Original Article

The impact of motion induced artifacts in the evaluation of HR-pQCT scans of the scaphoid bone: an assessment of inter- and intraobserver variability and quantitative parameters

Stefan Benedikt[1], Lukas Horling[1], Kerstin Stock[1], Gerald Degenhart[2], Johannes Pallua[1], Gernot Schmidle[1], Rohit Arora[1]

[1]Department of Orthopaedics and Traumatology, Medical University Innsbruck, Innsbruck, Austria; [2]Department of Radiology, Medical University Innsbruck, Innsbruck, Austria

Contributions: (I) Conception and design: S Benedikt, L Horling, K Stock, G Schmidle, R Arora; (II) Administrative support: R Arora, G Degenhart, J Pallua; (III) Provision of study materials or patients: S Benedikt, L Horling, K Stock; (IV) Collection and assembly of data: S Benedikt, L Horling, K Stock, G Degenhart; (V) Data analysis and interpretation: S Benedikt, G Degenhart, J Pallua, G Schmidle; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr.med.univ. Lukas Horling. Department of Orthopaedics and Traumatology, Medical University Innsbruck, Anichstrasse 35, AT-6020 Innsbruck, Austria. Email: lukas.horling@tirol-kliniken.at.

Background: *In-vivo* high-resolution peripheral quantitative computed tomography (HR-pQCT) has high potential in scaphoid bone pathologies' scientific and clinical fields. The manufacturer's visual grading scale (VGS) classifies motion artifacts and divides scans into five quality grades ranging from grade 1 (good quality) to grade 5 (poor quality). This prospective study aimed to investigate the feasibility of the VGS and the influence of image quality on bone density and microarchitecture parameters for the scaphoid bone.

Methods: Within one year, twenty-two patients with scaphoid fractures received up to six scans of their fractured and contralateral wrist (each consisting of three stacks) using second-generation HR-pQCT (total 256 scans). Three experienced observers graded each stack following the visual grading system, and inter- and intraobserver variability were assessed. The contralateral uninjured scaphoids were then compared pairwise within each patient to high-quality grade 1 scans to determine the influence of image quality on density and microarchitecture parameters.

Results: Inter- and intraobserver variability among the three observers significantly revealed fair to moderate agreement, P<0.001 and P<0.05, respectively. Bone volume (BV) fraction tended to increase with poorer image quality but did not exceed four percent. Trabecular bone mineral density (Tb.BMD) decreased with poorer image quality but did not exceed five percent. Trabecular number and trabecular thickness significantly increased by 15.5% and 6.8% at grade five (P<0.001), respectively, and trabecular separation significantly decreased by 13.7% at grade five (P<0.001).

Conclusions: This study revealed a considerable influence of motion on bone morphometry parameters of the scaphoid. Therefore, high image quality must be a central point in studies focusing on the histomorphometry of small objects. The high inter- and intraobserver variability limit the VGS. Future research may focus on other grading systems or automated techniques leading to more consistent and reproducible results. Currently, the use of microarchitectural analysis should be limited to cases without motion artefacts or, at most low graded motion artefacts.

Keywords: High-resolution peripheral quantitative computed tomography (HR-pQCT); visual grading; motion artifacts; histomorphometry; interobserver; intraobserver

Introduction

Scaphoid fractures are the most common carpal fractures. Depending on the localization and grade of dislocation, treatment consists of either conservative therapy with cast immobilization or surgical treatment with osteosynthesis. Missed scaphoid fractures are a significant problem as they might end up as non-unions, leading to severe osteoarthritis of the wrist joint (1). The overall risk for developing a scaphoid non-union range between 2–5% (2). Eighty percent of these non-unions are related to a delayed or incorrect diagnosis (3).

Established imaging methods for diagnosing scaphoid fractures are X-ray, bone scan, magnetic resonance imaging (MRI), and computed tomography (CT), with MRI being the most sensitive and specific method for fracture detection. The CT also has a high specificity but a lower sensitivity; however, it is often preferred to the MRI as it is cheaper and more readily available (1,4,5). High-resolution peripheral quantitative computed tomography (HR-pQCT) represents an innovative option in detecting scaphoid fractures (6-8). Since the first results have only recently been published, its use is not yet widely established in this field. Initially, HR-pQCT was designed to measure bone density and quantify the three-dimensional microarchitecture of bone (9). Its clinical value is remaining still marginal due to several reasons, including technical issues, a lack of standardization in scan acquisition and evaluation, and it's cost-related limited availability (10). However, in recent years HR-pQCT has made significant progress in many scientific fields, e.g., in the assessment of the influence of rheumatologic diseases on joint surfaces (11,12), of bone microarchitecture and bone strength in secondary osteoporosis and metabolic bone disorders (10), the effect of several anti-osteoporotic drugs on bone quality (13) as well as in the evaluation of fracture healing (14-16) and research on fracture mechanisms of the distal radius (17,18).

HR-pQCT has the best signal-to-noise ratio and the highest spatial resolution of all tools used in *in-vivo* routine clinical diagnostics. A resolution of 61 µm voxel can be achieved on peripheral extremities. Since radiosensitive organs are only marginally exposed, HR-pQCT generates only a small radiation exposure for the patient with less than 5 µSv effective radiation dose per stack (19-21).

Due to the high resolution, small fissures in the scaphoid bone can be detected more sensitively than with the conventional CT (6,22). Furthermore, the analysis of the bone mineral density and micro-architecture could help to provide insights into fracture mechanisms, the non-union etiology (7), and its biology (23) and quantify the healing process of scaphoid fractures (16).

However, a disadvantage of this imaging method is the comparatively long scan time compared to conventional CT, which predisposes the captured scans to significant motion artifacts (24). For classifying the severity of motion artifacts, the manufacturer provides a visual grading scale (VGS) described by Sode *et al.* (25), which divides HR-pQCT scans into five categories ranging from "no visible artifacts" to "severe motion artifacts". Pialat *et al.* (24) evaluated the influence of motion artifacts of the radius and the tibia scans referring to the VGS and demonstrated considerable error in density and microarchitecture measurements with subject motion.

The influence of motion artifacts on the microarchitecture of scaphoids using the classification described by Sode *et al.* (25) has not yet been evaluated. The scaphoid bone differs from the radius and tibia in both its shape and micro-architecture (26). Due to its thin cortex and the fine trabecular network motion might have a different impact on image quality. Moreover, the subjectivity when classifying the bone into grade 1 to grade 5 seems higher for the scaphoid than for the radius and tibia (7).

This study aimed to investigate the feasibility of the VGS for the scaphoid bone by assessing its inter-and intraobserver variability and the influence of motion-induced artifacts on bone density and microarchitectural parameters.

Methods

Study design and population

The current prospective study is based on scaphoid HR-pQCT scans of 22 patients with unilateral conservatively treated scaphoid fractures (aged >18 years) presenting between May 2018 and December 2020. All patients received up to six scans of their fractured and contralateral scaphoid within 1 year (2, 4, 6, 12 weeks, 6 months,

