














# Reliability and Change in Erosion Measurements by High-resolution Peripheral Quantitative Computed Tomography in a Longitudinal Dataset of Rheumatoid Arthritis Patients

Stephanie Finzel<sup>1</sup> , Sarah L. Manske<sup>2</sup> , Cheryl C.M. Barnabe<sup>3</sup> , Andrew J. Burghardt<sup>4</sup> ,  
Hubert Marotte<sup>5</sup> , Andrea Scharmga<sup>6</sup> , Ellen-Margrethe Hauge<sup>7</sup> , Roland Chapurlat<sup>8</sup> ,  
Klaus Engelke<sup>9</sup> , Xiaojuan Li<sup>10</sup> , Bente C.J. van Teeffelen<sup>11</sup> , Philip G. Conaghan<sup>12</sup> ,  
and Kathryn S. Stok<sup>13</sup> 

**ABSTRACT. Objective.** The aim of this multireader exercise was to assess the reliability and change over time of erosion measurements in patients with rheumatoid arthritis (RA) using high-resolution peripheral quantitative computed tomography (HR-pQCT).

**Methods.** HR-pQCT scans of 23 patients with RA were assessed at baseline and 12 months. Four experienced readers examined the dorsal, palmar, radial, and ulnar surfaces of the metacarpal head (MH) and phalangeal base (PB) of the second and third digits, blinded to time order. In total, 368 surfaces (23 patients' 16 surfaces) were evaluated per timepoint to characterize cortical breaks as pathological (erosion) or physiological, and to quantify erosion width and depth. Reliability was evaluated by intraclass correlation coefficients (ICC), percentage agreement, and Light k; change over time was defined by means  $\pm$  SD of erosion numbers and dimensions.

**Results.** ICC for the mean measurements of width and depth of the pathological breaks ranged between 0.819–0.883, and 0.771–0.907, respectively. Most physiological cortical breaks were found at the palmar PB, whereas most pathological cortical breaks were located at the radial MH. There was a significant increase in both the numbers and the dimensions of erosions between baseline and follow-up ( $P = 0.0001$  for erosion numbers, width, and depth in axial plane;  $P = 0.001$  for depth in perpendicular plane).

**Conclusion.** This exercise confirmed good reliability of HR-pQCT erosion measurements and their ability to detect change over time.

**Key Indexing Terms:** computed tomography, metacarpophalangeal joint, outcomes, rheumatoid arthritis

RELEX-2 was hosted by the University of San Francisco, California, USA. The meeting was sponsored in part by Scanco Medical AG. PGC is supported in part by the UK National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

<sup>1</sup>S. Finzel, MD, Senior Attending Physician, Department of Rheumatology and Clinical Immunology, Medical Center – University of Freiburg, Medical Faculty, University of Freiburg, Freiburg, Germany; <sup>2</sup>S.L. Manske, PhD, Assistant Professor, Department of Radiology, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; <sup>3</sup>C.C. Barnabe, MD, MSc, Associate Professor, Departments of Medicine and Community Health Sciences, University of Calgary, and McCaig Institute for Bone and Joint Health, University of Calgary, Calgary, Alberta, Canada; <sup>4</sup>A.J. Burghardt, BS, Research Specialist, Department of Radiology and Biomedical Imaging, University of California San Francisco, California, USA; <sup>5</sup>H. Marotte, MD, PhD, Professor, INSERM 1059, Université de Lyon, and Service de Rhumatologie, CHU de Saint-Etienne, Saint-Etienne, France; <sup>6</sup>A. Scharmga, PhD, Maastricht University, Maastricht, the Netherlands; <sup>7</sup>E.M. Hauge, MD, PhD, Professor, Department of Rheumatology, Aarhus University

