Cartilage $T_1p$ and $T_2$ relaxation times: longitudinal reproducibility and variations using different coils, MR systems and sites

X. Li †, V. Pedoia ‡, D. Kumar † ‡, J. Rivoire †, C. Wyatt †, D. Lansdown †, K. Amano †, N. Okazaki †, D. Savic †, M.F. Koff †, J. Felmlee ‡, S.L. Williams ‡, S. Majumdar †

† University of California, San Francisco, CA, USA
‡ Hospital for Special Surgery, New York, NY, USA
§ Mayo Clinic, Rochester, MN, USA
∥ Division of Physical Therapy, Medical University of South Carolina, Charleston, SC, USA

ARTICLE INFO

Article history:
Received 30 October 2014
Accepted 6 July 2015

Keywords:
Cartilage
Quantitative MRI
$T_1p$
$T_2$
Reproducibility
Multi-site study

SUMMARY

Objective: To evaluate the longitudinal reproducibility and variations of cartilage $T_1p$ and $T_2$ measurements using different coils, MR systems and sites.

Methods: Single-Site study: Phantom data were collected monthly for up to 29 months on four GE 3T MR systems. Data from phantoms and human subjects were collected on two MR systems using the same model of coil; and were collected on one MR system using two models of coils. Multi-site study: Three participating sites used the same model of MR systems and coils, and identical imaging protocols. Phantom data were collected monthly. Human subjects were scanned and rescanned on the same day at each site. Two traveling human subjects were scanned at all three sites.

Results: Single-Site Study: The phantom longitudinal RMS-CVs ranged from 1.8% to 2.7% for $T_1p$ and 1.8–2.8% for $T_2$. Significant differences were found in $T_1p$ and $T_2$ values using different MR systems and coils. Multi-Site Study: The phantom longitudinal RMS-CVs ranged from 1.3% to 2.6% for $T_1p$ and 1.2–2.7% for $T_2$. Across three sites ($n = 16$), the in vivo scan-rescan RMS-CV was 3.1% and 4.0% for $T_1p$ and $T_2$, respectively. Phantom $T_1p$ and $T_2$ values were significantly different between three sites but highly correlated ($R > 0.99$). No significant difference was found in $T_1p$ and $T_2$ values of traveling controls, with cross-site RMS-CV as 4.9% and 4.4% for $T_1p$ and $T_2$, respectively.

Conclusion: With careful quality control and cross-calibration, quantitative MRI can be readily applied in multi-site studies and clinical trials for evaluating cartilage degeneration.

© 2015 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

Introduction

Osteoarthritis (OA) constitutes a significant health burden affecting more than 27 million people in US alone, and has been recognized as one of the fastest growing medical conditions worldwide due to the increased prevalence of obesity and aging of society. The disease is characterized primarily by cartilage degeneration. MR techniques that quantify cartilage matrix changes have become more accessible, with the rationale that detecting early and subtle cartilage degeneration would be critical for allowing early intervention, monitoring treatment efficacy, and leading to prevention strategies for OA. Among these techniques, $T_1p$ and $T_2$ relaxation time quantification have gained significant attention because they do not need contrast agent injection nor special hardware, and can be feasibly performed in a clinical setting. $T_2$ mapping is a product sequence while $T_1p$ mapping prototype acquisitions are available from all major MR manufacturers. Numerous studies have shown that $T_1p$ and $T_2$ quantification techniques can detect early cartilage damage and degeneration in patients with OA, acute joint injury or cartilage damage.

Despite the promising results, the application of $T_1p$ and $T_2$ quantification in multicenter clinical studies and trials is very limited. One impedying factor is the limited documentation of potential variations of $T_1p$ and $T_2$ by using different MR systems, coils, and sites. Furthermore, longitudinal assessment of cartilage degeneration requires reproducible quantitative measurements over time. Previous studies of $T_1p$ and $T_2$ reproducibility were primarily limited to short-term reproducibility, except for the 3- and 7-day scan-rescan studies.
8-year T2 data as part of the OA Initiative (OAI) study quality control\textsuperscript{17,18}. Understanding and documenting these variations are critical for setting up multi-center longitudinal studies using T1\textsubscript{p} and T2 techniques.

Currently, a multi-center feasibility study of applying T1\textsubscript{p} and T2 quantification techniques in knees after acute ACL injury is being performed at three geographically remote centers. In this report, we first evaluated the longitudinal reproducibility and variations of T1\textsubscript{p} and T2 values using different MR systems and coils at one site (a single-site study), and then evaluated the reproducibility and cross-validation results among three sites (the multi-site study).

