Definition for Rheumatoid Arthritis Erosions Imaged with High Resolution Peripheral Quantitative Computed Tomography and Interreader Reliability for Detection and Measurement

Cheryl Barnabe, Dominique Toepfer, Hubert Marotte, Ellen-Margrethe Hauge, Andrea Scharmga, Roland Kocijan, Sebastian Kraus, Stephanie Boutroy, Georg Schett, Kresten Krarup Keller, Joost de Jong, Kathryn S. Stok, Stephanie Finzel, and the SPECTRA Collaboration

ABSTRACT. Objective. High-resolution peripheral quantitative computed tomography (HR-pQCT) sensitively detects erosions in rheumatoid arthritis (RA); however, nonpathological cortical bone disruptions are potentially misclassified as erosive. Our objectives were to set and test a definition for pathologic cortical bone disruptions in RA and to standardize reference landmarks for measuring erosion size.

Methods. HR-pQCT images of metacarpophalangeal joints of RA and control subjects were used in an iterative process to achieve consensus on the definition and reference landmarks. Independent readers (n = 11) applied the definition to score 58 joints and measure pathologic erosions in 2 perpendicular multiplanar reformations for their maximum width and depth. Interreader reliability for erosion detection and variability in measurements between readers [root mean square coefficient of variation (RMSCV), intraclass correlation (ICC)] were calculated.

Results. Pathologic erosions were defined as cortical breaks extending over a minimum of 2 consecutive slices in perpendicular planes, with underlying trabecular bone loss and a nonlinear shape. Interreader agreement for classifying pathologic erosions was 90.2%, whereas variability for width and depth erosion assessment was observed (RMSCV perpendicular width 12.3%, axial width 20.6%, perpendicular depth 24.0%, axial depth 22.2%; ICC perpendicular width 0.206, axial width 0.665, axial depth 0.871, perpendicular depth 0.783). Mean erosion width was 1.84 mm (range 0.16–8.90) and mean depth was 1.86 mm (range 0.30–8.00).

Conclusion. We propose a new definition for erosions visualized with HR-pQCT imaging. Interreader reliability for erosion detection is good, but further refinement of selection of landmarks for erosion size measurement, or automated volumetric methods, will be pursued. (J Rheumatol 2016; 43:1935–40; doi:10.3899/jrheum.160648)

Key Indexing Terms:
RHEUMATOID ARTHRITIS
EROSIONS
HR-PQCT

From Medicine and Community Health Sciences, University of Calgary, Calgary, Alberta, Canada; Institute of Medical Physics, University of Erlangen, Erlangen, Germany; Institut national de la santé et de la recherche médicale, Research Unit U1059, and INSERM UMR 1033, University of Lyon, Saint-Etienne; Rheumatology Department, University Hospital of Saint-Etienne, Saint-Etienne, France; Department of Rheumatology, Aarhus University Hospital, Aarhus, Denmark; Department of Rheumatology, Maastricht University Medical Centre, Maastricht, the Netherlands; St. Vincent Hospital — Medical Department II, The VINFORCE Study Group, Academic Teaching Hospital of Medical University of Vienna; Aarhus University Hospital; A. Scharmga, MSc, Department of Rheumatology, Maastricht University Medical Centre; R. Kocijan, MD, St. Vincent Hospital–Medical Department II, The VINFORCE Study Group, Academic Teaching Hospital of Medical University of Vienna; S. Kraus, MD, Department of Internal Medicine III, Rheumatology and Immunology, University of Erlangen-Nuremberg, and Department of Clinical Medicine, Aarhus University; K. K. Keller, MD, PhD, Department of Rheumatology, Aarhus University Hospital; J. de Jong, MSc, Department of Rheumatology, Maastricht University Medical Centre; K. S. Stok, PhD, ETH Zurich; S. Finzel, MD, Department of Internal Medicine III, Rheumatology and Immunology, University of Erlangen-Nuremberg, and Department for Rheumatology and Clinical Immunology, University Medical Center Freiburg; SPECTRA Collaboration members are listed in Appendix 1.

