thickness parameters.[7] Chiba et. al.[8] used 3D registration method on the repeated scans (1 and 4 weeks apart) of 15 healthy adults (ages 20-74 years) to demonstrate improved precision error compared to the CSA method for several HR-pQCT measurements. Plett et. al.[5] compared the *ex vivo* precision of 3D and CSA registrations for microFE outcomes using 10 cadaveric bones and showed marginal improvements after 3D registration. They showed that both methods improved precision compared to no registration, except for CSA failure load. In contrast, Ellouz et al. [9] reported that image registration (CSA or 3D) of repeated scans from 15 healthy adults (ages 21-47 years, scanning interval same day to one week), did not improve the precision of microFE outcomes compared to unregistered scans. Among the two studies evaluating 3D-registered microFE precision, different variations of 3D-registration were applied on different datasets and the reported outcomes were either *ex vivo* or *in vivo* precision, but not both, thereby precluding comparison of the methods. Finally, the specific details of each registration approach were not reported, likely leading to a lack of consensus on which 3D-registration method to use.

Furthermore, none of the abovementioned studies reported short-term *in vivo* multicenter precision. Instead, analyses were performed at a single center with highly trained personnel in a research setting rather than under "real world" clinical trial conditions. In a realistic multicenter clinical trial, each participant is longitudinally scanned at a single center and radiology technicians at the different clinical centers have varying levels of training using HR-pQCT. Only in the rare case that a HR-pQCT scanner became inoperable during a trial would a participant be scanned at multiple centers. Multicenter trials are common when investigating rare diseases such as osteogenesis imperfecta (OI), a collagen-related genetic disorder resulting in low bone mass, impaired quality, and increased fragility, due to low participant numbers at each clinical

center.[10,11] Although most studies using HR-pQCT are single center cross-sectional, two multi-center longitudinal studies have been reported, requiring the availability of multicenter precision data.[12,13] To date, only three studies have reported multicenter precision of HR-pQCT measurements, two of which examined *in vivo* precision,[12,14] while the other used a cadaveric bone phantom.[15] However, none of these studies were performed using the scans from individuals with OI, and they did not examine the effect of image registration methods or scanner generation on precision.

In the current study we investigated whether different image registration methods improve the short-term *in vivo* precision of HR-pQCT outcomes at distal radius and tibia in adults with OI. Our study was performed as part of a clinical trial including ten clinical scanning centers and involved radiology technicians with varying levels of training and experience using HR-pQCT. We also evaluated the influence of subject- and scanner-related factors on the extent to which registration improved precision errors in both scanner generations (XCT vs. XCT2) at the radius and tibia. We believe that our precision data resulting from direct comparisons between different registration methods as part of a "real world" multicenter clinical trial, along with the provided detailed methodology and recommendations will motivate the HR-pQCT community of users to more readily reach a consensus on how to analyze longitudinal data.

4.2. Methods:

4.2.1. Participants:

Twenty-nine adults with OI type I or IV participated in this study (**Table 1**; mean age [SD]: 41.7 [14] years; age range: 19-65 years). Scans were performed at 10 HR-pQCT imaging sites in North America and Europe as part of the ASTEROID Phase 2b, multinational, randomized, double-blind, dose-finding study in adults with OI (ClinicalTrials.gov Identifier: NCT03118570). Each imaging site scanned between 1-3 participants. The inclusion/exclusion criteria are provided in **supplemental methods**. The study was approved by the institutional review boards of each clinical center, and informed consent was obtained from each participant. We have previously used these data to report single and multi-center short-term *in vivo* precision for unregistered and cross-sectional area (CSA) registered density and microstructural outcomes.[16]

Table 1. Descriptive statistics of the participants and their scans in this study. Each scan set represents duplicate scans with complete limb repositioning between scans. OI, osteogenesis imperfecta; SD, standard deviation; XCT, XtremeCT; XCT2, XtremeCT II

Overview of participants			
Variable	**Statistic**		**Value(s)**
Age (years)	Mean (SD):		41.7 (14)
	min < med < max:		19 < 39.5 < 65
Weight (kg)	Mean (SD):		74 (21.4)
	min < med < max:		39 < 72.5 < 118
Radius Length (cm)	Mean (SD):		26.5 (2.4)
	min < med < max:		21 < 25.2 < 30
Tibia Length (cm)	Mean (SD):		37.9 (2.5)
	min < med < max:		33 < 38.5 < 42.5
Variable	**Feature**	**N**	**(%)**
Scanner	XCT	9	31.0
	XCT2	20	69.0
OI Type	Type I	23	79.3
	Type IV	6	20.7
Sex	Female	18	62.1
	Male	11	37.9
Anatomical Site	Radius	23	46.9
	Tibia	26	53.1

4.2.2. HR-pQCT imaging:

At 10 scanning sites (6 XCT, 4 XCT2), two HR-pQCT scans of the radius and tibia were acquired on the same day by the same operator after full repositioning. The reference line positioning was at the medial proximal margin of the radial articular surface and at the tibial plateau.[17] The mid-point of the scanned volume of interest was located at a distance of 4% (radius) and 7% (tibia) of the ulna or tibial length from the reference line.[1,17] Each scan contained 110 slices at 82 µm isotropic voxel size or 168 slices at 60.7 µm isotropic voxel size for XCT and XCT2 scanners, respectively.

4.2.3. Image quality:

After image quality assessment, 6/29 radial (20.7%) and 3/29 tibial (10.3%) scans were removed due to motion (score ≥ 4) or metal artifacts, and the remaining scans were retained for analysis.[18] At each site, 1-2 trained technicians created the periosteal contour. A single, trained technician at Shriners Hospital for Children (Montreal, Canada) performed the standard evaluation protocol recommended by Scanco and verified the periosteal and endosteal contours.

4.2.4. Image registration:

4.2.4.1. Density and microstructural analysis:

We reported density and microstructural outcomes from the volumes of interest (VOI) identified using three methods: a) unregistered, b) cross-sectional area (CSA) registered, and c) three-dimensional (3D) registered. A complete list of parameters calculated for each HR-pQCT scan is provided in **Supplemental Methods Table 1**.

Unregistered and CSA-registered measurements were produced automatically by the Scanco software. In CSA-registered images, the software automatically registered the two images by longitudinally translating the second image until reaching the maximum cross-correlation between the total cross-sectional area (Tt.Ar) of the two scans. For CSA registration, we only report measurements on images with an overlapping volume larger than 80%.

The first step of 3D registration using IPL (scripts available on Willie Lab Github repository) was to find the 4 × 4 transformation matrix that aligned the second grayscale image (moving image) with the first grayscale image (reference image) (**Figure** 16, *step 1*). We used initial alignment of the center of mass of the two images, a cross-correlation similarity metric, and the downhill simplex optimization scheme. To reduce the effect of noise on registration, only the volume within the periosteal contours were registered. The registration was performed in 3 stages (scale factors of 10, 4, and 1) to avoid registration errors (e.g., local minima) and to reduce computation time. For the second step, the 4 × 4 transformation matrix was used to transform a binary solid block (slab) of the moving image to the domain of the reference image to identify the largest common volume between their masks (**Figure** 16, *step 2*). Using the solid blocks, we prevented any error from image rotation (i.e., interpolation error) to occur on the bone periosteal surface. Finally, in the third step the common volume was inverse-transformed to the domain of the moving image to modify the original periosteal and endosteal contour lines (**Figure** 16, *step 3*). This approach prevented interpolation error that would result from rotating the grayscale images used to measure bone properties.[7,9] Nearest neighbor and cubic interpolations were used for the binary and grayscale images, respectively.[19]

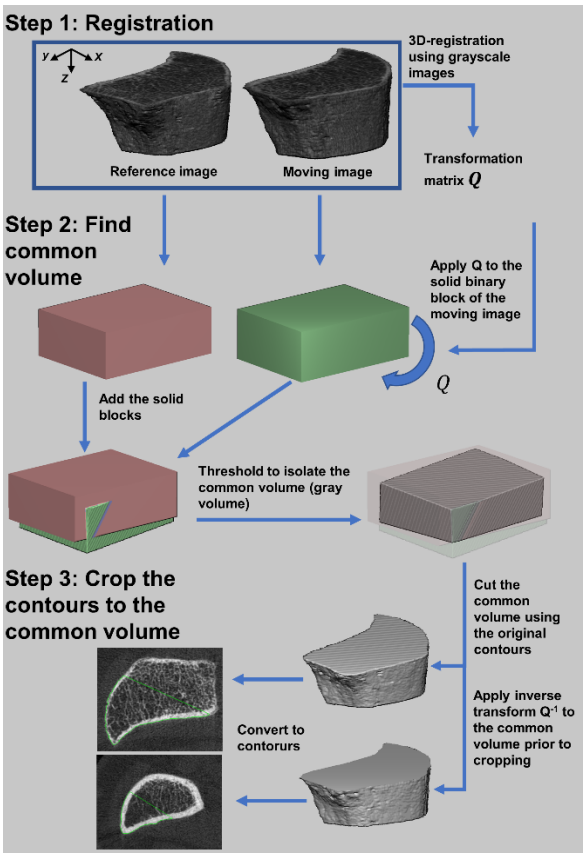


Figure 16**. 3D registration process to align two images and to find the common volume for density and microstructural outcomes**. The **first** step was to find the transformation matrix that would align the moving image onto the reference image. In the **second** step, solid blocks were created and aligned to find the common volume. Finally, during the **third** step, the common volume was used to modify the original contour lines (i.e., Scanco GOBJ files). Although only periosteal contours are shown, the same process was used to modify the endosteal and cortical contours. 3D, three-dimensional; Q, transformation matrix.

4.2.4.2. microFE analysis:

For micro-finite element (microFE) analyses, we simulated axial compression to derive stiffness, failure load, and apparent modulus using IPLFE v01.16 (Scanco Medical AG) (**Supplemental Methods Table 1**). We used a linear elastic modulus of E = 6829 MPa, and E = 8748 MPa for XCT and XCT2 scans, respectively.[2,20,21] Poisson's ratio was set to 0.3. We used Pistoia's failure criteria with critical volume of 2% and critical value of 7000 µstrain.[22,23]

We reported microFE analyses of the VOIs identified using five different registration methods (scripts available on Willie Lab Github repository): a) unregistered, b) cross-sectional area (CSA) registered, c) 3D-registered[9] (3D), d) matched-angle (MA) registered,[4] and e) 3D-registered with transformed boundary conditions (3D-TB).[5] The 3D, MA, and 3D-TB registration

methods correct for rotational misalignments; the 3D method does so by rotating the moving image to align it with the reference image. The MA method rotates both images, but in opposite directions, hence negating the bias introduced by rotating just one image, therefore attenuating the impact of interpolation error. The 3D-TB method does not rotate any of the image, rather it rotates the boundary conditions to match each of the images, hence eliminating interpolation error.

For the CSA-registered microFE analysis, we identified the slice offset between the scans from the microstructural analysis report and performed microFE over the common region identified during the analysis of the density and microstructural outcomes. For 3D registration of microFE data (**Figure 17**), we aligned the grayscale and binary mask of the moving image with the reference image. Then, the grayscale images were binarized using the standard Scanco's segmentation protocol (same for all registration methods) and segmented using their respective masks. For XCT scans, only the periosteal mask was required, while XCT2 scans required both the periosteal and endosteal masks, due to the dual-threshold segmentation process. Next, since the common volume created during 3D-registration did not have flat distal and proximal ends, it was cropped at the top and bottom to create flat surfaces (i.e., the slices containing partial bone cross-sections were identified visually and removed).

For the MA method,[4] we rotated the reference and moving images using rotation angles of half of the original angles, but in opposite directions. This way, the MA method aligns the two images in a middle domain (**Figure 17**). Briefly, using the registration function in IPL, we divided the rotation angles from registration by half to identify the middle domain. Then, each of the reference and moving images where registered to the middle domain.[3] Finally, we performed microFE on the flat common volume.

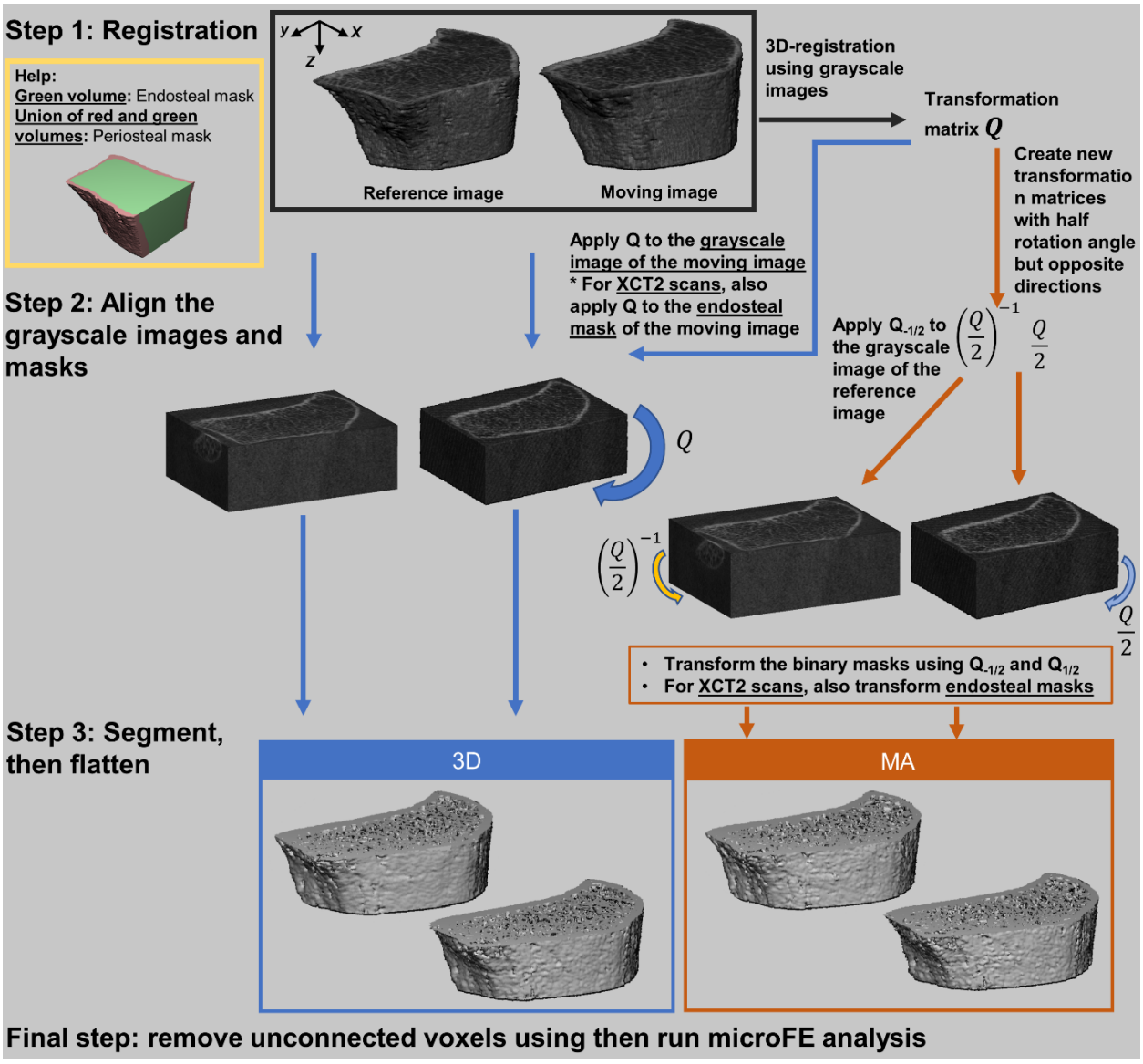


Figure 17. 3D and MA registration methods to align two images for microFE analysis on the common region. The **first** step was to find the transformation matrix that would align the moving image to the reference image. In the **second** step, the grayscale images were aligned. Finally, during the **third** step, grayscale images were segmented, binarized, and cropped to perform microFE. In 3D method, the moving image was transferred to the domain of the reference image, whereas in MA method, both images were transferred to a middle domain. 3D, three-dimensional; MA, matched-angle; Q, transformation matrix.

For the 3D-TB method (**Figure 18**), instead of rotating the moving image to align it with the reference image, the compressive displacement vector was inverse transformed to match the orientation of the moving image using the same transformation matrix obtained during 3D-registration (i.e., the resulting vector had components in x, y, and z directions, whereas the initial

vector only had z component). This way, we prevented introducing interpolation error into the images. An alternative volume flattening method to the one used for the 3D and MA models was required, because the coordinate system and the slicing of the images is defined with respect to the scanner axis rather than bone axis. Therefore, different number of slices will be cropped, and at a different orientation with respect to the scanner axis, thereby creating two different volumes. To address this, we created a flat common mask in the domain of the reference image, and inverse transformed it to the domain of the moving image. Then, to fill the cropped regions, we added a layer of stiff material with an elastic modulus of 410 GPa and Poisson's ratio of 0.3. The addition of the stiff material enabled the application of the boundary conditions in IPL. Next, during post-processing, reaction forces were projected onto the unit vector representing the direction of applied compressive load and aggregated to obtain the total reaction force (i.e., a simplified alternative is to extract the total reaction forces in x, y, and z directions from the POSTLIST file, then project them onto the unit vector of the applied loading). Finally, stiffness and failure load were calculated.

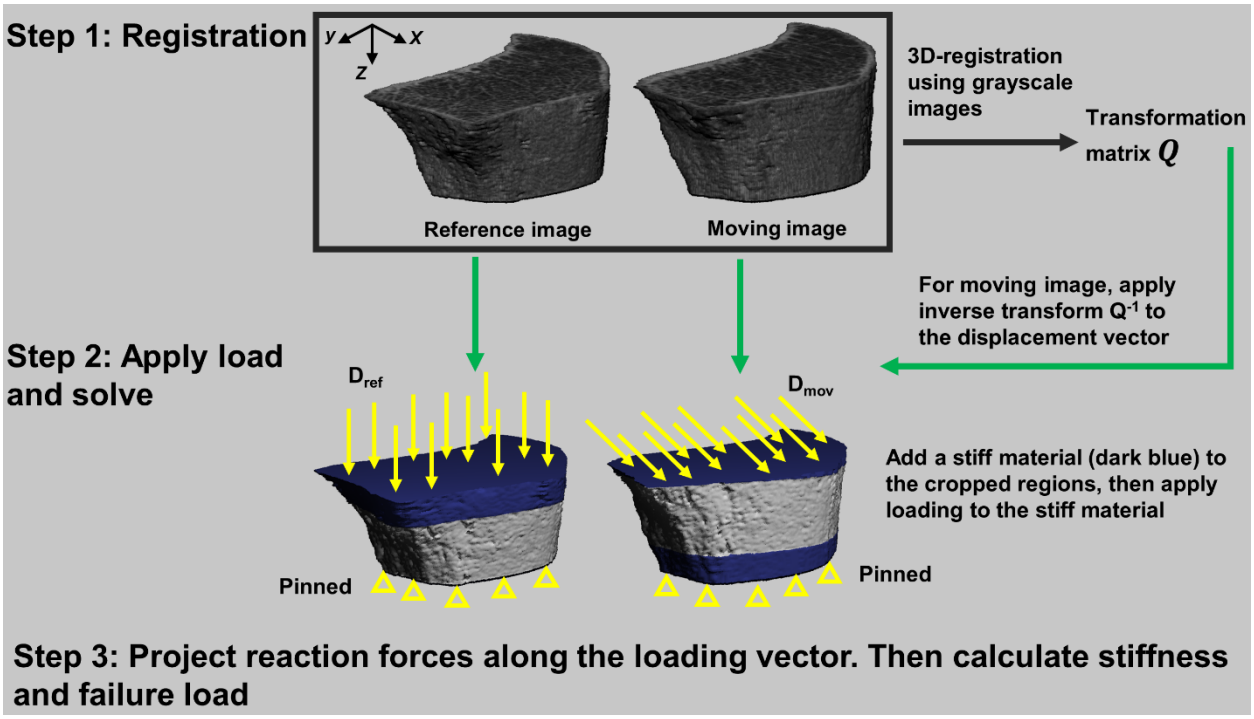


Figure 18. 3D registration methods with transformed boundary conditions to align the applied compressive strains for the two images for microFE analysis on the common region. The **first** step was to register the grayscale images. In the **second** step, a stiff material was added to the regions that were cropped after finding the 3D-registered common volume. The vector of applied displacement was also transformed using the transformation matrix obtained from 3D-

registration. In the **third** step, after solving the model, stiffness and failure load were calculated using Pistoia's criterion. The nodal reaction forces were projected along the loading vector. Q, transformation matrix; D_{ref}, displacement vector applied to the reference image; D_{mov}, displacement vector applied to the moving image.

We reported two types of overlaps. First, we reported the percentage of overlap for bone masks before and after registration to quantify the degree of repositioning error. The average overlap percentage (flattened overlap% in parentheses) of different sites and scanners were 91% ≤ CSA ≤ 96%, 91% (83%) ≤ 3D ≤ 95% (90%), 91% (83%) ≤ MA ≤ 96% (90%) (**Supplemental Methods Figure 1-A**). Second, we computed the overlap between bone voxels after alignment during 3D or MA registration, which provides an estimate of registration accuracy. The average percentage of matching bone voxels was 85% and 84% for 3D and MA-registration, respectively (**Supplemental Methods Figure 1-B and C**). Further, using visual inspection of the overlaid images from 3D or MA registrations, we verified that image registration did not fail. We also calculated the rotational misalignment between the two images around x-axis [tibia] (palmar [medial] – dorsal [lateral]), y-axis (medial [anterior] – lateral [posterior]), and z-axis (distal – proximal). Rotations around the axis in the transverse plane (x and y) were summed and named R_{xy}, while rotation around longitudinal axis was named R_z. R_{xy} and R_z were defined as separate terms because the latter only contributes to the interpolation errors, while the former contributes to both repositioning and interpolation errors.

4.2.5. Statistical analysis:

We reported short-term precision errors as root mean squared coefficients of variation ($CV\%_{rms}$)[24] and differences between registered and unregistered precision errors ($\Delta CV = CV_{registered} - CV_{unregistered}$) were used to evaluate (*i*) effect of registration on precision errors ($H_0: \Delta CV = 0$, where H_0 is the null hypothesis), and (*ii*) registration performance between registration methods ($H_0: \Delta CV_A - \Delta CV_B = 0$, where A and B represent different registration methods).

We constructed multiple linear regression models to evaluate the influence of participant- (i.e., sex, anatomical site, motion, unregistered precision), scanner- (i.e., scanner generation), and processing- (i.e., registration method, xy rotation [R_{xy}], z rotation [R_z]) related covariates on changes in precision errors following registration (ΔCV). Using variance estimates from each model we also estimated the proportion of variance explained by each covariate. Although our dataset comprised of scans obtained from different scanner generations, our analyses were

performed on pooled data to reflect the approach taken in multicenter trials, and to maximize our statistical power. In cases where the influence of individuate covariates were evaluated, we compared stratified ΔCV values using Wilcoxon signed rank test. By stratifying the outcomes based on covariates of interest (i.e., anatomical site, presence of motion, and scanner generation), we could determine the validity of the pooling.

The methods above evaluate the *average* effect of registration on precision error. To investigate how registration influenced precision at the individual-level, we present a novel visualization technique termed "*Registration performance plots*" (RPPs). For each registration method, we calculate the percentage change in precision following registration ($\Delta CV\% = \frac{CV_{registered} - CV_{unregistered}}{CV_{unregistered}} \times 100\%$). Then, for each $\Delta CV\%$ value, we calculate the performance of registration P_r, defined as $P_r = (-1) \times \Delta CV\%$ (e.g., $+100\%$ performance is equivalent to $\Delta CV\% = -100\%$). P_r is non-symmetrical and is bound by $[-Inf, 100]$, and is interpreted such that $P_r > 0$ represents favorable registration performance (i.e., improvement), $P_r < 0$ represents unfavorable registration performance (i.e., deterioration), and $P_r = 0$ represents no change following registration. Next, we calculated the fraction of scans for which $\Delta CV\%$ improved or deteriorated at least some specified amount, for P_r values ranging from -100% to $+100\%$. These data were then visualized as an RPP defined by an x-axis representing P_r values ranging from -100% (truncated from $-Inf$) to $+100\%$, and the y-axis representing the fraction of scans achieving at least P_r performance. Although plotted on a common axis, improvement ($P_r > 0$) and deterioration ($P_r < 0$) curves (i.e., component curves) were evaluated separately for the purpose of comparing the extent of improvement and deterioration offered by different methods. Component curves were compared across different registration methods using the area under curve (AUC) as the statistic of interest, for each side of the RPPs (i.e., improvement and deterioration). On the improvement side, a larger AUC indicates better performance, while the opposite is true for the deterioration side. To determine whether the AUCs were significantly different between registration methods, we bootstrapped the component curves (N = 1000) to estimate AUC variance terms. To complement these analyses, the y-intercept of each RPP component curve at $P_r = 0$ can be interpreted as the fraction of scans that resulted in improvement or deterioration after registration. Thus, in addition to each RPP, we also represented these fractions in bar plots.

P-values from multiple comparisons were corrected using Benjamini-Hochberg (BH) method. The significance level was $p<0.05$, and tendencies at $p<0.1$. In addition to $CV\%_{rms}$, we also report median CV ($CV\%_{med}$), which is robust to outliers. Further, as per recommendations by the International Society for Clinical Densitometry (ISCD), we computed the least significant change (LSC).[24–26]

4.3. Results:

4.3.1. Density and microstructural:

4.3.1.1. Aggregate registration performance:

We evaluated the effect of image registration on density and microstructural parameters using a combination of regression models (**Figure 19**, **Supplemental Results Figure 4-B**) and stratified pairwise analyses (**Supplemental Results Figure 3**). Regression analyses allowed us to investigate the influence of various factors affecting the impact of image registration on precision error, while handling the high degree of heterogeneity observed in our data (visualized using heatmaps; **Supplemental Results Figures 1 and 2**). For each outcome, the heatmaps shown in the **Supplemental Results Figures 1** indicate the absolute CV% values for each participant, site, and scanner generation, and registration method. The heatmaps in **Supplemental Results Figures 2** show changes in precision after image registration. We included baseline unregistered precision errors in our regression models because we observed a crude (unadjusted) negative association between unregistered precision errors and the degree of improvement following registration, suggesting that scans with larger precision errors prior to registration have larger margins for improvement following registration (**Figure 19**, **Supplemental Results Figure 4-A**). Importantly, the inclusion of unregistered precision in our models did not result in spurious associations arising from correlated error terms between unregistered precision and change in precision following registration were derived from the same images. The full regression results for all indices, including the residuals histogram can be found in **Supplemental Results Figure 4-B**. The model residuals were used to evaluate bias. This ensured that similar measurement error terms canceled each other out (see **supplemental methods** for details).

With respect to the change in precision error following registration, regression analyses explained varying degrees of observed variance, ranging from 8% for Tb.Th and Tt.vBMC to 68% for Tt.vBMD. Despite this variation, we found that baseline unregistered precision errors consistently explained the most variance (15-50%; except for Tt.vBMC [~0.5%] and Tb.Th

[~1.5%]), followed by image registration and xy rotation angles, both explaining between 1%-20% (**Figure 19**, **Supplemental Results Figure 4-B**). Importantly, after adjusting for other factors, CSA and 3D registration methods were independently associated with improvements in precision error for Tt.vBMD, Tb.vBMC, Ct.vBMC, and Ct.Th. CSA registration was also associated with improved precision in Tb.vBMD, Ct.vBMD, Tb.BV/TV, Tb.1/N.SD, Tb.aBMC, Ct.aBMC, Tb.Ar, Ct.Ar, and Ct.Pm. For Ct.Po, only 3D-registration was associated with improved precision. Larger xy rotation angles were associated with improved precision for Tb.vBMD, Tb.vBMC, Tb.N, Tb.Sp, and Tb.1/N.SD. Only Tb.Th at the tibia was associated with a reduction in precision error. Motion and scanner generation were not associated with changes in precision error following registration.

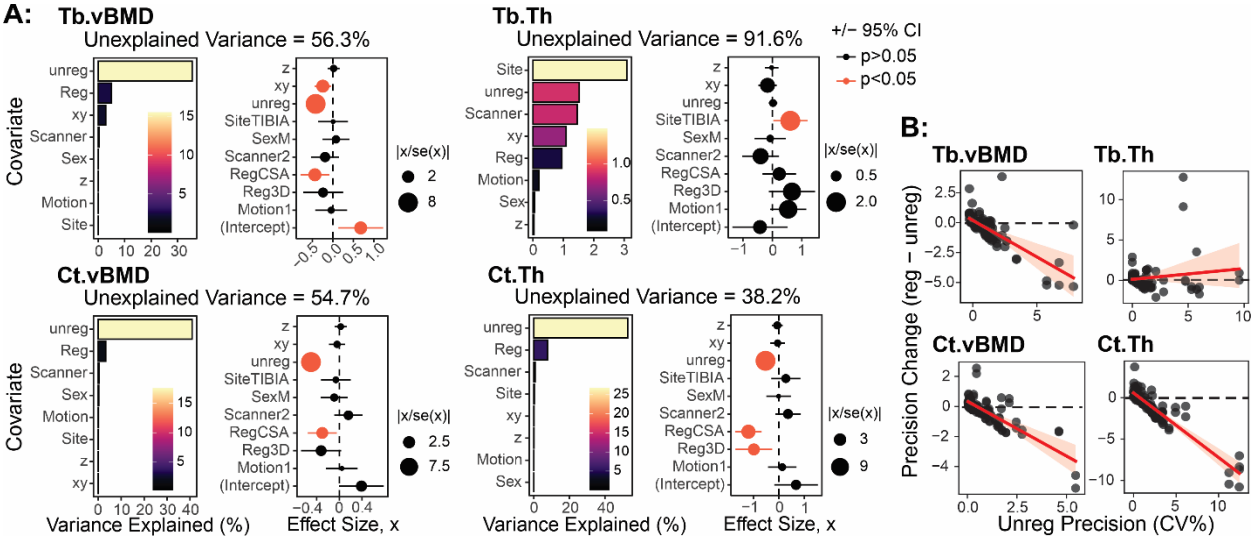


Figure 19. Regression analysis of registration-associated changes in precision errors for selected density and microstructural outcomes. A) For each HR-pQCT measure, **left panel** shows percentage of variance explained by each model covariate and **right panel** shows model covariate coefficients. The color bars indicate % of explained variance. For model covariate coefficient plots, effect sizes x were scaled by standard errors se(x) and error bars represent 95% confidence intervals (CI). Dependent variable in the regression model is the change in precision error following registration for each measurement, and independent variables are anatomical site (reference = radius), motion status (reference = 0 [no motion]), scanner generation (reference = 1 [XCT scanner]), unregistered precision error (unreg), registration method (Reg), rotation angles (xy and z), and sex (reference = F [female]). **B)** The **scatter plots** indicate the crude association between the unregistered precision error and changes in precision after registration. 3D, three-dimensional registered; CSA, cross-sectional area registered.

Regression analyses were further complemented through stratified pairwise analyses, specifically stratifying data be anatomical site, motion status or scanner generation, and comparing

precision errors across registration methods and versus unregistered values. Although results were consistent between the two approaches, pairwise analyses revealed additional nuanced differences between methods. For example, 3D and CSA methods significantly improved precision of Tt.vBMD, Tb.vBMC and Ct.vBMC at the radius and tibia, however the same methods only improved Ct.Th at the radius and Tb.vBMD, Tb.BV/TV, Tb.Sp, and Tb.1/N.SD at the tibia. 3D registration also had larger improvements in precision error compared to CSA for tibial Tt.vBMD, and Tb.Sp (**Supplemental Results Figure 3-A**). Additionally, pairwise comparisons for the data separated for radius and tibia prior to stratification based on motion status and scanner generation are available (**Supplemental Results Figure 3-B and 3-C**). These results were mostly consistent with the pooled data, although some differences were lost due to the lower power.

4.3.1.2. Subject-level registration performance:

To investigate the effect of image registration on the precision of density and microstructural parameters at the individual-level, we constructed registration performance plots (RPPs) for each measure stratified by anatomical site, motion status, or scanner generation (**Figure 20**, **Supplemental Results Figure 5**). The supplemental figures show the RPPs for all indices, while **Figure 20** only presents select indices (Tb.vBMD, Ct.vBMD, Tb.Th, and Ct.Th). These analyses provide information about *i*) the fraction of scans that become more or less precise following registration (RPP y-intercept), and *ii*) the extent of improvement or deterioration in precision error achieved by each registration method (difference in AUC between CSA and 3D methods).

Anatomical site. CSA and 3D registration improved precision at radial and tibial sites in >60-80% of scans for Tt.vBMD, Tb.vBMD, Ct.vBMD, Tb.aBMC, Ct.aBMC, Tb.BV/TV, Tb.Ar and Ct.Pm, indicating a relatively low risk of introducing error to these measures by registration (**Figure 20A**, **Supplemental Results Figure 5-A**). However, there were several parameters in which registration led to similar rates of deterioration (i.e., worse precision following registration) and improvement (i.e., improved precision following registration), including Ct.vBMD (radial and tibial), Tt.vBMC (radial and tibial), Tb.N (radial and tibial), Tb.Sp (radial) and Tb.1/N.SD (radial). In these cases, aggregate statistics (e.g., $CV\%_{RMS}$) suggest that registration is associated with little to no improvement in precision error. However, on an individual basis registration is introducing error to a substantial fraction of scans. This has implications in clinical practice where screening and monitoring HR-pQCT measures in individuals requires reliable reproducibility, that is not

deteriorated by image registration. Among scans that were less precise following registration, 3D registration introduced significantly less error than CSA registration to Tt.vBMD (radial), Tb.vBMC (radial), and Tb.BV/TV (tibial) measures (represented by $AUC_{3D} < AUC_{CSA}$ in the detrimental range of the RPP). Whereas CSA registration introduced less error than 3D registration for Ct.Th (radial and tibial; $AUC_{3D} > AUC_{CSA}$ in the detrimental range of the RPP). Conversely, among scans that resulted in improved precision following registration, the 3D method improved precision to a greater extent than CSA for Tt.vBMC (radial and tibial), Tb.vBMD (radial), Ct.vBMC (radial), Ct.Po (radial), Tb.N (tibial), Tb.Th (tibial), Tb.Sp (tibial), and Tb.1/N.SD (tibial). Contrary to findings in the radius, in the tibia CSA methods led to superior improvements in precision for Tt.vBMC.

Motion status. Independent of motion status, we found that each registration method improved precision in >60-80% of scans for Tt.vBMD, Tb.vBMD, Tb.vBMC, Ct.vBMC, Tb.aBMC, Ct.aBMC, Tb.BV/TV, Tb.Ar, and Ct.Pm (**Figure 20B**, **Supplemental Results Figure 5-B**). Among individual scans that were less precise following registration, 3D registration was less detrimental than CSA for Tb.vBMD (no motion), but more detrimental for Tb.Th (no motion) and Ct.vBMC (no motion), while the opposite was true for Tt.vBMC (motion) and Ct.Th (motion). Among scans that were more precise following registration, 3D registration offered greater improvements than CSA registration in Tb.vBMD (motion), Tb.BV/TV (motion), Tb.Th (no motion), Tb.Sp (motion), Tb.1/N.SD (motion) and Ct.Po (motion).

Scanner generation. Finally, examining the effects of scanner generation, we found similar patterns with motion and anatomical site stratified data. Among scans acquired using XCT scanners (**Supplemental Results Figure 5-C**), 3D registration was less detrimental than CSA for Tb.vBMD and Tb.BV/TV when registration resulted in lower precision. Among scans that experienced improved precision following registration, 3D-registration was superior to CSA registration for Tb.Sp. For XCT2 scans (**Supplemental Results Figure 5-C**), 3D registration resulted in a larger AUC on the improvement side for Tb.N, Tb.Th, Tb.Sp, Tb.1/N.SD, and Ct.Po.

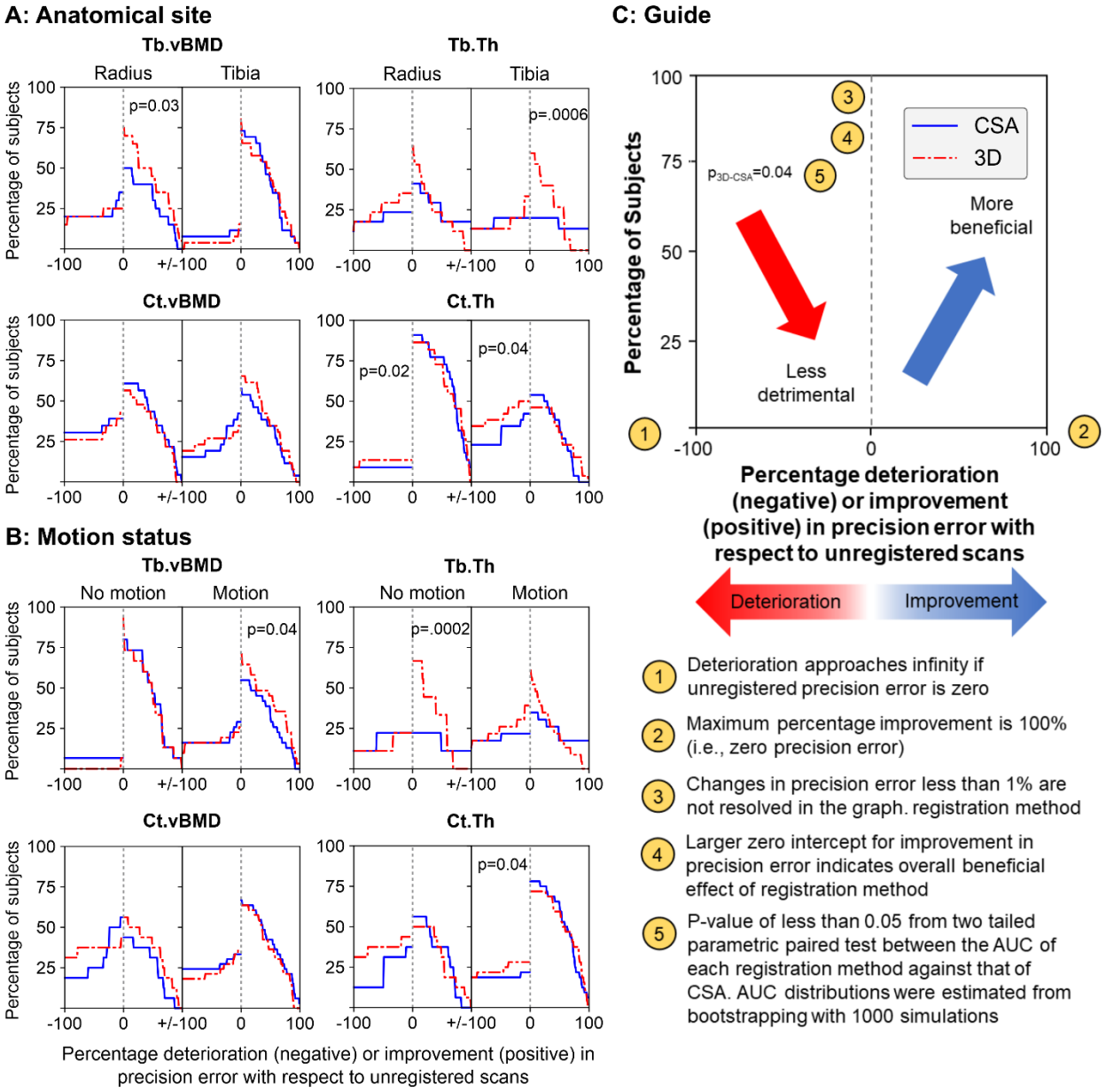


Figure 20. Registration performance plots (RPPs) of CSA and 3D registration methods for selected density and microstructural outcomes. A-B) RPP analyses comparing CSA and 3D methods by anatomical site (**A**) and motion status (**B**). X-axes show the percentage change in precision error for a registration method relative to unregistered precision error. Positive values of x represent improvement in precision error when using image registration, whereas negative values indicate deterioration. Y-axis shows the percentage of individuals for whom a percentage of improvement or deterioration is obtained using image registration. For positive values on x-axis, higher percentage of individuals means better performance, while for negative values on x-axis, lower percentage of individuals is better. The difference between 100% and the sum of negative and positive curves at x=0 is the percentage of individuals with improvements in precision less

than 1%. P-values indicate significant difference in area under the curve (AUC) separated by improvement or deterioration. **C)** RPP interpretation guide.

4.3.2. microFE:

4.3.2.1. Aggregate registration performance:

For microFE parameters, we evaluated the influence of anatomical site, motion status, scanner generation, baseline unregistered precision error (assessed for the same reason as for density and microstructural measures [**Figure 22**, **Supplemental Results Figure 9-A**]), registration method, rotation angles and sex on the performance of registration using regression analysis (**Figure 22**, **Supplemental Results Figure 9-B**). Similar to the density and microstructural measures, the heatmaps presented in **Supplemental Results Figures 6 and 7** present the absolute CV% and changes in precision after registration, respectively. Regression models explained ~41%, 36%, and 68% of the variance observed in registration-associated changes in precision for failure load, stiffness, and apparent modulus, respectively. Unregistered precision error alone explained ~35%-45% of the variance for microFE outcomes, followed by image registration, which explained <5% of variance for failure load and stiffness, and ~12% for apparent modulus. Similar to non-microFE parameters, unregistered precision was significantly and negatively associated with registration-associated changes in precision error for all microFE outcomes. CSA and 3D-TB registrations were associated with significant improvements in precision stiffness, while only 3D-TB was associated with improved failure load precision. CSA, 3D, and MA were associated with improvements in precision for apparent modulus. For failure load, scans acquired in males (vs. female), and XCT scans marginally resulted in further improvements in precision following registration (**Figure 22**).

From our multiple comparison results, we found that the precision of apparent modulus was significantly improved using all registration methods regardless of stratification, except for scans without motion (**Figure 21A-C, Supplemental Results Figure 8-A, B**). For the radial and tibial and XCT2 scans, the 3D-TB method improved the precision of failure load compared to unregistered scans. At the tibia, 3D-TB failure load was also more precise than CSA. However, these effects were significant only prior to p-value correction (**Figure 21A-C**). For scans with motion and radial scans, both the 3D-TB and MA methods improved failure load precision; however, at the radius, significance was lost after p-value adjustment. For stiffness calculated in scans with motion, precision significantly improved using all registration methods except 3D. For

XCT2 scans, only the 3D-TB and MA methods improved stiffness precision. At the radius and prior to p-value adjustment, all registration methods improved precision for stiffness, except the 3D method. Also, for XCT2 prior to p-value adjustment, MA stiffness was more precise than CSA. The pairwise comparisons separated for radius and tibia prior to stratification based on motion status and scanner generation are also available in **Supplemental Results Figure 8-A and B**, which show similar patterns as the pooled data.

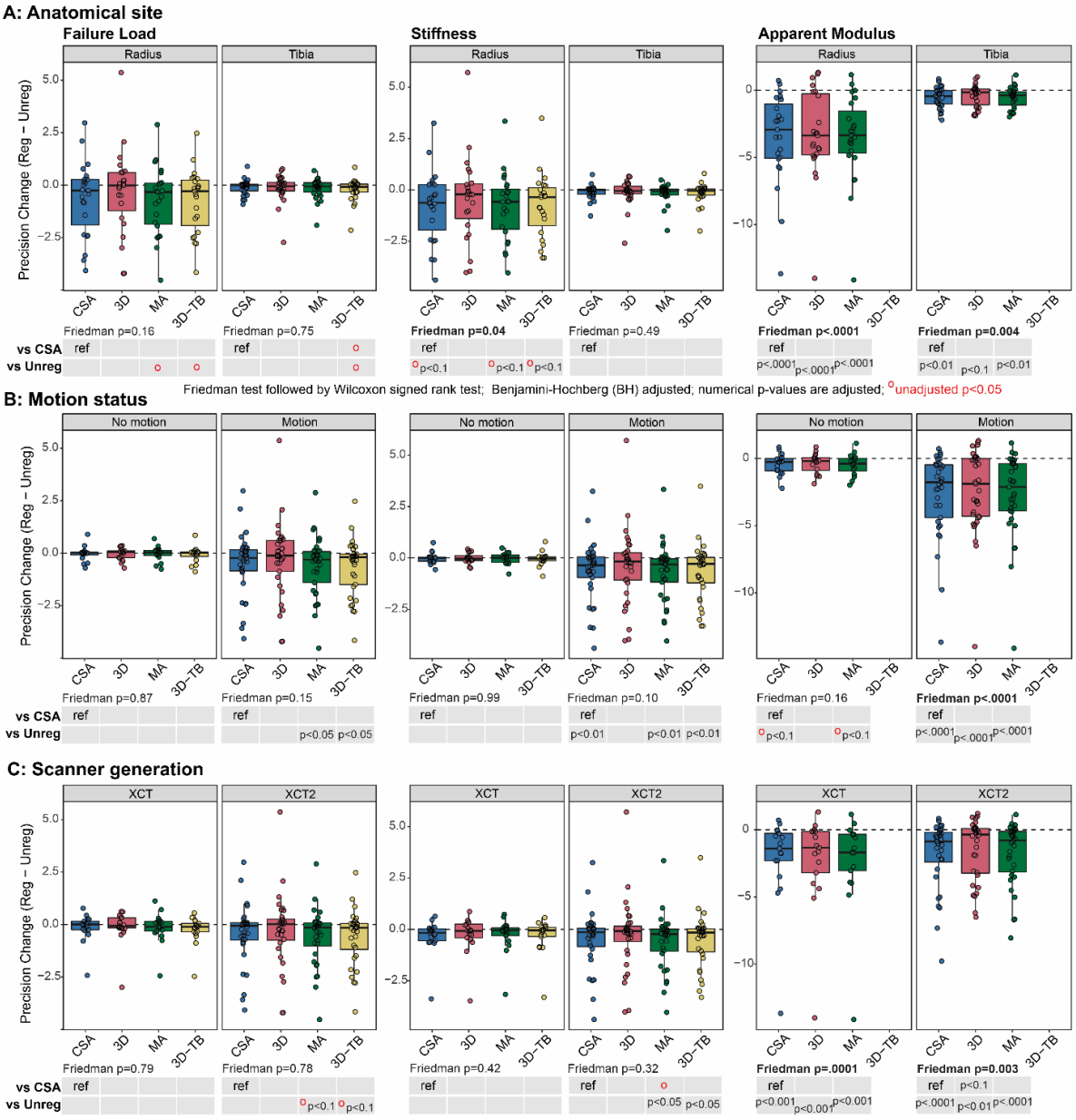


Figure 21. Pair-wise comparisons of registration-associated changes in precision errors for microFE outcomes. A-C) Comparative analyses for i) vs. CSA and ii) vs. unregistered methods were performed using Wilcoxon signed rank test for data stratified by anatomical site (**A**), motion

status (**B**), and scanner generation (**C**). P-values were corrected for multiple comparisons using Benjamini-Hochberg (BH) method. Significance level is $p<0.05$, while significance prior to p-value adjustment is shown as $^{o}p<0.05$. 3D, three-dimensional; 3D-TB, three-dimensional registered with transformed boundary conditions; CSA, cross-sectional area registered; MA, matched-angle registered; Unreg., unregistered. XCT, XtremeCT; XCT2, XtremeCT II.

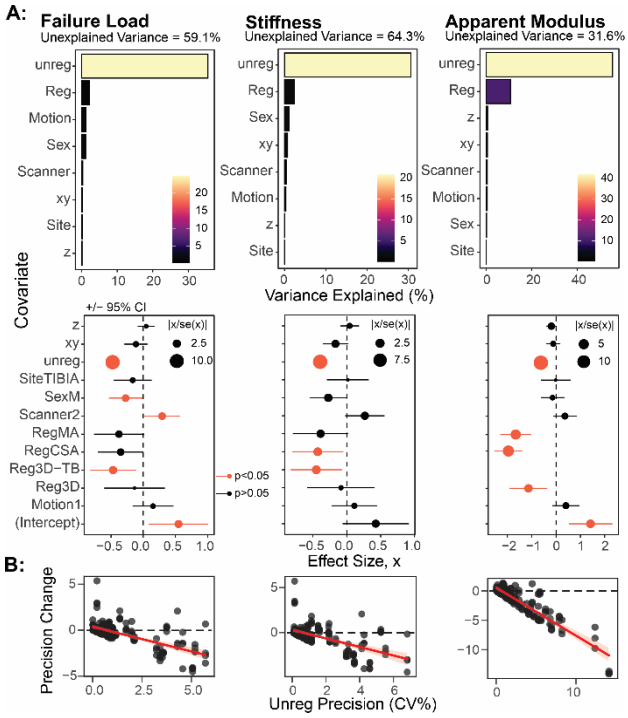


Figure 22. Regression analysis of registration-associated changes in precision errors for microFE outcomes. A) For each HR-pQCT measure, **top panel** shows percentage of variance explained by each model covariate and **middle panel** shows model covariate coefficients. The color bars indicate % of explained variance. For model covariate coefficient plots, effect sizes x were scaled by standard errors se(x) and error bars represent 95% confidence intervals (CI). Dependent variable in the regression model is the change in precision error following registration for each measurement, and independent variables are anatomical site (reference = radius), motion status (reference = 0 [no motion]), scanner generation (reference = 1 [XCT scanner]), unregistered precision error (unreg), registration method (Reg, reference = unregistered), rotation angles (xy and z), and sex (reference = F [female]). **B)** the **scatter plots** indicate the crude association between the unregistered precision error and changes in precision after registration. 3D, three-dimensional registered; 3D-TB, three-dimensional registered with transformed boundary conditions; CSA, cross-sectional area registered; MA, matched-angle registered; Unreg, unregistered.

4.3.2.2. Subject-level registration performance:

Anatomical site. At the radius, registration improved precision in 50-75% of failure load estimates, 70-75% of stiffness estimates, and 82-91% of apparent modulus estimates, and there was no significant difference in the magnitude of improvement or deterioration between registration

methods, as measured by the AUCs of the RPP (**Figure 23A**). At the tibia, registration improved the precision in 50-73% of failure load estimates, 55-62% of stiffness estimates, and 58-77% of apparent modulus estimates. We observed differences across registration methods in the tibia that were otherwise absent in the radius. Notably among tibial scans that resulted in poorer precision following registration, we found that CSA was significantly less detrimental than all other methods for failure load, while only 3D was more detrimental for stiffness. Among tibial scans that improved following registration, 3D, 3D-TB, and MA offered significantly greater improvement than CSA for failure load. For stiffness, only MA offered larger improvement than CSA.

Motion status. For scans with no motion, registration improved precision in 35-48% of failure load estimates, 48-59% of stiffness estimates, and 63-74% of apparent modulus estimates, whereas for scans with motion, registration improved precision in 55-74% of failure load estimates, 62-79% of stiffness estimates, and 74-95% of apparent modulus estimates (**Figure 23B**). Among scans without motion that were less precise following registration (**Figure 23B**), 3D and MA were more detrimental compared to CSA for failure load. Among scans with motion, there was no difference in the magnitude (i.e., AUC of RPP) of improvement or deterioration following registration between methods.

Scanner generation. For XCT scans, registration improved precision in 48-71% of failure load estimates, 60-74% of stiffness estimates, and 82-88% of apparent modulus estimates, whereas for XCT2 scans, registration improved precision in 48-72% of failure load estimates, 53-59% of stiffness estimates, and 63-81% of apparent modulus estimates (**Figure 23C**). For XCT2 scans, only MA registered stiffness had superior improvement compared to CSA among scans that were more precise following registration.

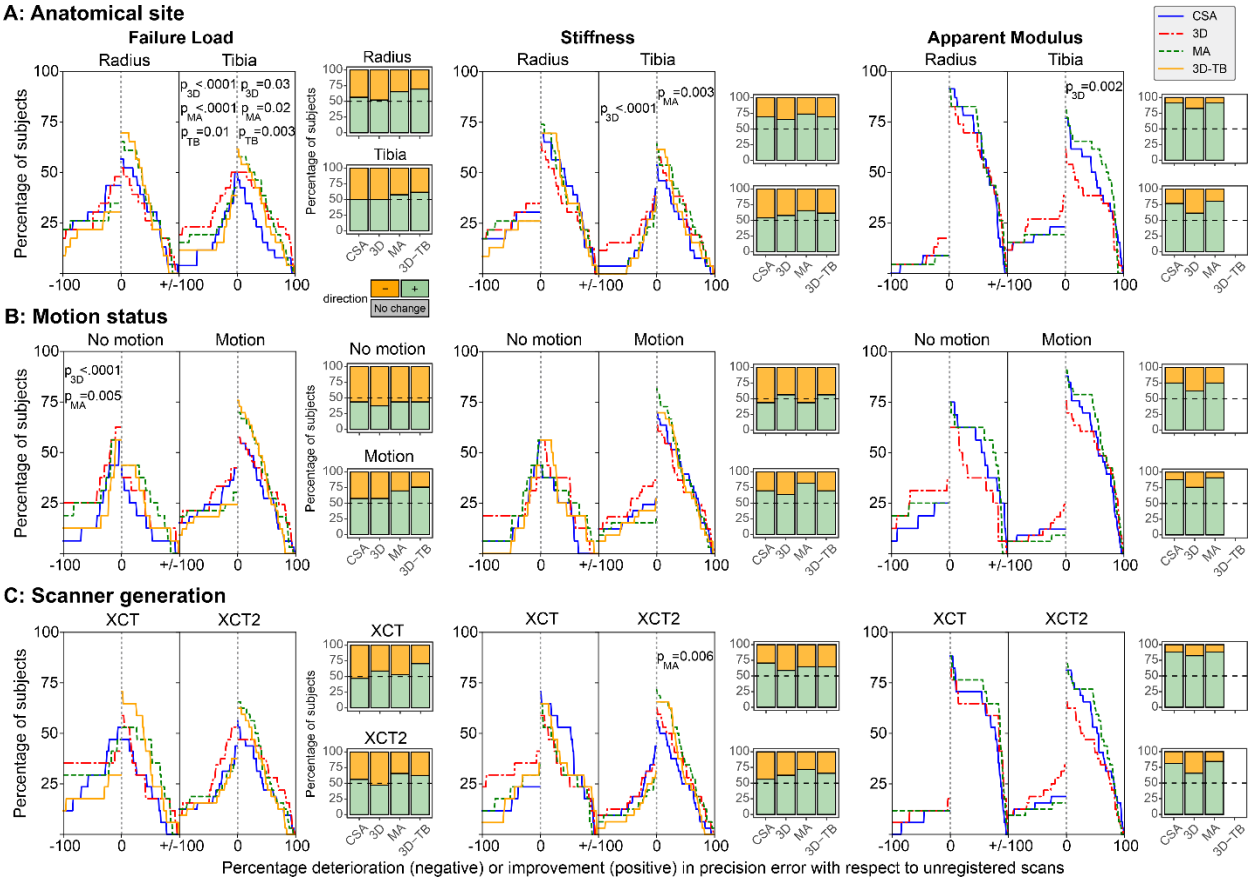


Figure 23. RPPs for microFE outcomes stratified by anatomical site, motion status, and scanner generation. A-B) RPP analyses comparing registration methods by anatomical site (**A**), motion status (**B**), and scanner generation (**C**). Apparent modulus was not reported for 3D-TB registration because area could not be calculated for scans with partial cross-sections and different orientations. For each parameter, **left panel** represents RPP plot and **right panel** represents the proportion of individual scans that experience improvement (green), no change (grey) or deterioration (orange) following registration (corresponds to y-axis intercept on RPP plot). For RPP plots, x-axes show the percentage change in precision error relative to unregistered precision error for a registration method. Positive values of x mean improvement in precision error when using image registration, whereas negative values indicate deterioration. Y-axis indicated the percentage of individuals for whom a percentage of improvement or deterioration is obtained using image registration. For positive values on x-axis, higher percentage of individuals means better performance, while for negative values on x-axis, lower percentage of individuals is better. The difference between 100% and the sum of negative and positive curves at x=0 is the percentage of individuals with improvements in precision less than 1%. P-values indicate significant difference in area under the curve (AUC) separated by improvement or deterioration. Please note that the AUC comparison is performed for precision changes ranging between $[-Inf, 100]$, whereas the

RPP plots are bound by $[-100, 100]$. Therefore, some of the statistical results from the deterioration side may not be acknowledged on the RPP plots.

4.4. Discussion:

Our study compared the effect of the CSA and various 3D image registration methods on the precision of HR-pQCT measurements. We assessed the influence of various covariates, including anatomical site, motion and scanner generation on the performance of registration methods using regression, pairwise comparisons and relative performance plot-based analyses.