Hospital and Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; <sup>8</sup>R. Chapurlat, MD, PhD, Professor, INSERM 1033, Hôpital Edouard Herriot, Lyon, France; <sup>9</sup>K. Engelke, PhD, Professor, Department of Medicine 3, FAU University Erlangen-Nürnberg and Universitätsklinikum Erlangen, Erlangen, Germany; <sup>10</sup>X. Li, PhD, Professor, Department of Biomedical Engineering, Cleveland Clinic, Cleveland, Ohio, USA; <sup>11</sup>B.C. van Teeffelen, Department of Biomedical Engineering, Melbourne School of Engineering, The University of Melbourne, Melbourne, Australia; <sup>12</sup>P.G. Conaghan, MD, PhD, Professor, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds and NIHR Leeds Biomedical Research Centre, Leeds, UK; <sup>13</sup>K.S. Stok, PhD, Senior Lecturer, Institute for Biomechanics, ETH Zurich, Zurich, Switzerland, and Department of Biomedical Engineering, University of Melbourne, Melbourne, Australia. The authors state that they have no conflicts of interest.

Address correspondence to Dr. S. Finzel, Head of Clinical Trials Unit Rheumatology, Department of Rheumatology and Clinical Immunology, University Medical Center, Medical Faculty, University of Freiburg, Hugstetter Strasse 55, 79110 Freiburg, Germany.  
Email: Stephanie.Finzel@uniklinik-freiburg.de.

Accepted for publication September 5, 2020.

High-resolution peripheral quantitative computed tomography (HR-pQCT) provides accurate detection of periarticular bone changes, which is required for diagnosis and therapeutic monitoring in rheumatoid arthritis (RA)<sup>1</sup>. Previously, the Study group for xtrEme Computed Tomography in Rheumatoid Arthritis (SPECTRA) collaboration presented a consensus definition for bone erosion, and a common approach for measuring erosion size, with feasibility and preliminary reliability tested in a cross-sectional dataset of RA metacarpophalangeal (MCP) joints (RELEX-1)<sup>2</sup>. Good agreement was demonstrated regarding the presence and nature of cortical breaks; however, agreement for measuring erosion dimensions needed refinement. We therefore performed this multireader HR-pQCT exercise in order to assess the reliability of erosion measurements and to evaluate change over time in patients with RA.

## MATERIALS AND METHODS

**Images.** Twenty-three seropositive RA patients underwent HR-pQCT imaging of the second and third digits of their dominant hand at baseline (0 months) and follow-up (12 months) at the University of Erlangen-Nuremberg, Université de Lyon, University of California San Francisco, and University of Calgary. Patients were selected according to the presence of bone erosions on radiographs as assessed by the Sharp/van der Heijde (SvdH) score, and the need to change therapy due to insufficient disease control. Local ethics approval and written informed consent were obtained prior to study entry (institutional review board numbers: Calgary REB15-0582; San Francisco 12-10418; Lyon CPP 13/083; Erlangen 3839). All participants were scanned using a first-generation HR-pQCT scanner (Scanco Medical AG) and standard acquisition settings were applied with an 82- $\mu$ m isotropic voxel size resolution<sup>3</sup>. Image datasets were viewed using Osirix (version 5.8). Readers were blinded to clinical status and time sequence of images.

Prior to image evaluation, the 4 readers calibrated measurements using a test set of images. All the readers participating had at least 3 years' experience in reading HR-pQCT datasets.

**Joint image evaluation.** The image evaluation algorithm involves assessing 8 surfaces at each of the second and third MCP joints, specifically the palmar, dorsal, radial, and ulnar surfaces of each of the proximal phalangeal base (PB) and the metacarpal head (MH)<sup>1</sup>. Only images of sufficient quality were evaluated<sup>4</sup>. Individual surfaces were analyzed for the presence of cortical breaks (present or absent) according to the SPECTRA definition: The cortical break should be present in 2 consecutive slices and 2 perpendicular planes, and should show a loss of underlying trabecular bone. The cortical break is characterized as being pathological (erosion) or physiological, with the former described as a nonlinear appearance typical of erosions, and the latter as a parallel/linear break typical of vessel channels<sup>3</sup>. Supplementary Figure 1 (available with the online version of this article) gives an example of typical pathological or physiological cortical breaks. For erosions, readers quantified the size of the break by measuring the maximum width and corresponding depth in both axial and perpendicular planes to each surface. The depth of the cortical break was recorded on the same slice where the maximal width was obtained. All measures were quantified in mm. Readers noted whether multiple cortical breaks were present on the same surface but only recorded measurements for the largest cortical break.