Address correspondence to Dr. C. Barnabe, 3330 Hospital Dr. NW, Calgary, Alberta T2N 4N1, Canada. E-mail: ccbarnab@ucalgary.ca
High-resolution peripheral quantitative computed tomography (HR-pQCT) is a technical modification of conventional CT designed to examine volumetric bone mineral density and bone microarchitecture. Reviews highlight the novelty of HR-pQCT in quantifying bone density and microarchitecture in both adults and children, health and disease states, and in response to treatment, while likely providing improved capability to predict fracture risk compared to dual-energy x-ray absorptiometry. Erosions are a defining feature of rheumatoid arthritis (RA). Members of SPECTRA (Study group for xtrEme Computed Tomography in Rheumatoid Arthritis, Appendix 1), an international collaboration, have published on the use of this technology to assess periaricular bone changes in a variety of arthritis conditions including RA and psoriatic arthritis. Moreover, HR-pQCT has the capacity to more sensitively detect erosions compared to plain radiography, ultrasound (US), and magnetic resonance imaging.

Early independent work in detecting erosions with HR-pQCT and quantifying their size resulted in multiple definitions and size measurement methods reported in the literature. To illustrate, the first publication by Stach, et al. defined erosion as a periaricular cortical bone shell break, and measured the maximum width of the cortical break in 2-dimensional images. This size was then applied to a semiquantitative scale to produce an erosion score, which distinguished RA and healthy subjects with a sensitivity of 0.73 and specificity of 0.83. Fouque-Aubert, et al. defined erosions as sharply margined bone lesions with periaricular localization and a cortical break in at least 2 adjacent slices. To calculate the erosion size, they manually defined a region of interest for each erosion, and used the mean area of the slice, the total number of slices, and the slice height in their calculation of an erosion volume. Further, Srikhum, et al. applied the same definition as Fouque-Aubert, but confirmed questionable cortical breaks with 3-D image reconstruction, and then, in keeping with the Stach methodology, measured the maximal dimensions of the erosion; subsequently, they applied different size criteria in proposing a semiquantitative score. Barnabe, et al. described erosions as a definite cortical break on a 2-dimensional image, extending over a minimum of 3 slices, and confirmed in all 3 perpendicular multiplanar reformations (MPR), but did not pursue quantification of the erosions. Albrecht, et al. used a semiellipsoid formula for quantification of erosions after applying the erosion definition of Stach. Lastly, Töpfer, et al. have pursued measuring volume, surface area, and shape variables of erosions in 3-D.

SPECTRA thus prioritized achieving consensus on a definition for erosions, and standardizing reference points for erosion size measurements in 2-dimensional images until a 3-D protocol can be validated. A first reliability exercise (RELEX-1) was performed at the University of Erlangen-Nuremberg in April 2013, to facilitate these objectives.

**MATERIALS AND METHODS**

*Images.* Subjects with RA, control subjects, and cadaveric specimens not known to have RA were imaged under research ethics approval for respective project protocols at each individual partner institution [including the University of Calgary Conjoint Health Research Ethics Board in Canada, Ethik-Kommission der Medizinenischen Fakultaet der FAU Erlangen-Nuremberg in Germany, Technical University of Eindhoven (Department of Biomedical Engineering) in the Netherlands, and Comité de Protection des Personnes SUD-EST IV in France]. Individual patient consent was obtained prior to participation. All images were acquired using XTreme CT 1 (Scanco Medical AG) at standard clinical settings, with a nominal isotropic voxel size of 82 μm. A SPECTRA-endorsed image acquisition protocol was used at all investigative sites. Images were viewed using OsiriX software (Version 5.8) on 27-inch cinema screen iMac computers.

*Reliability exercise participants.* Eleven readers participated in RELEX-1, with varied levels of experience reading HR-pQCT data sets (3 readers > 5 yrs, 4 readers 1-3 yrs, 4 readers no experience). Readers also varied in their training backgrounds, and included practicing rheumatologists, researchers, and either research or medical trainees. This variation in experience was purposeful in that discussion around the definition would include basic and obvious descriptions of the observations, refined by the expertise of the experienced readers in nonpathologic findings. Inexperienced readers were given the opportunity to familiarize themselves with the HR-pQCT images of joints and the imaging software in a half-day session before the meeting.

*HR-pQCT convention for joint image scoring.* Only images of acceptable quality (Scanco Grades 1-3) were evaluated. Eight surfaces for each second and third metacarpophalangeal (MCP) joint were assessed as described by Stach, et al., which included the palmar, dorsal, radial, and ulnar aspects of each of the proximal phalanx and metacarpal head.