Density and microstructural parameters. Consistent across all analytic approaches, we found that image registration improved the precision of total and trabecular bone mineral densities, trabecular and cortical bone mineral contents, area measurements, trabecular bone volume fraction, separation, and heterogeneity, as well as cortical thickness and perimeter (**Supplemental Results Figure 3-A**). These findings are in line with prior studies[6–9] and provide further support that image registration yields improved precision of HR-pQCT measurements and microFE outcomes, with the 3D method being marginally superior. For the radial density and microstructural outcomes, our data were consistent with Ellouz et. al.[9], showing precision improvements (i.e., compared to unregistered precision) following 3D or CSA registration of Tt.vBMD and Ct.Th. However, while Ellouz et al. showed improved precision of radial and tibial Ct.Po after 3D registration, we only observed an association between 3D registration and improvement in Ct.Po precision after adjustment for other factors (e.g., motion, scanner generation, and site). This could be due to their use of a fixed rather than relative to bone length volume of interest, as different porosity measurements have been observed between the two regions.[17,27,28] However, the use of the fixed region can introduce bias when properties are compared across individuals with different bone lengths, such as those suffering from OI. Also, their study population consisted of 15 healthy subjects (aged 21–47 years), with time intervals between repeated scans ranging between the same day and a week (i.e., potential drift and fluctuations in x-ray tube), whereas all our repeated scans were performed on the same day. Finally, they only included scans from a first-generation HR-pQCT (XCT), thus the lower resolution likely affected the porosity measurements. Chiba et. al.[8] did not report unregistered precision errors, and only compared 3D and CSA precision errors. Compared to CSA, they showed improved precision using 3D registration for radial Tb.BV/TV, Ct.Po, and Ct.Ar. At the tibia, they showed that 3D registration was more precise for Tb.Sp, Tb.Ar, Ct.Po, Ct.Pm, and Ct.Ar.[8] Despite

different populations, our observed differences between 3D and CSA registrations were consistent with those of Chiba et. al., except for radial Tb.BV/TV, tibial Tt.vBMD and Tt.vBMC, and radial and tibial Ct.Po. These discrepancies could be due to their use of a fixed, rather than relative, volume of interest. Kemp et. al.[7] showed improved precision of radius measurements after registration for all BMD indices, Tb.BV/TV, Tb.N (3D only), Tb.Sp, Tb.Ar (CSA only), Ct.Th, and Ct.Ar (CSA only). For the tibia, they observed similar effects except that, Ct.vBMD, Tb. N and Tb.Sp were not affected by registration. Our results were consistent with Kemp et. al. except for Tb.vBMD (radius), Ct.vBMD (radius), Ct.Th (tibia), Tb.N (radius), and Tb.Sp. Of note, the differences between studies were independent of scanner generation and rather could be a result of different scanned regions (Kemp et. al. scanned the fixed offset region, whereas we scanned the region with an offset relative to the bone length). In agreement with our findings, MacNeil and Boyd[6] also showed improved precision for 3D registration compared to CSA registration. Overall, there is overwhelming consensus that registration improves precision of density and microstructural parameters, however, there remains a lack of consensus to the extent of this improvement on each parameter due to variations across studies. With respect to the current study, these differences are likely due to the multicenter design versus the single-center design of all other studies (not representative of multicenter clinical trials), and anatomical region scanned (relative vs. fixed). Also, the patient population could represent a source of difference, although our regression analyses did not demonstrate an association between OI type and changes in precision error after registration.

MicroFE parameters. Regarding microFE, we showed the beneficial effect of registration, especially for apparent modulus. Ellouz et. al.[9] only reported results for stiffness where they showed similar precision between the unregistered, CSA, and 3D-registered scans for radius and tibia. This finding is consistent with ours, although we showed better radial precision for the CSA and 3D methods prior to p-value correction. Plett and colleagues[5] only examined *ex vivo* precision for CSA and 3D-TB methods using cadaveric radii, and showed that both methods significantly improved the precision of stiffness, and 3D-TB improved the precision for failure load compared to no registration. They reported no differences between the CSA and 3D methods although 3D registration tended to be more precise, consistent with our findings. For the first time, we showed that MA-registered microFE provides similar improvements in precision as 3D-TB (slightly better

than CSA), suggesting that efforts to remove the bias in interpolation error (MA), or prevent interpolation error (3D-TB) are beneficial. Of note, after adjusting for other factors, MA registration was not associated with improved precision for failure load and stiffness, while CSA and 3D-TB registrations were.

Covariates influencing registration performance. Our regression analyses identified several covariates that influenced the performance of registration. For microFE outcomes, radial scans were associated with larger registration-associated improvements in precision errors than tibial scans. Interestingly, the presence of motion artefacts was not associated with changes in precision after registration, whereas motion was associated with poorer precision compared to no motion, even after registration (**Supplemental Results Figure 10**). This suggests that motion is detrimental even if images are properly registered, which is because motion effectively distorts the image while image registration assumes rigid transformations. As a result, even if registration has high accuracy, the two bones sections will not be completely similar. We conclude that motion artifacts independently contribute to precision error, and this emphasizes the importance of acquiring scans with minimal motion.

Subject-level registration performance. While pairwise comparisons evaluate the aggregate performance of registration, our novel RPP analyses provide further insight into the subject-level performance. Collectively, our RPP analyses from the stratified data based on anatomical site, motion status, and scanner generation suggest image registration has a high likelihood of improving precision in individual scans for Tt.vBMD, Tb.vBMC, Ct.vBMC, Tb.aBMC, Ct.aBMC, Tb.BV/TV, Tb.Ar, Ct.Pm and Ct.Ar measures, while registration can introduce error under certain conditions in Ct.vBMD, Tt.vBMC, Tb.N, Tb.Th, Tb.Sp and Ct.Po. Our data suggest that the main reason behind this deterioration of precision is that the precision of unregistered scans is already high, hence it becomes more likely that the registration introduces, rather than corrects, error. The caveat associated with this finding is that since unregistered precision is seldom quantified in a clinical trial setting, all the information required to determine whether registration will be beneficial or detrimental will rarely be available. This makes individual-tailored registration recommendations unfeasible and registration recommendation will have to be made on the basis of practices that will yield better results *on average*. In this regard, we found that 3D registration offers marginal improvements over CSA registration for several

parameters. For apparent modulus, registration by any method leads to improved precision in most scans, and for failure load and stiffness, the 3D-TB and MA methods marginally outperformed other methods.

Precision error sources. The main sources of precision error for the density and microstructural outcomes are repositioning (translational and rotational), contouring, and motion. Considering that *i*) contouring errors and motion affected all registration methods in this study similarly because we used the same scans and did not change contours between registration methods, *ii*) all registration methods result in significant reductions in precision error, and *iii*) only marginal differences between 3D and CSA methods are observed, we can conclude that translational misalignment is the principal source of error corrected by registration of density and microstructural outcomes. The smaller effect of rotational misalignment is due to the relatively small xy rotation angles because of using the standard casts, as a rotational misalignment around z-axis is more likely to happen, yet it does not result in error. The reason is that in the presence of rotation around z-axis, the same bone length is still being imaged (i.e., the effect of rotation around the z-axis is similar to rotating the screen when viewing the image, hence not affecting the scanned volume). Similar reasoning applies to microFE outcomes. The efficacy of 3D registration for microFE outcomes depends on its implementation. In contrast to 3D-registered nonFE outcomes, the 3D or MA registered microFE outcomes require image rotation. Consequently, some possible sources of error for 3D-registered microFE are *i*) interpolation (i.e., rotating image in space) and *ii*) bias in interpolation (i.e., rotating only one of the images to align it with the other image versus rotating both images, each half-way and in opposite directions). These sources of error are applicable to the grayscale images and binary masks. In our supplementary analyses, we developed alternative variations of the 3D and MA-registered scans, by eliminating the effect of the periosteal mask (i.e., we dilated the mask of the 1st image and used it to segment both aligned images, hence prevented any effect from the differences in the masks, and their interpolation error). These methods are named as having the "same periosteal" mask (3D-SP and MA-SP) (more details in **supplemental methods**). Interestingly, we found comparable performance between the standard 3D and MA methods, and their alternatives (i.e., 3D-SP and MA-SP), suggesting that interpolation error on the edges of the periosteal mask is not a considerable source of error for microFE (**Supplemental Results Figure 11**). This further means that alternatively, the grayscale images

can be masked prior to rotation and segmentation (rather than masking after segmentation), hence no need for the rotation of periosteal masks. Our results also indicated that the interpolation error of the grayscale images (i.e., affecting both the cortical and trabecular structures), or bias in such interpolation is not considerably detrimental for 3D-registered microFE (3D), although preventing interpolation error or its bias is beneficial by either transforming the boundary conditions instead of the images (3D-TB), or rotating both images (MA).

Applications of image registration. Image registration is used to ensure that the same bone volume is evaluated in an individual overtime. A disadvantage of image registration is the reduced size of the analyzed region, which is even larger for microFE due to the flattening, and the dependency of stiffness on the sample height. Since image registration changes the analyzed volume, specifically for each set of paired scans, comparison of inter-subject HR-pQCT measures post-registration is not advised, unless relative changes are being evaluated. Needless to say, image registration is not relevant for cross-sectional studies since only one image is available per subject. Similarly, for any cross-sectional analyses within longitudinal studies, where changes over time are not of interest, image registration is not recommended.

Recommendations for longitudinal studies. Image registration depends on identifying common features to align scans, and consequently, the accuracy of registration deteriorates as scan sets become dissimilar due to biological changes arising during longitudinal studies. This loss of performance is expected to be smaller for 3D registration compared to CSA registration because the latter aligns the scans only based on matching cross-sectional-area, whereas 3D registration uses a collection of features including area and trabecular structure. In term of processing time, the CSA-method is faster than 3D-registration, however, the additional steps of 3D image registration and image transformations can be done within 5 to 15 minutes depending on the scanner generation and anatomical site (i.e., XCT2 and tibial scans require longer cpu-time). Therefore, although 3D-registration only provides marginal improvements in precision compared to the CSA-registration, the additional computational time for 3D-registration is also minor. Regarding 3D-registration for microFE, the 3D-TB and MA methods are superior despite their marginal improvement compared to the standard 3D registration (i.e., no significant difference). The 3D-TB method does not suffer from interpolation error, which despite being minimal, can affect the accuracy of the model. MA registration provides similar precision as 3D-TB, but also enables the calculation of apparent

modulus. Based on these considerations, we recommend using 3D (3D-TB or MA registration for microFE) registration for longitudinal studies. We must emphasize that even marginal changes in precision error translate to larger changes in the LSCs (i.e., 2.77 times), which is in the order of expected biological changes in bone. The summary of findings including $CV\%_{rms}$, $CV\%_{med}$, and LSCs are provided in **Supplemental Results Tables 1 and 2**.

Generalizing to other populations. The current study was conducted as part of a multi-center clinical trial on individuals with OI. The purposes of our precision study were firstly to enable the interpretation of the trial data. Additionally, we investigated the effect of various image registration methods on our precision errors, which are presented here. We anticipate that the findings of this study can be extended to populations other than individuals with OI. Firstly, the absolute precision errors of HR-pQCT measures obtained from OI participants in our study are comparable to those of healthy adults evaluated in a meta-analysis by Mikolajewicz and colleagues[29]. Moreover, our regression analyses did not demonstrate an association between OI type and changes in precision error after registration, suggesting that bone quality is not a covariate of consequence. Finally, we focused on changes in precision error after registration, and all registration methods were applied to the same images, hence eliminating the effect of the input image. Nevertheless, we acknowledge the possibility of differences and future studies on other populations are required for confirmation.

Strengths and limitations. Our study has several *strengths*, first of which is the inclusion of repeated scans from multiple scanning sites. This study design ensures proper translation of our precision errors to multicenter clinical trials. Moreover, we included radial and tibial scans from both scanner generations. We implemented a systematic approach to assess the precision of microFE outcomes by comparing multiple ways of performing the 3D-registration and provide detailed scripts documenting our methods. This approach enabled us to understand the limiting factors and identify the optimal 3D-registered microFE method in terms of precision. Nevertheless, we acknowledge that future studies that compare different variations of 3D-registration using the data obtained in research setting at a single scanning site will be beneficial. This study also has several *limitations,* including a moderate sample size, which could limit our ability to detect some of the smaller differences between registration methods. To mitigate this, we pooled (increased statistical power) and stratified (reduced statistical power) the data to determine the consistency of

our findings. Related to this, there were sample size imbalance across scanner generations, OI types, and genders. Furthermore, although planned initially, we could not obtain usable repeated scans from participants with OI type III. Moreover, our study was performed on participants with osteogenesis imperfecta type I and IV and thus further studies are required to definitively confirm that our results can be applied to other populations. Finally, this study focused on short-term reproducibility outcomes which does not allow for the evaluation of registration on precision errors following biological change.

In conclusion, our results emphasize the importance of image registration to improve the short-term precision of HR-pQCT measurements, especially in the context of longitudinal data. The main outcomes of this study are: **1)** Although 3D and CSA registration methods perform similarly, we recommend 3D methods due to marginal improvements over the CSA method, and its reduced dependence on bone area, which is subject to change over time; **2)** 3D registration for microFE analysis is most beneficial when done using the 3D-TB or MA methods, which prevent interpolation error or bias in interpolation error, respectively; **3)** this study contributes to the utilization and standardization of 3D registration for HR-pQCT image analysis by providing detailed methods to perform the various 3D-registration techniques; **4)** our analyses highlight the importance of obtaining scans with minimal motion, as our data showed that registration only partially corrects motion-related error.

Conflict of Interest Statement

This research was supported by Mereo Biopharma. BMW and FR have received institutional research support and materials and are consultants for Mereo BioPharma.

Author Contributions

Study conception and design: SH, EAZ, BMW; Data acquisition: SH, EAZ, MR, BS, CJ, BMW; Analysis and interpretation of data: SH, NM, EAZ, FR, BMW; Drafting of Manuscript: SH, NM, BMW. All authors contributed to the critical revision and approval of the final manuscript.

Acknowledgements

We thank all the study participants and staff at the imaging sites. The participating imaging sites are: Klinik for Knogleskørhed, Aarhus Universitets hospital; Children's Nutrition Research Center, Baylor College of Medicine; Cambridge Clinical Research Centre, Cambridge University Hospitals NHS Foundation Trust–Addenbrooke's Hospital; Hôpital Edouard Herriot; Rhumatologie Centre, Hôpital Lariboisière; Institute for Clinical and Translational Research, Johns Hopkins University; Oregon Health and Science University; Endokrinologisk Afdeling, Odense Universitetshospital; Shriners Hospitals for Children-Canada; Osteoporosis Centre, Southampton General Hospital NHS Foundation Trust; Osteoporosis Clinic, Toronto General Hospital; Institute of Imaging Science, Vanderbilt University; and the Center for Advanced Medicine, Division of Bone and Mineral Diseases, Washington University School of Medicine in St Louis. We also thank Drs. Sarah Manske and Steven Boyd for helpful discussions and assistance with programming related to the implementation of the 3D registration methods. The study was funded by Mereo BioPharma, Shriners Hospitals for Children, and FRQS Programme de bourses de chercheur (BMW, SH, MR).

References:

1. Whittier DE, Boyd SK, Burghardt AJ, Paccou J, Ghasem-Zadeh A, Chapurlat R, Engelke K, Bouxsein ML. Guidelines for the assessment of bone density and microarchitecture in vivo using high-resolution peripheral quantitative computed tomography. Osteoporos. Int. 2020;
2. 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.
3. van Rietbergen B, Biver E, Chevalley T, Ito K, Chapurlat R, Ferrari S. A Novel HR-pQCT Image Registration Approach Reveals Sex-Specific Changes in Cortical Bone Retraction With Aging. J. Bone Miner. Res. 2021;00(00):1–13.
4. de Bakker CMJ, Altman AR, Li C, Tribble MB, Lott C, Tseng WJ, Liu XS. Minimizing

Interpolation Bias and Precision Error in In Vivo μCT-Based Measurements of Bone Structure and Dynamics. Ann. Biomed. Eng. 2016;44(8):2518–28.
5. Plett RM, Kemp TD, Burt LA, Billington EO, Hanley DA, Boyd SK. Using 3D image registration to maximize the reproducibility of longitudinal bone strength assessment by HR-pQCT and finite element analysis. Osteoporosis International; 2021;
6. MacNeil JA, Boyd SK. Improved reproducibility of high-resolution peripheral quantitative computed tomography for measurement of bone quality. Med. Eng. Phys. England; 2008 Jul;30(6):792–9.
7. Kemp TD, de Bakker CMJ, Gabel L, Hanley DA, Billington EO, Burt LA, Boyd SK. Longitudinal bone microarchitectural changes are best detected using image registration. Osteoporos. Int. Osteoporosis International; 2020;
8. Chiba K, Okazaki N, Isobe Y, Miyazaki S, Yonekura A, Tomita M, Osaki M. Precision of 3D Registration Analysis for Longitudinal Study of Second-Generation HR-pQCT. J. Clin. Densitom. 2020;24(2).
9. Ellouz R, Chapurlat R, van Rietbergen B, Christen P, Pialat JB, Boutroy S. Challenges in longitudinal measurements with HR-pQCT: Evaluation of a 3D registration method to improve bone microarchitecture and strength measurement reproducibility. Bone. 2014;63:147–57.
10. Joan C Marini, Antonella Forlino, Hans Peter Bächinger, Nick J Bishop, Peter H Byers, Anne De Paepe, Francois Fassier, Nadja Fratzl-Zelman, Kenneth M Kozloff, Deborah Krakow, Kathleen Montpetit OS. Osteogenesis imperfecta. Nat. Rev. Dis. Prim. 2015;3:521–4.
11. Varga P, Willie BM, Stephan C, Kozlo KM, Zysset PK, Currey J. Finite element analysis of bone strength in osteogenesis imperfecta. 2020;133(January).
12. Chiba K, Yamada S, Yoda I, Era M, Yokota K, Okazaki N, Ota S, Isobe Y, Miyazaki S, Tashiro S, Nakashima S, Morimoto S, Sato S, Tsukazaki T, Watanabe T, Enomoto H, Yabe Y, Yonekura A, Tomita M, Ito M, Osaki M. Effects of monthly intravenous ibandronate on bone mineral density and microstructure in patients with primary osteoporosis after teriparatide treatment: The MONUMENT study. Bone. 2021;144(November 2020):115770.
13. Cheung AM, Majumdar S, Brixen K, Chapurlat R, Fuerst T, Engelke K, Dardzinski B, Cabal A, Verbruggen N, Ather S, Rosenberg E, De Papp AE. Effects of odanacatib on the radius and tibia of postmenopausal women: Improvements in bone geometry, microarchitecture, and estimated bone strength. J. Bone Miner. Res. 2014;29(8):1786–94.
14. Engelke K, Stampa B, Timm W, Dardzinski B, De Papp AE, Genant HK, Fuerst T. Short-term in vivo precision of BMD and parameters of trabecular architecture at the distal forearm and tibia. Osteoporos. Int. 2012;23(8):2151–8.
15. Burghardt AJ, Liu D, McKay HA, Khosla S, Patsch JM, Shane E, Valentinitsch A, Engelke K, Chapurlat R, Boyd SK, Szabo E, Boutroy S, Kazakia GJ, Pialat J-B, Bogado CE, Majumdar S, Zanchetta MB, Bouxsein ML. Multicenter precision of cortical and trabecular bone quality measures assessed by high-resolution peripheral quantitative computed tomography. J. Bone Miner. Res. 2012;28(3):524–36.
16. Mikolajewicz N, Zimmermann EA, Rummler M, Hosseinitabatabaei S, Julien C, Glorieux FH, Rauch F, Willie BM. Multisite longitudinal calibration of HR-pQCT scanners and precision in osteogenesis imperfecta. Bone. 2021;147(February):115880.

17. Bonaretti S, Majumdar S, Lang TF, Khosla S, Burghardt AJ. The comparability of HR-pQCT bone measurements is improved by scanning anatomically standardized regions. Osteoporos. Int. Osteoporosis International; 2017;1–14.
18. Pialat JB, Burghardt AJ, Sode M, Link TM, Majumdar S. Visual grading of motion induced image degradation in high resolution peripheral computed tomography: Impact of image quality on measures of bone density and micro-architecture. Bone. 2012;50(1):111–8.
19. Boyd STKB, Oser STM, Uhn MIK, Linck ROJK, Rauze PELK. Evaluation of Three-dimensional Image Registration Methodologies for In Vivo Micro-computed Tomography. 2006;34(10):1587–99.
20. 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.
21. Whittier DE, Manske SL, Kiel DP, Bouxsein M, Boyd SK. Harmonizing finite element modelling for non-invasive strength estimation by high-resolution peripheral quantitative computed tomography. J. Biomech. 2018 Oct;80:63–71.
22. Pistoia W, van Rietbergen B, Lochmüller E-M, Lill CA, Eckstein F, Rüegsegger P. Estimation of distal radius failure load with micro-finite element analysis models based on three-dimensional peripheral quantitative computed tomography images. Bone. United States; 2002 Jun;30(6):842–8.
23. Hosseinitabatabaei S, Kawalilak C, McDonald M, Kontulainen S, Johnston J. Distal radius sections offer accurate and precise estimates of forearm fracture load. Clin. Biomech. 2020;80(November 2019):105144.
24. Glüer CC, Blake G, Lu Y, Blunt BA, Jergas M, Genant HK. Accurate assessment of precision errors: How to measure the reproducibility of bone densitometry techniques. Osteoporos. Int. 1995;5(4):262–70.
25. Shuhart CR, Yeap SS, Anderson PA, Jankowski LG, Lewiecki EM, Morse LR, Rosen HN, Weber DR, Zemel BS, Shepherd JA. Executive Summary of the 2019 ISCD Position Development Conference on Monitoring Treatment, DXA Cross-calibration and Least Significant Change, Spinal Cord Injury, Peri-prosthetic and Orthopedic Bone Health, Transgender Medicine, and Pediatrics. J. Clin. Densitom. 2019;22(4):453–71.
26. Bunyamin A, Björkman K, Kawalilak C, Hosseinitabatabaei S, Teare A, Johnston J, Kontulainen S. Reliability of Annual Changes and Monitoring Time Intervals for Bone Strength, Size, Density, and Microarchitectural Development at the Distal Radius and Tibia in Children: A 1-Year HR-pQCT Follow-Up. J. Bone Miner. Res. 2019;34(7):1297–305.
27. Shanbhogue V V., Hansen S, Halekoh U, Brixen K. Use of relative vs fixed offset distance to define region of interest at the distal radius and tibia in high-resolution peripheral quantitative computed tomography. J. Clin. Densitom. 2015;18(2):217–25.
28. Okazaki N, Chiba K, Burghardt AJ, Kondo C, Doi M, Yokota K, Yonekura A, Tomita M, Osaki M. Differences in bone mineral density and morphometry measurements by fixed versus relative offset methods in high-resolution peripheral quantitative computed tomography. Bone. 2021;149(April):115973.
29. Mikolajewicz N, Bishop N, Burghardt AJ, Folkestad L, Hall A, Kozloff KM, Lukey PT, Molloy-Bland M, Morin SN, Offiah AC, Shapiro J, van Rietbergen B, Wager K, Willie

BM, Komarova S V, Glorieux FH. HR-pQCT Measures of Bone Microarchitecture Predict Fracture: Systematic Review and Meta-analysis. J. Bone Miner. Res. 2019;

Supplemental Methods:

Inclusion/exclusion criteria: Inclusion criteria included male and female participants with a clinical diagnosis of OI Type I, III or IV with a *COL1A1*/*COL1A2* defect confirmed by genetic testing, age greater than 18 years, and one or more non-traumatic long bone, rib, hand/feet and/or vertebral fracture(s) in the past five years. Female participants were ineligible if pregnant, breastfeeding, or following contraceptive guidance.

Participants were ineligible if they were greater than 75 years of age, or had a history of the following: skeletal malignancies/bone metastases; neural foraminal stenosis; uncontrolled diseases affecting bone metabolism; skeletal conditions leading to long bone deformities or increased fracture risk other than OI; bisphosphonate treatment 3 months prior to baseline; teriparatide, denosumab or other anabolic or anti-resorptive medication within 6 months prior to baseline; myocardial infarction, agina pectoris, ischaemic stroke, or transient ischaemic attack; alcohol or drug abuse in 12 months prior to dosing; significant psychiatric or medical disorder affecting compliance to study protocol; history of external radiation; participation in any clinical investigation within 4 weeks or 5 half-lives of the drug prior to dosing; or allergy to the study drug.

List of the reported parameters:

Supplemental Methods Table 1. Complete list of reported HR-pQCT outcomes, their descriptions, and units.

Parameter (Abbreviation)	Measurement/calculation method	Units
Volumetric bone mineral density (vBMD)		
1. Total (Tt.vBMD)	Mean mineral density of voxels within the periosteal contour	mg HA/cm^3
2. Cortical (Ct.vBMD)	Mean mineral density of voxels within the cortical compartment	mg HA/cm^3
3. Trabecular (Tb.vBMD)	Mean mineral density of voxels within the endosteal contour	mg HA/cm^3
Volumetric bone mineral content (vBMC)		
4. Total (Tt.vBMC)	Tt.vBMD multiplied by total volume	mg HA
5. Cortical (Ct.vBMC)	Ct.vBMD multiplied by cortical volume	mg HA
6. Trabecular (Tb.vBMC)	Tb.vBMD multiplied by trabecular volume	mg HA
Areal bone mineral content (aBMC)		
7. Cortical (Ct.aBMC)	Ct.vBMD multiplied by cortical area	mg HA/cm
8. Trabecular (Tb.aBMC)	Tb.vBMD multiplied by trabecular area	mg HA/cm
Area (Ar)		
9. Cortical (Ct.Ar)	Mean area of slices within the cortical compartment	mm^2
10. Trabecular (Tb.Ar)	Mean area of slices within the endosteal contour	mm^2
Cortical (Ct.) microstructure[1,2]		
11. Thickness (Ct.Th)	Mean distance between periosteal and endosteal contours calculated using distance transformation	mm
12. Porosity (Ct.Po)	Ratio of pore volume to total cortical volume	mm^3/mm^3 (%)
13. Perimeter (Ct.Pm)	Mean length of periosteal contour	mm
Trabecular (Tb.) microstructure[3,4]		
14. Bone volume fraction (Tb.BV/TV)	Tb.vBMD divided by 1200 (XCT) or ratio of bone voxels to total voxels within endosteal contour (XCT2)	mm^3/mm^3
15. Number (Tb.N)	Mean inverse spacing between the ridges of trabeculae	mm^{-1}
16. Thickness (Tb.Th)	Tb.BV/TV divided by Tb.N (XCT) or Mean spacing of segmented trabeculae measured using distance transformation (XCT2)	mm
17. Separation (Tb.Sp)	(1-Tb.BV/TV)/Tb.N (XCT) or Mean distance between trabeculae (XCT2)	mm
18. Inhomogeneity of trabecular network (Tb.1/N.SD)	The standard deviation of spacing between trabeculae ridges	mm

Micro Finite element analysis (FEA)		
19. Failure load	Estimated load at which a predefined percentage of voxels exceed a predefined level of strain	N
20. Stiffness	The total reaction force of bone divided by the applied displacement (1% strain)	N/mm
21. Apparent modulus	Ratio of apparent stress (reaction force divided by mean area) to apparent strain (1% of the model height)	kN/mm^2

HA, hydroxyapatite; XCT, XtremeCT; XCT2, XtremeCT II

Test statistics for area under the curve (AUC) of registration performance plots (RPP): For each pairwise comparison between registration methods, p-values for AUC differences were estimated from Z scores computed as $Z = \frac{AUC_A - AUC_B}{S_p}$, where $S_p = \sqrt{Var(AUC_A) + Var(AUC_B)}$, and A and B specify the registration methods being compared. $Var(AUC_X)$ indicates variance.

Justification for unbiased estimates of associations between unregistered precision errors and change in precision errors after registration (ΔCV) in the regression model: The common criticism of evaluating the association between a baseline measurement (e.g.., pre-registration precision), and changes in measurements from baseline to follow up (e.g., post-registration precision) is "mathematical coupling" and "regression to the mean".[5] This problem arises when one variable (e.g., dependent variable) directly or indirectly contains the whole or part of another (e.g., independent variable). Considering that each measurement is distorted by sampling error, a consequence of mathematical coupling is that both the dependent and independent variables contain common error terms. As a result, modeling the association between the independent and dependent variables inadvertently captures the relationship between error terms, which can cause biases in the association estimates. However, our regression model is protected from such bias because we derived paired unregistered and registered precisions estimates from the same images which ensures that common error terms negate each other when the difference between registered and unregistered precision errors is calculated. Below is a detailed proof:

There is a concern that since change is calculated from the baseline value, and that the measurement error (e_x) is present in both change and baseline estimates, baseline will always be correlated to change. Assume the following:

$$d = (y + e_y) - (x + e_x)$$

where d is change, y and x are post and pre-scores, and e_y and e_x are associated error terms, respectively. If we then construct a regression model evaluating the association between baseline score x and change d we will expect a correlation because e_x (t (the measurement error) is present in both terms (highlighted in yellow):

$$(y + e_y) - (x + e_x) \sim a(x + e_x)$$

where the left-hand side of the model represents change, and the right-hand side represents the baseline score x, its associated error e_x and the fitted coefficient a. Based on this construction, a will likely be negative because we are in part fitting the relationship between $-e_x$ and e_x. However, this model is misrepresentative of our study design since we never actually sample e_y. Instead e_y can be approximated to e_x since y is a transformation of x rather than a newly sampled data point. In reality e_y does not exactly equal e_x because we apply a transformation during the registration process. Nonetheless it should be highly correlated. The revised model then looks like this:

$$(y + e_{x\prime}) - (x + e_x) \sim a(x + e_x)$$

$$y - x \sim a(x + e_x)$$

where e_y is replaced by $e_{x\prime}$ which we approximate to equal e_x, and assume that the difference between $e_{x\prime}$ and e_x is approximately zero. Importantly, in this version the two e_x terms on the left-hand side cancel out, meaning that any correlation arising between e_x on the left- and right-hand sides of the equations is nullified (contrary to what was seen in the first version of the model that contains e_x and e_y terms). In other words, our model construction allows us to examine the relationship between baseline precision and change in precision, while controlling for any spurious associations arising from correlated errors.

Although including baseline values in a model describing change has been a matter of controversy, its justification depends on how the model is interpreted, as discussed by F. Lord[6] (see Lord's paradox). In brief, both models (a model with and without baseline precision estimates as covariates) are valid, however, depending on which model you choose, the interpretation changes:

I. ***Model 1***: model without baseline precision estimates as covariates
    - *Interpretation*: model describes average effect of registration on **group** precision errors (i.e., unconditional effect)

II. ***Model 2***: model with baseline precision estimates as covariates (*current study*)
    - *Interpretation*: model describes average effect of registration on **individual** precision errors (i.e., effect conditioned on baseline precision)

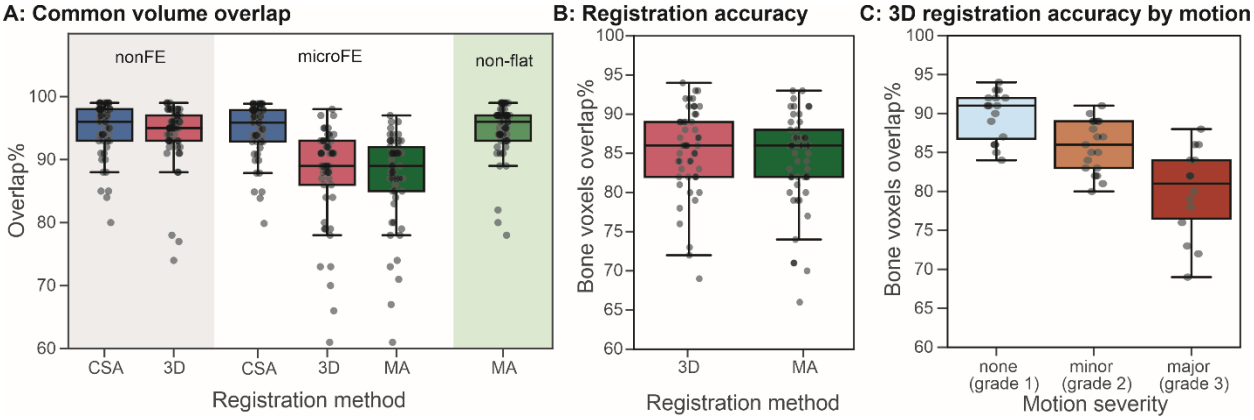
Therefore, considering our model does not suffer from biased estimates of associations due to common error terms, including the unregistered precision errors as model covariates allows us to precisely dissect the influence of unregistered precision error on the individual-level performance of registration. Adjusting for unregistered precision errors in our model further allows us to reveal the true relationship between other covariates (e.g., OI type or scanner generation) and registration-associated changes in precision errors.

Overlap percentages for different registration methods:

Two types of overlap percentages are reported. First, we reported the percentage of overlap for bone masks before and after registration to quantify the degree of repositioning error. Second, we computed the overlap between bone voxels after alignment during 3D or MA registration, which provides an estimate of registration accuracy.

For CSA registration, all of the scans had an overlap% of 80 and larger. For 3D registration, only 3 cases had an overlap of between 74-80%, and for MA only one case was 78%. The minimum overlap% for flattened volumes used for 3D or MA-registered microFE was 61%. However, we included these scans for several reasons, as long as the criteria for CSA was met. The first reason

was that typically it is more likely to have a 3D or MA overlap smaller than CSA, thus including those scans we minimized exclusions. Secondly, we did not observe any association between the overlap% of the masks and precision errors or changes in precision error.



Supplemental Methods Figure 1. A) Boxplots [median, interquartile range] showing the o**verlaps percentages** for all analyzed registration methods (CSA, 3D, MA). For CSA method, the overlap% is the same between microFE and nonFE outcomes since no flattening was required, whereas for 3D registration, microFE was performed on the flattened volumes. For MA registration was only used for microFE using a flattened volume, however the overlap% prior to flattening is also reported; **B)** The boxplots indicate the **overlap% of bone voxels** after 3D or MA registrations, indicating registration accuracy; **C)** Box plots showing **the overlap% of bone voxels after 3D registration** stratified by motion severity (none = grade 1, minor = grade 2, major = grade 3).

Alternative 3D or MA-registered microFE (same periosteal [SP]):

When performing 3D or MA-registered microFE, the periosteal mask can be treated differently to separate bone from soft-tissue: *i*) using different periosteal mask (the method used in the main study), or *ii*) using the same periosteal mask (SP). In the first method, the mask of each image, which is created after contouring is used to segment each bone from the surrounding tissue, whereas in the SP approach, the mask of the first image is dilated and used to segment both aligned images, hence preventing any effect from the differences in the masks, and their interpolation error on precision errors. Please note that using a dilated mask in fact eliminates the effect of periosteal mask, which could also be achieved by not using a periosteal mask at all. However, not using a periosteal mask would make the images unnecessarily large and increase the risk of error (e.g., including ulna in the region of interest). We found comparable performance between the standard

3D and MA methods, and their alternatives, suggesting that interpolation error is not a considerable source of error for microFE (**Supplemental Results Figure 11**). However, IPL scripts for performing the SP alternatives of 3D and MA-registered microFE can be found on Willie Lab Github repository.

References:

1. Burghardt AJ, Buie HR, Laib A, Majumdar S, Boyd SK. Reproducibility of direct quantitative measures of cortical bone microarchitecture of the distal radius and tibia by HR-pQCT. Bone. 2010;
2. 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.
3. Boutroy S, Bouxsein ML, Munoz F, Delmas PD. In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography. J. Clin. Endocrinol. Metab. 2005;90(12):6508–15.
4. Hildebrand T, Rüegsegger P. A new method for the model-independent assessment of thickness in three-dimensional images - Hildebrand - 2003 - Journal of Microscopy - Wiley Online Library. J. Microsc. [Internet]. 2015;185(1):67–75. Available from: http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2818.1997.1340694.x/abstract%5Cnhttp://files/158/abstract.html
5. Tu YK, Gilthorpe MS. Revisiting the relation between change and initial value: A review and evaluation. Stat. Med. 2007;26(2):443–57.
6. LORD FM. A PARADOX IN THE INTERPRETATION OF GROUP COMPARISONS. Psychol. Bull. 1967;

Supplemental Results:

Supplemental Results Table 1. Density and microstructural precision errors for different registration methods stratified by site and scanner. $CV\%_{rms}$, LSC%, and $CV\%_{median}$ are shown along with overlap percentages for each registration method. Difference between registration methods were tested using Wilcoxon signed rank test. 3D, three-dimensional; CSA, cross-sectional area; LSC, least significant change; n, number of scans; Unreg., unregistered. XCT, XtremeCT; XCT2, XtremeCT II.

	XCT-Radius (n=8)			XCT2-Radius (n=15)			XCT-Tibia (n=9)			XCT2-Tibia (n=17)		
	Unreg.	**CSA**	**3D**	**Unreg.**	**CSA**	**3D**	**Unreg.**	**CSA**	**3D**	**Unreg.**	**CSA**	**3D**
Overlap	N/A	91%	91%	N/A	95%	94%	N/A	94%	94%	N/A	96%	95%
Density [$CV\%_{rms}$ (LSC%) $\underline{CV\%_{median}}$]												
Tt.vBMD	3.76(10)	0.57(1.6)	0.22(0.6)[c]	4.65(13)	1.30(3.6)[c]	1.1(3.0)[b]	1.16(3.2)	0.32(0.9)	0.51(1.4)	1.52(4.2)	0.69(1.9)[c]	0.82(2.3)[c]
	2.39	0.27	0.20	1.55	0.71	0.57	1.14	0.24	0.52	0.65	0.19	0.20
Tb.vBMD	2.50(6.9)	1.23(3.4)	0.66(1.8)	1.82(5.4)	1.53(4.2)	1.12(3.1)	3.03(8.4)	2.45(6.8)	2.74(7.6)[b]	2.65(7.3)	1.34(3.7)[b]	0.90(2.5)[b]
	1.39	0.44	0.54	0.96	0.79	0.71	1.42	0.84	0.70	0.60	0.27	0.38
Ct.vBMD	1.41(3.9)	0.69(1.9)	0.70(1.9)	2.50(6.9)	1.64(4.5)	1.50(4.2)	0.60(1.7)	0.42(1.2)	0.43(1.2)	1.96(5.4)	1.44(4.0)	1.25(3.5)[b]
	1.03	0.49	0.22	0.83	0.54	0.62	0.42	0.27	0.31	0.38	0.35	0.24
Volumetric mass												
Tt.vBMC	0.40(1.1)	0.41(1.1)	0.19(0.5)	0.68(1.9)	0.68(1.9)	0.80(2.2)	0.28(0.8)	0.22(0.6)	0.35(1)	0.43(1.2)	0.33(0.9)	0.48(1.3)
	0.28	0.20	0.12	0.48	0.49	0.39	0.21	0.20	0.37	0.23	0.12	0.13
Tb.vBMC	5.22(14)	0.77(2.1)[b]	0.59(1.6)[b]	3.00(8.3)	1.03(2.9)[b]	0.67(1.9)[a]	1.47(4.1)	2.51(7)	1.87(5.2)	1.84(5.1)	0.63(1.7)[b]	0.44(1.2)[b]
	3.17	0.54	0.43	1.90	0.78	0.62	0.87	0.46	0.46	0.87	0.24	0.31
Ct.vBMC	2.21(6.1)	0.42(1.2)	0.39(1.1)	2.55(7.1)	0.80(2.2)[b]	0.71(2)[b]	0.81(2.2)	0.41(1.1)	0.55(1.5)	0.93(2.6)	0.36(1.0)[b]	0.52(1.4)[b]
	0.62	0.41	0.22	1.23	0.60	0.40	0.41	0.28	0.39	0.65	0.24	0.12
Areal mass												
Tb.aBMC	7.76(22)	1.16(3.2)[b]	N/A	4.24(12)	1.25(3.5)[b]	N/A	3.77(10)	2.47(6.8)	N/A	2.60(7.2)	0.88(2.5)[b]	N/A
	3.28	0.63		2.67	0.75		2.31	0.83		1.20	0.44	
Ct.aBMC	3.69(10)	0.63(1.7)	N/A	3.62(10)	1.1(3.1)[b]	N/A	1.29(3.6)	0.51(1.4)	N/A	1.31(3.6)	0.51(1.4)[b]	N/A
	1.30	0.50		1.76	0.94		0.56	0.31		0.92	0.35	
Area												
Tb.Ar	5.50(15)	0.18(0.5)[b]	N/A	5.23(14)	0.59(1.6)[a]	N/A	1.54(4.3)	0.19(0.5)[b]	N/A	1.74(4.8)	0.24(0.7)[b]	N/A

	3.39	0.14		2.33	0.22		1.22	0.03		0.84	0.10	
Ct.Ar	2.53(7)	0.88(2.4)[c]	N/A	3.97(11)	2.81(7.8)[b]	N/A	1.13(3.1)	0.92(2.5)	N/A	1.95(5.4)	1.80(5)	N/A
	0.85	0.31		1.24	0.92		0.52	0.61		0.74	0.65	
Trabecular microstructure												
Tb.BV/TV	2.71(7.5)	1.33(3.7)	0.66(1.8)	2.32(6.4)	2.12(5.9)	1.69(4.7)[c]	4.16(11)	2.65(7.3)	2.74(7.6)[c]	1.22(3.4)	0.65(1.8)[c]	0.41(1.1)[c]
	1.34	0.61	0.54	1.14	0.58	0.47	1.55	0.56	0.70	0.42	0.00	0.19
Tb.N	4.18(12)	3.29(9.1)	3.11(8.6)	2.48(6.9)	2.94(8.1)	3.22(8.9)	5.67(16)	5.35(15)	5.29(15)	3.63(10)	2.21(6.1)	1.94(5.4)
	3.75	2.76	2.76	0.80	0.80	0.77	1.68	1.46	1.48	0.73	0.59	0.40
Tb.Th	3.49(9.7)	3.14(8.7)	3.48(9.6)	1.36(3.8)	1.38(3.8)	0.99(2.7)	4.35(12)	6.79(19)	7.38(20)	0.66(1.8)	0.74(2)	0.76(2.1)
	1.94	2.71	3.01	0.34	0.38	0.35	1.30	0.94	0.91	0.00	0.26	0.19
Tb.Sp	4.29(12)	3.32(9.2)	3.08(8.5)	2.52(7)	2.84(7.9)	3.14(8.7)	5.75(16)	5.52(15)	5.25(14)	3.71(10)	2.14(5.9)[b]	2.03(5.6)[b]
	3.43	2.68	2.75	0.65	1.21	1.05	1.80	1.58	1.52	0.88	0.49	0.21
Tb.N.SD	4.01(11)	3.43(9.5)	2.27(6.3)	4.40(12)	4.26(12)	4.79(13)	3.59(9.9)	3.35(9.3)	3.37(9.3)	6.78(19)	4.89(13)[c]	4.40(12)[b]
	2.35	1.64	1.38	2.09	1.74	1.12	1.59	0.68	1.71	3.51	1.72	1.83
Cortical microstructure												
Ct.Th	4.44(12)	0.69(1.9)[b]	0.66(1.8)[b]	6.76(19)	2.97(8.2)[b]	2.12(5.9)[b]	1.70(4.7)	0.81(2.2)	1.02(2.8)	2.34(6.5)	1.81(5)	2.30(6.4)
	2.29	0.42	0.50	2.13	0.79	0.86	0.64	0.37	0.65	0.70	0.53	0.70
Ct.Pm	2.02(5.6)	0.24(0.7)[b]	N/A	2.34(6.5)	0.25(0.7)[b]	N/A	0.72(2)	0.09(0.2)[b]	N/A	0.76(2.1)	0.32(0.9)[b]	N/A
	1.21	0.12		0.97	0.10		0.33	0.07		0.31	0.12	
Ct.Po	9.17(25)	10.96(30)	9.06(25)	30.41(84)	30.05(83)	22.26(61)	3.85(11)	3.93(11)	4.01(11)	29.03(80)	30.05(83)	10.86(30)
	3.64	7.07	6.77	10.10	6.73	6.53	3.21	2.57	2.46	3.57	2.36	2.40

* Please note that no differences were observed between 3D and CSA methods.

Significance of Wilcoxon signed rank test after p-value correction are noted as follows:

[a] Significantly different from unregistered, $p<0.01$

[b] Significantly different from unregistered, $p<0.05$

[c] Tendency of difference from unregistered, $p<0.1$

Supplemental Results Table 2. MicroFE precision errors for different registration methods stratified by site and scanner. $CV\%_{rms}$, LSC%, and $CV\%_{median}$ are shown along with overlap percentages for each registration method. 3D, three-dimensional registered; CSA, cross-sectional area registered; LSC, least significant change; MA, matched-angle registered; n, number of scans; Unreg., unregistered. XCT, XtremeCT; XCT2, XtremeCT II.

$CV\%_{rms}$(LSC) $\underline{CV\%_{med}}$	**XCT-Radius (n=8)**					**XCT2-Radius (n=15)**				
	Unreg.	**CSA**	**3D**	**MA**	**3D-TB**	**Unreg.**	**CSA**	**3D**	**MA**	**3D-TB**
Overlap	N/A	91%	83%	83%	83%	N/A	95%	90%	90%	90%
Failure load	1.38(3.8) 0.47	0.75(2.1) 0.66	0.65(1.8) 0.58	0.81(2.2) 0.64	0.63(1.7) 0.63	4.48(12.4) 2.08	2.99(8.3) 1.42	3.16(8.8) 1.22	2.03(5.6) 1.32	1.95(5.4) 1.31
Stiffness	1.53(4.2) 0.90	0.68(1.9) 0.40	0.75(2.1) 0.38	0.83(2.3) 0.56	0.72(1.9) 0.43	4.60(12.7) 2.22	2.92(8.1) 1.27	3.20(8.9) 1.20	2.10(5.8) 1.32	2.18(6.0) 1.29
Apparent modulus	6.03(16.7) 3.67	0.96(2.7)[c] 0.76	0.88(2.4)[c] 0.39	0.83(2.3)[c] 0.32	N/A	8.20(22.7) 5.28	2.81(7.8)[a] 1.37	2.94(8.1)[a] 2.02	2.24(6.2)[a] 2.28	N/A
	XCT-Tibia (n=9)					**XCT2-Tibia (n=17)**				
Overlap	N/A	94%	85%	86%	85%	N/A	96%	89%	90%	89%
Failure load	1.41(3.9) 0.25	1.45(4.0) 0.21	1.45(4.0) 0.47	1.39(3.9) 0.13	1.22(3.4) 0.12	1.68(4.7) 0.48	1.44(4.0) 0.38	0.86(2.4) 0.24	0.84(2.3) 0.28	0.77(2.1) 0.29
Stiffness	1.91(5.3) 0.29	1.88(5.2) 0.21	1.90(5.3) 0.47	1.79(4.9) 0.18	1.81(5.0) 0.11	1.93(5.3) 0.42	1.70(4.7) 0.28	1.06(2.9) 0.30	1.02(2.8) 0.30	1.08(2.9) 0.30
Apparent modulus	2.36(6.5) 1.58	1.79(5.0)[b] 0.20	1.84(5.1)[c] 0.38	1.77(4.9)[c] 0.22	N/A	1.86(5.2) 0.94	1.18(3.3) 0.33	1.05(2.9) 0.37	0.88(2.5)[c] 0.24	N/A

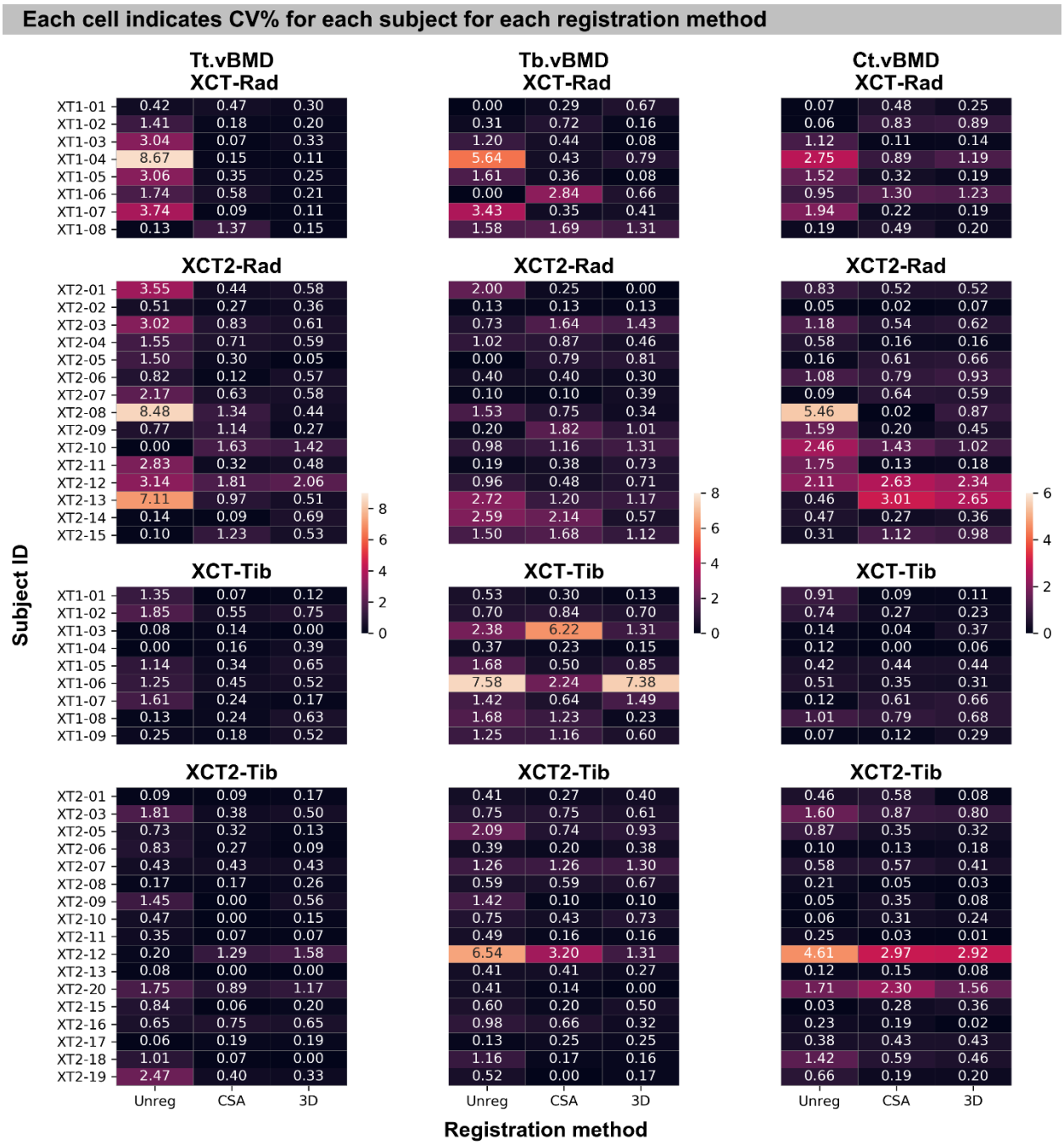
* Please note that no differences were observed between 3D and CSA methods.

Significance of Wilcoxon signed rank test after p-value correction are noted as follows:

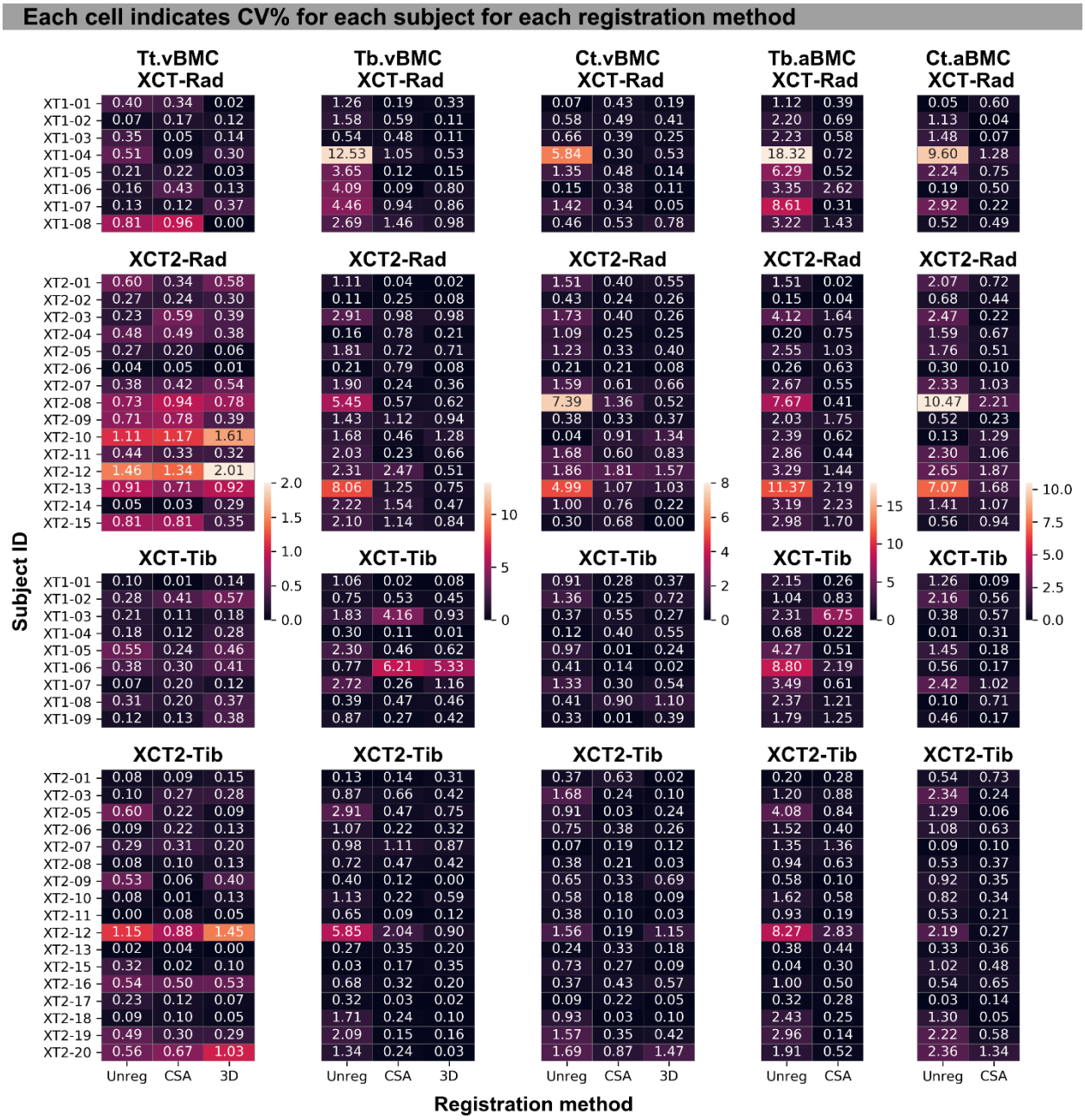
[a] Significantly different from unregistered, $p<0.01$

[b] Significantly different from unregistered, $p<0.05$

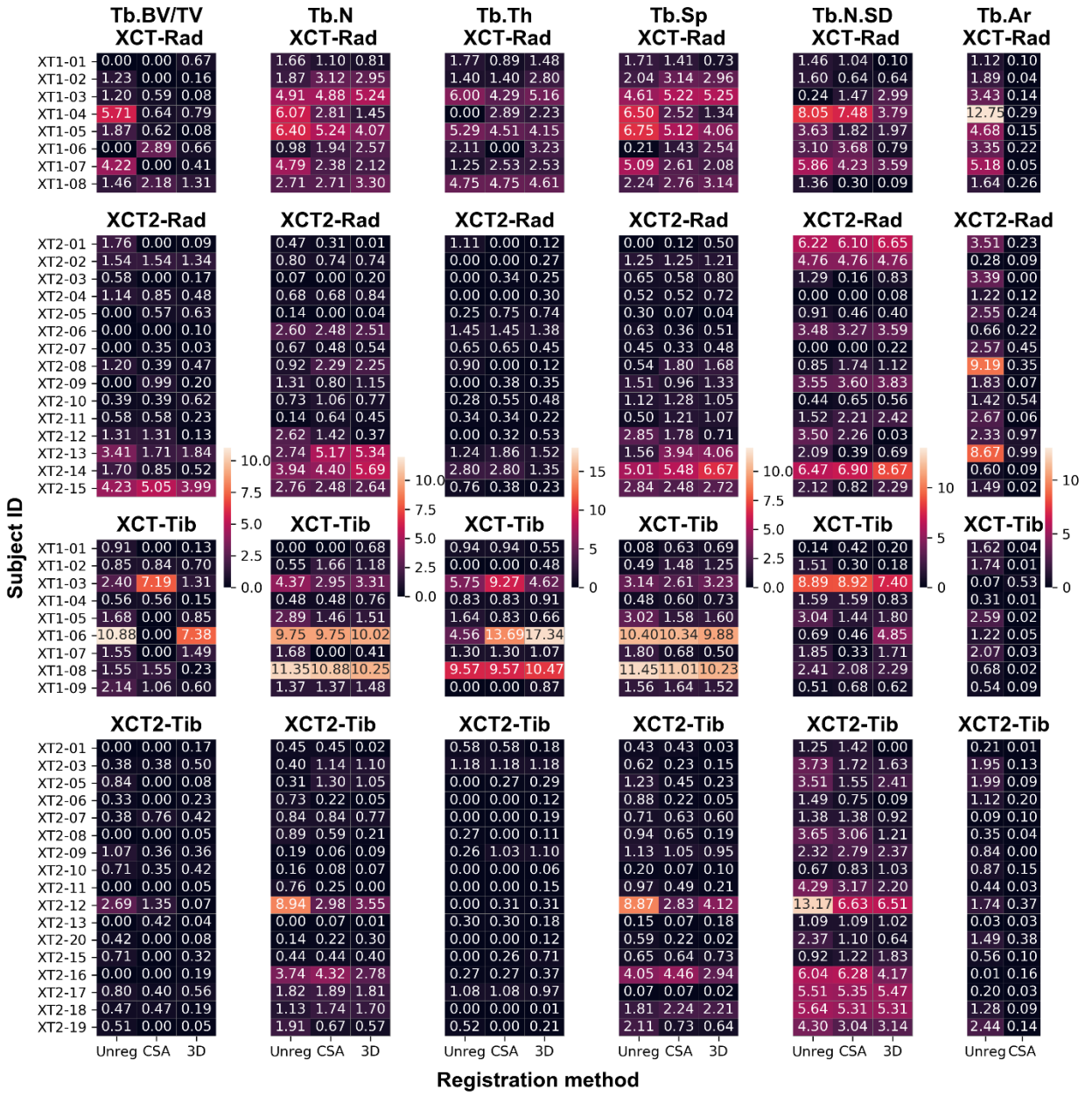
[c] Tendency of difference from unregistered, $p<0.1$



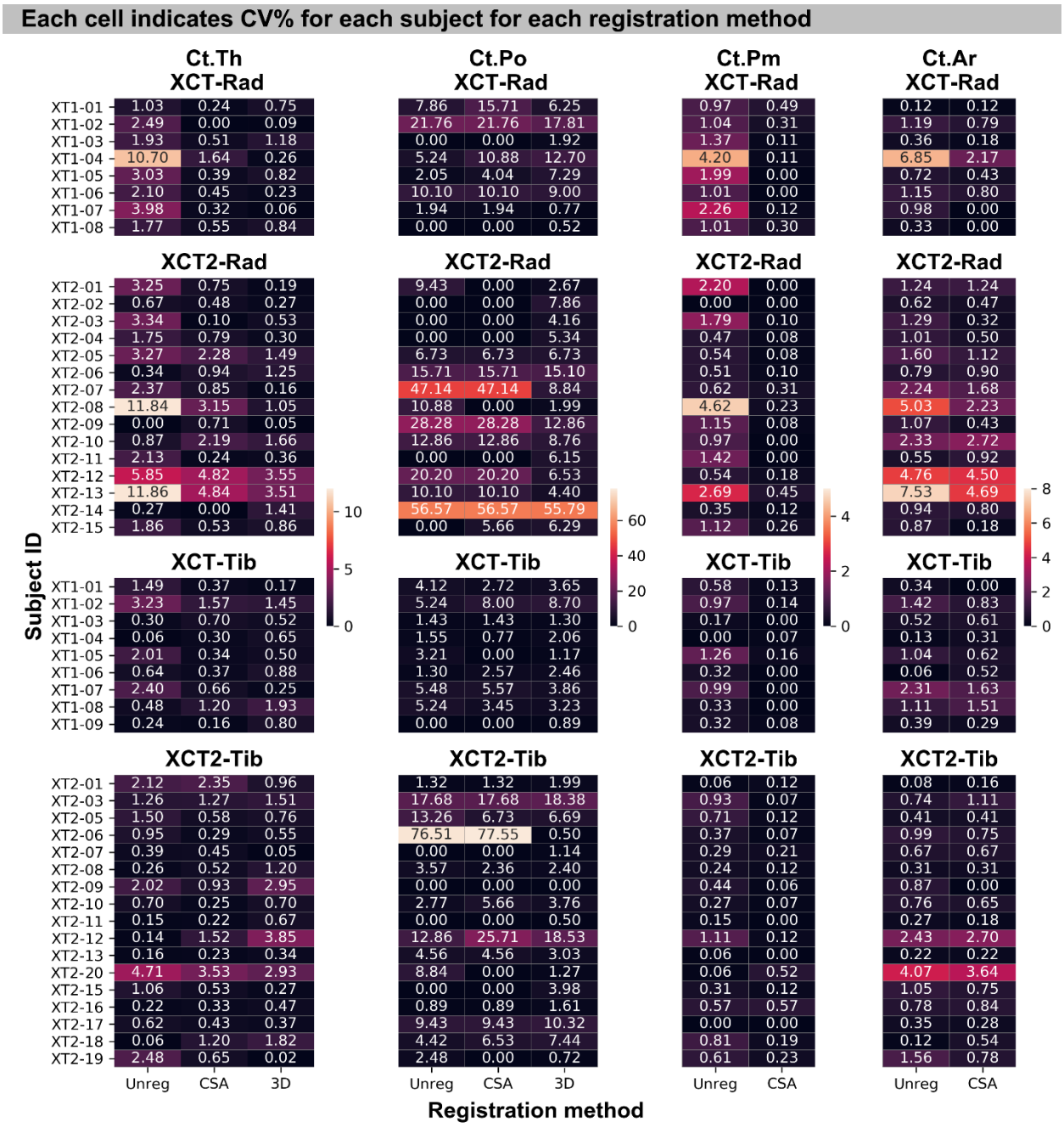
Supplemental Results Figure 1-A. Heatmap of CV% for volumetric BMD stratified by anatomical site and scanner generation. Each row contains CV%s for each individual, and each column shows each registration method. A common scale is used for different sites and scanners for each outcome. Numeric values of CV% are shown in each cell.