**Methods**

**Study design**

The overall study design is illustrated in Fig. 1, and is described in detail below in two sections: the single-site and multi-site study. This study was approved by the Committee for Human Research at all institutions participating in the study, and informed consent was obtained from all subjects prior to data acquisition.

**Single-site study**

The study was designed to evaluate: (1) short and long-term reproducibility of T1\textsubscript{p} and T2 values; (2) the variation of T1\textsubscript{p} and T2 values using different model MR systems from the same vendor; and (3) the variation of T1\textsubscript{p} and T2 values using different coils. All the data were collected between September 2011 and July 2014.

To evaluate short-term and long-term reproducibility, phantoms were scanned monthly using three models of GE 3T MR systems (GE Healthcare, Milwaukee, WI) using knee coils of the same model from the same vendor (quadrature transmit/8-channel phased-array receive knee coil, InVivo, Gainesville, FL, termed as ‘QT8PAR knee coil’ below) at a single institution: GE Signa HDx long bore (maximum gradient strength: 50 mT/m; slew rate 150 mT/m/s; bore size: 60 cm); GE MR750 (maximum gradient strength: 50 mT/m; slew rate 200 mT/m/s; bore size: 60 cm); GE MR750 wide bore (maximum gradient strength: 44 mT/m; slew rate: 200 mT/m/s; bore size: 70 cm).

To evaluate the variation of T1\textsubscript{p} and T2 values using different MR systems, phantom data was collected at four GE 3T MR systems at the same institution: the three MR systems above and a GE Signa HDx short bore (maximum gradient strength: 23 mT/m whole mode, 40 T/m zoom code; slew rate: 80 mT/m/s whole mode, 150 mT/m/s zoom mode; bore size: 60 cm). In addition, 10 healthy subjects were scanned on both the HDx long bore and the MR750 wide bore using the same model of QT8PAR knee coils within a period of 3-months.

To evaluate the variation of T1\textsubscript{p} and T2 values using different coils, five healthy subjects were scanned on the MR750 wide bore using a QT8PAR knee coil and a 16-channel phased-array receive only flex coil (GE Healthcare, termed as ‘16PAR flex coil’ below).

**Multi-site study**

All three sites used GE MR750 systems with QT8PAR knee coils. The study was designed to evaluate: (1) reproducibility of T1\textsubscript{p} and T2 values in phantoms scanned monthly at each site; (2) scan/re-scan (on the same day) reproducibility of T1\textsubscript{p} and T2 values in healthy controls at each site (n = 6, 5, 5 for site 1, 2, 3 respectively); and (3) cross-validation of T1\textsubscript{p} and T2 values in the same phantom sets and in the same volunteers across three sites. For phantom scans, one phantom set was scanned at all three sites at one time.
point. For human subject scans, two volunteers traveled and were scanned at all three sites at baseline and at 10-month follow-up. At all three sites, the same sequence and same imaging protocol was used for both phantom and in vivo scans as detailed below. All of the data were collected between November 2013 and October 2014.

**Imaging protocol**

**Phantom imaging protocol**

Phantoms were created by dissolving agarose powder in deionized water at different concentrations (weight/volume, 2%, 3%, 4%). Six phantom tubes (25 mm diameter, 2 for each concentration) were placed in a foam holder and named Phantoms #1–6. During each exam, phantoms were first scanned at isocenter, then left (70 mm off-center), and right (70 mm off-center) positions. At isocenter, T1 and T2 measurements were acquired separately with eight echoes each. At the left and right positions, T1 and T2 measurements were acquired in a combined sequence with four echoes each.62 (Table I).

**In vivo imaging protocol**

For the single-site study, the in vivo imaging protocol included high-resolution 3D fast spin-echo (FSE) images (CUBE) for cartilage segmentation, and T1 and T2 sequences. T1 and T2 measurements were acquired separately with eight echoes each.

For the multi-site study, a custom leg-holder was used during data acquisition to ensure consistent knee flexion during scanning. The holders for all three sites were made from a common cast mold. The foot was positioned in a U-shaped foam holder (GE Healthcare, Milwaukee, WI) and oriented vertically to minimize any internal or external rotation of the knee joint. For this study, CUBE images were used for cartilage segmentation, the combined T1 and T2 sequences were used with four echoes each, and additional fat-suppressed and non-fat-suppressed 2D FSE images were collected for clinical evaluation of any joint damage (Table I).

