



Three-Dimensional Surface-Based Analysis of Cartilage MRI Data in Knee Osteoarthritis: Validation and Initial Clinical Application

James W. MacKay, PhD,^{1,2*}  Joshua D. Kaggie, PhD,¹  Graham M. Treece, PhD,³
Stephen M. McDonnell, MD,⁴ Wasim Khan, PhD,⁴ Alexandra R. Roberts, PhD,^{5,6}
Robert L. Janiczek, PhD,⁵ Martin J. Graves, PhD,¹ Tom D. Turmezei, PhD,^{2,7†}
Andrew W. McCaskie, MD,^{4†} and Fiona J. Gilbert, MD^{1†}

Background: Traditional quantitative analysis of cartilage with MRI averages measurements (eg, thickness) across regions-of-interest (ROIs) which may reduce responsiveness.

Purpose: To validate and describe clinical application of a semiautomated surface-based method for analyzing cartilage relaxation times ("composition") and morphology on MRI, 3D cartilage surface mapping (3D-CaSM).

Study Type: Validation study in cadaveric knees and prospective observational (cohort) study in human participants.

Population: Four cadaveric knees and 14 participants aged 40–60 with mild–moderate knee osteoarthritis (OA) and 6 age-matched healthy volunteers, imaged at baseline, 1, and 6 months.

Field Strength/Sequence: 3D spoiled gradient echo, T₁rho/T₂ magnetization-prepared 3D fast spin echo for mapping of T₁rho/T₂ relaxation times and delayed gadolinium enhanced MRI of cartilage (dGEMRIC) using variable flip angle T₁ relaxation time mapping at 3T.

Assessment: 3D-CaSM was validated against high-resolution peripheral quantitative computed tomography (HRpQCT) in cadaveric knees, with comparison to expert manual segmentation. The clinical study assessed test–retest repeatability and sensitivity to change over 6 months for cartilage thickness and relaxation times.

Statistical Tests: Bland–Altman analysis was performed for the validation study and evaluation of test–retest repeatability. Six-month changes were assessed via calculation of the percentage of each cartilage surface affected by areas of significant change (%SC), defined using thresholds based on area and smallest detectable difference (SDD).

Results: Bias and precision (0.06 ± 0.25 mm) of 3D-CaSM against reference HRpQCT data were comparable to expert manual segmentation (−0.13 ± 0.26 mm). 3D-CaSM demonstrated significant (>SDD) 6-month changes in cartilage thickness and relaxation times in both OA participants and healthy controls. The parameter demonstrating the greatest 6-month change was T₂ relaxation time (OA median %SC [IQR] = 8.8% [5.5 to 12.6]).

Data Conclusion: This study demonstrates the construct validity and potential clinical utility of 3D-CaSM, which may offer advantages to conventional ROI-based methods.

Level of Evidence: 2.

Technical Efficacy Stage: 2.

J. MAGN. RESON. IMAGING 2020.

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.27193

Received Feb 20, 2020, Accepted for publication Apr 24, 2020.

*Address reprint requests to: J.W.M., Norwich Medical School, Bob Champion Research and Education Building, Norwich Research Park, Norwich NR4 7UQ, UK. E-mail: james.w.mackay@uea.ac.uk

†Joint senior authors.

From the ¹Department of Radiology, University of Cambridge, Cambridge, UK; ²Norwich Medical School, University of East Anglia, Norwich, UK; ³Department of Engineering, University of Cambridge, Cambridge, UK; ⁴Division of Trauma & Orthopaedics, Department of Surgery, University of Cambridge, Cambridge, UK; ⁵Clinical Imaging, GlaxoSmithKline, London, UK; ⁶Antaros Medical, Uppsala, Sweden; and ⁷Department of Radiology, Norfolk & Norwich University Hospital, Norwich, UK

Additional supporting information may be found in the online version of this article

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

CARTILAGE BREAKDOWN is a hallmark of osteoarthritis (OA), and its prevention remains a therapeutic target.¹ Therefore, the development and optimization of quantitative imaging biomarkers (QIBs) of cartilage health is desirable.

Quantification of cartilage morphology (volume and thickness) and relaxation times (eg, $T_1\rho$, T_2 , T_1) using magnetic resonance imaging (MRI) is well established and has been implemented in several clinical trials and observational studies.^{2–4} Analysis pipelines usually require expert manual segmentation of articular cartilage. While this can be accurate and reliable, it suffers from three main drawbacks. First, it is time-consuming and represents a considerable resource burden for researchers. Second, even in expert hands manual segmentation is, in general, less reproducible and sensitive to change than automated or semiautomated approaches, although this will vary depending on the exact algorithm/method used.^{5,6} Third, cartilage measurements tend to be averaged over large regions of interest (ROIs). While these averaged measurements may be attractively simple, such approaches limit responsiveness and average out potentially important focal changes.⁷ It has been shown that manual segmentation-based approaches for cartilage morphology are unlikely to be sufficiently responsive in clinical studies with 6 months or less follow-up.^{8,9}

These drawbacks limit the utility of cartilage QIBs in clinical trials, particularly in early-phase clinical trials where sample sizes are small and follow-up periods are short. However, with the shift towards biomarker-rich experimental medicine study designs, there is increasing interest in the potential utility of imaging biomarkers, including the ability to demonstrate early proof-of-concept.¹⁰

Improved utility in this setting requires improved responsiveness. This may be achieved by approaches that better reflect the spatial distribution of changes in cartilage and which include automated or semiautomated segmentation pipelines. Here we propose a method that meets these criteria, which we term 3D cartilage surface mapping (3D-CaSM). This is a modification of the previously described methods of cortical bone mapping (CBM) and joint space mapping (JSM). CBM has been validated and used extensively for the measurement of cortical bone thickness in osteoporosis.^{11,12} JSM has been validated for the 3D measurement of joint space width from clinical computed tomography (CT) data.¹³ The underlying measurement algorithm can overcome inherent inaccuracies in measurement of thin plate-like structures at clinical imaging resolutions related to slice thickness and the imaging system's point spread function via model-based deconvolution of the imaging data.^{14,15}

Therefore, the purpose of this study was to describe the validation and initial clinical application of 3D-CaSM, a novel surface-based semiautomatic analysis method for the assessment of cartilage on MRI.

Materials and Methods

Study Design

The described work consists of a validation study in cadavers (referred to as the validation study) and an in vivo study demonstrating the clinical application of this method (referred to as the clinical study). The validation study compared cartilage thickness measurements performed on cadaveric knees between MRI and the reference method of high-resolution peripheral quantitative computed tomography (HRpQCT). The clinical study assessed the interobserver reproducibility and test-retest repeatability of 3D-CaSM and sensitivity to change over 6 months in participants with knee OA and age-matched healthy volunteers.

Cadaveric studies were approved by the University of Cambridge Human Biology Research Ethics Committee, with all participants having given antemortem written consent for postmortem use of samples in research. Clinical studies were approved by the Local Research Ethics Committee, and written informed consent was provided by all participants.

Validation Study: Participants

Four embalmed (not frozen) human cadaveric knees (two female, aged 81–89 years old) were obtained from the University of Cambridge Human Anatomy Centre. Specimens consisted of intact articulated knee joints from distal femoral diaphysis to proximal tibial diaphysis and included all periarticular soft tissues. Details of the embalming process are provided in the Supplementary Materials.

Validation Study: Image Acquisition

Intact cadaveric knees underwent MRI on a clinical 3T system (MR 750; GE Healthcare, Waukesha, WI) using an 8-channel transmit/receive knee coil (Invivo, Gainesville, FL). A 3D fat-suppressed spoiled gradient echo (3D SPGR) sequence was performed aiming to optimize contrast resolution between articular cartilage, adjacent bone, and synovial fluid and to maximize spatial resolution within a clinically feasible acquisition time. Imaging parameters were as follows: field of view, $150 \times 150 \times 140 \text{ mm}^3$; matrix 512×380 (interpolated to 512×512) with in-plane spatial resolution $0.3 \times 0.3 \text{ mm}$; slice thickness 1 mm ; flip angle 25° ; repetition time 26 msec ; echo time 6.8 msec ; acquisition time $\sim 7 \text{ minutes}$.