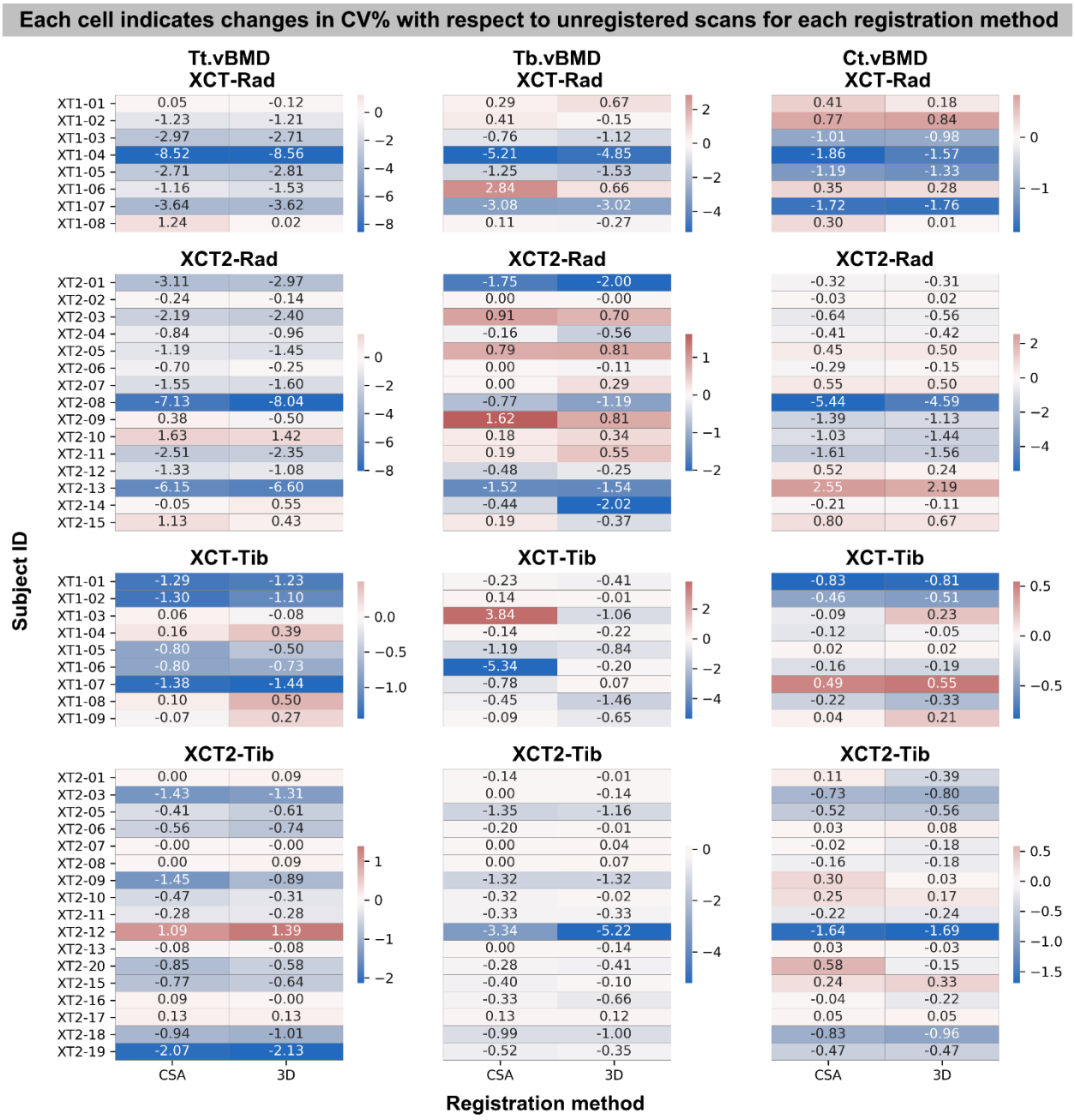
Supplemental Results Figure 1-B. Heatmap of CV% for volumetric and areal BMC stratified by anatomical site and scanner generation. Each row contains CV%s for each individual, and each column shows each registration method. A common scale is used for different sites and scanners for each outcome. Numeric values of CV% are shown in each cell.



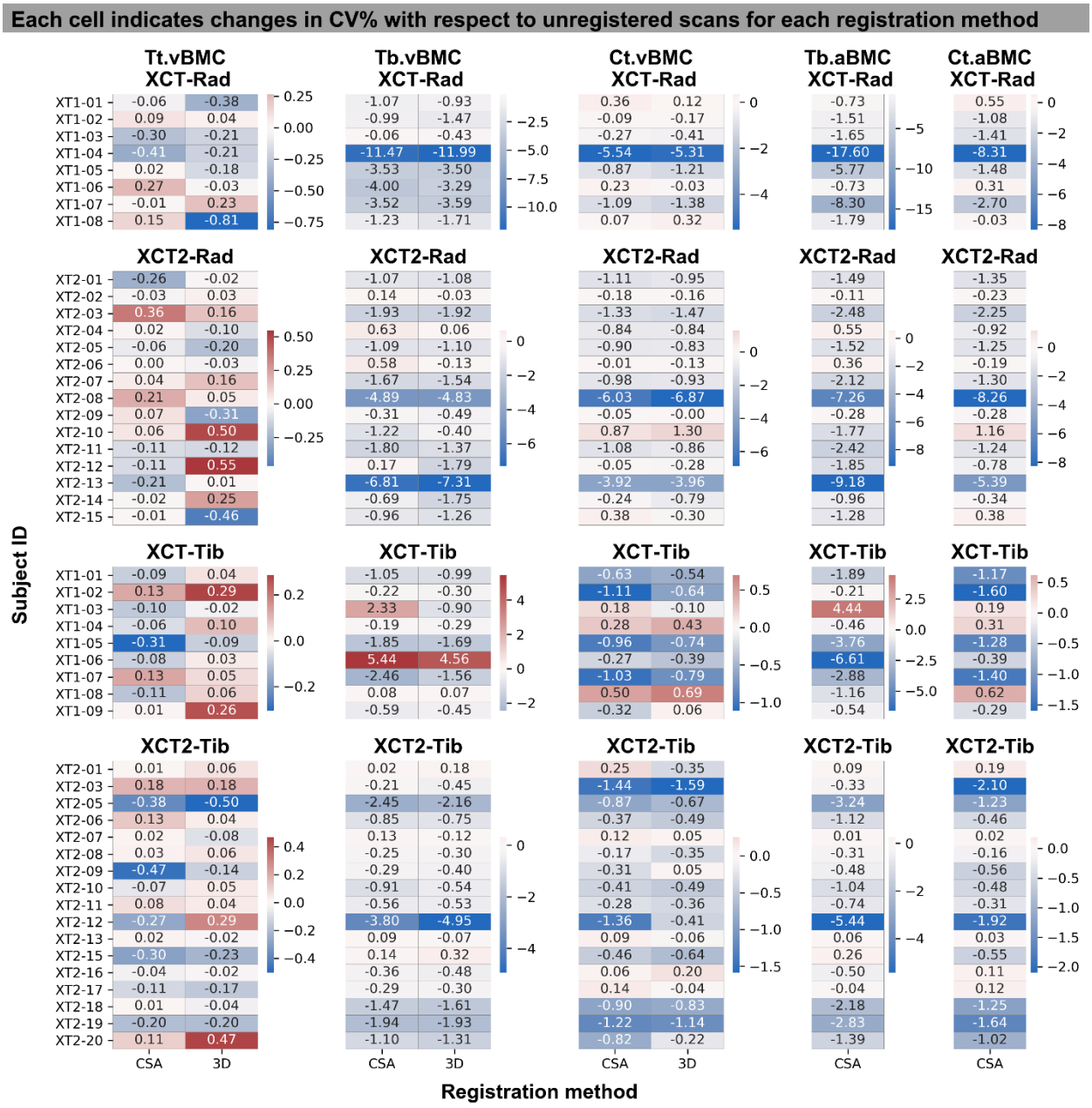
Supplemental Results Figure 1-C. Heatmap of CV% for trabecular microstructural outcomes stratified by anatomical site and scanner generation. Each row contains CV%s for each individual, and each column shows each registration method. A common scale is used for different sites and scanners for each outcome. Numeric values of CV% are shown in each cell.



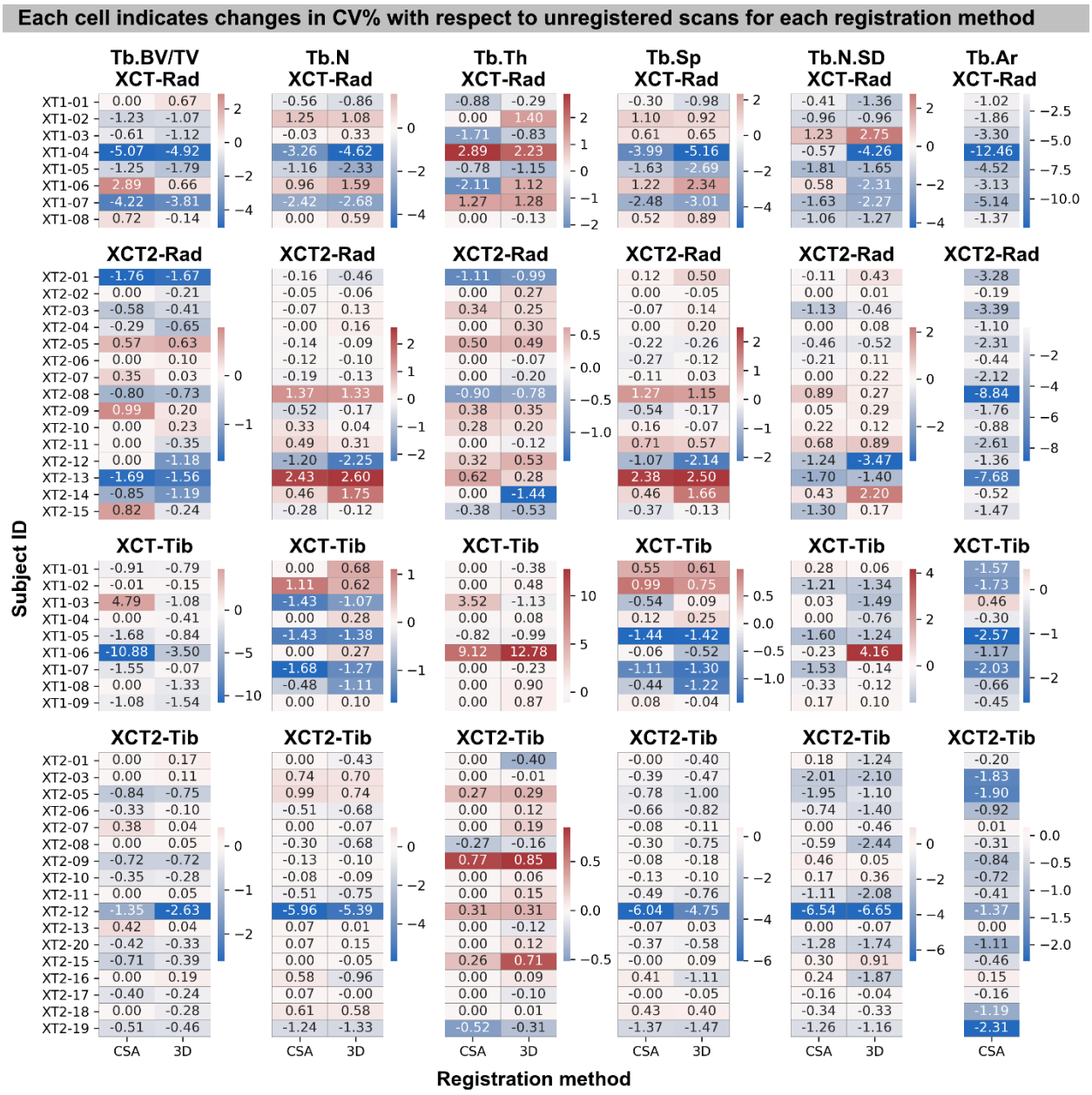
Supplemental Results Figure 1-D. Heatmap of CV% for cortical microstructural outcomes stratified by anatomical site and scanner generation. Each row contains CV%s for each individual, and each column shows each registration method. A common scale is used for different sites and scanners for each outcome. Numeric values of CV% are shown in each cell.



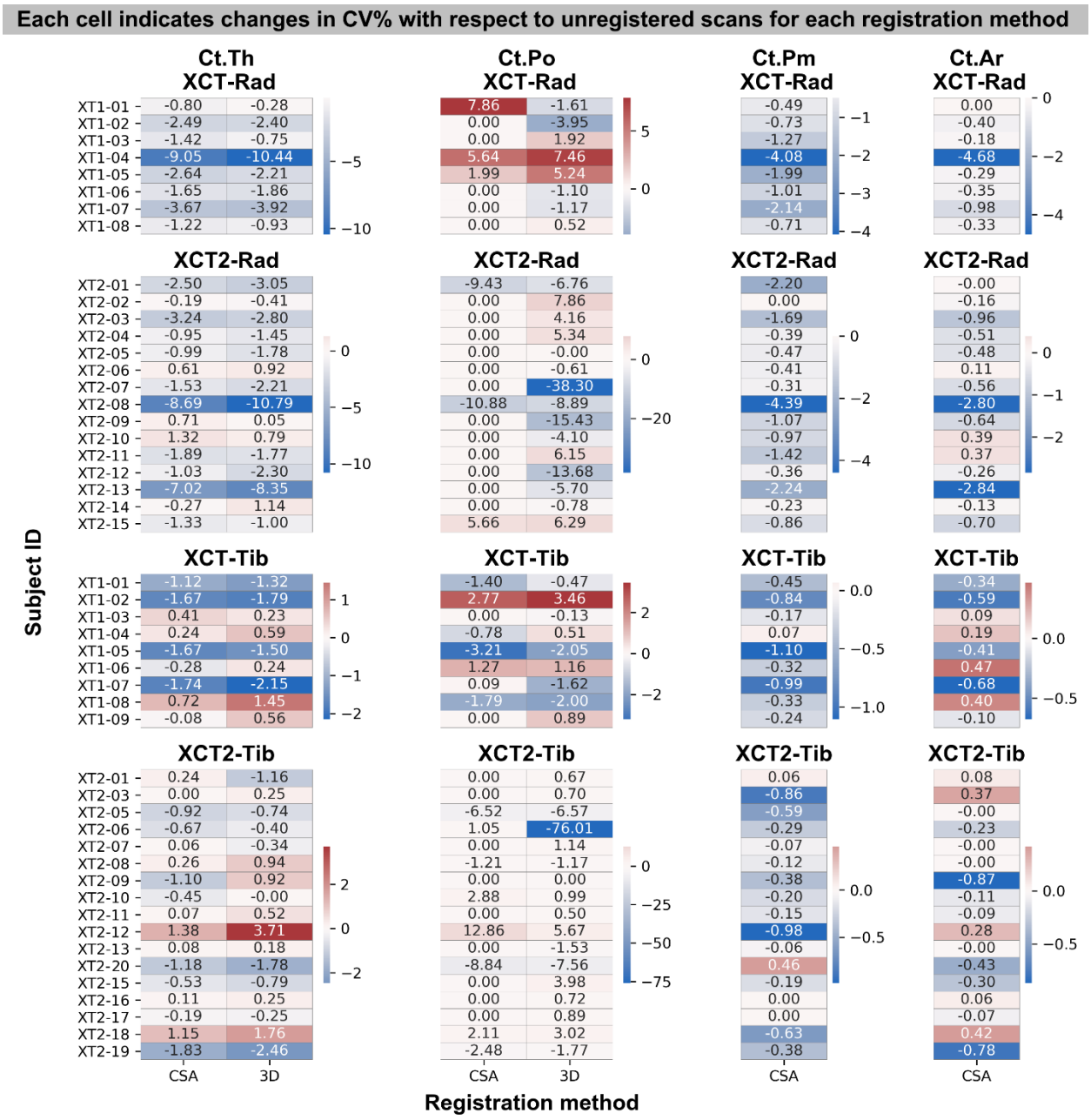
Supplemental Results Figure 2-A. Heatmap of changes in CV% from unregistered values for volumetric BMD stratified by anatomical site and scanner generation. Negative values (blue) indicate lower precision error after registration, while positive values (red) show higher errors. Each row contains change in CV%s for each individual, and each column shows each registration method. Numeric values of CV% are shown in each cell.



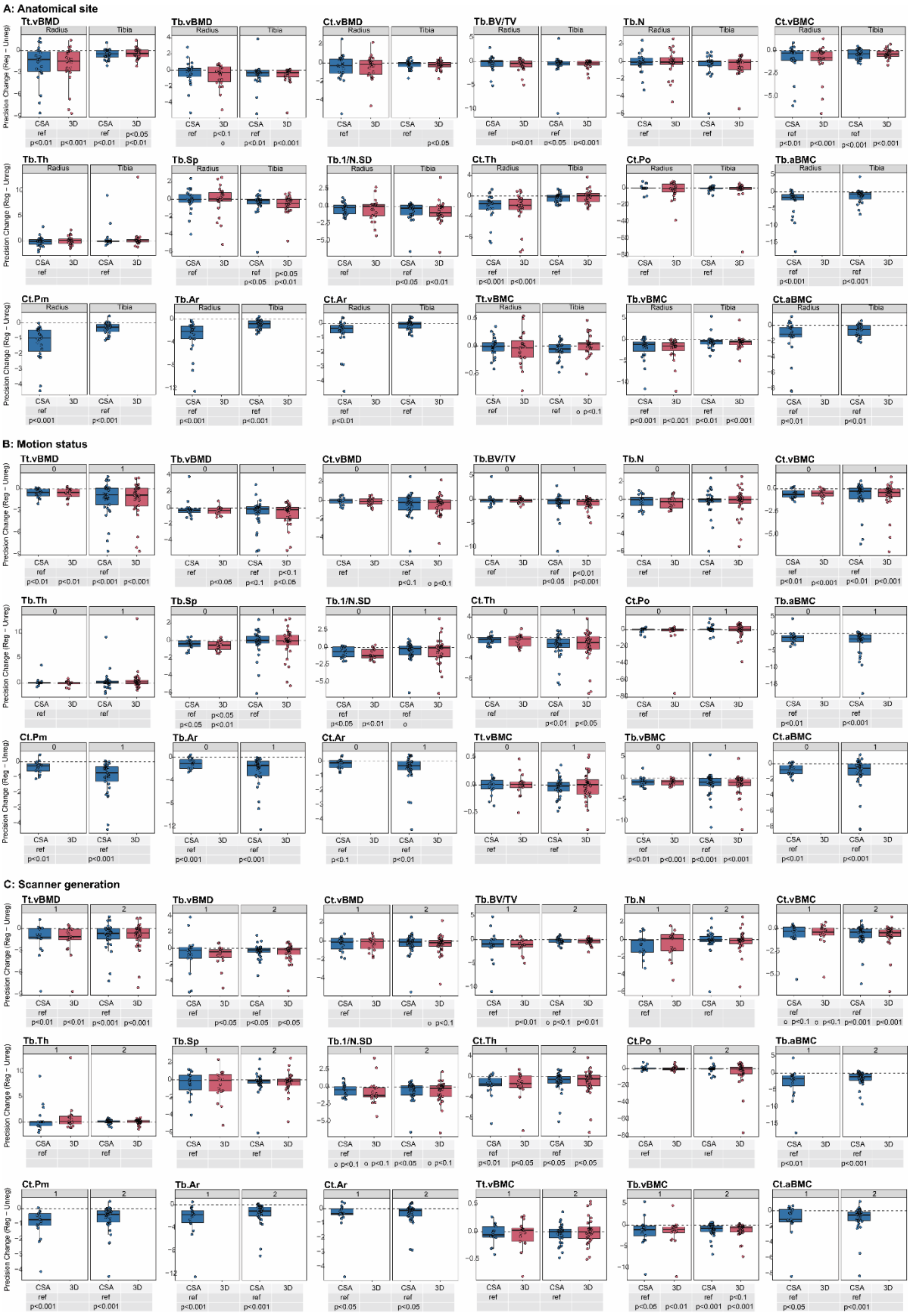
Supplemental Results Figure 2-B. Heatmap of changes in CV% from unregistered values for volumetric and areal BMC stratified by anatomical site and scanner generation. Negative values (blue) indicate lower precision error after registration, while positive values (red) show higher errors. Each row contains change in CV%s for each individual, and each column shows each registration method. Numeric values of CV% are shown in each cell.



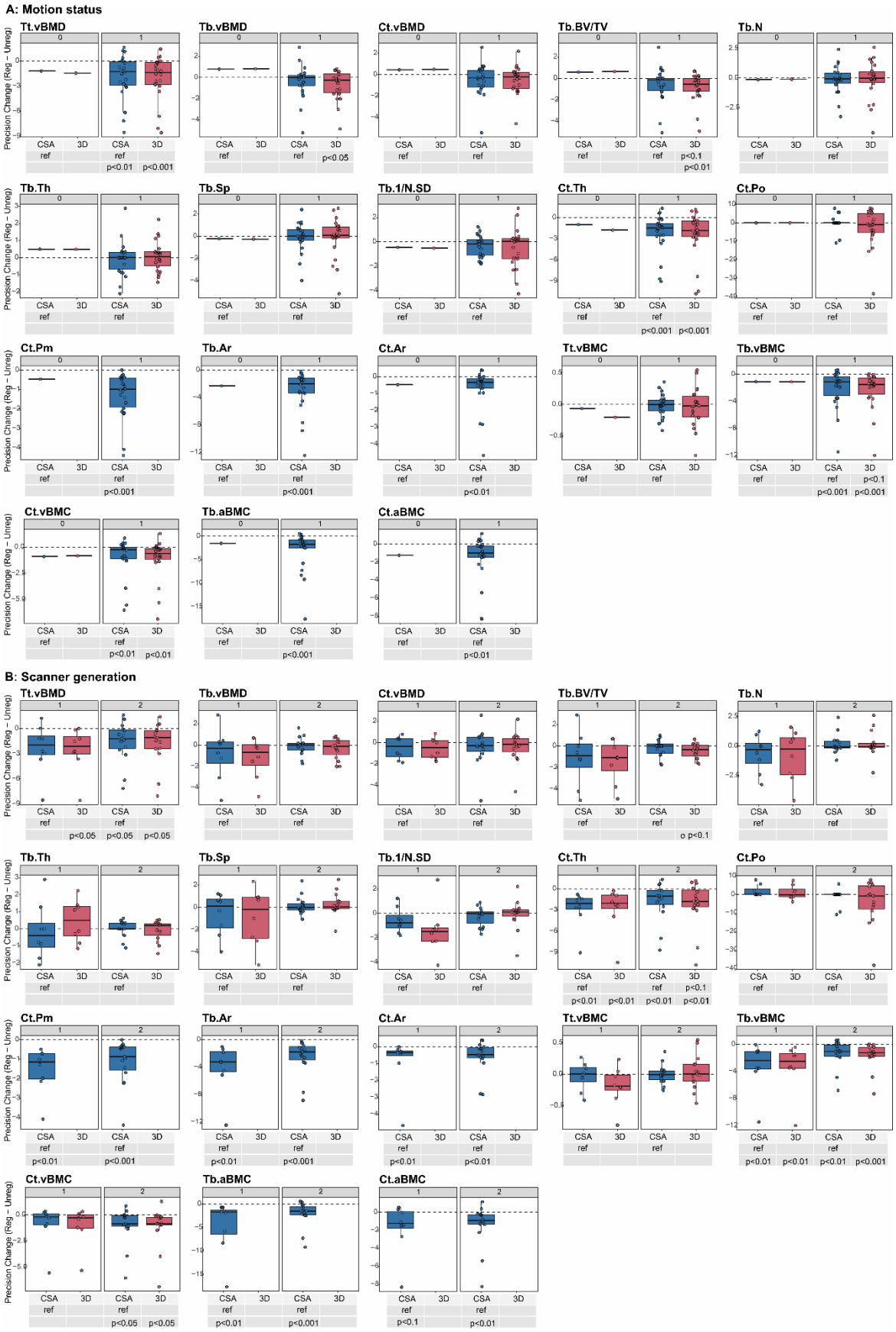
Supplemental Results Figure 2-C. Heatmap of changes in CV% from unregistered values for trabecular microstructural outcomes stratified by anatomical site and scanner generation. Negative values (blue) indicate lower precision error after registration, while positive values (red) show higher errors. Each row contains change in CV%s for each individual, and each column shows each registration method. Numeric values of CV% are shown in each cell.



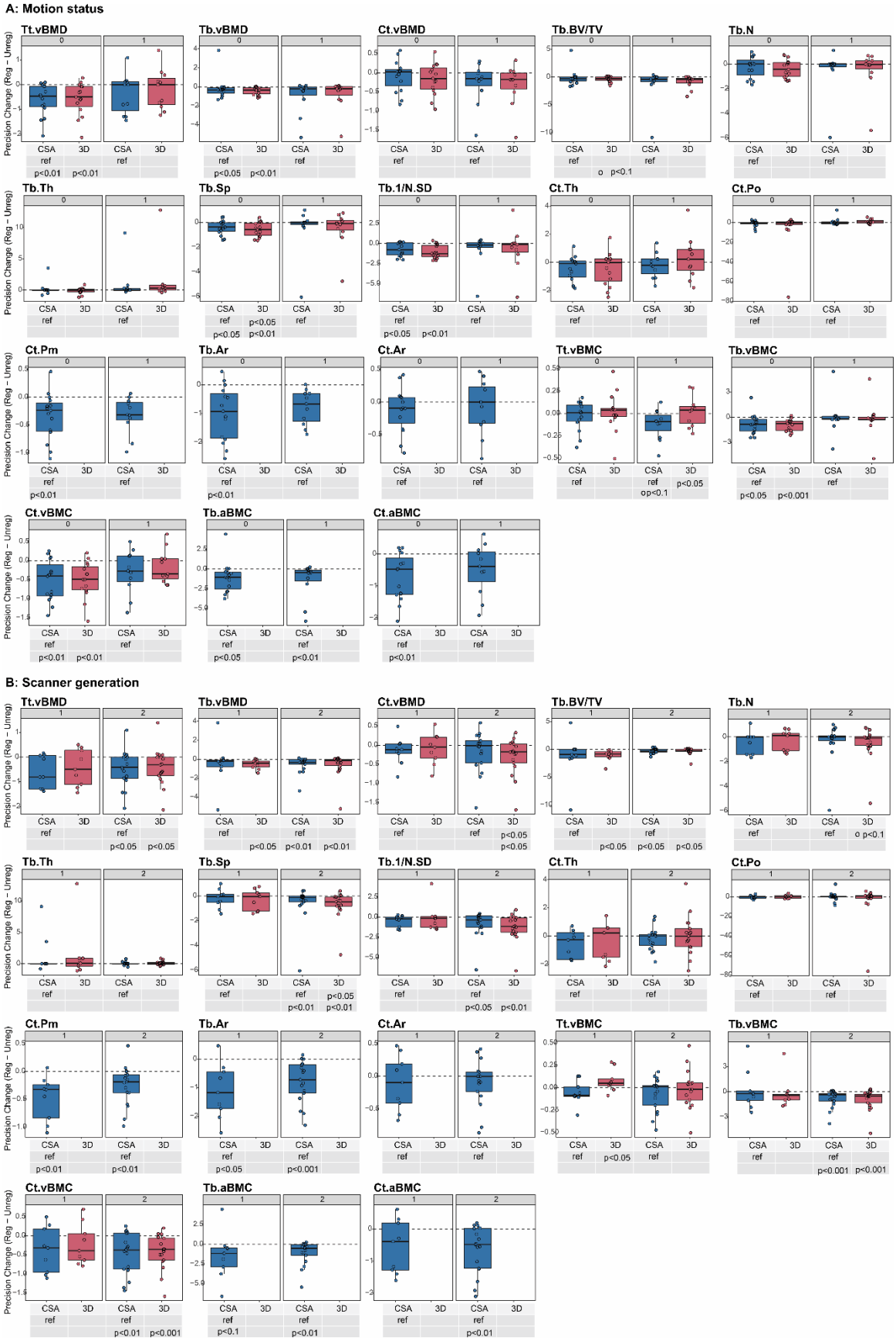
Supplemental Results Figure 2-D. Heatmap of changes in CV% from unregistered values for cortical microstructural outcomes stratified by anatomical site and scanner generation. Negative values (blue) indicate lower precision error after registration, while positive values (red) show higher errors. Each row contains change in CV%s for each individual, and each column shows each registration method. Numeric values of CV% are shown in each cell.


A: Anatomical site
B: Motion status
C: Scanner generation
Tt.vBMD
Tb.vBMD
Ct.vBMD
Tb.BV/TV
Tb.N
Ct.vBMC
Tb.Th
Tb.Sp
Tb.1/N.SD
Ct.Th
Ct.Po
Tb.aBMC
Ct.Pm
Tb.Ar
Ct.Ar
Tt.vBMC
Tb.vBMC
Ct.aBMC
Radius
Tibia
Precision Change (Reg - Unreg)
CSA
3D
ref

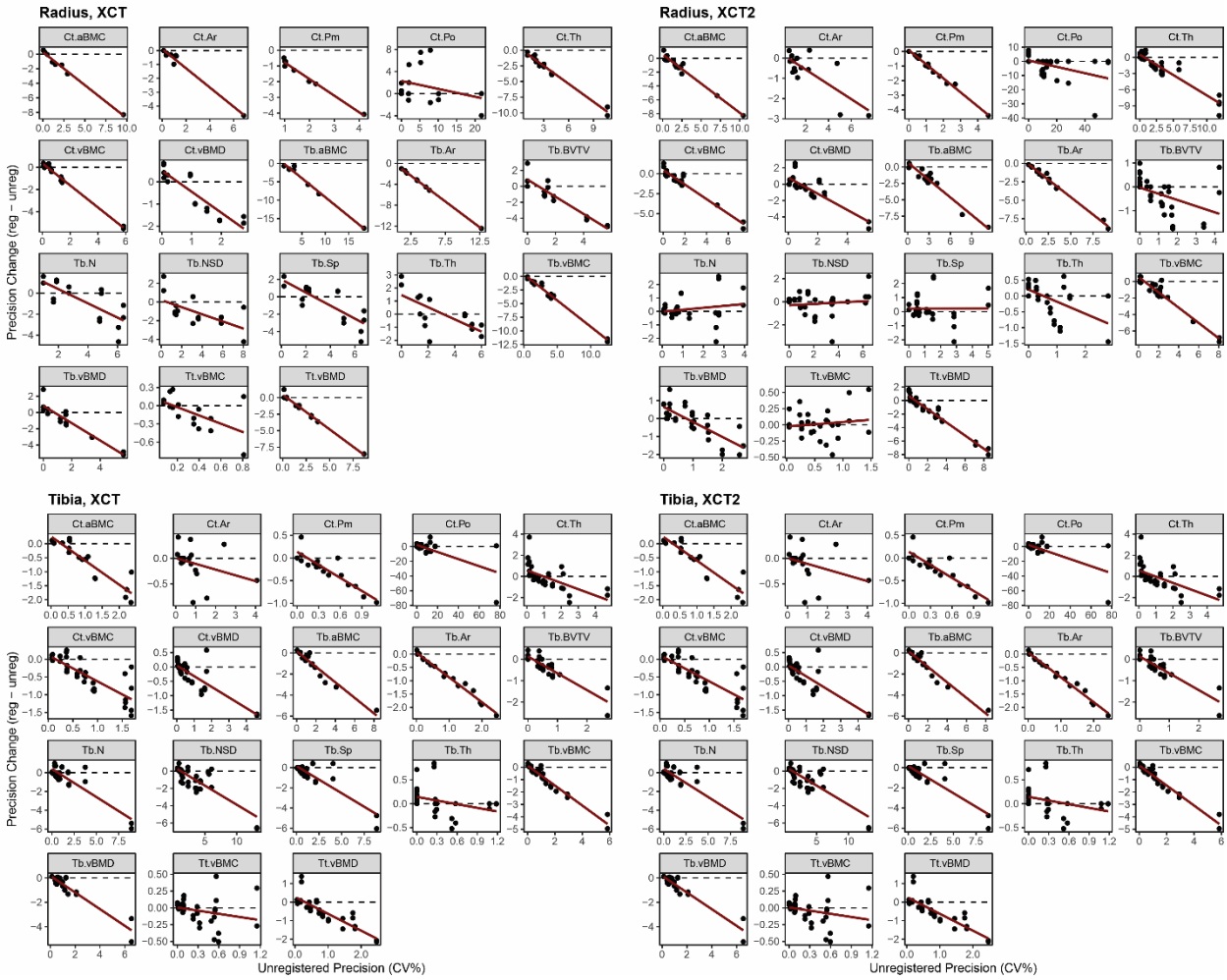
Supplemental Results Figure 3-A. Pair-wise comparisons of registration-associated changes in precision errors for selected density and microstructural outcomes. Comparisons were performed using Wilcoxon signed rank test for data stratified based on anatomical site, motion status, and scanner generation. Changes in precision error were also compared to zero. Since the comparisons are made for repeated measures, comparisons of changes in precision errors yield the same result as comparing the absolute precision errors. The p-values were corrected for multiple comparisons using Benjamini-Hochberg (BH) method. Significance level is $p<0.05$, while significance prior to p-value adjustment is shown as $^{o}p<0.05$.


A: Motion status
Tt.vBMD
Tb.vBMD
Ct.vBMD
Tb.BV/TV
Tb.N
Tb.Th
Tb.Sp
Tb.1/N.SD
Ct.Th
Ct.Po
Ct.Pm
Tb.Ar
Ct.Ar
Tt.vBMC
Tb.vBMC
Ct.vBMC
Tb.aBMC
Ct.aBMC
Precision Change (Reg − Unreg)
CSA
3D
ref
B: Scanner generation
Tt.vBMD
Tb.vBMD
Ct.vBMD
Tb.BV/TV
Tb.N
Tb.Th
Tb.Sp
Tb.1/N.SD
Ct.Th
Ct.Po
Ct.Pm
Tb.Ar
Ct.Ar
Tt.vBMC
Tb.vBMC
Ct.vBMC
Tb.aBMC
Ct.aBMC
Precision Change (Reg − Unreg)
CSA
3D
ref

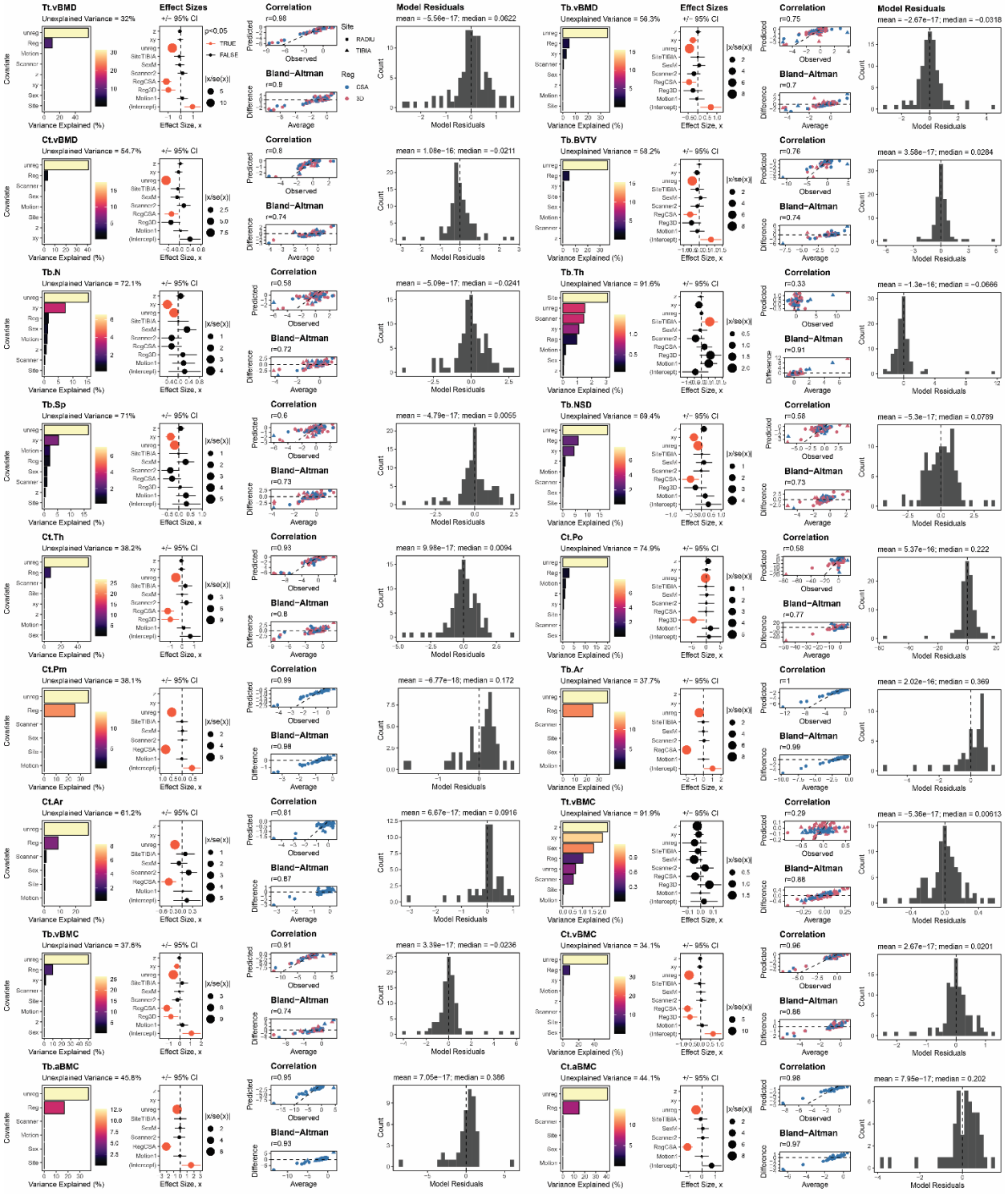
Supplemental Results Figure 3-B. Pair-wise comparisons of registration-associated changes in precision errors for selected density and microstructural outcomes at the radius. Comparisons were performed using Wilcoxon signed rank test for data stratified based on motion status and scanner generation. Changes in precision error were also compared to zero. Since the comparisons are made for repeated measures, comparisons of changes in precision errors yield the same result as comparing the absolute precision errors. The p-values were corrected for multiple comparisons using Benjamini-Hochberg (BH) method. Significance level is $p<0.05$, while significance prior to p-value adjustment is shown as $^{o}p<0.05$.


A: Motion status
Tt.vBMD
Tb.vBMD
Ct.vBMD
Tb.BV/TV
Tb.N
Tb.Th
Tb.Sp
Tb.1/N.SD
Ct.Th
Ct.Po
Ct.Pm
Tb.Ar
Ct.Ar
Tt.vBMC
Tb.vBMC
Ct.vBMC
Tb.aBMC
Ct.aBMC
Precision Change (Reg − Unreg)
CSA
3D
ref
B: Scanner generation

Supplemental Results Figure 3-C. Pair-wise comparisons of registration-associated changes in precision errors for selected density and microstructural outcomes at the tibia. Comparisons were performed using Wilcoxon signed rank test for data stratified based on motion status and scanner generation. Changes in precision error were also compared to zero. Since the comparisons are made for repeated measures, comparisons of changes in precision errors yield the same result as comparing the absolute precision errors. The p-values were corrected for multiple comparisons using Benjamini-Hochberg (BH) method. Significance level is $p<0.05$, while significance prior to p-value adjustment is shown as $^{o}p<0.05$.

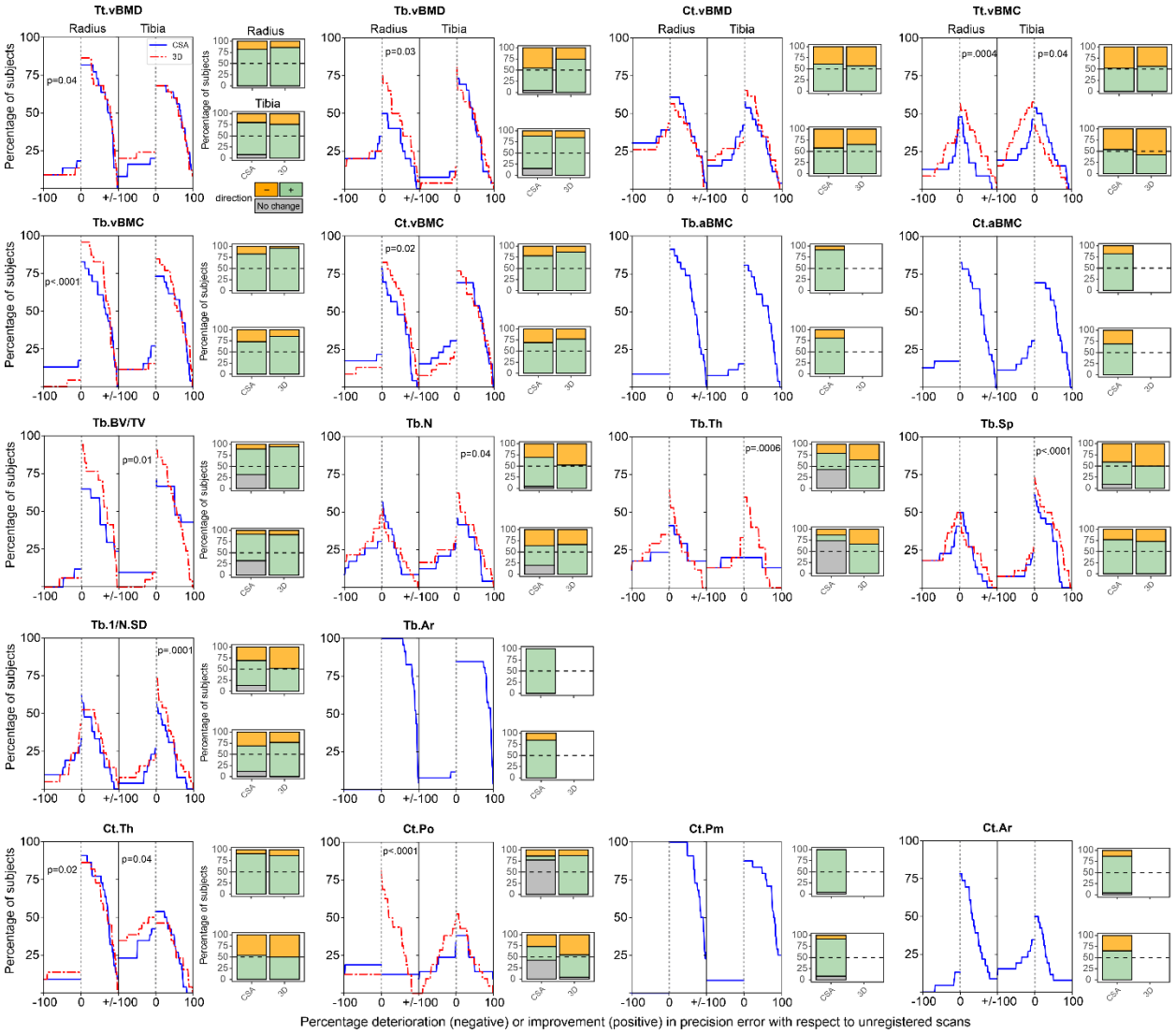


Supplemental Results Figure 4-A. Crude association between the precision error of unregistered scans and change in precision error after image registration for density and microstructural outcomes. Larger negative values on Y-axis indicate more improvement in precision error after image registration.

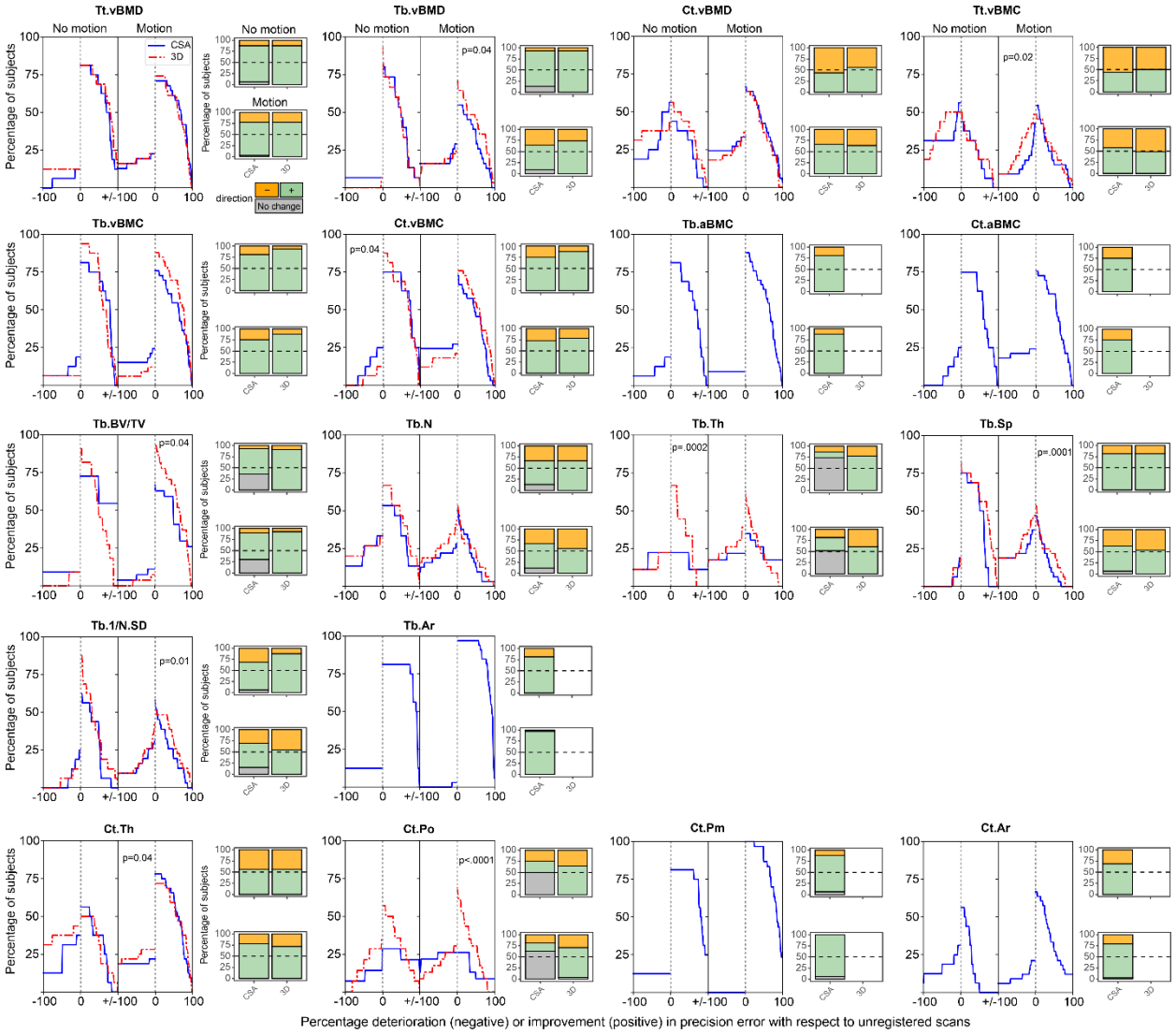


Supplemental Results Figure 4-B. Summary of linear multiple regression model for density and microstructural outcomes. Dependent variable is the change in precision error from unregistered scans for each outcome, and independent variables are anatomical site, motion status, scanner generation, unregistered precision error (unreg), registration method, rotation angles, and

sex. The effect size of factors with a significant effect of changes in precision error is shown in red. Model performance is shown in terms of the correlation and agreement between observed and predicted values of the dependent variable, and the distribution of model residuals. The colorbars indicate % of explained variance.

