BJR

https://doi.org/10.1259/bjr.20190143

Received: 08 February 2019 | Revised: 14 June 2019 | Accepted: 30 June 2019

Cite this article as:
Farrow M, Grainger AJ, Tan AL, Buch MH, Emery P, Ridgway JP, et al. Normal values and test–retest variability of stimulated-echo diffusion tensor imaging and fat fraction measurements in the muscle. *Br J Radiol* 2019; **92**: 20190143.

FULL PAPER

Normal values and test–retest variability of stimulated-echo diffusion tensor imaging and fat fraction measurements in the muscle

[1,2]MATTHEW FARROW, BSc, [1,2]ANDREW J GRAINGER, MD, PhD, [1,2]AI LYN TAN, MD, PhD, [1,2]MAYA H BUCH, MD, PhD, [1,2]PAUL EMERY, MD, PhD, [2,3]JOHN P RIDGWAY, PhD, [4]THORSTEN FEIWEIER, PhD, [2,3]STEVEN F TANNER, PhD and [2,3]JOHN BIGLANDS, PhD

[1]Leeds institute of Rheumatic and Musculoskeletal Medicine, Chapel Allerton Hospital, University of Leeds, United Kingdom
[2]NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom
[3]Medical Physics and Engineering, Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom
[4]Siemens Healthcare GmbH, Erlangen, Germany

Address correspondence to: John Biglands
E-mail: *j.biglands@nhs.net*

Objectives: To assess the test–retest variability of both diffusion parameters and fat fraction (FF) estimates in normal muscle, and to assess differences in normal values between muscles in the thigh.
Methods: 29 healthy volunteers (mean age 37 years, range 20–60 years, 17/29 males) completed the study. Magnetic resonance images of the mid-thigh were acquired using a stimulated echo acquisition mode-echoplanar imaging (STEAM-EPI) imaging sequence, to assess diffusion, and 2-point Dixon imaging, to assess FF. Imaging was repeated in 19 participants after a 30 min interval in order to assess test–retest variability of the measurements.
Results: Intraclass correlation coefficients (ICCs) for test–retest variability were 0.99 [95% confidence interval, (CI): 0.98, 1] for FF, 0.94 (95% CI: 0.84, 0.97) for mean diffusivity and 0.89 (95% CI: 0.74, 0.96) for fractional anisotropy (FA). FF was higher in the hamstrings than the quadriceps by a mean difference of 1.81% (95% CI:1.63, 2.00)%, $p < 0.001$. Mean diffusivity was significantly lower in the hamstrings than the quadriceps (0.26 (0.13, 0.39) x10^{-3} mm^2s^{-1}, $p < 0.001$) whereas fractional anisotropy was significantly higher in the hamstrings relative to the quadriceps with a mean difference of 0.063 (0.05, 0.07), $p < 0.001$.
Conclusions: This study has shown excellent test-retest, variability in MR-based FF and diffusion measurements and demonstrated significant differences in these measures between hamstrings and quadriceps in the healthy thigh.
Advances in knowledge: Test–retest variability is excellent for STEAM-EPI diffusion and 2-point Dixon-based FF measurements in the healthy muscle. Inter- and intra-observer variability were excellent for region of interest placement for STEAM-EPI diffusion and 2-point Dixon-based FF measurements in the healthy muscle. There are significant differences in FF and diffusion measurements between the hamstrings and quadriceps in the normal muscle.

INTRODUCTION

There is growing evidence supporting the use of diffusion tensor imaging (DTI) in skeletal muscle. DTI measurements have been shown to be sensitive to muscle changes due to myositis,[1–3] diabetes mellitus,[4] muscle injury,[5,6] exercise[7–9] and ageing.[10] The majority of these studies use spin echo DTI; however there are advantages in using a stimulated echo acquisition mode (STEAM) approach. The STEAM preparation stores transverse magnetization along the longitudinal axis before recovering it after a given mixing time (TM) during which signal decay is governed by T_1. This allows for long diffusion times without excessive loss of SNR due to T_2 decay.[6,11] Additionally, the need for high amplitude gradients is reduced, decreasing induced eddy currents that can spatially distort images.[12] Fat is suppressed during TM due to the short T_1 of fat, including the olefinic fat peak, whose resonant frequency is close to that of water and is usually not suppressed by chemical-shift-selective fat suppression techniques.[13] This is particularly important in muscle imaging where fat signal can potentially confound quantitative measurements.[14,15] Finally, a long TM allows a long diffusion time (TD), which may be necessary to ensure sensitivity to restricted diffusion in the muscle as the diameter of muscle fibres is relatively large (~50 μm).[16]

