Skeletal Radiol (2007) 36:315–320
DOI 10.1007/s00256-006-0208-z

SCIENTIFIC ARTICLE

Longitudinal in vivo reproducibility of cartilage volume and surface in osteoarthritis of the knee

M. H. Brem · J. Pauser · H. Yoshioka · A. Brenning · J. Stratmann · F. F. Hennig · R. Kikinis · J. Duryea · C. S. Winalski · P. Lang

Received: 31 May 2006 / Revised: 13 August 2006 / Accepted: 15 August 2006 / Published online: 12 January 2007

Abstract
Objective The aim of this study was to evaluate the longitudinal reproducibility of cartilage volume and surface area measurements in moderate osteoarthritis (OA) of the knee.
Materials and methods We analysed 5 MRI (GE 1.5T, sagittal 3D SPGR) data sets of patients with osteoarthritis (OA) of the knee (Kellgren Lawrence grade I–II). Two scans were performed: one baseline scan and one follow-up scan 3 months later (96±10 days). For segmentation, 3D Slicer 2.5 software was used. Two segmentations were performed by two readers independently who were blinded to the scan dates. Tibial and femoral cartilage volume and surface were determined. Longitudinal and cross-sectional precision errors were calculated using the standard deviation (SD) and coefficient of variation (CV%=100×[SD/mean]) from the repeated measurements in each patient. The in vivo reproducibility was then calculated as the root mean square of these individual reproducibility errors.
Results The cross-sectional root mean squared coefficient of variation (RMSE-CV) was 1.2, 2.2 and 2.4% for surface area measurements (femur, medial and lateral tibia respectively) and 1.4, 1.8 and 1.3% for the corresponding cartilage volumes. Longitudinal RMSE-CV was 3.3, 3.1 and 3.7% for the surface area measurements (femur, medial and lateral tibia respectively) and 2.3, 3.3 and 2.4% for femur, medial and lateral tibia cartilage volumes.
Conclusion The longitudinal in vivo reproducibility of cartilage surface and volume measurements in the knee using this segmentation method is excellent. To the best of our knowledge we measured, for the first time, the longitudinal reproducibility of cartilage volume and surface area in participants with mild to moderate OA.

Keywords Cartilage volume · Osteoarthritis · MR imaging · Knee joint

M. H. Brem · J. Pauser · H. Yoshioka · J. Stratmann · R. Kikinis · J. Duryea · C. S. Winalski · P. Lang (✉)
Musculoskeletal Division, Department of Radiology, ASB-1, L-1, Room 003E, Harvard Medical School,
Brigham and Women's Hospital,
75 Francis Street,
Boston, MA 02115, USA
e-mail: pklang@partners.org

M. H. Brem · F. F. Hennig
Division of Trauma Surgery and Orthopaedic Surgery,
Department of Surgery, Friedrich-Alexander,
University of Erlangen-Nuremberg,
Erlangen, Germany

A. Brenning
Department of Medical Informatics, Biometry and Epidemiology,
Friedrich-Alexander, University of Erlangen-Nuremberg,
Erlangen, Germany

C. S. Winalski
Division of Radiology, Cleveland Clinic Foundation,
Cleveland, OH, USA

Introduction

Injuries and degenerative changes in cartilage are significant causes of osteoarthritis (OA), leading to reduced quality of life. The high incidence and the increasing average lifespan of the population has resulted in enormous costs in the diagnosis and treatment of OA. In the USA, the annual total cost of arthritis and other rheumatic conditions was about $116.3 billion in 1997 [1]. The economic problem is of similar magnitude in European countries [2]. Thus, there is a significant need for a non-invasive method of monitoring OA to judge the success of potential chondroregenerative and surgical treatment. Magnetic resonance imaging (MRI)

is widely used as a non-invasive imaging tool for OA of the knee [3, 4]. The measurement of cartilage volume has been reported to be a measure to monitor disease progression [5]. The accuracy of cartilage volume and surface measurement has been validated for healthy and moderate to advanced OA in cross-sectional studies [6–10]. However, in assessing the reproducibility of a quantitative measurement technique, it is important not only to assess short-term reproducibility in cross-sectional, single time point data, but also to evaluate longitudinal reproducibility.