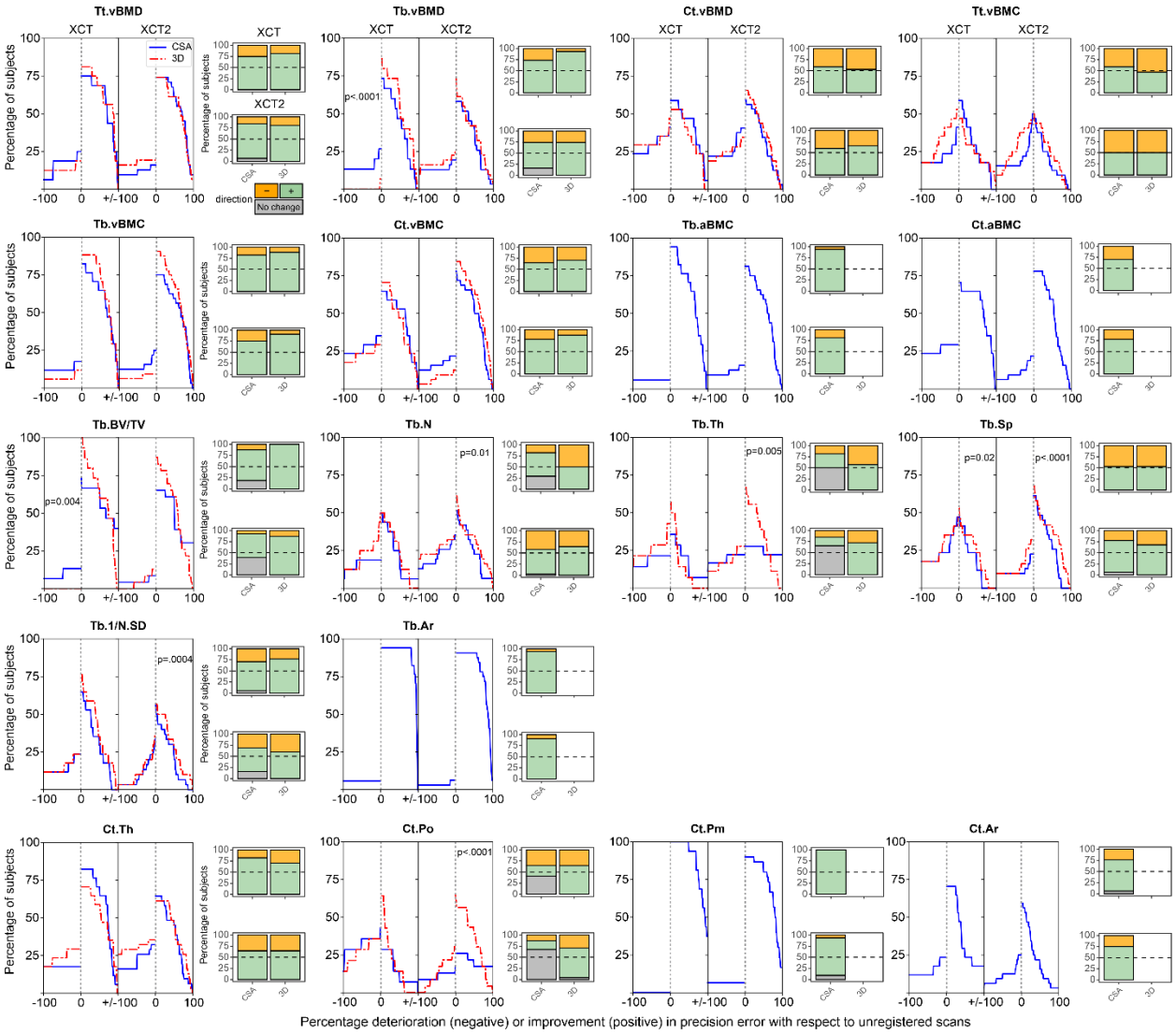


Supplemental Results Figure 5-A. Individual-level registration performance plots (RPPs) for CSA and 3D registration methods for the density and microstructural outcomes stratified by anatomical site. X-axis shows the percentage change in precision error of CSA and 3D registration relative to unregistered precision error. Positive values of x mean improvement in precision error when using image registration, whereas negative values indicate deterioration. Y-axis indicated the percentage of individuals for whom a percentage of improvement or deterioration is obtained using image registration. For positive values on x-axis, higher percentage of individuals means better performance, while for negative values on x-axis, lower percentage of individuals is better. The difference between 100% and the sum of negative and positive curves at x=0 is the percentage of individuals with improvements in precision less than 1%. P-values indicate significant difference in area under the curve (AUC) separated by improvement or deterioration. **Bar plots represents the proportion of individual scans that experience improvement (green), no change (grey) or deterioration (orange) following registration (corresponds to y-axis intercept on RPP plot).**

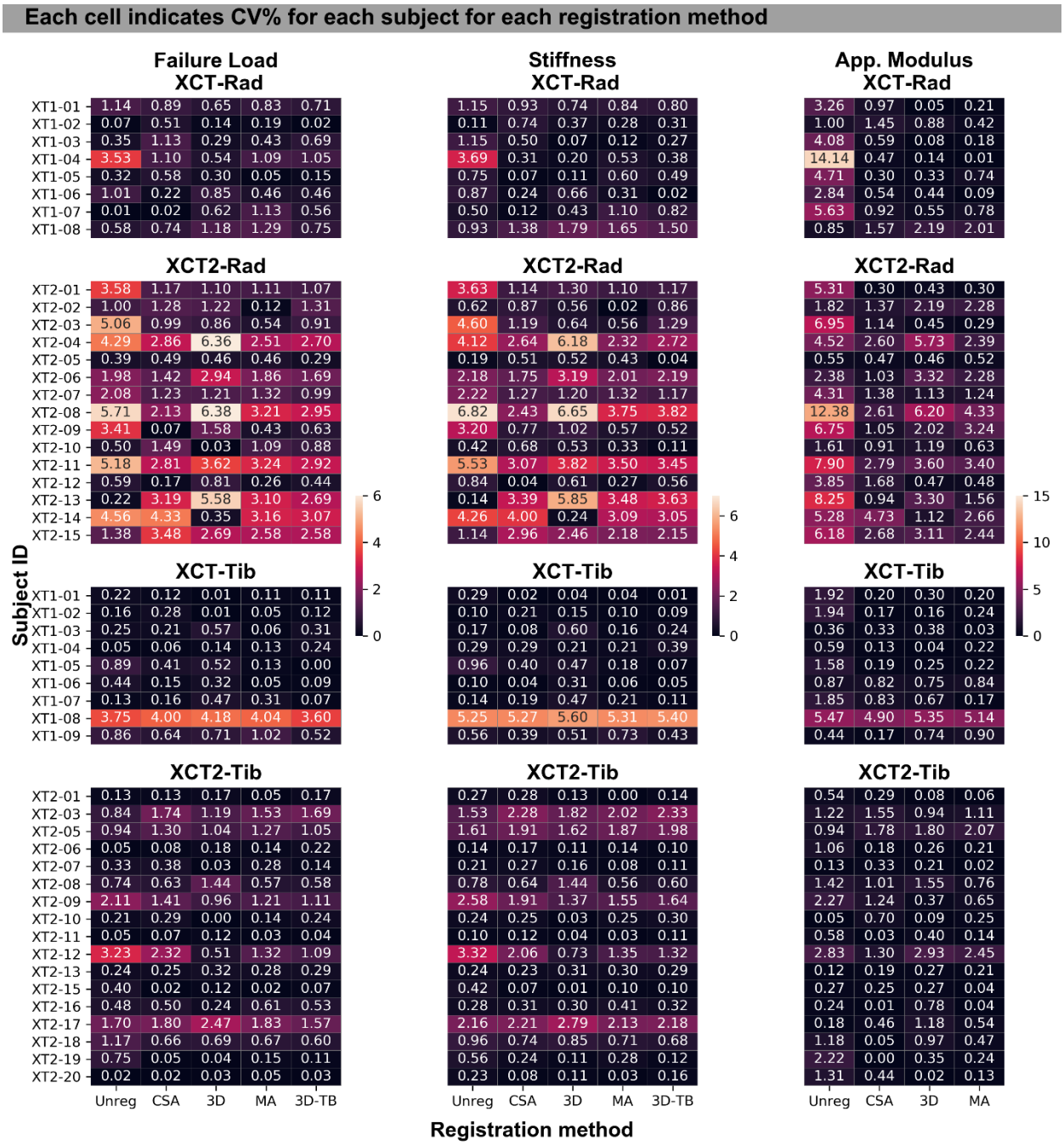


Supplemental Results Figure 5-B. Individual-level registration performance plots (RPPs) for CSA and 3D registration methods for the density and microstructural outcomes stratified by motion status. X-axis shows the percentage change in precision error of CSA and 3D registration relative to unregistered precision error. Positive values of x mean improvement in precision error when using image registration, whereas negative values indicate deterioration. Y-axis indicated the percentage of individuals for whom a percentage of improvement or deterioration is obtained using image registration. For positive values on x-axis, higher percentage of individuals means better performance, while for negative values on x-axis, lower percentage of individuals is better. The difference between 100% and the sum of negative and positive curves at x=0 is the percentage of individuals with improvements in precision less than 1%. P-values indicate significant difference in area under the curve (AUC) separated by improvement or deterioration. **Bar plots represents the proportion of individual scans that experience improvement (green), no**

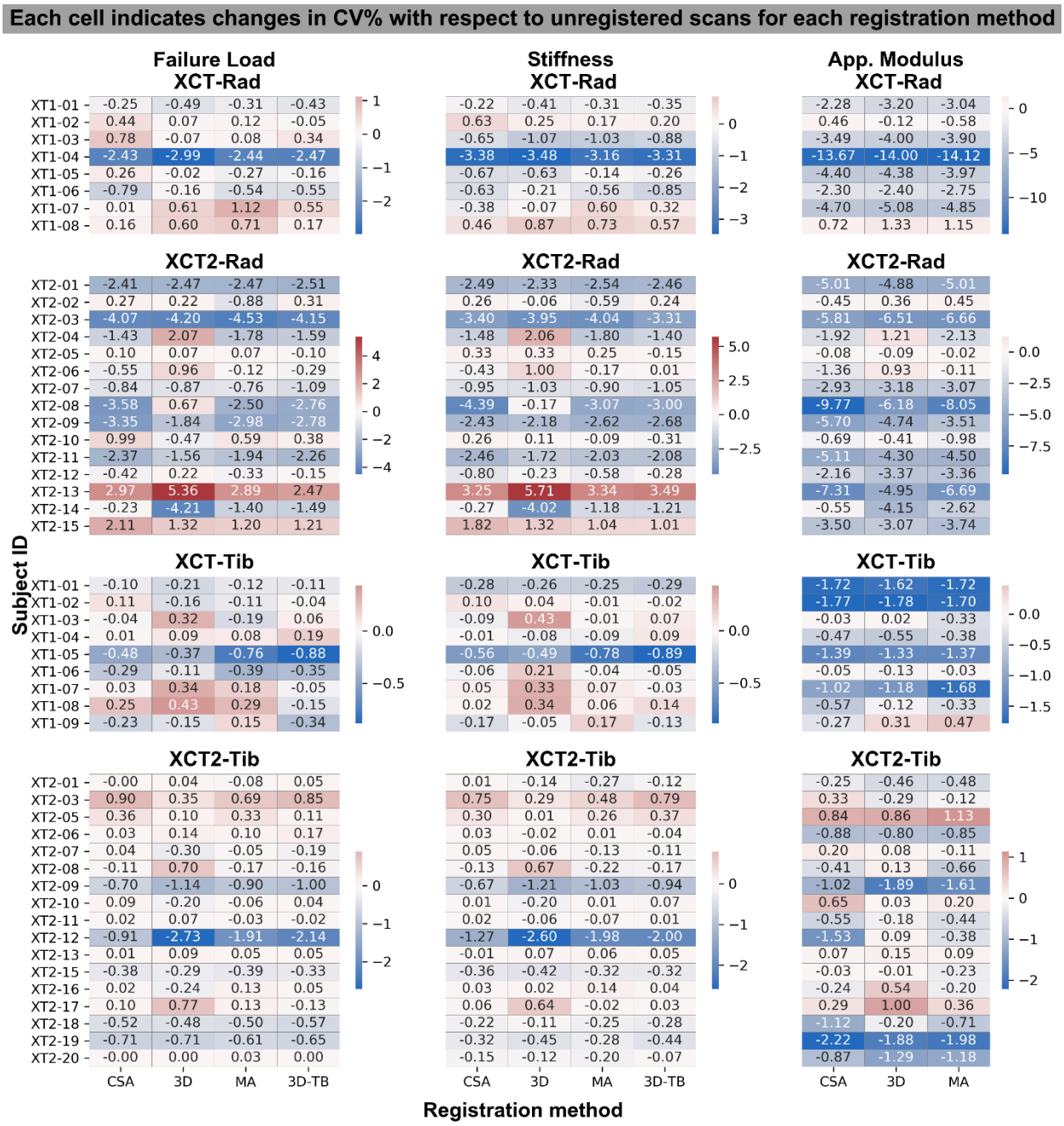
change (grey) or deterioration (orange) following registration (corresponds to y-axis intercept on RPP plot).



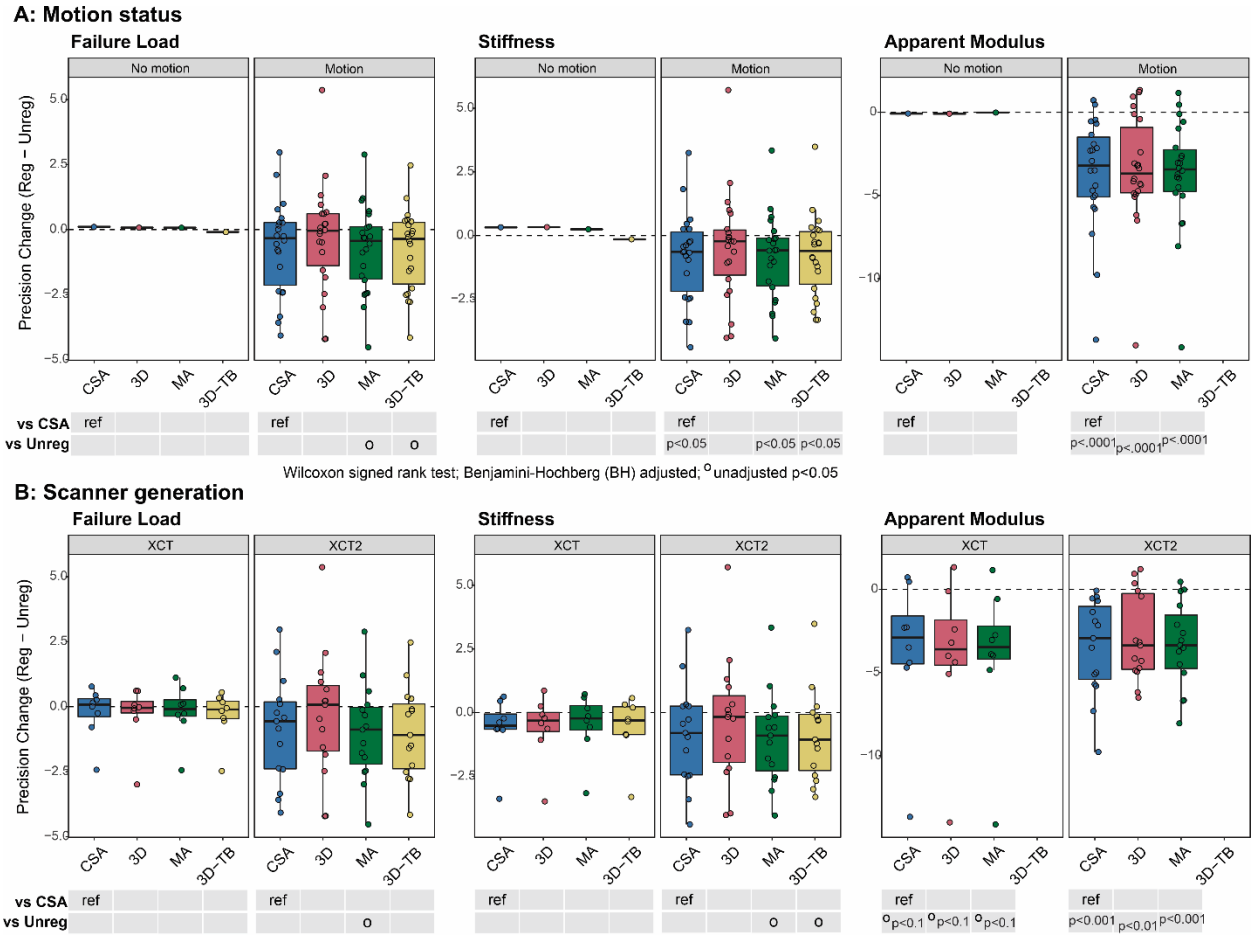
Supplemental Results Figure 5-C. Individual-level registration performance plots (RPPs) for CSA and 3D registration methods for the density and microstructural outcomes stratified by scanner generation. X-axis shows the percentage change in precision error of CSA and 3D registration relative to unregistered precision error. Positive values of x mean improvement in precision error when using image registration, whereas negative values indicate deterioration. Y-axis indicated the percentage of individuals for whom a percentage of improvement or deterioration is obtained using image registration. For positive values on x-axis, higher percentage of individuals means better performance, while for negative values on x-axis, lower percentage of individuals is better. The difference between 100% and the sum of negative and positive curves at x=0 is the percentage of individuals with improvements in precision less than 1%. P-values indicate significant difference in area under the curve (AUC) separated by improvement or deterioration. **Bar plots represents the proportion of individual scans that experience improvement (green), no change (grey) or deterioration (orange) following registration (corresponds to y-axis intercept on RPP plot).**



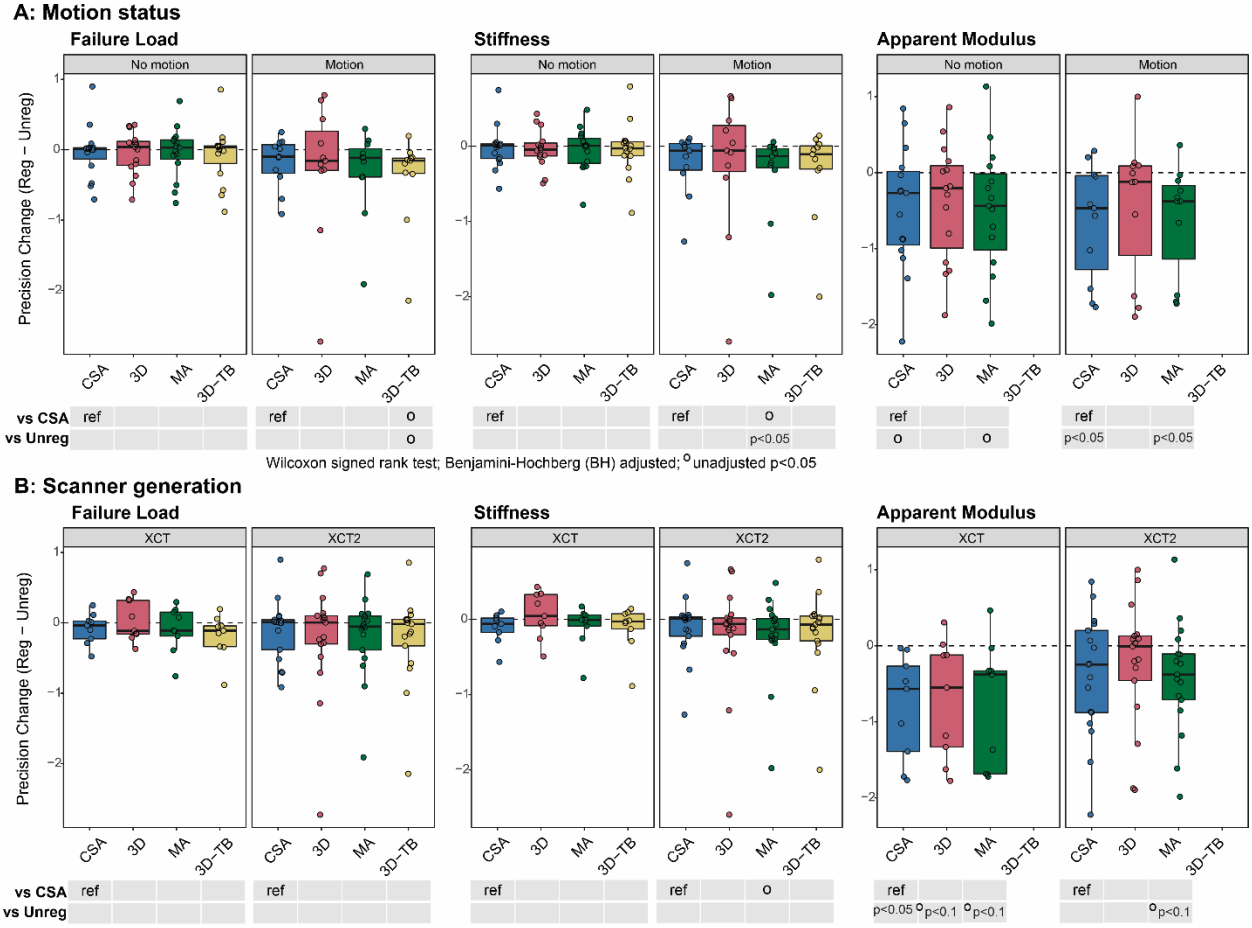
Supplemental Results Figure 6. Heatmap of CV% for microFE outcomes stratified by anatomical site and scanner generation. Each row contains CV%s for each individual, and each column shows each registration method. A common scale is used for different sites and scanners for each outcome. Numeric values of CV% are shown in each cell.



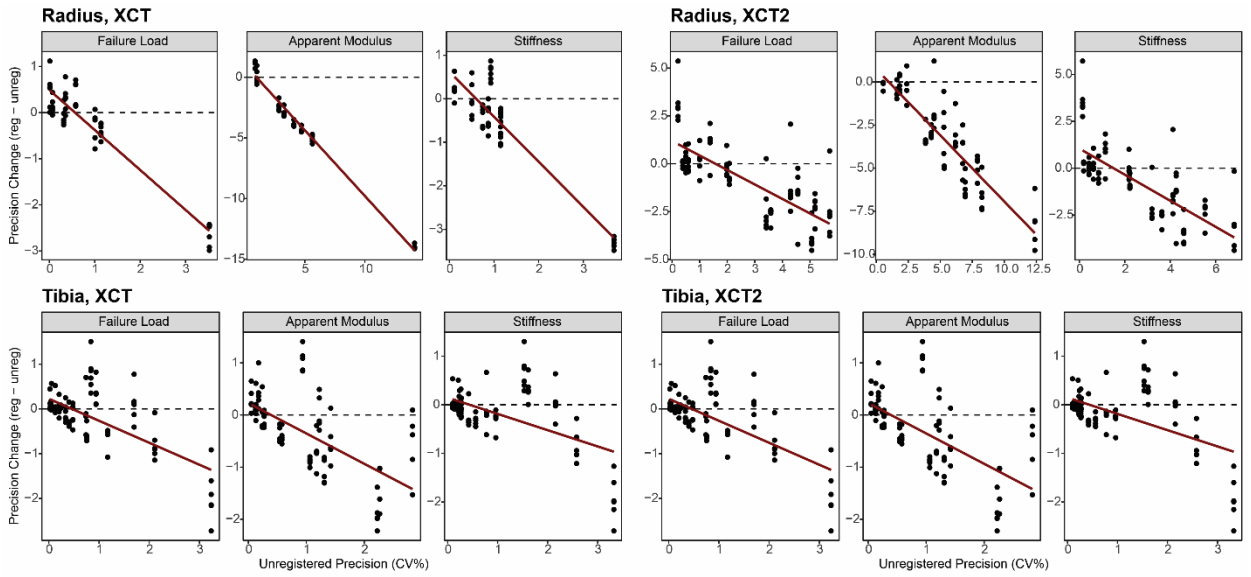
Supplemental Results Figure 7. Heatmap of changes in CV% from unregistered values for microFE outcomes stratified by anatomical site and scanner generation. Negative values (blue) indicate lower precision error after registration, while positive values (red) show higher errors. Each row contains change in CV%s for each individual, and each column shows each registration method. Numeric values of CV% are shown in each cell.



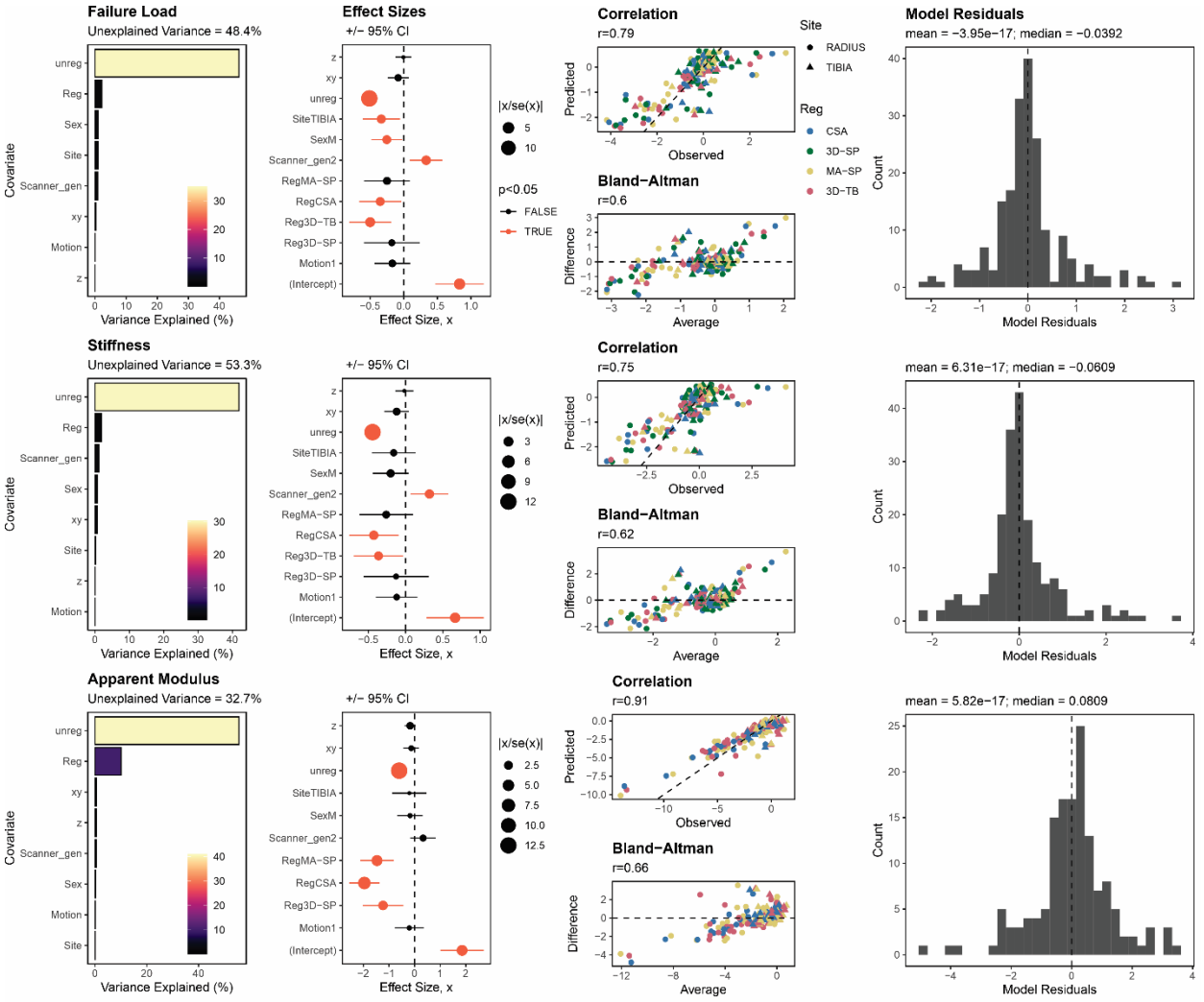
Supplemental Results Figure 8-A. Pair-wise comparisons of registration-associated changes in precision errors for microFE outcomes at the radius. Comparisons were performed using Wilcoxon signed rank test for data stratified based on motion status, and scanner generation. Changes in precision error were also compared to zero. Since the comparisons are made for repeated measures, comparisons of changes in precision errors yield the same result as comparing the absolute precision errors. The p-values were corrected for multiple comparisons using Benjamini-Hochberg (BH) method Significance level is $p<0.05$, while significance prior to p-value adjustment is shown as $^{o}p<0.05$.



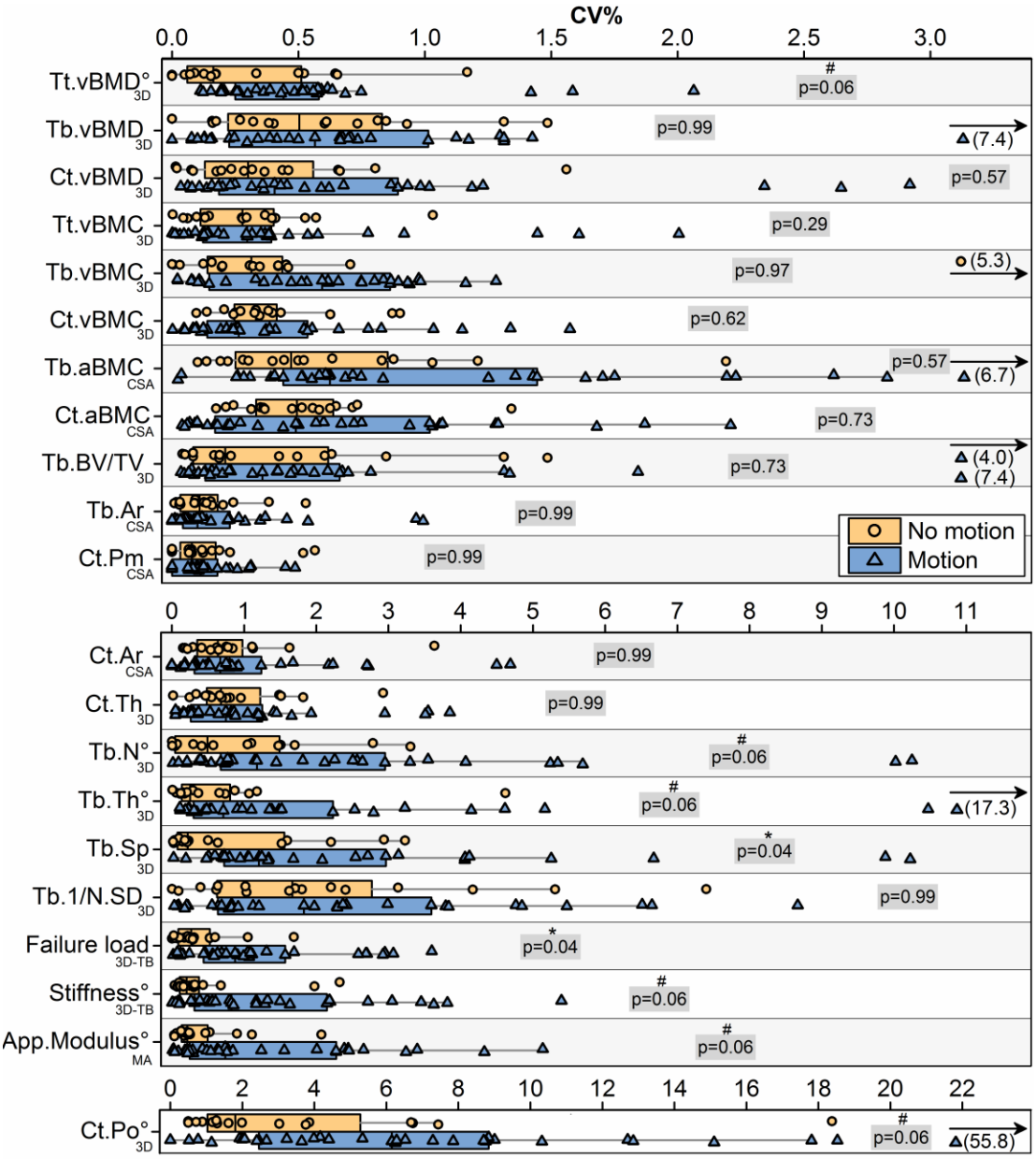
Supplemental Results Figure 8-B. Pair-wise comparisons of registration-associated changes in precision errors for microFE outcomes at the tibia. Comparisons were performed using Wilcoxon signed rank test for data stratified based on motion status, and scanner generation. Changes in precision error were also compared to zero. Since the comparisons are made for repeated measures, comparisons of changes in precision errors yield the same result as comparing the absolute precision errors. The p-values were corrected for multiple comparisons using Benjamini-Hochberg (BH) method. Significance level is $p<0.05$, while significance prior to p-value adjustment is shown as $^{o}p<0.05$.



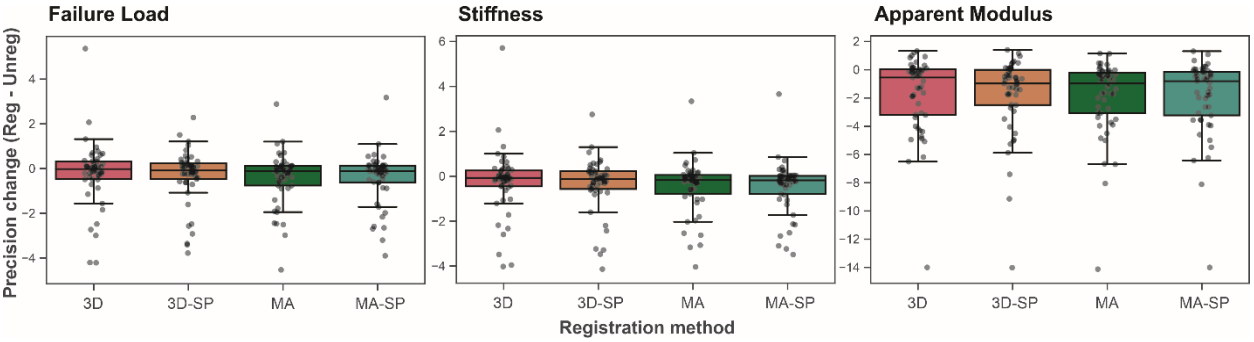
Supplemental Results Figure 9-A. Crude association between the precision error of unregistered scans and change in precision error after image registration for microFE outcomes. Larger negative values on Y-axis indicate more improvement in precision error after image registration.



Supplemental Results Figure 9-B. Summary of linear multiple regression model for microFE outcomes. Dependent variable is the change in precision error from unregistered scans for each outcome, and independent variables are anatomical site, motion status, scanner generation, unregistered precision error (unreg), registration method, rotation angles, and sex. The effect size of factors with a significant effect of changes in precision error is shown in red. Model performance is shown in terms of the correlation and agreement between observed and predicted values of the dependent variable, and the distribution of model residuals. The colorbars indicate % of explained variance.



Supplemental Results Figure 10. Absolute CV% values for HR-pQCT outcomes stratified by motion status. Parameters with comparable scale are grouped together. Arrows to the right-most of the graphs indicate the presence of data points out of the plot range with values shown in brackets. For each outcome, data from the registration method with the lowest amount of repositioning error were included (i.e., 3D-registered for density and microstructural outcomes except that CSA method was used for area outcome measures. For microFE, 3D-TB method was used except for apparent modulus). #p<0.1, *p<0.05 from comparison between scans with and without motion by Mann Whitney test. Outcomes shown by a "°" were significantly different only prior to p-value adjustment.



Supplemental Results Figure 11. Pair-wise comparisons of registration-associated changes in precision errors for microFE outcomes. Comparative analyses for i) 3D vs. 3D-SP, ii) MA vs. MA-SP were performed using Wilcoxon signed rank test for the pooled data. P-values were corrected for multiple comparisons using Benjamini-Hochberg (BH) method. Significance level is $p<0.05$. No difference was observed between the standard and SP methods.

Connection between chapters:

In Chapter 4, I investigated the effect of different image registration methods on the precision of HR-pQCT measurements, using same-day repeated scans acquired from adults with OI as part of a phase 2b multicenter clinical trial. Repositioning error in longitudinal HR-pQCT imaging can lead to different bone volumes assessed over time, which reduces precision. Increasing the precision of HR-pQCT measurements increases the reliability of this modality to identify changes in bone over time due to growth, aging, diseases, treatments, and so forth. The measurements discussed in the previous chapter were representative of average bone properties including bone density, morphology, and strength. Changes in these parameters indicate the net changes in bone, and proper image registration ensures that the measurements are monitored over the same bone region over time. Our goal was to identify the best image registration method for improved reliability of HR-pQCT measurements, and to contribute to the standardization of the methodology.

Image registration is also the basis for time-lapse HR-pQCT analysis that can be used to evaluate bone formation and resorption. These measurements can additionally provide information on local changes on bone to give better insight on the underlying mechanisms responsible for net bone changes. Nevertheless, a challenge regarding time-lapse HR-pQCT analysis is that it is not well standardized and validated. Similar to the previous chapter, this chapter also involved the comparison of different image registration methods. Additionally, in this chapter, I compared other parameters including the input image type, noise reduction method, density threshold, and the periosteal mask affecting the time-lapse analysis outcomes. Together, identifying the proper combination of parameters can lead to the improved reliability of time-lapse HR-pQCT analysis, and enable the standardization of this approach.

Together, Chapters 4 and 5 contribute to improving the reliability of HR-pQCT measurements in the longitudinal studies of adults including clinical trials. Our approach was to target two main challenges related to longitudinal HR-pQCT studies in the context of a multicenter clinical trial, where we reported clinically relevant data. The measurements obtained from studies included in Chapters 4 and 5 can provide deep insight into the effect of different conditions on bone over time, which can contribute to the standardized use of HR-pQCT as a reliable clinical

tool. While the population studied was adults with OI, the methodologies used or developed can be implemented for any adult population group.

Chapter 5. Spatiotemporal changes at the distal radius and tibia of adults with OI induced by setrusumab (anabolic drug):

Status: *Under preparation, will be submitted to Journal of Bone and Mineral Research (JBMR)*

Publication title:

Non-invasive quantification of bone remodeling dynamics in adults with osteogenesis imperfecta using time-lapse HR-pQCT

Seyedmahdi Hosseinitabatabaei[a,b,c], Isabela Vitienes[a,b,c], Maximillian Rummler[a,c], Annette Birkhold[d], Frank Rauch[a], Bettina M. Willie[a,b,c,†]

[a]*Research Centre, Shriners Hospital for Children-Canada, Montreal, Canada*
[b]*Department of Biomedical Engineering, McGill University, Montreal, Canada*
[c]*Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Canada*
[d]*Siemens Healthineers GmbH, Erlangen, Germany*

Abstract:
Time-lapse imaging using high-resolution peripheral quantitative computed tomography (HR-pQCT; XCT[2]) has emerged as a method to quantify bone (re)modelling. However, there is no consensus on how to perform the procedure. As part of a phase-2b multicenter trial, we used the same-day repeated scans from 29 adults (19-65yrs) with osteogenesis imperfecta (OI) to examine the influence of various parameters on HR-pQCT-derived bone formation and resorption to identify the preferred methodology. In repeated scans without motion (n=13), bone (re)modeling indicates error. Thus, we evaluated several parameters to minimize apparent (re)modeling, including the input image type (binary or grayscale), registration method (3D or matched-angle [MA]), and segmentation mask (original or dilated periosteal mask). Additionally for grayscale images, different values for the density difference between voxels to be considered formed or resorbed, the minimum size of formation/resorption clusters, and gaussian smoothing sigma were evaluated. Regardless of image registration, the binary method resulted in large errors ~13% and ~8% for XCT and XCT2, respectively. For the grayscale method, the errors were smaller for the 3D method than for MA, especially for original masks. For both XCT and XCT2, a density threshold of 200 mgHA/cm^3, and a cluster size of 0 resulted in formation/resorption volumes approaching zero, negligible effect of increasing the density threshold and cluster size, and negligible noise when combined with Gaussian noise reduction. We then validated the selected method using a combination of repeated and longitudinal scans, and showed that registration rotation angles and attenuation coefficient drift could not have produced bone formation and resorption. We finally used the selected and validated method, and found a positive dose-dependent effect of an anabolic drug (setrusumab) on bone formation and resorption, as well as net changes in bone at the distal radius and tibia of adults with OI.

5.1. Introduction:

Osteogenesis imperfecta (OI) is a collagen-related genetic disorder resulting in low bone mass, impaired quality, and increased fragility.[1,2] At least 18 genetic mutations can lead to an OI phenotype.[3] However, in most cases, OI is caused by dominant mutations in either the COL1A1 or COL1A2 gene, which code for collagen type I. The 2019 Nosology and Classification of Genetic Skeletal Disorders distinguishes four phenotypical OI types caused by these mutations: type I - mild, type II - perinatally lethal, type III - severe and type IV – moderate.[4]

Bone medications increase bone mass either by decreasing bone resorption (anti-catabolic) or increasing bone formation (anabolic).[5] Sclerostin is a protein that reduces bone formation by inhibiting WNT/b-catenin signaling in osteoblasts and is predominantly secreted by osteocytes. Several preclinical studies on OI mice models have collectively shown enhanced trabecular and cortical bone mass, and increased stiffness and strength, after treatment with sclerostin neutralizing antibody.[6–12] In an open-label, phase 2a trial, pharmacodynamics and safety of a sclerostin neutralizing antibody (setrusumab) was investigated in 14 adults with moderate OI.[13] That study showed that setrusumab stimulates bone formation, reduces bone resorption, and increases lumbar spine areal bone mineral density (aBMD) in adults with moderate OI (defined as participants with OI types I, III, or IV with a history of at least two fractures), with no treatment-related adverse events and fractures. A phase 2b, multicenter, multinational, double-blind, dose finding clinical trial on setrusumab recently showed a dose-dependent effect of setrusumab on radial and tibia volumetric BMD (vBMD) and strength using high-resolution peripheral-quantitative computed tomography (HR-pQCT), as well as spinal aBMD using DXA, in adults with OI.[14–16]

(HR-pQCT is a promising non-invasive imaging tool for measuring bone density, microstructure, and strength at the distal peripheral skeleton.[17–19] A less common application of HR-pQCT in a longitudinal setting is to monitor bone formation and resorption using a method known as timelapse HR-pQCT, by comparing to aligned images from two timepoints in a voxel-by-voxel fashion.[20,21] In contrast to the density, microstructure, and strength measurements that indicate net overall changes in bone, timelapse HR-pQCT can additionally help to elucidate the cellular mechanism behind the observed bone changes. Timelapse in vivo morphometry methods have been developed and validated in preclinical models to allow the monitoring of cortical and cancellous bone formation and resorption.[22–25] The qualitative approach was first developed by

Waarsing et al.,[25] and then expanded by other groups to quantify bone formation and resorption.[23,24,26]

To date, six research groups have implemented timelapse HR-pQCT.[21,27–33] Unlike the animal studies where segmented images were subtracted to identify bone formation and resorption, timelapse HR-pQCT studies have used binary or grayscale images as inputs. The first of four groups to use grayscale images as inputs, assessed longitudinal radial and tibial scans in nine postmenopausal women using first generation scanner to study the long-term bone remodeling with aging (XCT).[21,27] They also used repeated scans from 15 healthy adult females to evaluate errors associated with timelapse HR-pQCT and showed results with falsely determined bone remodelling of less than 0.5%, and that the least-detectable bone formation and bone resorption were 2.0 ± 1.0% and 2.2 ± 0.7% respectively. Using the same grayscale method, another group related bone strain to local changes in radius microstructure following 12 months of axial forearm loading in postmenopausal women.[28] A third group using the grayscale method showed significantly higher bone resorption at the distal radius in progressive adolescent idiopathic scoliosis compared to the non-progressive form.[29] The last group studied bone (re)modeling in metacarpophalangeal joints of participants with rheumatoid arthritis using grayscale images from the second generation scanner (XCT2) as inputs.[30] In contrast, two groups have used binary images as inputs. One quantified bone (re)modeling in the tibia of postmenopausal women induced by 6 months of high impact exercise[31], while the other used a multiple thresholding approach to assess bone formation and resorption at the distal radius of adults during fracture healing.[32,33] Although a few studies have reported the errors associated with using grayscale input images, no one has yet performed a thorough parametric analysis to determine a preferred method for timelapse analysis using HR-pQCT data.

All of the mentioned studies used the standard 3D-registration method, which is prone to interpolation error (i.e., due to image rotation).[19,34] It is not known if an alternative image registration method, known as matched-angle registration (MA),[19,34,35] can reduce the impact of interpolation error on timelapse HR-pQCT. Timelapse HR-pQCT is a promising alternative to non-site specific and invasive bone turnover markers in the blood[36] or highly invasive and labor intensive histomorphometry from bone biopsies that are limited to a single timepoint and to a non-load-bearing site that does not represent other skeletal sites.[37] However, there are remaining

questions regarding which settings to use for timelapse HR-pQCT analysis to minimize errors while maintaining sensitivity, and there is a high need for standardization, and more rigorous validations.

To date, no study has reported timelapse HR-pQCT in an OI population. Importantly, all of the mentioned studies were performed at a single center with highly trained personnel in a research setting rather than under "real world" clinical trial conditions. In a realistic multicenter clinical trial, each participant is longitudinally scanned at a single center and radiology technicians at the different clinical centers have varying levels of training using HR-pQCT. Multicenter trials are common when investigating rare diseases such as OI.

Therefore, our objectives were: 1) to identify the errors associated with timelapse analysis using different settings on the in-vivo repeated scans from the participants in the clinical trial, and to select the preferred approach based on the resulting errors; 2) to provide more rigorous validations for the selected method for timelapse analysis using the repeated scans and longitudinal data; and 3) to use the selected and validated method to evaluate bone formation and resorption induced by different doses of setrusumab in participants with osteogenesis imperfect in a clinical trial.

5.2. Methods:

5.2.1. Datasets:

Four sets of data were used for this study: 1) repeated scans from a subset of the participants (**Table 1**; mean age [SD]: 41.7 [14] years; age range: 19-65 years) of the clinical trial to identify the errors associated with different time-lapse settings, and accordingly select the preferred approach; 2) longitudinal data from the trial, from those participants (**Table 1**; mean age [SD]: 43.3 [12.6] years; age range: 19-67 years) who completed scans at baseline, 6months, and 12months, and had scans with acceptable motion (grades 1, 2, or 3). These data were used to investigate the effect of different doses of setrusumab on bone formation and resorption; 3) a smaller subset of longitudinal data from participants (**Table 1**; mean age [SD]: 42.5 [11.8] years; age range: 19-65 years) who had scans at all timepoint (baseline, 6-months, 12-months, 18-months, and 24-months), with acceptable motion; and 4) a combination of repeated and longitudinal scans from a subset of participants, in order to further validate the timelapse calculations by comparing the outcomes when using either of the repeated scans against a common scan from another timepoint. A

breakdown of the features (i.e., treatment dose, scanner generation, OI type, and sex) of the participants can be found in **Table 1**.

Table 2. Descriptive statistics of the participants and their scans in this study separated for different datasets. Each scan set for repeated scans represents duplicate scans with complete limb repositioning between scans, while 12 months and 24 months sets indicate scans from 3 and 5 timepoints, respectively. OI, osteogenesis imperfecta; SD, standard deviation; XCT, XtremeCT; XCT2, XtremeCT II

Overview of participants for each dataset					
		Repeated - all	**Repeated – no motion**	**12 months**	**24 months**
Variable	**Statistic**	**Value**	**Value**	**Value**	**Value**
Age (years)	Mean (SD)	41.7 (14.0)	39.2 (14.0)	43.3 (12.6)	42.5 (11.8)
	min < med < max	19 < 39.5 < 65	19 < 38 < 64	19 < 42 < 67	19 < 42 < 65
Weight (kg)	Mean (SD)	74.0 (21.4)	79.5 (16.9)	65.8 (19.1)	65.1 (17.5)
	min < med < max	39 < 72.5 < 118	53 < 78 < 118	20 < 62 < 118	20 < 62 < 115
Feature	**Categories**	**N**	**N**	**N**	**N**
Treatment Dose	2 mg/kg	10	6	34	26
	8 mg/kg	6	2	29	20
	20 mg/kg	13	5	62*	45*
Scanner	XCT	9	4	50	36
	XCT2	20	9	75	55
OI Type	Type I	23	10	82	58
	Type III	0	0	12	10
	Type IV	6	3	31	23
Sex	Female	18	5	86	63
	Male	11	8	39	28
Anatomical Site	Radius	23	1	70	48
	Tibia	26	12	55	43

** High-dose group includes open-label arm; hence ~2 times data compared to low and medium doses*

5.2.2. Participants and treatment:

One-hundred and thirty-one (N=131) adults were randomly assigned to either one of three different doses of setrusumab or to a placebo arm for the 24-months ASTEROID Phase 2b, multinational, randomized, double-blind, dose-finding study in adults with OI (ClinicalTrials.gov Identifier: NCT03118570). Any participants already receiving placebo were given 20 mg/kg open-label setrusumab, and new participants were randomized equally to monthly treatments by setrusumab until 12 months using 20, 8, 2 mg/kg, or 20 mg/kg open label. After 12 months timepoint, some of the participants were treated with zoledronic acid based on the discretion of the physician. The data from a subset of 55 participants that did not have any missing timepoints, and met the motion criteria (details below) were used for this study. A subset of 29 participants from 10 HR-pQCT imaging sites also had repeated scans. This subset consisted of adults with OI type I or IV. The

inclusion/exclusion criteria are provided in **supplemental methods**. The study was approved by the institutional review board of each clinical center, and informed consent was obtained from each participant.

5.2.3. HR-pQCT imaging:

The distal radius and tibia were scanned at 0, 6, 12, 18, and 24 months at 13 scanning sites (7 XCT, 6 XCT2) using high resolution peripheral quantitative computed tomography (HR-pQCT) (Xtreme CT, Scanco Medical AG, Bruettisellen, Switzerland). The scans were acquired from the nondominant arm and corresponding leg except in the case of a recent fracture or metal rod. The reference line positioning was at the medial proximal margin of the radial articular surface and at the tibial plateau.[38] The middle slice of the scanned volume of interest was located at a distance of 4% (radius) and 7% (tibia) of the ulna or tibial length from the reference line.[17,38] The XCT and XCT2 scans contained 110 slices at 82 μm and 168 slices at 60.7 μm isotropic voxel size, respectively. For a subset of participants at 10 scanning sites (6 XCT, 4 XCT2), two HR-pQCT scans of the radius and tibia were acquired on the same day by the same operator after full repositioning. Using the repeated scans, the precision errors of bone density, morphology, and micro-finite-element outcomes of the repeated scans have been reported by Mikolajewicz et al.[39] and Hosseinitabatabaei et al.[19].

5.2.4. Image quality:

At each imaging site, a trained technician graded the scout view scan, and repeated a scan (up to a total of 3 scans) if the motion grade was 4 or higher.[40] Motion grading was also performed at the central scanning center (Shriners Hospital for Children, Montreal, Canada; SHC), and scans were excluded due to a motion grade was 4 or higher, the presence of an active fracture callus in the volume of interest, or metal artifact in the volume of interest, in which case a rescan of the participant was requested. At each site, 1-2 trained technicians created the periosteal contour. A trained technician at SHC verified the periosteal and endosteal contours and modified them if needed to perform the standard evaluation protocol recommended by the manufacturer (Scanco Medical AG). All the technicians at the imaging sites and central site were blinded to the treatment dose.

5.2.5. Definition of outcomes:

The outcomes of HR-pQCT based dynamic morphometry were computed by comparing the images from two timepoints (I_t and I_{t+1}) in a voxel-by-voxel fashion using Python 3.7 (Python

scripts are available in the Willie Lab Github repository). The computed outcomes were normalized to corresponding bone measurements at the first timepoint (I_t). The 3D measure of bone formation included normalized newly mineralized bone volume (MV/BV [$\mu m^3/\mu m^3$]), defining the volume of formed bone between two timepoints. The 3D measure of bone resorption included normalized eroded bone volume (EV/BV [$\mu m^3/\mu m^3$]). Net change in volume fraction (ΔV/BV [$\mu m^3/\mu m^3$]) was defined as the difference between MV/BV and EV/BV. All outcomes were computed for total, cortical, and trabecular bone compartments, and were reported as percentages. To identify the best approach, for the repeated scans, the total bone outcomes were computed for different variations based on image registration method, periosteal mask, and input image type, as described in the following sections.

5.2.6. Image registration:

For time-lapse analysis, two images must be aligned in space, such that they can be compared in a voxel-by-voxel fashion. Two image registration methods were used to align the scans: 3-dimensional (3D), or matched-angle (MA).[19,34,35] The 3D method aligns the scans by rotating the moving image to the domain of the reference image. The MA method rotates both images, but in opposite directions, hence aiming to negate the bias introduced by rotating just one image, to attenuate the impact of interpolation error.

The first step of 3D registration was done using IPL (scripts available on Willie Lab Github repository). The first step was to find the 4 × 4 transformation matrix that aligned the second grayscale image (moving image) with the first grayscale image (reference image). We used initial alignment of the center of mass of the two images, a cross-correlation similarity metric, and the downhill simplex optimization scheme. To reduce the effect of noise on registration, only the volume within the periosteal masks were registered. The registration was performed in 3 stages (downsampling factors of 10, 4, and 1) to avoid registration errors (e.g., local minima) and to reduce computation time. For the second step, we aligned the grayscale and binary mask of the moving image with the reference image. Then, the grayscale images were binarized using the standard Scanco's segmentation protocol and segmented using their respective masks. Nearest neighbor and cubic interpolations were used for the binary and grayscale images, respectively.[41] Next, to ensure that all of the image slices contain closed cortical masks, and to prevent any artifacts on the partial slices created due to image rotations, it was cropped at the top and bottom

to create flat surfaces (i.e., the slices containing partial bone cross-sections were identified automatically during time-lapse analyses).

For the MA method,[35] the image registration settings were identical to the 3D-registration. We rotated the reference and moving images using rotation angles of half of the original angles, but in opposite directions. This way, the MA method aligns the two images in a middle domain. Briefly, using the registration function in IPL, we divided the rotation angles from registration by half to identify the middle domain. Then, each of the reference and moving images where registered to the middle domain.[34] Finally, we performed the analysis on the flat common volume.

5.2.7. Input image type:

Since there is no agreement on which input image type to use for time-lapse analysis, we performed our analyses using both the aligned grayscale and binary input images. Binary images were created after image registration, using the standard protocol for each scanner generation. When using the binary input images, bone formation and resorption can be identified using triangulation (Boolean operation). Briefly, when a voxel is foreground (i.e., bone) in the baseline image, and background (i.e., void) in the followup image, it is considered bone resorption, while the opposite means bone formation. Voxels common between the images are classified as quiescent.

For the grayscale approach, the density values of the baseline and follow-up images were subtracted voxel-by-voxel, resulting in an image containing the differences in local densities between two images. Since the images may contain artifact from different sources (e.g., differences in calibration between scans, image interpolation error introduced by image rotation during image registration, and noise in the images), not all the differences in density will indicate local biological bone changes. To reduce the effects of errors on the results, we used the common two-step filter approach.[21] In the first step, a global threshold expressed in the units of density ($mgHA/cm^3$) is defined and only voxels for which the change in bone density exceeds the positive and negative value of the threshold are classified as sites of bone formation and bone resorption, respectively. Next, to reduce the noise from the results, from the voxels classified as formation or resorption, only those voxels forming consistent clusters of at least a predefined number of voxels were remained, while the small clusters were removed. In addition to the minimum cluster size, we investigated using Gaussian smoothing for noise reduction, by reducing the noise in the images prior to subtraction and thresholding. To denoise the grayscale images, we applied a complete

Gaussian kernel and identified the regions of bone formation and resorption. Since the three variables (i.e., density threshold, minimum cluster size, and Gaussian smoothing), determine the outcome of the approach, we used several values for each variable on the repeated scans: density threshold {125, 150,…, 375}, minimum cluster size {0, 5, 10, 15}, and Gaussian σ {0, 0.4, 0.8, 1.2, 2.0}.

5.2.8. Periosteal mask:

The periosteal mask is created during the HR-pQCT image analysis to segment the bone from the soft-tissue. Since the images from different timepoints are required to be aligned in space, image rotation is required, which can introduce errors at the edges of the periosteal mask. Although such errors are shown to be negligible for microFE analysis[19] that are averaged from the whole bone structure, they may be more confounding for timelapse analysis due to its voxel-based nature. Therefore, we performed timelapse analysis using two variations of the periosteal mask: 1) using the original periosteal mask of each image to segment the bone; and 2) using the dilated mask (5 voxels) of each image, to ensure that the periosteal surface is not affected by the image rotation, or possible contouring errors.

5.2.9. Validation of timelapse calculations:

Since dynamic histomorphometry is not feasible in humans, there is no gold standard to validate regions of bone formation and resorption obtained from time-lapse HR-pQCT. Therefore, we used our repeated scans to perform an indirect validation. This could be achieved by aligning each of the repeated scans from a pair to a scan from another timepoint. In this case, similar outcomes from the two timelapse analysis on the same volume of interest indicates that the identified regions of bone formation and resorption are not due to random noise, but are true changes. This analysis was done on scans that did not have any motion, which resulted in a total of 21 tibial test cases from 6 participants (i.e., each test case consists of two repeated scans, and a scan from another timepoint).

5.2.10. Statistical analyses:

The statistical analyses were performed using SAS 9.4 (SAS Institute Inc., Cary, NC).

Least detectable changes from repeated scans: Any apparent bone formation and resorption from the repeated scans is considered error, or the least detectable changes. We computed the average and standard deviation of volume outcomes of formation (MV/BV) and resorption (EV/BV) from all pairs of repeated scans that were free from motion (n=13 [4 XCT])

for different variations based on image registration method, periosteal mask, and input image type. We used Bland-Altman plots and linear regression to assess the agreement between the 3D vs MA registrations, as well as using the same vs different periosteal mask. To identify the best settings, our target was to obtain the smallest possible errors (~ 0%). For the grayscale method, from all possible combinations of the density threshold, minimum cluster size, and Gaussian sigma, additional criteria were used including minimized sensitivity of outcomes to changes in variables, and the selection of the smallest possible values for the variables.

Validation of timelapse HR-pQCT method: We used a mixed-effect (multilevel) model to compare the outcomes between repeated scans, and pairs to account for the dependency of observations between the test cases from each participant (i.e., 21 test cases from 6 participants). The model included random intercept for each subject, and the need for random slope for each group was verified visually, and based on reduced Akaike information criterion (AIC). The test cases were nested within each subject. Pairwise comparisons were performed using the Games-Howell test to account for unequal variances of the estimated means between groups. The significance level was $p < 0.05$ and trends at $p < 0.1$. Moreover, we assessed whether differences in MV/BV and EV/BV between repeated and longitudinal scans are due to the differences in interpolation error from different rotation angles from 3D registration. This was done by categorizing the relationship between the differences in rotation angles and differences in MV/BV and EV/BV into two groups, supporting the positive relationship (i.e., [larger rotation angle with larger MV/BV and EV/BV] and [smaller rotation angle with smaller MV/BV and EV/BV) and a negative relationship (i.e., [larger rotation angle with smaller MV/BV and EV/BV] and [smaller rotation angle with larger MV/BV and EV/BV). Then we compared the distribution of the data to that of equal number of cases supporting positive and negative relationship using chi-squared test of goodness-of-fit ($p<0.05$).

setrusumab dose-dependent bone formation and resorption: Our aim was to investigate the effect of treatment dose while accounting for differences in scanner generation, OI type, and motion grade. Prior to modeling our data, we first visually confirmed a correlation between the measurements from a participant from different timepoints. Accordingly, we fit a generalized linear mixed model to account for the correlations.

To identify the best fit model, we first created an empty means model (also known as unconditional mean model) to determine the amount of variance attributed to between-person and within-person differences. We computed intra-class correlation (ICC) to determine the amount of variance attributed to the within- and between-subject differences, and model likelihood ratio test to determine the need for additional factors. We thus added fixed effects to the model, and assessed random intercept and random slope models. We determined the best fit model by the largest reduction in AIC, parsimony of the model, and visual inspection of the data, and accordingly included random intercept and slope (with respect to time) for subjects. The between-subject factors included scanner generation, treatment dose, OI type, and treatment × time interaction. Time was the within-subject factor. To account for the variations in motion grade between the different scans of each participant over time, motion was included as both the between- and within-subject factors (also known as time-varying predictor). The motion grade was calculated by adding the grades from each of two scans (e.g., motion grade of $2 = 1 + 1$ means no motion), and both the between- and within-subject grades were centered around their respective means.