**Statistical analysis.** The interreader reliability of the detection of cortical breaks was evaluated using the percentage of agreement and Light k for the chance-corrected agreement<sup>5,6</sup>. The intraclass correlation (ICC) was calculated as an indicator of variability in cortical break depth and width measurements between readers 1, 2, 3, and 4. Paired-sample *t*-test was used to evaluate the longitudinal changes between baseline and follow-up scans of each subject. Analyses were performed with SPSS (version 23; IBM Corp.).

## RESULTS

**Patient characteristics.** Mean age (SD) was 46 (13) years, 60% women, mean disease duration 2.3 (2.8) years, and mean Disease Activity Score in 28 joints at baseline was  $3.51 \pm 1.03$ . There was no significant change detectable over time in SvdH score. All patients received methotrexate; 18 patients were also treated with or started receiving a tumor necrosis factor- $\alpha$  inhibitor.

Images from 2 timepoints (baseline and follow-up) for 23 subjects were evaluated, resulting in 46 individual joints with 368 unique surfaces that were evaluated per timepoint. Thus, a total of 736 surfaces were evaluated.

**Evaluability of images.** The percentage agreement for evaluability of all the images between all readers was 80% (589/736). The chance-corrected agreement was fair (Cohen  $\kappa$  0.218; ranges for all individual reader pairs 0.005–0.519). Only the surfaces in which all 4 readers agreed that the image was evaluable were included beyond this step ( $n = 585$ ). Evaluability was affected by the presence of motion artifacts and/or technical artifacts such as stack artifacts.

**Presence of cortical breaks.** The percentage agreement for the presence or absence of cortical breaks on all evaluable images between all readers was 57% (334/585). The chance-corrected agreement resulted in a moderate *k* value of 0.493. Cohen  $\kappa$  for all individual reader pairs (reader 1 vs reader 2, etc.) ranged between 0.405–0.551.

**Characterization of cortical breaks.** In total, 99 cortical breaks were identified on the baseline and follow-up images. The percentage agreement for the appearance of a cortical break as pathological or physiological between all readers was 81% (80/99). The chance-corrected agreement resulted in a substantial *k* value of 0.796. Cohen  $\kappa$  for all individual reader pairs ranged between 0.765–0.838.

**Numbers and localizations of erosions and physiological cortical breaks.** Table 1 shows the number of breaks (total and erosions) for the 8 individual surfaces in which all readers agreed on the presence of a cortical break. The distribution of cortical breaks confirmed findings from previous publications<sup>3,7,8</sup>.

**Widths and depths of erosions.** There were 41 cortical breaks detected as erosions by all readers. Table 1 shows the mean dimensions with SD measured by all readers of these 41 erosions for the respective surfaces the erosions were detected in.

**Interreader agreement regarding measurements of the sizes of cortical breaks.** Numbers and dimensions of cortical breaks were determined on surfaces where all readers agreed that an erosion was present ( $n = 41$ ). ICC was calculated; for all 4 measures the ICC was high: mean values  $\pm$  SD and ICC for erosion numbers, axial width and depth, as well as perpendicular width and depth were  $1.39 \pm 0.62$  (ICC 0.803);  $2.31 \pm 1.39$  (ICC 0.883);  $1.85 \pm 0.86$  (ICC 0.907);  $1.99 \pm 0.87$  (ICC 0.819), and  $1.89 \pm 0.91$  (ICC 0.771), respectively (see Table 2 for details; for further measures of precision, see Supplementary Table 1, available with the online version of this article).

**Longitudinal change of cortical breaks over time.** All pairs of measures (baseline and follow-up) were evaluated and the mean

Table 1. Total and pathological cortical breaks per surface, as well as localization and size of pathological cortical breaks per surface, as measured by 4 readers.