A variety of factors can lead to errors in quantitative imaging measurements, including patient positioning and region of interest (ROI) placement. Such errors may mask subtle muscle changes in quantitative MR studies; therefore it is important to assess measurement variability. The existing literature evaluating variability of muscle DTI measurements is limited to small datasets ($n < 10$) using spin echo-based diffusion measurements.[17–19] As there are microstructure differences between muscles,[20–22] it is also important to quantify differences in diffusion parameters between healthy muscles so that these can be accounted for when studying changes due to disease. Although these differences have been reported using spin echo DTI,[23–25] they have not been previously reported using a STEAM technique in the normal thigh.

Muscle fat content is also an important factor in assessing muscle health. Fatty infiltration occurs in many pathological conditions affecting skeletal muscle and as part of the ageing process.[10] Fat fraction (FF) measurements have been shown to be sensitive to muscle changes in response to exercise,[26] muscular dystrophy,[27,28] and myositis.[29] However, the existing test–retest variability data in the literature tends to be based on small study sizes.[17,30,31]

The aim of this study was to measure the interscan variability, as well as the inter- and intraobserver variability of muscle ROI placement, from STEAM-EPI DTI and 2-point Dixon based muscle FF measurements. In addition, the differences in diffusion parameters and FF between healthy thigh muscles were assessed.

METHODS

Participants

This prospective study was conducted at the Leeds Teaching Hospitals NHS Trust between January 2015 and January 2017. 30 healthy volunteers gave written, informed consent to take part in the study with approval of the National Research Ethics Service (14-LO-1785 and 17-EM-0079). The study was conducted according to the Declaration of Helsinki. Sample size was based on published rules of thumb recommending between 12 and 30 participants per group of interest to estimate parameters for powering future clinical trials.[32] Out of the 30 participants, 19 agreed to return for a repeat scan to assess variability, with a randomly selected proportion also undergoing MR spectroscopy (MRS). Exclusion criteria for healthy volunteers were age <18 years, contraindications to MRI, previous history of muscle disorder or arthritis, spinal disease and neuropathy.

MR measurements

Axial MR data were acquired using a MAGNETOM Verio 3T MR scanner (Siemens Healthcare, Erlangen, Germany). Two small, four-channel, flex coils were wrapped around the right thigh and placed with the distal end of both coils positioned 4 cm from the superior edge of the patella.

Fat quantitation was performed using a 40-slice, volume interpolated breath-hold examination (VIBE), 2-point Dixon sequence: repetition time (TR) 11 ms, echo time (TE) 2.45 ms & 3.675 ms, slice thickness 5 mm, matrix 256 × 256, field of view (FOV) 300 mm, acquisition time 1 min 47 s, flip angle 15°.[17,28,33] 2-point VIBE Dixon was selected because of its wide availability and well documented recent use in the muscle.[26,29,33–38] Diffusion-weighted images (DWI) were acquired using a STEAM prototype sequence with an echoplanar imaging (EPI) readout[39] with spectral adiabatic inversion recovery fat suppression: TR/TE: 6300 ms/42.4 ms, slice thickness 5 mm, matrix 128 × 128, FOV 300 mm, partial Fourier 6/8, GRAPPA acceleration factor 2 (24 reference lines), six diffusion directions, bandwidth 1448 Hz/pixel, four slices, nominal *b*-values 0 & 500 s/mm^2, eight averages per b-value. The TM, the time between the second and third 90° RF pulses, was 981 ms. The TD, the time between the start of two diffusion encoding gradients, was 1000 ms, acquisition time was 6 min 12 s. With the given TM, contributions from imaging gradients to the *b*-matrix are no longer negligible and have been considered in the diffusion tensor estimation.

To ensure that the imaging volume was positioned consistently, the inferior edge of the VIBE Dixon volume was located at the insertion point of the distal rectus femoris muscle into the tendon. The diffusion acquisition slices were aligned with the central four slices of the VIBE Dixon volume. Participants returning for a second scan (visit 2) left the scanner room for thirty-minutes between scans. Participants did not undertake any activity during the 30 min intervals to avoid exercise-induced physiological muscle changes.