The model residuals were checked for normality around zero and homoscedasticity. Although residuals were normally distributed, in some cases, heteroscedasticity of residuals was observed. Moreover, when using the identity link function, some of the estimated means from the model were negative, which was not in agreement with the raw data and the nature of the outcomes. Therefore, for those cases, we created the model assuming a normal distribution, while using non-linear link functions, for which a cube root was adequate. The estimated means were conditional on the absence of motion. In case of significant effect, OI types were compared using the Tukey-Kramer test, while for comparisons between treatment doses at each timepoint, a t-test with Benjamini-Hochberg correction was used. The significance level was $^{**}p < 0.01$, $^{*}p < 0.05$, and trends at $^{\#}p < 0.1$, while in case of significance prior to p-value adjustment, the symbols were enclosed in circle.

Associations with bone density and strength: Using linear regression, we evaluated the explained variance in percentage changes in total volumetric bone mineral density (vBMD), stiffness, and failure load for the $\Delta V/BV$. The vBMD measurements were acquired from 3D-registered regions, while the stiffness and failure load were computed from micro finite-element (μFE) analysis on scans registered using cross-sectional-area (CSA) registration.[19]

5.3. Results:

5.3.1. Assessment of Methodology:

Binary method results in high errors in MV/BV and EV/BV

The average MV/BV and EV/BV for the repeated participant scans when using binary input images is ~13% for XCT, and ~8% for XCT2, regardless of the image registration method (Figure 24-A). These values indicate error in the method since no biological change should be observed over the ten-minute period when the two repeated scans for each participant were conducted. Thus this value is the smallest values of MV/BV and EV/BV that can be reliably detected in longitudinal scans. However, such large changes (8-13%) are not reasonable in bone over a 6-month period. When using the binary method, bone formation and resorption can be consistently observed all across bone surface in the repeated scans (Figure 24-B), which is due to partial volume effect, interpolation error, imperfection in the alignment of the two images after registration, and segmentation errors.

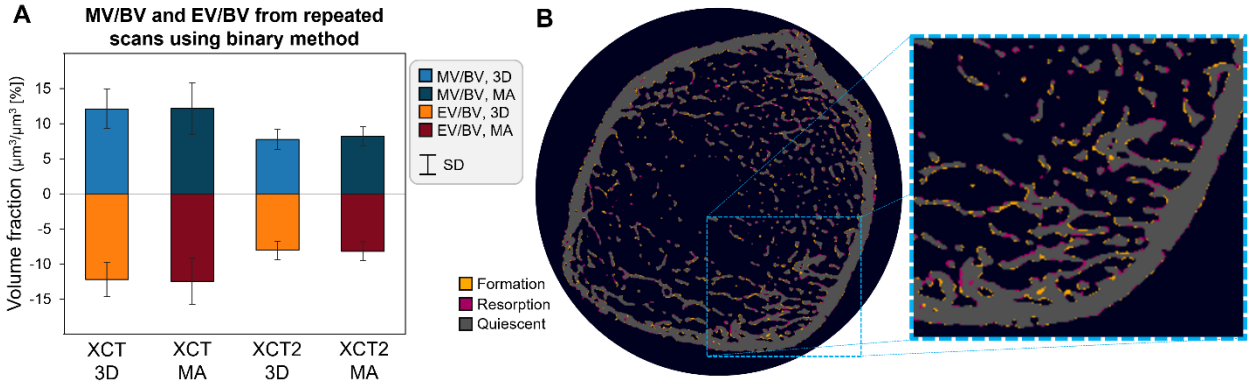


Figure 24. Relatively high errors associated with the binary method. A) The average mineralized volume fraction (MV/BV) and eroded volume fraction (EV/BV) from repeated scans using binary input images for each scanner generation (XCT and XCT2) and registration method (3D and matched-angle [MA]). The error bars indicate standard deviation (SD). **B)** an example of a slice from a repeated scan showing regions of bone formation (orange), resorption (purple), and quiescent (gray), present all across bone surface.

The grayscale method involving Gaussian smoothing is the preferred method

Without Gaussian smoothing, the density threshold that would result in MV/BV and EV/BV less than 1 $\mu m^3/\mu m^3$ [%] for the repeated scans was ~275 mgHA/cm^3, and ~375 mgHA/cm^3, for XCT (Figure 25-A, *left panel*) and XCT2 (Figure 25-B, *left panel*), respectively. A larger threshold for XCT2 was expected, due to larger errors associated with higher resolution and usual lower signal-

to-noise ratio (SNR), as more speckles were observed for XCT2 scans. In addition, these thresholds need to be followed by applying a minimum cluster size or 10 voxels (XCT) and 15 voxels (XCT2), which is rather subjective. On the other hand, when using Gaussian smoothing to reduce noise in the grayscale images prior to subtraction, errors less than 1 $\mu m^3/\mu m^3$ [%] could be achieved at a density threshold of 200 mgHA/cm^3 for both XCT and XCT2 (Figure 25-A and B, *middle panel*), with no need for a cluster size filtering. This threshold is also well below 320 mgHA/cm^3 and 450 mgHA/cm^3, which are the typical thresholds for segmenting the trabecular and cortical bone compartments, respectively. Therefore, the grayscale method on 3D-registered scans with the same dilated mask, and Gaussian smoothing with a threshold of 200 mgHA/cm^3, and no minimum cluster size were selected as the preferred settings. It must be noted that motion leads to larger error. This can be observed in the *right panel* of Figure 25, where the MV/BV and EV/BV of the selected method for the repeated scans with motion is seen to be ~6 $\mu m^3/\mu m^3$ [%] (XCT) and ~10 $\mu m^3/\mu m^3$ [%] (XCT2).

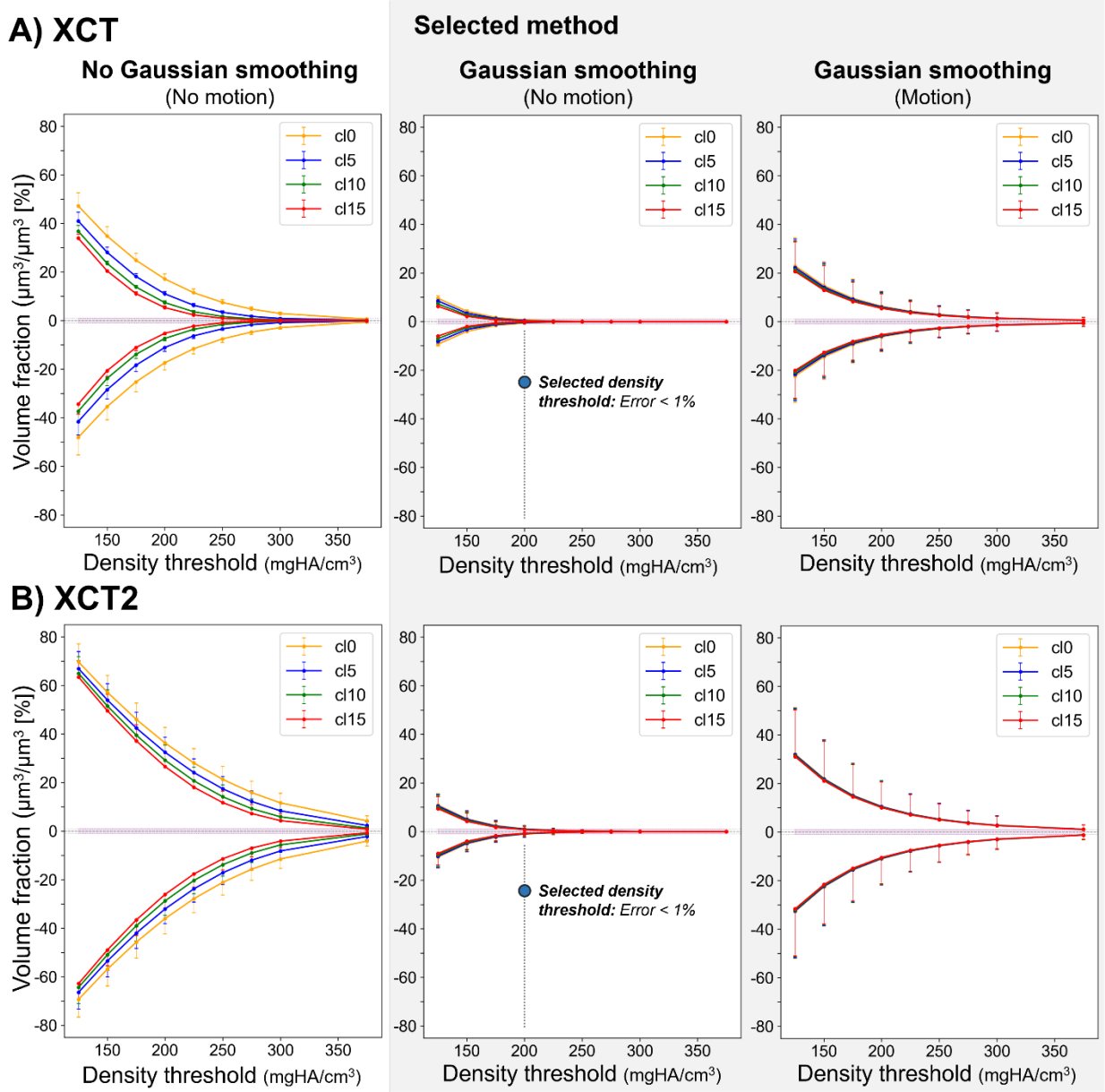


Figure 25. Identifying the superior method between method involving and not involving Gaussian smoothing. The average mineralized volume fraction (MV/BV) and eroded volume fraction (EV/BV) from repeated scans using grayscale input images for **A)** XCT**;** and **B)** XCT2. The left panel illustrates the method without Gaussian smoothing, while the middle panel indicates the method with Gaussian smoothing. The right panel indicates the results for the repeated scans with motion using the method with Gaussian smoothing. The x-axis indicates different density thresholds, while the y-axis indicates the MV/BV (positive), and EV/BV (negative). Each colored line shows different minimum cluster sizes. For each scanner, the lowest threshold and minimum cluster size that would results in MV/BV and EV/BV less than 1% was selected (middle panel). The error bars show standard deviation (SD).

3D image registration and dilated periosteal mask are superior

For the grayscale method, the Bland-Altman plots indicate that MV/BV and EV/BV of the repeated scans are almost identical, while slightly smaller for the 3D-registration compared to MA-registration when using the same dilated periosteal mask (Figure 26-A and B, *left panel*). However, when using the original periosteal masks of each image, the differences become larger (Mean difference: ~1.4% and ~0.4% for MV/BV and EV/BV, respectively) in favor of 3D-registration (Figure 26-A and B, *middle panel*). There was an almost perfect correlation between 3D and MA registration methods with $R^2 = 0.99$, regardless of the periosteal mask. The MV/BV and EV/BV were also larger when using the original periosteal masks compared to the dilated mask, with mean differences of ~0.9% and ~1.3% for MV/BV and EV/BV, respectively (Figure 26-A and B, *right panel*). Therefore, the grayscale method using the same dilated mask and 3D-registration resulted in the lowest error and thus are preferred.

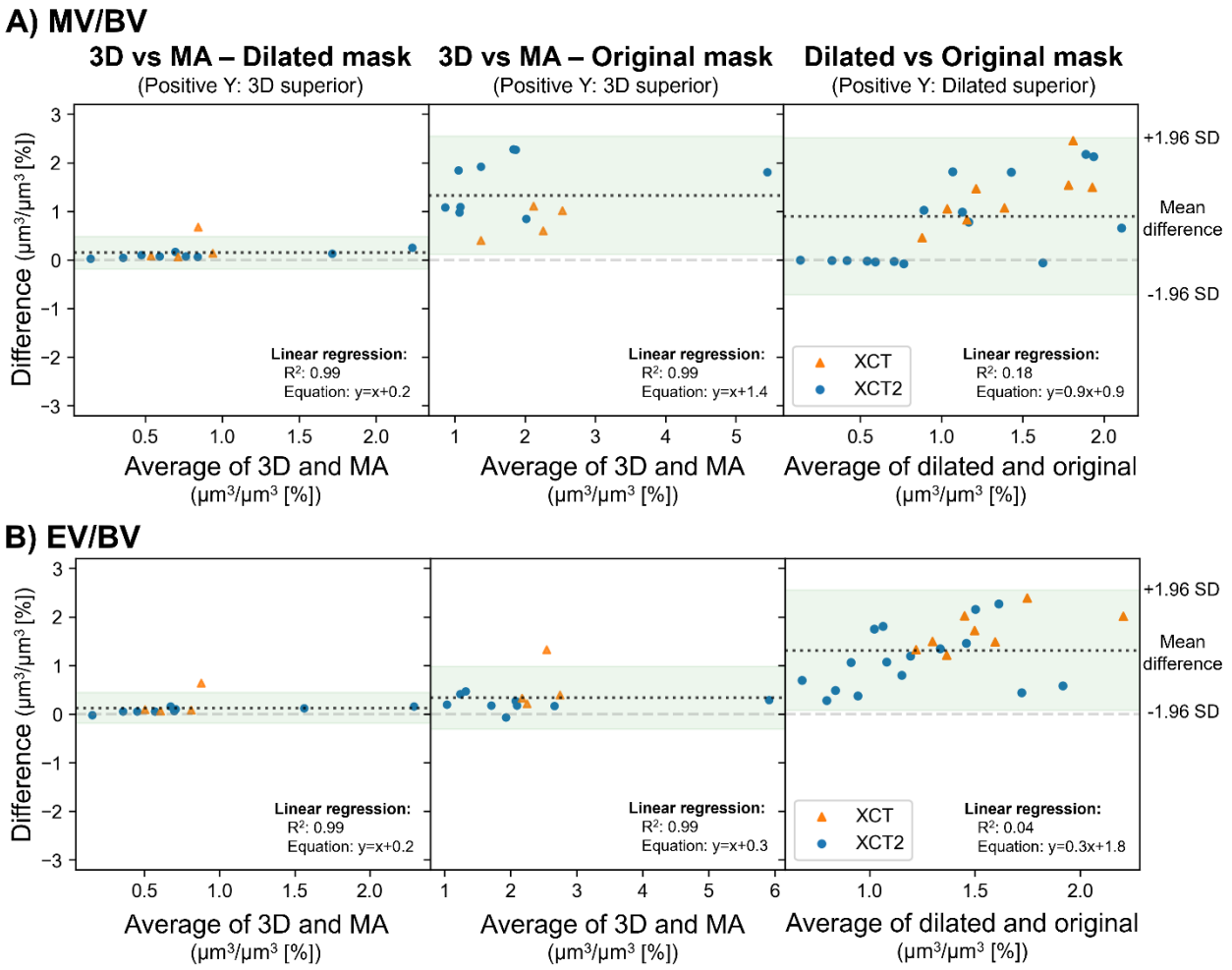


Figure 26. Identifying the superior image registration and periosteal mask for grayscale method using Bland-Altman plots from repeated scans indicating the agreement between A) mineralized volume fraction (MV/BV); and **B)** eroded volume fraction (EV/BV) from different image registration methods (3D vs matched-angle [MA]) using dilated masks (left panel) or original masks (middle panel), as well as between the dilated and original masks (right panel). The x-axis indicates the average value from the two methods, with the y-axis indicated the difference. The units are the same unit as the outcomes ($\mu m^3/\mu m^3$[%]). All data are from analysis using grayscale input images with Gaussian smoothing. Data shown for comparing the original and dilated masks in the right panel are from both 3D and MA registrations. Data points are separated for scanner generations, with orange rectangles for XCT, and blue circles for XCT2. The coefficient of determination (R^2) and linear regression equations are also shown on each graph.

The selected method can reliably identify changes over time

To validate the selected method, we used a data set consisting of repeated and longitudinal scans. More specifically, we matched each of the two repeated scans with a scan from the participant at

another timepoint, resulting in the calculation of 3 sets of MV/BV and EV/BV as follows: 1) between the two repeated scans (named “Repeated”); 2) between one of the repeated scans, and the scan from another timepoint (named “Pair 1”); and 3) between the other repeated scan, and the same scan from another timepoint (named “Pair 2”). MV/BV was significantly smaller for the repeated scans (mean: 0.75%) compared to pair 1 (mean: 2.88%, $p = 0.04$) and pair 2 (mean: 3.08%, $p = 0.02$), while there was no evidence of difference between pair 1 and pair 2 ($p = 0.95$) (Figure 27-A). EV/BV for the repeated scans (mean: 0.74%) tended to be smaller than pair 1 (mean: 3.35%, $p = 0.06$), while it was smaller than pair 2, at a borderline significance level (mean: 3.56%, $p = 0.05$) (Figure 27-A). There was no evidence of difference between pair 1 and pair 2 ($p = 0.98$). These results indicate that the selected method is capable of preventing various sources of error, while being sensitive to biological changes. Qualitative visual assessment of the scans confirms negligible bone formation and resorption between the repeated scans. Furthermore regions of formation and resorption were observed at the same locations within pair 1 and pair 2, thus confirming that the identified regions are not from errors in the images, but are in fact biological changes (Figure 27-B).

In ~90% of the cases, MV/BV and EV/BV were smaller in repeated scan compared to pair 1 and pair 2 (Figure 27-C). In the remaining ~10%, MV/BV and EV/BV were only slightly larger in the repeated scans (i.e., repeated – pair 1 < 0.3%). Differences in MV/BV and EV/BV between pair 1 and pair 2 were approximately zero for most cases.

There was a concern that rotational angle may affect the results, however from all cases, there was not enough evidence of differences between the number of supporting a positive relationship (Figure 27-C, quadrants 1 and 3) and supporting a negative relationship (Figure 27-C, quadrants 2 and 4) cases ($p = 0.17$). Similarly for EV/BV, 40% did not support the relationship between smaller rotation angle and smaller EV/BV, with no evidence of differences between the number of cases supporting positive and negative relationships ($p = 0.11$) (Figure 27-C). For the repeated scans, their MV/BV and EV/BV was smaller than those of pair 1 and pair 2 regardless of the rotation angles.

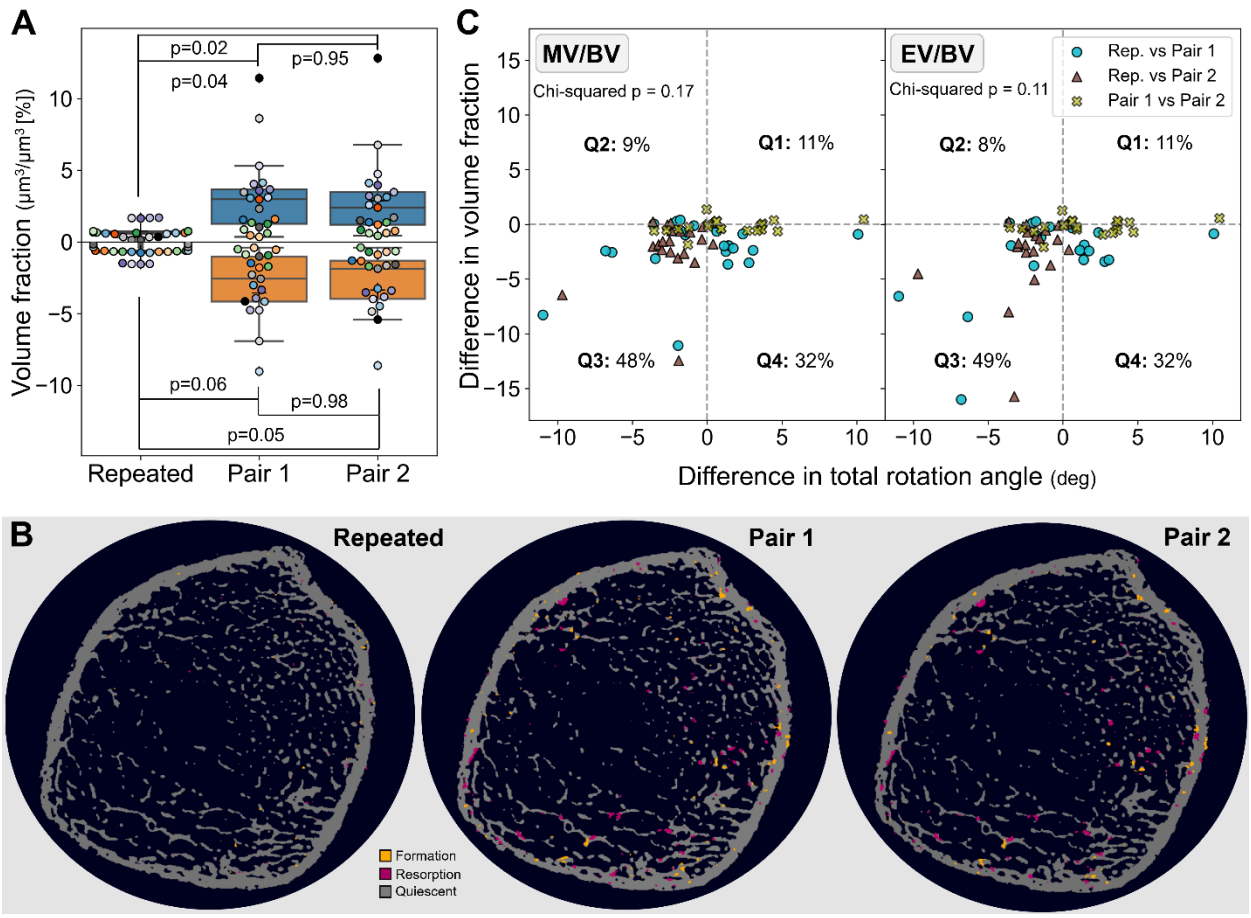


Figure 27. Validating the selected method involving grayscale images smoothed using Gaussian filter. A) Mineralized volume fraction (MV/BV) and eroded volume fraction (EV/BV) from repeated scans, pair 1, and pair 2. Each pair is created from one of the repeated scans, and a scan from another timepoint. The data from each case (i.e. repeated scans + scan from another timepoint) are shown by a unique color. P-values are from Games-Howell post-hoc test from the mixed model. **B)** An example of a case showing negligible formation and resorption between repeated scans, and similar formation and resorption from two pairs. **C)** Scatter plot showing the difference in MV/BV and EV/BV based on differences in total rotation angles between scans. The distribution of the data on each quadrant (Q1, Q2, Q2, Q4) of the graph (e.g., top left or bottom right) are shown in percentages. The Chi-squared goodness-of-fit p-value indicates no difference between the cases where smaller rotation angle supports smaller MV/BV and EV/BV (i.e., Q1 and Q3) versus the cases that it does not support it (i.e., Q2 and Q4), suggesting the absence of relationship between different in total rotation angles, and MV/BV and EV/BV.

The grayscale method is not sensitive to attenuation coefficient drift

To investigate whether the selected method is sensitive to attenuation coefficient drift (i.e., when the measured attenuation coefficient and associated voxel intensities decrease over time as the tube

ages), we created Shewhart control charts of the measurements from the first Scanco quality control phantom (QC1) over time, and fitted a linear line to identify the slope, indicating the rate at which the density measurements drifted per day (Figure 28-A). This was done separately for each rod in the QC1 phantom, since different rates of change per day were observed. Next, we plotted the computed slopes against the density of each rod to derive an equation explaining the attenuation drift per day based on voxel density (Figure 28-B). For instance, a voxel with a density of 400 mgHA/cm^3 would experience a decrease in density of 0.54 mgHA/cm^3, while a voxel of 1200 mgHA/cm^3, would decrease by 1.62 mgHA/cm^3 over 180 days (time interval between two timepoints in this study). Finally, we recalibrated the follow-up scan using the derived equation to simulate the computed drifts in density. The Bland-Altman plot of average MV/BV or EV/BV from scans with and without simulated drift against their differences showed negligible changes in MV/BV and EV/BV, with differences increasing linearly with increased average (e.g., difference of ~0.03% for MV/BV or EV/BV of 2%) (Figure 28-C, *top panel*). The differences were largest for cortical bone and smallest for trabecular bone, which was expected, since the drift was linearly related to density. Next, we conducted a worst-case analysis by simulating an artificially inflated drift of ~18 mgHA/cm^3 for a voxel of 800 mgHA/cm^3 over 6 months. This value was selected since at a drift of ~16 mgHA/cm^3, the scanner would typically require recalibration, as per the manufacturer. However, in practice, we are not aware of this large of a drift ever occurring over a 6-month period. In this extreme case, while the absolute errors were larger as expected, they were still small relative to MV/BV and EV/BV (e.g., difference of ~0.3% for MV/BV or EV/BV of 2%) (Figure 28-C, *bottom panel*).

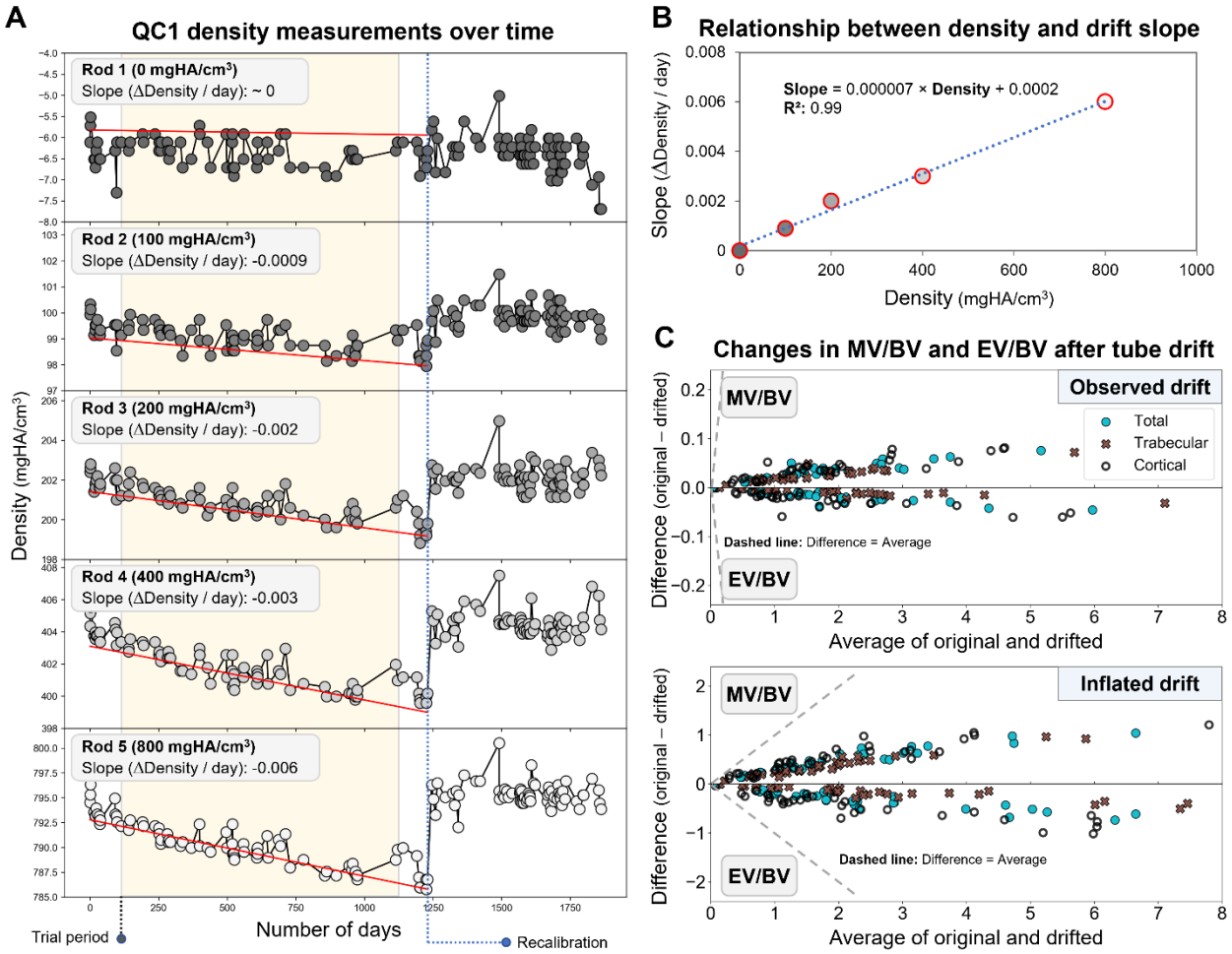


Figure 28. Illustration of lack of sensitivity of the selected method on x-ray tube drift. A) Shewhart control chart for the first quality control (QC1) phantom measurements over time shown separately for each rod. A linear line is fitted for the portion with drift prior to recalibration to identify the drift in density measurements per day. **B)** Plot of the slopes from part A (i.e., density drift per day) plotted against the average nominal density of each rod. A linear line is fitted to identify the relationship between density and the amount of density drift per day. **C)** Changes in MV/BV and EV/BV after applying the observed x-ray tube drift to the followup scans (C, top panel), and an inflated drift (i.e., ~16 times the values obtained from part B; C, bottom panel) against the average values for total, cortical, and trabecular bone compartments.

5.3.2. Implementation of Methodology on Clinical Data:

Dose-dependent effect of setrusumab on bone formation and resorption

For the 12-month analysis at the radius, for the baseline-6 month interval, there was a tendency ($p < 0.1$) toward higher formation and resorption for high dose versus medium dose. For the 6-12 month interval, the high dose had significantly higher ($p < 0.05$) formation and resorption than the

low dose (Figure 29-A). Nevertheless, the statistical significances and tendencies were nullified after p-value correction. While a dose-dependent increase was observed in both formation and resorption, the net change in volume fraction (ΔB/BV) was in favor of bone formation (i.e., positive ΔV/BV) for high dose in total and cortical bone compartments (statistical tendency of $p < 0.1$ for cortical bone before p-value correction) (Figure 29-B). On the other hand, for the low and medium doses, bone resorption tended to be larger than formation with no statistically significant difference. For the trabecular bone, all doses favored bone resorption, while the low dose had a significant domination of resorption ($p < 0.05$) (Figure 29-B).

For the 12-month analysis at the tibia, the medium and high doses resulted in similar MV/BV, while the high dose cortical formation was significantly higher than low dose at both baseline-6 month and 6-12 month intervals ($p < 0.05$) (Figure 29-C). All doses had similar resorption for the total and cortical bone compartments. For the trabecular bone, the medium dose had the highest formation and resorption, which was significantly higher than the low dose (tendency at $p < 0.1$ for resorption) (Figure 29-C). Similar to the radius, ΔV/BV was in favor of bone formation (i.e., positive ΔV/BV) in high dose for total and cortical bone compartments, neutral for the medium dose, and in favor of resorption for the low dose (Figure 29-D). However, the statistical significances were lost after correction for multiple comparisons. There was a significant difference between cortical ΔV/BV of the high and low doses at 6-12 month interval, even after p-value adjustment ($p < 0.05$). For the trabecular bone, a slight positive dose-dependent pattern existed (Figure 29-D).

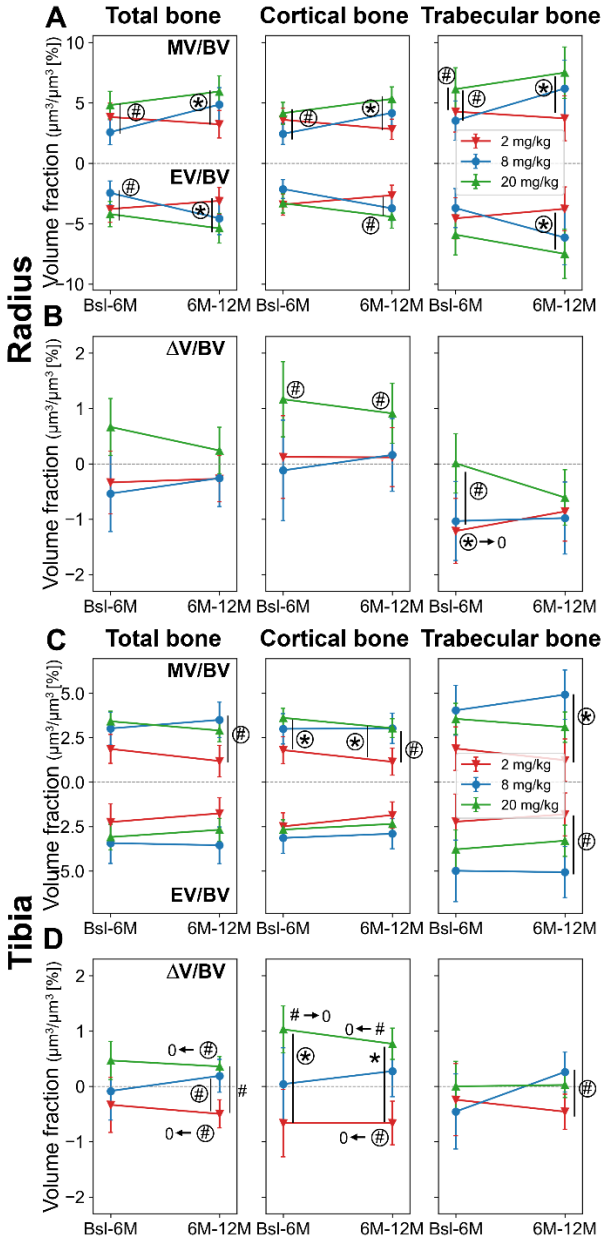


Figure 29. Mineralized volume fraction (MV/BV) and eroded volume fraction (EV/BV) for different timepoints until 12 months for different treatment doses (red: 2 mg/kg, blue: 8 mg/kg, and green: 20 mg/kg) at the **A)** radius, and **C)** tibia. Net changes in volume fraction (ΔV/BV) are shown at the **B)** radius, and **D)** tibia. The left, middle, and right panels show the results for total, cortical, and trabecular bone compartments, respectively. *p<0.05 and #p<0.1 comparing the estimated means (error bars: standard error of mean [SEM]) for different treatment doses are shown from the mixed model. For ΔV/BV in addition to comparing treatment doses, each dose was compared to zero, with p-values shown by "→ 0" or "0 ←" sign. The symbols enclosed in circles indicate prior to p-value adjustment.

The 24-month results at the radius, which included fewer participants mostly due to s higher rate of missing timepoints (since one inclusion criterion was no missing timepoint), were consistent with the 12-month results in suggesting a positive dose-dependent effect of setrusumab on bone formation and resorption, regardless of the bone compartment, especially during the drug administration phase (i.e., until 12 months) (Figure 30-A). The high dose had significantly higher formation and resorption at the 6-12 month time interval compared to the low dose for the total ($p < 0.05$) and cortical ($p < 0.01$) bone compartments, except for total bone resorption with a tendency ($p < 0.1$) (Figure 30-A). Formation for the high dose was also higher than medium dose ($p < 0.05$) in the cortical bone. There was no dose-dependent significant difference for $\Delta V/BV$ observed, while the high dose showed a pattern of formation-favored changes for the total and cortical bone compartments during the drug administration phase (Figure 30-B). The $\Delta V/BV$ for the trabecular bone was in favor of resorption for all doses.

For the 24-month results at the tibia, there were no differences between formation and resorption from different doses, except for high trabecular resorption in high dose compared to medium dose ($p <0.05$) (Figure 30-C). At the baseline-6 month time interval, all doses showed formation-favored changes, which was diminished at later time intervals (Figure 30-D). At the 18-24 month time interval, the high dose showed significantly resorption-dominated changes at all bone compartments ($p < 0.05$) (Figure 30-D). All of the p-values from the 24-months analyses were nullified after correction.

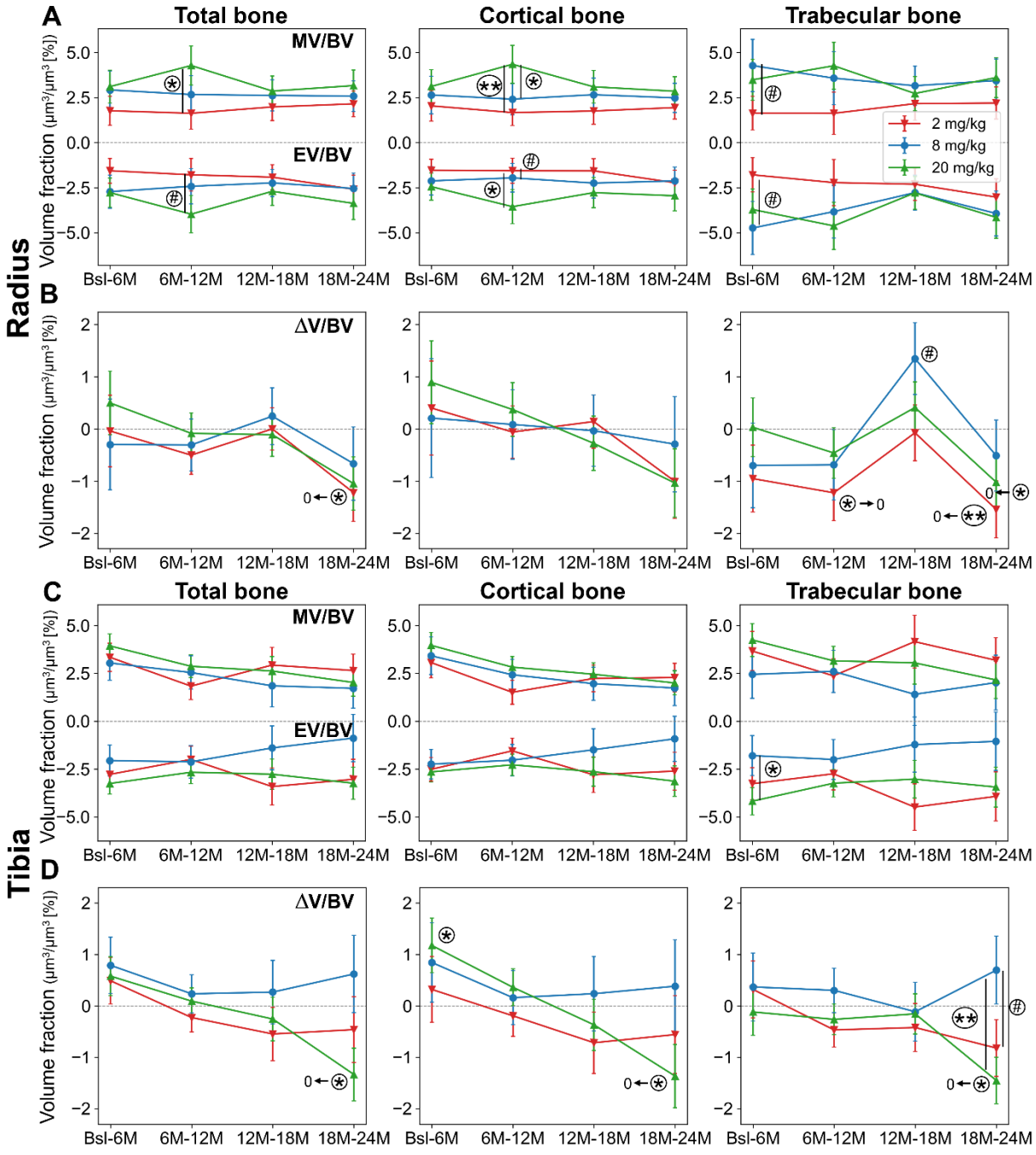


Figure 30. Mineralized volume fraction (MV/BV) and eroded volume fraction (EV/BV) for different timepoints until 24 months for different treatment doses (red: 2 mg/kg, blue: 8 mg/kg, and green: 20 mg/kg) at the **A)** radius, and **C)** tibia. Net changes in volume fraction (ΔV/BV) are shown at the **B)** radius, and **D)** tibia. The left, middle, and right panels show the results for total, cortical, and trabecular bone compartments, respectively. **p<0.01, *p<0.05 and #p<0.1 comparing the estimated means (error bars: standard error of mean [SEM]) for different treatment doses are shown from the mixed model. For ΔV/BV in addition to comparing treatment

doses, each dose was compared to zero, with p-values shown by "→ 0" or "0 ←" sign. The symbols enclosed in circles indicate prior to p-value adjustment.

Effect of OI type on bone formation and resorption

For the tibia from the 12-months dataset, total bone ΔV/BV was bigger for OI type I ($p < 0.01$) and type III ($p < 0.01$) than type IV.

For the 24-months data at the tibia, compared to OI types I and IV, OI type III had bigger total MV/BV ($p < 0.001$, $p < 0.001$), total ΔV/BV ($p < 0.001$, $p < 0.0001$), cortical MV/BV ($p < 0.01$, $p < 0.01$), cortical ΔV/BV ($p < 0.01$, $p < 0.0001$), trabecular MV/BV ($p < 0.001$, $p < 0.01$), trabecular EV/BV ($p < 0.01$, $p < 0.05$), and trabecular ΔV/BV ($p < 0.01$, $p < 0.01$). For the 24-months tibia, OI type I was bigger than type IV for total ΔV/BV ($p < 0.05$), and cortical ΔV/BV ($p < 0.01$). At the radius, OI type was significant only for the trabecular ΔV/BV for the 24-months dataset, with OI type III bigger than type IV ($p < 0.01$).

The net change in volume fraction is associated with changes in bone density and strength:

The net change in volume fraction (ΔV/BV) explained ~50% of the variance in the percentage changes in total bone mineral density (Tt.vBMD; $R^2 = 0.49$, $p < 0.0001$), stiffness ($R^2 = 0.50$, $p < 0.0001$), and failure load ($R^2 = 0.50$, $p < 0.0001$) in a significant positive linear correlation (Figure 31). Furthermore, the average ΔV/BV at each timepoint follows similar patterns and values as the average percentage changes in Tt.vBMD (Supplemental Figure 1-A), stiffness (Supplemental Figure 1-B), and failure load (Supplemental Figure 1-C) for all treatment doses.

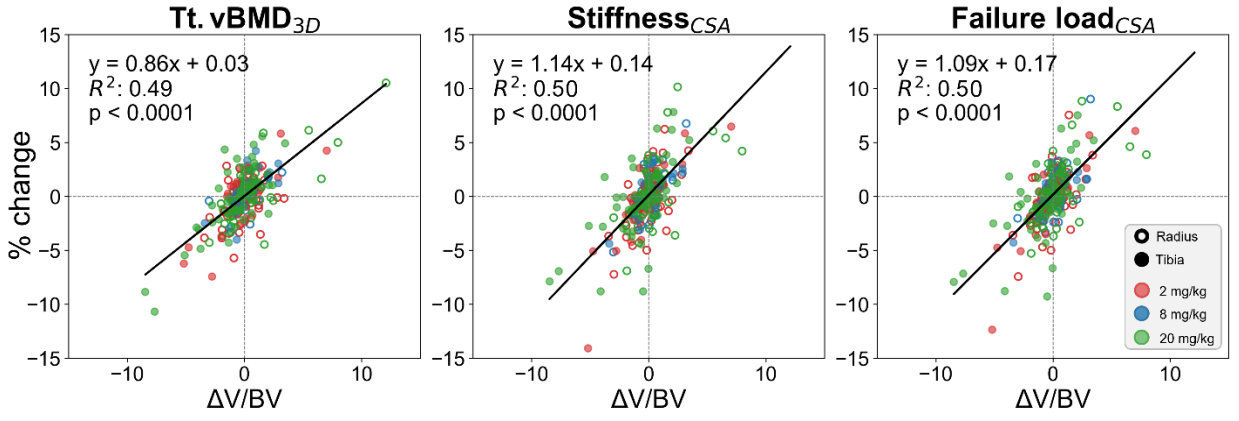


Figure 31. Correlation between net change in volume fraction (ΔV/BV) [x-axis] and percentage change in total volumetric bone mineral density (Tt.vBMD) [y-axis, left panel], micro finite-element (μFE) derived stiffness [y-axis, middle panel], and μFE derived failure load [y-axis, right panel]. Tt.vBMD was calculated from 3D-registered scans, while μFE outcomes were derived from scans registered using the scanners default cross-sectional-area (CSA) registration. Coefficient of determination (R^2) and p-values testing the significance of the correlation are shown on each graph. Hollow markers indicate radius, while filled markers are tibia. Data from different doses are color-coded.

5.4. Discussion:

In this study, we used a systematic approach to compare different variables when performing timelapse HR-pQCT analysis as a non-invasive tool to quantify bone formation and resorption using repeated scans. Following the identification of the preferred approach based on the minimal errors from the repeated scans, we performed several validations of the preferred method. Finally, we implemented the preferred and validated method to assess bone formation and resorption in a dose-finding clinical trial on osteogenesis imperfecta (OI).

Since some studies used grayscale input images, while others used binary input images, we first compared these two methods. The average MV/BV and EV/BV for the repeated participant scans when using binary input images is ~13% for XCT, and ~8% for XCT2 (Figure 24). Considering these errors, MV/BV and EV/BV values smaller than ~10-15% may not be identified reliably, which is not reasonable in many cases, unless for extremely long time intervals between two scans. The first of the two groups that used the binary method[31] reported root-mean-squared coefficient of variation (RMSCV%) of 2.1% and least-significant-change (LSC) of 1%/month for bone remodeling rate of the same-day repeated scans from XCT scans from older adults. However, comparing our results to this study is not feasible. First, they reported bone remodeling rate as percentage changes per month, while it is not clear how the results from same-day scans were converted to changes per months. In the case of multiplying the results from the repeated scans by 30 days, the reported LSC translates to 0.03% (0.03% = 1% ÷ 30 days) for the same-day repeated scans (works only if assuming 1 day between scans instead of 0 days), which is an extremely small error (compared to ~10% for this study). It may be worth noting that while they used the same segmentation algorithm, they used a different image interpolation method (i.e., Lanczos vs cubic) although such large differences are not expected. Second, it is not clear how RMSCV% was

computed for timelapse outcomes using only two scans. The other study used a multi-threshold binary method on fracture healing,[32] as well as contralateral limbs. While they did not have same-day scans, they included scans with two-week time intervals, and reported MV/BV and EV/BV of ~18% and 8% for the trabecular and cortical compartments in the absence of motion, which is comparable to our results. More recently, the same group used the multi-threshold method on the same dataset to study the relationship between bone remodeling and mechanical loading from finite-element analysis.[33] The higher trabecular results are expected due to higher surface to volume ratio. By visual inspection, we observed consistent bone formation and resorption all across bone surface in the repeated scans using the binary method, which is due to partial volume effect, interpolation error, imperfection in the alignment of the two images after registration, and segmentation errors. This consistent formation and resorption is also hardly distinguishable from true changes. Accordingly, we conclude that the binary method is not a reliable method, especially to assess the magnitude of bone formation and resorption.

Using grayscale input images is more common for time-lapse HR-pQCT. In the method used in the literature, two unfiltered grayscale images are subtracted, and the difference image is thresholded to identify formed or resorbed voxels. Next, the identified clusters of formed or resorbed voxels are filtered to reduce the effect of noise. In the first study introducing timelapse HR-pQCT using XCT scanner,[21] a density threshold of 225 mgHA/cm^3 and a minimum cluster size of 30 voxels were selected based on the criteria of MV/BV and EV/BV < 2.5% from a sensitivity study using *in vivo* repeated scans (time interval between two scans varied between one day and one week) from healthy adults. Later, the same group[27] repeated the sensitivity analysis using same day *ex vivo* scans based on combined MV/BV and EV/BV < 0.5%, and recommended the same density threshold of 225 mgHA/cm^3, while reduced the minimum cluster size to 5 voxels for the XCT scanner. The *in vivo* errors for the modified settings were ~2% for the scans without motion and ~5-9% for the scans with motion. The other study that used the XCT scanner,[28] reported ~10% and ~5% error for the trabecular and cortical compartments when using the 225 mgHA/cm^3 density threshold, and minimum cluster size of 5 for *in vivo* repeated scans acquired one week apart. While this method was not the selected method in this study, the errors from our XCT same-day *in vivo* repeated scans using the same method that was used by the other two studies were ~6.5% for MV/BV and EV/BV for the scans without motion, while they increased to ~14%

for the repeated scans with motion (all grades of motion combined). The comparison of the results from scans with motion is not feasible due to the different possible combinations and the distribution of motion grades. For the scans without motion, our errors are higher than those of Christen et. al.,[27], which can partly be due to the fact that the repeated scans for this study were acquired during a multicenter clinical trial, with different operators, and different scanners. The errors reported by Mancuso and Troy[28] seem to be between the errors from this study and those of Christen et. al. Regardless of the study, the errors associated with the commonly used method may not be sufficiently low to reliably identify typical biological changes.

The only study on grayscale timelapse HR-pQCT using an XCT2 scanner was performed on metacarpophalangeal joints of individuals with rheumatoid arthritis.[30] They recommended a different density threshold from XCT to account for differences in resolution and anatomical site. They recommended a density threshold of 125 mgHA/cm^3 and a minimum cluster size of 5. While their threshold seems lower than that of previous studies, it is in fact larger than previous studies if converted into values that would have the same definition as previous studies. More specifically, the 125 mgHA/cm^3 threshold includes the intercept in density calibration. Therefore, the equivalent threshold to compare with other studies (including our study) is to subtract the density calibration intercept, which results in a threshold of ~450 mgHA/cm^3 (i.e., 450 = 125 – (-325) assuming an intercept of -325 mgHA/cm^3). To ensure a density-independent threshold, the intercept of the difference image can be set to zero using IPL's "/header_log_set" function. Using this threshold, their reported median errors were ~0% for *in vivo* same day repeated scans. These smaller errors for such a high threshold of ~450 mgHA/cm^3 are expected and in line with our finding of errors ~0%. However, such a high threshold will result in a high rate of false negative, which reduces the utility of the timelapse method, especially for shorter time intervals, or cases where more subtle changes are expected.

Using the commonly used timelapse analysis using grayscale images with no pre-subtraction noise reduction is not ideal since the errors associated with it are relatively high, unless in the case of using high thresholds which diminish the sensitivity of the method. This high threshold in fact exceeds the segmentation threshold of the trabecular bone for XCT2. Further, applying a minimum cluster size is rather subjective, and may result in the exclusion of small cluster of formation and resorption. Accordingly, in this study, we used Gaussian smoothing prior

to grayscale image subtraction for the first time. One benefit of this method was that the outcomes were approximately independent of the minimum cluster size, hence reducing subjectivity (Figure 25). Secondly, by adapting the Gaussian kernel sigma, we obtained similar density thresholds of 200 mgHA/cm^3 for XCT and XCT2, which is desired since the threshold is defined in terms of bone density, which is independent of scanner generation (Figure 25). Of note, the adaptation of the Gaussian sigma is reasonable in the physical sense (distance of 0.8 × 82 μm is comparable to 1.2 × 60.7 μm). Of note, the blurring of the edges is not a concern since the same filter is applied to both images that will be subtracted, hence the blurring effect cancels out. Related to this, we also tested an edge-preserving bilateral denoising filter on a small set of scans, and obtained similar results to the Gaussian filter (*results not shown*). However, a Gaussian filter is a better option since it is more common, easier to implement, and has fewer parameters.

In the current study, we also investigated the effect of image registration method, and the definition of periosteal mask on timelapse HR-pQCT outcomes. For the binary method, the errors of MV/BV and EV/BV were mostly similar but only slightly smaller than those of MA registration (Figure 24). For the grayscale method, when using the dilated periosteal masks, the errors from the two registration methods were almost identical (Figure 26, *left panel*), although they were slightly in favor of 3D registration. When using different periosteal masks, the difference was larger and in favor of 3D registration (Figure 26, *middle panel*). While no HR-pQCT study has previously investigated MA registration for timelapse analysis, a study by de Bakker et. al.[35] developed MA registration for microCT scans of rat tibia, and showed that MA registration reduced the errors of bone resorption volume, while formation volume was not changed. This reduction in bone resorption resulted in reduced bias since 3D registration had higher resorption errors compared to formation. The differences could mainly be due to the differences in scanner resolution and noise level, as well as the implementation of MA registration. The implementation of MA registration for HR-pQCT involves several steps, which could accumulate errors. Moreover, the differences between 3D and MA registration were higher for original masks, indicating that the interpolation error on both masks are more detrimental than error on one mask only. Briefly, during MA registration, the fixed image is transferred to the middle domain. Then, each of the fixed and moving images were registered to the middle domain separately. Of note, a recent study by van Rietbergen et. al.[34] developed and used the same MA registration method

used in this study to measure cortical retraction. However, their method was based on comparing the image masks versus the voxel-by-voxel comparison of density.

We also observed higher errors for MV/BV and EV/BV when using original periosteal masks versus using dilated masks (Figure 26, *right panel*). This indicates that the masks may crop out some of the surface voxels, hence increasing the errors. Another benefit with using dilated masks is to ensure that interpolation error during the rotation of the masks does not increase errors. The use of dilated masks has also been touched upon in the study by Mancuso and Troy,[28] while no formal results were presented.

In addition to the systematic approach for selecting the preferred method by comparing errors from a variety of methods, this study also validated the selected method. We first matched each of the two repeated scans with a scan from the participant at another timepoint. This resulted in 3 groups (1 repeated + 2 longitudinal). We showed that the MV/BV and EV/BV were close to zero for the repeated scans, and were significantly smaller than the longitudinal scans confirming the sensitivity of the method (Figure 27-A). Further, we showed that there is no difference between the two pairs of longitudinal scans (Figure 27-A and B), suggesting that the detected regions of bone formation and resorption are true biological changes, rather than random noise. Further, our data did not provide evidence that the smaller MV/BV and EV/BV were due to smaller rotation angles despite being slightly in favor of this hypothesis (Figure 27-C). In other words, although rotation angle may have partially affected bone formation and resorption computations, it is not a deriving factor. Finally, by quantifying the density-dependent attenuation coefficient drift (Figure 28-A and B), and inducing the drift on the scans, as well as inflated drift of ~16 times the observed drift, we showed that the selected method is only slightly sensitive to the drifts in density, confirming that the drifts between two timepoints are not deriving the identified formation and resorption regions (Figure 28-C).

Using the selected and validated method, we evaluated the effect of different doses of an anabolic drug (setrusumab) on bone formation and resorption of individuals with OI. For the analysis of clinical data, we used two different datasets based on the major timepoints of 12 months (duration of treatment with setrusumab) and 24 months (end of study). For each dataset at each anatomical site, after accounting for other factors including motion, scanner generation, and OI type, we observed an overall positive effect of higher dose on bone formation and resorption. At

the radius from both datasets, the high dose had on average larger MV/BV and EV/BV compared to low dose. These differences were larger for the 6-12 months time period in both datasets, which reached statistical significance prior to p-value correction (Figure 29-A and Figure 30-A).

The net change in volume fraction (ΔV/BV) is computed by subtracting EV/BV from MV/BV. At the radius from the 12-month dataset, we found that ΔV/BV was positive (formation being larger than resorption) for the high dose only (Figure 29-B). These results were consistent with the first 12 months of the 24-month dataset (Figure 30-B). The ΔV/BV also showed a downward trend with time, which is expected after stopping setrusumab treatment. Attention is needed for the interpretation of the outcome between each two timepoints, as they indicate changes between two consecutive timepoints, and not the change from baseline. For instance, if ΔV/BV is 2% and -1% for baseline-6 month and 6 month-12 month time intervals, respectively, it does not mean that at 12 month there is less bone than baseline, rather it means that at 12 month, there is less bone than at 6 months. Similarly, ΔV/BV of ~1% at both baseline-6 months, and 6-12 months intervals indicate a total ΔV/BV of > 2% during a 12 months period (e.g., cortical bone), which is a considerable change at the radius and tibia considering the typically observed changes at these sites.

By decomposing the results into cortical and trabecular compartments, we observed that the two compartments had different patterns for ΔV/BV. While at the radius from both datasets, ΔV/BV tended to be ≥ 0 for the cortical bone, it was mostly < 0 for the trabecular compartment (Figure 29-B and Figure 30-B). These results suggest that the overall positive effect of setrusumab at the distal radius is based on enhancing cortical bone.

At the tibia for the 12-month analysis, the high and medium dose had similar MV/BV average values that were bigger than that for the low dose, regardless of bone compartment, although it was statistically significant for the cortical bone prior to p-value correction (Figure 29-C). Bone resorption was largest for the medium dose, although the difference was not statistically significant (Figure 29-C). The ΔV/BV was positively associated with treatment dose, with largest and statistically significant differences being for the cortical compartment (Figure 29-D). As for the radius, ΔV/BV was positive only for the high dose. For the 24-month dataset, until the 6-12 months time interval, all doses had similar MV/BV and EV/BV, while from 12-24 interval months, the medium dose had similar MV/BV to the other doses, but smaller EV/BV (Figure 30-C),

translating into a positive ΔV/BV (Figure 30-D). This observation was not due to errors in timelapse computations because the same patterns could be observed for the average changes in total BMD, stiffness, and failure load (Supplemental Figure 1). One underlying reason could be the heterogeneous nature of OI. Despite this heterogeneity, we observed patterns of larger timelapse outcomes for OI type III compared to other OI types, which may partly be due to the tendency for OI type III to have lower amount of bone.