In the present study, we evaluated the longitudinal reproducibility of cartilage volume and surface in patients with early mild to moderate OA with Kellgren-Lawrence stage I– II changes.

Materials and methods

We analysed MR images from 5 patients (2 female and 3 male) with osteoarthritis (OA) of the knee. The ages ranged from 40–73 years (mean age 64.3 years). These patients were enrolled in the placebo arm in a clinical trial evaluating a chondroregenerative drug for the treatment of OA, which was approved by the Institutional Review Board. All had a cartilage lesion in a weight-bearing area of the knee (Kellgren–Lawrence grade I–II). We have previously described detailed inclusion criteria for this clinical trial [11].

The MR images were obtained with a 1.5 T MR scanner (Signa, software version 8.3; GE Healthcare, Milwaukee, WI, USA) using a transmit–receive extremity coil. In each patient, sagittal fat-suppressed 3D spoiled gradient recalled (SPGR) images (repetition time [TR] ms/echo time [TE] ms/flip angle = 60/5/40°) were obtained with continuous 1.5-mm section thickness, a partition of 64, an image matrix of 256×160, a 120-mm field of view (FOV), a band width of 16 kHz, and one excitation for a total acquisition time of 10 min 18 s. Two scans were performed: one baseline scan and one follow-up scan 3 months later (96± 10 days).

The MR image data were transferred to a workstation using our in-house software 3D Slicer version 2.5 (http://www.Slicer.org). For each of the two scans a manual segmentation was performed (Fig. 1). With manual segmentation, the reader draws a line around the cartilage borders of the femur and tibia cartilage on every single MRI slice with a computer mouse. A second round of segmentations was performed 2 weeks later to avoid memory recall. Both segmentation sessions were done by each reader independently and blinded to the scan dates. The segmentations for baseline and follow-up scans were not segmented simultaneously. Each reader segmented all the baseline scans first and all the follow-up scans in the second segmentation round. The segmentation and the re-segmentation were performed by two experts (MHB and JP) experienced in cartilage imaging and segmentation. Both readers underwent extensive training (more than 1 year) on the recognition of cartilage damage prior to the study. Furthermore, they underwent 4 months' training on cartilage segmentation. The volume of the femoral and the

