ORIGINAL ARTICLE

JBMR®

HR-pQCT Measures of Bone Microarchitecture Predict Fracture: Systematic Review and Meta-Analysis

Nicholas Mikolajewicz,[1,2] Nick Bishop,[3] Andrew J Burghardt,[4] Lars Folkestad,[5] Anthony Hall,[6] Kenneth M Kozloff,[7] Pauline T Lukey,[8] Michael Molloy-Bland,[9] Suzanne N Morin,[10] Amaka C Offiah,[3] Jay Shapiro,[11] Bert van Rietbergen,[12] Kim Wager,[13] Bettina M Willie,[1,14] Svetlana V Komarova,[1,2] and Francis H Glorieux[1]

[1]Research Center, Shriners Hospital for Children, Montreal, Canada
[2]Department of Dentistry, McGill University, Montreal, Canada
[3]Department of Oncology & Metabolism, University of Sheffield, Sheffield, UK
[4]Department of Radiology & Biomedical Imaging, University of California, San Francisco, San Francisco, CA, USA
[5]Department of Clinical Research, Odense University Hospital, Odense, Denmark
[6]Mereo BioPharma Group PLC, London, UK
[7]Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, MI, USA
[8]Target to Treatment Consulting Ltd, Stevenage, UK
[9]Oxford PharmaGenesis, Melbourne, Australia
[10]Department of Medicine, McGill University, Montreal, Canada
[11]Department of Bone and Osteogenesis Imperfecta, Kennedy Krieger Institute, Baltimore, MD, USA
[12]Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands
[13]Oxford PharmaGenesis, Oxford, UK
[14]Department of Pediatric Surgery, McGill University, Montreal, Canada

ABSTRACT

High-resolution peripheral quantitative computed tomography (HR-pQCT) is a noninvasive imaging modality for assessing volumetric bone mineral density (vBMD) and microarchitecture of cancellous and cortical bone. The objective was to (1) assess fracture-associated differences in HR-pQCT bone parameters; and (2) to determine if HR-pQCT is sufficiently precise to reliably detect these differences in individuals. We systematically identified 40 studies that used HR-pQCT (39/40 used XtremeCT scanners) to assess 1291 to 3253 and 3389 to 10,687 individuals with and without fractures, respectively, ranging in age from 10.9 to 84.7 years with no comorbid conditions. Parameters describing radial and tibial bone density, microarchitecture, and strength were extracted and percentage differences between fracture and control subjects were estimated using a random effects meta-analysis. An additional meta-analysis of short-term in vivo reproducibility of bone parameters assessed by XtremeCT was conducted to determine whether fracture-associated differences exceeded the least significant change (LSC) required to discern measured differences from precision error. Radial and tibial HR-pQCT parameters, including failure load, were significantly altered in fracture subjects, with differences ranging from −2.6% (95% confidence interval [CI] −3.4 to −1.9) in radial cortical vBMD to −12.6% (95% CI −15.0 to −10.3) in radial trabecular vBMD. Fracture-associated differences reported by prospective studies were consistent with those from retrospective studies, indicating that HR-pQCT can predict incident fracture. Assessment of study quality, heterogeneity, and publication biases verified the validity of these findings. Finally, we demonstrated that fracture-associated deficits in total and trabecular vBMD and certain tibial cortical parameters can be reliably discerned from HR-pQCT-related precision error and can be used to detect fracture-associated differences in individual patients. Although differences in other HR-pQCT measures, including failure load, were significantly associated with fracture, improved reproducibility is needed to ensure reliable individual cross-sectional screening and longitudinal monitoring. In conclusion, our study supports the use of HR-pQCT in clinical fracture prediction. © 2019 American Society for Bone and Mineral Research.

KEY WORDS: ANALYSIS/QUANTITATION OF BONE; BONE QCT; CLINICAL TRIALS; FRACTURE RISK ASSESSMENT; STATISTICAL METHODS

Received in original form March 6, 2019; revised form September 19, 2019; accepted October 13, 2019. Accepted manuscript online October 24, 2019.
Address correspondence to: Francis H Glorieux, MD, PhD, Research Center, Shriners Hospital for Children - Canada, 1003 Decarie Blvd, Montreal, Quebec, H4A 0A9, Canada. E-mail: glorieux@shriners.mcgill.ca
Additional Supporting Information may be found in the online version of this article.

Journal of Bone and Mineral Research, Vol. 35, No. 3, March 2020, pp 446–459.
DOI: 10.1002/jbmr.3901

Introduction

The use of bone imaging outcomes to predict fracture risk is important in clinical assessment of bone health. Currently, the gold standard for clinical imaging of bone mass is dual-energy X-ray absorptiometry (DXA). DXA-derived areal BMD (aBMD) is a significant predictor of fracture risk; however, its predictive value is limited because (1) the aBMD in most older individuals that experience a fracture is outside the osteoporotic range (T-score < −2.5)[1,2]; (2) it does not differentiate between the cortical and cancellous bone compartment; and (3) additional information on bone microarchitecture—which is an indicator of bone quality and predictor of fracture[3–6]—cannot be determined.

High-resolution peripheral quantitative computed tomography (HR-pQCT) has emerged as a noninvasive imaging modality with an isotropic voxel size of 82 μm (XtremeCT; XCT, Scanco Medical, Bruttisellen, Switzerland) or 61 μm (XtremeCT II; XCT II, Scanco Medical), which allows for assessment of volumetric bone density and microarchitecture of cortical and cancellous bone compartments. Although early studies on the association between HR-pQCT measures and fracture have been promising, HR-pQCT scanners are predominantly used in a research setting.[7] Until recently,[3–6] HR-pQCT studies have been retrospective by design,[8] raising uncertainty about whether the fracture-associated differences observed by HR-pQCT are a cause or consequence of the fracture event. Moreover, there have been concerns about whether HR-pQCT scans, which are restricted to metaphyseal (XCT) radial and tibial sites, are informative about clinically relevant sites such as spine or hip. Next-generation scanners (XCT II) allow imaging of diaphyseal sites, but little data are currently available.

To detect differences in bone between baseline and follow-up measures in an individual patient, imaging techniques must be highly reproducible (ie, precise). A system's precision can be quantified using repeat measures of imaging phantoms, such as the European Forearm Phantom (EFP).[9,10] However, when scanning patients, precision error tends to be much higher than that assessed by phantoms because of motion artifacts, reference line positioning, and operator skill. Thus, the reproducibility of HR-pQCT measures is best assessed using repeat measures from individual patients. Precision errors are parameter- and scanner-dependent and, once characterized, the least significant change (LSC) can be estimated, which then informs the user of the smallest difference that can confidently be discerned from precision error.[11] For HR-pQCT to be reliably used in a clinical setting for longitudinal monitoring and cross-sectional screening,[12] fracture risk-associated differences in bone must exceed the LSC.

The objectives of this study were to (1) conduct a systematic review and meta-analysis of studies that investigated the association between fracture and HR-pQCT measures in otherwise healthy individuals with and without age-related osteoporosis; and (2) determine whether HR-pQCT is sufficiently precise to detect fracture-associated differences in individual patients. In short, if deficits in HR-pQCT measures are associated with fracture, can these deficits be detected at the individual patient level—rather than the cohort level—in the presence of measurement error?

Materials and Methods

The present study was conducted in a PRISMA-compliant manner[13] See Supplemental Table S1 in for the PRISMA checklist.

Software

Endnote X7 (Thomson Reuters, Toronto, Canada) was used to manage references; the MetaLab meta-analysis tool box[14] in MATLAB (MathWorks, Natick, MA, USA) and JASP 0.9.01 (JASP Team, Amsterdam, The Netherlands) were used to conduct the meta-analysis; Excel 2016 (Microsoft, Redmond, WA, USA) was used for data storage; and CorelDRAW X8 (Corel, Ottawa, Canada) was used for figure preparation.

Search strategy and inclusion criteria

The search strategy was constructed using key terms for "HR-pQCT" and "fracture risk" (Supplemental Table S2). Medline, Embase, PROSPERO, and Cochrane DSR were searched on February 14, 2018, and articles were exported to Endnote. Titles and abstracts were screened independently by two authors (MMB and KW). Studies were included if HR-pQCT was used to compare bone parameters between individuals of all ages with or without fracture. Only individuals that were reported as apparently healthy, or with possible age-related osteoporosis, were considered. Studies were excluded if outcome reporting was deficient (eg, unclear or absent measures of effect, error, or sample size). Only studies published in English were included. Reviews, books, letters, editorials, and conference proceedings were excluded. No restrictions were imposed on study design, type of fracture assessed, or treatments taken by study participants (eg, bisphosphonates). Eligibility was confirmed by full-text screening of articles. Screening conflicts were resolved by consensus as required. The bibliographies of studies identified in the primary search were screened for additional studies using the same selection criteria as above.

Quality assessment

Study quality was assessed using a nine-item quality checklist[14]: (1) HR-pQCT short-term reproducibility assessed; (2) fracture history verified by primary report (eg, radiograph, medical report); (3) recruitment from same source (fracture and non-fracture individuals recruited from same population); (4) age-matched; (5) random sampling; (6) random matching; (7) informed consent; (8) ethics committee approval; and (9) conflict of interest statements. Data were stratified by aggregate quality score to assess the influence of study quality on outcomes.

Data extraction

Data were extracted by a single reviewer (MMB) and verified by another (NM). For each study, we extracted sample sizes (n_f = number of fracture subjects; n_c = number of non-fracture subjects), raw outcomes (θ_f = fracture subject outcome; θ_c = non-fracture subject outcome); and error measures [$se(\theta)$ = standard errors; $sd(\theta)$ = standard deviations] for 11 HR-pQCT parameters (Table 1) from individuals with and without fracture. Trabecular bone volume fraction (Tb. BV/TV, %) was also extracted from studies and converted to trabecular vBMD (Tb.vBMD) [Tb. vBMD = (1200 mg hydroxyapatite (HA)/cm^3 · Tb. BV/TV)/100], assuming fully mineralized bone has a mineral density of 1200 mg HA/cm^3. The MetaLab data extraction module was used for graphical data extraction.[14] Study characteristics including age, sex, fracture site, and trauma degree were recorded for all studies.