Figure 1 The fractured and the contralateral scaphoids were scanned 2, 4, 6, 12 weeks, 6 months and 1 year after trauma. Cast immobilization of the fractured wrist ranged between 4 to 12 weeks (blue line). T, trauma; S1–6, scan 1–6.
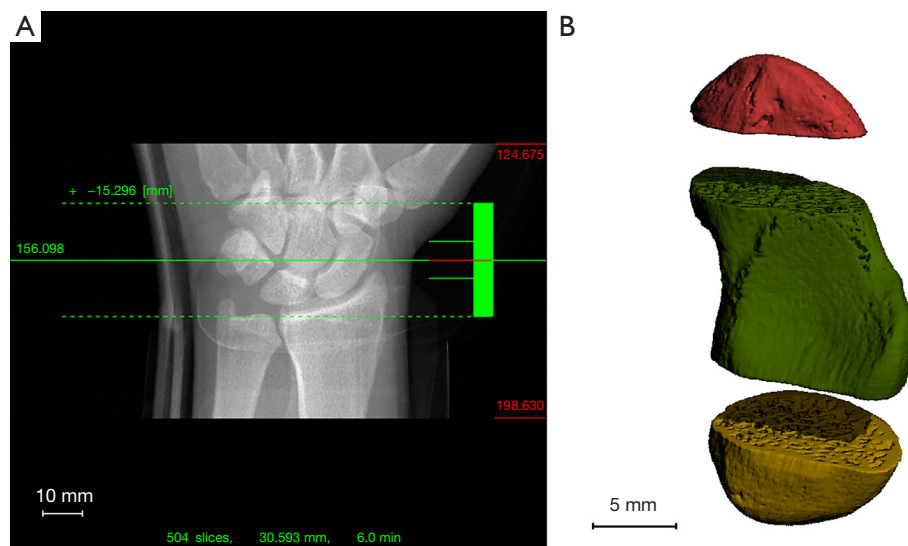
Figure 2 Scans visualized with three stacks of 10.2 mm with the scaphoid centered. (A) HR-pQCT scout view of a right wrist with the scaphoid centered between three stacks of 10.2 mm each. (B) 3D reconstruction of a scaphoid with the proximal stack colored in yellow, the middle stack in green, and the distal stack in red. HR-pQCT, high-resolution peripheral quantitative computed tomography.

and 1 year after trauma). The rationale of scanning the contralateral side during every follow-up visit was to get as much image data as possible to explore the influence of motion artifacts. The fractured wrists were immobilized for four to twelve weeks depending on the fracture type, the clinical symptoms, and the bone consolidation in the follow-up X-rays (*Figure 1*). Depending on the length of immobilization, the first, second and third scans of the fractured side were performed with the patient immobilized in a bellow elbow fibreglass cast, including the thumb; all other scans were performed without a cast.

Patients were excluded in case of pregnancy, a previous ipsilateral scaphoid fracture in medical history, or in case of pre-existing conditions that affect the musculoskeletal system in any form.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics board of the Medical University of Innsbruck, Austria (No. 1259/2017) and informed consent was taken from all the patients.

Scan acquisition

Scans were obtained with a second-generation HR-pQCT (XtremeCTII, Scanco Medical, Switzerland). The scans were visualized with three stacks of 10.2 mm with the scaphoid centered (*Figure 2A,2B*). During scanning, the wrist was immobilized in a thumb-up position with the manufacturer's standard motion restraining holders and the appropriate inflatable pads as previously described (7,24) to obtain reproducible scans and minimize motion during acquisition (*Figure 3A-3C*). Pre-settings of all scans included a resolution of 60.7 μm isovoxels resulting in 504 slices, an integration time of 46 ms, and a voltage and intensity of
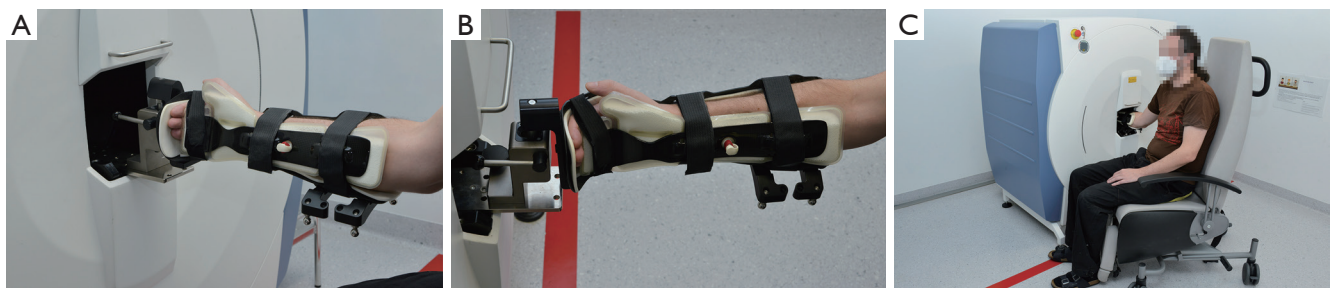
Figure 3 Positioning of the patient (A-C). The wrist immobilized in a thumb-up position with the manufacturer's standard motion restraining holders and the appropriate inflatable pads to obtain reproducible scans and minimize motion during acquisition. These images are published with the participant's consent.

68 kV and 1,460 µA, respectively. The scanning time for one scaphoid was approximately six minutes. According to the manufacturers settings, the radiation dose was 5 µSv per stack, so the maximum radiation dose was 180 µSv in the case of six bilateral scans.

The Scanco medical software package was used for direct post-processing, containing VMS (multiprocessing virtual memory-based operating system, ©Hewlett-Packard, Palo Alto, USA) and image processing language IPL (Image Processing Language, Scanco Medical AG, Bruttisellen, Switzerland).

Image quality grading and variability assessment

Visual grading was performed on fully reconstructed scans with the image processing software ImageJ Version 1.49 (https://imagej.nih.gov/ij/docs/faqs.html). All scans were assessed by three experienced observers (SB, LH, KS) frequently performing research with HR-pQCT over years. To avoid detection bias, none of the selected observers was involved in the scan acquisition before, and all data were anonymized previously. Scans were graded according to the manufacturer's VGS (Scanco Medical, Switzerland), ranging from grade 1 (no motion artifacts) to grade 5 (extreme motion artifacts) (*Figure 4A-4E*). The scaphoid of each stack was evaluated slide by slide and graded according to the most severe occurring motion artifact.

To examine the intraobserver variability, the second evaluation of 33 randomly chosen scans (99 stacks) was performed by each observer at least 2 weeks after the initial reading. The interobserver variability was calculated from the first results of the observers, the intraobserver variability was calculated from the first and the second reading of each observer.

Selection of paired exams of the non-fractured scaphoids for quantitative analysis

All fractured scaphoids were excluded for quantitative analysis since the natural remodelling in the fractured scaphoids changed the bone structure over the repeated scans. Repeated acquisitions from each patient were considered in pairs as already published by Pialat *et al.* (24). Each patient's chronologically first grade 1 scan served as an internal reference. Five pairs were constituted if a patient had six scans (with at least one grade 1 scan). In case of discrepancies in visual grading between the observers, image quality was defined by the median value of the three observer's ratings.

Contouring, segmentation, and quantitative analysis of the non-fractured scaphoids

Depending on the severity of motion artifacts, every 5th to 10th slide was contoured using the manufacturer's automatic contouring algorithm. Due to contouring problems previously described by Bevers *et al.* (7), the lower threshold for binarization was changed from the default to 105 per 1,000, and a manual correction was performed in all cases. The slices in between were interpolated using the manufacturer's morphing algorithm and, if needed, manually corrected.