The net change in volume fraction (ΔV/BV) can be associated with the changes in bone static measurements such as density and strength. We observed high and significant correlations between ΔV/BV and changes in total BMD, stiffness, and failure loads, which are deemed to have relatively high reliability amongst HR-pQCT measurements.[17,19,42] The ΔV/BV explained ~50% of the total variance in those three measurements (Figure 31). Of note, the R^2 from the regression between changes in total BMD and stiffness, as well as failure load was ~0.34, which is much smaller than 0.5. This suggests that ΔV/BV explains the changes in bone both in terms of changes in bone density and morphological changes. Additionally, we assessed the average value of ΔV/BV against the average value of total BMD, stiffness, and failure load at each timepoint, for each anatomical site, and for each treatment dose. Our qualitative assessment indicated similar average values for ΔV/BV and those outcomes (Supplemental Figure 1). Importantly, it is not possible to acquire reliable and meaningful ΔV/BV values if EV/BV was not reliable. This suggests that EV/BV values were in fact true biological changes. Together, these results support the reliability of timelapse HR-pQCT in monitoring bone changes, with the added benefit of getting more insight into the dynamics and locality of such changes.

This study has several strengths. First, this is the first study to provide a comprehensive comparison of different possible methods for conducting timelapse HR-pQCT in terms of input image type, image registration method, definition of the periosteal mask, as well has density threshold and noise reduction methods. Further, this is the first study to present results from both scanner generations, sexes, and anatomical sites. Therefore, our results would help with standardization of timelapse HR-pQCT analysis. Second, we used Gaussian smoothing for noise reduction as an alternative to the subjective minimum cluster sizes. Gaussian smoothing was also beneficial in reducing the density thresholds relative to the segmentation thresholds. Third, we provided rigorous indirect validations for timelapse HR-pQCT method using repeated and

longitudinal scans, and assessed the effect of potential confounding factors such as image rotation angles, and attenuation coefficient drift. These validations significantly increased the confidence of the findings from the clinical trial data. Lastly, our data were acquired from a multicenter clinical trial, which is considered a more realistic scenario, especially for rare diseases such as OI.

One limitation of this study is the lack of a placebo arm. While the study initially started with a placebo arm, it was later switched to an open label high-dose group because of ethical concerns. A second limitation is the amount of missing data, especially for the 24-month dataset. We decided to analyze the complete-case data instead of imputing the missing data for several reasons. First, the missing timepoints were not associated with the treatment dose, which was the main studied factor in this study. Rather, the exclusions were due to motion artifacts, scanner breakdown, operator error, fracture callus, or rods. Therefore, we do not expect any bias specifically related to the dose-dependent findings. Second, the complexity of the data, the presence of several variables, and the heterogeneity in the data increases the uncertainty associated with data imputation. Considering the complexity of the data, we however acknowledge that the study could benefit from a larger sample size. Finally, the relative amount of missing data was not small, making imputation less effective and uncertain. The third limitation was the unbalanced number of participants for different OI types (i.e., more OI type I compared to type IV and type III). However, this lack of balance is typical in studies on OI which is mainly due to the fact that OI type I is the most common type.

In conclusion, we used a systematic approach to identify the preferred settings to perform time-lapse HR-pQCT among different possible variations. Using same-day *in vivo* repeated scans, we indicated that using 3D-registered grayscale input images that are segmented using dilated periosteal masks, along with Gaussian smoothing for noise reduction, and a density threshold of 200 mgHA/cm^3 produce small errors. We then used a rigorous approach to validate the selected method by showing that the errors from repeated scans were significantly smaller than the calculated formation and resorption from matching each of the repeated scans with a scan from another timepoint. We also showed no difference between the results of each of the matched scan pairs, confirming that the computed bone changes were in fact biological changes. We further showed small effect of confounding factors such as registration rotation angle and x-ray attenuation coefficient drift on the results of the selected method. We finally used the selected and

validated method and showed a positive dose-dependent effect of setrusumab on bone formation and resorption, as well improved net changes in bone at the radius and tibia of adults with OI.

Conflict of Interest Statement:

This research was supported by Mereo Biopharma and Shriners Hospitals for Children. BMW and FR have received institutional research support and materials and are consultants for Mereo BioPharma.

Author Contributions:

Study conception and design: SH, BMW; Methodology and software: SH; Data acquisition: SH, MR, BMW; Analysis and interpretation of data: SH, IV, MR, AB, FR, BMW; Drafting of Manuscript: SH, BMW. All authors contributed to the critical revision and approval of the final manuscript.

Acknowledgements:

We thank all the study participants and staff at the imaging sites. The participating imaging sites are: Klinik for Knogleskørhed, Aarhus Universitets hospital; Children's Nutrition Research Center, Baylor College of Medicine; Cambridge Clinical Research Centre, Cambridge University Hospitals NHS Foundation Trust–Addenbrooke's Hospital; Hôpital Edouard Herriot; Rhumatologie Centre, Hôpital Lariboisière; Institute for Clinical and Translational Research, Johns Hopkins University; Oregon Health and Science University; Endokrinologisk Afdeling, Odense Universitetshospital; Shriners Hospitals for Children-Canada; Osteoporosis Centre, Southampton General Hospital NHS Foundation Trust; Osteoporosis Clinic, Toronto General Hospital; Institute of Imaging Science, Vanderbilt University; and the Center for Advanced Medicine, Division of Bone and Mineral Diseases, Washington University School of Medicine in St Louis. We also thank Dr. Elizabeth Zimmermann and Catherine Julien for assisting with data collection. We thank members of Ultragenyx and Mereo BioPharma for assistance with some data organization. The study was funded by Mereo BioPharma, Shriners Hospitals for Children, and FRQS Programme de bourses de chercheur (BMW, SH, MR).

References:

1. Marini JC. Osteogenesis imperfecta. Prim. Metab. Bone Dis. Disord. Miner. Metab. 2018.
2. Oakley I, Reece LP. Anesthetic implications for the patient with osteogenesis imperfecta. AANA J. 2010.
3. Tauer JT, Robinson M-E, Rauch F. Osteogenesis Imperfecta: New Perspectives From Clinical and Translational Research. JBMR Plus. 2019;3(8):e10174.
4. Mortier GR, Cohn DH, Cormier-Daire V, Hall C, Krakow D, Mundlos S, Nishimura G, Robertson S, Sangiorgi L, Savarirayan R, Sillence D, Superti-Furga A, Unger S, Warman

ML. Nosology and classification of genetic skeletal disorders: 2019 revision. Am. J. Med. Genet. Part A. 2019;179(12):2393–419.
5. Marom R, Lee YC, Grafe I, Lee B. Pharmacological and biological therapeutic strategies for osteogenesis imperfecta. Am. J. Med. Genet. Part C Semin. Med. Genet. 2016;172(4):367–83.
6. Nijhuis WH, Weinans HH, Bank RA, Hvid I, Eastwood DM, Allgrove J, Sakkers RJ. Current concepts in osteogenesis imperfecta: bone structure, biomechanics and medical management. J. Child. Orthop. 2019;13(1):1–11.
7. Sinder BP, White LE, Salemi JD, Ominsky MS, Caird MS, Marini JC, Kozloff KM. Adult Brtl/+ mouse model of osteogenesis imperfecta demonstrates anabolic response to sclerostin antibody treatment with increased bone mass and strength. Osteoporos. Int. 2014;25(8):2097–107.
8. Sinder BP, Salemi JD, Ominsky MS, Caird MS, Marini JC, Kozloff KM. Rapidly growing Brtl/+ mouse model of osteogenesis imperfecta improves bone mass and strength with sclerostin antibody treatment. Bone [Internet]. Elsevier Inc.; 2015;71:115–23. Available from: http://dx.doi.org/10.1016/j.bone.2014.10.012
9. Sinder BP, Eddy MM, Ominsky MS, Caird MS, Marini JC, Kozloff KM. Sclerostin antibody improves skeletal parameters in a Brtl/+ mouse model of osteogenesis imperfecta. J. Bone Miner. Res. 2013;
10. Grafe I, Alexander S, Yang T, Lietman C, Homan EP, Munivez E, Chen Y, Jiang MM, Bertin T, Dawson B, Asuncion F, Ke HZ, Ominsky MS, Lee B. Sclerostin Antibody Treatment Improves the Bone Phenotype of Crtap -/- Mice, a Model of Recessive Osteogenesis Imperfecta. J. Bone Miner. Res. 2016;
11. Roschger A, Roschger P, Keplingter P, Klaushofer K, Abdullah S, Kneissel M, Rauch F. Effect of sclerostin antibody treatment in a mouse model of severe osteogenesis imperfecta. Bone. 2014;
12. Lee B, Smith R, Orwoll E, Wang Y, Shapiro J, Nicol L, Sloan J, Nagamani SC. Serum Sclerostin Levels in Adults With Osteogenesis Imperfecta: Comparison With Normal Individuals and Response to Teriparatide Therapy. J. Bone Miner. Res. 2017;33(2):307–15.
13. Glorieux FH, Devogelaer JP, Durigova M, Goemaere S, Hemsley S, Jakob F, Junker U, Ruckle J, Seefried L, Winkle PJ. BPS804 Anti-Sclerostin Antibody in Adults With Moderate Osteogenesis Imperfecta: Results of a Randomized Phase 2a Trial. J. Bone Miner. Res. 2017;32(7):1496–504.
14. Javaid MK, Saag K, Lewiecki M, Civitelli R, Cohen-Solal M, Willie B, Hosseinitabatabaei S Zimmermann E, Clancy J, MacKin A, Black I MA. Baseline patient demographics and disease characteristics in adults with Osteogenesis Imperfecta (OI) in the Phase 2b ASTEROID Study. Am. Soc. Bone Miner. Res. Annu. Meet. San Diego, CA: American Society for Bone and Mineral Research; 2021.
15. Glorieux FH, Javaid MK, Langdahl B, Chapurlat R, Jan De Beur S, Reid Sutton V, Poole K, Orwoll ES, Willie B, Mikolajewicz N, Clancy J, MacKin A, Mistry A, Ominsky M SC. setrusumab for the Treatment of Osteogenesis Imperfecta (OI): Results from the Phase 2b ASTEROID Study. Am. Soc. Bone Miner. Res. Annu. Meet. San Diego, CA: American Society for Bone and Mineral Research; 2021.
16. Javaid MK, Langdahl B, Chapurlat R, Jan De Beur S, Reid Sutton V, Poole K, Dahir K,

Orwoll ES, Willie B, Mikolajewicz N, Zimmermann E, Ominsky M, Saville C, Clancy J, MacKin A GF. Changes in BMD and Bone Turnover Markers Following Discontinuation of setrusumab After 1 Year of Treatment in Adults with Osteogenesis Imperfecta. Am. Soc. Bone Miner. Res. Annu. Meet. Austin, TX: American Society for Bone and Mineral Research; 2022.
17. Whittier DE, Boyd SK, Burghardt AJ, Paccou J, Ghasem-Zadeh A, Chapurlat R, Engelke K, Bouxsein ML. Guidelines for the assessment of bone density and microarchitecture in vivo using high-resolution peripheral quantitative computed tomography. Osteoporos. Int. 2020;
18. 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.
19. Hosseinitabatabaei S, Mikolajewicz N, Zimmermann EA, Rummler M, Steyn B, Julien C, Rauch F, Willie BM. 3D image registration marginally improves the precision of HR-pQCT measurements compared to cross-sectional-area registration in adults with osteogenesis imperfecta . J. Bone Miner. Res. 2022;00(00):1–17.
20. Christen P, Müller R. In vivo Visualisation and Quantification of Bone Resorption and Bone Formation from Time-Lapse Imaging. Curr. Osteoporos. Rep. Current Osteoporosis Reports; 2017;15(4):311–7.
21. Christen P, Ito K, Ellouz R, Boutroy S, Sornay-Rendu E, Chapurlat RD, Van Rietbergen B. Bone remodelling in humans is load-driven but not lazy. Nat. Commun. [Internet]. Nature Publishing Group; 2014;5:1–5. Available from: http://dx.doi.org/10.1038/ncomms5855
22. Checa S, Birkhold AI, Weinkamer R, Zaslansky P, Willie BM, Razi H, Duda GN. Skeletal maturity leads to a reduction in the strain magnitudes induced within the bone: A murine tibia study. Acta Biomater. [Internet]. Acta Materialia Inc.; 2014;13:301–10. Available from: http://dx.doi.org/10.1016/j.actbio.2014.11.021
23. Birkhold AI, Razi H, Duda GN, Weinkamer R, Checa S, Willie BM. Mineralizing surface is the main target of mechanical stimulation independent of age: 3D dynamic in vivo morphometry. Bone [Internet]. Elsevier Inc.; 2014;66:15–25. Available from: http://dx.doi.org/10.1016/j.bone.2014.05.013
24. Schulte FA, Lambers FM, Kuhn G, Müller R. In vivo micro-computed tomography allows direct three-dimensional quantification of both bone formation and bone resorption parameters using time-lapsed imaging. Bone [Internet]. Elsevier Inc.; 2011;48(3):433–42. Available from: http://dx.doi.org/10.1016/j.bone.2010.10.007
25. Waarsing JH, Day JS, Linden JC Van Der, Ederveen AG, Spanjers C, Clerck N De, Sasov A, Verhaar JAN, Weinans H. Detecting and tracking local changes in the tibiae of individual rats : a novel method to analyse longitudinal in vivo micro-CT data. 2004;34:163–9.
26. Birkhold AI, Razi H, Duda GN, Weinkamer R, Checa S, Willie BM. The influence of age on adaptive bone formation and bone resorption. Biomaterials [Internet]. Elsevier Ltd; 2014;35(34):9290–301. Available from: http://dx.doi.org/10.1016/j.biomaterials.2014.07.051
27. Christen P, Boutroy S, Ellouz R, Chapurlat R, Van Rietbergen B. Least-detectable and age-related local in vivo bone remodelling assessed by time-lapse HR-pQCT. PLoS One. 2018;13(1):1–11.

28. Mancuso ME, Troy KL. Relating Bone Strain to Local Changes in Radius Microstructure following 12 Months of Axial Forearm Loading in Women. J. Biomech. Eng. 2020;142(11):1–11.
29. Zhang J, Wang Y, Cheng K, Cheuk K, Lam T, Hung ALH, Cheng JCY, Qiu Y, Müller R, Christen P, Lee WYW. Association of higher bone turnover with risk of curve progression in adolescent idiopathic scoliosis. Bone [Internet]. Elsevier; 2020;(September):115655. Available from: https://doi.org/10.1016/j.bone.2020.115655
30. Brunet SC, Kuczynski MT, Bhatla JL, Lemay S, Pauchard Y, Salat P, Barnabe C, Manske SL. The utility of multi-stack alignment and 3D longitudinal image registration to assess bone remodeling in rheumatoid arthritis patients from second generation HR-pQCT scans. BMC Med. Imaging. BMC Medical Imaging; 2020;20(1):1–10.
31. Du J, Hartley C, Brooke-Wavell K, Paggiosi MA, Walsh JS, Li S, Silberschmidt V V. High-impact exercise stimulated localised adaptation of microarchitecture across distal tibia in postmenopausal women. Osteoporos. Int. 2020;
32. Atkins PR, Stock K, Ohs N, Collins CJ, Horling L, Benedikt S, Degenhart G, Lippuner K, Blauth M, Christen P, Müller R. Formation Dominates Resorption With Increasing Mineralized Density and Time Postfracture in Cortical but Not Trabecular Bone: A Longitudinal HRpQCT Imaging Study in the Distal Radius. JBMR Plus. 2021;5(6):1–17.
33. Collins CJ, Atkins PR, Ohs N, Blauth M, Lippuner K, Müller R. Clinical observation of diminished bone quality and quantity through longitudinal HR-pQCT-derived remodeling and mechanoregulation. Sci. Rep. [Internet]. Nature Publishing Group UK; 2022;12(1):1–13. Available from: https://doi.org/10.1038/s41598-022-22678-z
34. van Rietbergen B, Biver E, Chevalley T, Ito K, Chapurlat R, Ferrari S. A Novel HR-pQCT Image Registration Approach Reveals Sex-Specific Changes in Cortical Bone Retraction With Aging. J. Bone Miner. Res. 2021;00(00):1–13.
35. de Bakker CMJ, Altman AR, Li C, Tribble MB, Lott C, Tseng WJ, Liu XS. Minimizing Interpolation Bias and Precision Error in In Vivo μCT-Based Measurements of Bone Structure and Dynamics. Ann. Biomed. Eng. 2016;44(8):2518–28.
36. Borgen TT, Solberg LB, Lauritzen T, Apalset EM, Bjørnerem Å, Eriksen EF. Target Values and Daytime Variation of Bone Turnover Markers in Monitoring Osteoporosis Treatment After Fractures. JBMR Plus. 2022;6(6):1–9.
37. Compston JE, Vedi S, Stellon AJ. Inter-observer and intra-observer variation in bone histomorphometry. Calcif. Tissue Int. 1986;
38. Bonaretti S, Majumdar S, Lang TF, Khosla S, Burghardt AJ. The comparability of HR-pQCT bone measurements is improved by scanning anatomically standardized regions. Osteoporos. Int. Osteoporosis International; 2017;1–14.
39. Mikolajewicz N, Zimmermann EA, Rummler M, Hosseinitabatabaei S, Julien C, Glorieux FH, Rauch F, Willie BM. Multisite longitudinal calibration of HR-pQCT scanners and precision in osteogenesis imperfecta. Bone. 2021;147(February):115880.
40. Pialat JB, Burghardt AJ, Sode M, Link TM, Majumdar S. Visual grading of motion induced image degradation in high resolution peripheral computed tomography: Impact of image quality on measures of bone density and micro-architecture. Bone. 2012;50(1):111–8.
41. Boyd STKB, Oser STM, Uhn MIK, Linck ROJK, Rauze PELK. Evaluation of Three-dimensional Image Registration Methodologies for In Vivo Micro-computed

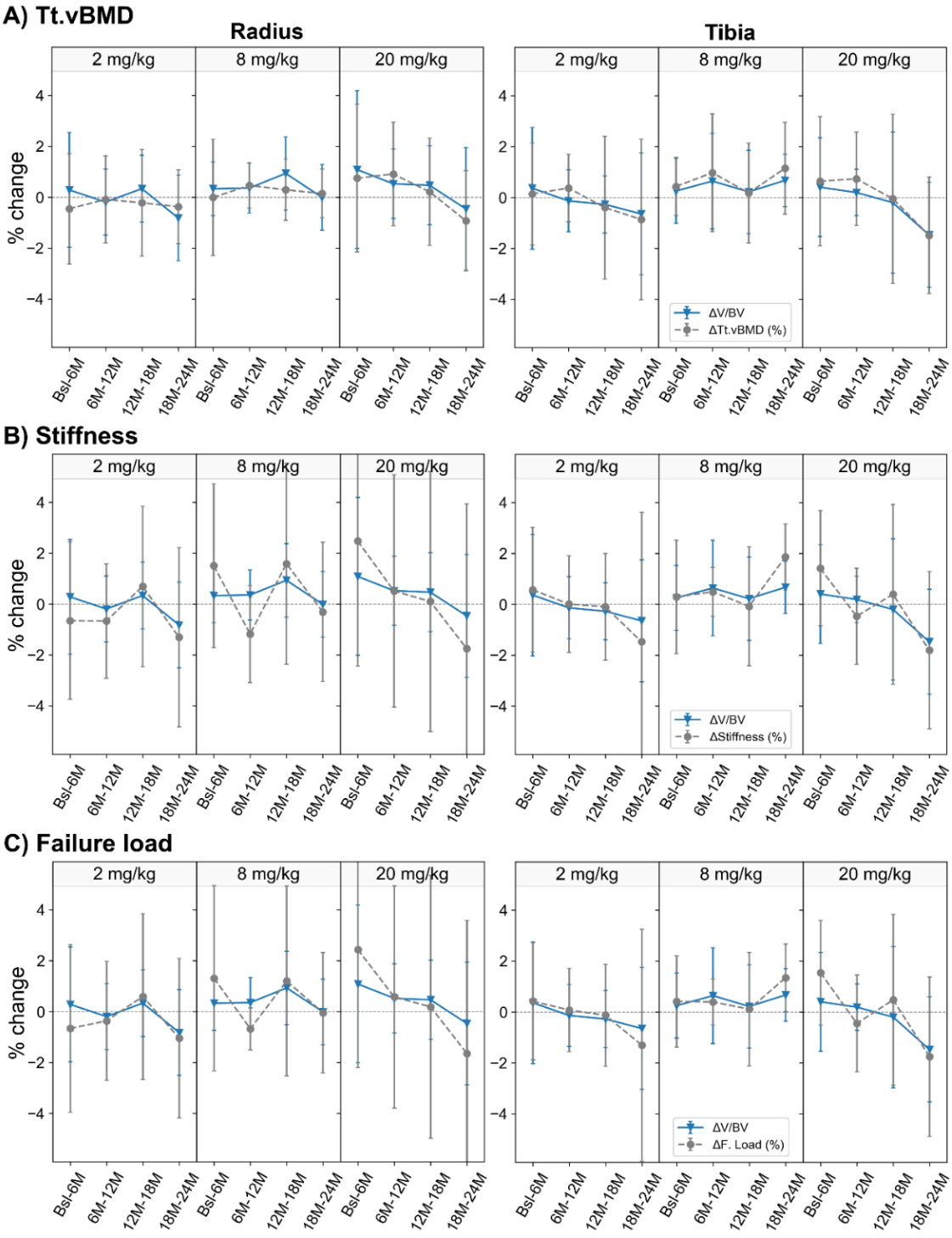
Tomography. 2006;34(10):1587–99.
42. Mikolajewicz N, Bishop N, Burghardt AJ, Folkestad L, Hall A, Kozloff KM, Lukey PT, Molloy-Bland M, Morin SN, Offiah AC, Shapiro J, van Rietbergen B, Wager K, Willie BM, Komarova S V, Glorieux FH. HR-pQCT Measures of Bone Microarchitecture Predict Fracture: Systematic Review and Meta-analysis. J. Bone Miner. Res. 2019;

Supplemental Methods:

Inclusion/exclusion criteria: Inclusion criteria included male and female participants with a clinical diagnosis of OI Type I, III or IV with a *COL1A1*/*COL1A2* defect confirmed by genetic testing, age greater than 18 years, and one or more non-traumatic long bone, rib, hand/feet and/or vertebral fracture(s) in the past five years. Female participants were ineligible if pregnant, breastfeeding, or following contraceptive guidance.

Participants were ineligible if they were greater than 75 years of age, or had a history of the following: skeletal malignancies/bone metastases; neural foraminal stenosis; uncontrolled diseases affecting bone metabolism; skeletal conditions leading to long bone deformities or increased fracture risk other than OI; bisphosphonate treatment 3 months prior to baseline; teriparatide, denosumab or other anabolic or anti-resorptive medication within 6 months prior to baseline; myocardial infarction, agina pectoris, ischaemic stroke, or transient ischaemic attack; alcohol or drug abuse in 12 months prior to dosing; significant psychiatric or medical disorder affecting compliance to study protocol; history of external radiation; participation in any clinical investigation within 4 weeks or 5 half-lives of the drug prior to dosing; or allergy to the study drug.

Supplemental Results:



Supplemental Figure 1. Similar outcomes from timelapse analysis and density and micro-finite-element analysis. Average net change in volume fraction (ΔV/BV) [*blue solid line*] plotted

along with percentage changes in **A)** Tt.vBMD; **B)** stiffness; and **C)** failure load [*gray dashed line*] at different timepoints, separated by site (*radius: left column; tibia: right column*) and treatment dose.

Connection between chapters:

In Chapters 4 and 5, I targeted two challenges facing longitudinal HR-pQCT analysis in the context of a multicenter clinical trial on adults with OI. These two challenges were: 1) identifying the best image registration method for improving the reliability of HR-pQCT-based measurement of changes in average bone properties such as density, morphology, and strength; and 2) identifying the best image registration method combined with other parameters that would improve the reliability of HR-pQCT-derived measurements of bone formation and resorption as local bone changes.

While plenty of HR-pQCT data is available for adults, such data is limited for children. In particular, for some population groups, such as children with OI, such data is rare, and only one study has reported HR-pQCT data in children with OI. In line with the limited data, the understanding of the challenges specific to longitudinal imaging in children is also limited. For instance, due to growth, image registration is complex in children, and the common rigid registration may not be able to properly register images. However, the limitations of image registration in children are not well discussed in the literature.

An important step in improving the reliability of pediatric HR-pQCT studies is to produce more data since it facilitates the creation of normative data that would result in more meaningful interpretation of clinical studies.

The previous two chapters covered methods for improving the reliability of longitudinal HR-pQCT studies in adults by reducing repositioning error (Chapter 4) and providing settings for timelapse analysis that would results in lowest errors while being sensitive (Chapter 5). The next chapter (Chapter 6) will provide the first set of 1-year longitudinal data from children with OI and age- and sex-matched healthy controls. Further, image registration will be examined to illustrate the challenges. Finally, scanning additional regions of bone are discussed along with the challenges, and relevant solutions.

Chapter 6. Studying the natural history of the peripheral bones of children with OI and healthy controls using longitudinal HR-pQCT analysis:

Status: *In preparation*

Publication title:

Natural history of the peripheral bones in children with osteogenesis imperfecta and age- and sex-matched healthy controls using longitudinal HR-pQCT analysis

Seyedmahdi Hosseinitabatabaei[a,b,c], Samantha McCluskey[a,b,c], Elizabeth A. Zimmermann[a,c], Francis H. Glorieux[a], Fredrick Charbonneau[a], Frank Rauch[a], Bettina M. Willie[a,b,c,†]

[a]*Research Centre, Shriners Hospital for Children-Canada, Montreal, Canada*
[b]*Department of Biomedical Engineering, McGill University, Montreal, Canada*
[c]*Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Canada*

Abstract:

Osteogenesis imperfecta (OI) is one of the most common bone dysplasias, characterized by increased bone fragility with frequent fractures, especially in children. Although some studies have used peripheral-quantitative computed tomography (pQCT) to examine bone density in children with OI, less is known concerning bone microstructure and strength due to limitations with that imaging modality and few longitudinal studies have been performed. High-resolution pQCT (HR-pQCT) is a non-invasive imaging tool for measuring bone density, microstructure, and strength at the distal peripheral skeleton. So far, only one study has reported HR-pQCT data for children with OI, but they provided no longitudinal data. Furthermore, their measurements were limited to a single image stack at the metaphysis, although long bone fractures in children with OI most often occur in the diaphysis.

We compared HR-pQCT measurements and their changes over 1 year between 7 children with OI and 7 age- and sex-matched healthy controls using longitudinal double-stack scans at the metaphysis, and single stack scans at the diaphysis. Scans were acquired using a second-generation scanner (XtremeCTII, Scanco Medical AG, Bruettisellen, Switzerland). Our 1-year data showed that deteriorations in the OI group in trabecular microstructural measurements were more prominent at the tibia compared to the radius. The changes over the ~1-year period were mostly comparable between the OI and control groups at both the radius and tibia. At the diaphysis, we found similar results at the radius and tibia, where differences between the two groups were mainly driven by area, especially larger marrow cavity area in the control group.

Additionally, we used the longitudinal scans to investigate the feasibility of image registration in growing children using the commonly used rigid registration. We showed that longitudinal image registration in growing bones using the available rigid registration algorithm is not feasible. Finally, we investigated the benefits and challenges related to acquiring double-stack scans. We observed misalignment between the stacks of our double-stack scans caused by subject movement. While we could minimize the misalignment, we could not eliminate it. However, such misalignment does not affect density and morphological outcomes, while micro-finite-element (FE) needs to be done on stacks separately.

6.1. Introduction:

High-resolution peripheral-quantitative computed tomography (HR-pQCT) is a promising non-invasive imaging tool for measuring bone density, microstructure, and strength at the distal peripheral skeleton.[1,2]

While plenty of HR-pQCT data is available for the adult population, such data is limited for children, particularly longitudinal HR-pQCT data. Further, limited HR-pQCT data is available for some disease populations, one of which is osteogenesis imperfecta (OI). OI, or brittle bone disease, is a collagen-related genetic disorder resulting in bone fragility.[3] The reduced bone strength in OI is complex, and requires understanding of OI bone properties at different length scales. However, only four studies have reported bone geometry, density, and microarchitecture in the distal radius and tibia in adults with OI using HR-pQCT,[4–7] aside from a case-report.[8] Two studies have also reported the short-term precision of HR-pQCT measurements in adults with OI, using the same dataset.[9,10] In contrast, only one study has reported HR-pQCT data for children with OI.[11] Their participants consisted of 9 children with mild (*N=7*) and severe (*N=2*) OI, aged from 9-15 years old and age- and sex-matched healthy controls. The goal of this cross-sectional study was to investigate the feasibility of HR-pQCT imaging for children with OI considering the fragility, length and deformity of their bones. The scans were acquired with an XCT scanner using the standard imaging and analysis protocols. It is known that trabecular microstructural outcomes can be measured more accurately and independently of density using the second generation scanner.[12,13] Further, this study provided no longitudinal information on HR-pQCT imaging of pediatric OI population. Thus it is clear that additional studies using HR-pQCT to examine bone microstructure in children with OI are warranted.

Many techniques used in adults to improve reliability, such as image registration, may not be readily applicable to children due to bone growth. Combining these two factors, longitudinal HR-pQCT studies in children are poorly standardized, and the challenges and opportunities are not well understood. An important consideration with HR-pQCT is that it has been limited to the bone metaphysis. Using peripheral quantitative computed tomography (pQCT), several studies have shown that the diaphysis of long bones from children and adults with OI have a smaller periosteal circumference compared to healthy controls.[14,15] This is of particular importance in terms of diaphyseal fractures, which are common in children with OI. While pQCT has provided valuable information on children with OI, it is limited because of low resolution that cannot resolve bone

microstructure (200-400 µm vs 61-82 µm for HR-pQCT). Further, the typical pQCT scanning protocol only captures one slice that averages bone properties over a single 2 mm slice (compared to a length of ~10 mm samples at 61-82 µm for HR-pQCT). Considering the mentioned limitations, pQCT is also not suited for finite-element (FE) analysis, while HR-pQCT works seamlessly with microFE analysis that takes into account bone microstructure. Thus, HR-pQCT can provide us with the opportunity to examine both cortical and trabecular bone microstrucuture, density, and strength at the long bone metaphysis as well as cortical bone at the diaphysis.

To address the discussed gaps, we performed longitudinal scans at the metaphysis and diaphysis, to perform a more comprehensive comparison between children with OI and age- and sex-matched healthy controls. We also investigated the feasibility of using image registration for growing children to clarify its limitations and potentials,. To our knowledge, this is the first longitudinal HR-pQCT dataset for children with OI.

6.2. Methods:

6.2.1. Participants:

As part of an ongoing study involving 20 children with osteogenesis imperfecta (OI) and 20 age- and sex-matched healthy controls. At this point, 7 children with OI and 7 healthy controls have been fully analyzed and will be discussed in this thesis. Children with OI were recruited during one of their routine clinic visits. For each recruited participant with OI, an age- and sex-matched healthy control was recruited. The inclusion criteria were males of females with age of 5 to 18 years old. OI participants had to be clinically diagnosed with either Type I, III or IV and a confirmed COL1A1/COL1A2 mutation. The OI participants were treated with long-acting intravenous bisphosphonates (e.g., zoledronate or pamidronate) ongoing for at least two years prior to enrolment and were expected to continue bisphosphonate treatment unchanged for the duration of the study. The exclusion criteria were confounding skeletal or other medical conditions that might interfere with the objectives of the study (as determined by the investigator). Further, participants with prescription medications that might affect growth or bone mineral accrual within the last 6 months prior to baseline. (e.g., teriparatide, denosumab) were excluded. The baseline characteristics of the participants can be found in Table 1. The study was approved by the Institutional Review Board overseeing clinical research at McGill University and Shriners Hospital for Children.

Table 3. Descriptive statistics of the participants' characteristics at baseline stratified by OI status and type. Differences between characteristic of participants with OI and their corresponding controls were tested using paired t-test following non-significant Shapiro-Wilk test results. No difference was observed between the OI-All and Control groups. No test was done for OI-I and OI-IV groups separately

		OI-All *(N=7)*	**OI-I** *(N=4)*	**OI-IV** *(N=3)*	**Control** *(N=7)*
Variable	**Statistic**	**Value**	**Value**	**Value**	**Value**
Age (years)	Mean (SD)	11.3 (3.0)	11.7 (3.4)	10.7 (2.8)	11.6 (3.0)
	min < med < max	7.2 < 11.1 < 15.1	7.2 < 12.3 < 15.1	8.5 < 9.8 < 13.9	7.7 < 11.2 < 15.6
Weight (kg)	Mean (SD)	40.7 (13.9)	44.7 (11.3)	35.3 (17.8)	41.7 (17.4)
	min < med < max	24.2 <35.6< 55.9	34.2 <44.8< 55.0	24.2 <25.8< 55.9	23.6 <41.8< 65.0
Height (cm)	Mean (SD)	139.2 (14.5)	146.1 (11.6)	129.9 (14.1)	145.4 (17.3)
	min < med < max	117 < 145 < 154	129 < 151 < 154	117 < 128 < 145	128 < 139 < 172
Ulna length	Mean (SD)	226.4 (34.2)	243.7 (23.2)	203.3 (36.2)	225.4 (33.4)
(mm)	min < med < max	180 < 235 < 265	215 < 247 < 265	180 < 185 < 245	185 < 230 < 280
Tibia length	Mean (SD)	312.8 (47.6)	337.5 (33.0)	280.0 (48.2)	325.3 (47.9)
(mm)	min <med <max	245 < 335 < 360	290 < 350 < 360	245 < 260 < 335	270 < 310 < 405
Feature	**Categories**	**N**	**N**	**N**	**N**
Sex	Female	1	0	1	1
	Male	6	4	2	6

**There were no differences between baseline characteristics of participants with OI and controls from paired t-test*

6.2.2. HR-pQCT imaging:

Using a second-generation high resolution peripheral quantitative computed tomography (HR-pQCT) (XtremeCTII, Scanco Medical AG, Bruettisellen, Switzerland), we scanned the metaphysis and diaphysis of the radius and tibia at baseline and 1-year followup. While 5 out of 7 participants with OI were scanned between ~11.5 to ~12.7 months, one participant was scanned after ~14 months and another one after ~19 months due to availability issues. The scans were acquired from the nondominant arm and corresponding leg except in case of a recent fracture or metal rod. The reference line location was based on the status of the growth plate Figure 32. In the case of an open growth plate, the reference line was placed at the most distal margin of the distal growth plate for both the radius and tibia. In case of fused growth plate, the reference line was placed at the medial proximal margin of the radial articular surface and at the tibial plateau. The top slice of the scanned region at the metaphysis and diaphysis was at the 4% and 30% of ulna or tibia length from the reference line, respectively. The metaphyseal scans contained 336 slices (i.e., two consecutive stacks of 168 slices each referred to as double stack) at 60.7 μm isotropic voxel size, while the diaphyseal scans contained 168 slices (i.e., single stack). There was no overlap between the stacks of the double stack scans. For each scan, the scout view scan was graded for motion, and repeat scans (up to a total of 3 scans) were performed if motion grade was 4 or higher.[16]

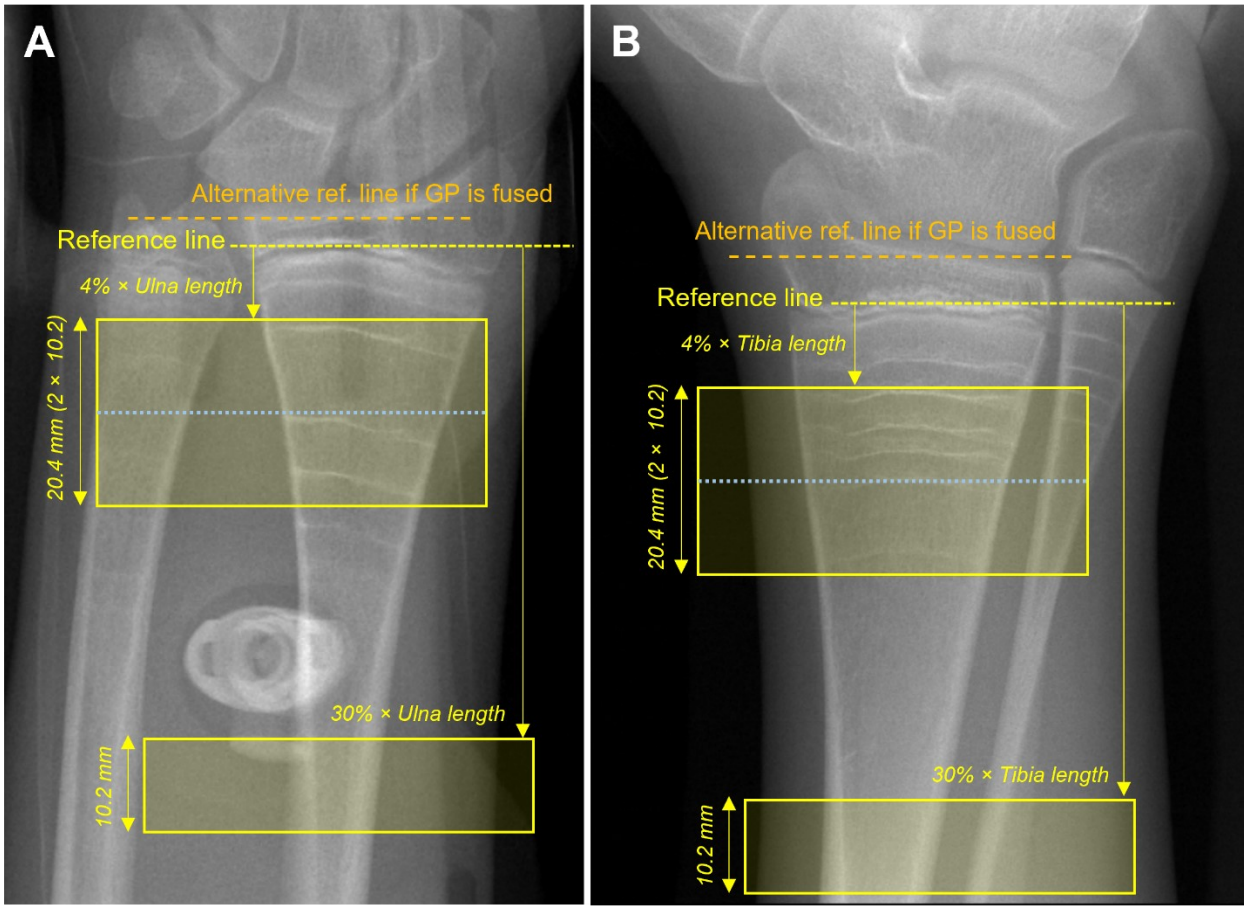


Figure 32. Scout views illustrating the positioning of the reference line and scanned regions: (A) At the radius metaphysis, two adjacent stacks of 168 slices each (10.2 mm) are scanned with the top slice located 4% of ulna length away from the reference line. At the radius diaphysis, one stack of 168 slices is scanned with the top slice located at 30% of ulna length from the reference line. In case of open growth plate, the reference line is located at the distal margin of growth plate, while for cases of fused growth plate, the reference line is located at the medial proximal margin of radial endplate. (B) At the tibia, similar regions as radius are scanned, except that their distance to the reference line is relative to tibia length. In case of open growth plate, the reference line is located at the distal margin of growth plate, while for cases of fused growth plate, the reference line is located at the tibial plateau

6.2.3. HR-pQCT Image analysis:

For all images, we used the manufacturer's standard *in vivo* protocol to first automatically contour the periosteal surface of the larger bone, followed by manual correction. To separate bone from soft-tissue, we applied the manufacturer's low-pass Gaussian filter (sigma 0.8, support 1.0) and a fixed dual threshold to extract the trabecular (320 mg HA/cm^3) and cortical (450 mg HA/cm^3) bones. At the metaphysis, we report total (Tt.), trabecular (Tb.), and cortical (Ct.) volumetric bone

mineral density (vBMD), while for the diaphysis, no trabecular outcome was reported except marrow cavity area (Mr.Ar (mm^2), equivalent to trabecular area). At the metaphysis, we assessed microarchitecture directly using voxel-based measurements based on the distance transformation[17] to calculate trabecular thickness (Tb.Th; mm), number (Tb.N; mm−1) and separation (Tb.Sp; mm), as well as inhomogeneity of the trabecular network (Tb.1/N.SD; mm) and area (Tb.Ar; mm^2). Trabecular bone volume fraction (BV/TV; %) was computed as the ratio of voxels in the mineralized bone phase to the total number of voxels in the trabecular region. For cortical bone at both regions, cortical thickness (Ct.Th; mm), porosity (Ct.Po; %), and area (Ct.Ar; mm^2) were reported.

6.2.4. microFE analysis:

We performed microFE analysis to estimate bone stiffness, failure load, and apparent modulus. The analysis simulated axial compression (IPLFE v01.16, Scanco Medical AG).[2] A linear elastic modulus of E = 8748 MPa was applied and a Poisson's ratio was set to 0.3. A Pistoia's failure criteria with critical volume of 2% and critical value of 7000 μstrain was used.[18,19]

6.2.5. Feasibility of 3D image registration:

During growth, bone structure morphs into a different shape, at both the macro and micro scales. At the macro level, due to the increase in bone length (i.e., moving target), the location of an equivalent anatomical site relative to the reference line is not constant. It is important to note that an equivalent anatomical site does not mean the same bone material (e.g., not the same trabeculae) because the distance between the existing bone and growth plate increases over time (i.e., existing bone in metaphysis ends up in bone shaft). This moving bone material also changes shape. Thus overall, changes in growing bone are the sum of "moving target" and "moving and morphing material". At the same time, errors in limb repositioning occur in growing bone, similar to mature bone. This multi-component change in bone makes image registration difficult. While the nature of the bone changes is not rigid, we investigated whether the available rigid 3D registration algorithm can adequately identify alignment based on trabecular bone features.

The image registration using Scan sooftware IPL involved the computation of the 4 × 4 transformation matrix that aligned the followup image (moving image) with the baseline image (reference image). We applied an initial axial translation to the followup image at the beginning of image registration. The initial translation was ~10 mm, or based on the estimated distance from

matching sclerotic lines, which are horizontal trabeculae containing some degree of cartilage thought to be caused by the temporary interruption of growth plate cartilage resorption, caused by bisphosphonate treatment. We also used initial rotation along with a cross-correlation similarity metric, and the downhill simplex optimization scheme. To reduce the effect of noise on registration, only the volume within the periosteal masks were registered. To avoid registration errors (e.g., local minima) and to reduce computation time, the registration was performed in 3 stages (downsampling of 10, 4, and 1). Nearest neighbor and cubic interpolations were used for the binary and grayscale images, respectively.[20]

6.2.6. Feasibility of stack alignment:

Scanning two image stacks is useful when scanning children over time, as their long bone length typically grows ~1cm/year, which is similar to the height of a single stack. However, scanning two stacks requires at two separate instances of gantry rotation. Therefore, there is a chance of slight misalignment between the consecutive image stacks. We scanned double stacks on a human cadaveric radius phantom (Relaxed hand sectional phantom [XA231R], The Phantom Laboratory, Salem, NY, USA) to identify if the misalignment is scanner or subject (i.e., motion) related. For cases with misalignment, we used image registration to align the stacks. To do this we first separated the two grayscale stacks. Next, we took the bottom-most slice of the top stack, and top-most slice of the bottom stack to use for registration. These two slices are the adjacent slices between the two stacks. To enable the initialization and prevent the failure of 3D registration, the two adjacent slices were duplicated to create an overlap between them, and to create larger volumes. Then, 3D registration was performed while suppressing translation along the longitudinal axis (i.e., assuming no physical gap nor overlap between the two adjacent slices), and only rotation around the transverse axes (i.e., x and y axes). Using the computed transformation matrix, the corresponding images of the bottom stack were then transformed using cubic and nearest neighbor resampling methods for the grayscale and binary components, respectively.

6.2.7. Statistical analysis:

The outcome of 3D image registration on longitudinal scans, as well as stack alignment were assessed qualitatively by visual inspection. Hence, no statistical analysis was performed.

To compare baseline characteristics of OI and control participants (age, weight, height, ulna length, and tibia length), we first checked the normality of the differences between the paired

measures of each group using Shapiro-Wilk test, followed by paired t-test or Wilcoxon-signed-rank-test for normal and non-normal distributions, respectively.

For each HR-pQCT and microFE outcome, changes over time were modelled linearly (linearity enforced by having only two timepoints) using a mixed effect model with random intercept (random slope was not included due to having only two timepoints per participant) and a variance-covariance model that incorporated correlations for the observations from the same participant. The restricted maximum likelihood estimation (REML) with a Kenward-Roger correction for small samples were used to minimize small sample bias in estimation and to prevent inflation of Type-I error rates.[21] The OI status (OI vs control), time, and their interaction were included as fixed effects. Time in months was included as a continuous variable to account for differences in time interval between the baseline and follow up scans among participants. Other factors were not included in the model due to the low sample size and model parsimony. Due to the low sample size, no statistical sub-analysis was performed to compare OI types. In future, these comparisons will be investigated in the study with complete data set of N=20 per group. The adequacy of the models was evaluated based on residuals and by the visual inspection of plots of predicted vs measured outcomes. The mean intercept (i.e., difference between OI and control at time = 0), slope values (i.e., OI status × time interaction) between groups, and the effect of time on the measurements of each group were tested to compare the baseline measurements and growth of OI and control groups. For each outcome, the raw (not predicted from the model) individual growth curves were plotted, overlayed by the mean estimated (from the model) group wise growth curves and their 95% confidence limits. Mean and standard deviation for HR-pQCT measurements separated for OI and control groups, and time point are shown in supplementary tables. For the fixed effects, $p < 0.05$ was considered statistically significant. No correction for multiple comparisons was performed as each outcome measure did not have multiple comparisons beyond the significance of fixed effects, the exploratory nature of the study, and to minimize the inflation of Type-II error rates, considering the low sample size.[22,23] The SAS statistical program (version 9.4; SAS Institute, Inc, Cary, NC) was used for all statistical analyses and visualizations.

6.3. Results:

6.3.1. Baseline characteristics:

The Shapiro-Wilk test did not reject the normality of differences of the baseline characteristics (age, weight, height, ulna length, tibia length) between the OI and control groups ($p > 0.05$). From paired t-test comparisons, there were no differences between baseline characteristics of participants with OI and controls ($p > 0.05$) (Table 1).

6.3.2. 3D image registration of longitudinal scans:

In healthy children, the rigid 3D-registration failed to properly align the scans from two time points. In the children with OI, the rigid registration was successful only if accompanied by the presence of specific landmarks that could be matched between two time points, as well as assignment of initial translation by the user (Figure 33). In this case, the landmark was the sclerotic lines. The sclerotic lines first enabled us to visually identify the correct registration, and secondly enabled proper registration by providing sufficient common features.

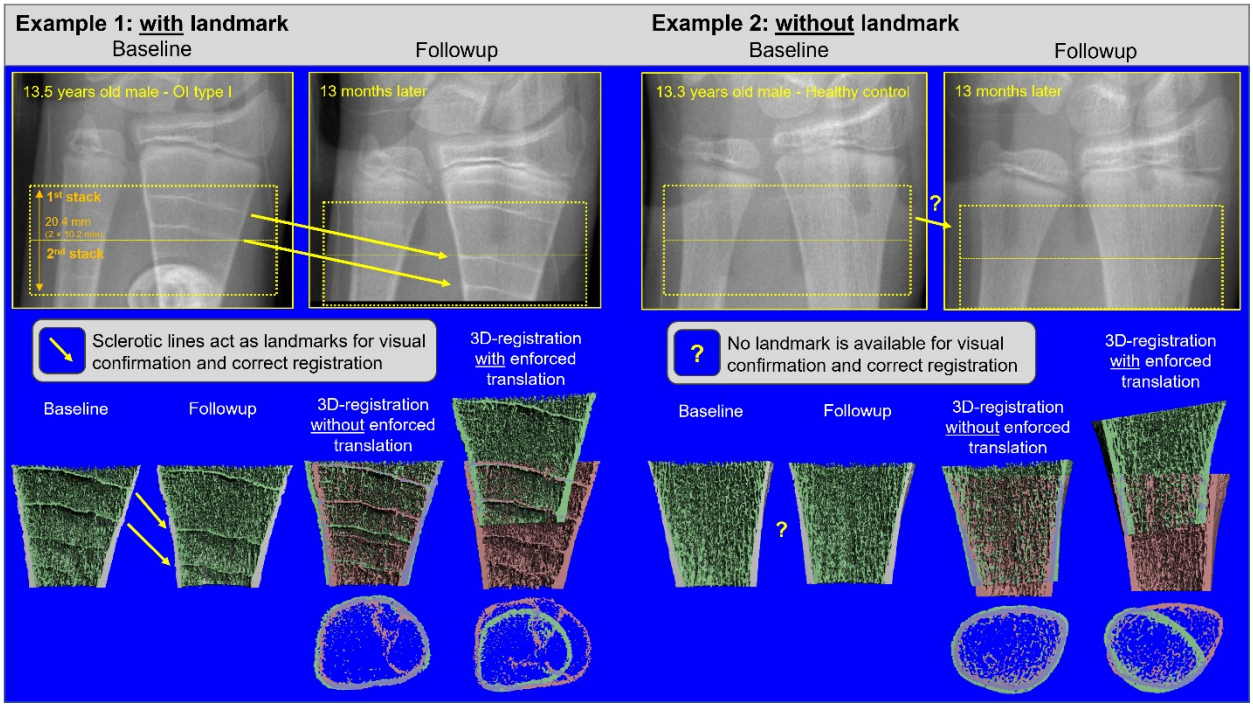


Figure 33. Examples of 3D image registration at the distal radius with and without landmarks. In the example with landmark, a 13.5 year old boy with type I OI is scanned at two timepoints, 12 months apart. Due to bisphosphonate treatment, the remnants of growth plate are

visible on the scans as distinct lines. These lines provided enough features between the scans for proper image registration, when accompanied by proper initial translation enforced by the user. The line could also be used for visual verification of the registration. One the other hand, for the case without landmark, a 13.3 year old healthy boy was scanned at two timepoints, 13 months apart. In this case, 3D registration failed to properly align the scans. Due to the lack of any distinct landmark, proper registration could not be visually verified

6.3.3. Stack alignment:

We did not observe any misalignment in the double stack image of the human cadaveric radius phantom, which confirmed that the misalignment between consecutive stacks is not scanner-related (Figure 34, bottom row). In contrast, the majority of *in vivo* double stack images had misalignments, suggesting participant motion was likely the cause. While the misalignment could be reduced using image registration, it could not be eliminated due to the presence of a gap between the stacks (Figure 34, top and middle rows).

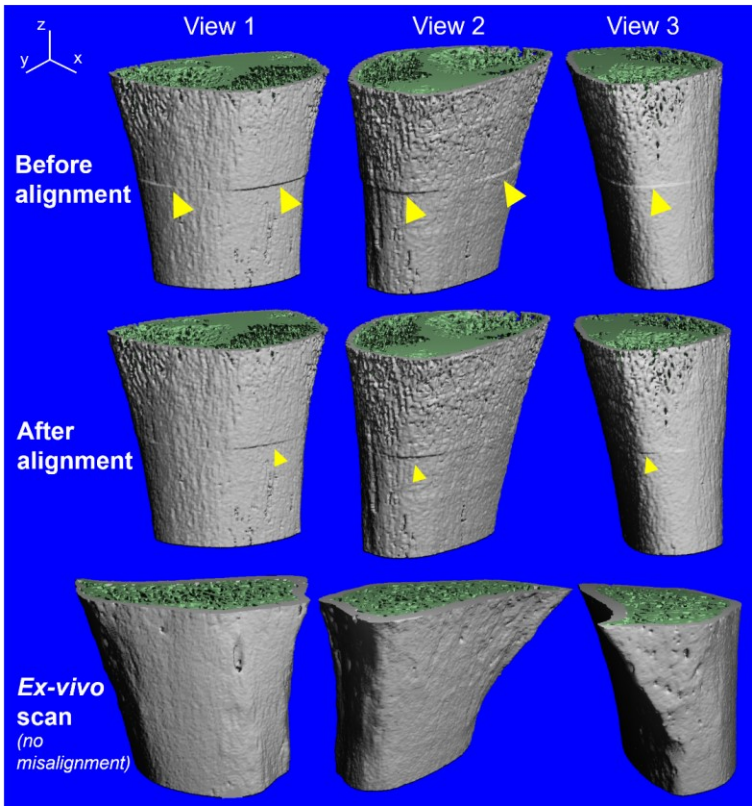


Figure 34. An example of stack alignment for double-stack scans without overlap. The top row shows the original double-stack scan with considerable misalignment (yellow arrows) from

three different views. The double-stack scan was acquired at the distal radius of a 11-year old boy, without any overlap between the two stacks. The middle row shows reduced misalignment between the stacks after 3D image registration. While the misalignment was reduced considerably, it was not removed completely. The bottom row indicates an ex-vivo double-stack scan from a cadaveric forearm phantom showing no misalignment

6.3.4. Differences between OI and controls:

Based on the results of our feasibility study on 3D registration, no registration was performed in this study. We also opted to analyze scans without correction for stack misalignment as we decided to report results for stacks separately. At the distal stack of tibia metaphysis (Figure 35, Supplementary Table 1), the baseline measurements (i.e., model intercept) were lower for the OI group compared to the control groups for Tt.vBMD, Tb.vBMD, and all trabecular morphological measurements except Tb.Th and Tb.Ar. Measurements of Tb.Sp and Tb.1/N.SD were higher for the OI group, where higher is worse. For cortical measurements, only Ct.Po had significantly different baseline measurements between the two groups. All of the microFE outcomes (i.e., Failure load, stiffness, and apparent modulus) were lower for the OI group. Increases in Ct.vBMD, Ct.Th, Ct.Ar, failure load, and stiffness were significant over time for the OI group. For the control group, the increases were significant for Ct.Ar, failure load, and stiffness. There was no difference between changes in bone measurements over time between the OI and controls groups (i.e., model time × OI status interaction term).

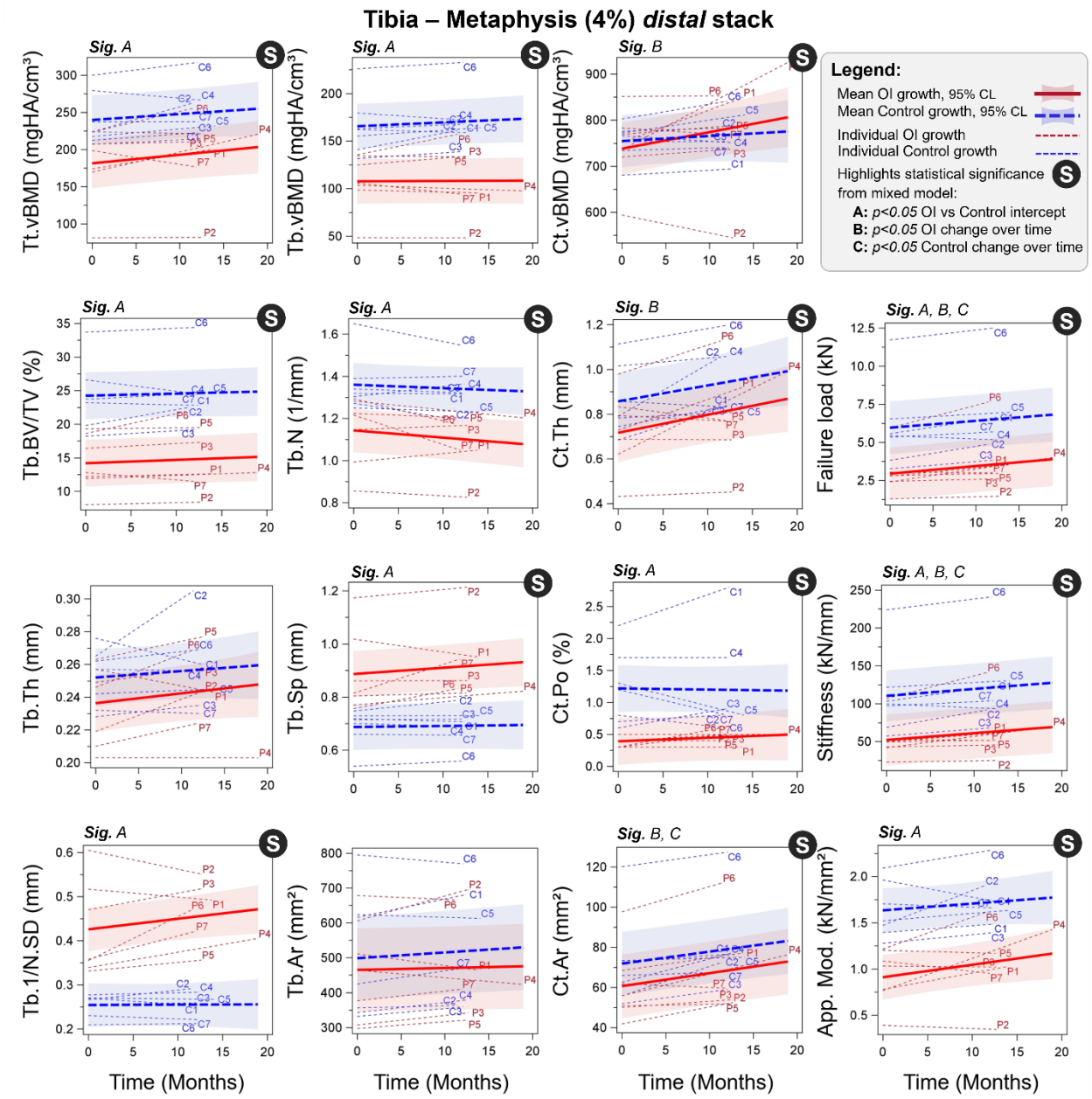


Figure 35. Linear growth curve for HR-pQCT outcomes at the distal stack of tibia metaphysis separated for participants with OI (red) and age- and sex-matched healthy controls (blue). The thick lines indicate the group mean growth curves for OI (solid) and controls (dashed) from the random intercept model, with shaded bands indicating the 95% confidence intervals. The thin dashed lines indicate raw growth curves for each participant, with labels "P" and "C" for participants with OI and controls, respectively. The icon indicating the letter "S" inside a circle highlight a statistically significant result. **Sig. A** p<0.05 OI vs Control intercept; **B** p<0.05 OI change over time; and **C** p<0.05 Control change over time. OI Osteogenesis Imperfecta

At the proximal stack of tibia metaphysis (Figure 36, Supplementary Table 2), the baseline measurements of OI group were lower for Tt.vBMD, Tb.vBMD, and all of trabecular morphological measurements (Tb.Sp and Tb.1/N.SD were higher in OI) except Tb.Th and Tb.Ar. For cortical measurements, Ct.Th and Ct.Ar at baseline were significantly smaller in the OI group. All of the microFE outcomes were smaller at baseline for the OI group. Increases in Tt.vBMD, Ct.vBMD, Ct.Th, Ct.Ar, failure load, stiffness, and apparent modulus were significant over time for the OI group. For the control group, the increases were significant for Ct.Ar, failure load, and stiffness. There was no difference between changes in bone measurements over time between the OI and controls groups.