To validate the Dixon fat measurements, single-voxel MR spectra were acquired on a subset of the recruited population. A 10×10×10 mm^3 voxel was positioned in the vastus lateralis avoiding regions of vasculature or fascial fat. The acquisition employed a STEAM sequence with TR: 4000 ms, TE: 20 ms, 30 ms, 40 ms, 50 ms and 70 ms, 1024 data points, bandwidth: 1500 Hz and 24 averages.

Data analysis

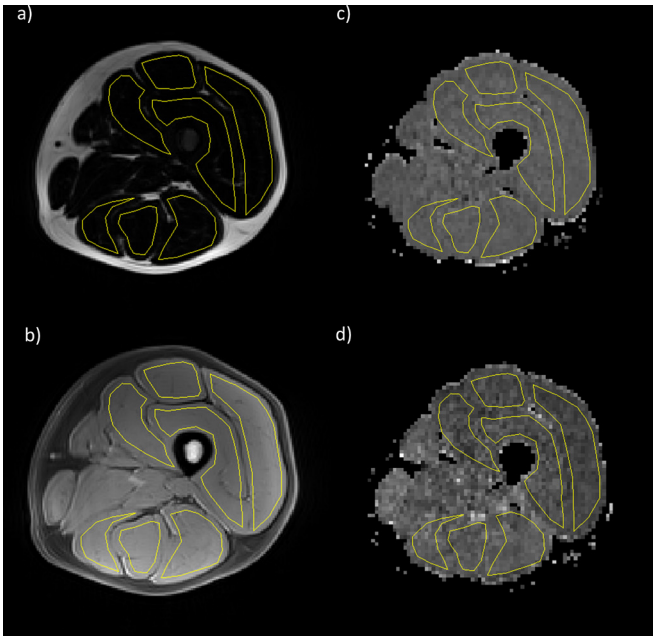
Fat and water images were generated from the in- and out-of-phase images using the scanner vendor's software. B_0 variations due to changes in magnetic susceptibility at tissue interfaces were corrected for using a phase-correction method.[40] Differences in T_1 between fat and water were corrected for using assumed values for water (T_{1w}=1420 ms) and fat (T_{1f}=371ms)[41] as described by Liu at al..[42] The FF was then calculated from the adjusted fat (Sf) and water (Sw) signals as follows:

$$fat\ fraction\ (FF) = \frac{S_f}{S_f+S_w} \times 100\%$$

Diffusion-weighted images (DWI) were converted to mean diffusivity (MD), fractional anisotropy (FA) and diffusion eigenvalue (λ_1, λ_2, λ_3) maps using the scanner vendor's software.

ROIs were contoured by two raters (AG with 20 years of experience; MF with 1 year of experience) using Osirix imaging software (v. 4.0; open-source DICOM viewer, www.osirix-viewer.com). Regions depicting the individual hamstring muscles (semi-tendinosus, semi-membranosus, biceps femoris) and quadriceps muscles (rectus femoris, vastus lateralis, vastus medialis and vastus intermedius) were drawn on the middle slice (slice 20) of the in-phase VIBE Dixon volume for each participant, avoiding fascial

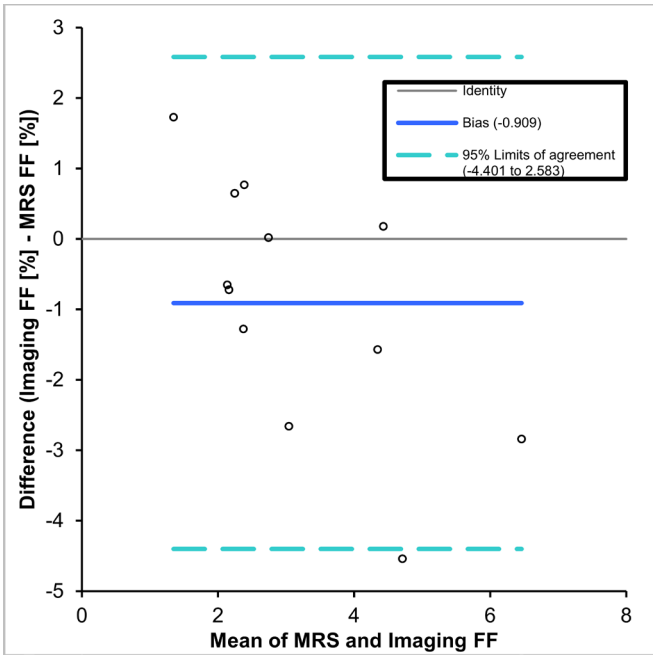
Figure 1. Example images from a healthy volunteer showing VIBE Dixon (a) fat and (b) water images and STEAM diffusion maps (c) MD and (d) FA. Regions of interest (shown in yellow) were drawn corresponding to the individual muscles of the hamstrings and quadriceps. FA,fractional anisotropy; MD, mean diffusivity; STEAM, stimulated echo acquisition mode; VIBE, volume interpolated breath-hold examination.