Table 1. HR-pQCT Parameters Included in Meta-Analysis

Parameter (abbreviation)	Description	Units
Volumetric bone mineral density (vBMD) measures		
1. Total (Tt.vBMD)	Total volumetric density	mg HA/cm^3
2. Cortical (Ct.vBMD)	Cortical volumetric density	mg HA/cm^3
3. Trabecular (Tb.vBMD)	Trabecular volumetric density	mg HA/cm^3
Cortical (Ct.) measures		
4. Area (Ct.Ar)	Mean area occupied by cortical bone	mm^2
5. Thickness (Ct.Th)	Mean cortical thickness, calculated indirectly as ratio of cortical bone volume to outer bone surface[37] or directly[38]	mm
6. Porosity (Ct.Po)	Cortical porosity, calculated using void-voxel[38] or density-based method[34]	%
Trabecular (Tb.) measures		
7. Thickness (Tb.Th)	Mean thickness of trabeculae	mm
8. Number (Tb.N)	Mean number of trabeculae per unit length	mm^{-1}
9. Separation (Tb.Sp)	Mean distance between trabeculae	mm
Finite-element analysis (FEA) measures		
10. Stiffness	Whole bone stiffness	N/mm
11. Failure load	Estimated maximum load	N

HA = hydroxyapatite.

Study-level outcomes

All measures were extracted as raw values and calculated as percentage differences (*dif*):

$$dif\,(\%) = \frac{\theta_f - \theta_c}{\theta_c} \times 100\% \quad (1)$$

and standard errors *se*(*dif*)[15]:

$$se(dif) = \sqrt{\left(\frac{se(\theta_c)}{\theta_c}\right)^2 + \left(\frac{se(\theta_f)}{\theta_f}\right)^2} \times 100\% \quad (2)$$

Assessment of short-term in vivo reproducibility

A selection of methodological studies from the authors' library were used to conduct a rapid review and meta-analysis of short-term in vivo reproducibility of HR-pQCT measures acquired from individual patients by XCT scanners. These included studies in which within-patient reproducibility was assessed between scans after limb repositioning, since error can arise from repositioning of reference lines, motion artifacts, and non-overlapping volumes of interest (VOI). For each study, the root-mean-squared percent coefficient of variance ($CV\%_{RMS}$), number of subjects (ie, precision error), number of repeat scans, and scanned regions were extracted. Region-specific (radius, tibia) reproducibility was calculated as a weighted mean. Weights were degrees of freedom $df = m(n-1)$, where m = number of subjects and n = number of repeat scans. Parameter- and region-specific $CV\%_{RMS}$ estimates were used to compute the LSC required to reliably detect a difference between HR-pQCT measures acquired from two separate scans in an individual patient, at a 95% confidence level:

$$LSC = 1.96 \cdot \sqrt{CV\%_{RMS}^2 + CV\%_{RMS}^2} = 2.77 \cdot CV\%_{RMS} \quad (3)$$

Throughout this article, "reliability of HR-pQCT measures" refers to how fracture-associated differences relate to the LSC, where the LSC is interpreted as a metric that assesses whether there is sufficient signal over noise to deem a measurement reliable.[12] HR-pQCT measures for which fracture-associated differences exceeded the LSC were deemed reliable predictors of fracture at the individual patient level.

Results

Overview of studies

Database searches identified 516 unique citations; of these, 36 were eligible for inclusion (Fig. 1). An additional 13 studies were included from a bibliographic review of the 36 included studies, and two more were added from the authors' library. From the 51 identified studies, 11 studies focused on populations with comorbid illness/conditions (kidney disease[16–20] [$n = 5$], systemic lupus erythematosus[21,22] [$n = 2$], idiopathic osteoporosis[23] [$n = 1$], Sjögren's syndrome[24] [$n = 1$], ankylosing spondylitis[25] [$n = 1$], and oligo/amenorrhea[26] [$n = 1$]) and were excluded. A study in individuals with diabetes was also identified; however, it was included because nondiabetic cohort data were also reported.[27] In total, 40 studies were included in our meta-analysis (Fig. 1).

Characteristics of studies included in the analysis are presented in Supplemental Table S3. Most studies were conducted in Europe (20/40 [50%]) or North America (17/40 [42.5%]), and the remaining studies (3/40 [7.5%]) were conducted in Australia, China, or Israel/USA. The number of individuals included in our meta-analysis varied for different HR-pQCT parameters, ranging from 1291 to 3253 fracture subjects and 3389 to 10,687 non-fracture subjects. All studies used the first-generation XCT scanner (isotropic voxel size of 82 μm) to assess the association between tibial and/or radial measures and fracture, with the exception of Fink and colleagues, who used the second-generation XCT II scanner (isotropic voxel size of 61 μm).[28] Almost all studies in adults reported using standard 9.5 mm (42/42 data sets [100%]) or 22.5 mm (38/39 data sets [97.4%]) proximal offsets from the radial and tibial reference lines, respectively. In contrast, pediatric studies tended to use smaller absolute (1.0 mm,[29] 2.0 mm,[30] or 9.5 mm[31]) or relative (4% total length[32]) proximal radial offsets. Four studies reported using the Strax method for image segmentation and analysis;[6,32–34] however, one of these also reported data acquired

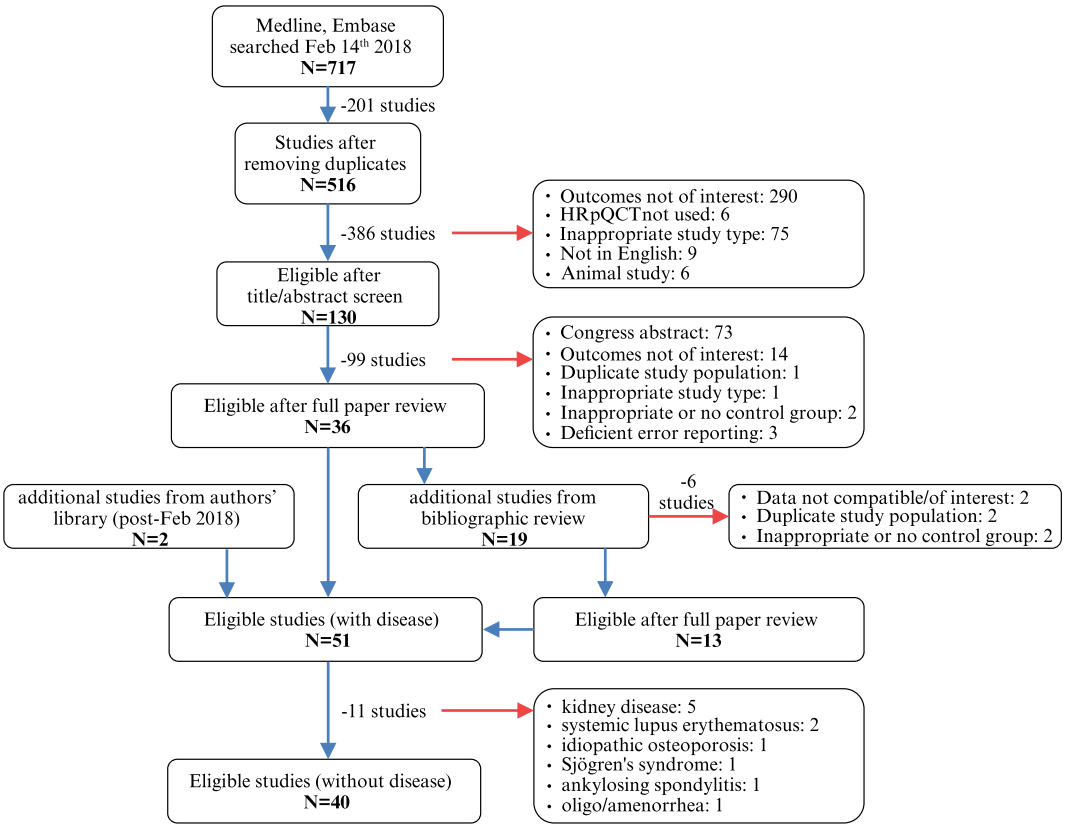

Fig. 1. PRISMA diagram of flow of information through systematic review and meta-analysis. N = number of studies; blue arrows = studies included; red arrows = studies omitted.

by the Scanco method so Strax data were omitted in this case.[6] Although there are known inconsistencies between values estimated by the Strax and Scanco algorithms, most parameter estimates are linearly related, thereby ensuring that proportional changes in either set of values are consistent and amenable to being pooled in the current study as percentage differences.[35] The most frequently reported fracture sites were "any" (n = 20 studies), forearm (n = 12), and vertebral/spine (n = 5).

Cross-sectional (retrospective) studies were defined as those in which HR-pQCT scans were performed after the fracture event. Most studies (36/40 [90%]) were cross-sectional, of which half (18/36 [50%]) were case–control studies. Cross-sectional studies tended to focus on older female subjects (women [n data sets = 28; median age 64.9 years; interquartile range (IQR) 47.6–70.1], men [n = 9; median age 57.0 years; IQR 30.0–71.0], girls [n = 3; median age 11.8 years; IQR 11.1–11.8], boys [n = 3; median age 12.6 years; IQR 12.0–15.2]). The time between fracture and HR-pQCT assessment varied substantially across the cross-sectional studies—some reported that fracture occurred at any time before the scan (10/40 studies [25%]) or after menopause (5/40 [12.5%]), whereas others reported time frames ranging from <4 weeks to 35.8 years between fracture and scan. Four studies were prospective (4/40 [10%]), in which HR-pQCT scans were acquired at baseline and then patients were followed (for 5.0 to 9.4 years) for fracture events. Three prospective studies investigated females (median age 67.1 years; IQR 65.0–68.4), whereas Szulc and colleagues[5] was the only prospective study on males (mean age 72.1 years). All four prospective studies included low-trauma fractures (ie, fragility fractures, defined as a fall from a standing position or less) and excluded moderate- to high-trauma fractures.

Radial and tibial HR-pQCT bone parameters discriminate fracture status

Fracture-associated differences across all studies were assessed in radial and tibial Tb.vBMD (Fig. 2) and finite element analysis (FEA)-derived failure load (Fig. 3), as well as additional density (Supplemental Figs. S1 and S2), microarchitectural (Supplemental Figs. S3–S8), and FEA parameters (Supplemental Fig. S9).