Surface	Palmar PB	Ulnar PB	Dorsal PB	Radial PB	Palmar MH	Ulnar MH	Dorsal MH	Radial MH	Total
Total no. breaks (no. pathological breaks)	31 (0)	3 (3)	1 (0)	3 (2)	16 (1)	8 (6)	8 (4)	29 (25)	99 (41)
Width axial, mm	NA	1.76 ± 0.38	NA	1.89 ± 0.30	4.05 ± 0.59	1.98 ± 0.38	2.16 ± 0.81	2.38 ± 1.46	2.37 ± 0.65
Depth axial, mm	NA	1.27 ± 0.45	NA	1.64 ± 0.80	2.31 ± 0.67	1.90 ± 0.45	1.36 ± 0.64	1.97 ± 0.94	1.74 ± 0.66
Width perpendicular, mm	NA	1.53 ± 0.62	NA	1.28 ± 0.33	2.04 ± 0.66	2.72 ± 1.52	1.93 ± 0.65	2.43 ± 1.38	1.99 ± 0.86
Depth perpendicular, mm	NA	1.40 ± 0.51	NA	2.07 ± 1.78	2.31 ± 1.89	1.96 ± 0.59	1.58 ± 0.69	1.94 ± 0.89	1.88 ± 1.06

Values are mean ± SD unless otherwise specified. Numbers and measures are given for full consensus on presence and appearance of breaks only. Values refer to baseline measures. MH: metacarpal head; NA: not applicable; PB: phalangeal base.

Table 2. Erosions and variability in measurements between all readers (n = 41 per reader, full consensus only).

	Reader 1	Reader 2	Reader 3	Reader 4	ICC
No. erosions	1.35 ± 0.58	1.53 ± 0.75	1.40 ± 0.63	1.28 ± 0.51	0.803
Width axial, mm	2.06 ± 0.99	2.40 ± 1.33	2.38 ± 1.40	2.38 ± 1.48	0.883
Depth axial, mm	1.80 ± 0.80	1.84 ± 0.83	1.67 ± 0.92	2.07 ± 0.88	0.907
Width perpendicular, mm	1.87 ± 0.82	1.79 ± 0.84	1.87 ± 0.87	2.44 ± 0.94	0.819
Depth perpendicular, mm	1.68 ± 0.74	2.03 ± 0.87	1.66 ± 0.94	2.18 ± 1.08	0.771

Values are mean values (± SD). ICC: intraclass correlation coefficient.

values were compared to test for significant differences over time. In total, all pairs from all readers gave 285 pairs. Mean values (± SD) of erosion numbers, widths, and depths are shown in Table 3. There was a significant increase in both the numbers of erosions and the dimensions of the cortical breaks between baseline and follow-up scans (all  $P < 0.01$ ).

## DISCUSSION

In this multireader responsiveness exercise, we applied HR-pQCT imaging to assess reliability and change over time of erosion measurements in a dataset of patients with RA. We applied our consensus definition of bone erosion as well as a previously agreed evaluation algorithm<sup>3</sup>. The exercise yielded good reliability for HR-pQCT measurements (ICC > 0.771) and a significant increase was observed in both number and dimensions of erosions between baseline and follow-up ( $P < 0.01$ ). Further, most physiological cortical breaks were found at the palmar PB, whereas most erosions were located at the radial MH; the distribution of erosions and physiological cortical breaks confirmed the findings from earlier studies<sup>3,7,8</sup>.

Agreement (ICC) for erosion numbers, width, and depth

of cortical breaks were high, and k for appearance of cortical breaks were good. The reliability measures in this study revealed better results than in the RELEX-1 exercise<sup>3</sup>. For the current exercise, we used only 4, not 11, readers as in the first exercise with prestudy calibration<sup>3</sup>. It should be noted that we used 4 readers, unlike the 2 readers typically used in a clinical trial, and that images were read in unknown time order, which may also reduce responsiveness. A limitation of the study might be that only those surfaces were analyzed further, in which all readers agreed that the image quality was sufficient, and a cortical break was present, which reduced the number of analyzable surfaces. This emphasizes the need for adequate training before reading HR-pQCT images. On the other hand, the lack of training could be overcome by developing semiautomated algorithms allowing for volumetric assessment of pathological cortical breaks.