Anatomical dissection was performed following MRI to allow disarticulation of the knees into individual bones. The periarticular soft tissues, ligaments, capsule, and menisci were removed to leave the bones with cartilage surfaces exposed. Dissection was necessary because 1) the contrast resolution for cartilage on CT in intact joints is poor due to the presence of adjacent structures with similar attenuation coefficients, and 2) intact knees would not fit inside the bore of the HRpQCT scanner. To prevent desiccation, all specimens remained stored in embalming fluid and kept in a refrigerated facility (between 10 and 12°C) when not being imaged.

Disarticulated femurs and tibias were then imaged with HRpQCT (Xtreme CT; Scanco Medical, Brüttisellen, Switzerland) with peak voltage of 59.4 kV , tube current of 0.18 mAs . Images were acquired with an isotropic voxel size of 0.082 mm , providing a higher spatial resolution than achievable using clinical MRI (voxel size $0.3 \times 0.3 \times 1 \text{ mm}$). Disarticulated specimens were mounted in an acrylic holder and secured using padding material prior to scanning, ensuring that an air–cartilage interface was maintained across

the surface and that the articular cartilage was not in contact with the holder at any point.

Validation Study: Image Analysis

3D SPGR images were imported in Digital Imaging and Communications in Medicine (DICOM) format into Stradwin v. 5.4a (University of Cambridge Department of Engineering, Cambridge, UK, now freely available as 'StradView' at <http://mi.eng.cam.ac.uk/Main/StradView/>) for 3D-CaSM. The initial 3D-CaSM analysis process is summarized in Fig. 1 with a full description in the Supplementary Material. This results in ~6000 thickness measurements for a single knee, together with accurately located inner and outer cartilage surfaces.

Full manual cartilage segmentation was performed by a musculoskeletal radiologist with 6 years' experience in OA research (J.M.). Segmentation was done using every second slice (ie, every 2 mm). This takes ~3–4 hours for a single knee. Manual thickness values were obtained by taking the distance between the intersections of the outer and inner surface on the manual segmentation with an automatically generated vector normal to each vertex on the cartilage patches generated by 3D-CaSM for that knee, ensuring spatial correspondence of 3D-CaSM and manual thickness values.

Detailed methodology for the generation of HRpQCT measurements is provided in the Supplementary Material. In brief, the process resulted in a set of spatially corresponding HRpQCT and MRI thickness measurements.

Validation Study: Statistical Analysis

The thickness value at each vertex on each cadaveric MRI surface was compared to the thickness value at the identical location on the corresponding HRpQCT surface for both 3D-CaSM and manual segmentation. Bland–Altman analysis was performed to calculate the mean bias, 95% limits of agreement (LOA), and repeatability coefficients (RC) for MRI compared to HRpQCT thickness measurements (Supplementary Methods). For the purposes of this study, the LOA is the interval $[-RC, RC]$ where the difference between two measurements under repeatability conditions for a randomly selected vertex is expected to be 95% of the time.¹⁶

Data were analyzed separately for each cartilage surface and also with data combined from all surfaces. By performing appropriate spatial normalization (surface-to-surface combined similarity and thin-plate spline registration performed in wxRegSurf v. 18; University of Cambridge Department of Engineering, Cambridge, UK, freely available at <http://mi.eng.cam.ac.uk/~ahg/wxRegSurf/>), the spatial distribution of MRI-HRpQCT error values across all cadaveric subjects could be displayed on representative canonical (template) surfaces.

Clinical Study: Participants

Fourteen participants with mild–moderate knee OA and six healthy volunteers (HV) were imaged at baseline and 6 months. One-to-one matching of OA and HV participants was not performed as it was felt that the HV population were likely to be more homogeneous in terms of their cartilage measurements. Fifteen of these participants

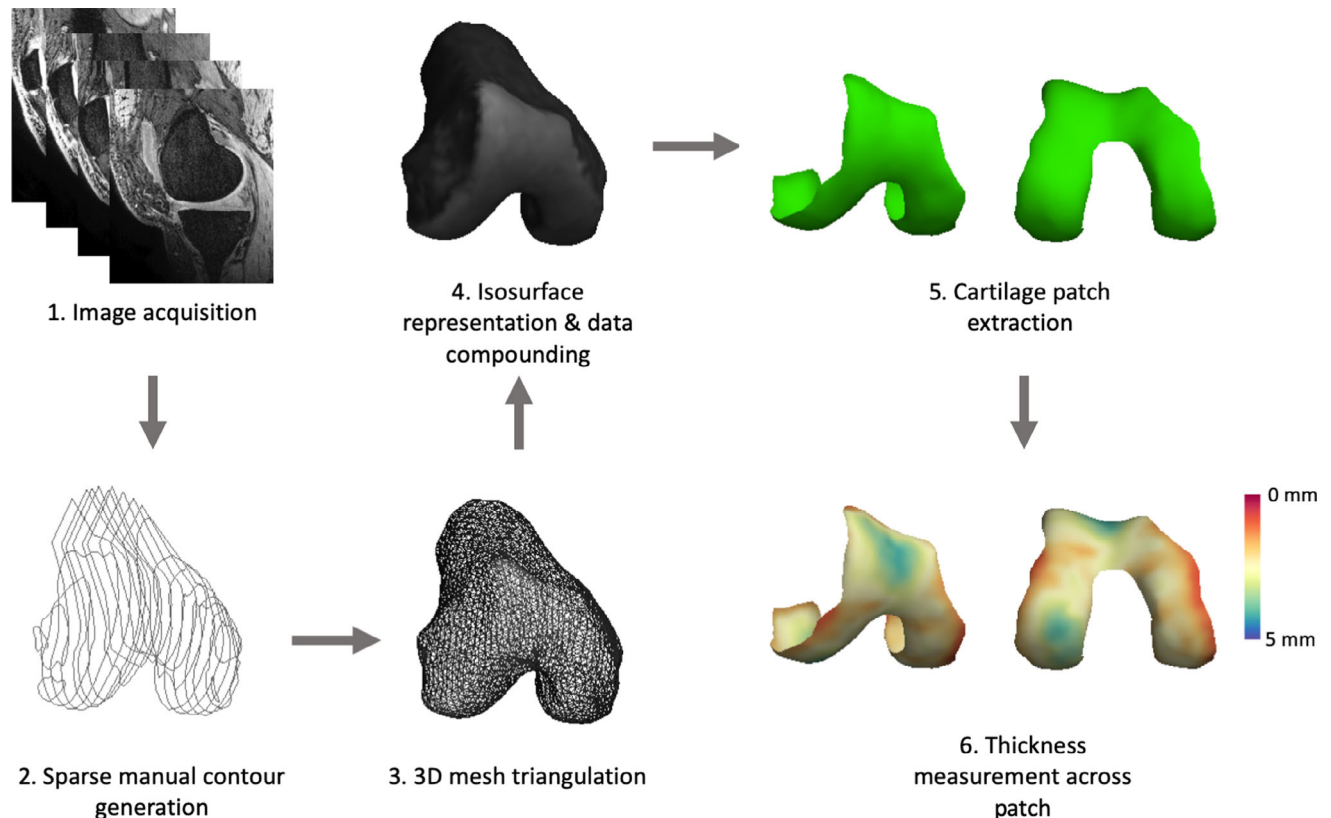


FIGURE 1: Outline of initial steps in 3D-CaSM pipeline, from image acquisition to thickness measurement. The surface generated in steps 2–4 is only used as a guide for cartilage location, and therefore does not need to be highly accurately segmented. Step 6 results in the generation of accurate inner and outer cartilage surfaces between which the thickness measurements were made. Femur used for demonstration purposes; 3D-CaSM was also performed for tibial cartilage surfaces.