HR-pQCT-derived radial and tibial parameters were significantly different among fracture and control subjects. Fracture-associated differences increased with age (note that forest plots are ordered by age) and were consistently larger in the radius than in the tibia, especially for trabecular measures Tb.vBMD (p = 0.02), trabecular thickness (Tb.Th) (p = 0.15), and trabecular number (Tb.N) (p = 0.04). Differences varied considerably across bone parameters, ranging from 0.9% (95% confidence interval [CI] −2.6 to 4.5) in radial cortical porosity (Ct.Po) (Supplemental Fig. S5) to −12.6% (95% CI −15.0 to −10.3) in radial Tb.vBMD (Fig. 2). Thus, bone quality assessment by HR-pQCT can discriminate fracture status in individuals.

Quality analysis

We assessed how study quality influences outcomes. The median study-level quality score was 7 of 9. In general, fracture-associated differences were overestimated in lower-quality studies (Fig. 4;

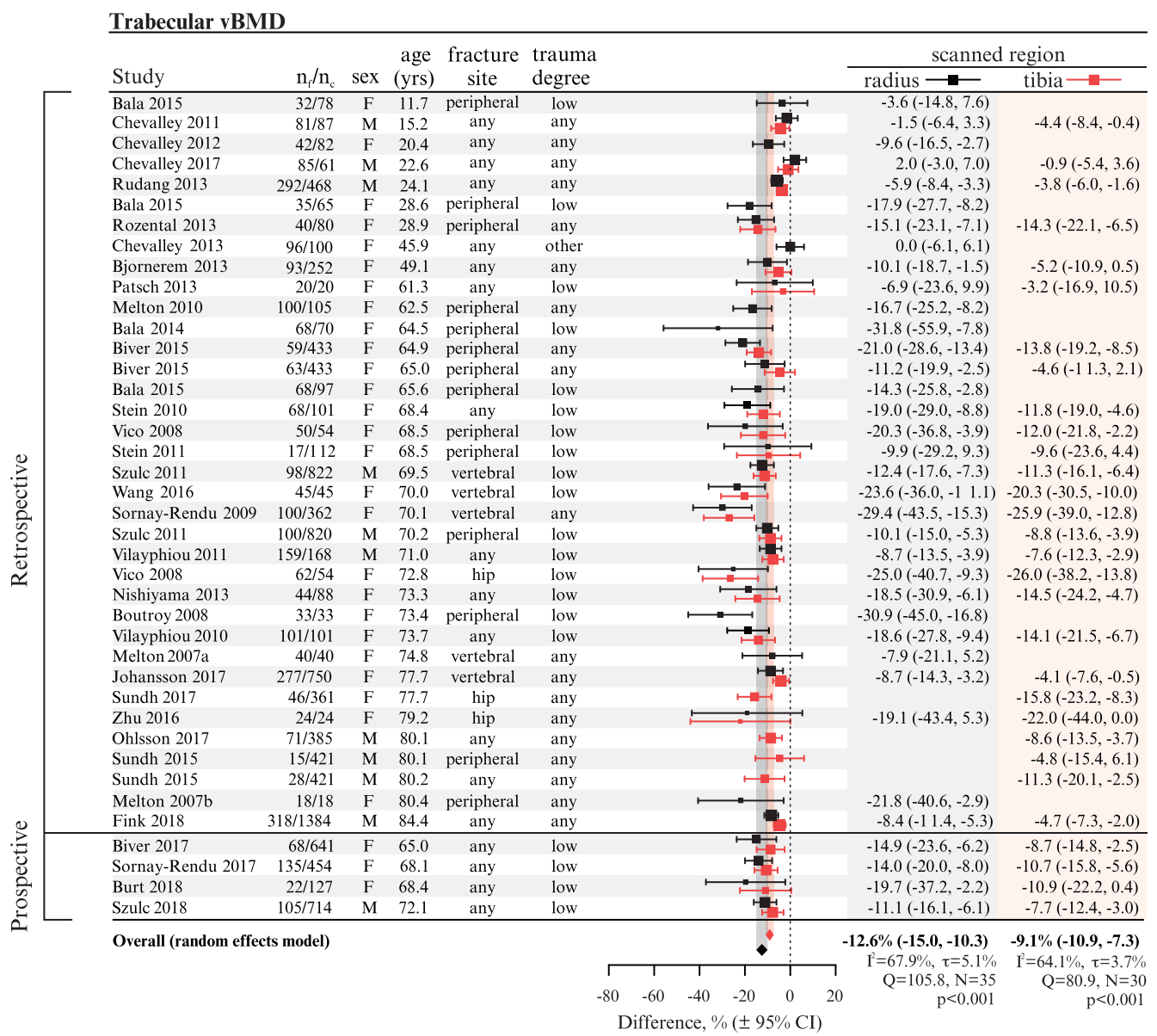

Fig. 2. Forest plot of fracture-associated differences in radial and tibial Tb.vBMD. Data are study-level percent differences between individuals with and without fracture ±95% CI, stratified by retrospective and prospective study design and sorted by participant age within each stratum. Red markers = tibial Tb.vBMD; black markers = radial Tb.vBMD; dashed line = no fracture reference. I^2, τ, and Q = heterogeneity statistics; N = number of data sets; p = *p* value for Q heterogeneity test; n_f = number of fracture subjects; n_c = number of control subjects; Tb.vBMD = trabecular volumetric BMD. Marker sizes are proportional to study-level weights.

Supplemental Fig. S10). Study quality explained 11.2% (cortical volumetric BMD [Ct.vBMD]) to 30.3% (trabecular thickness [Tb.Th]) of the observed heterogeneity between studies (Fig. 4; Supplemental Fig. S10). Nonetheless, higher variances reported in lower-quality studies ensured the overall outcome was not distorted.

Risk of bias

Because sampling error is assumed to be random, all study-level outcomes should be symmetrically distributed around the most precise estimates. However, parameter-specific funnel plots revealed asymmetries for several HR-pQCT measures (Fig. 5; Supplemental Fig. S11), which is suggestive of biased reporting practices that arise when small studies refrain from publishing negative results.[36] We used trim-and-fill analysis to examine the effect of publication bias and to compute "unbiased" estimates. Bias had a negligible influence on all parameter estimates, except Ct.vBMD (Fig. 5*C*). We concluded that there was minimal risk that the fracture-associated differences computed in this meta-analysis are an artifact of bias.

Heterogeneity analysis

Overall heterogeneity in study-level fracture-associated differences was moderate to high, accounting for 39.1% (radial Ct.Po;

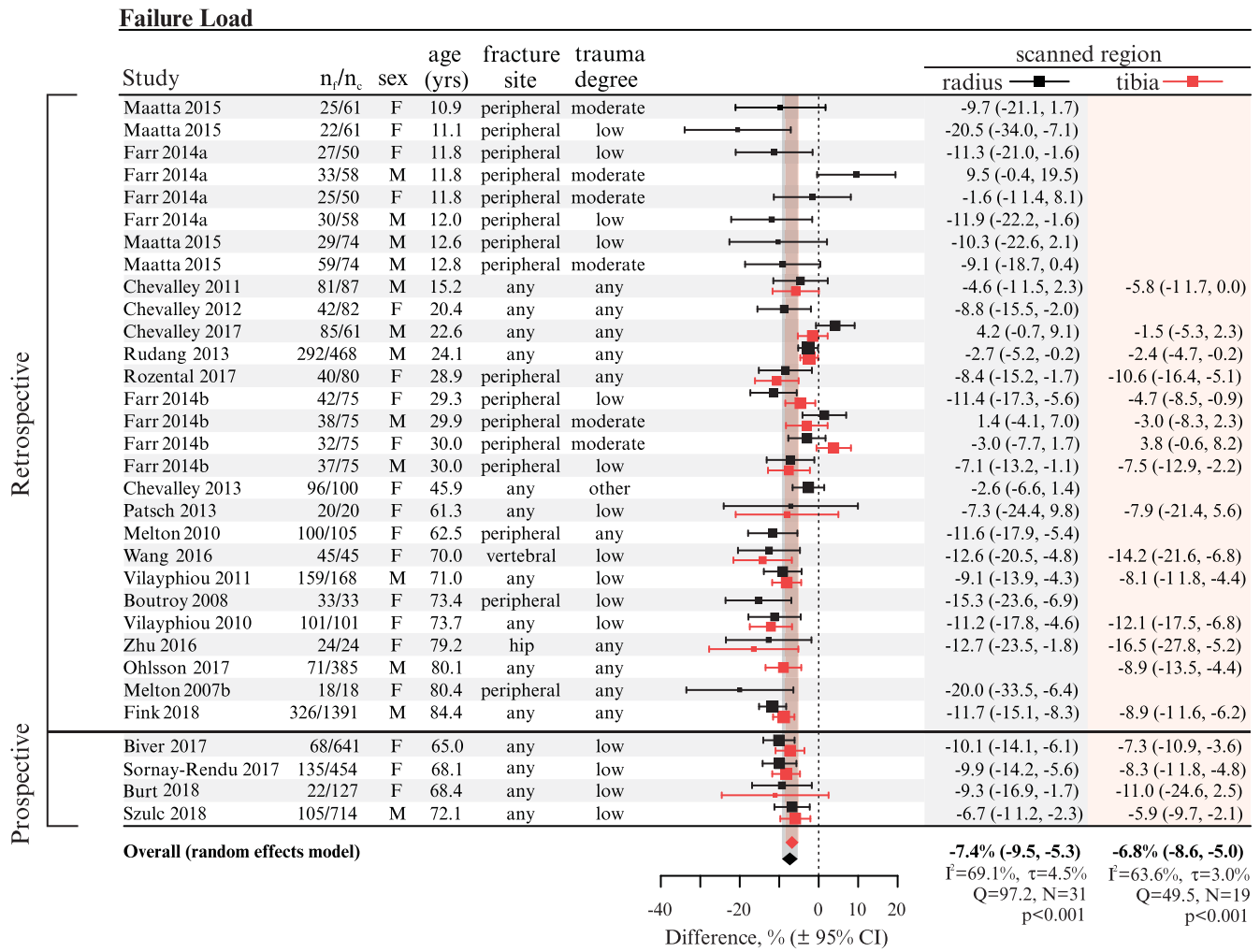

Fig. 3. Forest plot of fracture-associated differences in radial and tibial failure load. Data are study-level percent differences between individuals with and without fracture history ±95% CI, stratified by retrospective and prospective study design and sorted by participant age within each stratum. Red markers = tibial failure load; black markers = radial failure load; dashed line = no fracture reference. I^2, τ, and Q = heterogeneity statistics; N = number of data sets, p = *p* value for Q heterogeneity test; n_f = number of fracture subjects; n_c = number of control subjects. Marker sizes are proportional to study-level weights.