The manufacturers' dual-threshold algorithm did not allow for reliable and accurate distinctions between cortical and cancellous bone. This might be due to the high variability of the cortical layer thickness and low differences between cortical and cancellous bone mineralisation and thickness. Also, manual contouring of the inner border of the cortex was not feasible as a clear distinction between cortex and dense cancellous bone was not always possible.
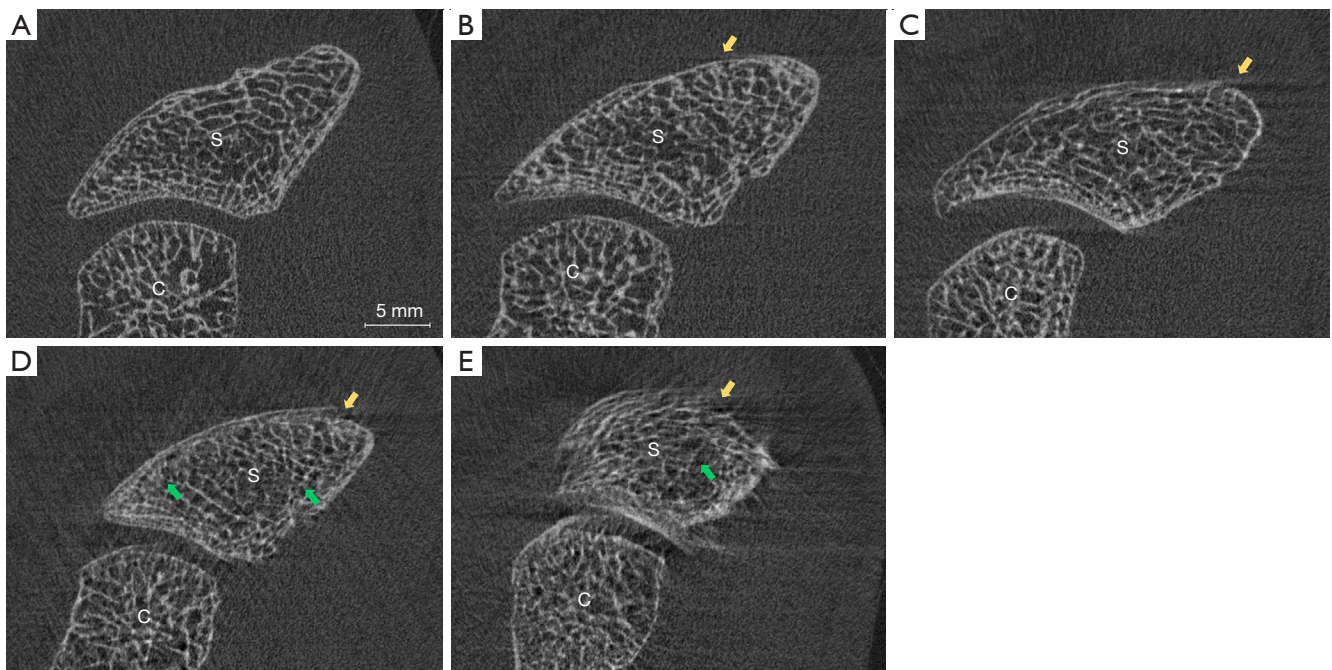
Figure 4 Visual grading scale of the scaphoid. (A) Grade 1, no visible motion artifacts. (B) Grade 2, slight horizontal streaks (yellow arrow). (C) Grade 3, prominent horizontal streaks are visible, but the cortex is intact (yellow arrow). (D) Grade 4, prominent horizontal streaks, minor disruptions of the cortex continuity (yellow arrow), and minor trabeculae smearing (green arrows). (E) Grade 5, prominent horizontal streaking, considerable disruption of the cortical continuity (yellow arrow), considerable trabecular smearing (green arrow). S, scaphoid; C, capitate.

Therefore, the contouring process of the cancellous bone was modified. Semi-automatic contouring of the outer shell combined with a ten-voxel erosion step provided reproducible and sufficient contouring of the cancellous bone compartment (*Figure 5A-5C*). The erosion step was based on the try and error procedure. The value of 10 voxels was the best compromise to remove as much cortical bone as possible while preserving the cancellous bone.

Contoured bones were quantitatively analyzed using the manufacturers' provided evaluation software. The lower threshold was set at 320 mgHA/ccm with a Gaussian filter at σ=0.8 and support =1.0 according to the standard image processing specifications (27).

Since the positioning of the wrists within the manufacturer's restraining holders was not exactly reproducible among the repeated scans, the respective stacks of one patient were not entirely identical. Therefore, of the six repeated scans, only those axial slides were compared that were depicted in all proximal, middle, and distal stacks of this patient, respectively (*Figure 6*).

Measurement parameters included the total volume (TV),

the bone volume (BV), trabecular bone volume fraction (Tb. BV/TV), trabecular number (Tb.N), trabecular thickness (Tb.Th), trabecular separation (Tb.Sp), inhomogeneity of trabecular network (Tb.1/N.SD), connective density (Conn.D), structural model index (SMI), mean trabecular bone mineral density (Tb.BMD) and apparent density. TV and BV were used to calculate Tb.BV/TV but were not included in the statistic since small positioning discrepancies among the six scans resulted in different volumes within the stacks.

Statistical analysis

Statistical analysis was performed with SPSS (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0, Armonk, NY, USA).

Interobserver variability between the three observers was assessed using the Fleiss'-Kappa test. For intraobserver variability, Cohen's Kappa coefficient was calculated. P values <0.05 were considered significant. The strength of agreement was determined according to the classification of
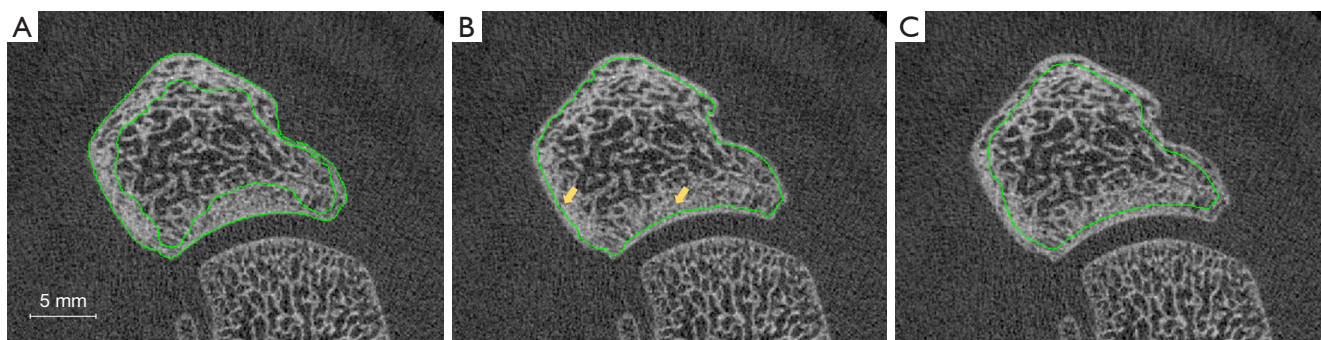
Figure 5 Representation of different contouring strategies. (A) In the scaphoid, the provided dual-threshold algorithm did not achieve good contouring between the complex structure of cortical (outer green circle) and cancellous bone (inner green circle). This might be due to the high variability of the cortical layer thickness and low differences between the mineralization and thickness of cortical and cancellous bone. (B) Also, manual contouring of the inner border of the cortex (green circle) was not feasible as a clear distinction between cortex and dense cancellous bone was not always possible (yellow arrows). (C) Solution: semi-automatic contouring of the outer shell combined with a ten-voxel erosion step provided reproducible and sufficient contouring (green circle) of the cancellous bone compartment.
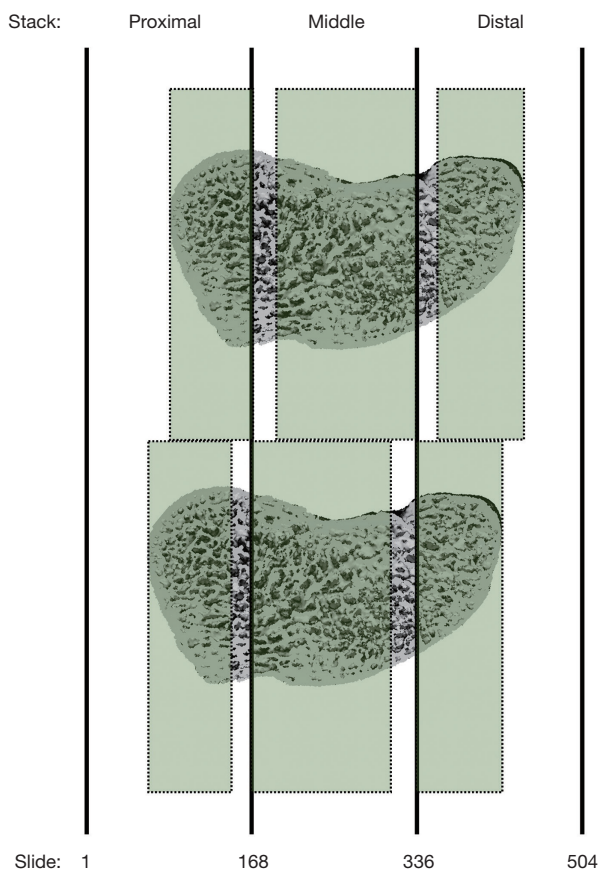
Figure 6 Only those axial slides were compared that the proximal, middle and distal stacks had in common. green field, compared/evaluated area.