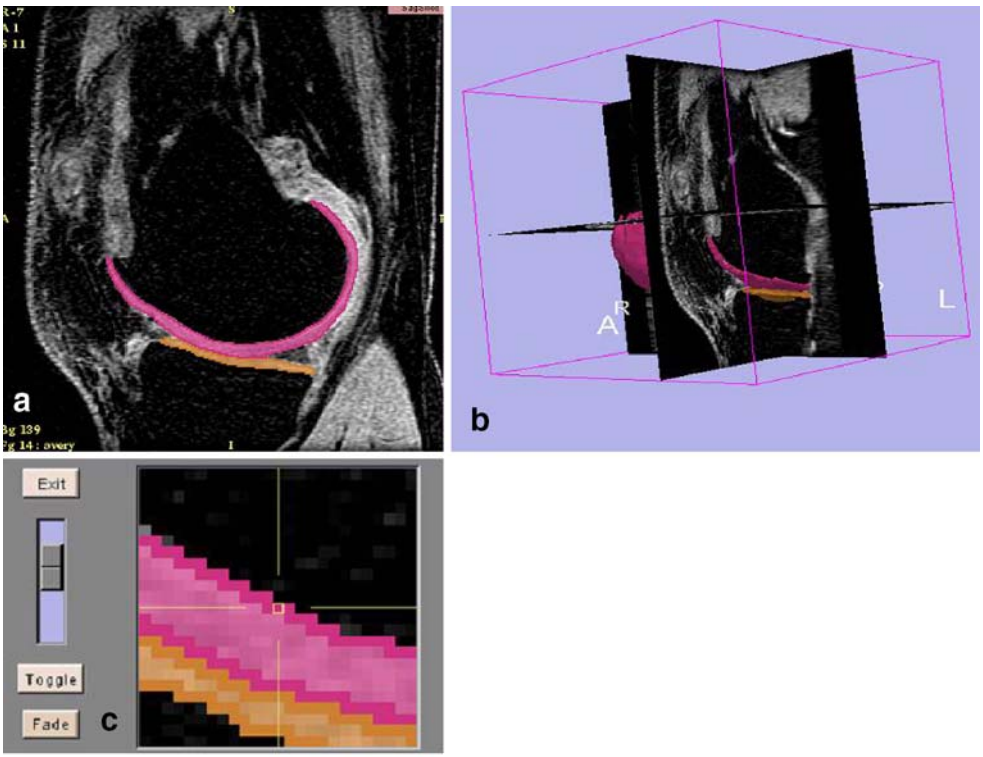

Fig. 1 Screen shots of the segmentation software. **a** Sagittal slice of the knee with segmented femur and tibial cartilage. **b** 3D view of the same knee with 3D models of the femur and tibial cartilage. **c** Zoom view of the region, which is segmented (*violet red* femur cartilage, *green* tibia cartilage). The zoom view offers the possibility of segmenting precisely, pixel by pixel

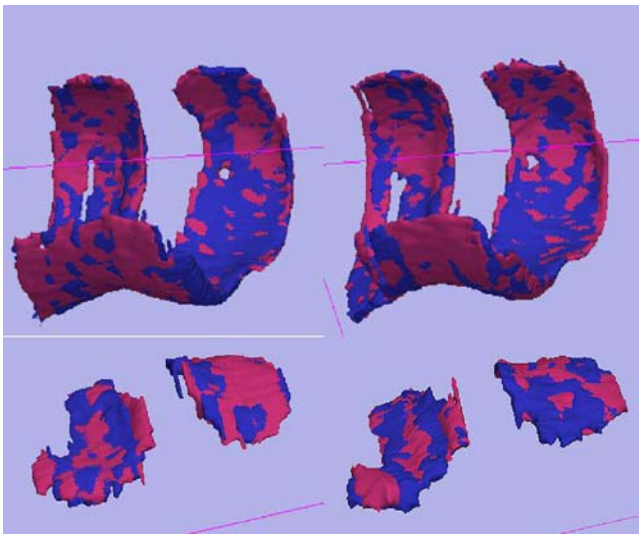

Fig. 2 3D models of the segmented cartilage (*pink* Reader 1, *blue* Reader 2). The models of time point 1 (*left*) and time point 2 are projected over each other. The "spotting" appearance of the superimposed models shows that essentially neither of the readers segmented differently from the other

medial and lateral tibial cartilage was measured and the total surface area of the femoral and the medial and lateral tibial cartilage 3D model was calculated (Fig. 2). The area and volume measurement were computed using the marching cubes algorithm [12, 13].

Measurements of reproducibility and precision errors were calculated as described by Glüer et al. for test–retest (short term/cross-sectional) reproducibility for a single time point and longitudinal (long-term) reproducibility for dual time points [14].

According to this approach, precision error is expressed as the root mean squared standard deviations of the individual patients (RMSE-SD) and the corresponding coefficient of variation (RMSE-CV) [14]. The individual standard deviations are each based on a total of four readings made by two readers for the same scan. While short-term precision refers to the reproducibility of measurements derived from the same scan, the long-term precision error represents the reproducibility between two scans (time point 2 versus 1), including variability that depends on the readers, image quality, software and MRI machine. We computed each patient's standard error of estimation (SEE), taking into account the appropriate degrees of freedom, and obtained the total estimated long-term longitudinal precision error by computing the root mean squared error (RMSE-SEE; see Glüer et al. [14]). For data analyses the open-source data analysis environment R was used [15].

Results

The RMSE-CV for the short-term precision of the femoral volume cartilage ranged from 1.4% for time point 1 (t=1) to 1.3% for time point 2 (t=2). The RMSE-SD was 0.21 for t=1 and 0.19 for t=2. The results for the short-term precision of the medial tibial cartilage volume RMSE-CV ranged from 1.1% (t=1) to 1.8% (t=2) and we found a RMSE-SD=0.03 for t=1 and RMSE-SD=0.05 for t=2. The RMSE-CV for the short-term precision of the lateral tibial cartilage volume ranged from 1.3% (t=1) to 1.1% (t=2) and showed RMSE-SD of 0.03 for t=1 and 0.03 for t=2.