The analysis of change over time yielded highly significant values for mean ± SD of number, width, and depth of cortical breaks. Our findings showed responsiveness over time despite having small sample sizes and achieving disease control.

Ongoing work has evaluated the nature of small cortical breaks. Boutroy, *et al*<sup>9</sup> performed a perfusion study on a cadaveric

Table 3. Erosions of all baseline and follow-up measurements of pathological cortical breaks of all baseline and follow-up measurements in all 736 surfaces examined.

	Baseline	12 Months	P
No. erosions	0.73 ± 0.67	1.02 ± 0.90	0.0001
Width axial, mm	1.31 ± 1.56	1.79 ± 1.55	0.0001
Depth axial, mm	0.87 ± 1.06	1.25 ± 1.19	0.0001
Width perpendicular, mm	1.07 ± 1.14	1.69 ± 1.45	0.0001
Depth perpendicular, mm	0.88 ± 1.07	1.20 ± 1.15	0.001

Values are mean ± SD. All 285 pairs of baseline and follow-up measures from all readers were considered.



hand using contrast perfusion, confirming the location of vascular foramen and their comparative frequency in periarticular bone. This provides construct validity for the SPECTRA erosion definition. Scharmga, *et al* compared vascular foramen in histology and in HR-pQCT<sup>10</sup>. Perhaps unsurprisingly, due to differences in spatial resolution, there was a substantially higher number of vessel channels found in histology than in HR-pQCT. It needs to be assessed further, however, whether uniquely identified HR-pQCT small cortical breaks are of added value in RA monitoring.

While the assessment of radiographic joint space width in HR-pQCT may be semiautomated<sup>11,12,13</sup>, the evaluation algorithm of cortical breaks still requires training and time<sup>14,15</sup>. Therefore, our collaboration is pursuing the investigation of a common technical algorithm for semi- or fully automated erosion detection and measurement allowing for volumetric erosion assessment<sup>16</sup>.

In conclusion, HR-pQCT evaluation using trained readers allows for highly reliable and precise detection of cortical breaks and facilitates differentiation of pathological from physiological cortical breaks. Reading by less experienced readers results in fair *k* values with regard to evaluability and break detection. Moreover, our results suggest that HR-pQCT shows good responsiveness of erosion measures over time.

## ACKNOWLEDGMENT

This paper was written on behalf of all SPECTRA collaboration members.

## ONLINE SUPPLEMENT

Supplementary material accompanies the online version of this article.

## REFERENCES

1. Stach CM, Bauerle M, Englbrecht M, Kronke G, Engelke K, Manger B, et al. Periarticular bone structure in rheumatoid arthritis patients and healthy individuals assessed by high-resolution computed tomography. *Arthritis Rheum* 2010;62:330-9.
2. Barnabe C, Toepfer D, Marotte H, Hauge EM, Scharmga A, Kocijan R, et al; SPECTRA Collaboration. Definition for rheumatoid arthritis erosions imaged with high resolution peripheral quantitative computed tomography and interreader reliability for detection and measurement. *J Rheumatol* 2016;43:1935-40.
3. Barnabe C, Feehan L; SPECTRA (Study Group for XTrEme-CT in RA). High-resolution peripheral quantitative computed tomography imaging protocol for metacarpophalangeal joints in inflammatory arthritis: the SPECTRA collaboration. *J Rheumatol* 2012; 39:1494-5.
4. 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.
5. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychol Bull* 1971;76:365-77.
6. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* 2005;85:257-68.
7. Finzel S, Ohrndorf S, Englbrecht M, Stach C, Messerschmidt J, Schett G, et al. A detailed comparative study of high-resolution ultrasound and micro-computed tomography for detection of arthritic bone erosions. *Arthritis Rheum* 2011;63:1231-6.