Landis and Koch 1977 (28).

Mean percent difference and standard deviation between motion degraded pairs (1-2, 1-3, 1-4, 1-5) and the reference pairs (1-1) were calculated, and Dunnet's test was used to confirm significant changes in image quality among the quality grades. P values <0.05 were considered significant.

To prove that the time interval of one year between the first and last scan did not influence bone density and microarchitecture, the first measurement results of the non-fractured scaphoids were compared with those one year after trauma using a paired *t*-test.

Results

Patients' demographics are shown in supplemental digital content 1 (Table S1).

Descriptive data

A total of 771 stacks of five to six repeated scans of 22 patients' contralateral and fractured wrists were included in the study. After the exclusion of three scans of different patients, each consisting of four stacks (due to an operator error), 759 stacks (253 proximal, middle, and distal stacks, respectively) were included for variability assessment (*Figure* 7). Among the 759 stacks, 62.5% were rated as grade 1 or grade 2, grade 4 or grade 5 were allocated in only 18.6%. *Figure 8* shows the percentages of acquisitions
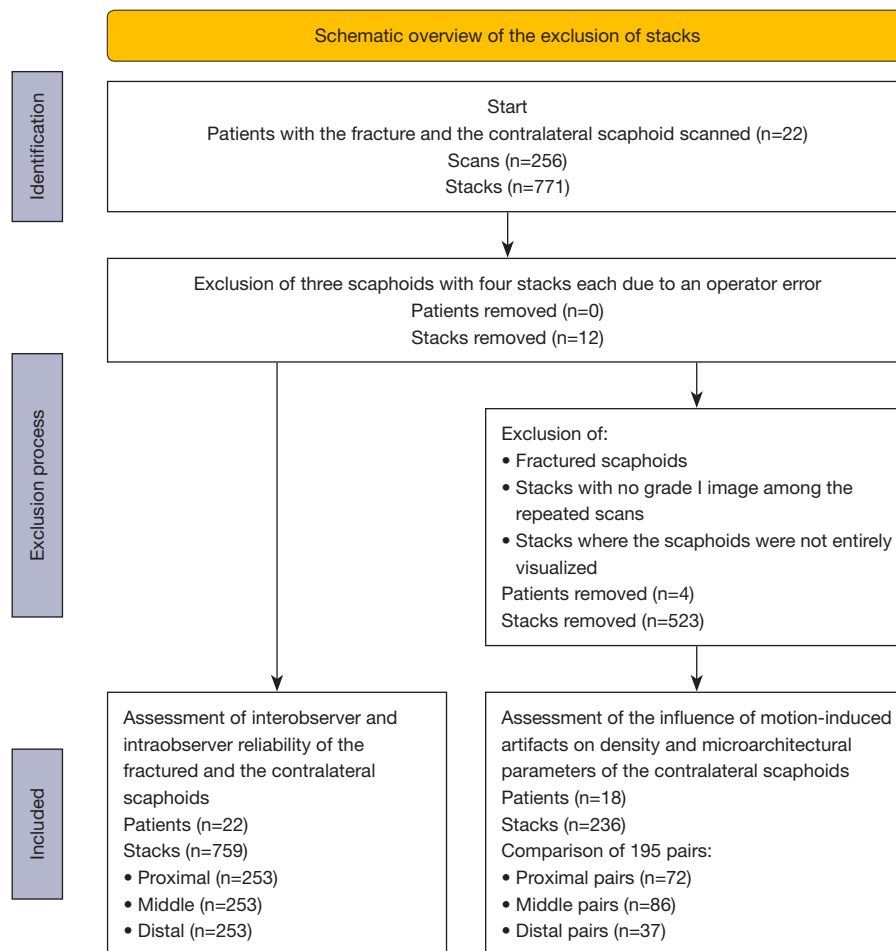
Figure 7 Schematic overview of the exclusion of stacks.

across the quality grades for the proximal, middle, distal, and totality of all stacks.

Among the 378 stacks of the fractured scaphoids, 62.7% were rated as grade 1 or grade 2, grade 4 or grade 5 were allocated in 16.7%. Among the 174 stacks of the fractured scaphoids immobilized with a cast, 70.5% were rated as grade 1 or grade 2, grade 4 or grade 5 were allocated in 10.3%. Among the 381 stacks of the contralateral non-fractured scaphoids, 62.2% were rated as grade 1 or grade 2, grade 4 or grade 5 were allocated in 20.1%.

Interobserver variability

For variability analysis, fractured and contralateral scaphoids were included. The overall interobserver agreement among the three observers for both the fractured and the contralateral scaphoids (n=253) revealed fair to a moderate

agreement with kappa values of 0.47, 0.40, and 0.39 for the proximal, middle, and the distal stack, respectively (P<0.001 in all cases). For only the fractured scaphoids (n=126, ~50%), agreement was fair to moderate with kappa values of 0.43, 0.35, and 0.42 for the proximal, middle, and distal stack respectively (P<0.001). For those fractured scaphoids immobilized with a cast (n=59, ~24%), agreement was fair to moderate with kappa values of 0.41, 0.36, and 0.40 for the proximal, middle, and distal stack, respectively (P<0.001). For the contralateral scaphoids (n=127, ~50%), agreement was fair to moderate with kappa values of 0.50, 0.45, and 0.36 for the proximal, middle, and distal stack, respectively (P<0.001). The lowest interobserver reliability was observed for grade 3. *Table 1* shows the kappa value at each quality grade.

Merging observer's gradings into subgroups with good-quality scans (stacks rated with grades 1 and 2) and poor-
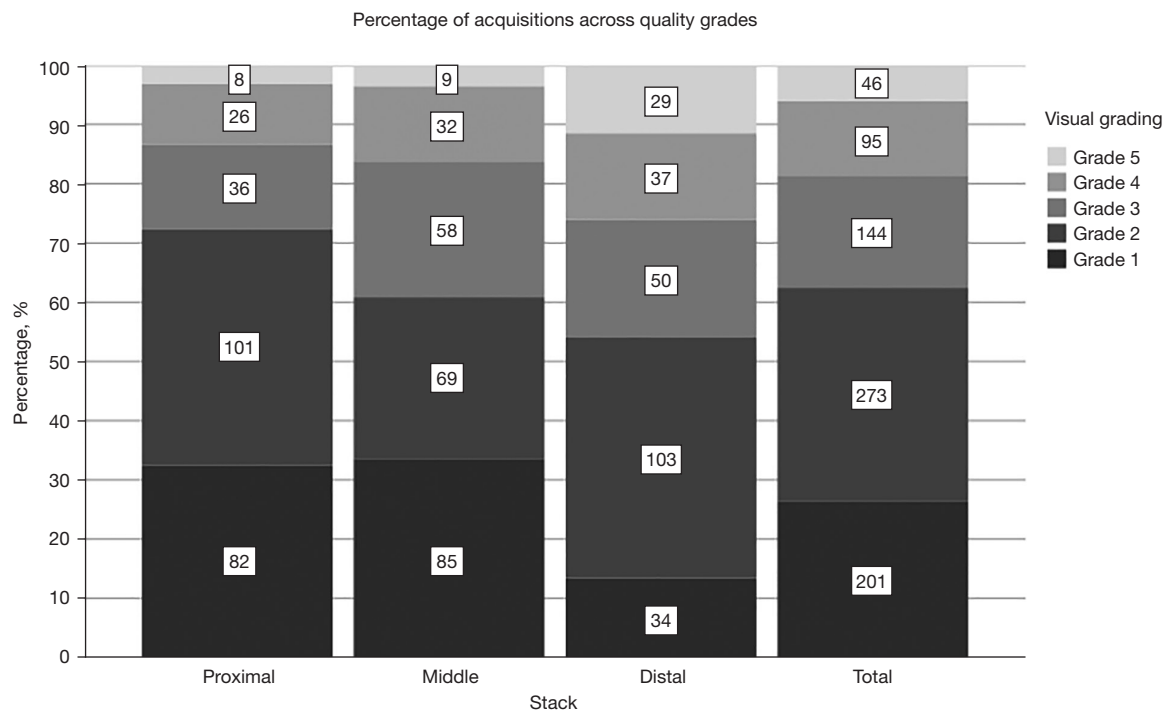
Figure 8 The percentages of acquisitions across the quality grades for the proximal, middle, distal, and totality of all stacks. The y-axis shows the relative results, the values within the bars represent the absolute number.