The RMSE-CV for the short-term precision of the femoral surface area RMSE-CV was 1.0% (t=1) and 1.2% (t=2). The RMSE-SD was 137.29 (t=1) and 162.34 (t=2). The RMSE-SV for the short-term precision for the surface area of the medial tibial cartilage was 2.2% for t=1 and 1.6% for t=2 (RMSE-SD = 55.92 and 40.06 respectively). The RMSE-CV for the short-term precision for the surface area of the lateral tibial cartilage was 1.5% for t=1 and 2.4% for t=2. The RMSE-SD were 36.06 and 58.41 respectively.

For the long-term precision of the femoral volume the RMS-CV was 2.3% and the RMSE-SEE was 0.33. The long-term precision for the medial tibial volume showed a RMSE-CV of 3.3% and a RMS-SEE of 0.08. For the long-term precision of the lateral tibial volume the RMSE-CV was 2.3% and the RMS-SEE 0.06. The RMSE-CV for the long-term precision of the femoral surface area was 3.3% and the RMSE-SEE was 446.66.

Table 1 Results of the calculation of the root mean square (RMSE) for short-term and long-term precision

	Short-term/single time point (t=1)		Short-term/single time point (t=2)		Long-term/dual time point	
	RMS-SD	RMS-CV%	RMS-SD	RMS-CV%	RMS-SEE	RMS-CV%
Femur volume	0.21 mm^3	1.4	0.19 mm^3	1.3	0.33 mm^3	2.3
Tibia medial volume	0.03 mm^3	1.1	0.05 mm^3	1.8	0.08 mm^3	3.3
Tibia lateral volume	0.03 mm^3	1.3	0.03 mm^3	1.1	0.06 mm^3	2.3
Femur surface	137.29 mm^2	1.0	162.34 mm^2	1.2	446.66 mm^2	3.3
Tibia medial surface	55.92 mm^2	2.2	40.06 mm^2	1.6	78.29 mm^2	3.1
Tibia lateral surface	36.06 mm^2	1.5	58.41 mm^2	2.4	92.12 mm^2	3.7

RMS-SD root mean squared standard deviations, *RMS-CV* root mean squared coefficient of variation.

Table 2 Mean values of volume and surface measurement

	Volume measurement (mm^3)			Surface measurement (mm^2)		
	t=1	t=2	Difference in %	t=1	t=2	Difference in %
Femur	14.86	14.94	<1	13,677.69	13,520.33	1.2
Tibia medial	2.48	2.45	1.2	2,500.23	2,522.59	<1
Tibia lateral	2.58	2.6	<1	2,466.24	2,431.87	1.4

For the long-term precision of the medial tibial cartilage surface the RMS-CV was 3.1% and the RMSE-SEE was 78.29. For the long-term precision of the lateral tibia we found a RMS-CV of 3.7% and the RMSE-SEE was 92.12 (Table 1).

The mean results of both readers differed from each other by only a small range between less than 1% and a maximum of 2%. This shows a high inter-reader precision for the cartilage volume as well as for the cartilage surface (Fig. 2, Tables 2, 3).

Discussion

To our knowledge, this is the first time that long-term reproducibility in patients with mild to moderate OA has been evaluated. The RMS-CVs expressed in CV% ranged between 2.3 and 3.3% for the long-term volume measurement and between 3.1 and 3.7% for the long-term surface area measurement for all areas.

Previous studies evaluated the longitudinal reproducibility of cartilage volume with data acquired from healthy volunteers [9, 16–18]. They achieved precision errors of up to 3.8% (3 healthy volunteers, baseline scan and 2-month follow-up scan) [18]. Another study showed results ranging from 2.5 to 3.6% CV for reproducibility of volume and 2.1 to 3.1% CV for reproducibility of surface measurements [9]. Our data indicate that longitudinal reproducibility similar to that obtained in healthy participants can be achieved in patients with mild to moderate OA. This is important since this is the most likely target group for future pharmacologic intervention.

The manual segmentation technique in this study used the "3D-Slicer" software. This application is widely applicable and can be downloaded free of charge (http://www.Slicer.org). Semi-automated segmentation is feasible in healthy participants [17]. It may not be feasible in patients with cartilage defects, due to changes in cartilage signal and loss of borders between cartilage and surrounding tissue. Especially in the posterior femoral region non-uniform signal intensity can cause difficulties in the detection of cartilage borders as reported by Yoshioka et al. [11]. This non-uniform signal intensity needs direct the intervention of the reader.