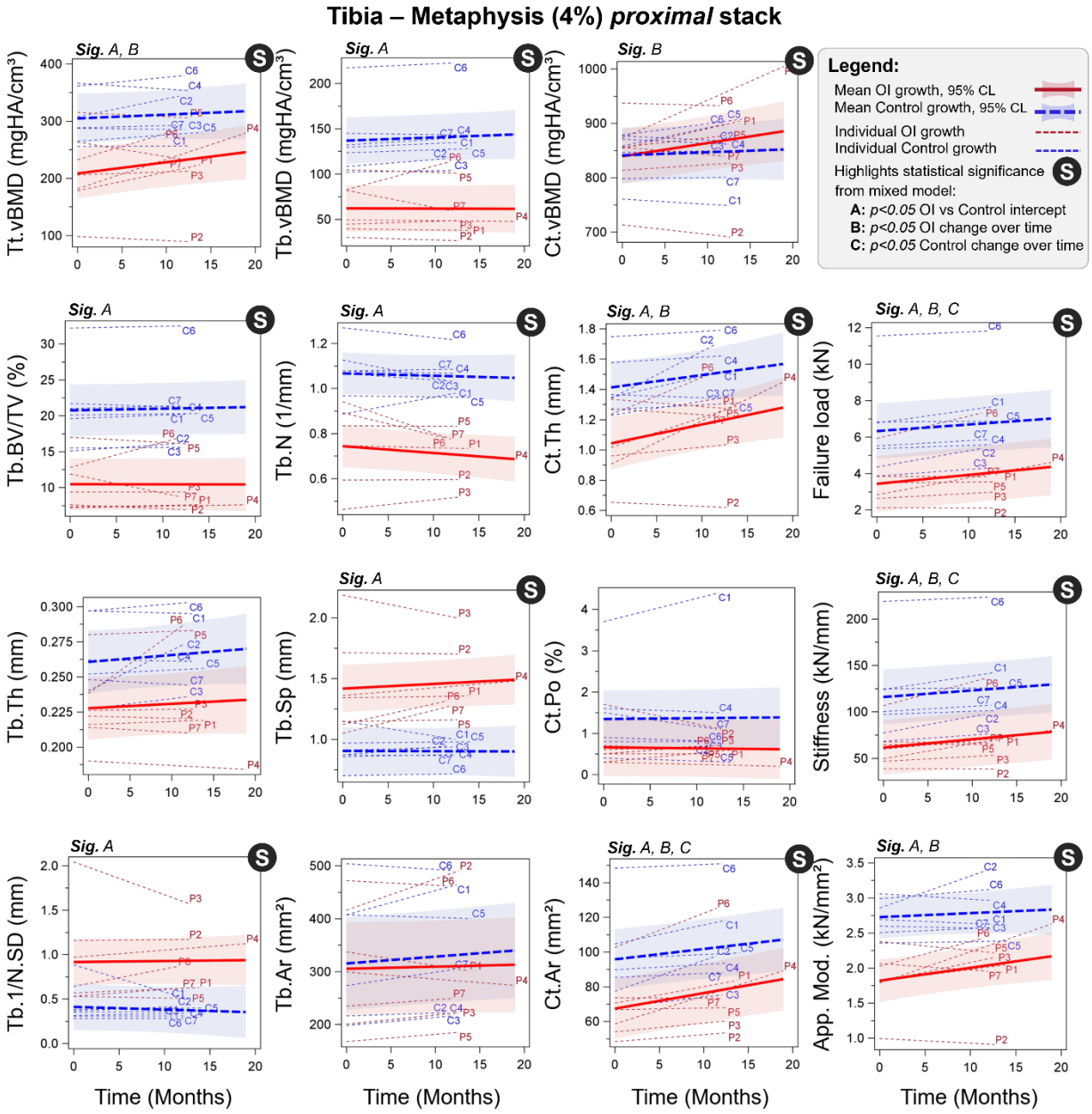


Figure 36. Linear growth curve for HR-pQCT outcomes at the proximal stack of tibia metaphysis separated for participants with OI (red) and age- and sex-matched healthy controls (blue). The thick lines indicate the group mean growth curves for OI (solid) and controls (dashed) from the random intercept model, with shaded bands indicating the 95% confidence intervals. The thin dashed lines indicate raw growth curves for each participant, with labels "P" and "C" for participants with OI and controls, respectively. The icon indicating the letter "S" inside a circle highlight a statistically significant result. **Sig. A** p<0.05 OI vs Control intercept; **B** p<0.05 OI change over time; and **C** p<0.05 Control change over time. OI Osteogenesis Imperfecta

At the tibia diaphysis (Figure 37, Supplementary Table 3), a significant difference was only found at baseline between the OI and control groups for Mr.Ar, where Mr.Ar was lower in the OI group. Increases in Ct.vBMD, Ct.Ar, failure load, and stiffness were significant over time for the OI group. For the control group, the increases were significant for Mr.Ar, Ct.Th, Ct.Ar, failure load, and stiffness. Changes over time were different between the OI and control groups for Ct.vBMD, where Ct.vBMD increased for OI group, and decreased for the control group.

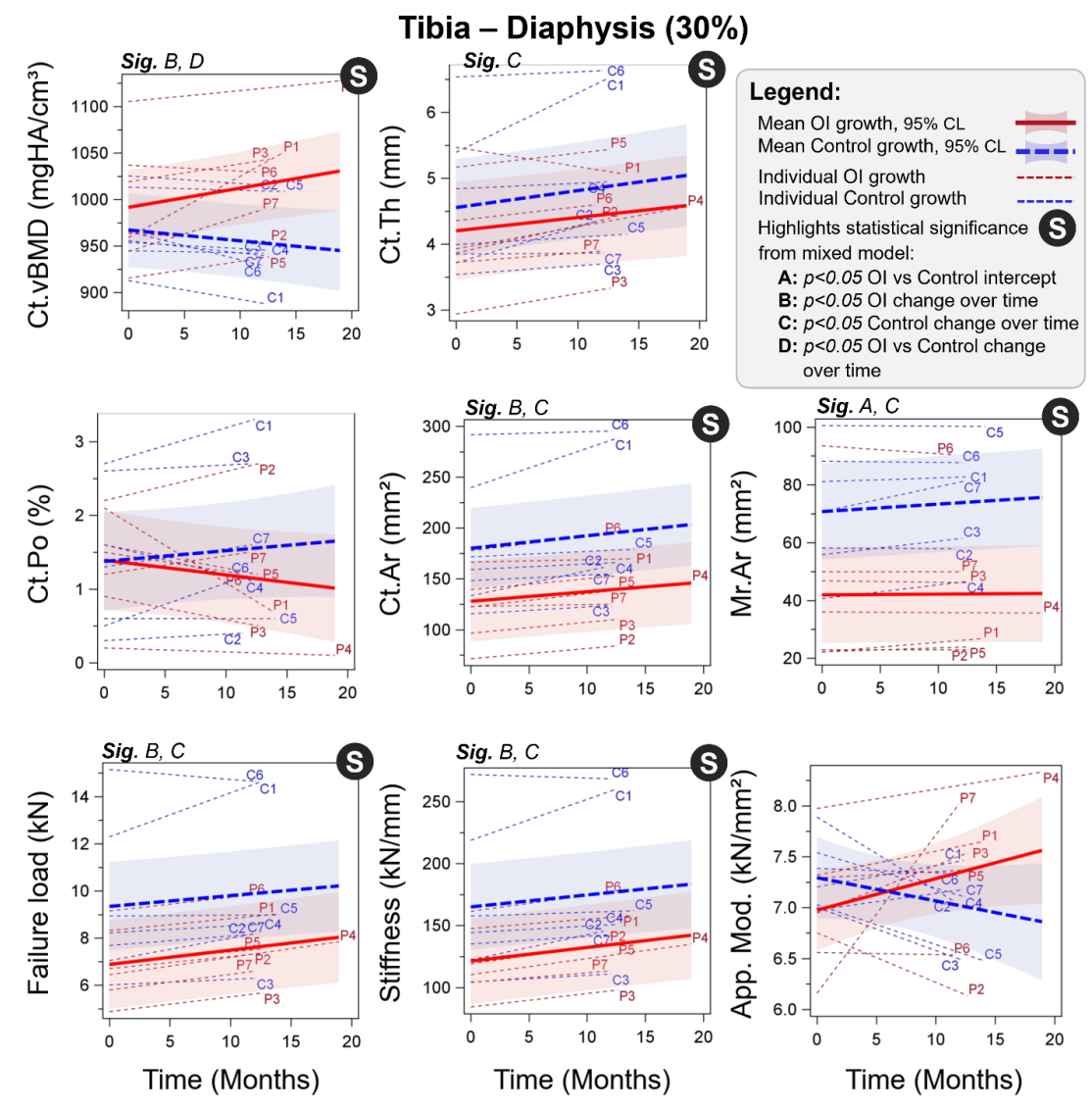


Figure 37. Linear growth curve for HR-pQCT outcomes at the tibia diaphysis separated for participants with OI (red) and age- and sex-matched healthy controls (blue). The thick lines indicate the group mean growth curves for OI (solid) and controls (dashed) from the random intercept model, with shaded bands indicating the 95% confidence intervals. The thin dashed lines

indicate raw growth curves for each participant, with labels “P” and “C” for participants with OI and controls, respectively. The icon indicating the letter “S” inside a circle highlight a statistically significant result. **Sig. A** $p<0.05$ OI vs Control intercept; **B** $p<0.05$ OI change over time; **C** $p<0.05$ Control change over time; and **D** $p<0.05$ OI vs Control change over time. OI Osteogenesis Imperfecta

At the distal stack of radius metaphysis (Figure 38, Supplementary Table 4), baseline measurements were higher for the OI group only for Tb.1/N.SD, where higher is worse. Ct.Th, Ct.Ar, and failure load increased significantly with time for both the OI and control groups. Increases in Tt.vBMD were significant only for the control group, while significant only for the OI group for Ct.vBMD.

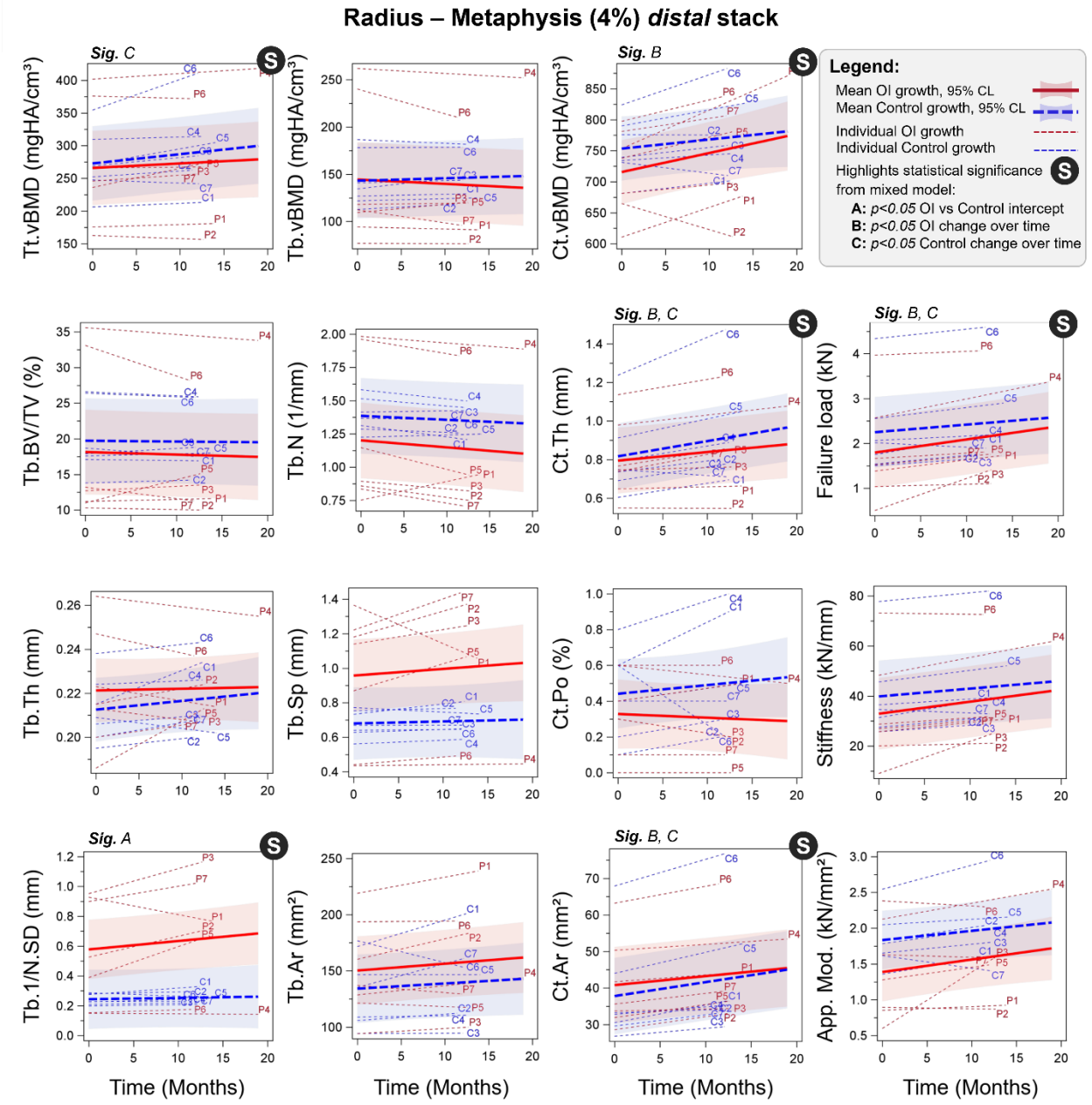


Figure 38. Linear growth curve for HR-pQCT outcomes at the distal stack of radius metaphysis separated for participants with OI (red) and age- and sex-matched healthy controls (blue). The thick lines indicate the group mean growth curves for OI (solid) and controls (dashed) from the random intercept model, with shaded bands indicating the 95% confidence intervals. The thin dashed lines indicate raw growth curves for each participant, with labels "P" and "C" for participants with OI and controls, respectively. The icon indicating the letter "S" inside a circle highlight a statistically significant result. **Sig. A** $p<0.05$ OI vs Control intercept; **B** $p<0.05$ OI change over time; and **C** $p<0.05$ Control change over time. OI Osteogenesis Imperfecta

At the proximal stack of radius metaphysis (Figure 39, Supplementary Table 5), baseline measurements were higher for the OI group only for Tb.1/N.SD. Tb.Sp, and Ct.Po. Tb.Sp, Ct.Ar, failure load, and stiffness increased significantly with time for the OI groups. For the control group, changes in Tt.vBMD, Ct.Th, Ct.Ar, failure load, and stiffness were significant.

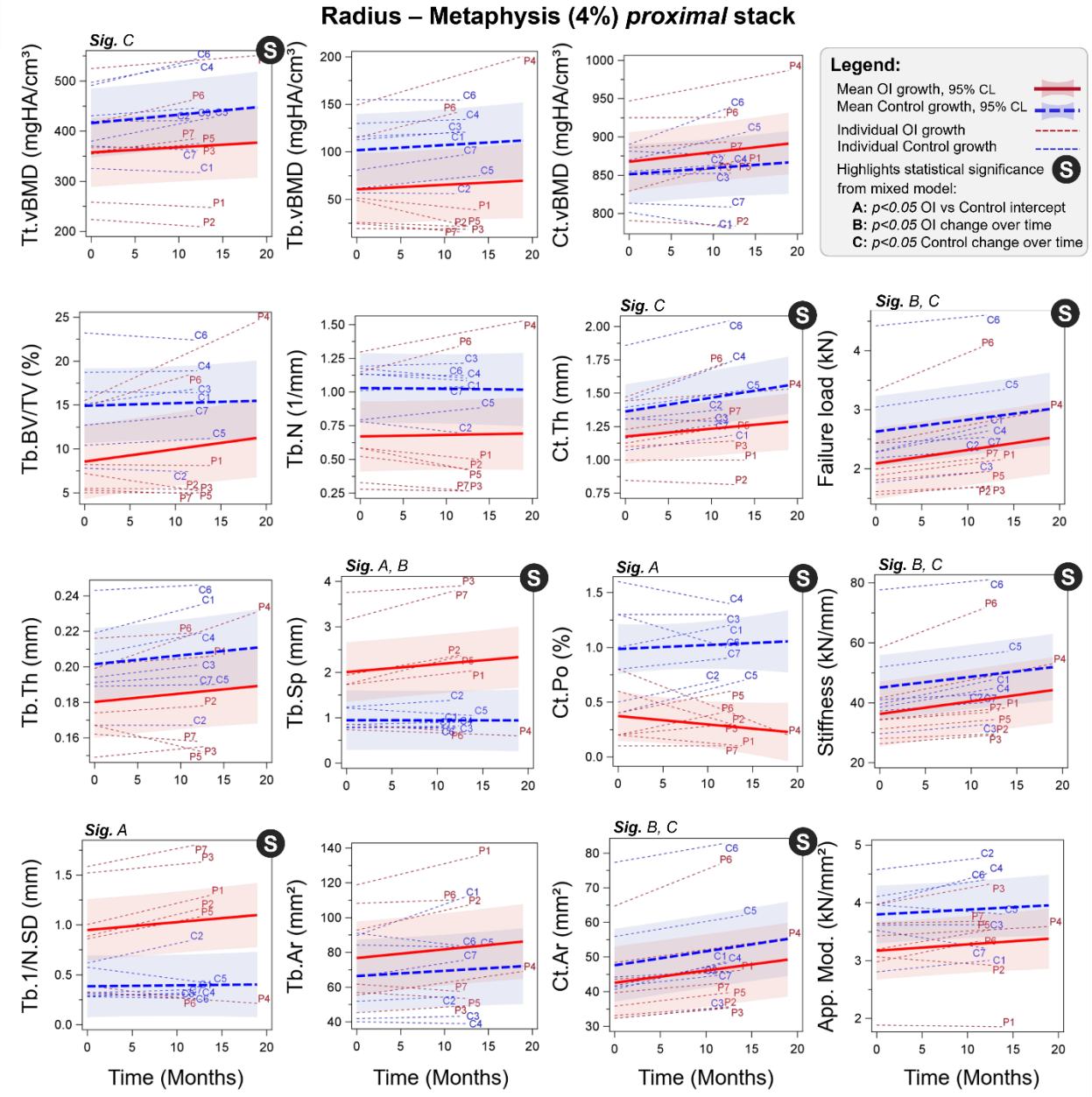


Figure 39. Linear growth curve for HR-pQCT outcomes at the proximal stack of radius metaphysis separated for participants with OI (red) and age- and sex-matched healthy controls (blue). The thick lines indicate the group mean growth curves for OI (solid) and controls (dashed) from the random intercept model, with shaded bands indicating the 95% confidence

intervals. The thin dashed lines indicate raw growth curves for each participant, with labels "P" and "C" for participants with OI and controls, respectively. The icon indicating the letter "S" inside a circle highlight a statistically significant result. **Sig. A** $p<0.05$ OI vs Control intercept; **B** $p<0.05$ OI change over time; and **C** $p<0.05$ Control change over time. OI Osteogenesis Imperfecta

At the radius diaphysis (Figure 40, Supplementary Table 6), baseline measurements for Mr.Ar, and apparent modulus were lower for the OI group. Ct.vBMD significantly increased for the OI group, while Ct.Po and apparent modulus significantly decreased. For the control group, only failure load and stiffness increased significantly. Changes over time were different between the OI and control groups for Ct.vBMD and apparent modulus.

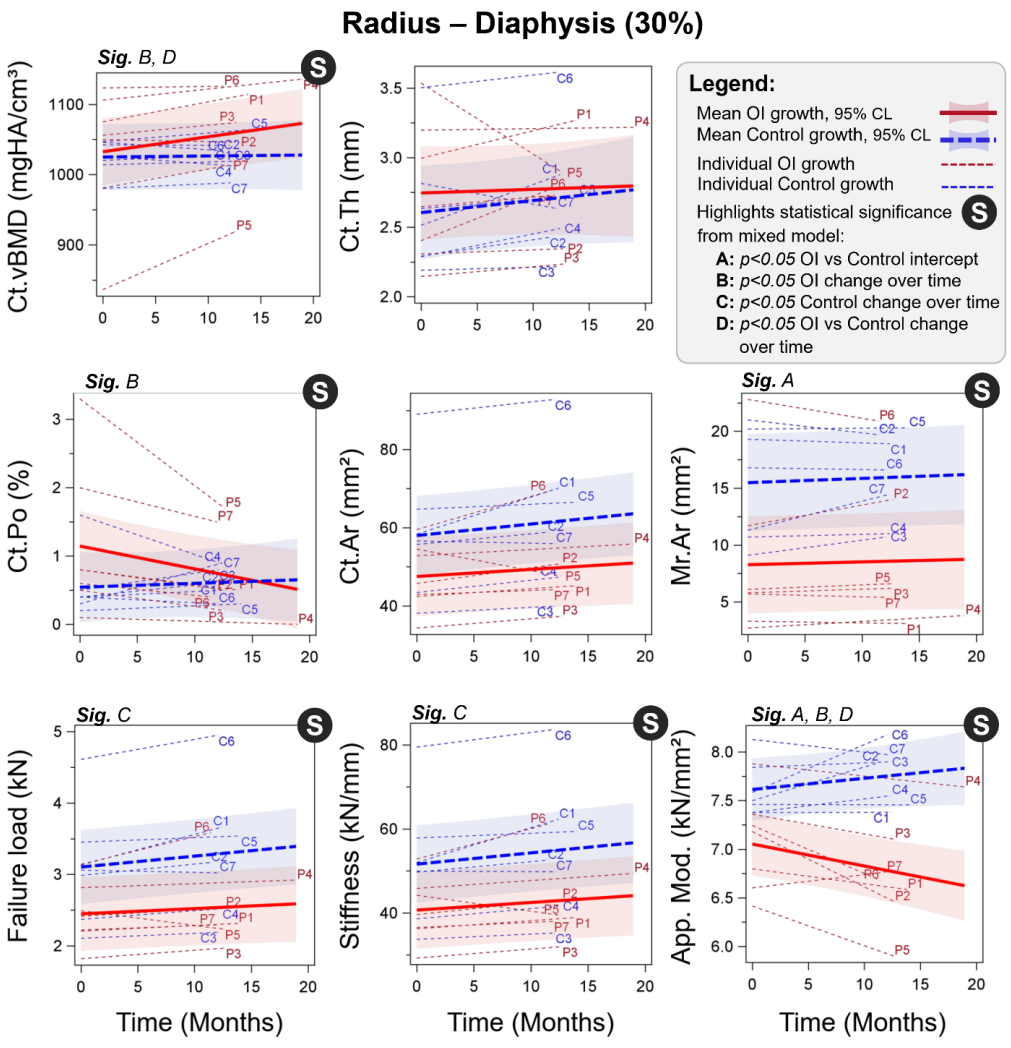


Figure 40. Linear growth curve for HR-pQCT outcomes at the radius diaphysis separated for participants with OI (red) and age- and sex-matched healthy controls (blue). The thick lines indicate the group mean growth curves for OI (solid) and controls (dashed) from the random intercept model, with shaded bands indicating the 95% confidence intervals. The thin dashed lines indicate raw growth curves for each participant, with labels "P" and "C" for participants with OI and controls, respectively. The icon indicating the letter "S" inside a circle highlight a statistically significant result. **Sig. A** $p<0.05$ OI vs Control intercept; **B** $p<0.05$ OI change over time; **C** $p<0.05$ Control change over time; and **D** $p<0.05$ OI vs Control change over time. OI Osteogenesis Imperfecta

6.4. Discussion:

In this study, we presented the first set of longitudinal HR-pQCT data from children with OI and age- and sex-matched healthy controls. At the metaphysis, our results suggest that the differences between the OI population and controls are more prominent at the tibia than at the radius. This finding is not surprising, as the tibia is a weight bearing site that is more susceptible to deformity due to the weight, and likely influenced by limited activity in individuals with OI compared to the controls. At the tibia metaphysis, we found that the baseline measurements were lower in the OI and control groups for Tt.vBMD, and most of trabecular outcomes, except thickness and area regardless of the image stack. Similar results were obtained for microFE outcomes, while differences between OI and controls were larger for cortical measurements for the proximal stack, where cortical bone is thicker. Cortical measurements and microFE results had significant changes over time for the OI group (except cortical porosity and apparent modulus) for both image stacks. The control group had significant increases over time only for the microFE outcomes. In contrast, at the radius metaphysis, only Tb.1/N.SD showed differences at baseline between the OI and control groups, as well as Tb.Sp and Ct.Po for the proximal stack. Changes over time were similar between the groups, and were mostly for microFE and cortical measurements. The patterns observed in this study are mostly in line with previous studies from adults, while the previous pediatric study did not perform any statistical analysis. While the numeric effect sizes (i.e., difference between OI and control groups) for cortical measurements seem comparable between this study and those of Fennimore et. al, they are generally larger for the trabecular bone and microFE in this study. These differences may be explained by the differences in scanner generations (e.g., Fennimore et. al. used XCT vs XCT2 in this study) that are known to cause the largest differences in trabecular outcomes,[13] and scanned region (e.g., Fennimore et. al. scanned a single stack at 1 mm proximal from the proximal point of the distal

epiphyseal growth plate, whereas we used a relative ROI positioning). It is possible that the use of a fixed region of interest by Fennimore resulted in scanned regions being more distal for the control group in case of longer bones for the control group, although we do not know if the limb length was different. Considering that metaphysis is a site of large changes in bone macro-scale properties,[24] we anticipate that the differences in the scanned region are the main reason behind possible differences. Unfortunately, bone length, bone area, and trabecular volume fraction were not reported in the study by Fennimore to facilitate the examination of our hypothesis. The observation of lower bone density and deteriorated bone microstructure in OI group is not unique to children, as similar findings have been reported in HR-pQCT studies comparing adults with OI and healthy controls.[4–7]

In a pQCT study on the radius of children and adolescents with type I OI, and age- and sex-matched healthy controls, Rauch et. al.,[14] showed similar differences at the metaphysis (i.e., lower BMD and cortical thickness) between the OI and control groups. Although the two studies used different modalities, the scanned regions of interest at the metaphysis started at a similar location, although scan length was different at 10 mm vs 2 mm for HR-pQCT and pQCT.

In our study, all of the participants in the OI group had been treated with bisphosphonates for at least two years prior to their participation. Studies on adults with OI have shown that bisphosphonates increase BMD, and total bone volume.[25–27] In growing children and adolescents, it has been shown that bisphosphonates significantly increase cortical width, bone volume per tissue volume, and trabecular number.[28] Accordingly, we anticipate that treatment with bisphosphonates could have minimized differences between the OI and control groups. For instance, we observed that cortical porosity was lower in the OI group compared to controls at the metaphysis. Lower cortical porosity typically contributes to higher strength. However, this study is not the first to report such a pattern.[4,6] This could be a result of bisphosphonate treatment in the OI group, that reduces cortical porosity. Bisphosphonates can further impact the differences between the OI and control groups by creating sclerotic lines. Sclerotic lines are horizontal trabeculae containing some degree of cartilage and are thought to be caused by the temporary interruption of growth plate cartilage resorption, which can remain in bone even into adulthood.[29] Sclerotic lines appear as bright areas on images that resemble bone and falsely increase the measured bone mass. Sclerotic lines are the product of bisphosphonates which are commonly

given to children with OI. The presence of sclerotic lines can reduce the reliability of HR-pQCT scans because these artifacts result in measurements that may not be representative of true bone properties. In our dataset, we observed sclerotic lines in all OI scans at both the radius and tibia. We anticipate that the difference between HR-pQCT measurements of children with OI and healthy controls will become larger after removing the sclerotic lines. Future studies are needed to test this hypothesis.

In this study, we also reported HR-pQCT measurements from the diaphysis, which has not been done before to our knowledge. Studying the diaphysis can provide additional information as it undergoes different modes of loading and has distinct morphology compared to the metaphysis. In fact, we found that at both the radius and tibia, Ct.vBMD increased over the course of the study for the OI group, while it had a downward trend for the control group (changes for the control group were not significant, but the treatment-time interaction was significant). On the other hand, the control group had larger cross-sectional area. It has been suggested that the smaller cross-sectional area in the diaphysis of long bones in children with OI is associated with the high rate of fractures at the diaphysis.[30,31] On the other hand, low trauma diaphyseal fractures are not common in healthy children. Rather, healthy children may experience fractures at the distal radial metaphysis during a fall, it has been suggested that this is because of the insufficient increase in cortical thickness and subsequent lag in the increase in distal radius strength during their peak growth period.[32] In fact, we found that the strength measurements were larger for the control group, although the difference was not statistically significant. Interestingly, apparent modulus decreased for the OI group, while it increased for the control group. These results suggest the importance of bone morphology at the radius and tibia diaphysis for its strength. However, it should be noted that our microFE models assign a single material to bone, and different strength results may be obtained when using microFE models with density-based material properties.[18,19] In the pQCT study on children and adolescents with type I OI, and age- and sex-matched healthy controls, Rauch et. al.,[14] showed a slight but statistically significant increased cortical vBMD in the diaphysis of the radius for the OI group. While we found a similar pattern, our results did not reach statistical significance. This could be due to differences in scanned region, as well as the statistical power of the study considering the relatively small effect size.

In this study, we also investigated the feasibility of rigid 3D-registration using Scanco image processing language (IPL) on scans taken at baseline and 1-year follow up in children with osteogenesis imperfecta (OI), and age and sex-matched healthy controls. In healthy children, the rigid 3D-registration failed to properly align the metaphyseal scans from two time points. In the children with OI, the rigid registration was successful when the sclerotic lines were present in the images, as well as assignment of initial translation by the user (Figure 33). The sclerotic lines first enabled us to visually identify the correct registration, and secondly enabled proper registration by providing sufficient common features. Therefore, in the absence of common landmarks between subsequent scans, rigid 3D-registration is inadequate for growing bone (Figure 33). Considering that changes in bone due to growth are less dramatic at the diaphysis, 3D registration at the diaphysis should also be investigated. More sophisticated 3D registration methods such as non-rigid registration may be able to align growing bone even in the absence of clear landmarks. In addition, registration of growing bone requires a sufficient image stack length to be scanned. For instance, for a one-year interval between two scans, at least two stacks (~2 cm) are needed to ensure that the same bone material can be covered in both scans, considering ~ 1 cm/year growth.

Regardless of the feasibility of image registration, the need for image registration depends on the purpose of the longitudinal study. In adults, image registration is used to identify the same exact bone volume (which corresponds to the same anatomical site) of interest by correcting repositioning errors. In children, the same anatomical region (e.g., metaphysis) and the same exact volume of bone over time are not identical because as bone grows, the distance between the existing bone and growth plate increases. Since repositioning error also occurs in children and image registration cannot correct it, longitudinal changes in children's bones need to be interpreted with respect to unregistered short-term precision errors, and with additional caution. In this regard, data on HR-pQCT precision errors for children with OI are needed to assess the monitoring time interval required between longitudinal scans to be able to detect changes using HR-pQCT.[33]

In this study, we also investigated the misalignment between the stacks of double-stack HR-pQCT scans, and the possibility of alignment. Scanning two image stacks is useful when scanning children over time, considering the ~1cm/year growth in bone length, which is similar to the height of a single stack. We first collected a double image stack on a cadaveric phantom and confirmed that the misalignment between consecutive stacks is not scanner-related by observing

no misalignment between the two stacks of ex-vivo scans (Figure 34). Misalignment occurred in almost all *in-vivo* double-stack scans (Figure 34). Thus, the misalignment between consecutive stacks is caused by limb movement, which can create misalignment by two main mechanisms. In one case, limb movement may occur during the idle time (approximately 5 seconds) between the end of first stack, and the beginning of the second stack. With this type of movement, no motion artifact will be present. In another case, movement can occur during the scanning of either of the stacks or both. The main difference between the two mechanisms is that the latter creates motion artifact, while the former does not. In practice, misalignment between the two image stacks can be a result of a combination of both types of movement.

For average-based HR-pQCT outcomes including density, geometry, and microstructure, misalignment between the two stacks may be left uncorrected. However, microFE is sensitive to misalignment, as it can affect the connectivity of the bone. In those situations, either the two stacks can be analyzed separately or misalignment can be corrected. Nevertheless, it is worth noting that the effect of misalignment of model connectivity may not be much different from the effect of motion on microFE analysis.

The possibility of correction of misalignment between the stacks depends on the scanning protocol and the source of misalignment. Firstly, in the case of motion artifact during scanning, misalignment cannot be perfectly corrected since the images are in fact deformed. In the absence of motion artifact during scanning, out observations suggest that the misalignment cannot be removed entirely when there is no overlap between the stacks. A more conservative method creating an overlap between stacks may enable the correction of misalignment caused by movement during the idle time. While we do not have scans with overlaps, in a study by Brunet et. al.,[34] to study bone remodeling at the metacarpophalangeal joints of adults with rheumatoid arthritis, leaving an overlap seemed to enable proper alignment of image stacks. They acquired three consecutive stacks with 25% overlap between them, and used the overlap to align the stacks using 3D image registration. However, this method results in overall longer scanning time with subsequently more radiation exposure, and smaller scanned region due to redundancy, although a smaller overlap may still enable proper alignment. Therefore, it should be used only when correct alignment of stacks is needed. The alternative multi-stack scanning protocol does not have any

overlap between the stacks, similar to the one used in this study. For our analyses, we separated the stacks for consistency, and improved interpretability.

This study has several strengths. First, the position of the scanned regions with respect to the reference line was relative to bone lengths, ensuring that the same anatomical site was scanned among participants. This is of particular importance when comparing groups with different limb lengths such as OI. Related to this, our reference line positioning resulted in regions as distal as possible, while making sure to not target the growth plate. Secondly, we scanned two stacks at the metaphysis. This ensured that we could capture the same bone material during the course of the study and perform image registration. Similarly, scanning the long bone diaphysis not only provided us with more insight from a bone region with different loading modes and morphology, but is also clinically relevant considering that diaphysis is the primary site of long bone fracture in children with OI.

The limitations of this study should also be acknowledged. The first limitation is the small sample size and the consequent low power and the inflated estimates of effect size or "winner's curse". While some of the estimated effect sizes were detected even with the low sample size, a larger sample size will provide more accurate estimates. It should be noted that this manuscript is part of an ongoing study, where the sample size will increase to 20 per group. Similarly, in the current dataset, only one female was included compared to 6 males, which is again resolved in the complete dataset. Another limitation was the varying time interval (i.e., 9-14 months) between the baseline and follow up scans among participants. However, this was inevitable due to the complexity of recruitment and coordinating scanning sessions during a global pandemic. To minimize the impact, we approximately matched the time interval between each OI participant and the corresponding control. Also, we accounted for different time intervals in our statistical model. A third limitation is that, the image segmentation algorithm and microFE properties were the same as those of adults. Considering that the average thickness of trabeculae in children tends to be smaller than in adults, a segmentation protocol that most accurately preserves the fine trabeculae might be beneficial, which needs to be investigated in future studies. However, obtaining pediatric-specific microFE material properties and failure criterion is not easy due to the lack of pediatric cadaveric specimens for experimental validation studies. Nevertheless, the absolute values of outcomes are not problematic for the purpose of comparison between populations.

In this study, we report the first set of longitudinal HR-pQCT data at the metaphysis and diaphysis of radius and tibia from children with OI, and compared them with age- and sex-matched healthy controls. Our key findings in this 1-year data were (1) At the metaphysis, the differences between the OI and control groups were more prominent for the tibia, while the longitudinal changes over the ~1-year period were mostly comparable between the two groups. (2) Our data suggest that the difference between the radius and tibia at the metaphysis is mainly in trabecular microstructure (i.e., more deteriorated in the OI group at the tibia). At the diaphysis, we found that the radius and tibia were mostly similar. The differences between the two groups at the diaphysis were mainly driven by area, in particular larger marrow cavity area in the control group. 3) We showed that longitudinal image registration in growing bones is not feasible at the proximal metaphysis in the absence of landmarks such as sclerotic lines, at least using the common rigid registration methods. 4) We found that subject movement causes misalignment between the stacks of our double-stack scans. While we could minimize the misalignment using image registration, we could not eliminate it. However, such misalignment does not affect density and morphological outcomes, while microFE needs to be done on stacks separately.

Conflict of Interest Statement:

This research was supported by Mereo Biopharma and Shriners Hospitals for Children. BMW and FR have received institutional research support and materials and are consultants for Mereo BioPharma.

Author Contributions:

Study conception and design: SH, EZ, FG, BMW; Study coordination: FC; Methodology and software: SH; Data acquisition: SH, BMW; Analysis and interpretation of data: SH, SM, FR, BMW; Drafting of Manuscript: SH, BMW. All authors contributed to the critical revision and approval of the final manuscript.

Acknowledgements:

We thank all the study participants and staff at the Shriners Hospital for Children-Canada. We also thank Catherine Julien for assisting with data collection, and Carolyn Denton for assisting with data analysis. We thank members of Ultragenyx and Mereo BioPharma for assistance with some data organization. The study was funded by Mereo BioPharma, Shriners Hospitals for Children, and FRQS Programme de bourses de chercheur (BMW, SH).

References:

1. Whittier DE, Boyd SK, Burghardt AJ, Paccou J, Ghasem-Zadeh A, Chapurlat R, Engelke K, Bouxsein ML. Guidelines for the assessment of bone density and microarchitecture in vivo using high-resolution peripheral quantitative computed tomography. Osteoporos. Int. 2020;

2. 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.

3. Marini JC. Osteogenesis imperfecta. Prim. Metab. Bone Dis. Disord. Miner. Metab. 2018.

4. Folkestad L, Hald JD, Hansen S, Gram J, Langdahl B, Abrahamsen B, Brixen K. Bone geometry, density, and microarchitecture in the distal radius and tibia in adults with osteogenesis imperfecta type i assessed by high-resolution pQCT. J. Bone Miner. Res. 2012;27(6):1405–12.

5. Rolvien T, Stürznickel J, Schmidt FN, Butscheidt S, Schmidt T, Busse B, Mundlos S, Schinke T, Kornak U, Amling M, Oheim R. Comparison of Bone Microarchitecture Between Adult Osteogenesis Imperfecta and Early-Onset Osteoporosis. Calcif. Tissue Int. [Internet]. Springer US; 2018;103(5):512–21. Available from: http://dx.doi.org/10.1007/s00223-018-0447-8

6. Kocijan R, Muschitz C, Haschka J, Hans D, Nia A, Geroldinger A, Ardelt M,

Wakolbinger R, Resch H. Bone structure assessed by HR-pQCT, TBS and DXL in adult patients with different types of osteogenesis imperfecta. Osteoporos. Int. 2015;

7. Hald JD, Folkestad L, Harsløf T, Lund AM, Duno M, Jensen JB, Neghabat S, Brixen K, Langdahl B. Skeletal phenotypes in adult patients with osteogenesis imperfecta—correlations with COL1A1/COL1A2 genotype and collagen structure. Osteoporos. Int. [Internet]. Osteoporosis International; 2016;27(11):3331–41. Available from: http://dx.doi.org/10.1007/s00198-016-3653-0

8. Plachel F, Renner U, Kocijan R, Muschitz C, Lomoschitz F, Resch H. Osteogenesis imperfecta Typ III und hypogonadotroper Hypogonadismus führen zu massiven Knochenverlust: ein Fallbericht. Wiener Medizinische Wochenschrift. 2015;165(13–14):285–9.

9. Mikolajewicz N, Zimmermann EA, Rummler M, Hosseinitabatabaei S, Julien C, Glorieux FH, Rauch F, Willie BM. Multisite longitudinal calibration of HR-pQCT scanners and precision in osteogenesis imperfecta. Bone. 2021;147(February):115880.

10. Hosseinitabatabaei S, Mikolajewicz N, Zimmermann EA, Rummler M, Steyn B, Julien C, Rauch F, Willie BM. 3D image registration marginally improves the precision of HR-pQCT measurements compared to cross-sectional-area registration in adults with osteogenesis imperfecta. J. Bone Miner. Res. 2022;00(00):1–17.

11. Fennimore DJ, Digby M, Paggiosi M, Arundel P, Bishop NJ, Dimitri P, Offiah AC. High-resolution peripheral quantitative computed tomography in children with osteogenesis imperfecta. Pediatr. Radiol. Pediatric Radiology; 2020;3(XtremeCT II).

12. Manske SL, Zhu Y, Sandino C, Boyd SK. Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. Bone. 2015;79:213–21.

13. Manske SL, Davison EM, Burt LA, Raymond DA, Boyd SK. The Estimation of Second-Generation HR-pQCT From First-Generation HR-pQCT Using In Vivo Cross-Calibration. J. Bone Miner. Res. 2017;32(7):1514–24.

14. Rauch F, Land C, Cornibert S, Schoenau E, Glorieux FH. High and low density in the same bone: A study on children and adolescents with mild osteogenesis imperfecta. Bone. 2005;37(5):634–41.

15. Rauch F, Cornibert S, Cheung M, Glorieux FH. Long-bone changes after pamidronate discontinuation in children and adolescents with osteogenesis imperfecta. Bone. 2007;40(4):821–7.

16. Pialat JB, Burghardt AJ, Sode M, Link TM, Majumdar S. Visual grading of motion induced image degradation in high resolution peripheral computed tomography: Impact of image quality on measures of bone density and micro-architecture. Bone. 2012;50(1):111–8.

17. Hildebrand T, Rüegsegger P. A new method for the model-independent assessment of thickness in three-dimensional images - Hildebrand - 2003 - Journal of Microscopy - Wiley Online Library. J. Microsc. [Internet]. 2015;185(1):67–75. Available from: http://onlinelibrary.wiley.com/doi/10.1046/j.1365-2818.1997.1340694.x/abstract%5Cnhttp://files/158/abstract.html

18. 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.

19. Hosseinitabatabaei S, Kawalilak C, McDonald M, Kontulainen S, Johnston J. Distal radius sections offer accurate and precise estimates of forearm fracture load. Clin. Biomech. 2020;80(November 2019):105144.

20. Boyd STKB, Oser STM, Uhn MIK, Linck ROJK, Rauze PELK. Evaluation of Three-dimensional Image Registration Methodologies for In Vivo Micro-computed Tomography. 2006;34(10):1587–99.

21. McNeish D. Small Sample Methods for Multilevel Modeling: A Colloquial Elucidation of REML and the Kenward-Roger Correction. Multivariate Behav. Res. [Internet]. Taylor & Francis; 2017;52(5):661–70. Available from: https://doi.org/10.1080/00273171.2017.1344538

22. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol. Opt. 2014.

23. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990;

24. Boyd SK. Site-Specific Variation of Bone Micro-Architecture in the Distal Radius and Tibia. J. Clin. Densitom. 2008;11(3):424–30.

25. Adami S, Gatti D, Colapietro F, Fracassi E, Braga V, Rossini M, Tatò L. Intravenous neridronate in adults with osteogenesis imperfecta. J. Bone Miner. Res. 2003;

26. Shapiro JR, McCarthy EF, Rossiter K, Ernest K, Gelman R, Fedarko N, Santiago HT, Bober M. The effect of intravenous pamidronate on bone mineral density, bone histomorphometry, and parameters of bone turnover in adults with type IA osteogenesis imperfecta. Calcif. Tissue Int. 2003;

27. Chevrel G, Schott AM, Fontanges E, Charrin JE, Lina-Granade G, Duboeuf F, Garnero P, Arlot M, Raynal C, Meunier PJ. Effects of oral alendronate on BMD in adult patients with osteogenesis imperfecta: A 3-year randomized placebo-controlled trial. J. Bone Miner. Res. 2006;

28. Rauch F, Travers R, Plotkin H, Glorieux FH. The effects of intravenous pamidronate on the bone tissue of children and adolescents with osteogenesis imperfecta. J. Clin. Invest. 2002;

29. Rauch F, Travers R, Munns C, Glorieux FH. Clinical vignette - Sclerotic metaphyseal lines in a child treated with pamidronate: Histomorphometric analysis. J. Bone Miner.

Res. 2004;

30. Palomo T, Fassier F, Ouellet J, Sato A, Montpetit K, Glorieux FH, Rauch F. Intravenous Bisphosphonate Therapy of Young Children with Osteogenesis Imperfecta: Skeletal Findings during Follow Up Throughout the Growing Years. J. Bone Miner. Res. 2015;

31. Palomo T, Vilacą T, Lazaretti-Castro M. Osteogenesis imperfecta: Diagnosis and treatment. Curr. Opin. Endocrinol. Diabetes Obes. 2017.

32. Rauch F, Neu C, Manz F, Schoenau E. The development of metaphyseal cortex - Implications for distal radius fractures during growth. J. Bone Miner. Res. 2001;16(8):1547–55.

33. Bunyamin A, Björkman K, Kawalilak C, Hosseinitabatabaei S, Teare A, Johnston J, Kontulainen S. Reliability of Annual Changes and Monitoring Time Intervals for Bone Strength, Size, Density, and Microarchitectural Development at the Distal Radius and Tibia in Children: A 1-Year HR-pQCT Follow-Up. J. Bone Miner. Res. 2019;34(7):1297–305.

34. Brunet SC, Kuczynski MT, Bhatla JL, Lemay S, Pauchard Y, Salat P, Barnabe C, Manske SL. The utility of multi-stack alignment and 3D longitudinal image registration to assess bone remodeling in rheumatoid arthritis patients from second generation HR-pQCT scans. BMC Med. Imaging. BMC Medical Imaging; 2020;20(1):1–10.

Supplemental results:

Supplementary Table 1: Mean ± standard deviation for HR-pQCT measurements at the distal stack of the tibia metaphysis separated for OI and control groups, and timepoint. Statistical inferences are based on the mixed model for repeated measures (see the bottom of table).

Parameter	OI *(N=7)*		Control *(N=7)*	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**
Density (mg/cm³)				
Tt.vBMD[A]	182.5 ± 49.6	195.8 ± 55.6	239.6 ± 35.9	249.8 ± 35.9
Tb.vBMD[A]	107.3 ± 30	108.5 ± 35.9	165.4 ± 30.9	170.7 ± 29.3
Ct.vBMD[B]	743.4 ± 78.4	780.7 ± 121.6	754.8 ± 39.1	767.8 ± 48.6
Area (mm²)				
Tb.Ar	462.9 ± 146.4	474.6 ± 149.2	498.5 ± 180.4	519.1 ± 168.3
Ct.Ar[B, C]	60.9 ± 18.3	69 ± 21.8	71.7 ± 22.8	79.3 ± 22.1
Trabecular microstructure				
Tb.BV/TV (%)[A]	14.2 ± 4.1	14.9 ± 4.8	24.2 ± 5	24.6 ± 4.7
Tb.N (1/mm)[A]	1.14 ± 0.16	1.1 ± 0.14	1.36 ± 0.14	1.34 ± 0.12
Tb.Th (mm)	0.24 ± 0.02	0.25 ± 0.03	0.25 ± 0.02	0.26 ± 0.03
Tb.Sp (mm)[A]	0.89 ± 0.15	0.92 ± 0.14	0.69 ± 0.07	0.69 ± 0.07
Tb.1/N.SD (mm)[A]	0.43 ± 0.11	0.46 ± 0.07	0.25 ± 0.03	0.26 ± 0.03
Cortical microstructure				
Ct.Th (mm)[B]	0.72 ± 0.17	0.82 ± 0.22	0.86 ± 0.15	0.95 ± 0.15
Ct.Po (%)[A]	0.39 ± 0.09	0.47 ± 0.1	1.21 ± 0.58	1.2 ± 0.79
microFE				
Failure load (kN)[A, B, C]	2.95 ± 1.43	3.59 ± 1.95	5.95 ± 2.78	6.52 ± 2.85
Stiffness (kN/mm)[A, B, C]	52.3 ± 26	63.8 ± 36.8	110.2 ± 54.6	121.7 ± 56.6
App. modulus (kN/mm²)[A]	0.92 ± 0.3	1.08 ± 0.4	1.63 ± 0.3	1.73 ± 0.3

Significance from the mixed model:

A *p<0.05 OI vs Control intercept*

B *p<0.05 OI change over time*

C *p<0.05 Control change over time*

D *p<0.05 OI vs Control change over time.*

OI: Osteogenesis Imperfecta

Supplementary Table 2: Mean ± standard deviation for HR-pQCT measurements at the proximal stack of the tibia metaphysis separated for OI and control groups, and timepoint. Statistical inferences are based on the mixed model for repeated measures (see the bottom of table)

Parameter	OI *(N=7)*		Control *(N=7)*	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**
Density (mg/cm³)				
Tt.vBMD[A, B]	210.7 ± 69.2	232.6 ± 72.3	304.4 ± 43.9	313 ± 45.6
Tb.vBMD[A]	61.9 ± 27.8	62 ± 32.6	136.8 ± 38.3	141.3 ± 38.3
Ct.vBMD[B]	843.2 ± 68.1	868.1 ± 98.3	841.4 ± 45.1	847.8 ± 55.9
Area (mm²)				
Tb.Ar	303.9 ± 112.5	312.1 ± 118.4	315.2 ± 123.2	331.4 ± 119.5
Ct.Ar[A, B, C]	67.7 ± 17.9	79 ± 23.8	95.6 ± 26.2	103.3 ± 24.6
Trabecular microstructure				
Tb.BV/TV (%)[A]	10.5 ± 3.7	10.4 ± 4.2	20.7 ± 5.7	21 ± 5.6
Tb.N (1/mm)[A]	0.74 ± 0.17	0.7 ± 0.11	1.07 ± 0.12	1.05 ± 0.08
Tb.Th (mm)	0.23 ± 0.03	0.23 ± 0.04	0.26 ± 0.03	0.27 ± 0.03
Tb.Sp (mm)[A]	1.42 ± 0.4	1.47 ± 0.29	0.9 ± 0.14	0.9 ± 0.1
Tb.1/N.SD (mm)[A]	0.92 ± 0.55	0.93 ± 0.38	0.41 ± 0.21	0.37 ± 0.07
Cortical microstructure				
Ct.Th (mm)[A, B]	1.05 ± 0.24	1.2 ± 0.3	1.41 ± 0.19	1.52 ± 0.19
Ct.Po (%)	0.66 ± 0.48	0.63 ± 0.27	1.34 ± 1.14	1.37 ± 1.39
microFE				
Failure load (kN)[A, B, C]	3.46 ± 1.23	4.06 ± 1.64	6.32 ± 2.56	6.78 ± 2.47
Stiffness (kN/mm)[A, B, C]	61.9 ± 22.2	73.1 ± 30.8	115.7 ± 50.2	124.8 ± 48.3
App. modulus (kN/mm²)[A, B]	1.84 ± 0.43	2.05 ± 0.55	2.72 ± 0.26	2.8 ± 0.36

Significance from the mixed model:

A *p<0.05 OI vs Control intercept*

B *p<0.05 OI change over time*

C *p<0.05 Control change over time*

D *p<0.05 OI vs Control change over time.*

OI: Osteogenesis Imperfecta

Supplementary Table 3: Mean ± standard deviation for HR-pQCT measurements at the tibia diaphysis separated for OI and control groups, and timepoint. Statistical inferences are based on the mixed model for repeated measures (see the bottom of table)

Parameter	OI *(N=7)*		Control *(N=7)*	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**
Density (mg/cm³)				
Ct.vBMD[B, D]	992 ± 65.3	1018.9 ± 64.5	967.4 ± 39.3	952.7 ± 45.2
Area (mm²)				
Mr.Ar[A, C]	41.9 ± 25.6	42.2 ± 23.8	70.7 ± 20.8	74 ± 19
Ct.Ar[B, C]	128.1 ± 37.2	140.4 ± 37	179.9 ± 63.1	195.4 ± 67.7
Cortical microstructure				
Ct.Th (mm)[C]	4.21 ± 0.87	4.47 ± 0.7	4.55 ± 1.1	4.88 ± 1.22
Ct.Po (%)	1.39 ± 0.7	1.13 ± 0.84	1.37 ± 0.99	1.56 ± 1.08
microFE				
Failure load (kN)[B, C]	6.89 ± 1.45	7.69 ± 1.4	9.33 ± 3.24	9.91 ± 3.32
Stiffness (kN/mm)[B, C]	121.3 ± 26.1	136 ± 26	164.9 ± 59.4	177.1 ± 61.4
App. modulus (kN/mm²)	7 ± 0.59	7.37 ± 0.78	7.29 ± 0.34	7.02 ± 0.39

Significance from the mixed model:

A *p<0.05 OI vs Control intercept*

B *p<0.05 OI change over time*

C *p<0.05 Control change over time*

D *p<0.05 OI vs Control change over time.*

OI: Osteogenesis Imperfecta

Supplementary Table 4: Mean ± standard deviation for HR-pQCT measurements at the distal stack of the radius metaphysis separated for OI and control groups, and timepoint. Statistical inferences are based on the mixed model for repeated measures (see the bottom of table)

Parameter	OI *(N=7)*		Control *(N=7)*	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**
Density (mg/cm³)				
Tt.vBMD[C]	266.2 ± 91.9	274.9 ± 94.1	272.9 ± 47.5	290.2 ± 63.8
Tb.vBMD	144.7 ± 74.2	138 ± 66.6	143.1 ± 28.4	146.4 ± 25.2
Ct.vBMD[B]	718 ± 67.9	754.4 ± 94.3	753.7 ± 46.8	771.4 ± 64.5
Area (mm²)				
Tb.Ar	150.6 ± 43.5	158.2 ± 49.2	134.2 ± 32.7	139.9 ± 36.7
Ct.Ar[B, C]	40.7 ± 12.2	44.2 ± 12.8	37.8 ± 14.3	42.5 ± 16.7
Trabecular microstructure				
Tb.BV/TV (%)	18.2 ± 11.1	17.6 ± 9.4	19.7 ± 4.9	19.6 ± 4.5
Tb.N (1/mm)	1.2 ± 0.54	1.13 ± 0.51	1.39 ± 0.13	1.35 ± 0.11
Tb.Th (mm)	0.22 ± 0.03	0.22 ± 0.02	0.21 ± 0.01	0.22 ± 0.02
Tb.Sp (mm)	0.95 ± 0.38	1.02 ± 0.4	0.68 ± 0.08	0.69 ± 0.08
Tb.1/N.SD (mm)[A]	0.57 ± 0.36	0.66 ± 0.39	0.24 ± 0.04	0.26 ± 0.04
Cortical microstructure				
Ct.Th (mm)[B, C]	0.79 ± 0.2	0.85 ± 0.23	0.82 ± 0.21	0.91 ± 0.27
Ct.Po (%)	0.33 ± 0.23	0.3 ± 0.23	0.44 ± 0.24	0.5 ± 0.33
microFE				
Failure load (kN)[B, C]	1.81 ± 1.14	2.18 ± 1.1	2.25 ± 0.99	2.46 ± 1.03
Stiffness (kN/mm)	33.1 ± 21.2	39.2 ± 19.8	39.9 ± 18	43.7 ± 18.5
App. modulus (kN/mm²)	1.39 ± 0.68	1.62 ± 0.63	1.83 ± 0.38	1.99 ± 0.49

Significance from the mixed model:

A *p<0.05 OI vs Control intercept*

B *p<0.05 OI change over time*

C *p<0.05 Control change over time*

D *p<0.05 OI vs Control change over time.*

OI: Osteogenesis Imperfecta

Supplementary Table 5: Mean ± standard deviation for HR-pQCT measurements at the proximal stack of the radius metaphysis separated for OI and control groups, and timepoint. Statistical inferences are based on the mixed model for repeated measures (see the bottom of table)

Parameter	**OI** *(N=7)*		**Control** *(N=7)*	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**
Density (mg/cm³)				
Tt.vBMD[C]	357.4 ± 99.4	371.3 ± 117	416.2 ± 63.3	436.4 ± 83.6
Tb.vBMD	62 ± 50.4	65.4 ± 74.5	101.8 ± 36.5	108.3 ± 34.4
Ct.vBMD	868.5 ± 54.5	883.2 ± 62.8	851.4 ± 33.4	860.9 ± 54.4
Area (mm²)				
Tb.Ar	76.9 ± 29.1	83.1 ± 34	66.3 ± 22.3	70.1 ± 26
Ct.Ar[B, C]	42.5 ± 11.4	47.3 ± 14.7	47.6 ± 14.8	52.5 ± 15.5
Trabecular microstructure				
Tb.BV/TV (%)	8.8 ± 4.5	10.2 ± 8	14.9 ± 5.2	15.3 ± 5
Tb.N (1/mm)	0.68 ± 0.39	0.68 ± 0.53	1.03 ± 0.18	1.02 ± 0.17
Tb.Th (mm)	0.18 ± 0.02	0.19 ± 0.03	0.2 ± 0.02	0.21 ± 0.03
Tb.Sp (mm)[A, B]	1.98 ± 1.11	2.25 ± 1.31	0.94 ± 0.2	0.94 ± 0.21
Tb.1/N.SD (mm)[A]	0.94 ± 0.49	1.06 ± 0.62	0.38 ± 0.15	0.4 ± 0.2
Cortical microstructure				
Ct.Th (mm)[C]	1.17 ± 0.23	1.26 ± 0.31	1.36 ± 0.25	1.49 ± 0.31
Ct.Po (%)[A]	0.36 ± 0.26	0.29 ± 0.18	0.99 ± 0.45	1.03 ± 0.28
microFE				
Failure load (kN)[B, C]	2.09 ± 0.62	2.39 ± 0.85	2.63 ± 0.88	2.87 ± 0.87
Stiffness (kN/mm)[B, C]	36.3 ± 10.9	41.8 ± 15.3	45.1 ± 15.8	49.5 ± 15.8
App. modulus (kN/mm²)	3.17 ± 0.66	3.32 ± 0.77	3.79 ± 0.55	3.9 ± 0.68

Significance from the mixed model:

A *p<0.05 OI vs Control intercept*

B *p<0.05 OI change over time*

C *p<0.05 Control change over time*

D *p<0.05 OI vs Control change over time.*

OI: Osteogenesis Imperfecta

Supplementary Table 6: Mean ± standard deviation for HR-pQCT measurements at the radius diaphysis separated for OI and control groups, and timepoint. Statistical inferences are based on the mixed model for repeated measures (see the bottom of table)

Parameter	OI *(N=7)*		Control *(N=7)*	
	Baseline	**Follow-up**	**Baseline**	**Follow-up**
Density (mg/cm³)				
Ct.vBMD[B, D]	1032.4 ± 97.9	1061.2 ± 77	1025.2 ± 23.7	1026.6 ± 23.8
Area (mm²)				
Mr.Ar[A]	8.3 ± 7	8.6 ± 6.5	15.5 ± 5	15.9 ± 4
Ct.Ar	47.4 ± 8.6	50 ± 10.2	58 ± 16.4	61.6 ± 17.3
Cortical microstructure				
Ct.Th (mm)	2.75 ± 0.51	2.78 ± 0.39	2.6 ± 0.45	2.71 ± 0.45
Ct.Po (%)[B]	1.16 ± 1.11	0.69 ± 0.65	0.54 ± 0.48	0.61 ± 0.26
microFE				
Failure load (kN)[C]	2.44 ± 0.43	2.55 ± 0.54	3.1 ± 0.81	3.29 ± 0.9
Stiffness (kN/mm)[C]	40.6 ± 7.7	43.1 ± 9.5	51.6 ± 14.8	54.9 ± 15.8
App. modulus (kN/mm²)[A, B, D]	7.07 ± 0.5	6.73 ± 0.54	7.61 ± 0.28	7.76 ± 0.3

Significance from the mixed model:

A *p<0.05 OI vs Control intercept*

B *p<0.05 OI change over time*

C *p<0.05 Control change over time*

D *p<0.05 OI vs Control change over time.*

OI: Osteogenesis Imperfecta

Chapter 7. Summary:

7.1. General discussion and future directions:

High-resolution peripheral-quantitative computed tomography (HR-pQCT) is a non-invasive imaging tool for measuring bone microstructure and density in the peripheral skeleton. Combine with finite-element analysis (FEA), HR-pQCT can be used to estimate strength. Additionally, the high resolution of HR-pQCT offers the opportunity to quantify bone formation and resorption to better understand local changes in bone. Considering the extensive measurements that are provided using HR-pQCT, it is a promising tool in longitudinal studies. However, there are several challenges related to HR-pQCT application in longitudinal settings, resolution of which is important to improve its reliability.