(nine OA, six HV) were also imaged at 1 month for assessment of test–retest repeatability. The main inclusion criteria for both groups were age 40–60 years, body mass index ≤ 35 kg/m², and the imaged knee able to fit within the knee MRI coil (diameter ~18 cm). Additional inclusion criteria for OA participants were a clinical diagnosis of OA per American College of Rheumatology criteria and medial compartment predominant disease with a Kellgren–Lawrence grade of 2–3 as assessed on a posteroanterior fixed flexion knee radiograph using a positioning device (SynaFlexer; BioClinica, Newtown, PA).^{17–19} Main exclusion criteria for both groups were a history of previous ipsilateral lower limb fracture or surgery (including arthroscopy), history of significant soft-tissue knee injury (defined as being unable to walk normally for more than 1 week), history of metabolic bone disease or inflammatory arthritis or contraindication to MRI. We excluded any MRI dataset that demonstrated excessive artifact precluding quantitative assessment as judged by a musculoskeletal radiologist. The most common reasons for exclusion were phase wrap (most commonly in the medial–lateral direction for 3D volumetric acquisitions) and excessive motion artifact. At each study visit participants completed the Knee Injury and Osteoarthritis Outcome Score (KOOS) to assess symptoms and had their body mass index (BMI) recorded.²⁰

Clinical Study: Image Acquisition

All participants underwent MRI on the same 3T platform and using the same knee coil as used for cadaveric imaging. The symptomatic knee was imaged in OA participants. For HV participants a knee was selected using a random number generator. The same 3D SPGR sequence as used for cadaveric knees with identical parameters was also used for assessment of cartilage morphology. We also performed quantitative T₁rho (longitudinal relaxation in the presence of a radiofrequency field) and T₂ (transversal relaxation) relaxation time mapping using T₁rho/T₂ magnetization prepared pseudosteady-state 3D fast spin echo (FSE) sequences, with imaging after a period of unloading of at least 45 minutes.²¹ We also performed delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) by administering intravenous gadolinium-based contrast agent (GBCA; Dotarem; Guerbet, Paris, France) at a dose of 0.2 mmol/kg at the end of the first imaging session. After leaving the scanner room, participants cycled for 10 minutes on a stationary cycle to promote GBCA penetration into the knee joint followed by an 80-minute rest period to allow distribution of GBCA within the cartilage. Participants then returned for a second imaging session consisting of quantitative T₁ (longitudinal relaxation) relaxation time mapping using a variable flip angle approach. Sequence parameters for all MRI sequences are provided in Table 1. A detailed description of the correlation between these relaxation time methods with cartilage health/disease states is outside the scope of this article and can be found elsewhere.^{22,23} In brief, T₁rho relaxation time is sensitive to alterations in the proteoglycan content, collagen orientation, and water content of cartilage, T₂ relaxation time is sensitive to alterations in collagen orientation and water content of cartilage, and dGEMRIC is sensitive to alterations in glycosaminoglycan content, although the reported correlation between these methods and individual tissue components is variable.^{24–26}

TABLE 1. MRI Pulse Sequence Parameters

Purpose	Sequence	TR/TE (msec)	Flip angle (°)	RBW (kHz)	Acquired matrix (pixels)	FOV (mm)	ST (mm)	NEX	Acquisition time (min:s)
Morphology	3D SPGR FS	26/6.8	25	11.9	512 × 380 ^b	150	1	0.5 ^c	06:50
T ₁ rho	3D FSE FS CubeQuant	1500/1,10,20,35 ^a	90	62.5	320 × 256 ^b	160	3	0.5 ^c	05:00
T ₂	3D FSE FS CubeQuant	1500/7,13,27,40	90	62.5	320 × 256 ^b	160	3	0.5 ^c	05:00
dGEMRIC	3D SPGR	5/1.7	2,6,14	62.5	320 × 256 ^b	160	3	3	03:30

TR: repetition time; TE: echo time; RBW: receiver bandwidth; FOV: field of view; ST: slice thickness; NEX: number of excitations (signal averages); SPGR: spoiled gradient echo; FS: fat-saturated; FSE: fast spin echo.
^aValues given are spin lock time (TSL) values rather than TE values. Spin lock frequency was 500 Hz.
^bInterpolated to 512 × 512 with zero filling.
^cHalf-Fourier using phase-conjugate symmetry.

Clinical Study: Image Analysis

Thickness measurement with 3D-CaSM was done using the 3D SPGR sequence following the process described for the validation study. For analysis of cartilage relaxation times, parameter maps for $T_1\rho$ and T_2 relaxation times and dGEMRIC were created from source images registered to the 3D SPGR sequence. Rigid registration of source to 3D SPGR images was performed using the elastix registration tool without interpolation of the source data.²⁷ The femur and tibia were registered separately using masking to allow for different degrees of knee flexion between sequences and improve registration accuracy, as described previously.²⁸ Parameter maps were then generated for each relaxation time measurement by fitting the observed signal for each pixel to the appropriate equation (Supplementary Table 1). Goodness-of-fit statistics were extracted for each pixel, and pixels with poor fits (operationally defined as $R^2 < 0.8$) or implausible values (Supplementary Table 1) were excluded from subsequent analysis. In practice, very few (<1%) cartilage voxels were excluded.

Inner and outer cartilage surfaces generated by the thickness measurement process were imported into the registered parameter maps. Relaxation time parameters at each vertex on the cartilage surface were then measured by sampling the data along the surface normal connecting each vertex on the inner surface to its correspondent on the outer surface and taking the mean value (Supplementary Fig. 2).

Cartilage surfaces from each timepoint from each participant were spatially normalized to a canonical (template) surface using a combined similarity and thin-plate-spline transformation. All surface data from all participants could then be mapped to the same template surface to facilitate further spatially-corresponded analysis.

Two independent observers (J.M. and T.T., a musculoskeletal radiologist with 10 years' experience) performed 3D-CaSM on 10 randomly selected knees for assessment of interobserver reproducibility. Both observers also performed manual segmentation of the femoral and tibial cartilage to allow comparison of interobserver reproducibility between 3D-CaSM and manual segmentation. Manual thickness values were obtained as described for the validation study.

Clinical Study: Statistical Analysis

In the presentation of 3D-CaSM results, we draw the distinction between *vertexwise* analyses where the values from all surface vertices (total ~6000 per participant) are used, and *surfacewise* analyses where a surface-averaged value is used.

For analysis of interobserver reproducibility, we performed surfacewise and vertexwise comparison between the two observers' measurements and calculated root mean square coefficients of variation (RMSCVs) for each cartilage surface. The RMSCV is calculated as:

$$RMSCV = \sqrt{\frac{\sum_{i=1}^n \left(\frac{\sigma_i}{\mu_i}\right)^2}{n}} \quad (1)$$

where σ_i is the standard deviation of two measurements for subject i , and μ_i is the mean of the two measurements.