We did not evaluate sub-regions of femoral and tibial cartilage since our purpose was to measure the precision of the segmentation method. We did not want to confound our results by introducing precision errors introduced by the choice of sub-region landmarks.

Short-term errors are within the range reported previously by other authors for healthy participants and OA patients [6, 10, 18]. A recently published study evaluated the longitudinal change in cartilage volume in OA (Kelgren I–III) over a time span of 2 years [16]. For coronal reformatted planes they achieved intra-observer values (CV) of 2.3% for lateral tibia cartilage to 2.8% for the lateral femoral cartilage, similar to our values for repeated measurements of the same knee at the same time points. We found the smallest precision error by calculating the RMS-CV (1.1%) for the femoral volume measurement and the highest by calculating the RMSE-CV for the medial tibial surface (2.4%). The relative high precision error may be related to the small mean value, as described above.

Table 3 Mean values of surface and volume by readers and differences between the mean values in percent

	t=1			t=2		
	Reader 1	Reader 2	Difference in %	Reader 1	Reader 2	Difference in %
Femur volume in mm^3 (mean)	15.01	14.71	2	15.09	14.79	2
Medial tibial volume in mm^3 (mean)	2.48	2.49	<1	2.47	2.43	1.4
Lateral tibial volume in mm^3 (mean)	2.59	2.57	<1	2.61	2.59	<1
Femur surface in mm^2 (mean)	13,734.92	13,620.45	<1	13,582.81	13,457.85	<1
Tibia medial surface in mm^2 (mean)	2,503.36	2,497.1	<1	2,530.28	2,514.9	<1
Tibia lateral surface in mm^2 (mean)	2,467.32	2,465.15	<1	2,427.92	2,435.82	<1

In a 2-year study Wluka et al. calculated an average annual volume loss in the medial and lateral tibial compartment of about 4.7 and 5.3% respectively, assuming a linear loss [5]. Using the same assumptions, the maximum loss observed during the 3-month observation period would be 1.3%. Of note, percent differences between time points 1 and 2 were much smaller in this study, suggesting that no significant progression had occurred during the longitudinal observation period. Thus, the longitudinal observation period selected in this study appears to be ideally suited to assessing machine or imaging technique-related longitudinal errors, while minimising the effects of disease progression on the measurement.

We chose the statistical technique, as described above, to quantify short- and long-term variability to be consistent with a statistical technique that has been applied in published studies on this subject. It would therefore permit comparable measures between similar studies [9, 10]. For estimating long-term precision errors, we made the reasonable assumption that cartilage volumes and surfaces do not change considerably during the 3-month period between the two scans. Wluka et al. reported a mean volume loss of about 5.3% tibial cartilage on an annual basis [5]. Assuming linear progression cartilage loss during a 3-month period, cartilage loss would be 1.3% during the observation period in this study. This is well below reproducible errors of measurement of cartilage volume in cross-sectional single time point studies. Based on this assumption, we considered each patient's cartilage volume and surface at time point 2 as being not significantly changed compared with the mean of the measurements made at time point 1.

We did not calculate inter-reader correlations because we do not consider it a meaningful measure for inter-reader reproducibility in the present context. For example, a high correlation could be the result of great differences between baseline and follow-up measurements. Paradoxically, if there is no systematic difference between baseline and follow-up measurements and between first and second segmentations (as suggested by our results), the inter-reader correlation for each patient would inevitably vary randomly around zero as a consequence of non-systematic random measurement/segmentation error.

This study has several limitations. The data were acquired from a single MRI manufacturer and at 1.5 T with a single type of acquisition. Results may be different with other equipment, field strength and pulse sequence. This study did not specifically address the time taken to draw and measure cartilage and volume. It is a technique that may take a couple of hours.

In conclusion, our study showed that the excellent longitudinal reproducibility of cartilage volume and surface area measurements can be obtained in patients with mild to moderate OA, the primary target group for future pharmacologic intervention. This method is ideally suited for longitudinal assessment of progression of disease in OA or therapeutic response in longitudinal drug trials. This is an essential requirement for any studies utilising MRI for monitoring therapeutic efficacy. MRI-based measurements of cartilage volume and surface are well suited for longitudinal studies.