**Image analysis**

All images were analyzed at one center under stringent quality control procedures. Acquisition parameters were first automatically checked to ensure consistency of imaging protocols, followed by visual evaluation of image quality (including orientation, coverage and artifacts) before quantitative analysis. Images with significant artifacts were excluded from analysis.

**Table I**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>TR/TE (ms)</th>
<th>ETL</th>
<th>FOV (cm)</th>
<th>Matrix</th>
<th>Slice thickness</th>
<th>NEX</th>
<th>Acceleration</th>
<th>Bandwidth (KHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Single-site Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sagittal fat-saturated 3D FSE (CUBE)</td>
<td>1500/25</td>
<td>50</td>
<td>14</td>
<td>384 x 384</td>
<td>1 mm</td>
<td>0.5</td>
<td>ARC Phase AF = 2</td>
<td>50</td>
</tr>
<tr>
<td>Sagittal T1ρ and T2</td>
<td>7–9/min Full</td>
<td>–</td>
<td>14</td>
<td>256 x 128</td>
<td>4 mm</td>
<td>1</td>
<td>ARC Phase AF = 2</td>
<td>64.5</td>
</tr>
<tr>
<td>Other parameters for T1ρ and T2</td>
<td>Views Per Segment – 64, time of recovery – 1.2 s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SNR calculation**

In phantom scans, a second TSL = 0 image was acquired. The difference between the two TSL = 0 images was used for evaluating noise. SNR was calculated as the mean signal within phantom ROIs/SD of noise. In vivo, this analysis was not possible, instead the noise SD was estimated from the background region outside the knee and below the patella as proposed in reference 22. To limit the center-to-edge variability of SNR caused by a phased array receive coil, the center five slices were used in phantoms and the center four slices for medial and lateral femoral condyles were used in human subjects.

**T1ρ and T2 quantification**

The first step registered all in vivo images to the TLS = 0 image to minimize motion between different echoes. For images acquired with separate T1ρ and T2 sequences (8 TSLs for T1ρ and 8 TEIs for T2), T1ρ and T2 maps were reconstructed by fitting the T1ρ- and T2-weighted images voxel-by-voxel to the equations below (three-parameter fitting):

\[
S(TSL) = A \times \exp(-TSL/T_{1\rho}) + B
\]  
(1)

\[
S(TE) = A \times \exp(-TE/T_{2}) + B
\]  
(2)

For images acquired with combined T1ρ and T2 sequence (4 TSLs for T1ρ and 4 TEIs for T2), T1ρ and T2 maps were reconstructed by fitting the T1ρ- and T2-weighted images voxel-by-voxel to the equations below (two-parameter fitting, because the three-parameter fitting would be suboptimal with only four echoes):

\[
S(TSL) = A \times \exp(-TSL/T_{1\rho})
\]  
(3)

\[
S(TE) = A \times \exp(-TE/T_{2})
\]  
(4)

For phantom images, an automatic program was applied to generate a circular ROI for each phantom in the middle four slices. The mean and standard deviation (SD) of T1ρ and T2 relaxation times were calculated in ROIs for each phantom.

For in vivo data, the high resolution CUBE images were rigidly registered to the TSL = 0 images using the VTK CIGS registration Toolkit. Cartilage was segmented semi-automatically using software developed in-house21 on the registered CUBE images into six compartments: lateral/medial femur (LF/MF), the lateral/medial tibia (LT/MT), trochlea (TrR) and patella (P). The 3D regions of interest (ROIs) were then overlaid on the T1ρ and T2 maps. The mean and SD T1ρ and T2 values were calculated for each compartment.
Statistical analysis

The longitudinal and scan-rescan (short-term) reproducibility of \( T_{1p} \) and \( T_2 \) values were evaluated using root-mean-square coefficients of variation (RMS-CV, %). The fitting errors were evaluated with RMS error normalized to the signal intensity of the TSL = 0 image. The differences of \( T_{1p} \) and \( T_2 \) values obtained at different positions (center vs left vs right) in the magnet, using different MR systems or different coils, and at different sites were evaluated using ANOVA, Bland–Altman, and pooled RMS analyses. Correlations between \( T_{1p} \) and \( T_2 \) values within the subjects as well as correlations of \( T_{1p} \) and \( T_2 \) values between MR systems were evaluated using the Spearman correlation coefficient \( R \).