Supplemental Fig. S5) to 81.1% (tibial Ct.vBMD; Supplemental Fig. S2) of the observed variance in the data. Single-study exclusion analysis did not identify any influential outliers (Fig. 6*A*, *B*; Supplemental Fig. S12*A*). Using a cumulative-study exclusion approach, the homogeneity threshold T_H was found to be below 30% for all parameters except Tb.N and trabecular separation (Tb.Sp), for which T_H was 33% and 43%, respectively (Fig. 6*C*, *D*; Supplemental Fig. S12*B*). Importantly, estimates from homogeneous subsets of the data were consistent with overall estimates, except Tb.Sp.

For cortical thickness (Ct.Th) and Ct.Po measures, we assessed if differences in parameter estimation methods contributed to heterogeneity. For Ct.Th, the indirect estimation method[37] was used in ~50% of studies and yielded significantly lower estimates compared with the direct method,[38] used in ~35% of studies (Supplemental Fig. S4*B*, *C*)—even after adjusting for age and sex differences (Supplemental Table S4). For Ct.Po, the majority of studies (90.1%) used a void voxel-based method,[38] which provided similar estimates of fracture-associated differences compared with the density-based method[39] (Supplemental Fig. S5*B*, *C*).

Taken together, we determined that meta-analytic outcomes were largely unaffected by the moderate to high heterogeneity observed between data sets.

Differences in HR-pQCT bone parameters can predict incident fracture

Retrospective studies do not inform us whether poor bone quality observed in fracture subjects is a cause or consequence of the fracture event. To address this, we separately examined fracture-associated differences reported by prospective and retrospective studies (Fig. 7; Supplemental Table S5). All parameters, except Ct. Po, were significantly associated with fractures in retrospective and prospective studies. Prospective studies were limited to a relatively homogeneous group of older females and males (I^2 = 0%) with similar fracture types and trauma (Supplemental Table S3); therefore,

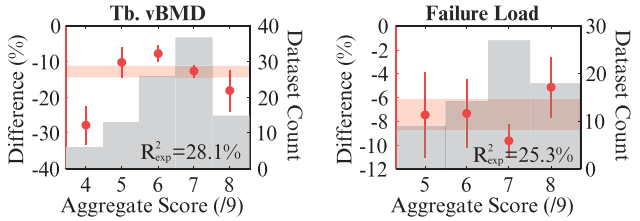

Fig. 4. Influence of aggregate study quality score on fracture-associated differences in Tb.vBMD and failure load. Red band = 95% CI for overall estimate; red markers = score-specific estimate ±95% CI; gray bars = number of data sets that received indicated aggregate quality score. R^2_{exp} specifies percentage of heterogeneity explained by quality score. Tb. vBMD = trabecular volumetric BMD. Analyses of influence of quality on fracture-associated differences in other HR-pQCT parameters are provided in Supplemental Fig. S10.

these findings may not be generalized to younger populations. Nevertheless, the consistency between prospective and retrospective outcomes is a strong indicator that HR-pQCT parameters are predictive of incident fracture, even in populations for whom prospective data are lacking (eg, pediatric and young adult populations).

Reproducibility of HR-pQCT measures and implications for fracture prediction

For HR-pQCT parameters to be useful in fracture prediction, they must not only be significantly associated with fracture but must

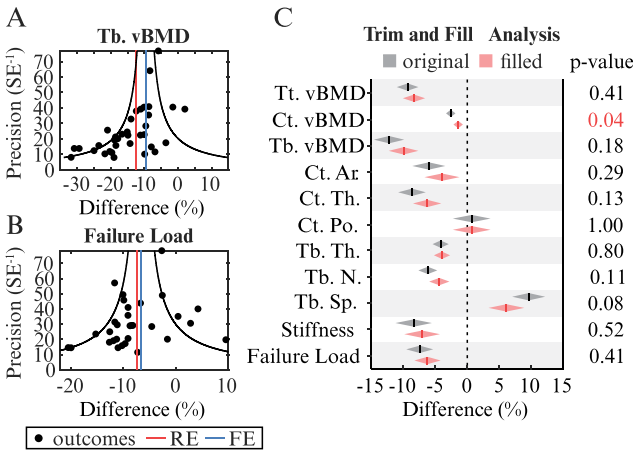

Fig. 5. Risk of bias. (*A*, *B*) Funnel plots of study-level fracture-associated differences for Tb.vBMD (*A*) and failure load (*B*). Black markers = study-level data; blue lines = fixed effect (FE) estimates; red lines = random effects (RE) estimates; black lines = theoretical 95% CI for FE estimate in absence of bias. Funnel plots for other HR-pQCT parameters are provided in Supplemental Fig. S11. (*C*) Comparison of original (gray) and trim-and-filled (red) random effects estimates. Significance between original and filled estimates was assessed by *Z* test and *p* values are reported. Ct.Ar = cortical area; Ct.Po = cortical porosity; Ct.Th = cortical thickness; Ct.vBMD = cortical volumetric BMD; Tb.N = trabecular number; Tb. Sp = trabecular separation; Tb.Th = trabecular thickness; Tb.vBMD = trabecular volumetric BMD; Tt.vBMD = tibial volumetric BMD.

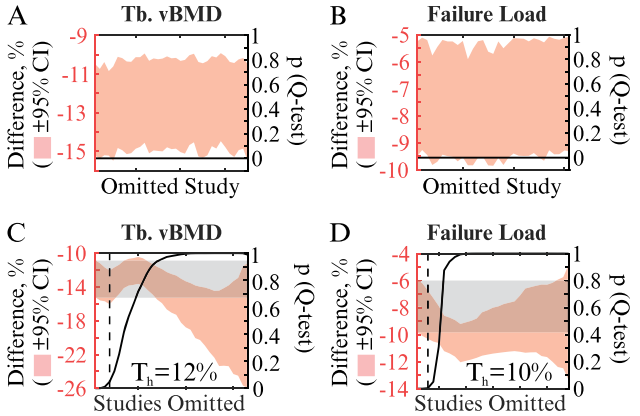

Fig. 6. Heterogeneity and sensitivity analyses for Tb.vBMD and failure load. (*A*, *B*) Single-study exclusion sensitivity analysis of Tb.vBMD (*A*) and failure load (*B*). Individual data sets were omitted from meta-analysis and influence on fracture-associated difference ± 95% CI (red left axis, shaded bands) and heterogeneity (black right axis, solid curve) was assessed. (*C*, *D*) Cumulative-study exclusion sensitivity analysis of Tb. vBMD (*C*) and failure load (*D*). Data sets were cumulatively removed according to maximal Q-reduction criteria and homogeneity threshold (T_h, dashed black line) was estimated as percentage of studies that must be removed to obtain homogeneous data set ($p > 0.05$ by *Q* test). Red left axis, shaded bands = fracture-associated difference ± 95% CI; black right axis, solid curve = *p* values for Q heterogeneity test; gray bands = overall 95% CI; Tb.vBMD = trabecular volumetric BMD. Sensitivity analyses for other HR-pQCT parameters are provided in Supplemental Fig. S12.

also be highly reproducible. We investigated the short-term reproducibility of HR-pQCT measures. Short-term reproducibility was assessed in 58% of studies (Supplemental Fig. S13*A*); however, the quality of reporting was poor because most studies tended to report precision ranges for multiple parameters rather than parameter-specific precision estimates. Nonetheless, short-term reproducibility for density-based measures ($CV\%_{RMS}$ = 1.2%) was significantly better ($p < 0.001$) than for structural measures ($CV\%_{RMS}$ = 3.9%) (Supplemental Fig. S13*B*).

To assess parameter-specific reproducibility, we conducted a rapid review and meta-analysis of in vivo short-term reproducibility of HR-pQCT measures acquired from individual patients, with complete limb repositioning between scans (Fig. 8*A–D*; Supplemental Table S6).[38,40–50] Consistent with ranges reported in HR-pQCT fracture studies (Supplemental Fig. S13*B*), density-related measurements ($CV\%_{RMS}$ = 0.8 – 2.0%; Fig. 8*A*) were the most precise, whereas Ct. Po ($CV\%_{RMS}$ = 6.2 – 12.5%; Fig. 8*B*) and trabecular microarchitecture measures ($CV\%_{RMS}$ = 4.1 – 4.9%; Fig. 8*C*) were the least. By comparison, FEA parameters were moderately precise ($CV\%_{RMS}$ = 2.0 – 3.5%; Fig. 8*D*). Of interest, reproducibility was consistently higher for tibial compared with radial parameters, especially for Ct.vBMD (p = 0.04), Ct.Th. (p = 0.08), and Ct.Po ($p < 0.001$) (Fig. 8*A–D*; Supplemental Table S6).