Table 1 Kappa values of the interobserver variability at each grade according to the visual grading scale for scan quality for the p, m, and d stacks

Grade	All scaphoids			Fractured scaphoids			Non-fractured scaphoids		
	p	m	d	p	m	d	p	m	d
Grade 1	0.690	0.676	0.538	0.694	0.676	0.601	0.684	0.673	0.461
Grade 2	0.432	0.288	0.377	0.348	0.274	0.364	0.516	0.302	0.389
Grade 3	0.228	0.258	0.239	0.199	0.187	0.283	0.258	0.335	0.199
Grade 4	0.371	0.304	0.320	0.244	0.166	0.337	0.431	0.435	0.298
Grade 5	0.481	0.366	0.550	0.555	0.347	0.590	0.407	0.384	0.515

P value of all results <0.001. p, proximal; m, middle; d, distal.

quality scans (stacks rated with grades 3, 4 and 5) revealed a substantial interobserver agreement with kappa values of 0.65, 0.62, and 0.63 for the proximal, middle, and distal stacks, respectively (P<0.001 in all cases).

Merging observer's gradings into subgroups with good-quality scans (stacks rated with grades 1, 2 and 3) and poor-quality scans (stacks rated with grades 4 and 5) revealed a moderate to a substantial interobserver agreement with kappa values of 0.58, 0.54, and 0.66 for the proximal, middle, and distal stacks, respectively (P<0.001 in all cases).

Intraobserver variability

For variability analysis, fractured and contralateral scaphoids were included. The overall average intraobserver variability was fair to moderate, with kappa values of 0.42, 0.43, and 0.37 for the proximal, middle, and distal stack, respectively (P<0.05). *Table 2* shows the intraobserver variability of the three observers in detail.

For only the fractured scaphoids (n=16/33), agreement was fair to moderate with kappa values of 0.33, 0.41, and

1344

Benedikt et al. Influence of motion in HR-pQCT scans of the scaphoid bone

Table 2 Kappa values of the intraobserver variability for visual grading according to the visual grading scale for scan quality for the proximal, middle, and distal stack

Observer	Proximal	Middle	Distal
Observer 1	0.517	0.665	0.408
Observer 2	0.331	0.151	0.225
Observer 3	0.410	0.480	0.463

P value of all results <0.05.

0.23 for the proximal, middle, and distal stack, respectively. Most kappa values were not significant. For those fractured scaphoids immobilized with a cast (n=7/33), agreement was slight to moderate with kappa values of 0.20, 0.44, and 0.25 for the proximal, middle, and distal stack, respectively. Most kappa values were not significant.

For the contralateral scaphoids (n=17/33), agreement was moderate with kappa values of 0.50, 0.44, and 0.46 for the proximal, middle, and distal stack, respectively (P<0.05).

Merging observer's gradings into subgroups with good-quality scans (stacks rated with grades 1 and 2) and poor-quality scans (stacks rated with grades 3, 4 and 5) revealed a moderate to a substantial intraobserver agreement with average kappa values of 0.67, 0.69, and 0.50 for the proximal, middle and distal stack, respectively (P<0.05).

Merging observer's gradings into subgroups with good-quality scans (stacks rated with grades 1, 2 and 3) and poor-quality scans (stacks rated with grades 4 and 5) revealed a moderate to a substantial intraobserver agreement with average kappa values of 0.59, 0.69, and 0.68 for the proximal, middle, and distal stacks, respectively (P<0.05 in all cases).

Paired analysis of the contralateral non-fractured sides

After excluding the scans of the fractured sides, stacks with no grade 1 scan among the repeated scans and stacks where the scaphoids were not entirely visualized, 236 stacks (87 proximal, 104 middle, and 45 distal stacks) of 18 patients were included for pairwise analysis; 195 pairs (72 proximal, 86 middle, and 37 distal pairs) were compared (*Figure 7*). The average relative differences in parameters of each quality grade compared to the internal grade 1 reference are presented in *Figure 9*. Supplemental digital content 2 shows the detailed results in Table S2.

The relative difference of Tb.BV/TV tended to increase with poorer image quality. The difference did not exceed

four percent, and Dunnett's test reached significance between the relative differences of grade 1 and grade 3 and between grade 1 and grade 4 with a P value of <0.001 and 0.026, respectively.

Tb.N and Tb.Th increased by 15.5% (±5.9%) and 6.8% (±1.8%) from grade 1 to grade 5, respectively, Tb.Sp and Tb.1/N.SD decreased by 13.7% (±6.2%) and 21.4% (±10.4%) from grade 1 to grade 5, respectively. Conn. D increased by 21.6% (±29.9%) and Tb.BMD decreased by 4.9% (±2.2%) from grade 1 to grade 5. All differences between grade 1 and grade 5 were significant (P<0.001). Relative differences of apparent density at grade 1 was −0.5% (±2.0%), between grade 1 and grade 3 2.0% (±3.4%) and between grade 1 and grade 5 1.7% (±3.0%). A significant difference was only observed between grade 1 and grade 3 (P<0.05). The mean SMI at grade 1 was −1.5 (±1.7) and slightly increased with poorer image quality (supplemental digital content 3: Table S3).

Influence of time

There was no significant difference between the first and last scans of the non-fractured scapoids after one year for Tb.BV/TV, Tb.N, Tb.Th, Tb.Sp, Tb.1/N.SD, Conn.D, Tb.BMD and apparent density indicating that the one-year follow-up period did not influence density and microstructural parameters in our patient population.

Discussion

Motion artifacts in *in-vivo* HR-pQCT scans are challenging, especially in small bones such as the scaphoid. Motion during scan acquisition can lead to considerable degraded image quality and introduce error, particularly regarding microarchitecture (24,25,29,30). Moreover, the high subjectivity and consequently low reproducibility in artifact grading might limit the current grading system (7,27).