Acknowledgements Part of this work was funded by the NIH: NIH P41-RR13218, National Alliance for Medical Image Computing (NAMIC), http://www.na-mic.org, and through the NIH Roadmap for Medical Research, NIH U54 EB005149.

References

1. Cisternas M, Yelin E, Trupin L, Murphy L, Helmick CG. Direct and indirect costs of arthritis and other rheumatic conditions—United States, 1997. Report. Atlanta: Centers for Disease Control and Prevention, November 21, 2003.
2. Pientka L. Arthrose als Volkskrankheit. Klin Forsch 2000;6 Suppl:2:2–3.
3. Disler DG, McCauley TR, Wirth CR, Fuchs MD. Detection of knee hyaline cartilage defects using fat-suppressed three-dimensional spoiled gradient-echo MR imaging: comparison with standard MR imaging and correlation with arthroscopy. AJR Am J Roentgenol 1995;165 (2):377–82.
4. Peterfy CG, van Dijke CF, Janzen DL, Glüer CC, Namba R, Majumdar S et al. Quantification of articular cartilage in the knee with pulsed saturation transfer subtraction and fat-suppressed MR imaging: optimization and validation. Radiology 1994;192 (2):485–91.
5. Wluka AE, Stuckey S, Snaddon J, Cicuttini FM. The determinants of change in tibial cartilage volume in osteoarthritic knees. Arthritis Rheum 2002;46 (8):2065–72.
6. Cicuttini F, Forbes A, Asbeutah A, Morris K, Stuckey S. Comparison and reproducibility of fast and conventional spoiled gradient-echo magnetic resonance sequences in the determination of knee cartilage volume. J Orthop Res 2000;18 (4):580–4.
7. Cicuttini F, Forbes A, Morris K, Darling S, Bailey M, Stuckey S. Gender differences in knee cartilage volume as measured by magnetic resonance imaging. Osteoarthr Cartil 1999;7 (3):265–71.
8. Eckstein F, Westhoff J, Sittek H, Maag KP, Haubner M, Faber S et al. In vivo reproducibility of three-dimensional cartilage volume and thickness measurements with MR imaging. AJR Am J Roentgenol 1998;170 (3):593–7.
9. Eckstein F, Heudorfer L, Faber SC, Burgkart R, Englmeier KH, Reiser M. Long-term and resegmentation precision of quantitative cartilage MR imaging (qMRI). Osteoarthr Cartil 2002;10 (12):922–8.
10. Eckstein F, Charles HC, Buck RJ, Kraus VB, Remmers AE, Hudelmaier M et al. Accuracy and precision of quantitative assessment of cartilage morphology by magnetic resonance imaging at 3.0T. Arthritis Rheum 2005;52 (10):3132–6.
11. Yoshioka H, Stevens K, Genovese M, Dillingham MF, Lang P. Articular cartilage of knee: normal patterns at MR imaging that mimic disease in healthy subjects and patients with osteoarthritis. Radiology 2004;231 (1):31–8.
12. Loresen We, Cline HE. Marching cubes: a high resolution 3D surface reconstruction algorithm. Comput Graph (ACM) 1987; 21 (4):163–9.
13. Sommerville DMY. Analytical geometry of three dimensions. Cambridge: Cambridge University Press; 1959.

14. Glüer CC, Blake G, Lu Y, Blunt BA, Jergas M, Genant HK. Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. Osteoporos Int 1995;5 (4):262–70.
15. Development CoreTeam R. R: a language and environment for statistical computing. In: R Foundation for Statistical Computing, http://www.R-project.org. Vienna; 2005.
16. Cicuttini FM, Wluka AE, Wang Y, Stuckey SL. Longitudinal study of changes in tibial and femoral cartilage in knee osteoarthritis. Arthritis Rheum 2004;50 (1):94–7.
17. Stammberger T, Eckstein F, Michaelis M, Englmeier KH, Reiser M. Interobserver reproducibility of quantitative cartilage measurements: comparison of B-spline snakes and manual segmentation. Magn Reson Imaging 1999;17 (7): 1033–42.
18. Kshirsagar AA, Watson PJ, Tyler JA, Hall LD. Measurement of localized cartilage volume and thickness of human knee joints by computer analysis of three-dimensional magnetic resonance images. Invest Radiol 1998;33 (5):289–99.