For analysis of test-retest repeatability, we calculated RMSCVs for surfacewise and vertexwise differences between

baseline and 1-month data. We determined vertexwise smallest detectable differences (SDD) for each surface-parameter combination via Bland-Altman analysis of test-retest data, taking the Bland-Altman RC as the SDD (Supplementary Material). This value represents the magnitude of change over time at a single vertex that has a less than 5% chance of being due to measurement error alone.²⁹

For assessment of 6-month changes, we performed vertexwise comparison between baseline and 6-month data. Because these data were mapped to the same canonical surface, we could create individual "change surfaces" for each parameter at each timepoint. We then defined regions of significant change by applying magnitude (change greater than SDD calculated from the test-retest repeatability data for that surface/parameter combination) and area (operationally defined as occurring across a cluster covering at least 1% of the cartilage surface) thresholds to these change surfaces. The percentage of each surface with areas of significant change was calculated and used to define three summary metrics: %SC_{pos}, %SC_{neg}, %SC_{total}, defined respectively as the percentage of the surface with areas of positive significant change (ie, parameter increase), negative significant change (parameter decrease), and any significant change (regardless of sign). These metrics were compared between OA and HV groups using descriptive statistics (medians and interquartile ranges [IQR]).

As well as analyzing surface-based changes at an individual level, the fact that all data were registered to a canonical surface permits group-averaged analysis of vertexwise 6-month changes and performance of statistical parametric mapping (SPM) to determine the statistical significance of these changes. SPM was performed using SurfStat (<http://www.math.mcgill.ca/keith/surfstat/>), a MatLab (MathWorks, Natick, MA) toolbox for the statistical analysis of surface data using linear models and random field theory. Linear mixed models were constructed with timepoint as a fixed effect and subject as a random effect. We considered a type 1 error rate of 10% ($P < 0.1$) acceptable given the exploratory nature of this analysis.

Statistical analyses other than SPM were performed in R v. 3.6.1.³⁰

Results

Validation Study

Data from the four cadaveric specimens gave a total of 17,335 surface vertices for comparison between MRI and HRpQCT. 3D-CaSM MRI thickness measurements demonstrated a small positive mean bias (ie, a small systematic overestimation of cartilage thickness) when compared to HRpQCT data, whereas manual MRI thickness measurements demonstrated a small systematic underestimation (Table 2). RCs were similar for both methods and were sub-millimeter in all cases. The spatial distribution of MRI errors was fairly uniform across the cartilage surfaces for both methods with the exception of higher errors for 3D-CaSM at the extreme medial aspect of the lateral tibial plateau, possibly due to residual undissected ligamentous or meniscal tissue confounding measurement at this site or partial volume effects (Fig. 2).

TABLE 2. Results of Bland–Altman Analyses for Comparison of MRI (3D-CaSM and Manual Segmentation) and HRpQCT Thickness Values

Surface	Method	Mean bias MRI – HRpQCT (mm)	95% LOA (mm)	RC (mm)
Femur	3D-CaSM	0.05	−0.40, 0.50	0.45
	Manual	−0.1	−0.58, 0.38	0.48
Medial tibia	3D-CaSM	0.15	−0.21, 0.52	0.37
	Manual	−0.17	−0.53, 0.19	0.36
Lateral tibia	3D-CaSM	0.11	−0.64, 0.85	0.74
	Manual	−0.25	−0.95, 0.45	0.70
All (combined)	3D-CaSM	0.06	−0.43, 0.56	0.50
	Manual	−0.13	−0.64, 0.38	0.51

HRpQCT: high-resolution peripheral quantitative computed tomography; LOA: limits of agreement; RC: repeatability coefficient.

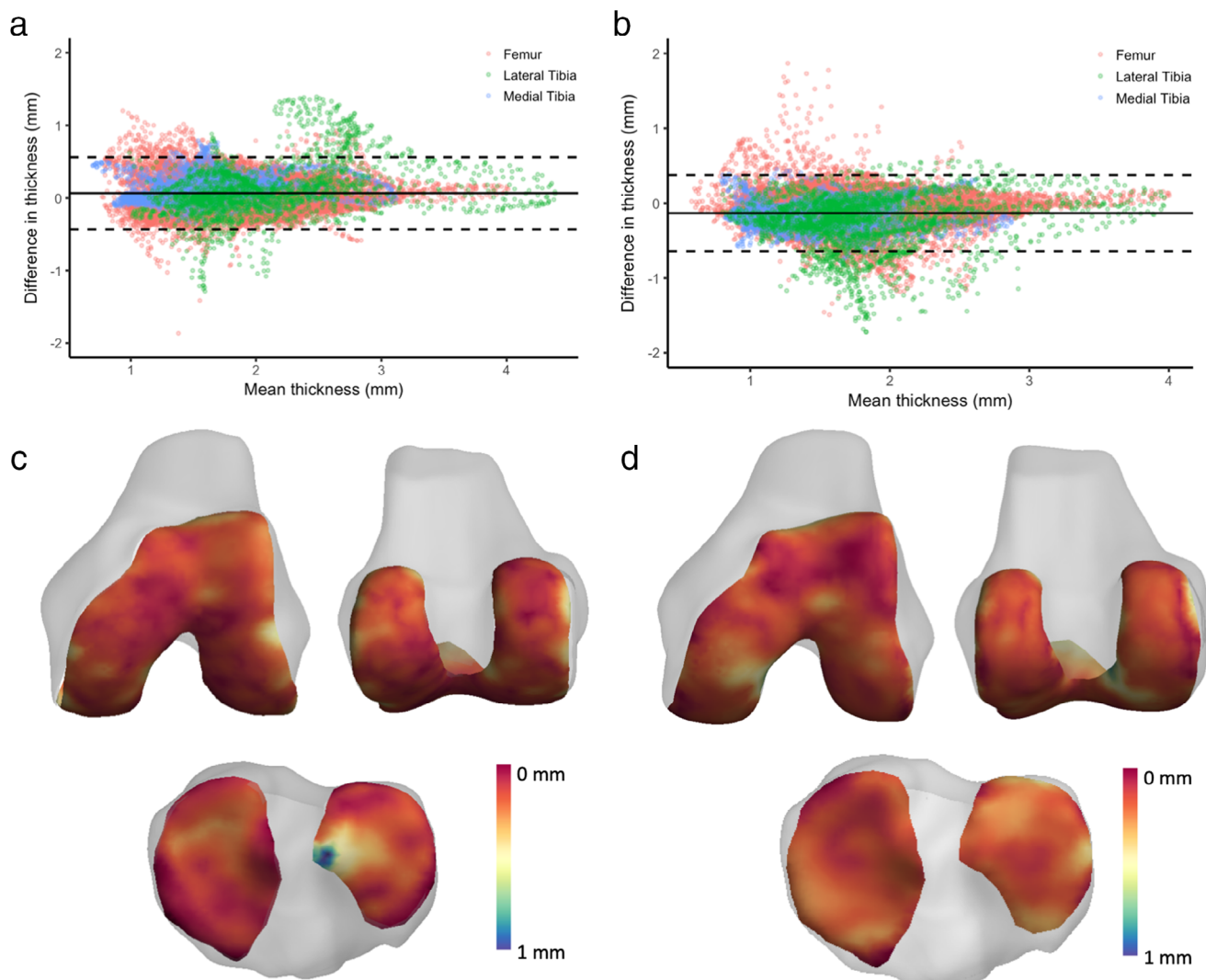


FIGURE 2: Bland–Altman plots assessing agreement between (a) 3D-CaSM and (b) manual segmentation with HRpQCT-derived thickness measurements. Spatial distribution of average error for all cadaver knees is displayed on canonical surfaces for (c) 3D-CaSM and (d) manual segmentation.

TABLE 3. Participant Characteristics at Baseline

	OA participants (<i>n</i> = 14)	HV participants (<i>n</i> = 6)
Age (years) ^a	51 (49–56)	54 (54–57)
Sex (M:F)	8:6	2:4
BMI (kg/m ²) ^a	29.3 (25.7–32.8)	29.3 (27.9–30.9)
Kellgren–Lawrence grade (2:3)	10:4	N/A
KOOS pain ^a	54 (43–80)	100 (98–100)
KOOS symptoms ^a	57 (45–67)	96 (93–100)
KOOS ADL ^a	68 (49–86)	100 (100–100)
KOOS sport & recreation ^a	35 (16–55)	98 (91–100)
KOOS quality of life ^a	25 (19–47)	99 (95–100)

HV: healthy volunteer; BMI: body mass index; KOOS: Knee injury and Osteoarthritis Outcome Score; ADL: activities of daily living.
^aValues are medians (interquartile ranges).