tissue and subcutaneous fat. The medial compartment of the thigh was not included as individual muscles in this compartment can be difficult to discriminate. To assess test–retest variability, the visit 1 and visit 2 MRI data sets were contoured by the same researcher (AG). To assess inter-rater variability, the visit 1 MRI data sets were contoured by two researchers separately (AG and MF), blinded to each other's ROIs. To assess intra-rater variability, the visit 1 MRI data sets were contoured twice by the same researcher (AG), with a 6 month interval between contours. ROIs were copied to the corresponding diffusion parameter maps, accounting for differences in image resolution, and the mean value within each ROI was taken (Figure 1).

To obtain FF measurements from the MRS data the acquired free-induction-decays were zero-filled (times 2) and line-broadened using a Gaussian filter. After Fourier transformation, resulting spectra were baseline- and phase-corrected. The areas under the fat and water spectral peaks were measured by integrating between the 0.5 to 3.5 ppm range for lipid and 3.6 to 5.8 ppm range for water. To correct for differences in T_2 relaxation between water and fat, water T_2 values were determined using the area under the water peak from spectra acquired with echo times of 20, 30, 40, 50 and 70 ms and a mono-exponential fit was used to derive T_2. Analogous T_2 measurements for fat were unreliable because the fat peaks were too small. Therefore, an assumed T_2 for lipid of 59.1 ms was used.[28] Finally, the FF was determined from the area under the peak values from the fat and water signals in the TE = 20 ms spectrum, corrected for T_2 effects. Spectroscopy based FF measurements were

Figure 2. Bland-Altman agreement plot showing the comparison of FF values measured by spectroscopic and Dixon imaging techniques. FF, fat fraction; MRS, MR spectroscopy.



compared against the imaging FF measurements taken from the vastus lateralis.

Statistical analyses

Offline image analysis was performed using MATLAB software (R2018a, Mathworks, Nattick, MA). Statistical analyses were performed using SPSS (IBM SPSS Statistics for Windows, v. 22.0. Armonk, NY: IBM Corp.). ICC and Bland–Altman plots were generated in MedCalc (MedCalc Software bvba, Ostend, Belgium; http://www.medcalc.org; 2017). ANOVA analysis was used to evaluate differences between individual muscles. A Greenhouse–Geisser correction was used in cases where sphericity was violated, and the Bonferroni correction was used in post-hoc analyses. FF values were log-transformed to a normal distribution before statistical analysis. Agreement was measured using Bland–Altman plots. Variability was assessed using the intraclass correlation coefficient (ICC) using a two-way mixed model with absolute agreement. ICC values above 0.60 were classed as good and values above 0.75 as excellent variability between the two measurements.[43] All data are presented as: mean (95% confidence interval), *p*-value, unless stated otherwise.

RESULTS

Participants

30 healthy volunteers were recruited, (mean age 36.5 years, range 20–60 years, 17/30 males) and 20 returned for a second scan to assess test–retest variability. Imaging data from one healthy volunteer was excluded due to misalignment between the fat and diffusion image at acquisition. The remaining 29 healthy volunteers had a mean age of 37 years, range 20–60 years, 17/29 males. For test–retest variability, one participant declined a second scan leaving 19 participants in the variability study (mean age 37, range 20–60

Table 1. Measured values (mean ± standard deviation) for individual muscles and muscle groups for FF, MD, FA and eigenvalue measurements