8. Finzel S, Rech J, Schmidt S, Engelke K, Englbrecht M, Schett G. Interleukin-6 receptor blockade induces limited repair of bone erosions in rheumatoid arthritis: a micro CT study. *Ann Rheum Dis* 2013;72:396-400.
9. Boutroy S, Chapurlat R, Vanden-Bossche A, Loeckle H, Thomas T, Marotte H. Erosion or vascular channel? *Arthritis Rheumatol* 2015;67:2956.
10. Scharmga A, Keller KK, Peters M, van Tubergen A, van den Bergh JP, et al. Vascular channels in metacarpophalangeal joints: a comparative histologic and high-resolution imaging study. *Sci Rep* 2017;7:8966.
11. Barnabe C, Buie H, Kan M, Szabo E, Barr SG, Martin L, et al. Reproducible metacarpal joint space width measurements using 3D analysis of images acquired with high-resolution peripheral quantitative computed tomography. *Med Eng Phys* 2013;35:1540-4.
12. Burghardt AJ, Lee CH, Kuo D, Majumdar S, Imboden JB, Link TM, et al. Quantitative in vivo HR-pQCT imaging of 3D wrist and metacarpophalangeal joint space width in rheumatoid arthritis. *Ann Biomed Eng* 2013;41:2553-64.
13. Stok KS, Finzel S, Burghardt AJ, Conaghan PG, Barnabe C; SPECTRA Collaboration. The SPECTRA Collaboration OMERACT Special Interest Group: current research and future directions. *J Rheumatol* 2017;44:1911-5.
14. Töpfer D, Finzel S, Museyko O, Schett G, Engelke K. Segmentation and quantification of bone erosions in high-resolution peripheral quantitative computed tomography datasets of the metacarpophalangeal joints of patients with rheumatoid arthritis. *Rheumatology* 2014;53:65-71.
15. Töpfer D, Gerner B, Finzel S, Kraus S, Museyko O, Schett G, et al. Automated three-dimensional registration of high-resolution peripheral quantitative computed tomography data to quantify size and shape changes of arthritic bone erosions. *Rheumatology* 2015;54:2171-80.
16. Peters M, Scharmga A, de Jong J, van Tubergen A, Geusens P, Arts JJ, et al. An automated algorithm for the detection of cortical interruptions on high resolution peripheral quantitative computed tomography images of finger joints. *PLoS One* 2017;12:e0175829.

**APPENDIX 1.** List of SPECTRA Collaboration members: Cheryl Barnabe, University of Calgary; Anne-Birgitte Blavnsfeldt, University of Aarhus; Stephanie Boutroy, Université de Lyon; Steven K. Boyd, University of Calgary; Andrew Burghardt, University of California San Francisco; Roland Chapurlat, Université de Lyon; Angela Cheung, University of Toronto; Ko Chiba, University of Nagasaki; Joost de Jong, Maastricht University Medical Centre; Klaus Engelke, University of Erlangen; Stephanie Finzel, University of Freiburg; Ursula Heilmeyer, University of California San Francisco; Harry Genant, University of California, San Francisco; Piet Geusens, Maastricht University; Ellen-Margrethe Hauge, Aarhus University Hospital; Joost de Jong, Maastricht University; Denis Julien, Université de Lyon; Rashid Kapadia, Scanco, USA; Kresten Keller, Aarhus University Hospital; Roland Kocijan, University of Vienna; Sebastian Kraus, Kantonsspital Baden; Eric Lespessailles, University of Orleans; Xiaojuan Li, University of Cleveland; Sarah Manske, University of Calgary; Hubert Marotte, Saint-Etienne, Université de Lyon; Liam Martin, University of Calgary; Michiel Peters, Maastricht University; Valentina Pedoia, University of California San Francisco; Andrea Scharmga, Maastricht University; Georg Schett, University of Erlangen; Kathryn S. Stok, The University of Melbourne; Nikolay Tzaribachev, Bad Bramstedt; Joop van den Bergh, Maastricht University; Bert van Rietbergen, Eindhoven University of Technology; Tomohiro Shimizu, University of California San Francisco; Lai-Shan Tam, University of Hong Kong; Karen Troy, University of Worcester; Mira van Veenendaal, University of Toronto; Nicolas Vilaythi, Scanco, Switzerland; Paul Willems, Maastricht University Medical Centre; Rae Yeung, University of Toronto.