TABLE 4. Interobserver Reproducibility, Test–Retest Repeatability, and Smallest Detectable Difference Data

		Interobserver RMSCV (%)				Test–retest RMSCV (%) 3D-CaSM		Vertexwise SDD
		Surfacewise		Vertexwise				
		Surface	Parameter	3D-CaSM	Manual	3D-CaSM	Manual	
Femur	Thickness	1.6	1.9	8.7	13.7	3.8	11.2	0.62 mm
	T ₁ rho	2.0	1.6	7.7	11.4	4.8	11.0	17.3 msec
	T ₂	2.1	1.0	6.5	12.2	4.3	9.9	12.4 msec
	dGEMRIC	3.1	1.4	9.7	12.0	8.1	10.0	152.3 msec
Medial Tibia	Thickness	5.2	5.1	15.9	14.7	5.2	14.1	0.70 mm
	T ₁ rho	2.7	7.7	7.3	23.2	6.9	12.6	21.0 msec
	T ₂	3.6	8.4	8.2	28.4	6.7	11.7	14.5 msec
	dGEMRIC	1.3	3.2	4.2	24.0	7.8	7.7	124.0 msec
Lateral Tibia	Thickness	3.9	4.5	12.9	13.4	3.9	10.4	0.61 mm
	T ₁ rho	5.2	1.4	10.1	14.4	7.7	12.7	18.2 msec
	T ₂	4.4	2.6	9.0	15.0	6.0	10.9	12.5 msec
	dGEMRIC	2.2	2.7	4.9	14.7	8.6	8.7	133.6 msec

SDD: smallest detectable difference; RMSCV: root-mean-square coefficient of variation.

Clinical Study

Clinical study participant characteristics at baseline are provided in Table 3. Two participants did not attend their scheduled 6-month follow-up visit (one OA participant and

one HV) and were excluded from 6-month change analyses. All 18 participants who completed baseline and 6-month visits had analyzable datasets for thickness data. Seventeen out of these 18 participants had analyzable datasets at both

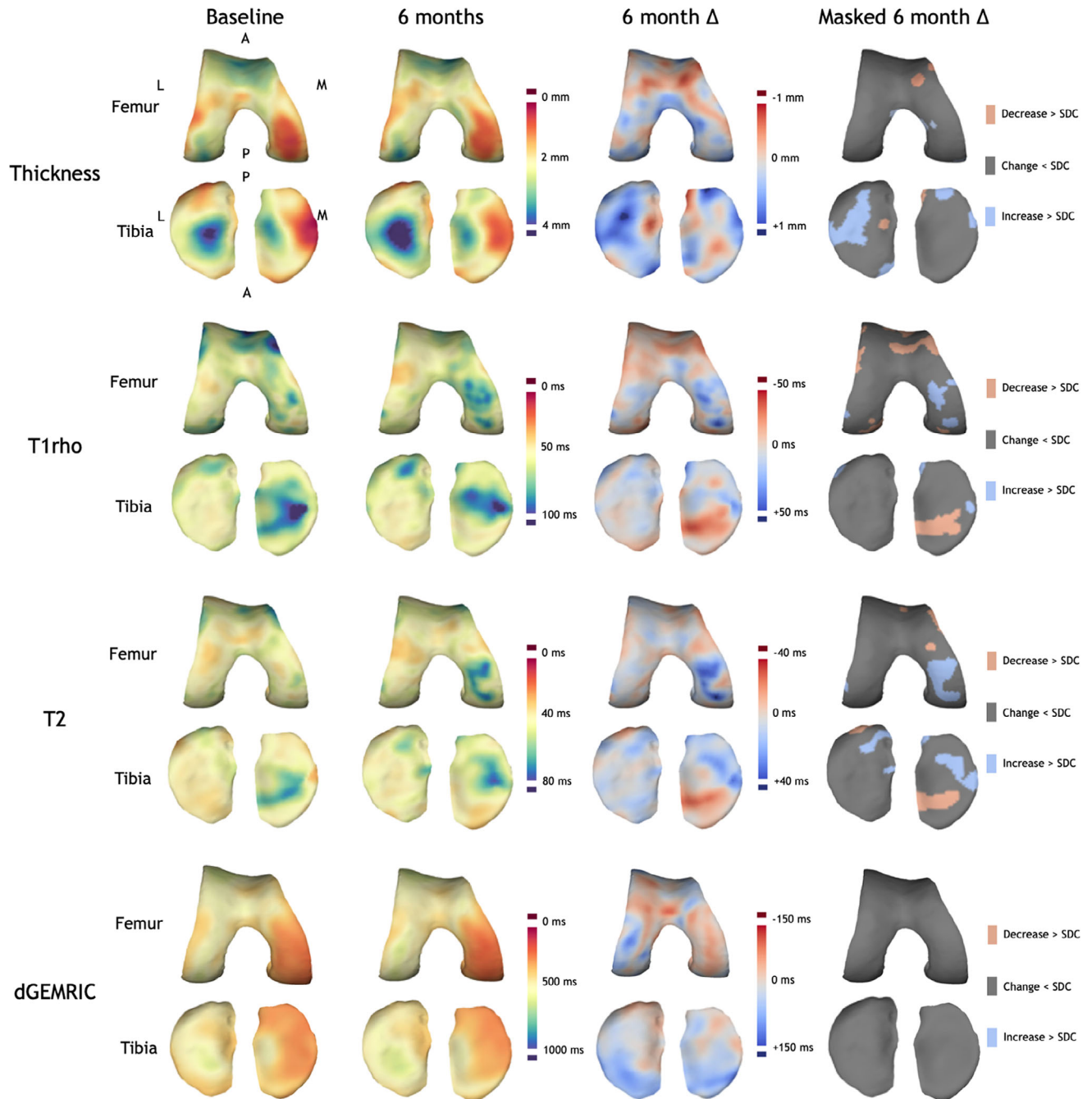


FIGURE 3: Baseline and 6-month follow-up thickness and relaxation time data for a single OA participant displayed on canonical femoral and tibial surfaces. Note the spatial heterogeneity of changes and the co-occurrence of both significant positive and negative changes in thickness and T₁rho/T₂ relaxation times.

visits for T₁rho relaxation time and 16 had analyzable datasets at both visits for T₂ relaxation time. A total of 16 participants (10 OA/6 HV) underwent dGEMRIC imaging at baseline; 13 of these individuals had analyzable datasets at both baseline and 6 months.

Surfacewise interobserver RMSCVs were similar for measurements performed using 3D-CaSM and those performed using manual segmentation. For example, femoral surfacewise interobserver RMSCVs for 3D-CaSM/manual segmentation were 1.6/1.9% for thickness, 2.0/1.6% for

T₁rho relaxation time, 2.1/1.0% for T₂ relaxation time, and 3.1/1.4% for dGEMRIC. However, vertexwise interobserver RMSCVs were lower for 3D-CaSM for all but one measurement (medial tibial thickness) than for manual segmentation (Table 4). For example, femoral vertexwise RMSCVs for 3D-CaSM/manual segmentation were 8.7/13.7% for thickness, 7.7/11.4% for T₁rho relaxation time, 6.5/12.2% for T₂ relaxation time, and 9.7/12.0% for dGEMRIC.