Results

Single-site study

Long-term reproducibility

Up to 29 months of data were collected from the three MR systems: HDx long bore (13 time points); MR750 (29 time points); MR750 wide bore (20 time points). Figure 2A and B shows the scatter plot of the \( T_{1p} \) and \( T_2 \) at the magnet center position. Table II A summarizes the RMS-CV of \( T_{1p} \) and \( T_2 \) values at the center, left and right positions. Table III summarizes the number of voxels, mean \( T_{1p} \) and \( T_2 \) values, pooled SD and fitting errors within each phantom ROI, as well as the pooled RMS of inter-location variation and long-term reproducibility.

Variations in \( T_{1p} \) and \( T_2 \) values using different MR systems

In phantoms, \( T_{1p} \) and \( T_2 \) values between any two MR systems were highly correlated (\( R > 0.9 \)). However, significant differences were observed in \( T_{1p} \) and \( T_2 \) values between MR systems (Fig. 2C and D), with MR750 having the highest values relative to the other systems. The MR750 wide bore had significantly lower \( T_{1p} \) values than HDx long bore (\( P = 0.02, 95\% \) CI \((-3.1, -0.4)) \). However no significant difference in \( T_2 \) values were found between these two MR systems (\( P = 0.16, 95\% \) CI \((-0.2, 0.9)) \).

The HDx short bore had significantly higher SNR, and the MR750 wide bore had significantly lower SNR, compared to HDx long bore and MR750. For example, 3% agarose phantom images with TSL/TE = 0 had SNR 127.8, 89.9, 80.0 and 61.4 for HDx...
short bore, HDx long bore, MR750, MR750 wide bore, respectively.

The global in vivo cartilage T1p and T2 values were significantly higher using HDx long bore compared to MR750 wide bore (34.3 ± 3.0 ms vs 31.5 ± 2.9 ms, \(P = 0.00003, 95\% \text{ CI } (2.1, 3.5)\) for T1p, and 25.2 ± 2.0 vs 22.3 ± 2.3 ms, \(P = 0.002, 95\% \text{ CI } (1.5, 4.4)\) for T2), (Fig. 3A and B). T1p values between the two MR systems were highly correlated (R = 0.91) while T2 values were less well correlated (R = 0.64). The T1p and T2 values within subjects using the same MR system were moderately correlated (R = 0.55) The in vivo SNR of T1p- and T2-weighted images were significantly higher using the HDx long bore compared to the MR750 wide bore (Table IV), in agreement with phantom results. Variations in T1p and T2 values using different coils

In phantoms, T1p and T2 values were significantly higher using the 16PAR flex coil than those using the QT8PAR knee coil \((P = 0.0009, 95\% \text{ CI } (0.4, 1.5)\) for T1p; \(P = 0.02, 95\% \text{ CI } (0.4, 3.0)\) for T2). The difference in T1p was 0.6 ms, 0.7 ms and 1.6 ms, and the difference in T2 was 0.9 ms, 1.7 ms and 2.6 ms for the 4%, 3% and 2% phantoms respectively. The SNR was significantly higher (the average SNR of TSL = 0/TE = 0 images was 165.7 vs 97.0 for 16-channel and 8-channel coils respectively) while the fitting errors were significantly lower (the average fitting error was 0.0032 vs 0.0054 for T1p, 0.0037 vs 0.0062 for T2 for 16 channel and 8-channel coils respectively) using the 16PAR flex coil than those using the QT8PAR knee coil.

The global in vivo cartilage T1p and T2 values were significantly higher using the 16PAR flex coil than those using the QT8PAR knee coil (32.9 ± 3.9 ms vs 30.1 ± 3.1 ms, \(P = 0.018, 95\% \text{ CI } (0.7, 3.6)\) for T1p, 27.4 ± 1.8 vs 23.4 ± 2.8, \(P = 0.012, 95\% \text{ CI } (1.7, 6.3)\) for T2) (Fig. 4A and B). No significant differences in fitting errors were observed between the two coils. The in vivo SNR of T1p - and T2- weighted images using the 16PAR flex coil were significantly higher compared to using the QT8PAR knee coil (Table IV).

Multi-site study

Longitudinal Phantom reproducibility

Table II summarizes the phantom RMS-CV for T1p and T2 values of Site 1 (7 months), Site 2 (4 months) and Site 3 (8 months).