Muscle	FF [%]	MD [$x10^{-3}mm^2s^{-1}$]	λ_1 [$x10^{-3}mm^2s^{-1}$]	λ_2 [$x10^{-3}mm^2s^{-1}$]	λ_3 [$x10^{-3}mm^2s^{-1}$]	FA
Semi-Mem	4.02 ± 1.67	1.30 ± 0.06	1.91 ± 0.10	1.09 ± 0.07	0.90 ± 0.07	0.39 ± 0.04
Semi-Tend	4.34 ± 2.26	1.25 ± 0.05	1.92 ± 0.06	1.02 ± 0.08	0.82 ± 0.08	0.44 ± 0.05
Bic Fem	4.17 ± 1.63	1.29 ± 0.04	1.87 ± 0.06	1.08 ± 0.06	0.91 ± 0.06	0.38 ± 0.04
Hamstrings	**4.21 ± 1.85**	**1.28 ± 0.04**	**1.90 ± 0.05**	**1.06 ± 0.06**	**0.87 ± 0.06**	**0.41 ± 0.04**
Vast Med	2.07 ± 0.53	1.35 ± 0.05	1.87 ± 0.07	1.22 ± 0.07	0.96 ± 0.05	0.34 ± 0.03
Vast Int	2.43 ± 0.47	1.33 ± 0.06	1.86 ± 0.06	1.17 ± 0.09	0.96 ± 0.07	0.35 ± 0.04
Vast Lat	2.53 ± 0.61	1.30 ± 0.06	1.80 ± 0.07	1.12 ± 0.09	0.97 ± 0.07	0.33 ± 0.04
Rec Fem	1.80 ± 0.66	1.26 ± 0.05	1.77 ± 0.07	1.09 ± 0.06	0.90 ± 0.07	0.35 ± 0.04
Quadriceps	**2.21 ± 0.48**	**1.31 ± 0.05**	**1.83 ± 0.07**	**1.15 ± 0.08**	**0.95 ± 0.06**	**0.34 ± 0.03**

FA, fractional anisotropy; FF, fat fraction; MD, mean diffusivity.

years 11/19 males). The semimembranosus muscle was not visible on the relevant slice in three healthy volunteers, so this muscle did not contribute to the hamstring measurement in those participants.

Differences between muscles

The Bland–Altman plot (Figure 2) showed that imaging underestimated FF relative to spectroscopy with a mean absolute FF bias of −0.91% (-4.4% to 2.6%). Mean values for FF and diffusion parameters for each muscle are shown in Table 1 and boxplots for the hamstrings and quadriceps muscle groups are shown in Figure 3. The FF in the hamstrings was approximately twice that in the quadriceps with a mean difference in FF between the two muscle groups of 1.81 (1.63, 2.00) %, $p < 0.001$). The MD was significantly lower in the hamstrings than in the quadriceps with a mean difference

Figure 3. Boxplots showing difference between hamstrings and quadriceps muscle groups in (a) fat fraction [%], (b) mean diffusivity [$x10^{-3}mm^2s^{-1}$], (c) fractional anisotropy, (d) λ_1 [$x10^{-3}mm^2s^{-1}$], (e) λ_2 [$x10^{-3}mm^2s^{-1}$], (f) λ_3 [$x\ 10^{-3}mm^2s^{-1}$].