3D-CaSM test–retest repeatability data and SDD values are provided in Table 4. Thickness measurements had slightly

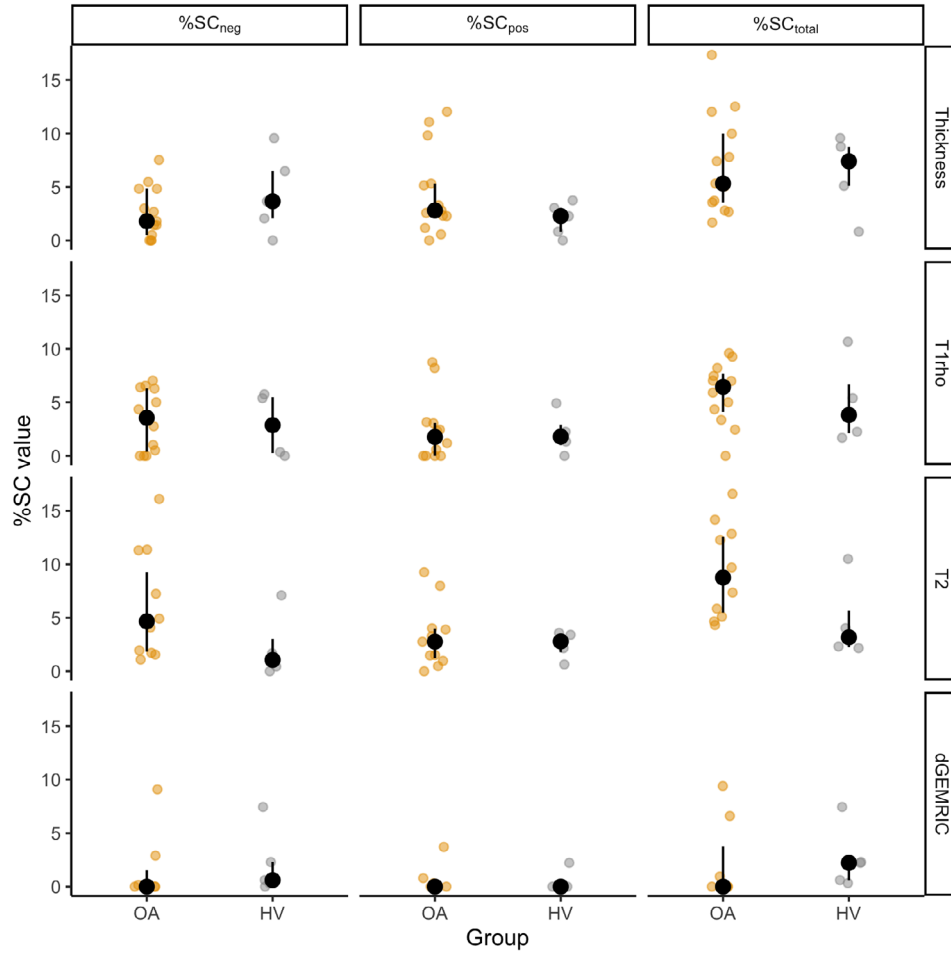


FIGURE 4: Six-month $\%SC_{neg}$, $\%SC_{pos}$, and $\%SC_{total}$ values for each parameter. Small colored dots represent individual participant values. Black dots represent median values for each group, with IQR error bars. Values are averaged across all surfaces analyzed (femur, medial, and lateral tibia) for ease of display.

better surfacewise repeatability (range 3.8–5.2%) than relaxation time measurements (range 4.3–8.6%) across all surfaces, but vertexwise repeatability was similar for all parameters (thickness range 10.4–14.1%, relaxation time range 7.7–12.7%).

Vertexwise analysis of 6-month changes revealed substantial spatial heterogeneity between participants, with areas of concurrent increase and decrease (eg, cartilage thickening and thinning) visible in some participants (Fig. 3). The parameter with the highest $\%SC_{total}$ value at 6 months in OA participants was T_2 , with median (IQR) $\%SC_{total}$ of 8.8 (5.5, 12.6); that is, significant changes were observed on average across about 9% of the total cartilage surface area in OA participants.

Plots of $\%SC_{pos}$, $\%SC_{neg}$, and $\%SC_{total}$ values for each parameter are displayed in Fig. 4. The median $\%SC_{total}$ was higher in OA participants than HV participants for $T_1\rho$ and T_2 relaxation times but lower for thickness and dGEMRIC (Table 5).

Analysis of 6-month change data was also performed using a surfacewise rather than a vertexwise approach to

enable comparison of the sensitivity to change of surfacewise and vertexwise approaches. Only one individual demonstrated significant surfacewise changes (greater than the calculated surfacewise SDD) in cartilage thickness, with no detectable changes in cartilage relaxation time parameters.

TABLE 5. Summary $\%SC_{total}$ Values for Each Parameter by Group; n is the Number of Participants With Analyzable 6-Month Change Data for Each Parameter

Parameter	Median (IQR) $\%SC_{total}$			
	OA	n	HV	n
Thickness	5.3 (3.5, 10.0)	13	7.4 (5.1, 8.8)	5
$T_1\rho$	6.5 (4.1, 7.7)	12	3.8 (2.1, 6.7)	4
T_2	8.8 (5.5, 12.6)	11	3.2 (2.2, 5.7)	4
dGEMRIC	0 (0, 3.8)	7	2.2 (0.6, 2.3)	5

HV: healthy volunteer.

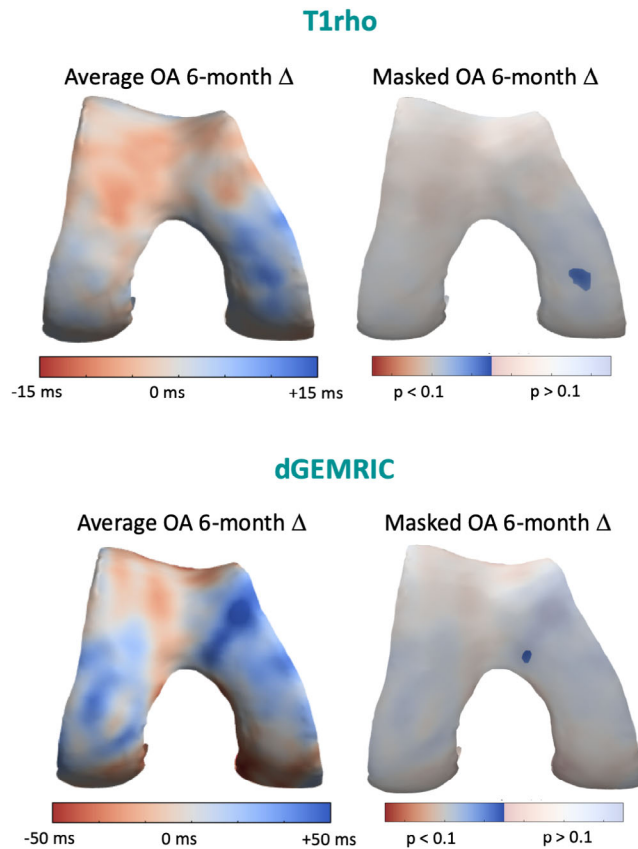


FIGURE 5: Average OA group 6-month change in femoral $T_{1\rho}$ and dGEMRIC. Focal regions of statistical significance ($P < 0.1$) are demonstrated at the central medial femoral condyle ($T_{1\rho}$) and close to the medial femoral sulcus (dGEMRIC). Scale for masked surface is the same as for the unmasked, but with areas without statistical significance washed out.

SPM revealed significant group-averaged 6-month changes in $T_{1\rho}$ and dGEMRIC in the femur in OA participants. A significant focal increase in $T_{1\rho}$ was demonstrated in the central weight-bearing medial femoral condyle and a significant focal increase in dGEMRIC was demonstrated close to the medial femoral sulcus (Fig. 5). No other statistically significant group-averaged changes were detected by SPM.