Scan-rescan reproducibility of healthy controls at each site

Across all three sites (n = 16), the scan-rescan RMS-CV was 3.1% and 4.0% for compartment T1p and T2 values, respectively. The RMS-CV in each compartment ranged from 2.3% to 3.9% for T1p, and ranged 3.2–5.3% for T2 (Fig. 5). Table III summarizes the number of voxels, mean T1p and T2 values, pooled SD and fitting errors within each compartment, as well as the pooled RMS of scan-rescan (short-term) reproducibility.

Cross-validation of T1p and T2 values among three sites

In phantoms. Phantom T1p and T2 values were significantly different among the three sites but highly correlated (R > 0.99). The mean CV was 2.9% and 4.1% for T1p and T2 values respectively.

In healthy controls. No significant differences were found in T1p and T2 values in the traveling controls between the three sites, with 4.9% and 4.4% RMS-CV for T1p and T2, respectively. No significant differences were found in T1p and T2 values between baseline and 10-month follow-up, with 4.4% and 5.1% RMS-CV for T1p and T2, respectively.

Table III summarizes the pooled T1p and T2 RMS for phantoms and human subjects using the different MR systems, different coils, and different sites.

Discussion

Quantitative evaluation of articular cartilage matrix composition using T1p and T2 mapping can potentially provide early markers of cartilage degeneration. These methods, present significant challenges to make accurate measurements on a thin curved structure. This study evaluated the short and longitudinal reproducibility, as well as variations of T1p and T2 values measured using different MR systems, coils and sites.

In our single-site study, the longitudinal RMS-CV of T1p and T2 values were <3% over periods from 13 to 29 months, indicating excellent longitudinal reproducibility. The T2 results are in agreement with a multi site study with 1.7–5.4% RMS-CV over an 8 year period. Factors that can introduce longitudinal variations of relaxation time measurements include any external variations of environment in the scanner room (temperature for example), MR system software or hardware upgrades, fluctuations in the MR system and coil performance, as well as changes in the phantom composition (primarily dehydration which will decrease T1 and T2 values). In the present study, no obvious system drift was observed, suggesting that relaxation times can be measured reliably using modern MR systems over 29 months.