The least significant change (Equation 3) required to reliably detect a difference between HR-pQCT measures acquired from separate scans was overlaid with fracture-associated differences in radial (Fig. 8*E*) and tibial (Fig. 8*F*) HR-pQCT

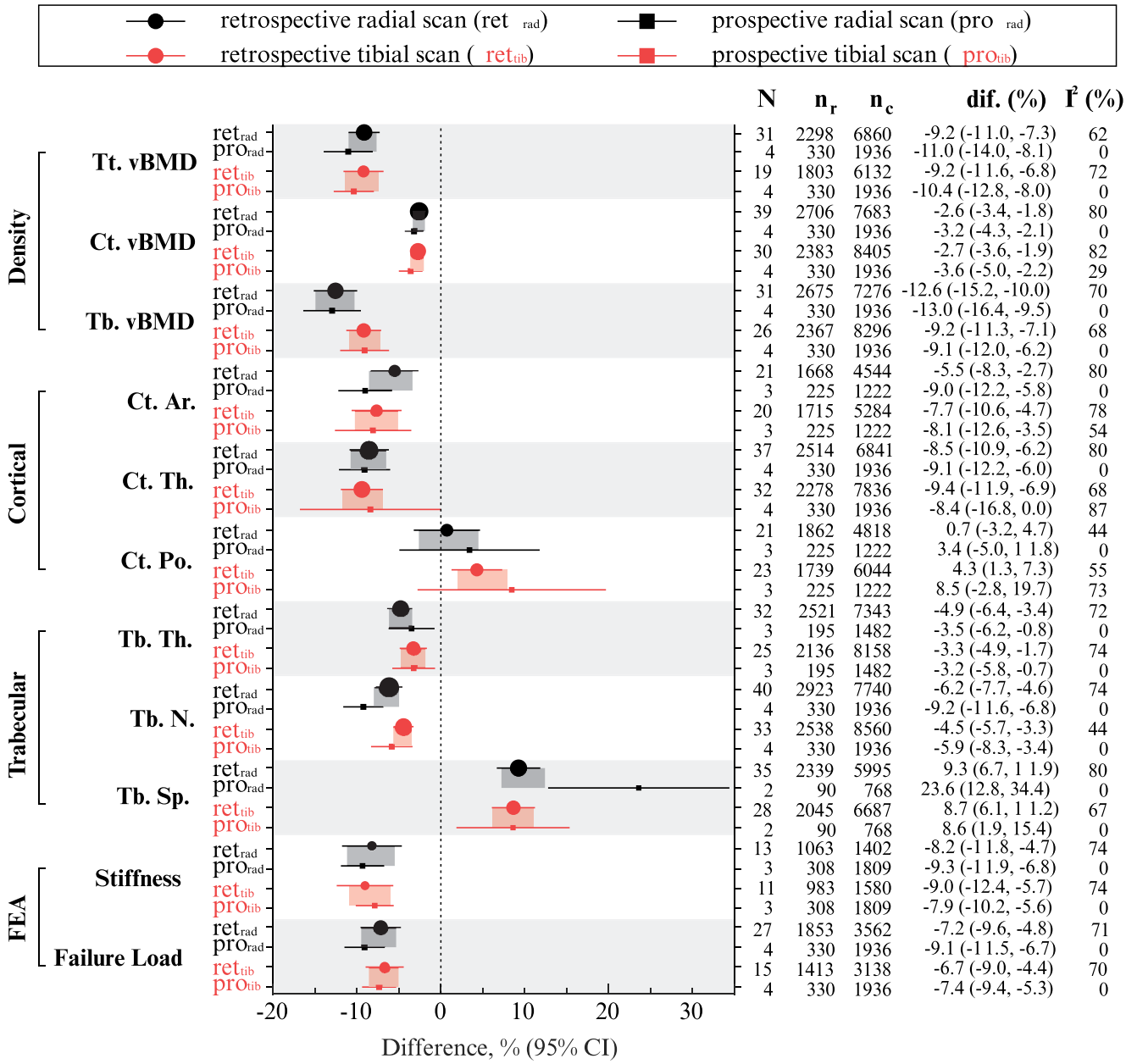

Fig. 7. Fracture-associated differences in radial and tibial HR-pQCT parameters acquired from prospective and retrospective studies. Fracture-associated differences (dif) ± 95% CI in HR-pQCT radial (black) and tibial (red) measures acquired from retrospective (ret; round markers) and prospective (pro; square markers) studies. Shaded red/black bands = overall estimate ± 95% CI for specified parameter and scanned region; dashed line = no fracture reference; Ct.Ar = cortical area; Ct.Po = cortical porosity; Ct.Th = cortical thickness; Ct.vBMD = cortical volumetric BMD; I^2 = heterogeneity statistic; N = number of data sets (marker sizes are proportional to N); n_c = number of non-fracture subjects; n_r = number of fracture subjects; Tb.N = trabecular number; Tb.Sp = trabecular separation; Tb.Th = trabecular thickness; Tb.vBMD = trabecular volumetric BMD; Tt.vBMD = tibial volumetric BMD. Detailed statistics are provided in Supplemental Table S5.

measures. Fracture-associated differences in radial and tibial vBMD (Tt.vBMD) and Tb.vBMD were significantly larger than the LSC (Fig. 8*E, F*). Additionally, differences in Ct.vBMD and Ct.Th exceeded the LSC in the tibia (Fig. 8*F*) but not the radius (Fig. 8*E*). There were several HR-pQCT measures that were in similar ranges as the LSC, specifically radial Ct.Th, stiffness, and failure load (Fig. 8*E*), as well as tibial cortical area (Ct.Ar), Tb.Sp, stiffness, and failure load (Fig. 8*F*). These borderline cases in which fracture-associated differences nearly exceed the LSC demonstrate a need for improved reproducibility, which may then qualify this subset of HR-pQCT parameters for fracture prediction.

Because of the poor reproducibility and relatively smaller fracture-associated differences, Ct.Po, Tb.Th, and Tb.N are unlikely to serve as reliable predictors of fracture.

We concluded that radial and tibial Tt.vBMD and Tb.vBMD, as well as tibial Ct.vBMD and Ct.Th can reliably detect fracture-associated differences.

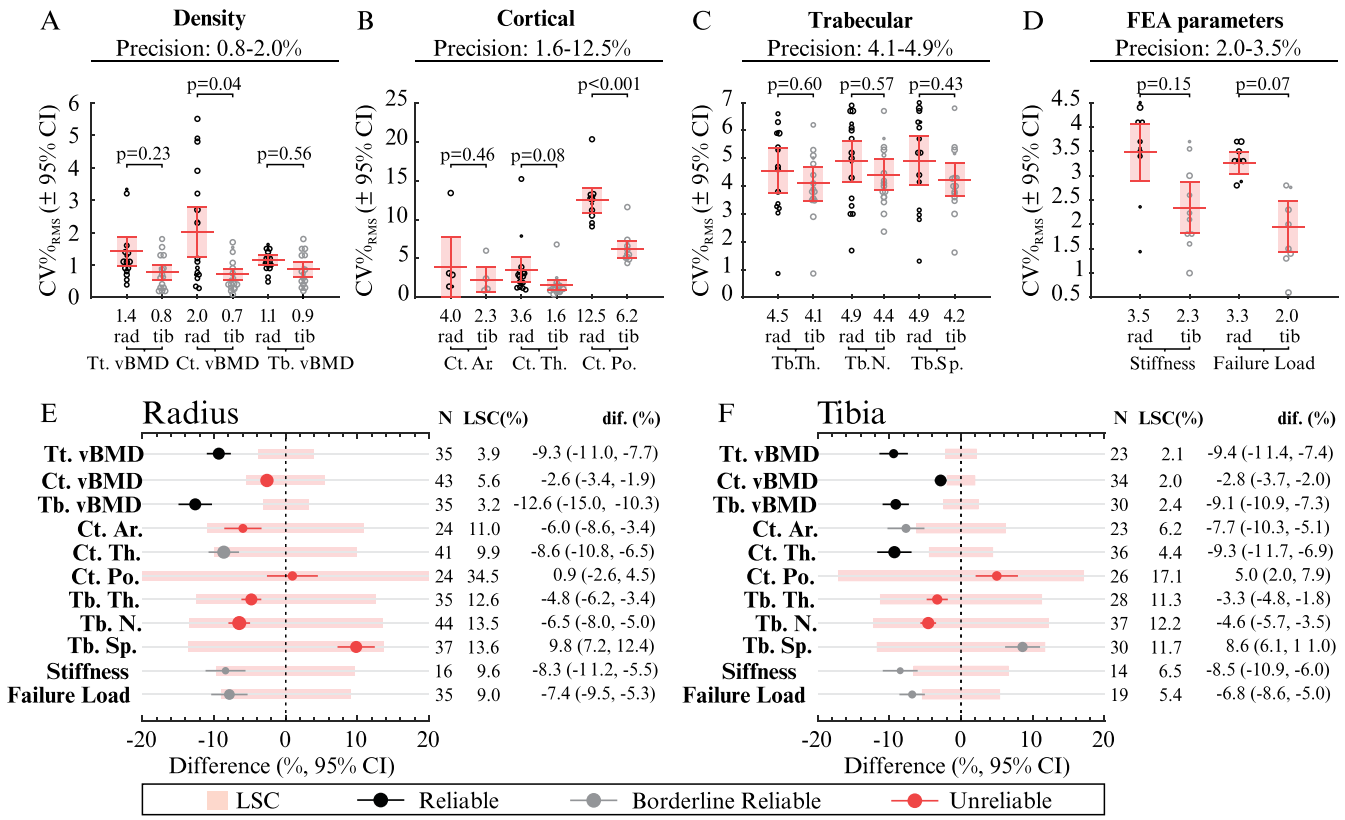

Fig. 8. In vivo short-term reproducibility of HR-pQCT measures and implications for fracture prediction. (*A–D*) Meta-analytic estimates of in vivo HR-pQCT reproducibility of density (*A*), cortical (*B*), trabecular (*C*), and FE model (*D*) parameters obtained from radial (rad.; black markers) and tibial (tib.; gray markers) scans. Study-level root-mean-squared percent coefficients of variance ($CV\%_{RMS}$) were pooled as weighted means. Weights were degrees of freedom = $m\,(n-1)$, where m = number of subjects, n = repeat measures. Red lines/bands = means ± 95% CI; radial and tibial reproducibility was compared by Z test, p values are shown. (*E*, *F*) Forest plots of fracture-associated differences (dif.) ± 95% CI in radial (*E*) and tibial (*F*) HR-pQCT measures overlaid with least-significant change (red bands); LSC = 2.77 ($CV\%_{RMS}$). Black markers = reliable measures (no overlap with LSC); gray markers = borderline reliable measures (partial overlap with LSC); red markers = unreliable measures (complete overlap with LSC); black dashed line = no fracture reference, markers are proportional to number of data sets N. Ct.Ar = cortical area; Ct.Po = cortical porosity; Ct.Th = cortical thickness; Ct.vBMD = cortical volumetric BMD; LSC = least significant change; Tb.N = trabecular number; Tb.Sp = trabecular separation; Tb.Th = trabecular thickness; Tb.vBMD = trabecular volumetric BMD; Tt.vBMD = tibial volumetric BMD.