*Achieving consensus on erosion definition and landmarks.* All proposed definitions and measurement techniques were reviewed at the beginning of the meeting. An initial definition for erosion synthesized from the existing literature (as summarized in the introduction) was proposed, and images from 15 subjects (120 surfaces) were scored. A general discussion on the characteristic properties of erosions and refinement of the definition ensued to increase specificity. A second round of image scoring using the refined definition was performed on images from a second set of 15 subjects (120 surfaces), while incorporating proposed landmarks for the measurement of erosion dimensions based on the manual approach described by Albrecht, et al. From this second round of scoring, for the purposes of the exercise, it was deemed necessary to note whether multiple erosions were present on the same surface and to record measurements for the largest erosion only. We also proposed simplifying the manual measurements by recording the depth of the erosion on the same slice as where the maximal width was obtained. Refined measurements on the recording sheet were made for efficiency. The final round of scoring applied these consensus-achieved definitions on 58 joints not previously scored in the exercise and blinded to disease status (control, RA case, or cadaveric specimen) to reduce bias.

*Statistical analysis.* For results from round 3, the percentage agreement for the absence or presence of an erosion and a κ score were calculated. Descriptive statistics were applied to provide an assessment of minimal, maximal, and mean erosion dimensions and to identify the presence of any significant outliers. The root mean square coefficient of variation (RMSCV) and intraclass correlation (ICC) with a mixed-effects model were used as indicators of variability in erosion depth and width measurement between readers. Analysis was performed on STATA IC (version 11.0).

**RESULTS**

*Definition for erosion (Figure 1).* The following definition, which has been accepted by SPECTRA, was applied in this reliability assessment:

- Presence of a definite interruption in the cortical bone

---

**Table:**

<table>
<thead>
<tr>
<th>Definition for Erosion</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of a definite interruption in the cortical bone</td>
<td></td>
</tr>
</tbody>
</table>
• Cortical break must extend over at least 2 consecutive slices
• Cortical break must be detectable in 2 perpendicular planes
• Loss of underlying trabecular bone at the cortical break
• Nonlinear in shape (to differentiate from vascular channels penetrating the cortices (Figure 2)

Interreader agreement for erosion detection. In total, 58 joints were scored, with 94.9% of surfaces (440/464) of sufficient quality for evaluation. The percentage agreement for the presence or absence of erosions between all readers was 90.2% (397/440 surfaces). The chance corrected agreement yielded a moderate $\kappa$ value of 0.515 (95% CI 0.373–0.632). The $\kappa$ value was highest among experienced readers at 0.748.

Landmarks for measurement (in 2 dimensions; Figure 3). All measurements reflect the Euclidian distance in millimeters between the landmarks based on the pixel size (1 pixel = 0.082 mm). First, in both the axial and corresponding perpendicular planes independently, the maximum width of the cortical break is determined by drawing and measuring the length of a straight line between the 2 margins of the break. On the same slice where this maximum width is determined, the depth is measured. The depth measurement is taken at a 90° angle to the drawn width line, extending to the deepest point of the erosion.

Erosion dimensions and CV. Erosion size and the variability in this measurement were determined on surfaces where at least 5 readers agreed that an erosion was present (n = 10). Erosion size ranged from 0.16 to 8.9 mm in maximal width (mean 1.84 mm) and 0.3 to 8.0 mm in maximal depth (mean 1.86 mm) in any MPR. The RMSCV varied from 12.3% for the perpendicular width, 20.6% for axial width, 22.2% for axial depth, and 24.0% for perpendicular depth. We did not observe that the RMSCV varied by erosion size (Figure 4). ICC were 0.206 for perpendicular width, 0.665 for axial width, 0.871 for axial depth, and 0.783 for perpendicular depth. When restricted to experienced readers (> 5 yrs experience), ICC were > 0.9.

Following round 3, we identified 8 surfaces where there was still disagreement by multiple readers on the presence of an erosion at that surface. Participants gathered during the European League Against Rheumatism meeting in June 2013 to review these images and proposed further studies to confirm the etiology of overlapping bony pathologies that mimic erosions, such as vascular channels near overhanging osteophytes.

DISCUSSION
In this first international reliability exercise using HR-pQCT technology for the assessment of erosions in RA, consensus was reached for a definition easily applied to the images with...
good agreement (> 90% agreement and a moderate \( \kappa \) value), and consistent with the definition used for US\textsuperscript{16}. Erosions imaged with HR-pQCT thereby are defined as a definite, nonlinear interruption in the cortical bone, extending over at least 2 consecutive slices and observed in 2 perpendicular MPR, where there is loss of underlying trabecular bone. In contrast, parallel breaks in cortical bone are usually vessel channels that occur physiologically\textsuperscript{9}. Because the difference between pathological cortical breaks (i.e., erosions) and physiological cortical breaks (i.e., vascular channels) is
currently based on appearance on HR-pQCT images and some preliminary studies using US, confirmation by histology and perfusion studies is required in the course of the validation of the method. It is important to note this definition may apply in other forms of inflammatory as well as degenerative arthritis, but does require validation in those other disease states, because shape, frequency, distribution, and accompanying bone catabolic or anabolic factors might vary between disease entities.