We observed significant differences in T1p and T2 values between different models of MR systems and coils. The MR systems
<table>
<thead>
<tr>
<th>Phantoms</th>
<th>Agarose Concentration (weight/volume, %)</th>
<th>Average# of Voxels Within ROI (n = 52)</th>
<th>Mean T1r (ms) within ROI (n = 52)</th>
<th>Pooled SD (ms) within ROI (n = 52)</th>
<th>Normalized RMS Fitting Error (n = 52)</th>
<th>Pooled RMS (ms) of Inter-location Variation (n = 52)</th>
<th>Pooled RMS (ms) of Long-term Reproducibility (n = 52)</th>
<th>Pooled RMS (ms) with Different MR Systems (n = 2)</th>
<th>Pooled RMS (ms) with Different Coils (n = 2)</th>
<th>Pooled RMS (ms) with Different Sites (n = 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T1r</strong></td>
<td>4%</td>
<td>5390</td>
<td>28.1</td>
<td>1.1</td>
<td>0.0056</td>
<td>0.5</td>
<td>0.6</td>
<td>0.4</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>3%</td>
<td>5390</td>
<td>37.9</td>
<td>1.2</td>
<td>0.0056</td>
<td>0.4</td>
<td>0.7</td>
<td>0.7</td>
<td>0.5</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>2%</td>
<td>5390</td>
<td>58.8</td>
<td>1.8</td>
<td>0.0050</td>
<td>0.4</td>
<td>1.2</td>
<td>1.2</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td><strong>Human Subjects</strong></td>
<td>Cartilage Compartment</td>
<td>Average# of Voxels Within ROI (n = 16)</td>
<td>Mean T1r (ms) within ROI (n = 16)</td>
<td>Pooled SD (ms) within ROI (n = 16)</td>
<td>Normalized RMS Fitting Error (n = 16)</td>
<td>Pooled RMS (ms) of Short-term Scan/rescan Reproducibility (n = 16)</td>
<td>Pooled RMS (ms) of Long-term Reproducibility (n = 4)</td>
<td>Pooled RMS (ms) with Different MR Systems (n = 10)</td>
<td>Pooled RMS (ms) with Different Coils (n = 5)</td>
<td>Pooled RMS (ms) with Different Sites (n = 4)</td>
</tr>
<tr>
<td>LF</td>
<td>2693</td>
<td>45.5</td>
<td>11.8</td>
<td>0.009</td>
<td>1.7</td>
<td>1.7</td>
<td>1.6</td>
<td>1.5</td>
<td>2.1</td>
<td>1.7</td>
</tr>
<tr>
<td>LT</td>
<td>1778</td>
<td>39.5</td>
<td>10.2</td>
<td>0.013</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>2.0</td>
<td>3.3</td>
<td>1.6</td>
</tr>
<tr>
<td>MF</td>
<td>3981</td>
<td>44.9</td>
<td>10.6</td>
<td>0.006</td>
<td>1.5</td>
<td>1.2</td>
<td>1.5</td>
<td>2.3</td>
<td>3.8</td>
<td>1.9</td>
</tr>
<tr>
<td>MT</td>
<td>1871</td>
<td>39.0</td>
<td>12.9</td>
<td>0.014</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.7</td>
<td>2.6</td>
<td>1.7</td>
</tr>
<tr>
<td>P</td>
<td>1739</td>
<td>46.9</td>
<td>9.4</td>
<td>0.013</td>
<td>1.1</td>
<td>1.7</td>
<td>1.7</td>
<td>2.3</td>
<td>1.4</td>
<td>2.3</td>
</tr>
<tr>
<td>TrF</td>
<td>2764</td>
<td>46.5</td>
<td>9.5</td>
<td>0.008</td>
<td>1.2</td>
<td>2.1</td>
<td>1.9</td>
<td>1.9</td>
<td>1.6</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>T2r</strong></td>
<td>4%</td>
<td>5390</td>
<td>27.2</td>
<td>0.8</td>
<td>0.0056</td>
<td>0.5</td>
<td>0.7</td>
<td>0.1</td>
<td>0.6</td>
<td>1.1</td>
</tr>
<tr>
<td>3%</td>
<td>5390</td>
<td>36.4</td>
<td>1.0</td>
<td>0.0069</td>
<td>0.5</td>
<td>0.8</td>
<td>0.3</td>
<td>1.2</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>2%</td>
<td>5390</td>
<td>56.0</td>
<td>1.8</td>
<td>0.0061</td>
<td>1.2</td>
<td>1.6</td>
<td>0.5</td>
<td>1.8</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>Human Subjects</strong></td>
<td>Cartilage Compartment</td>
<td>Average# of Voxels Within ROI (n = 16)</td>
<td>Mean T2 (ms) within ROI (n = 16)</td>
<td>Pooled SD (ms) within ROI (n = 16)</td>
<td>Normalized RMS Fitting Error (n = 16)</td>
<td>Pooled RMS (ms) of Short-term Scan/rescan Reproducibility (n = 16)</td>
<td>Pooled RMS (ms) of Long-term Reproducibility (n = 4)</td>
<td>Pooled RMS (ms) with Different MR Systems (n = 10)</td>
<td>Pooled RMS (ms) with Different Coils (n = 5)</td>
<td>Pooled RMS (ms) with Different Sites (n = 4)</td>
</tr>
<tr>
<td>LF</td>
<td>2693</td>
<td>32.7</td>
<td>9.7</td>
<td>0.014</td>
<td>1.4</td>
<td>1.3</td>
<td>1.3</td>
<td>2.5</td>
<td>0.2</td>
<td>1.2</td>
</tr>
<tr>
<td>LT</td>
<td>1778</td>
<td>28.3</td>
<td>8.6</td>
<td>0.012</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2.6</td>
<td>2.2</td>
<td>1.3</td>
</tr>
<tr>
<td>MF</td>
<td>3981</td>
<td>32.3</td>
<td>8.9</td>
<td>0.012</td>
<td>1.4</td>
<td>1.6</td>
<td>1.6</td>
<td>1.8</td>
<td>1.8</td>
<td>0.9</td>
</tr>
<tr>
<td>MT</td>
<td>1871</td>
<td>29.5</td>
<td>9.7</td>
<td>0.012</td>
<td>1.5</td>
<td>1.5</td>
<td>2.1</td>
<td>1.4</td>
<td>1.2</td>
<td>1.2</td>
</tr>
<tr>
<td>P</td>
<td>1739</td>
<td>31.9</td>
<td>7.3</td>
<td>0.009</td>
<td>1.0</td>
<td>1.4</td>
<td>2.0</td>
<td>0.2</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>TrF</td>
<td>2764</td>
<td>34.4</td>
<td>9.2</td>
<td>0.011</td>
<td>1.3</td>
<td>1.3</td>
<td>2.7</td>
<td>0.7</td>
<td>1.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>
used in the single-site study had different hardware systems including peak gradient amplitude, gradient slew rate and bore size, which resulted in different pulse width and minimum TR/TE in T₁ and T₂ sequences. Also, the transmit gain differed and likely introduced flip angle variations. In addition, the coupling between the knee and body RF coils caused by the construction techniques as well as bone diameter affects the efficiency of B₀, flip angles, and SNR. All of these factors may affect relaxation time quantification.