Discussion

Overview

The objective of this study was to demonstrate whether HR-pQCT-derived bone parameters can reliably predict fracture. We report that radial and tibial HR-pQCT parameters, including failure load, were significantly altered in fracture subjects. Differences in HR-pQCT measures reported by prospective studies were consistent with those from retrospective studies, indicating that HR-pQCT parameters are predictive of incident fracture. Assessment of study quality, heterogeneity, and publication biases verified the validity of these findings. To further support the utility of HR-pQCT, we evaluated whether expected fracture-associated differences can be discerned from measurement error at the individual level. Based on our assessment of published in vivo short-term reproducibility data, we demonstrate that fracture-associated differences in radial or tibial Tt.vBMD and Tb.vBMD, as well as tibial Ct.vBMD and Ct.Th can be reliably detected in individual patients using the XCT scanner. We conclude that there is strong evidence supporting the use of HR-pQCT for fracture prediction in a clinical setting.

Mechanistic insights

The current study offers several mechanistic and biological insights. Findings reported by retrospective and prospective studies were consistent, indicating that deficits in peripheral bone are evident before fracture and are not due to deterioration after the fracture event. Consistent with a prior meta-analysis of retrospective fracture association studies,[8] we found that HR-pQCT can predict fracture at sites other than the periphery, supporting the hypothesis that peripheral bone quality is reflective of clinically relevant sites such as hip and spine.[3,51] We also found that fracture-associated trabecular (but not cortical) deficits were consistently more pronounced in the radius compared with the weight-bearing tibia. However, this does not necessarily mean that the radius has better predictive value than the tibia. A cadaver study by Kroker and colleagues showed that tibial parameters were better correlated with femoral and lumbar failure loads compared with radial parameters.[51] Wang and colleagues showed that the association between vertebral fractures and FEA-derived failure load was slightly better in the tibia (odds ratio [OR] = 1.41, 95% CI 1.04 to 1.78) than in the radius (OR = 1.15, 95% CI 0.98 to 1.32).[52] Similarly, Zhu and

colleagues reported that tibial failure load (OR = 3.85, 95% CI 1.46 to 10.0) was more strongly associated with hip fracture than radial failure load (OR = 2.54, 95% CI 1.09 to 5.92).[53] These data suggest that FEA parameters for the load-bearing tibia may be more representative of clinically relevant fracture sites, which are often load-bearing (eg, femur, vertebra) unlike the radius.

HR-pQCT reproducibility

The short- and long-term intra- and interscanner reproducibility of HR-pQCT measurements is necessary to detect and monitor changes in bone density and microstructure over time in individual patients, compare patient data with normative reference populations for diagnostic classification,[12,54] and pool data from different scanners in multicenter trials. A sufficiently powered large trial can discern treatment effects regardless of good or poor HR-pQCT reproducibility because measurement error is random, and the average error tends to zero as the sample size increases. However, poor reproducibility has significant implications on early proof-of-principle and hypothesis-generating studies with small sample sizes and short monitoring time intervals, as well as in the clinic where patients are scanned once, and sample size cannot be exploited to offset measurement errors.

We conducted a rapid meta-analysis of short-term in vivo reproducibility of HR-pQCT measures assessed by the XCT scanner.[38,40–50] Density-related measures had the best reproducibility of 0.8–2.0%, which was comparable to DXA-derived aBMD reproducibility ($CV\%_{RMS}$ = 0.8–2.3%).[55] The signal-to-noise ratio (SNR) of HR-pQCT and DXA can be calculated as the ratio of fracture-associated differences to LSC, thus allowing for a direct comparison of DXA and HR-pQCT performance. For DXA-derived femoral neck aBMD, the reported fracture-associated difference was −6.3%[56] and LSC was 2.2–6.4%, which coincides with a SNR of 1.0 to 2.9. In comparison, for HR-pQCT, Tt.vBMD fracture-associated deficits in the radius were −9.3% and LSC was 3.9%, resulting in a SNR of 2.4. Similarly, in the tibia, Tt.vBMD fracture-associated deficits were −9.4% and LSC was 2.1, resulting in a SNR of 4.5. These data suggest that signal-to-noise ratio achieved by HR-pQCT is at least equivalent, if not better, than that achieved by the gold-standard DXA.

Although there were significant fracture-associated deficits in microarchitectural and FEA parameters, the short-term reproducibility reported for the XCT scanner (82 μm voxel size) limits the use of several parameters in clinical screening or monitoring applications. By comparison, the higher-resolution XCT II scanner (61 μm voxel size) offers substantial improvements in reproducibility for all trabecular parameters (XCT II: 0.8–2.4%; XCT: 4.1–4.9%) as well as Ct.Th (XCT II: 1.1–1.2%; XCT: 1.6–3.6%) but not density-based measures (XCT II: 0.6–1.5%; XCT: 0.8–2.0%) or Ct.Po (XCT II: 11.0–13.3%; XCT: 6.2–12.5%).[49] The XCT II scanner has a nominal voxel size that enables direct measurement and improved reproducibility of trabecular parameters,[57] unlike indirect methods used by the XCT, which have limited accuracy and precision.[37,41,58] The lack of improvement in Ct.Po precision was unexpected considering the improved precision in other parameters. An explanation offered by Chiba and colleagues was that Ct.Po estimates are small values, and so even the slightest deviation in measurement can manifest as a disproportionately pronounced error.[49] Ellouz and colleagues also suggested that the default use of 2D area-matching, rather than 3D image registration, can result in angular deviations and, consequently, analysis of slightly different regions, thereby impacting Ct.Po measurement precision.[43] Although additional studies are needed to verify the precision of XCT II-derived parameters, the improved reproducibility offered by the second-generation scanner is promising. Due to these ongoing improvements in technology and methodology, we must emphasize that establishing universal parameter-specific LSC thresholds for evaluating the reliability of measurements is impractical. Instead, we urge that individual centers derive their own LSC thresholds and use those reported in this study as a reference.

Beyond scanner resolution, the reproducibility of HR-pQCT parameters is influenced by several factors, including image grade (motion artifacts, noise, density),[42] operator skill,[48] registration protocol (3D image registration improved precision by 8–23% compared with 2D area-matching),[41,43] reference line (positioning of VOIs),[42,48] and scanned region (radius versus tibia).[42,45] Regrettably, there remains a lack of consensus and standardization of methods for HR-pQCT imaging and analysis, as well as for estimation methods, such as for Ct.Po, Ct.Th, and FEA measures. Only recently has the community coalesced toward scanning a relative position at 4.0% (radius) and 7.3% (tibia) of total bone length in adults to avoid limb length biases that arise when scanning at a fixed position.[59] Although efforts have been made in terms of providing training tools[48] and establishing normative databases, there is an urgent need for the HR-pQCT community to consolidate current practices and establish standardized and highly reproducible imaging and analysis protocols and reporting guidelines akin to those widely adopted for rodent μCT studies.[60]

Fracture prediction

Having established that deficits in HR-pQCT measures are significantly associated with fracture, we assessed which parameters had fracture-associated deficits that could be discerned from measurement error at the individual level, as opposed to the cohort level where measurement errors are a negligible factor. Parameters for which fracture-associated deficits exceeded the LSC were deemed reliable predictors of fracture at the individual level.

Tt.vBMD and Tb.vBMD (and related Tb.BV/TV) were found to be reliable predictors of fracture, owing to the high reproducibility of density-based measures and the strong association between Tt.vBMD or Tb.vBMD and fracture, which was consistent with the meta-analysis by Wong.[8] The same meta-analysis also reported that deficits in radial and tibial Ct.vBMD and Ct.Th. were strongly associated with all fractures; however, after accounting for the reproducibility attained by the XtremeCT, only tibial Ct.vBMD and Ct.Th could reliably discern fracture-associated deficits from precision error. Importantly, the Bone Microarchitecture International Consortium (BoMIC), which combined individual-level prospective data from eight cohorts (7254 individuals, mean age 69 ± 9 years), had reported that these parameters—which we showed to be reliable—improve fracture prediction beyond femoral neck aBMD or fracture risk assessment tool (FRAX) scores alone.[56]

Validation studies have demonstrated that FEA parameters are better predictors of bone strength than bone density.[61–63] This was confirmed by the BoMIC study, which reported superior fracture risk prediction by failure load (hazard ratio [HR] = 1.82 to 1.98) compared with all other HR-pQCT parameters (HR = 1.09 to 1.44), even after adjusting for DXA femoral neck aBMD.[56] However, our work suggests that the superior performance of failure load in fracture risk prediction is currently limited by

measurement precision. Thus, although there is strong evidence demonstrating that failure load is the best predictor of incident fracture at the cohort level, HR-pQCT cannot reliably discern fracture-associated differences in failure load from precision error at the individual level. We estimate that failure load LSC must be improved to below 5% for failure load to be used as a reliable predictor in a clinical setting.

HR-pQCT and bone diseases

This meta-analysis was restricted to individuals with no comorbidities or underlying conditions, aside from uncomplicated age-related and/or postmenopausal osteoporosis. However, many pathologies are associated with a heightened risk of fracture, and HR-pQCT may be useful in guiding decisions to treat patients as well as to monitor the efficacy of treatment over time. For example, several groups have compared adults with osteogenesis imperfecta (OI) to apparently healthy subjects and showed severe trabecular deficits in radial and tibial parameters, especially in Tb.vBMD (−21% to −38.5%), Tb.N (−21.4% to −49.4%), and Tb.Sp. (+54.5% to +77.8%).[64–66] Although alterations were more severe in moderate–severe OI (type III/IV) than in mild OI (type I),[65] the trabecular deterioration in all types of OI exceeded that observed in fracture subjects in the current study. Recently, Rolvien and colleagues used HR-pQCT to compare bone microarchitecture in adults with OI and sex- and age-matched subjects with early-onset osteoporosis (EOOP; a condition distinct from postmenopausal osteoporosis in that it is characterized by low bone mass and the occurrence of fragility fractures before the age of 50 years) and healthy subjects. In accordance with other publications, HR-pQCT revealed significant differences in radial and tibial Tb.vBMD as well as in a number of bone geometry and microstructural parameters in the OI cohort compared with healthy subjects. The EOOP cohort also showed a number of significant differences compared with healthy subjects, but the pattern was somewhat different from that for the OI cohort.[66]

Together these studies suggest that pathological deficits in HR-pQCT bone parameters are at least the same if not larger than deficits associated with fracture in apparently healthy individuals and that HR-pQCT assessment may be valuable in longitudinally monitoring disease progression and treatment efficacy.