Discussion

Bias and precision compared to reference HRpQCT data and interobserver agreement of 3D-CaSM were comparable to manual segmentation, the current standard. However, 3D-CaSM has the added benefits of reducing the time taken for segmentation (5 minutes vs. 3–4 hours for manual segmentation) and the possibility of performing surface-based analysis.

3D-CaSM could detect significant changes in cartilage morphology and relaxation times over a 6-month period in this study and highlighted the spatially heterogeneous and bidirectional nature of changes. Our results suggest that the use of 3D-CaSM and similar pipelines is likely to offer improved responsiveness when compared to approaches that fail to recognize this heterogeneity; for example, approaches

that involve analysis of only a single subregion and assume unidirectional change.

In this study we attempted to select participants representative of a significant unmet need in OA: individuals who are experiencing significant symptoms but are unlikely to be recommended for joint replacement surgery due to their young age. This also matches the demographic that is likely to be of interest in therapeutic trials.³¹

Interpretation of Results

Validation study results demonstrate a tendency for 3D-CaSM to overestimate cartilage thickness by a small amount, and for manual segmentation to underestimate thickness by a small amount. This phenomenon has been described previously with other cartilage segmentation methods.^{6,32} Interobserver reproducibility of 3D-CaSM is similar to expert manual segmentation when values are averaged across the entire cartilage surface. However, when measurement at each individual vertex is considered, reproducibility is better for 3D-CaSM for all but one surface/parameter combination. Test–retest repeatability data are also commensurate with those reported for alternative methods, where pooled (surfacewise) repeatability coefficients of variation of between 1.8–3.0%, 2.3–6.1%, 2.3–6.5%, and 4.2–7.4% for cartilage thickness, $T_{1\rho}$ relaxation time, T_2 relaxation time, and dGEMRIC, respectively, have been reported.^{23,33} Vertexwise RMSCVs for both interobserver reproducibility and test–retest repeatability were in all cases higher than the corresponding surfacewise value. This reflects the trade-off between granularity of analysis and precision error, as would be expected based on previous work.¹³ The finding of 6-month changes exceeding the SDD threshold suggests that the magnitude of expected biological change over this interval exceeds test–retest error.

The advantage of the vertexwise analysis performed as part of 3D-CaSM is demonstrated in the analysis of 6-month change data in this study. Areas of significant change are spatially heterogeneous between participants. Therefore, metrics such as %SC, which are location agnostic, may be more useful than group-averaged (SPM) analyses, which assume that regions of change are spatially consistent between individuals. However, this makes the implicit assumption that change in one location is as meaningful as change in any other, which may not be the case. Moreover, the predictive and concurrent validity of metrics such as %SC is unknown, in contrast to conventional manual segmentation/ROI-based approaches, for which a large amount of literature exists.³⁴

3D-CaSM detected fewer regions of significant change in cartilage relaxation time with dGEMRIC compared to $T_{1\rho}$ and T_2 relaxometry. Possible reasons for this include the fact that the data were obtained during a second session after a delay leading to increased variability, or the fact that dGEMRIC may just be a poorer technique in vivo.³⁵

Interpretation in the Context of Previous Studies

The results of 3D-CaSM in this study should be interpreted in the context of the advantages it confers when compared to previously described similar approaches. Methods that aim to reflect/exploit the 3D spatial distribution of cartilage parameters include ACRAAC (anatomically corresponded regional analysis of cartilage), cluster analysis, and voxel-based relaxometry.^{36–38} 3D-CaSM has some potential advantages over these previously described methods. First, it can simultaneously analyze relaxation time (“compositional”) and morphological data, although it should be noted that this should be possible with any of these methods in theory. This may be of particular interest in interventional studies; for example, determining whether an increase in cartilage thickness represents genuine cartilage regeneration or just cartilage swelling, which may in fact represent disease worsening.^{2,39} The output of 3D-CaSM can also easily be combined with data obtained from CT via CBM and JSM due to the identical formats of the output, permitting easy multimodal analysis. Again, multimodal analysis should be possible with the other methods but may not be as straightforward.

While technically an ROI-based method (an approach we have specifically sought to avoid here), the location agnostic “ordered value” approach of Eckstein et al is in some ways conceptually similar to 3D-CaSM.⁴⁰ This involves ordering each subregion of cartilage analyzed according to where the greatest changes have occurred, then using the region of greatest change (regardless of location) to compare between groups or associate with an outcome. Our %SC metric takes the “ordered value” approach to its logical conclusion by applying a similar concept but removing the need for arbitrary subregion definition, which would be expected to improve responsiveness based on previous work.⁷

Limitations

The main limitation of this study is the lack of longitudinal validation of 3D-CaSM and the metrics derived from it (eg, %SC). While the method may offer improved responsiveness to conventional ROI-based methods, it should be borne in mind that changes detected by these conventional methods have been linked to longer-term clinical and radiological outcomes, whereas changes detected by 3D-CaSM have not. Unless there is clinical meaning to the changes detected, the improved sensitivity of 3D-CaSM may not be advantageous for OA research. Implementation of 3D-CaSM in larger cohorts with longer-term follow-up is planned to help establish the clinical meaning of metrics such as %SC.

Another limitation is the lack of histological or biological validation of the findings. Therefore, the suggestion made in preceding paragraphs that 3D-CaSM may be able to depict abnormal “cartilage turnover” should be regarded as somewhat speculative. The %SC metric was defined based on magnitude (>SDD) and areal (>1% of cartilage surface) criteria, aiming to create a stringent threshold for genuine

disease-related change. However, it should be noted that the process of mapping an individual’s data to the canonical surface involves some averaging (smoothing) of the data, meaning that a large change at a single vertex could result in a change covering 1% of the surface. SPM was used to assess the significance of group-averaged changes over 6 months. While this approach is valid and has been used extensively in the neuroimaging community, larger group sizes are usual (hence, the more permissive allowable type 1 error rate of 10% in this exploratory study). Power calculations for SPM are difficult, but simulation studies suggest that sample sizes of 18, 12, 21, and 14 participants would be sufficient to demonstrate significant changes in thickness, $T_1\rho$ relaxation time, T_2 relaxation time, and dGEMRIC, respectively, given the effect sizes seen in this study and assuming a within-subjects repeated measures design with an acceptable type 1 error rate of 5%. However, as mentioned above it should also be noted that SPM may not be the optimal analysis method for 3D-CaSM data, given the spatial heterogeneity of the changes demonstrated. Location agnostic measures such as %SC may provide better responsiveness.

Another limitation of this study is the relatively long test–retest interval (1 month). The measured variability may therefore include contributions from biological changes in cartilage morphology and relaxation times, meaning true methodological variability is likely to be lower.

Finally, while we have speculated on some advantages that 3D-CaSM may confer over alternative automated and semiautomated cartilage analysis pipelines, we have not performed a dedicated head-to-head comparison in this study.

Conclusion

This study demonstrates the construct validity and initial clinical implementation of 3D-CaSM. This analysis pipeline is able to detect and map changes in cartilage morphology and relaxation times over 6 months and demonstrates bias, precision, test–retest repeatability, and interobserver reproducibility comparable to existing gold-standard methods.

Acknowledgments

The study team thanks Dr. Cecilia Brassett, the University Clinical Anatomist, and Maria Wright, senior dissecting room technician, for help in procuring and preparing the cadaveric specimens. We thank Dr. Ken Poole and Michael O’Breasil for their help with arranging access to the HRpQCT scanner at the MRC Elsie Widdowson Laboratory. The Addenbrooke’s Hospital Magnetic Resonance Imaging and Spectroscopy (MRIS) staff are thanked for their help with arranging and conducting the clinical study MRI examinations. We also acknowledge the support of the Addenbrooke’s Charitable Trust and the National Institute for Health Research Cambridge Biomedical Research

Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Contract grant sponsor: Experimental Medicine Initiative PhD Studentship from the University of Cambridge [grant number RG81329]; Contract grant sponsor: GlaxoSmithKline [grant number RG87552].