The RF coil transmits uniformity, which is influenced by both coil design and electric loading with subjects, can be another key factor contributing to variations in relaxation time. In general, the body transmit provides more uniform RF fields compared to a local transmit coil, but deposits higher energy (SAR) and is thus restrictive. The transmit B₁ non-uniformity will introduce spatial variations in flip angles then in relaxation times. The QT8PAR knee coil used in this study however has been documented to have a fairly good transmit uniformity²². In addition, the T₁ and T₂ sequence was used because different acquisition sequences can introduce significant differences in relaxation times²³. The authors speculated the low SNR resulted in underestimated T₁ and T₂ values for different model MR systems and coils, we allowed only GE MR750 3T MR systems with QT8PAR knee coils to be used in the multi-site post-ACL injury study. In addition, the identical T₁ and T₂ sequence was used because different acquisition sequences can introduce significant differences in relaxation times²³. A custom leg-holder, with the foot was positioned vertically in a foot holder was used to ensure consistent flexion and minimize joint rotation during scanning. Further, standardization of image acquisition was achieved by onsite training. Lastly, image analysis was performed centrally with stringent quality control. These efforts achieved the goal of acquiring accurate and reproducible

Table IV
SNR of control knees using different MR systems and different coils

<table>
<thead>
<tr>
<th>Images</th>
<th>LF</th>
<th>LT</th>
<th>MF</th>
<th>MT</th>
<th>P</th>
<th>TrF</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSL/TE = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDx LB (8Ch)</td>
<td>48.7 ± 12.3</td>
<td>43.2 ± 10.6</td>
<td>51.9 ± 11.3</td>
<td>45.5 ± 11.2</td>
<td>77.2 ± 19.2</td>
<td>65.9 ± 11.1</td>
</tr>
<tr>
<td>MR750W (8Ch)</td>
<td>43.8 ± 15.5</td>
<td>31.5 ± 17.9</td>
<td>47.9 ± 17.0</td>
<td>32.9 ± 20.0</td>
<td>54.6 ± 17.3</td>
<td>48.6 ± 13.1</td>
</tr>
<tr>
<td>MR750W (16Ch)</td>
<td>67.6 ± 21.4</td>
<td>58.8 ± 19.8</td>
<td>74.6 ± 29.2</td>
<td>61.5 ± 19.5</td>
<td>70.6 ± 20.2</td>
<td>73.4 ± 26.0</td>
</tr>
<tr>
<td>TSL = 80 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDx LB (8Ch)</td>
<td>11.4 ± 1.5</td>
<td>10.8 ± 2.4</td>
<td>12.3 ± 3.2</td>
<td>12.8 ± 2.6</td>
<td>14.5 ± 5.3</td>
<td>17.4 ± 2.8</td>
</tr>
<tr>
<td>MR750W (8Ch)</td>
<td>8.2 ± 0.4</td>
<td>7.7 ± 2.6</td>
<td>8.5 ± 1.0</td>
<td>7.5 ± 3.2</td>
<td>11.7 ± 1.6</td>
<td>11.6 ± 2.6</td>
</tr>
<tr>
<td>MR750W (16Ch)</td>
<td>13.0 ± 4.0</td>
<td>9.4 ± 3.7</td>
<td>13.7 ± 4.7</td>
<td>12.2 ± 4.8</td>
<td>12.3 ± 4.5</td>
<td>17.9 ± 7.3</td>
</tr>
<tr>
<td>TE = 54.7 ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDx LB (8Ch)</td>
<td>11.6 ± 2.6</td>
<td>12.6 ± 3.1</td>
<td>15.5 ± 4.2</td>
<td>15.3 ± 4.2</td>
<td>15.3 ± 5.0</td>
<td>21.1 ± 3.4</td>
</tr>
<tr>
<td>MR750W (8Ch)</td>
<td>8.8 ± 2.0</td>
<td>9.5 ± 2.4</td>
<td>9.8 ± 2.3</td>
<td>9.1 ± 2.3</td>
<td>11.7 ± 2.9</td>
<td>14.4 ± 3.6</td>
</tr>
<tr>
<td>MR750W (16Ch)</td>
<td>12.7 ± 3.3</td>
<td>11.5 ± 5.9</td>
<td>14.7 ± 6.3</td>
<td>14.1 ± 7.1</td>
<td>12.5 ± 6.4</td>
<td>18.8 ± 7.9</td>
</tr>
</tbody>
</table>

8Ch: Quadrature transmit and 8-channel phased array receive knee coil; 16Ch: receive only 16-channel phased array flex coil.
quantitative relaxation time values from each site and to enable pooling the data from all sites.