Study strengths and limitations

The principal strength of this study was being able to relate expected fracture-associated differences to what can be reproducibly detected in a clinical setting using the XtremeCT, while at the same time being PRISMA-compliant and statistically powered and considering the influences of study quality, bias, and heterogeneity. However, our study is not without limitations. Foremost, we must emphasize that our primary outcome, the percentage difference between HR-pQCT parameters obtained from fracture and control subjects, is not a measure of fracture risk, nor should it be interpreted as such (by contrast, the BoMIC study reports hazard ratios that convey the association between fracture-associated changes in HR-pQCT measures and incident fracture risk[56]). For instance, it would be inappropriate to conclude that tibial Tb.vBMD (difference = −9.1%) is a better fracture-risk predictor than tibial Ct.vBMD (difference = −2.8%) on the basis that it has a larger fracture-associated difference, since we have no information on how a unit difference in either measure relates to overall difference in fracture risk. However, using the same example, we can conclude that individuals at risk of fracture will, on average, have a −9.1% and −2.8% deficit in tibial Tb.vBMD and Ct.vBMD, respectively. The studies included in this meta-analysis were predominantly retrospective, and although four prospective studies were included, none focused on pediatric or young adult populations, representing a current gap in knowledge. Owing to the lack of individual-level data, the current meta-analysis was restricted to cohort-level data, which introduces a risk of aggregation bias[67] and limits investigation of individual-level covariates (eg, bisphosphonate use) and redundancies in predictive performance arising from multicollinearities across parameters.[6,40,68] Use of aggregate data also prevented harmonization of FEA parameters, which are derived using a range of different boundary conditions and tissue properties across studies.[69] By comparison, the BoMIC study harmonized failure loads across cohorts by linearly calibrating values (across various boundary conditions and tissue moduli) to approximate axial conditions with a tissue modulus of 6.829 GPa.[56] Finally, the precision errors used in this study to establish LSC thresholds for reliability were pooled from multiple studies. Although it is possible that certain HR-pQCT parameters may be reliably estimated in one center and not the other, our conclusions pertaining to the reliability of parameters reflect the average performance of each measure and should not be used in place of LSC thresholds derived by individual centers.

Concluding remarks

HR-pQCT has emerged as a powerful noninvasive bone imaging modality capable of assessing volumetric BMD, microarchitecture, and strength, and distinguishing cancellous and cortical bone. To maximize the full potential of HR-pQCT in fracture-risk prediction, standardized imaging and analysis protocols and reporting guidelines for HR-pQCT are urgently needed, and improvement of reproducibility needs to be addressed. In conclusion, this meta-analysis confirms that HR-pQCT has promising clinical utility for fracture-risk prediction and will be pivotal in furthering our understanding of how disease and treatment contribute to changes in bone mass and architecture and ultimately fracture risk.

Disclosures

ACO has received grant funding from Alexion and honoraria from Alexion, BioMarin, and Infomed. AJB has received institutional research support from Ultragenyx. BMW is a consultant for Mereo BioPharma and has received research grants from Mereo BioPharma. BvR is a consultant for Scanco Medical AG. FHG is a consultant to Amgen, Mereo BioPharma, and Novartis, and has received research grants from Amgen and Mereo BioPharma. KMK has received institutional research support and materials from Mereo BioPharma, research materials from Amgen and UCB Pharma, and institutional research support from Cayman Chemical for unrelated studies. KW is an employee of Oxford PharmaGenesis, which received funding to conduct this systematic review, and for medical writing support, from Mereo Biopharma. MMB is an independent contractor for Oxford PharmaGenesis, which received funding to conduct this systematic review, and for medical writing support, from Mereo Biopharma. NB has received grant funding from Alexion, Amgen, and Biomarin, honoraria from Alexion, Kyowa Kirin, and Mereo BioPharma, and is a consultant for Mereo BioPharma and UCB. PTL is an independent consultant to various biotech and academics,

including Mereo BioPharma. SNM has received research grants paid to the research institution by Amgen and Merck.

Acknowledgments

This study was supported by the Mereo BioPharma Group.

Authors' roles: Conceptualization: AH. Methodology: KW, MMB, NM, SVK, and AH. Data collection: MMB, KW, and NM. Formal analysis: MMB, KW, NM, and SVK. Interpretation: NM, NB, AJB, LF, AH, KMK, PTL, ACO, JS, BvR, SNM, BMW, SVK, FHG, and DL. Manuscript preparation: MMB, KW, NM, and SVK. All authors contributed to the critical revision and approval of the final manuscript.

References

1. Kanis JA, McCloskey EV, Johansson H, Oden A, Melton LJ III, Khaltaev N. A reference standard for the description of osteoporosis. Bone. 2008;42(3):467–75.
2. Siris ES, Chen YT, Abbott TA, et al. Bone mineral density thresholds for pharmacological intervention to prevent fractures. Arch Intern Med. 2004;164(10):1108–12.
3. Burt LA, Manske SL, Hanley DA, Boyd SK. Lower bone density, impaired microarchitecture, and strength predict future fragility fracture in postmenopausal women: 5-year follow-up of the Calgary CaMos cohort. J Bone Miner Res. 2018;33(4):589–97.
4. Sornay-Rendu E, Boutroy S, Duboeuf F, Chapurlat RD. Bone microarchitecture assessed by HR-pQCT as predictor of fracture risk in postmenopausal women: the OFELY study. J Bone Miner Res. 2017;32(6):1243–51.
5. Szulc P, Boutroy S, Chapurlat R. Prediction of fractures in men using bone microarchitectural parameters assessed by high-resolution peripheral quantitative computed tomography—the prospective STRAMBO study. J Bone Miner Res. 2018;33(8):1470–9.
6. Biver E, Durosier-Izart C, Chevalley T, van Rietbergen B, Rizzoli R, Ferrari S. Evaluation of radius microstructure and areal bone mineral density improves fracture prediction in postmenopausal women. J Bone Miner Res. 2018;33(2):328–37.
7. Nishiyama KK, Shane E. Clinical imaging of bone microarchitecture with HR-pQCT. Curr Osteoporos Rep. 2013;11(2):147–55.
8. Wong AK. A comparison of peripheral imaging technologies for bone and muscle quantification: a technical review of image acquisition. J Musculoskelet Neuronal Interact. 2016;16(4):265–82.
9. Ruegsegger P, Kalender WA. A phantom for standardization and quality control in peripheral bone measurements by PQCT and DXA. Phys Med Biol. 1993;38(12):1963.
10. Cheung AM, Majumdar S, Brixen K, et al. 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.
11. Gluer 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.
12. Blake GM, Fogelman I. How important are BMD accuracy errors for the clinical interpretation of DXA scans? J Bone Miner Res. 2008;23(4):457–62.
13. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6(7):e1000097.
14. Mikolajewicz N, Komarova SV. Meta-analytic methodology for basic research: in review. 2019;10:203.
15. Vesterinen HM, Sena ES, Egan KJ, et al. Meta-analysis of data from animal studies: a practical guide. J Neurosci Methods. 2014;221:92–102.
16. Bielesz B, Patsch JM, Fischer L, et al. Cortical porosity not superior to conventional densitometry in identifying hemodialysis patients with fragility fracture. PLoS One. 2017;12(2):e0171873.
17. Cejka D, Patsch JM, Weber M, et al. Bone microarchitecture in hemodialysis patients assessed by HR-pQCT. Clin J Am Soc Nephrol. 2011;6(9):2264–71.
18. Jamal S, Cheung AM, West S, Lok C. Bone mineral density by DXA and HR pQCT can discriminate fracture status in men and women with stages 3 to 5 chronic kidney disease. Osteoporos Int. 2012;23(12):2805–13.
19. Nickolas TL, Stein E, Cohen A, et al. Bone mass and microarchitecture in CKD patients with fracture. J Am Soc Nephrol. 2010;21(8):1371–80.
20. West SL, Lok CE, Langsetmo L, et al. Bone mineral density predicts fractures in chronic kidney disease. J Bone Miner Res. 2015;30(5):913–9.
21. Li EK, Zhu TY, Tam LS, et al. Bone microarchitecture assessment by high-resolution peripheral quantitative computed tomography in patients with systemic lupus erythematosus taking corticosteroids. J Rheumatol. 2010;37(7):1473–9.
22. Paupitz JA, Lima GL, Alvarenga JC, Oliveira RM, Bonfa E, Pereira RMR. Bone impairment assessed by HR-pQCT in juvenile-onset systemic lupus erythematosus. Osteoporos Int. 2016;27(5):1839–48.
23. Ostertag A, Collet C, Chappard C, et al. A case-control study of fractures in men with idiopathic osteoporosis: fractures are associated with older age and low cortical bone density. Bone. 2013;52(1):48–55.
24. Pasoto SG, Augusto KL, Alvarenga JC, et al. Cortical bone density and thickness alterations by high-resolution peripheral quantitative computed tomography: association with vertebral fractures in primary Sjogren's syndrome. Rheumatology (Oxford, England). 2016;55(12):2200–11.
25. Klingberg E, Lorentzon M, Göthlin J, et al. Bone microarchitecture in ankylosing spondylitis and the association with bone mineral density, fractures, and syndesmophytes. Arthritis Res Ther. 2013;15(6):R179.
26. Ackerman KE, Cano Sokoloff N, De Nardo Maffazioli G, Clarke HM, Lee H, Misra M. Fractures in relation to menstrual status and bone parameters in young athletes. Med Sci Sports Exerc. 2015;47(8):1577–86.
27. Patsch JM, Burghardt AJ, Yap SP, et al. Increased cortical porosity in type 2 diabetic postmenopausal women with fragility fractures. J Bone Miner Res. 2013;28(2):313–24.
28. Fink HA, Langsetmo L, Vo TN, Orwoll ES, Schousboe JT, Ensrud KE. Association of high-resolution peripheral quantitative computed tomography (HR-pQCT) bone microarchitectural parameters with previous clinical fracture in older men: the Osteoporotic Fractures in Men (MrOS) study. Bone. 2018;113:49–56.
29. Chevalley T, Bonjour JP, van Rietbergen B, Ferrari S, Rizzoli R. Fractures during childhood and adolescence in healthy boys: relation with bone mass, microstructure, and strength. J Clin Endocrinol Metab. 2011;96(10):3134–42.
30. Farr JN, Amin S, Melton LJ 3rd, et al. Bone strength and structural deficits in children and adolescents with a distal forearm fracture resulting from mild trauma. J Bone Miner Res. 2014;29(3):590–9.
31. Maatta M, Macdonald HM, Mulpuri K, McKay HA. Deficits in distal radius bone strength, density and microstructure are associated with forearm fractures in girls: an HR-pQCT study. Osteoporos Int. 2015;26(3):1163–74.
32. Bala Y, Bui QM, Wang XF, et al. Trabecular and cortical microstructure and fragility of the distal radius in women. J Bone Miner Res. 2015;30(4):621–9.
33. Bala Y, Zebaze R, Ghasem-Zadeh A, et al. Cortical porosity identifies women with osteopenia at increased risk for forearm fractures. J Bone Miner Res. 2014;29(6):1356–62.
34. Bjornerem A, Bui QM, Ghasem-Zadeh A, Hopper JL, Zebaze R, Seeman E. Fracture risk and height: an association partly accounted for by cortical porosity of relatively thinner cortices. J Bone Miner Res. 2013;28(9):2017–26.
35. Zebaze R, Ghasem-Zadeh A, Mbala A, Seeman E. A new method of segmentation of compact-appearing, transitional and trabecular compartments and quantification of cortical porosity from high resolution peripheral quantitative computed tomographic images. Bone. 2013;54(1):8–20.