The anatomy of the scaphoid compared to the distal radius and the proximal tibia is different. Macroscopically, they differ in size and shape. Regarding the density and microarchitecture, scaphoids seem to have a thinner cortical shell and a higher degree of mineralization (7,26,31,32). The thin cortical shell may cause cortical interruptions to be detected earlier than in the radius and tibia during motion, resulting in poorer grading of scans according to the manufacturer's visual grading system (Scanco Medical, Switzerland). Considering quantitative analysis, motion could have a different influence on bone density and

Influence of image quality on quantitative parameters
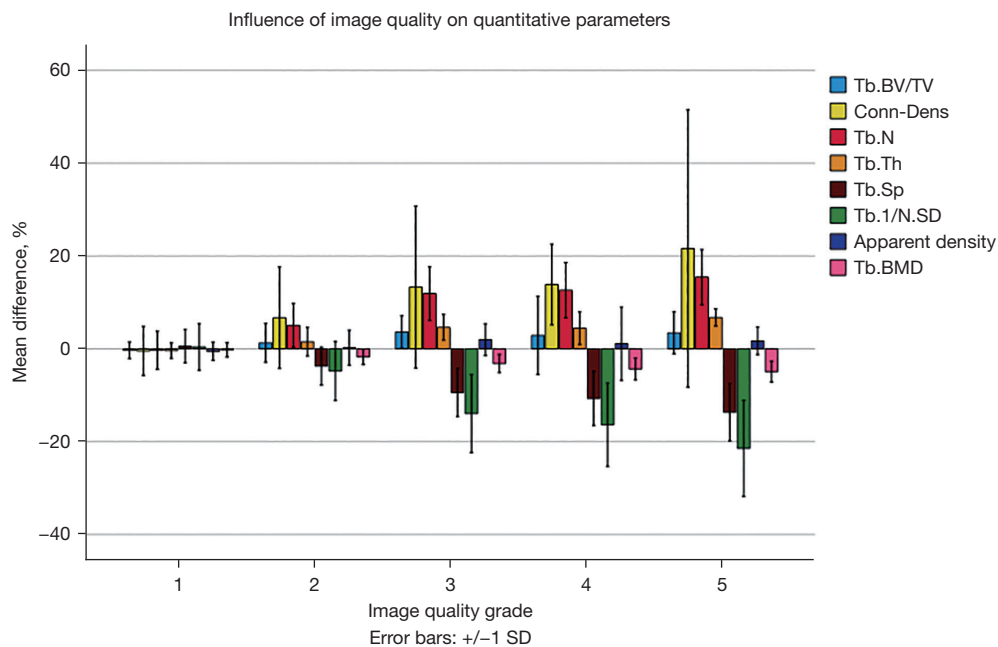
Figure 9 Relative differences between the image quality grades and the internal grade 1 reference of Tb.BV/TV, Tb.N, Tb.Th, Tb.Sp, inhomogeneity of Tb.1/N.SD, Conn.D, apparent density and Tb.BMD. Tb.BV/TV, trabecular bone volume fraction; Tb.N, trabecular number; Tb.Th, trabecular thickness; Tb.Sp, trabecular separation; Tb.1/N.SD, inhomogeneity of trabecular network; Conn.D, connective density; Tb.BMD, trabecular bone mineral density.

microstructural parameters.

This study examined the influence of motion-induced artifacts on the inter-and intraobserver variability for visual grading and on the histomorphometry of the scaphoid.

Inter- and intraobserver variability analysis revealed high subjectivity in rating. If each grade was considered individually, the interobserver and intraobserver agreement among the three observers was only fair to moderate. When pooling the results in two groups with excellent (grades 1 and 2) and poor quality (grades 3 to 5), substantial agreement for interobserver variability and moderate to substantial agreement for intraobserver variability was obtained. Interestingly lowest kappa value for interobserver reliability was reached ad grade 3. This could be related to the difficulty distinguishing between slight and prominent horizontal streaking and between no or minimal trabecular smearing. Regarding the intraobserver reliability lowest kappa value was reached by observer 2. Both aspects underline the high subjectivity and the revealed limitation of this grading. Bevers *et al.* (7) analyzed 85 scaphoids regarding image quality with a one observer standard and post hoc grading. Scans were separated into a good quality group, including grades 1–3, and a poor quality group,

including grades 4 and 5. Subjectivity in grading was higher than previously described as the standard grading missed 85.7% of the scans assessed as poor quality in the post hoc assessment (versus 12% described for the distal radius). However, since for the standard grading, only the single preview slice with a lower resolution, and for the post hoc grading, multiple high-resolution slices were evaluated, results are not comparable to our intraobserver variability. In the study of Pialat *et al.* (24) inter- and intraobserver variability for the distal radius and the distal tibia acquired with a first-generation HR-pQCT were higher than our results. The Cohen's Kappa score between two trained observers was 0.57 and for intraobserver variability 0.68 and 0.74, respectively. The study of Pauchard *et al.* (29) examined the interclass correlation (ICC) between four graders of the own laboratory and two operators from two extern laboratories on distal radii and tibiae using the first generation HR-pQCT. ICC between the intern observers was 0.75 and between all six operators 0.77, indicating good rater agreement.

The poorer results regarding interobserver- and intraobserver variability in this study might be emerged from the different anatomy compared to the radius and the

1346

Benedikt et al. Influence of motion in HR-pQCT scans of the scaphoid bone

tibia (26). Small cortical interruptions are earlier visible than in the radius or the tibia due to the low cortical thickness. Moreover, in extremely fine trabecular network as in the proximal and distal pole of the scaphoid, trabecular smearing can only be assessed to a limited extent.

For quantitative analysis, the fractured wrists were excluded since the natural remodelling in the fractured scaphoids changes the bone structure over time. The non-fractured scaphoid motion artefacts significantly influenced the densitometric and microarchitecture results. Tb.BV/TV, Tb.N, Tb.Th, Conn.D, and apparent density increased, while Tb.Sp, Tb.1/N.SD and Tb.BMD decreased with poorer image quality. In the literature, the influence of motion on measurements of the distal radius and the distal tibia is higher for most parameters than our scaphoid measurements. In the study of Pialat *et al.* (24), the mean percent difference and standard deviation of Tb.N, Tb.Th, Tb.Sp and Tb.1/N.SD for grade 3 were 12.8% (SD ±6.4%), –10.4% (SD ±5.0%), –11.1% (SD ±5.0%) and –8.7% (SD ±7.5%), respectively and for grade 5 25.8% (SD ±10.7%), –14.7% (SD ±6.4%), –20.4% (SD ±6.8%) and –12.9% (SD ±9.7%), respectively. Mean percent difference for Tb.BMD was 0.8% (SD ±1.5%) for grade 3 and 6.6% (SD ±3.6%) for grade 5. Interestingly, opposite our results Tb.Th decreased while Tb.BMD increased with poorer image quality. However, it must be noted that the study was conducted with a first-generation HR-pQCT which defers technically in several aspects. Firstly, first-generation HR-pQCT has a longer scanning time in the standard patient evaluation protocol. Secondly, its X-ray beam characteristics are different, potentially leading to differences in beam hardening. Thirdly, it uses a different filtering approach; while the first-gernation HR-pQCT applies an edge-enhancing filter, the second generation HR-pQCT applies a smoothing filter. Forthly, for direct measurements, first-generation HR-pQCT uses different thresholds for segmentation of bone from soft tissue (33).

Any motion artifact (scans classified with grade 2 or higher) will affect the density measurements and microarchitecture of the scaphoid. A nearly error-free quantitative analysis in this context can therefore only be achieved with grade 1 scans. The decision on which grades to include for analysis should be based on the research question. The motion-induced discrepancies must be evaluated in the context of possible biologically relevant changes based on therapeutic or metabolic influences (24). Since the scaphoid has rarely been the focus of HR-pQCT studies, little is known about the biological variability of the

individual parameters under physiological conditions and pathological or iatrogenic influences. The separation into a good and poor quality group results in higher feasibility regarding quality assessment. Especially for grade 1 and grade 2, high feasibility and minor microarchitecture errors are to be expected. However, a more strict selection leads to a higher dropout rate and a higher necessary number of study patients to achieve a valid number of cases.

High image quality is essential for histomorphometry analysis and should be a central point in every HR-pQCT study. A rigorous, standardized, and reproducible study protocol and a well-coordinated team form the basis for reliable results. After repeated scanning, previous studies showed an overall higher image quality (24,29). Therefore, for poor-quality scans, rescans are recommended. Bevers *et al.* (7) improved the quality of scaphoid scans when using a cast with a thumb part. Also, in this study, scans of wrists immobilized with a cast achieved slightly better results than wrists without cast immobilization. Automated methods for assessing motion artifacts (29,34) and contouring (7) can help create objective reproducible results but face practical problems in small objects like the scaphoid. Overcoming this limitation should be the focus of future studies.

The strength of this study includes the high number of analyzed stacks and the number of repeated scans per patient, which allowed up to five pairwise comparisons per stack.