The longitudinal phantom $T_1r$ and $T_2$ RMS-CVs from each site are comparable to previous multi-site $T_2$ studies\(^\text{17}\), indicating good longitudinal stability. The overall scan/re-scans RMS-CV in human subjects (3.1% for $T_1r$ and 4.0% for $T_2$) was comparable to single-site CVs\(^\text{24–26}\) and better than a multi-site study\(^\text{19}\). These CVs were less than the group differences between healthy and OA cartilage relaxation times, with a CV of 9–10% between the two groups\(^\text{8}\). Our good multi-site in vivo reproducibility is attributed to our stringent study design.

Significant differences were observed for phantom $T_1r$ and $T_2$ values between the three sites and are attributed to different performances of the MR systems and RF coils as well as environmental influences. The phantom $T_1r$ and $T_2$ values between sites were highly correlated ($R > 0.99$), suggesting the differences maybe corrected and allow pooling data for multi-site analysis.

No significant differences were observed for in vivo $T_1r$ and $T_2$ values for the traveling control subjects between the three sites despite the significant differences in phantoms values. The differences maybe due to different coil loading between phantom and human subjects, and the small systematic differences between sites maybe masked by in vivo measurement variations. No significant differences in $T_1r$ and $T_2$ values were found in the traveling controls from baseline to follow-up, and the longitudinal CVs were comparable to cross-sectional CVs, suggesting good in vivo longitudinal reproducibility. This study is limited by the small number of human subjects for the cross-validation of $T_1r$ and $T_2$ between sites. In addition, the relatively low resolution of $T_1r$ and $T_2$ images (0.6 mm in plane with 4 mm slices) may introduce bias to $T_1r$ and $T_2$ values due to the partial volume effect. Advanced acceleration techniques can be applied in the future to obtain $T_1r$ and $T_2$ images with higher resolutions within clinically acceptable acquisition time\(^\text{42}\). The reproducibility was evaluated only in healthy controls and should be evaluated in OA subjects in future studies. Different fitting methods generate significantly different quantification values\(^\text{28,29}\).

**Fig. 4.** Variations in in vivo $T_1r$ (A) and $T_2$ (B) values and Bland–Altman plots of $T_1r$ (C) and $T_2$ (D) values using different coils. $T_1r$ and $T_2$ measured using the 16PAR flex coil were significantly higher than those measured using the QT8PAR knee coil.

**Fig. 5.** In vivo scan/re-scan reproducibility of cartilage $T_1r$ and $T_2$ values of the multi-center study. (A) Overall CVs and CVs for each site; (B) CVs for each compartment.
and have different sensitivity to SNR and may yield different bias between MR systems or spatially across coils, which was not discussed in this study because the data were processed centrally using the same fitting method. Segmentation also introduces variation\(^\text{30}\). The intra- and inter-operator variation using the same post-processing software have been previously reported\(^\text{11}\).

In conclusion, minimizing variation has enabled good reproducibility and cross-validation to be achieved between sites for cartilage T1\(_g\) and T2 quantification. This is an essential step prior to initiating multi-site longitudinal studies or clinical trials. The results from this study identify quality control and cross-calibration methods required for quantitative MRI to be applied in multi-site studies for evaluating cartilage degeneration. Future studies are needed to expand the multi-site study to include MR systems from multiple manufacturers.

**Author contributions**

Study and Concept Design: XL, SM.

Data Collection and Coordination: VP, DK, JR, CW, DL, DS, MFK, JF, SLW.

Data Analysis and Results Interpretation: XL, VP, DK, JR, CW, DL, NO, KA.

Manuscript Draft: XL.

Critical Review, Edit and Proof of Manuscript: all authors.

**Conflict of interest**

No conflict of interest to the study from all authors.

**Acknowledgments**

The study was supported by NIH/NIAMS P50 AR060752 and the Arthritis Foundation (AF). The authors would like to thank the AF multi-site ACL study team.

**References**