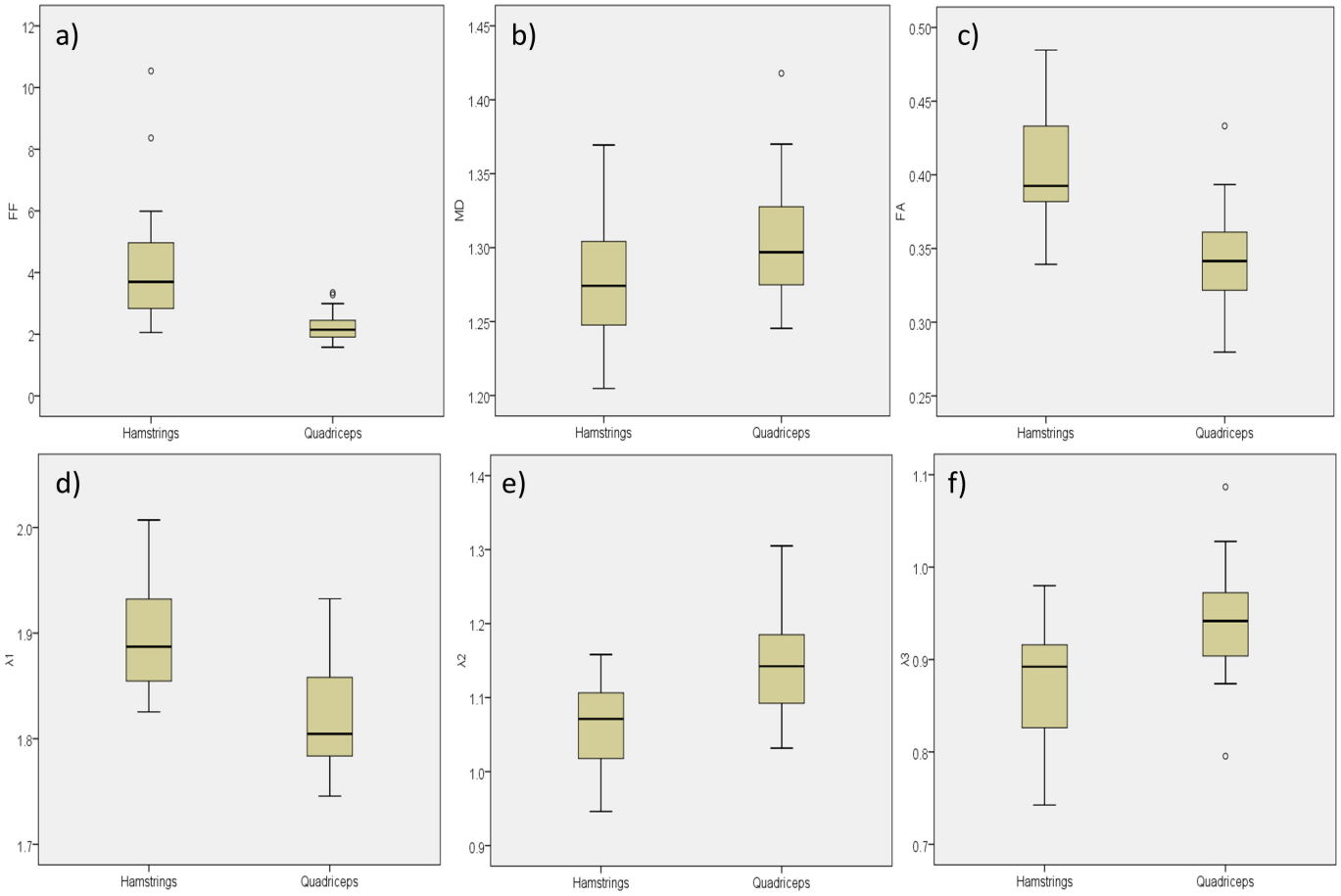
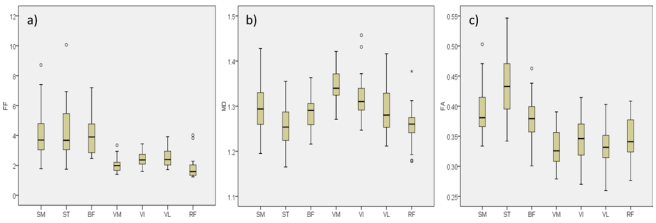


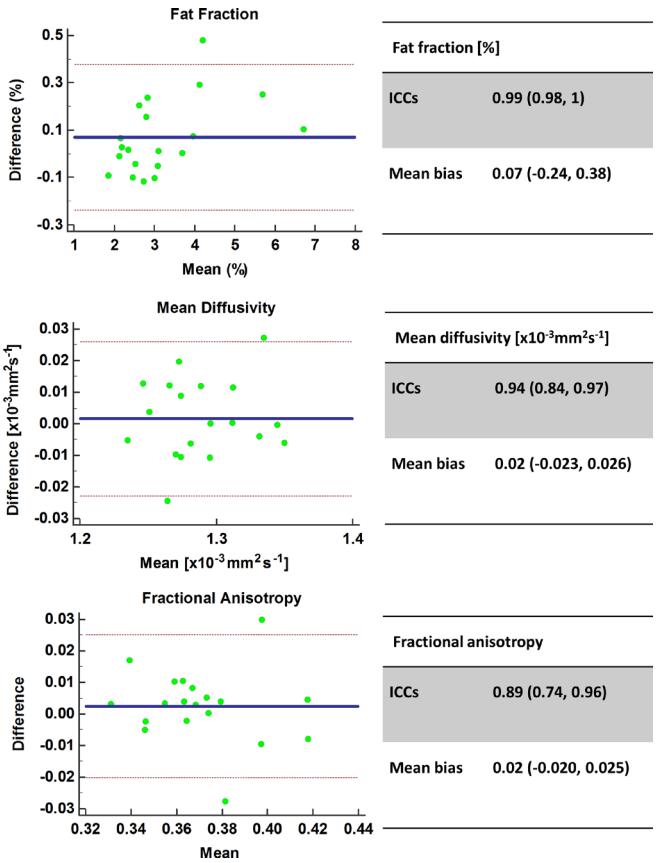
Figure 4. Boxplots showing distribution of values for each individual muscle for (a) fat fraction [%], (b) mean diffusivity [x10^{-3}mm^{2}s^{-1}], (c) fractional anisotropy, for the SM, ST, BF, VM, VI, VL and RF. RF,rectus femoris; SM, semi-membraneous;ST, semi-tendinosus; BF, bicepsfemoris; VM, vastus medialis; VI, vastus intermedius; VL,vastus lateralis.