Limitations of the study include the time intervals between the scans. The last imaging was performed one year after the first acquisition, which could lead to a selectional bias. Bone microarchitecture changes over time, e.g., due to illnesses and medication or changes in repetitive mechanical stress on the wrist joint (11-13,31,35). Kawalilak *et al.* (31) showed significant changes in the microarchitecture of the distal radius in postmenopausal women within one year. However, the present study population consisted of a healthy and on average younger study population. Moreover, a comparison of the density and the microarchitecture parameters between the first and the last follow-up revealed no significant differences.

Furthermore, since the positioning of the scaphoids within the restraining holders was not reproducible, one patient's stacks were not identical among the repeated scans. Therefore, of the six controls, only those areas were compared that the proximal, middle, and distal stacks, respectively, had in common. However, rotation errors or stack shifts could not be compensated thereby.

In clinical practice, it may be more relevant to evaluate

the entire scaphoid rather than the three stacks separately. Moreover, analyzing the whole bone may alleviate the problem of rotation errors. However, this study intended to evaluate the inter- and intraobserver reliability of the VGS in anatomically different areas of the scaphoid, with the poles in the proximal and the distal stack and the scaphoid waist in the middle stack. Since image quality can vary considerably within a scan due to the long scanning time, the authors considered it reasonable to evaluate the stacks individually to obtain more accurate results.

Conclusions

This study shows a significant influence of motion on the densitometric and microarchitectural parameters of the scaphoid. High image quality must be a central point in studies focusing on histomorphometry of small objects like the scaphoid.

The high inter- and intraobserver variability limit the VGS. Our results suggest strict selection criteria regarding the quality assessment of HR-pQCT scans to avoid misinterpretation of quantitative analysis in this emerging field of science. Future research may focus on other classification systems or automated techniques leading to more consistent and reproducible results. Under the present circumstances, the use of density and microarchitectural data of scaphoids should be limited to cases without or with minimal motion artifacts.

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims. amegroups.com/article/view/10.21037/qims-22-345/coif). Dr. SB reports that this work was supported by Johnson & Johnson (to Dr. Stefan Benedikt) (No. GMAFS20353). The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by institutional ethics board of the Medical University of Innbruck, Austria (No. 1259/2017) and informed consent was taken from all the patients.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Tada K, Ikeda K, Okamoto S, Hachinota A, Yamamoto D, Tsuchiya H. Scaphoid Fracture--Overview and Conservative Treatment. Hand Surg 2015;20:204-9.

2. Jørgsholm P, Ossowski D, Thomsen N, Björkman A. Epidemiology of scaphoid fractures and non-unions: A systematic review. Handchir Mikrochir Plast Chir 2020;52:374-81.

3. Reigstad O, Grimsgaard C, Thorkildsen R, Reigstad A, Røkkum M. Scaphoid non-unions, where do they come from? The epidemiology and initial presentation of 270 scaphoid non-unions. Hand Surg 2012;17:331-5.

4. Yin ZG, Zhang JB, Kan SL, Wang XG. Diagnostic accuracy of imaging modalities for suspected scaphoid fractures: meta-analysis combined with latent class analysis. J Bone Joint Surg Br 2012;94:1077-85.

5. Ty JM, Lozano-Calderon S, Ring D. Computed tomography for triage of suspected scaphoid fractures. Hand (N Y) 2008;3:155-8.

6. Daniels AM, Bevers MSAM, Sassen S, Wyers CE, van Rietbergen B, Geusens PPMM, Kaarsemaker S, Hannemann PFW, Poeze M, van den Bergh JP, Janzing HMJ. Improved Detection of Scaphoid Fractures with High-Resolution Peripheral Quantitative CT Compared with Conventional CT. J Bone Joint Surg Am 2020;102:2138-45.

7. Bevers MSAM, Daniels AM, Wyers CE, van Rietbergen

B, Geusens PPMM, Kaarsemaker S, Janzing HMJ, Hannemann PFW, Poeze M, van den Bergh JPW. The Feasibility of High-Resolution Peripheral Quantitative Computed Tomography (HR-pQCT) in Patients with Suspected Scaphoid Fractures. J Clin Densitom 2020;23:432-42.

8. Daniels AM, Wyers CE, Janzing HMJ, Sassen S, Loeffen D, Kaarsemaker S, van Rietbergen B, Hannemann PFW, Poeze M, van den Bergh JP. The interobserver reliability of the diagnosis and classification of scaphoid fractures using high-resolution peripheral quantitative CT. Bone Joint J 2020;102-B:478-84.

9. Scanco Medical. XtremeCT II User's Guide; 2018.

10. van den Bergh JP, Szulc P, Cheung AM, Bouxsein M, Engelke K, Chapurlat R. The clinical application of high-resolution peripheral computed tomography (HR-pQCT) in adults: state of the art and future directions. Osteoporos Int 2021;32:1465-85.

11. Wu D, Griffith JF, Lam SHM, Wong P, Yue J, Shi L, Li EK, Cheng IT, Li TK, Hung VW, Qin L, Tam LS. Comparison of bone structure and microstructure in the metacarpal heads between patients with psoriatic arthritis and healthy controls: an HR-pQCT study. Osteoporos Int 2020;31:941-50.

12. Yang H, Yu A, Burghardt AJ, Virayavanich W, Link TM, Imboden JB, Li X. Quantitative characterization of metacarpal and radial bone in rheumatoid arthritis using high resolution- peripheral quantitative computed tomography. Int J Rheum Dis 2017;20:353-62.

13. Chiba K, Yamada S, Yoda I, Era M, Yokota K, Okazaki N, et al. Effects of monthly intravenous ibandronate on bone mineral density and microstructure in patients with primary osteoporosis after teriparatide treatment: The MONUMENT study. Bone 2021;144:115770.

14. de Jong JJ, Willems PC, Arts JJ, Bours SG, Brink PR, van Geel TA, Poeze M, Geusens PP, van Rietbergen B, van den Bergh JP. Assessment of the healing process in distal radius fractures by high resolution peripheral quantitative computed tomography. Bone 2014;64:65-74.

15. Nishino Y, Chiba K, Era M, Okazaki N, Miyamoto T, Yonekura A, Tomita M, Osaki M. Analysis of fracture healing process by HR-pQCT in patients with distal radius fracture. J Bone Miner Metab 2020;38:710-7.

16. Bevers MSAM, Daniels AM, van Rietbergen B, Geusens PPMM, van Kuijk SMJ, Sassen S, Kaarsemaker S, Hannemann PFW, Poeze M, Janzing HMJ, van den Bergh JP, Wyers CE. Assessment of the healing of conservatively-treated scaphoid fractures using HR-pQCT. Bone

2021;153:116161.

17. Daniels AM, Janzing HMJ, Wyers CE, van Rietbergen B, Vranken L, Van der Velde RY, Geusens PPMM, Kaarsemaker S, Poeze M, Van den Bergh JP. Association of secondary displacement of distal radius fractures with cortical bone quality at the distal radius. Arch Orthop Trauma Surg 2021;141:1909-18.

18. Hosseini HS, Dünki A, Fabech J, Stauber M, Vilayphiou N, Pahr D, Pretterklieber M, Wandel J, Rietbergen BV, Zysset PK. Fast estimation of Colles' fracture load of the distal section of the radius by homogenized finite element analysis based on HR-pQCT. Bone 2017;97:65-75.

19. Deutschmann J, Patsch J, Valentinitsch A, Pietschmann P, Varga P, Dall'Ara, E, Zysset P, Weber G, Resch H, Kainberger F. Research Network Osteology Vienna: Hochauflösende und Mikro-Computertomographie in der Wiener Osteologie. Journal für Mineralstoffwechsel 2010;17:104-9.

20. Krug R, Burghardt AJ, Majumdar S, Link TM. High-resolution imaging techniques for the assessment of osteoporosis. Radiol Clin North Am 2010;48:601-21.