36. Schwarzer G, Carpenter JR, Rücker G. Meta-analysis with R. Cham: Springer International Publishing; 2015 pp 107–41.
37. Laib A, Hauselmann HJ, Ruegsegger P. In vivo high resolution 3D-QCT of the human forearm. Technol Health Care. 1998;6(5–6):329–37.
38. 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;47(3):519–28.
39. Zebaze R, Seeman E, Mbala A, Ghasemzadeh A, Mackie E, Bohte A. Method and system for image analysis of selected tissue structures. Google Patents; 2015.
40. 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.
41. MacNeil JA, Boyd SK. Improved reproducibility of high-resolution peripheral quantitative computed tomography for measurement of bone quality. Med Eng Phys. 2008;30(6):792–9.
42. Engelke K, Stampa B, Timm W, et al. 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.
43. 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.
44. Paggiosi MA, Eastell R, Walsh JS. Precision of high-resolution peripheral quantitative computed tomography measurement variables: influence of gender, examination site, and age. Calcif Tissue Int. 2014;94(2):191–201.
45. Kawalilak CE, Johnston JD, Olszynski WP, Leswick DA, Kontulainen SA. Comparison of short-term in vivo precision of bone density and microarchitecture at the distal radius and tibia between postmenopausal women and young adults. J Clin Densitom. 2014;17(4):510–7.
46. Kawalilak CE, Johnston JD, Olszynski WP, Kontulainen SA. Least significant changes and monitoring time intervals for high-resolution pQCT-derived bone outcomes in postmenopausal women. J Musculoskelet Neuronal Interact. 2015;15(2):190–6.
47. Kawalilak CE, Kontulainen SA, Amini MA, Lanovaz JL, Olszynski WP, Johnston JD. In vivo precision of three HR-pQCT-derived finite element models of the distal radius and tibia in postmenopausal women. BMC Musculoskel Dis. 2016;17(1):389.
48. Bonaretti S, Vilayphiou N, Chan CM, et al. Operator variability in scan positioning is a major component of HR-pQCT precision error and is reduced by standardized training. Osteoporos Int. 2017;28(1): 245–57.
49. Chiba K, Okazaki N, Kurogi A, et al. Precision of second-generation high-resolution peripheral quantitative computed tomography: intra- and intertester reproducibilities and factors involved in the reproducibility of cortical porosity. J Clin Densitom. 2018;21(2): 295–302.
50. Kawalilak CE, Bunyamin AT, Bjorkman KM, Johnston JD, Kontulainen SA. Precision of bone density and micro-architectural properties at the distal radius and tibia in children: an HR-pQCT study. Osteoporos Int. 2017;28(11):3189–97.
51. Kroker A, Plett R, Nishiyama KK, McErlain DD, Sandino C, Boyd SK. Distal skeletal tibia assessed by HR-pQCT is highly correlated with femoral and lumbar vertebra failure loads. J Biomech. 2017;59:43–9.
52. Wang J, Stein EM, Zhou B, et al. Deterioration of trabecular plate-rod and cortical microarchitecture and reduced bone stiffness at distal radius and tibia in postmenopausal women with vertebral fractures. Bone. 2016;88:39–46.
53. Zhu TY, Hung VWY, Cheung W-H, Cheng JCY, Qin L, Leung K-S. Value of measuring bone microarchitecture in fracture discrimination in older women with recent hip fracture: a case-control study with HR-pQCT. Sci Rep. 2016;6:34185.
54. Kiebzak GM, Faulkner KG, Wacker W, Hamdy R, Seier E, Watts NB. Effect of precision error on T-scores and the diagnostic classification of bone status. J Clin Densitom. 2007;10(3):239–43.
55. El Maghraoui A, Achemlal L, Bezza A. Monitoring of dual-energy X-ray absorptiometry measurement in clinical practice. J Clin Densitom. 2006;9(3):281–6.
56. Samelson EJ, Broe KE, Xu H, et al. 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. Lancet Diabetes Endocrinol. 2019;7(1):34–43.
57. 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.
58. Krause M, Museyko O, Breer S, et al. Accuracy of trabecular structure by HR-pQCT compared to gold standard μCT in the radius and tibia of patients with osteoporosis and long-term bisphosphonate therapy. Osteoporos Int. 2014;25(5):1595–606.
59. 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. 2017;28(7): 2115–28.
60. Bouxsein ML, Boyd SK, Christiansen BA, Guldberg RE, Jepsen KJ, Muller R. Guidelines for assessment of bone microstructure in rodents using micro-computed tomography. J Bone Miner Res. 2010;25(7):1468–86.
61. Varga P, Dall'Ara E, Pahr DH, Pretterklieber M, Zysset PK. Validation of an HR-pQCT-based homogenized finite element approach using mechanical testing of ultra-distal radius sections. Biomech Model Mechanobiol. 2011;10(4):431–44.
62. 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.
63. Pistoia W, van Rietbergen B, Lochmüller EM, 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. 2002;30(6): 842–8.
64. Folkestad L, Hald JD, Hansen S, et al. 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.
65. Kocijan R, Muschitz C, Haschka J, et al. Bone structure assessed by HR-pQCT, TBS and DXL in adult patients with different types of osteogenesis imperfecta. Osteoporos Int. 2015;26(10):2431–40.
66. Rolvien T, Sturznickel J, Schmidt FN, et al. Comparison of bone microarchitecture between adult osteogenesis imperfecta and early-onset osteoporosis. Calcif Tissue Int. 2018;103(5):512–21.
67. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. Heart Lung Vessel. 2013;5(4):219–25.
68. Boutroy S, Van Rietbergen B, Sornay-Rendu E, Munoz F, Bouxsein ML, Delmas PD. Finite element analysis based on in vivo HR-pQCT images of the distal radius is associated with wrist fracture in postmenopausal women. J Bone Miner Res. 2008;23(3):392–9.
69. 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;80:63–71.
70. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Introduction to meta-analysis. Chichester: John Wiley & Sons; 2011.

Appendix A

Supplemental Materials and Methods

Analysis of heterogeneity and bias

To quantify the extent of inconsistency (ie, heterogeneity) between data sets, we calculated Q and I^2 heterogeneity statistics. Q is a measure of total variation and was calculated as the

sum of the weighted squared differences between study-level differences dif_i and the fixed effect estimate $\hat{dif}_{FE}$:

$$Q = \sum_{i=1}^{N} \left(se(dif_i)^{-2} \cdot \left(dif_i - \hat{dif}_{FE} \right)^2 \right) \quad (A1)$$

where i is the study index, $\hat{dif}_{FE} = \frac{\sum_i se(dif_i)^{-2} dif_i}{\sum_i se(dif_i)^{-2}}$ and N is the number of studies.

Q is a chi-square-distributed statistic with $N - 1$ degrees of freedom (df). The p value corresponding to the Q statistic was used to test the null hypothesis that all data sets reported the same effect. Q statistics derived for subgroups were denoted Q_{within}. I^2 is a related heterogeneity statistic that describes the percentage of variance that is due to heterogeneity:

$$I^2 = \left(\left(\frac{Q}{df} \right) - 1 \right) \left(\frac{Q}{df} \right)^{-1} \quad (A2)$$

Heterogeneity was further assessed by single-study and cumulative-study exclusion plots.[14] The homogeneity threshold T_H was calculated from cumulative exclusion analysis and specifies the percentage of studies that need to be removed (according to maximal Q-reduction criteria) before a homogeneous set of studies is attained, as determined by the p value P_Q corresponding to the Q heterogeneity statistic.[14] Biases were visually assessed using funnel plots, and the theoretical impact of bias was determined by trim-and-fill analysis.[70]

Meta-analysis

Study-level outcomes were synthesized under the assumptions of a random effects model to obtain an overall outcome $\hat{dif}$:

$$\hat{dif} = \frac{\sum_i (dif_i \cdot w_i)}{\sum_i (w_i)} \quad (A3)$$

Where the random effects study-level weights w_i were estimated as:

$$w_i = \frac{1}{se(dif_i)^2 + \tau^2} \quad (A4)$$

And the interstudy variance τ^2 was approximated using the DerSimonian–Laird estimator:

$$\tau^2 = \frac{Q - (N-1)}{c} \quad (A5)$$

where $c = \sum_i se(dif_i)^{-2} - \frac{\sum_i \left(se(dif_i)^{-2} \right)^2}{\sum_i se(dif_i)^{-2}}$

Q is the heterogeneity statistic (Equation A1), c is a scaling factor, and N is the number of data sets. The standard error $se\left(\hat{dif}\right)$ corresponding to the overall outcome was estimated as:

$$se\left(\hat{dif}\right) = \frac{1}{\sqrt{\sum_i w_i}} \quad (A6)$$

and confidence intervals were constructed using critical values $z_{1-\alpha_2}$ obtained from a z-distribution:

$$\pm CI = \pm z_{1-\alpha/2} \cdot se\left(\hat{dif}\right) \quad (A7)$$

where $\alpha = 0.05$ corresponds to a 95% significance level. Outcomes were compared using the Z test.[70]