The SPECTRA collaboration includes investigators pursuing 3-D methodologies for volumetric erosion and surface area measurements and shape characterization in an automated fashion, as well as image coregistration and slice matching, which will be important for application of this technology in longitudinal studies. These, however, are still under development and not ready for use at this time. Therefore, we have proposed standard landmark reference points in 2 separate perpendicular MPR.

Our results indicate moderate variability in erosion measurement between readers, and thus indicate that there is still a need to improve setting of landmarks for erosion size determination, which is especially important given early indications of the ability to measure erosion healing as a result of biologic therapy. Future SPECTRA activities will focus on providing more precision in the landmark selection. However, we are encouraged that our preliminary validation work is reminiscent of that in the early magnetic resonance imaging studies.

The use of HR-pQCT in arthritis conditions is expanding. Since our initial SPECTRA meeting, additional investigation sites have begun research into this application. Interest remains in defining periarticular bone changes in early and late phases of RA, osteoarthritis, psoriatic arthritis, and crystal arthropathies. Multimodality imaging is being adopted, and other groups have developed methods to define the joint space width. This unique collaboration between rheumatologists, radiologists, biomedical engineers, physicists, and clinical researchers in bone metabolism and osteoimmunology provides a fertile ground for cross-disciplinary research, which will ultimately improve our ability to quantify bone damage in inflammatory and degenerative arthritis conditions. As we develop the technology, feasibility issues will be resolved so that the methodology can be applied in trials conducted at sites with HR-pQCT capabilities.

Following appropriate training under the direction of experienced readers, it is likely that this technology can be applied in future studies. This consensus definition for an erosion arising from our first international collaboration study will allow individual investigators to ensure they are describing the same bony pathologies and ensure homogeneity of result interpretation across multicenter studies as we pursue use of HR-pQCT as a novel imaging biomarker in rheumatology.

**APPENDIX 1.**

List of study collaborators. SPECTRA Collaboration Members: Cheryl Barnabe, University of Calgary; Susan Barr, University of Calgary; [Figure 4. Plot of root mean square coefficient of variation (RMSCV) by mean erosion size. This graph demonstrates the distribution of variability in the RMSCV among readers by the mean measured dimensions. Erosion size does not appear to affect the RMSCV.]
Stephanie Boutrou, 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; Joost de Jong, Maastricht University Medical Centre; Klaus Engelke, University of Erlangen; Lynne Feehan, University of British Columbia; Stephanie Finzel, Department of Rheumatology and Clinical Immunology, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg; Piet Geusens, Maastricht University Medical Centre; Ellen-Margrethe Hauge, Aarhus University Hospital; Yebin Jiang, University of Michigan; Kresten Keller, Aarhus University Hospital; Roland Kocijan, University of Vienna; Sebastian Kraus, University of Erlangen; Xiaojuan Li, University of California San Francisco; Hubert Marotte, Université de Lyon; Lian Martin, University of Calgary; Rosa Pereira, University of Sao Paolo; Andrea Scharmag, Maastricht University Medical Centre; Georg Schett, Maastricht University Medical Centre; Kristian Stengaard-Pedersen, Aarhus University Hospital; Kathryn S. Stok, ETH Zurich; Nikolay Tzaribachev, SCANCO Medical AG; Joop van den Bergh, Department of Rheumatology and Clinical Immunology, University Medical Center Freiburg, Faculty of Medicine, University of Freiburg; Piet Geusens, Maastricht University Medical Centre; Ellen-Margrethe Hauge, Aarhus University Hospital; Kristian Stengaard-Pedersen, Aarhus University Hospital; Kathryn S. Stok, ETH Zurich; Nikolay Tzaribachev, SCANCO Medical AG; Joop van den Bergh, Maastricht University Medical Centre; Bert van Rietbergen, Eindhoven University of Technology; Astrid van Tubergen, Maastricht University Medical Centre; Mira van Veenendaal, University of Toronto; Paul Willems, Maastricht University Medical Centre; Piet Williams, Maastricht University Medical Centre; Jeroen Williams, Maastricht University Medical Centre; Rae Yeung, University of Toronto.

REFERENCES