21. Link TM. Osteoporosis imaging: state of the art and advanced imaging. Radiology 2012;263:3-17.

22. Bevers MSAM, Wyers CE, Daniels AM, Audenaert EA, van Kuijk SMJ, van Rietbergen B, Geusens PPMM, Kaarsemaker S, Janzing HMJ, Hannemann PFW, Poeze M, van den Bergh JP. Association between bone shape and the presence of a fracture in patients with a clinically suspected scaphoid fracture. J Biomech 2021;128:110726.

23. Schmidle G, Ebner HL, Klauser AS, Fritz J, Arora R, Gabl M. Correlation of CT imaging and histology to guide bone graft selection in scaphoid non-union surgery. Arch Orthop Trauma Surg 2018;138:1395-405.

24. 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;50:111-8.

25. Sode M, Burghardt AJ, Pialat JB, Link TM, Majumdar S. Quantitative characterization of subject motion in HR-pQCT images of the distal radius and tibia. Bone 2011;48:1291-7.

26. Lee SB, Kim HJ, Chun JM, Lee CS, Kim SY, Kim PT, Jeon IH. Osseous microarchitecture of the scaphoid: Cadaveric study of regional variations and clinical implications. Clin Anat 2012;25:203-11.

27. 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;31:1607-27.

28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

29. Pauchard Y, Liphardt AM, Macdonald HM, Hanley DA, Boyd SK. Quality control for bone quality parameters affected by subject motion in high-resolution peripheral quantitative computed tomography. Bone 2012;50:1304-10.

30. 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;23:2151-8.

31. Kawalilak CE, Johnston JD, Olszynski WP, Kontulainen SA. Characterizing microarchitectural changes at the distal radius and tibia in postmenopausal women using HR-pQCT. Osteoporos Int 2014;25:2057-66.

32. Mata-Mbemba D, Rohringer T, Ibrahim A, Adams-Webberc T, Moineddin R, Doria AS, Vali R. HR-pQCT imaging in children, adolescents and young adults: Systematic review and subgroup meta-analysis of normative data. PLoS One 2019;14:e0225663.

33. 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;32:1514-24.

34. Pauchard Y, Ayres FJ, Boyd SK. Automated quantification of three-dimensional subject motion to monitor image quality in high-resolution peripheral quantitative computed tomography. Phys Med Biol 2011;56:6523-43.

35. Schipilow JD, Macdonald HM, Liphardt AM, Kan M, Boyd SK. Bone micro-architecture, estimated bone strength, and the muscle-bone interaction in elite athletes: an HR-pQCT study. Bone 2013;56:281-9.

Table S1 Supplemental digital content 1: Patient's demographics

ID	Age	Gender	Fractured side	Dominant side	Cause of accident	Profession	Smoker	Fracture location	Time of cast imobilization (days)	1st scan 2 weeks after trauma	2nd scan 4 weeks after trauma	3rd scan 6 weeks after trauma	4th scan 12 weeks after trauma	5th scan 6 months after trauma	6th scan 1 year after trauma
1	28	m	left	right	Fall	Blue-collar worker	yes	waist	67	✓*	✓*	✓*	✓	✓	✓
2	48	m	left	right	Traffic accident	White-collar worker	yes	waist	80	✓*	✓*	✓*	✓	✓	✓
3	20	w	left	right	Traffic accident	White-collar worker	no	waist	78	✓*	✓*	✓*	✓	✓	✓
4	63	w	left	both	Sport accident	Retired	no	distal pole	54	✓*	✓*	✓*	✓	✓	✓
5	28	w	left	right	Sport accident	White-collar worker	yes	distal pole	51	✓*	✓*	✓*	✓	✓	✓
6	19	m	left	right	Sport accident	Student	yes	waist	69	✓*	✓*	✓*	✓	✓	✓
7	72	w	left	right	Fall	Retired	no	waist	70	✓*	✓*	✓*	✓	✓	✓
8	56	m	left	right	Fall	Blue-collar worker	no	waist	85	✓*	✓*	✓*	✓	✓	✓
9	47	m	left	right	Fall	White-collar worker	no	distal pole	54	✓*	✓*	✓*	✓	✓	✓
10	19	m	right	right	Sport accident	Student	no	distal pole	41	✓*	✓*	✓	✓	✓	x
11	24	m	left	right	Sport accident	Blue-collar worker	yes	waist	93	✓*	✓*	✓*	✓	✓	✓
12	23	m	right	right	Sport accident	Student	no	waist	67	✓*	✓*	✓*	✓	✓	✓
13	30	m	left	right	Fall	Student	no	distal pole	30	✓*	✓*	✓	✓	✓	✓
14	72	w	left	right	Fall	Retired	no	waist	80	✓*	✓*	✓*	✓	✓	✓
15	27	m	right	right	Fall	White-collar worker	n.s.	waist	85	✓*	✓*	✓*	✓	✓	✓
16	32	m	right	right	Fall	Blue-collar worker	yes	waist	79	✓*	✓*	✓*	✓	x	✓
17	24	m	left	right	Sport accident	White-collar worker	yes	waist	69	✓*	✓*	✓*	✓	✓	x
18	27	w	right	right	Sport accident	White-collar worker	no	waist	84	✓*	✓*	✓*	✓	✓	✓
19	28	m	right	right	Sport accident	Blue-collar worker	yes	waist	80	✓*	✓*	✓*	✓	✓	✓
20	58	w	right	right	Fall	Housewife	yes	distal pole	42	✓*	✓*	✓	✓	✓	✓
21	56	w	left	both	Sport accident	White-collar worker	no	distal pole	40	✓*	✓*	✓	x	✓	✓
22	21	m	left	right	Fall	Blue-collar worker	no	distal pole	43	✓*	✓*	✓	✓	✓	✓

*, fractured side scanned with cast; ✓, scanned; X, not scanned (patient did not show up for the visit). m, male; f, female; n.s, not stated.

Table S2 Supplemental digital content 2: Mean relative differences (standard deviation) of the densitometric and micro–architectural parameters at each grade in percent compared to the internal grade 1 reference.

	1–1 (n=58)	1–2 (n=85)	1–3 (n=31)	1–4 (n=15)	1–5 (n=6)
Tb.BV/TV	−0.3 (1.8)	1.3 (4.2)	3.6 (3.5)	2.9 (8.4)	3.5 (4.5)
Tb.N	−0.3 (4.1)	5.1 (4.7)	11.9 (5.8)	12.6 (5.9)	15.5 (5.9)
Tb.Th	−0.4 (1.7)	1.5 (3.1)	4.7 (2.8)	4.5 (3.5)	6.8 (1.8)
Tb.Sp	0.6 (3.6)	−3.7 (4.1)	−9.4 (5.2)	−10.7 (5.8)	−13.7 (6.2)
Tb.1/N.SD	0.4 (5.0)	−4.8 (6.4)	−13.9 (8.4)	−16.3 (9.0)	−21.4 (10.4)
Conn.Dens.	−0.4 (5.3)	6.7 (10.9)	13.3 (17.5)	13.9 (8.7)	21.6 (29.9)
Apparent denisty	−0.5 (2.0)	0.3 (3.7)	2.0 (3.4)	1.1 (7.9)	1.7 (3.0)
Tb.BMD	−0.2 (1.5)	−1.7 (1.7)	−3.2 (1.9)	−4.3 (2.4)	−4.9 (2.2)

Tb.BV/TV, trabecular bone volume fraction, Tb.N, trabecular number, Tb.Th, trabecular thickness, Tb.Sp, trabecular separation, Tb.1/N.SD, inhomogeneity of trabecular network, Conn.D, connective density, Tb.BMD, trabecular bone mineral density.

Table S3 Supplemental digital content 3: Mean (standard deviation) of each grade's structure modeling index (SMI)

	1–1 (n=58)	1–2 (n=85)	1–3 (n=31)	1–4 (n=15)	1–5 (n=6)
SMI	−1.5 (1.7)	−1.0 (1.6)	−0.5 (1.7)	−0.4 (1.9)	−0.7 (1.9)