References

- Hunter DJ. Pharmacologic therapy for osteoarthritis—The era of disease modification. *Nat Rev Rheumatol* 2011;7:13-22.
- Eckstein F, Wirth W, Guermazi A, Maschek S, Aydemir A. Brief report: Intraarticular sprifermin not only increases cartilage thickness, but also reduces cartilage loss: Location-independent post hoc analysis using magnetic resonance imaging. *Arthritis Rheumatol* 2015;67:2916-2922.
- Pedroia V, Su F, Amano K, et al. Analysis of the articular cartilage T1ρ and T2 relaxation times changes after ACL reconstruction in injured and contralateral knees and relationships with bone shape. *J Orthop Res* 2017;35:707-717.
- Joseph GB, McCulloch CE, Nevitt MC, et al. Medial femur T2 Z-scores predict the probability of knee structural worsening over 4-8 years: Data from the osteoarthritis initiative. *J Magn Reson Imaging* 2017;46:1128-1136.
- Bowes MA, Guillard GA, Vincent GR, Brett AD, Wolstenholme CBH, Conaghan PG. Precision, reliability, and responsiveness of a novel automated quantification tool for cartilage thickness: Data from the osteoarthritis initiative. *J Rheumatol* 2020;47:282-289.
- Dam EB, Lillholm M, Marques J, Nielsen M. Automatic segmentation of high- and low-field knee MRIs using knee image quantification with data from the osteoarthritis initiative. *J Med Imaging* 2015;2:024001.
- Jørgensen DR, Lillholm M, Genant HK, Dam EB. On subregional analysis of cartilage loss from knee MRI. *Cartilage* 2013;4:121-130.
- Eckstein F, Maschek S, Wirth W, et al. One year change of knee cartilage morphology in the first release of participants from the osteoarthritis initiative progression subcohort: Association with sex, body mass index, symptoms and radiographic osteoarthritis status. *Ann Rheum Dis* 2009;68:674-679.
- Hunter DJ, Bowes MA, Eaton CB, et al. Can cartilage loss be detected in knee osteoarthritis (OA) patients with 3-6 months' observation using advanced image analysis of 3T MRI? *Osteoarthritis Cartil* 2010;18:677-683.
- Littman BH, Williams SA. The ultimate model organism: progress in experimental medicine. *Nat Rev Drug Discov* 2005;4:631-638.
- Poole KES, Treece GM, Mayhew PM, et al. Cortical thickness mapping to identify focal osteoporosis in patients with hip fracture. *PLoS One* 2012;7:e38466.
- Treece G, Gee A. Cortical bone mapping: Measurement and statistical analysis of localised skeletal changes. *Curr Osteoporos Rep* 2018;16:617-625.
- Turmezei TD, Treece GM, Gee AH, Houlden R, Poole KES. A new quantitative 3D approach to imaging of structural joint disease. *Sci Rep* 2018;8:9280.
- Treece GM, Gee AH, Mayhew PM, Poole KES. High resolution cortical bone thickness measurement from clinical CT data. *Med Image Anal* 2010;14:276-290.
- Treece GM, Gee AH. Independent measurement of femoral cortical thickness and cortical bone density using clinical CT. *Med Image Anal* 2015;20:249-264.
- Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: A review of statistical methods for technical performance assessment. *Stat Methods Med Res* 2015;24:27-67.
- Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis* 1957;16:494-502.
- Kothari M, Guermazi A, von Ingersleben G, et al. Fixed-flexion radiography of the knee provides reproducible joint space width measurements in osteoarthritis. *Eur Radiol* 2004;14:1568-1573.
- Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 1986;29:1039-1049.
- Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee injury and osteoarthritis outcome score (KOOS)—Development of a self-administered outcome measure. *J Orthop Sports Phys Ther* 1998;28:88-96.
- Chen W, Takahashi A, Han, E. 3D quantitative imaging of T1ρ and T2. In: *Proc 19th Annual Meeting ISMRM, Montreal*; 2011. p 231.
- Guermazi A, Alizai H, Crema MD, Trattnig S, Regatte RR, Roemer FW. Compositional MRI techniques for evaluation of cartilage degeneration in osteoarthritis. *Osteoarthritis Cartil* 2015;23:1639-1653.
- MacKay JW, Low SBL, Smith TO, Toms AP, McCaskie AW, Gilbert FJ. Systematic review and meta-analysis of the reliability and discriminative validity of cartilage compositional MRI in knee osteoarthritis. *Osteoarthritis Cartil* 2018;26:1140-1152.
- Akella SVS, Reddy Regatte R, Gougoutas AJ, et al. Proteoglycan-induced changes in T1ρ-relaxation of articular cartilage at 4T. *Magn Reson Med* 2001;46:419-423.
- Bashir A, Gray ML, Hartke J, Burstein D. Nondestructive imaging of human cartilage glycosaminoglycan concentration by MRI. *Magn Reson Med* 1999;41:857-865.
- Lüsse S, Claassen H, Gehrke T, et al. Evaluation of water content by spatially resolved transverse relaxation times of human articular cartilage. *Magn Reson Imaging* 2000;18:423-430.
- Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: A toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* 2010;29:196-205.
- Bron EE, van Tiel J, Smit H, et al. Image registration improves human knee cartilage T1 mapping with delayed gadolinium-enhanced MRI of cartilage (dGEMRIC). *Eur Radiol* 2013;23:246-252.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-310.
- R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019.
- Watt FE, Gulati M. New drug treatments for osteoarthritis: What is on the horizon? *Eur Med J Rheumatol* 2017;2:50-58.
- Norman B, Pedroia V, Majumdar S. Use of 2D U-net convolutional neural networks for automated cartilage and meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology* 2018;288:177-185.
- Eckstein F, Kunz M, Schutler M, et al. Two year longitudinal change and test-retest-precision of knee cartilage morphology in a pilot study for the osteoarthritis initiative. *Osteoarthritis Cartil* 2007;15:1326-1332.
- Eckstein F, Collins JE, Nevitt MC, et al. Brief report: Cartilage thickness change as an imaging biomarker of knee osteoarthritis progression: Data from the Foundation for the National Institutes of Health osteoarthritis biomarkers consortium. *Arthritis Rheumatol* 2015;67:3184-3189.
- Burstein D, Velyvis J, Scott KT, et al. Protocol issues for delayed Gd(DTPA)2-enhanced MRI (dGEMRIC) for clinical evaluation of articular cartilage. *Magn Reson Med* 2001;45:36-41.
- Williams TG, Holmes AP, Bowes M, et al. Measurement and visualisation of focal cartilage thickness change by MRI in a study of knee osteoarthritis using a novel image analysis tool. *Br J Radiol* 2010;83:940-948.

37. Monu UD, Jordan CD, Samuelson BL, Hargreaves BA, Gold GE, McWalter EJ. Cluster analysis of quantitative MRI T2 and T1 ρ relaxation times of cartilage identifies differences between healthy and ACL-injured individuals at 3T. *Osteoarthr Cartil* 2017;25:513-520.
38. Pedoia V, Li X, Su F, Calixto N, Majumdar S. Fully automatic analysis of the knee articular cartilage T1 ρ relaxation time using voxel-based relaxometry. *J Magn Reson Imaging* 2016;43:970-980.
39. Cotofana S, Buck R, Wirth W, et al. Cartilage thickening in early radiographic knee osteoarthritis: A within-person, between-knee comparison. *Arthritis Care Res* 2012;64:1681-1690.
40. Eckstein F, Buck R, Wirth W. Location-independent analysis of structural progression of osteoarthritis—Taking it all apart, and putting the puzzle back together makes the difference. *Semin Arthritis Rheum* 2017;46:404-410.