The focus of this thesis was on HR-pQCT imaging of individuals with osteogenesis imperfecta (OI). OI or brittle bone disease, is a collagen-related genetic disorder resulting in phenotypes through out the body where collagen is found. However, the most devastating and important phenotype of OI is bone fragility(Marini, 2018). Despite its rarity with an occurrence of 1 in 15,000-20,000 births (Oakley & Reece, 2010), OI is a devastating condition that significantly affects the mobility of patients, and results in numerous fractures in their bones. Understanding the deteriorated strength in OI is challenging. First, OI is a heterogenous disease, which can occur by a variety of genetic mutations (at least 18 genetic mutations), as well as its varied phenotypical classification (type I - mild, type II - perinatally lethal, type III - severe and type IV – moderate). Furthermore, the severity of phenotypes also depends on whether the mutations result in quantitative defects of collagen production (i.e., reduced amount of normal type I collagen production) or qualitative defects (i.e., production of collagen molecules with altered structure). Quantitative defects result in lower bone mass with normal quality and are associated with milder osteogenesis imperfecta such as type I, while qualitative defects can cause the more severe types of OI such as type II, III, and IV.

Considering the complexities associated with OI, and the hierarchical and complex nature of bone, it is not surprising that our understanding of the underlying reasons behind diminished bone strength in individuals with OI is limited. Studying the low bone strength in adult population needs to be done at multiple length scales. The brittle and weak structure of bone in OI subjects originates from brittle bone tissue properties (tissue level properties). At this scale, lower strength

in OI bones is associated with compromised toughening mechanisms. At the meso- and macro-scales, deteriorated structural and geometrical properties are considered as reasons behind lower strength. Despite the importance of meso- and macro-scale properties, and the fact that HR-pQCT can be used to study these parameters in detail, HR-pQCT data for the OI population is limited.

The first two studies in this thesis (Chapters 4 and 5) were conducted as part of a phase 2b dose-finding multicenter clinical trial of an anabolic drug known as setrusumab in adults with OI. There are currently no curative treatments for OI, and treatment is focused on decreasing pain and fractures and increasing bone mass and mobility. setrusumab is among a recent group of sclerostin-inhibitory antibodies (Scl-Ab), which aim to increase bone formation by inhibiting sclerostin- a protein that reduces bone formation by inhibiting WNT/b-catenin signaling in osteoblasts and is predominantly secreted by osteocytes.

The first study (Chaper 4) focused on the short-term precision of HR-pQCT measurements. To properly interpret the observed changes in HR-pQCT outcomes over time, knowing the short-term precision of such measurements in crucial. For instance, in case of precision errors of 1% for a specific measurement, if the longitudinal changes over the time period of a study are 1%, it is not clear whether the observed changes are random errors, or actual changes. On the other hand, observing changes larger than 3% would suggest some degree of biological change. The first challenge that was discussed in this thesis was repositioning error in longitudinal HR-pQCT imaging that can lead to different bone volumes being assessed over time. The scanning procedure uses standard casts to secure the limb in the scanner, and hence minimize the repositioning error to small degrees. However, considering the realistic changes that occur in bone over time, even slight improvements in the precision of HR-pQCT measurements is valuable toward improving its reliability. For instance, reducing the precision error from 1% to 0.5% means that even changes as small as 1.5% can be interpreted as true biological changes (versus 3%).

Accordingly, proper alignment of the longitudinal scans using image registration can be beneficial to minimize errors. In this thesis, I performed a comprehensive comparison of different image registration methods using the same-day repeated scans from adults with osteogenesis imperfecta (OI), acquired during a multicenter clinical trial. Overall, our results suggested that image registration significantly improves the precision of HR-pQCT measurements, in line with other studies on other populations. Our extensive analyses also revealed that 3D image registration

slightly outperforms the default cross-sectional-area (CSA) registration. Even though the differences between these two methods were small, 3D registration is preferred as even small improvements increase the reliability of HR-pQCT measurements. Furthermore, 3D registration is robust to changes in bone area that may occur in longitudinal studies, whereas CSA registration is solely based on bone area. We also found that the 3D and CSA registration methods had comparable common volume between the repeated scans. This means that 3D registration does result in much smaller common volumes to account for rotational misalignment, although it could also mean that rotational misalignment was not substantial. We also showed that the accuracy of image registration decreased for scans with more motion. This is because motion effectively distorts the image while image registration assumes rigid transformations. As a result, even if registration has high numerical accuracy, the two bones sections will not be similar. We concluded that motion artifacts independently contribute to precision error, and this emphasizes the importance of acquiring scans with minimal motion. Our group is the first to report HR-pQCT precision errors in the OI population. We found that the short-term precision errors computed from repeated scans of adults with OI were comparable to those of healthy adults, suggesting that our results may be extended to other populations.

An important consideration for image registration is that the registration results in a reduction of the size of the analyzed region, which is even greater for microFE due to the flattening. Therefore, comparison of inter-subject HR-pQCT measures should not be performed on images that are trimmed by image registration. In other words, image registration is only proper and relevant for studying intra-subject changes. Similarly, for any cross-sectional analyses within longitudinal studies, where changes over time are not of interest, image registration is not recommended.

In this study, we provided extensive details about image registration, and shared our scripts on Github to facilitate the implementation of these methods. In our study, our repeated scans were collected as part of a multicenter clinical trial from different scanners and operators, rather than in a controlled research setting. This can make our results more translatable to the clinical settings. More studies using realistic data are needed to compare image registration methods. Reaching a consensus regarding the best image registration method in the community is also crucial, and

streamlining the analyses is required. Future clinical studies should implement proper image registration for improved reliability.

The reliability of using timelapse HR-pQCT as an imaging biomarker of bone formation and resorption was also studied in this thesis (Chapter 5). In recent years, time-lapse imaging using HR-pQCT has emerged as a non-invasive imaging biomarker of bone (re)modeling. This method works by comparing aligned images from two timepoints in a voxel-by-voxel fashion. In contrast to the density, microstructure, and strength measurements that indicate net overall changes in bone, timelapse HR-pQCT can additionally contribute to elucidating the cellular mechanisms behind the observed bone changes.

Aside from being non-invasive, this imaging biomarker has several other advantages compared to serum biomarkers of bone turnover and invasive histomorphometry of bone biopsies. For instance, serum biomarkers are non-site specific, meaning that they only provide an overview of bone turnover throughout the body, while timelapse imaging can reveal changes occurring at each bone site. Further, serum biomarkers only provide indirect information about bone formation and resorption via relative changes in the production or degradation of collagen and are sensitive to diurnal rhythms. On the other hand, timelapse imaging provides a direct assessment of mineralization and resorption of bone. Dynamic histomorphometry is highly invasive, labor intensive, limited to small bone sections, and painful for the donor, and it requires specialized training for the clinician performing the biopsy. Nevertheless, a challenge with timelapse HR-pQCT is the lack of consensus on what settings to use, and lack of validation. In this thesis, I presented a systematic analysis of different settings for timelapse HR-pQCT using the same-day repeated scans from adults with OI to identify the combination of image registration method, input image type, proper definition of the periosteal mask, and proper methods for minimal noise and error. Further, for the first time, I validated the timelapse analysis using a combination of repeated and longitudinal scans, and further evaluated the possible influence of confounding factors such as the magnitude of rotation angle or the drift of attenuation coefficient on timelapse outcomes.

Collectively, the results of these analyses revealed that the timelapse analysis performed using 3D registered grayscale input images, with dilated periosteal masks, and with noise reduction using Gaussian smoothing, had low errors, could identify longitudinal changes reliably, and was insensitive to rotation angle and attenuation coefficient drift. Finally, using the selected and

validated method, I found a positive dose-dependent effect of an anabolic drug (setrusumab) on bone formation and resorption, as well as net changes in bone at the distal radius and tibia of adults with OI. This study also showed the added value of timelapse analysis in clinical settings, as it provided more insight into how bone changes are achieved rather than only showing the net changes. We also found moderate to strong correlations between the net change in bone from timelapse analysis and changes in bone density and strength. In fact, these outcomes had similar patterns in their mean changes over time. This can not only be an indirect indication for the validity of the timelapse analysis, but also compliment the existing measurements by additionally showing "how" the overall changes were achieved.

Accordingly, combining timelapse analysis with the standard HR-pQCT analysis can make clinical trials more informative and reliable, potentially reducing or eliminating the need for invasive bone biopsies. It is important that future studies investigate timelapse HR-pQCT with the settings recommended in this thesis, to identify if similar results can be obtained in terms of the magnitude of errors, and sensitivity to longitudinal changes. Further, extended analysis may be developed in future studies to better identify region specific changes using timelapse analysis, as well as identifying statistical metrics that can increase the amount of information provided by timelapse analysis. Finally, associating the sites of bone formation and resorption with the mechanical strains throughout the bone volume using realistic finite-element models can further increase our understanding of bone mechanoregulation in humans.

In the last study (Chapter 6), I targeted the limited available data for children, which is intertwined with limited understanding of the challenges related to longitudinal HR-pQCT imaging in children. In particular, in children with OI, as only one study has reported HR-pQCT data for children with OI, and they provided no longitudinal data. We compared HR-pQCT measurements and their changes during 1-year between children with OI and age- and sex-matched healthy controls. In this study, we performed our analysis on several regions of the radius and tibia to achieve better understanding of differences between the OI and healthy groups. These regions include two consecutive image stacks at the distal metaphysis of the radius and tibia, as well as a single image stack at the diaphysis (~1/3 of bone length). Including the diaphyseal region can provide additional insight considering that the bone undergoes a different mode of loading at this region compared to the metaphysis. Further, bone morphology in entirely different at the

metaphysis and diaphysis. Studying the diaphysis is also important as long bone fractures in children with OI most often occur in this region.

Our data showed more significant differences at the tibia metaphysis compared to the radius, which was mostly driven by trabecular deterioration. We also found that the changes over the 1-year period were mostly comparable between the OI and control groups at both the radius and tibia. At the diaphysis, we found that OI bones had smaller area than controls.

Our results suggest that a 1-year period is not sufficient to reveal the differences between growth in the OI and healthy children, although our small sample size could also be the reason behind this finding. Future studies that follow children with OI and controls over a longer period of time can provide additional insight into the natural history of the distal radius and tibia of children with OI. Future studies are also needed to increase the amount of available data for this population and to investigate the use of alternative segmentation methods in children. This is because the average thickness of trabeculae in children tends to be smaller than those of adults, a segmentation protocol that most accurately preserves the fine trabeculae might be beneficial. So far, the same binarization method has been used in children as well.

It is noteworthy that all of the participants with OI were taking bisphosphonates. This could affect the outcomes of the studies in two ways that could reduce the differences between the OI and control groups. First, it has been shown that bisphosphonates increase bone mass. Secondly, bisphosphonates results in the appearance of sclerotic lines, which are horizontal trabeculae containing some degree of cartilage thought to be caused by the temporary interruption of growth plate cartilage resorption. In fact, we observed sclerotic lines in all the OI scans. As such, we anticipate that the differences between untreated OI bones and healthy children will be larger that what is suggested in this study. Future studies can study this effect to some extent by removing the sclerotic lines from the scans, and comparing the OI and control groups on the corrected scans. The data presented in this study included a small sample size of 7 per group. Thus, some of the results are prone to being underpowered or having inflated significance. Despite this, some of the effect sizes were large compared to their variance. These results are part of an ongoing study that will eventually increase the sample size to 20 per group.

Additionally, we showed that longitudinal image registration in growing bones using rigid registration algorithm is not adequate. This means that the longitudinal imaging of children

includes some degree of error due to repositioning error that cannot be eliminated post imaging. Finally, we illustrated the limitation associated with scanning consecutive image stacks, namely misalignment between stacks, and discussed how it influences HR-pQCT outcomes.

Overall, this thesis highlights some of the under-appreciated challenges in longitudinal HR-pQCT imaging, with a focus on realistic data from a clinically important patient population. Methodological solutions with detailed explanations were also developed to resolve or minimize such challenges. It is important that the bone research and HR-pQCT community carefully consider these technical aspects of bone analysis, and come to an agreement on a standardized analysis, that can be implemented as automated workflows for more consistent findings across studies. This is of considerable importance considering the value of high quality data that can results in better high-level decision making in future.

7.2. Conclusions:

The purpose of this thesis was to discuss several challenges related to longitudinal HR-pQCT studies in both the adult and pediatric OI populations, and to provide methodological approaches for reliable HR-pQCT image analysis. In Chapter 4, repositioning error was discussed as a factor reducing the precision of longitudinal HR-pQCT measurements, and 3D image registration was shown to be most beneficial in improving the precision of HR-pQCT measurements in adults with OI. In Chapter 5, timelapse HR-pQCT as an imaging biomarker of bone formation and resorption was investigated for settings that would results in the most reliable outcomes, validated, and applied to the data from a multicenter clinical trial on adults with OI to show positive dose-dependent effects of an anabolic drug on formation and resorption at the radius and tibia. In Chapter 6, challenges specific to longitudinal HR-pQCT imaging in children was discussed, namely image registration, where I showed that the common rigid image registration is not feasible. Also, the first set of longitudinal HR-pQCT data for children with OI was reported. We compared HR-pQCT measurements and their changes during 1-year between children with OI and age- and sex-matched healthy controls and showed that deteriorated microstructure as the main difference between the two groups.

The following specific conclusions were drawn:

Study #1:

- Image registration significantly improves the precision of HR-pQCT measurements of density, morphology, and strength, regardless of the registration method
- For density and morphology outcomes, 3D registered slightly outperformed the CSA method, and is robust to changes in bone area, hence preferred
- Similarly, for microFE, 3D-TB or MA registration methods slightly outperformed the standard 3D and CSA methods
- Registration only partially corrects motion-related error, thus, obtaining scans with minimal motion is important
- Precision errors for adults with OI were comparable to those of other populations

Study #2:

- 3D registration and matched angle registration provided almost identical results for timelapse analysis
- 3D registration is preferred for timelapse analysis due to being more straightforward
- Using binary input images for timelapse analysis results in high errors that appear all across bone surface
- Considering the expected realistic changes in bone, the high errors associated with the binary method suggest that it is not a reliable method
- For both XCT and XCT2, the grayscale method with a density threshold of 200 mgHA/cm^3, and a cluster size of 0 resulted formation/resorption volumes approaching zero, negligible effect of increasing the density threshold and cluster size, and negligible noise when combined with Gaussian noise reduction (Gaussian sigma of 0.8 for XCT and 1.2 for XCT2)
- Using Gaussian smoothing for noise reduction is preferred as it removes the need for applying a minimum cluster size, which is somewhat arbitrary

- With the validation study, we showed that while there was negligible formation and resorption between the same-day repeated scans, matching either of the repeated scans with the same baseline scan result in similar regions identified as formation and resorption, confirming that the identified regions were not due to random noise, rather were true biological changes
- Our results did not show enough evidence for a relationship between the rotation angle between scans, and the magnitude of bone formation and resorption
- We found that the effect of drift in attenuation coefficient within the allowed range over time has a negligible to small effect on the outcomes of timelapse analysis
- There is a positive dose-dependent effect of setrusumab on bone formation and resorption, as well improved net changes in bone at the radius and tibia of adults with OI

Study #3:

- At the metaphysis, differences between the children with OI and healthy controls are more prominent at the tibia compared to the radius
- At the tibia metaphysis, the baseline measurements were different between children with OI and controls for Tt.vBMD, and most of trabecular outcomes, except thickness and area, and microFE outcomes
- Changes the 1-year period were similar between the OI and control groups, and were mostly for microFE and cortical measurements
- At the diaphysis, the main difference between the OI and healthy bones was larger bone area in the healthy group
- Cortical bone density was similar between the OI and control groups at the diaphysis
- Rigid 3D image registration is not capable of properly aligning scans of growing children in the absence of landmarks such as sclerotic lines
- The misalignment between consecutive image stacks is not scanner-related and is caused by limb movement

- For average-based HR-pQCT outcomes including density, geometry, and microstructure, misalignment between the two stacks is not problematic. However, microFE is sensitive to misalignment. In those situations, either the double stacks can be analyzed separately or misalignment may be corrected
- If alignment of consecutive image stacks is necessary, leaving an overlap between the stacks is recommended

References

Atkins, P. R., Stock, K., Ohs, N., Collins, C. J., Horling, L., Benedikt, S., Degenhart, G., Lippuner, K., Blauth, M., Christen, P., & Müller, R. (2021). Formation Dominates Resorption With Increasing Mineralized Density and Time Postfracture in Cortical but Not Trabecular Bone: A Longitudinal HRpQCT Imaging Study in the Distal Radius. *JBMR Plus*, *5*(6), 1–17. https://doi.org/10.1002/jbm4.10493

Beaupré, G. S., Orr, T. E., & Carter, D. R. (1990). An approach for time-dependent bone modeling and remodeling—theoretical development. *Journal of Orthopaedic Research*. https://doi.org/10.1002/jor.1100080506

Bellido, T., Plotkin, L. I., & Bruzzaniti, A. (2019). Basic and Applied Bone Biology. In *Basic and Applied Bone Biology*. https://doi.org/10.1016/c2016-0-03781-6

Birkhold, A. I., Razi, H., Duda, G. N., Weinkamer, R., Checa, S., & Willie, B. M. (2014a). Mineralizing surface is the main target of mechanical stimulation independent of age: 3D dynamic in vivo morphometry. *Bone*, *66*, 15–25. https://doi.org/10.1016/j.bone.2014.05.013

Birkhold, A. I., Razi, H., Duda, G. N., Weinkamer, R., Checa, S., & Willie, B. M. (2014b). The influence of age on adaptive bone formation and bone resorption. *Biomaterials*, *35*(34), 9290–9301. https://doi.org/10.1016/j.biomaterials.2014.07.051

Bonewald, L. F. (2011). The amazing osteocyte. *Journal of Bone and Mineral Research*. https://doi.org/10.1002/jbmr.320

Boskey, A. L. (2013). Bone composition: relationship to bone fragility and antiosteoporotic drug effects. *BoneKEy Reports*. https://doi.org/10.1038/bonekey.2013.181

Boutroy, S., Bouxsein, M. L., Munoz, F., & Delmas, P. D. (2005). In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography. *Journal of Clinical Endocrinology and Metabolism*, *90*(12), 6508–6515. https://doi.org/10.1210/jc.2005-1258

Brand, R. A. (2011). The classic: The architecture of the trabecular bone (tenth contribution on the mechanics of the human skeletal framework). *Clinical Orthopaedics and Related Research*. https://doi.org/10.1007/s11999-011-2042-4

Brunet, S. C., Kuczynski, M. T., Bhatla, J. L., Lemay, S., Pauchard, Y., Salat, P., Barnabe, C., & Manske, S. L. (2020). The utility of multi-stack alignment and 3D longitudinal image

registration to assess bone remodeling in rheumatoid arthritis patients from second generation HR-pQCT scans. *BMC Medical Imaging*, *20*(1), 1–10. https://doi.org/10.1186/s12880-020-00437-8

Burghardt, A. J., Buie, H. R., Laib, A., Majumdar, S., & Boyd, S. K. (2010). Reproducibility of direct quantitative measures of cortical bone microarchitecture of the distal radius and tibia by HR-pQCT. *Bone*. https://doi.org/10.1016/j.bone.2010.05.034

Burghardt, A. J., Kazakia, G. J., Sode, M., De Papp, A. E., Link, T. M., & Majumdar, S. (2010). A longitudinal HR-pQCT study of alendronate treatment in postmenopausal women with low bone density: Relations among density, cortical and trabecular microarchitecture, biomechanics, and bone turnover. *Journal of Bone and Mineral Research*, *25*(12), 2282–2295. https://doi.org/10.1002/jbmr.157

Burr, D. B. (2002). The contribution of the organic matrix to bone's material properties. *Bone*. https://doi.org/10.1016/S8756-3282(02)00815-3

Burt, L. A., Bhatla, J. L., Hanley, D. A., & Boyd, S. K. (2017). Cortical porosity exhibits accelerated rate of change in peri- compared with post-menopausal women. *Osteoporosis International*, *28*(4), 1423–1431. https://doi.org/10.1007/s00198-016-3900-4

Burt, Lauren A., Hanley, D. A., & Boyd, S. K. (2017). Cross-sectional Versus Longitudinal Change in a Prospective HR-pQCT Study. *Journal of Bone and Mineral Research*, *32*(7), 1505–1513. https://doi.org/10.1002/jbmr.3129

Burt, Lauren A., Manske, S. L., Hanley, D. A., & Boyd, S. K. (2018). Lower Bone Density, Impaired Microarchitecture, and Strength Predict Future Fragility Fracture in Postmenopausal Women: 5-Year Follow-up of the Calgary CaMos Cohort. *Journal of Bone and Mineral Research*, *33*(4), 589–597. https://doi.org/10.1002/jbmr.3347

Caksa, S., Yuan, A., Rudolph, S. E., Yu, E. W., Popp, K. L., & Bouxsein, M. L. (2019). Influence of soft tissue on bone density and microarchitecture measurements by high-resolution peripheral quantitative computed tomography. *Bone*, *124*(April), 47–52. https://doi.org/10.1016/j.bone.2019.04.008

Carriero, A., Zimmermann, E. A., Paluszny, A., Tang, S. Y., Bale, H., Busse, B., Alliston, T., Kazakia, G., Ritchie, R. O., & Shefelbine, S. J. (2014). How tough is brittle bone? Investigating osteogenesis imperfecta in mouse bone. *Journal of Bone and Mineral*

Research, *29*(6), 1392–1401. https://doi.org/10.1002/jbmr.2172

Checa, S., Birkhold, A. I., Weinkamer, R., Zaslansky, P., Willie, B. M., Razi, H., & Duda, G. N. (2014). Skeletal maturity leads to a reduction in the strain magnitudes induced within the bone: A murine tibia study. *Acta Biomaterialia*, *13*, 301–310. https://doi.org/10.1016/j.actbio.2014.11.021

Christen, P., Boutroy, S., Ellouz, R., Chapurlat, R., & Van Rietbergen, B. (2018). Least-detectable and age-related local in vivo bone remodelling assessed by time-lapse HR-pQCT. *PLoS ONE*, *13*(1), 1–11. https://doi.org/10.1371/journal.pone.0191369

Christen, P., Ito, K., Ellouz, R., Boutroy, S., Sornay-Rendu, E., Chapurlat, R. D., & Van Rietbergen, B. (2014). Bone remodelling in humans is load-driven but not lazy. *Nature Communications*, *5*, 1–5. https://doi.org/10.1038/ncomms5855

Christen, P., & Müller, R. (2017). In vivo Visualisation and Quantification of Bone Resorption and Bone Formation from Time-Lapse Imaging. *Current Osteoporosis Reports*, *15*(4), 311–317. https://doi.org/10.1007/s11914-017-0372-1

Collins, C. J., Atkins, P. R., Ohs, N., Blauth, M., Lippuner, K., & Müller, R. (2022). Clinical observation of diminished bone quality and quantity through longitudinal HR-pQCT-derived remodeling and mechanoregulation. *Scientific Reports*, *12*(1), 1–13. https://doi.org/10.1038/s41598-022-22678-z

de Bakker, C. M. J., Altman, A. R., Li, C., Tribble, M. B., Lott, C., Tseng, W. J., & Liu, X. S. (2016). Minimizing Interpolation Bias and Precision Error in In Vivo μCT-Based Measurements of Bone Structure and Dynamics. *Annals of Biomedical Engineering*, *44*(8), 2518–2528. https://doi.org/10.1007/s10439-015-1527-9

Du, J., Hartley, C., Brooke-Wavell, K., Paggiosi, M. A., Walsh, J. S., Li, S., & Silberschmidt, V. V. (2020). High-impact exercise stimulated localised adaptation of microarchitecture across distal tibia in postmenopausal women. *Osteoporosis International*. https://doi.org/10.1007/s00198-020-05714-4

Ellouz, R., Chapurlat, R., van Rietbergen, B., Christen, P., Pialat, J. B., & Boutroy, S. (2014). Challenges in longitudinal measurements with HR-pQCT: Evaluation of a 3D registration method to improve bone microarchitecture and strength measurement reproducibility. *Bone*, *63*, 147–157. https://doi.org/10.1016/j.bone.2014.03.001

Engelke, K., Stampa, B., Timm, W., Dardzinski, B., De Papp, A. E., Genant, H. K., & Fuerst, T. (2012). Short-term in vivo precision of BMD and parameters of trabecular architecture at the distal forearm and tibia. *Osteoporosis International*, *23*(8), 2151–2158. https://doi.org/10.1007/s00198-011-1829-1

Fennimore, D. J., Digby, M., Paggiosi, M., Arundel, P., Bishop, N. J., Dimitri, P., & Offiah, A. C. (2020). High-resolution peripheral quantitative computed tomography in children with osteogenesis imperfecta. *Pediatric Radiology*, *3*(XtremeCT II). https://doi.org/10.1007/s00247-020-04736-8

Folkestad, L., Hald, J. D., Hansen, S., Gram, J., Langdahl, B., Abrahamsen, B., & Brixen, K. (2012). Bone geometry, density, and microarchitecture in the distal radius and tibia in adults with osteogenesis imperfecta type i assessed by high-resolution pQCT. *Journal of Bone and Mineral Research*, *27*(6), 1405–1412. https://doi.org/10.1002/jbmr.1592

Fratzl-Zelman, N., Schmidt, I., Roschger, P., Glorieux, F. H., Klaushofer, K., Fratzl, P., Rauch, F., & Wagermaier, W. (2014). Mineral particle size in children with osteogenesis imperfecta type I is not increased independently of specific collagen mutations. *Bone*, *60*, 122–1128. https://doi.org/10.1016/j.bone.2013.11.023

Fratzl, P., Gupta, H. S., Fischer, F. D., & Kolednik, O. (2007). Hindered crack propagation in materials with periodically varying young's modulus - Lessons from biological materials. *Advanced Materials*. https://doi.org/10.1002/adma.200602394

Fratzl, P., & Weinkamer, R. (2007). Nature's hierarchical materials. In *Progress in Materials Science*. https://doi.org/10.1016/j.pmatsci.2007.06.001

Frost, H. M. (1987). Bone "mass" and the "mechanostat": A proposal. *The Anatomical Record*. https://doi.org/10.1002/ar.1092190104

Frost, H. M. (1994). Wolff's Law and bone's structural adaptations to mechanical usage: an overview for clinicians. In *Angle Orthodontist* (Vol. 64, Issue 3, pp. 175–188). https://doi.org/10.1043/0003-3219(1994)064<0175:WLABSA>2.0.CO;2

Gabel, L., Liphardt, A. M., Hulme, P. A., Heer, M., Zwart, S. R., Sibonga, J. D., Smith, S. M., & Boyd, S. (2021). Pre-flight exercise and bone metabolism predict unloading-induced bone loss due to spaceflight. *British Journal of Sports Medicine*, *1*, 1–9. https://doi.org/10.1136/bjsports-2020-103602

Geusens, P., Chapurlat, R., Schett, G., Ghasem-Zadeh, A., Seeman, E., De Jong, J., & Van Den Bergh, J. (2014). High-resolution in vivo imaging of bone and joints: A window to microarchitecture. *Nature Reviews Rheumatology*, *10*(5), 304–313. https://doi.org/10.1038/nrrheum.2014.23

Glorieux, F. H., Devogelaer, J. P., Durigova, M., Goemaere, S., Hemsley, S., Jakob, F., Junker, U., Ruckle, J., Seefried, L., & Winkle, P. J. (2017). BPS804 Anti-Sclerostin Antibody in Adults With Moderate Osteogenesis Imperfecta: Results of a Randomized Phase 2a Trial. *Journal of Bone and Mineral Research*, *32*(7), 1496–1504. https://doi.org/10.1002/jbmr.3143

Glüer, C. C., Blake, G., Lu, Y., Blunt, B. A., Jergas, M., & Genant, H. K. (1995). Accurate assessment of precision errors: How to measure the reproducibility of bone densitometry techniques. *Osteoporosis International*, *5*(4), 262–270. https://doi.org/10.1007/BF01774016

Grafe, I., Alexander, S., Yang, T., Lietman, C., Homan, E. P., Munivez, E., Chen, Y., Jiang, M. M., Bertin, T., Dawson, B., Asuncion, F., Ke, H. Z., Ominsky, M. S., & Lee, B. (2016). Sclerostin Antibody Treatment Improves the Bone Phenotype of Crtap -/- Mice, a Model of Recessive Osteogenesis Imperfecta. *Journal of Bone and Mineral Research*. https://doi.org/10.1002/jbmr.2776

Gupta, H. S., & Zioupos, P. (2008). Fracture of bone tissue: The “hows” and the “whys.” *Medical Engineering and Physics*. https://doi.org/10.1016/j.medengphy.2008.09.007

Gupta, Himadri S., Seto, J., Wagermaier, W., Zaslansky, P., Boesecke, P., & Fratzl, P. (2006). Cooperative deformation of mineral and collagen in bone at the nanoscale. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.0604237103

Gupta, Himadri S., Wagermaier, W., Zickler, G. A., Aroush, D. R. Ben, Funari, S. S., Roschger, P., Wagner, H. D., & Fratzl, P. (2005). Nanoscale deformation mechanisms in bone. *Nano Letters*. https://doi.org/10.1021/nl051584b

Hald, J. D., Folkestad, L., Harsløf, T., Lund, A. M., Duno, M., Jensen, J. B., Neghabat, S., Brixen, K., & Langdahl, B. (2016). Skeletal phenotypes in adult patients with osteogenesis imperfecta—correlations with COL1A1/COL1A2 genotype and collagen structure. *Osteoporosis International*, *27*(11), 3331–3341. https://doi.org/10.1007/s00198-016-3653-0

Hosseinitabatabaei, S., Mikolajewicz, N., Zimmermann, E. A., Rummler, M., Steyn, B., Julien, C., Rauch, F., & Willie, B. M. (2022). 3D image registration marginally improves the precision of HR-pQCT measurements compared to cross-sectional-area registration in adults with osteogenesis imperfecta. *Journal of Bone and Mineral Research*, *00*(00), 1–17. https://doi.org/10.1002/jbmr.4541

Indermaur, M., Casari, D., Kochetkova, T., Peruzzi, C., Zimmermann, E., Rauch, F., Willie, B., Michler, J., Schwiedrzik, J., & Zysset, P. (2021). Compressive Strength of Iliac Bone ECM Is Not Reduced in Osteogenesis Imperfecta and Increases With Mineralization. *Journal of Bone and Mineral Research*, *00*(00), 1–12. https://doi.org/10.1002/jbmr.4286

Joan C Marini, Antonella Forlino, Hans Peter Bächinger, Nick J Bishop, Peter H Byers, Anne De Paepe, Francois Fassier, Nadja Fratzl-Zelman, Kenneth M Kozloff, Deborah Krakow, Kathleen Montpetit, O. S. (2015). Osteogenesis imperfecta. *Nature Reviews Disease Primers*, *3*, 521–524. https://doi.org/10.1007/978-1-4471-6572-9_117

Kawalilak, C. E., Kontulainen, S. A., Amini, M. A., Lanovaz, J. L., Olszynski, W. P., & Johnston, J. D. (2016). In vivo precision of three HR-pQCT-derived finite element models of the distal radius and tibia in postmenopausal women. *BMC Musculoskeletal Disorders*, *17*(1), 389–400. https://doi.org/10.1186/s12891-016-1238-x

Kocijan, R., Muschitz, C., Haschka, J., Hans, D., Nia, A., Geroldinger, A., Ardelt, M., Wakolbinger, R., & Resch, H. (2015). Bone structure assessed by HR-pQCT, TBS and DXL in adult patients with different types of osteogenesis imperfecta. *Osteoporosis International*. https://doi.org/10.1007/s00198-015-3156-4

Koester, K. J., Ager, J. W., & Ritchie, R. O. (2008). The true toughness of human cortical bone measured with realistically short cracks. *Nature Materials*. https://doi.org/10.1038/nmat2221

Kolednik, O., Predan, J., Fischer, F. D., & Fratzl, P. (2014). Improvements of strength and fracture resistance by spatial material property variations. *Acta Materialia*. https://doi.org/10.1016/j.actamat.2014.01.034

Kolednik, Otmar, Predan, J., Fischer, F. D., & Fratzl, P. (2011). Bioinspired design criteria for damage-resistant materials with periodically varying microstructure. *Advanced Functional Materials*. https://doi.org/10.1002/adfm.201100443

Laib, A., & Rüegsegger, P. (1999). Calibration of trabecular bone structure measurements of in vivo three-dimensional peripheral quantitative computed tomography with 28-µm-resolution microcomputed tomography. *Bone*, *24*(1), 35–39. https://doi.org/10.1016/S8756-3282(98)00159-8

Land, C., Rauch, F., Munns, C. F., Sahebjam, S., & Glorieux, F. H. (2006). Vertebral morphometry in children and adolescents with osteogenesis imperfecta: Effect of intravenous pamidronate treatment. *Bone*. https://doi.org/10.1016/j.bone.2006.04.004

Landis, W. J., Song, M. J., Leith, A., McEwen, L., & McEwen, B. F. (1993). Mineral and organic matrix interaction in normally calcifying tendon visualized in three dimensions by high-voltage electron microscopic tomography and graphic image reconstruction. *Journal of Structural Biology*. https://doi.org/10.1006/jsbi.1993.1003

Lee, B., Smith, R., Orwoll, E., Wang, Y., Shapiro, J., Nicol, L., Sloan, J., & Nagamani, S. C. (2017). Serum Sclerostin Levels in Adults With Osteogenesis Imperfecta: Comparison With Normal Individuals and Response to Teriparatide Therapy. *Journal of Bone and Mineral Research*, *33*(2), 307–315. https://doi.org/10.1002/jbmr.3312

MacNeil, J A, & Boyd, S. K. (2008). Improved reproducibility of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Medical Engineering & Physics*, *30*(6), 792–799. https://doi.org/10.1016/j.medengphy.2007.11.003; 10.1016/j.medengphy.2007.11.003

MacNeil, Joshua A, & Boyd, S. K. (2007). Accuracy of high-resolution peripheral quantitative computed tomography for measurement of bone quality. *Medical Engineering & Physics*, *29*(10), 1096–1105. https://doi.org/10.1016/j.medengphy.2006.11.002

Mancuso, M. E., & Troy, K. L. (2020). Relating Bone Strain to Local Changes in Radius Microstructure following 12 Months of Axial Forearm Loading in Women. *Journal of Biomechanical Engineering*, *142*(11), 1–11. https://doi.org/10.1115/1.4048232

Manske, S. L., Zhu, Y., Sandino, C., & Boyd, S. K. (2015). Human trabecular bone microarchitecture can be assessed independently of density with second generation HR-pQCT. *Bone*, *79*, 213–221. https://doi.org/10.1016/j.bone.2015.06.006

Marini, J. C. (2018). Osteogenesis imperfecta. In *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*. https://doi.org/10.1002/9781119266594.ch113

Marom, R., Lee, Y. C., Grafe, I., & Lee, B. (2016). Pharmacological and biological therapeutic strategies for osteogenesis imperfecta. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics*, *172*(4), 367–383. https://doi.org/10.1002/ajmg.c.31532

McClung, M. R., Grauer, A., Boonen, S., Bolognese, M. A., Brown, J. P., Diez-Perez, A., Langdahl, B. L., Reginster, J. Y., Zanchetta, J. R., Wasserman, S. M., Katz, L., Maddox, J., Yang, Y. C., Libanati, C., & Bone, H. G. (2014). Romosozumab in postmenopausal women with low bone mineral density. *New England Journal of Medicine*. https://doi.org/10.1056/NEJMoa1305224

Mikolajewicz, N., Bishop, N., Burghardt, A. J., Folkestad, L., Hall, A., Kozloff, K. M., Lukey, P. T., Molloy-Bland, M., Morin, S. N., Offiah, A. C., Shapiro, J., van Rietbergen, B., Wager, K., Willie, B. M., Komarova, S. V, & Glorieux, F. H. (2019). HR-pQCT Measures of Bone Microarchitecture Predict Fracture: Systematic Review and Meta-analysis. *Journal of Bone and Mineral Research : The Official Journal of the American Society for Bone and Mineral Research*. https://doi.org/10.1002/jbmr.3901

Mortier, G. R., Cohn, D. H., Cormier-Daire, V., Hall, C., Krakow, D., Mundlos, S., Nishimura, G., Robertson, S., Sangiorgi, L., Savarirayan, R., Sillence, D., Superti-Furga, A., Unger, S., & Warman, M. L. (2019). Nosology and classification of genetic skeletal disorders: 2019 revision. *American Journal of Medical Genetics, Part A*, *179*(12), 2393–2419. https://doi.org/10.1002/ajmg.a.61366

Mueller, T. L., Stauber, M., Kohler, T., Eckstein, F., Müller, R., & van Lenthe, G. H. (2009). Non-invasive bone competence analysis by high-resolution pQCT: An in vitro reproducibility study on structural and mechanical properties at the human radius. *Bone*, *44*(2), 364–371. https://doi.org/10.1016/j.bone.2008.10.045

Nalla, R. K., Kinney, J. H., & Ritchie, R. O. (2003). Mechanistic fracture criteria for the failure of human cortical bone. In *Nature Materials*. https://doi.org/10.1038/nmat832

Nalla, R. K., Kruzic, J. J., Kinney, J. H., Balooch, M., Ager, J. W., & Ritchie, R. O. (2006). Role of microstructure in the aging-related deterioration of the toughness of human cortical bone. *Materials Science and Engineering C*. https://doi.org/10.1016/j.msec.2005.08.021

Nijhuis, W. H., Weinans, H. H., Bank, R. A., Hvid, I., Eastwood, D. M., Allgrove, J., & Sakkers, R. J. (2019). Current concepts in osteogenesis imperfecta: bone structure, biomechanics and

medical management. *Journal of Children's Orthopaedics*, *13*(1), 1–11. https://doi.org/10.1302/1863-2548.13.180190

Nishiyama, K. K., Pauchard, Y., Nikkel, L. E., Iyer, S., Zhang, C., McMahon, D. J., Cohen, D., Boyd, S. K., Shane, E., & Nickolas, T. L. (2015). Longitudinal HR-pQCT and image registration detects endocortical bone loss in kidney transplantation patients. *Journal of Bone and Mineral Research*, *30*(3), 456–463. https://doi.org/10.1002/jbmr.2358

Nishiyama, K. K., & Shane, E. (2013). Clinical imaging of bone microarchitecture with HR-pQCT. *Current Osteoporosis Reports*, *11*(2), 147–155. https://doi.org/10.1007/s11914-013-0142-7

Oakley, I., & Reece, L. P. (2010). Anesthetic implications for the patient with osteogenesis imperfecta. In *AANA Journal*.

Paggiosi, M. A., Eastell, R., & Walsh, J. S. (2014). Precision of high-resolution peripheral quantitative computed tomography measurement variables: Influence of gender, examination site, and age. *Calcified Tissue International*, *94*(2), 191–201. https://doi.org/10.1007/s00223-013-9798-3

Parfitt, A. M. (1994). Osteonal and hemi-osteonal remodeling: The spatial and temporal framework for signal traffic in adult human bone. *Journal of Cellular Biochemistry*. https://doi.org/10.1002/jcb.240550303

Peterlik, H., Roschger, P., Klaushofer, K., & Fratzl, P. (2006). From brittle to ductile fracture of bone. *Nature Materials*. https://doi.org/10.1038/nmat1545

Peters, M., van den Bergh, J. P., Geusens, P., Scharmga, A., Loeffen, D., Weijers, R., van Rietbergen, B., & van Tubergen, A. (2019). Prospective Follow-Up of Cortical Interruptions, Bone Density, and Micro-structure Detected on HR-pQCT: A Study in Patients with Rheumatoid Arthritis and Healthy Subjects. *Calcified Tissue International*, *104*(6), 571–581. https://doi.org/10.1007/s00223-019-00523-2

Plachel, F., Renner, U., Kocijan, R., Muschitz, C., Lomoschitz, F., & Resch, H. (2015). Osteogenesis imperfecta Typ III und hypogonadotroper Hypogonadismus führen zu massiven Knochenverlust: ein Fallbericht. *Wiener Medizinische Wochenschrift*, *165*(13–14), 285–289. https://doi.org/10.1007/s10354-015-0367-4

Plett, R. M., Kemp, T. D., Burt, L. A., Billington, E. O., Hanley, D. A., & Boyd, S. K. (2021).

Using 3D image registration to maximize the reproducibility of longitudinal bone strength assessment by HR-pQCT and finite element analysis.

Plotkin, H., Rauch, F., Bishop, N. J., Montpetit, K., Ruck-Gibis, J., Travers, R., & Glorieux, F. H. (2000). Pamidronate treatment of severe osteogenesis imperfecta in children under 3 years of age. *Journal of Clinical Endocrinology and Metabolism*. https://doi.org/10.1210/jc.85.5.1846

Rauch, F., Travers, R., Munns, C., & Glorieux, F. H. (2004). Clinical vignette - Sclerotic metaphyseal lines in a child treated with pamidronate: Histomorphometric analysis. *Journal of Bone and Mineral Research*. https://doi.org/10.1359/JBMR.040303

Recker, R. R., Benson, C. T., Matsumoto, T., Bolognese, M. A., Robins, D. A., Alam, J., Chiang, A. Y., Hu, L., Krege, J. H., Sowa, H., Mitlak, B. H., & Myers, S. L. (2015). A randomized, double-blind phase 2 clinical trial of blosozumab, a sclerostin antibody, in postmenopausal women with low bone mineral density. *Journal of Bone and Mineral Research*. https://doi.org/10.1002/jbmr.2351

Reznikov, N., Shahar, R., & Weiner, S. (2014). Bone hierarchical structure in three dimensions. *Acta Biomaterialia*, *10*(9), 3815–3826. https://doi.org/10.1016/j.actbio.2014.05.024

Rho, J. Y., Kuhn-Spearing, L., & Zioupos, P. (1998). Mechanical properties and the hierarchical structure of bone. *Medical Engineering and Physics*, *20*(2), 92–102. https://doi.org/10.1016/S1350-4533(98)00007-1

Rinaldi, G., Wisniewski, C. A., Setty, N. G., & LeBoff, M. S. (2011). Peripheral Quantitative Computed Tomography: Optimization of Reproducibility Measures of Bone Density, Geometry, and Strength at the Radius and Tibia. *Journal of Clinical Densitometry*. https://doi.org/10.1016/j.jocd.2011.05.002

Ritchie, R. O. (2011). The conflicts between strength and toughness. *Nature Materials*. https://doi.org/10.1038/nmat3115

Robling, A. G., Daly, R., Fuchs, R. K., & Burr, D. B. (2019). Basic and Applied Bone Biology. In *Basic and Applied Bone Biology*. https://doi.org/10.1016/c2016-0-03781-6

Robling, A. G., Niziolek, P. J., Baldridge, L. A., Condon, K. W., Allen, M. R., Alam, I., Mantila, S. M., Gluhak-Heinrich, J., Bellido, T. M., Harris, S. E., & Turner, C. H. (2008). Mechanical stimulation of bone in vivo reduces osteocyte expression of Sost/sclerostin.

Journal of Biological Chemistry. https://doi.org/10.1074/jbc.M705092200

Rolvien, T., Stürznickel, J., Schmidt, F. N., Butscheidt, S., Schmidt, T., Busse, B., Mundlos, S., Schinke, T., Kornak, U., Amling, M., & Oheim, R. (2018). Comparison of Bone Microarchitecture Between Adult Osteogenesis Imperfecta and Early-Onset Osteoporosis. *Calcified Tissue International*, *103*(5), 512–521. https://doi.org/10.1007/s00223-018-0447-8

Roschger, A., Roschger, P., Keplingter, P., Klaushofer, K., Abdullah, S., Kneissel, M., & Rauch, F. (2014). Effect of sclerostin antibody treatment in a mouse model of severe osteogenesis imperfecta. *Bone*. https://doi.org/10.1016/j.bone.2014.06.015

Samelson, E. J., Broe, K. E., Xu, H., Yang, L., Boyd, S., Biver, E., Szulc, P., Adachi, J., Amin, S., Atkinson, E., Berger, C., Burt, L., Chapurlat, R., Chevalley, T., Ferrari, S., Goltzman, D., Hanley, D. A., Hannan, M. T., Khosla, S., … Bouxsein, M. L. (2019). Cortical and trabecular bone microarchitecture as an independent predictor of incident fracture risk in older women and men in the Bone Microarchitecture International Consortium (BoMIC): a prospective study. *The Lancet Diabetes and Endocrinology*, *7*(1), 34–43. https://doi.org/10.1016/S2213-8587(18)30308-5

Schulte, F. A., Lambers, F. M., Kuhn, G., & Müller, R. (2011). In vivo micro-computed tomography allows direct three-dimensional quantification of both bone formation and bone resorption parameters using time-lapsed imaging. *Bone*, *48*(3), 433–442. https://doi.org/10.1016/j.bone.2010.10.007

Sinder, B. P., White, L. E., Salemi, J. D., Ominsky, M. S., Caird, M. S., Marini, J. C., & Kozloff, K. M. (2014). Adult Brtl/+ mouse model of osteogenesis imperfecta demonstrates anabolic response to sclerostin antibody treatment with increased bone mass and strength. *Osteoporosis International*, *25*(8), 2097–2107. https://doi.org/10.1007/s00198-014-2737-y

Sinder, Benjamin P., Eddy, M. M., Ominsky, M. S., Caird, M. S., Marini, J. C., & Kozloff, K. M. (2013). Sclerostin antibody improves skeletal parameters in a Brtl/+ mouse model of osteogenesis imperfecta. *Journal of Bone and Mineral Research*. https://doi.org/10.1002/jbmr.1717

Sinder, Benjamin P., Salemi, J. D., Ominsky, M. S., Caird, M. S., Marini, J. C., & Kozloff, K. M. (2015). Rapidly growing Brtl/+ mouse model of osteogenesis imperfecta improves bone mass and strength with sclerostin antibody treatment. *Bone*, *71*, 115–123.

https://doi.org/10.1016/j.bone.2014.10.012

Swinford, R. R., & Warden, S. J. (2010). Factors affecting short-term precision of musculoskeletal measures using peripheral quantitative computed tomography (pQCT). *Osteoporosis International*. https://doi.org/10.1007/s00198-009-1151-3

Tauer, J. T., Robinson, M.-E., & Rauch, F. (2019). Osteogenesis Imperfecta: New Perspectives From Clinical and Translational Research. *JBMR Plus*, *3*(8), e10174. https://doi.org/10.1002/jbm4.10174

Tsai, J. N., Nishiyama, K. K., Lin, D., Yuan, A., Lee, H., Bouxsein, M. L., & Leder, B. Z. (2017). Effects of Denosumab and Teriparatide Transitions on Bone Microarchitecture and Estimated Strength: the DATA-Switch HR-pQCT study. *Journal of Bone and Mineral Research*. https://doi.org/10.1002/jbmr.3198

Tsai, J. N., Uihlein, A. V., Burnett-Bowie, S. A. M., Neer, R. M., Zhu, Y., Derrico, N., Lee, H., Bouxsein, M. L., & Leder, B. Z. (2015). Comparative effects of teriparatide, denosumab, and combination therapy on peripheral compartmental bone density, microarchitecture, and estimated strength: The DATA-HRpQCT study. *Journal of Bone and Mineral Research*. https://doi.org/10.1002/jbmr.2315

Tsai, J. N., Uihlein, A. V., Burnett-Bowie, S. M., Neer, R. M., Derrico, N. P., Lee, H., Bouxsein, M. L., & Leder, B. Z. (2016). Effects of two years of teriparatide, denosumab, or both on bone microarchitecture and strength (DATA-HRpQCT study). *Journal of Clinical Endocrinology and Metabolism*. https://doi.org/10.1210/jc.2016-1160

van Rietbergen, B., & Ito, K. (2015). A survey of micro-finite element analysis for clinical assessment of bone strength: The first decade. *Journal of Biomechanics*, *48*(5), 832–841. https://doi.org/10.1016/j.jbiomech.2014.12.024

van Rietbergen, Bert, Biver, E., Chevalley, T., Ito, K., Chapurlat, R., & Ferrari, S. (2021). A Novel HR-pQCT Image Registration Approach Reveals Sex-Specific Changes in Cortical Bone Retraction With Aging. *Journal of Bone and Mineral Research*, *00*(00), 1–13. https://doi.org/10.1002/jbmr.4285

Varga, P., Willie, B. M., Stephan, C., Kozlo, K. M., Zysset, P. K., & Currey, J. (2020). *Finite element analysis of bone strength in osteogenesis imperfecta*. *133*(January). https://doi.org/10.1016/j.bone.2020.115250

Waarsing, J. H., Day, J. S., Linden, J. C. Van Der, Ederveen, A. G., Spanjers, C., Clerck, N. De, Sasov, A., Verhaar, J. A. N., & Weinans, H. (2004). *Detecting and tracking local changes in the tibiae of individual rats : a novel method to analyse longitudinal in vivo micro-CT data*. *34*, 163–169. https://doi.org/10.1016/j.bone.2003.08.012

Wagermaier, W., Klaushofer, K., & Fratzl, P. (2015). Fragility of Bone Material Controlled by Internal Interfaces. In *Calcified Tissue International*. https://doi.org/10.1007/s00223-015-9978-4

Whittier, D. E., Boyd, S. K., Burghardt, A. J., Paccou, J., Ghasem-Zadeh, A., Chapurlat, R., Engelke, K., & Bouxsein, M. L. (2020). Guidelines for the assessment of bone density and microarchitecture in vivo using high-resolution peripheral quantitative computed tomography. *Osteoporosis International*. https://doi.org/10.1007/s00198-020-05438-5

Willie, B. M., Zimmermann, E. A., Vitienes, I., Main, R. P., & Komarova, S. V. (2020). Bone adaptation: Safety factors and load predictability in shaping skeletal form. *Bone*, *131*(October 2019), 115114. https://doi.org/10.1016/j.bone.2019.115114

Wong, A. (2016). A comparison of peripheral imaging technologies for bone and muscle quantification: A technical review of image acquisition. *Journal of Musculoskeletal Neuronal Interactions*, *16*(4), 265–282.

Wong, A. K. O., Beattie, K. A., Min, K. K. H., Merali, Z., Webber, C. E., Gordon, C. L., Papaioannou, A., Cheung, A. M. W., & Adachi, J. D. (2015). A trimodality comparison of volumetric bone imaging technologies. part ii: 1-yr change, long-term precision, and least significant change. *Journal of Clinical Densitometry*. https://doi.org/10.1016/j.jocd.2014.07.004

Yang, L., Burton, A. C., Bradburn, M., Nielson, C. M., Orwoll, E. S., & Eastell, R. (2012). Distribution of bone density in the proximal femur and its association with hip fracture risk in older men: The osteoporotic fractures in men (MrOS) study. In *Journal of Bone and Mineral Research*. https://doi.org/10.1002/jbmr.1693

Zhang, J., Wang, Y., Cheng, K., Cheuk, K., Lam, T., Hung, A. L. H., Cheng, J. C. Y., Qiu, Y., Müller, R., Christen, P., & Lee, W. Y. W. (2020). Association of higher bone turnover with risk of curve progression in adolescent idiopathic scoliosis. *Bone*, *September*, 115655. https://doi.org/10.1016/j.bone.2020.115655

Zioupos, P., & Currey, J. D. (1998). Changes in the stiffness, strength, and toughness of human cortical bone with age. *Bone*. https://doi.org/10.1016/S8756-3282(97)00228-7

Zysset, P., Qin, L., Lang, T., Khosla, S., Leslie, W. D., Shepherd, J. A., Schousboe, J. T., & Engelke, K. (2015). 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. *Journal of Clinical Densitometry*, *18*(3), 359–392. https://doi.org/10.1016/j.jocd.2015.06.011