of 0.26 (0.13, 0.39) x 10^{-3} mm^{2}s^{-1}, $p < 0.001$. The FA was significantly higher in the hamstrings than the quadriceps with a mean difference of 0.063 (0.05, 0.07), $p < 0.001$. The diffusion eigenvalues showed that λ_1 was significantly higher 0.076 (0.059, 0.092) x10^{-3} mm^{2}s^{-1}, $p < 0.001$, and λ_2 and λ_3 were lower in the hamstrings relative to the quadriceps; 0.083 (0.064, 1.03) x 10^{-3} mm^{2}s^{-1}, $p < 0.001$ and 0.068 (0.051, 0.085) x 10^{-3} mm^{2}s^{-1}, $p < 0.001$, respectively.

Boxplots for the individual muscles measured are shown in Figure 4. There were significant differences in FF between muscles (ANOVA $p < 0.001$). Pairwise post-hoc analysis showed that all significant differences were explained by the differences between the hamstrings and quadriceps groups. That is, any of the individual muscles that make up the hamstrings had a significantly different FF to any of the muscles that make up the quadriceps but was not significantly different from any other muscles within the hamstrings, and vice-versa. Similarly, there were significant differences in FA between individual muscles (ANOVA: $p < 0.001$) and all significant differences in FA were explained by the difference between the hamstrings and quadriceps groups. There were also significant differences in individual muscles in MD (ANOVA: $p < 0.001$). However, with MD the differences between individual muscles were not so clearly separated. Similar values in MD were observed between some muscles in the quadriceps and hamstrings; *e.g.* vastus intermedius was not significantly different to semimembranosus or biceps femoris.

Figure 5. Test–retest variability Bland–Altman agreement plots for FF, MD and FA. FA,fractional anisotropy; FF, fat fraction; MD, mean diffusivity.



Variability

Figure 5 depicts the Bland–Altman agreement plots between visits. Variability was excellent with FF ICC: 0.99 (0.98, 1), mean bias 0.07 (-0.24, 0.38)%, $p < 0.05$, MD ICC: 0.94 (0.84, 0.97), mean bias 0.02 (-0.023, 0.026) x10^{-3}mm^{2}s^{-1}, $p < 0.05$ and FA ICC: 0.89 (0.74, 0.96), mean bias 0.02 (-0.020, 0.025), $p < 0.05$). Test–retest variability analyses were also performed separately for the hamstrings and quadriceps muscle groups, and no substantial differences in variability were found (Table 2).

Figure 6 depicts the Bland–Altman agreement plots between different ROI raters. Variability was excellent with for FF ICC: 0.97 (0.93, 0.99), mean bias −0.09 (-0.60, 0.42)%, $p < 0.05$, MD ICC: 0.99 (0.97, 1), mean bias −0.013 (-0.017, 0.009) x10^{-3}mm^{2}s^{-1}, $p < 0.05$, FA 0.99 (0.98, 1), mean bias −0.001 (-0.010, 0.007), $p < 0.05$. Inter-rater variability analyses were also performed separately for the hamstrings and quadriceps muscle groups, and no substantial differences in variability were found (Table 2). Figure 7 depicts the Bland–Altman agreement plots between the same ROI rater with a 6 month interval between rating. Variability was excellent with for FF ICC: 0.99 (0.96, 1), mean bias 0.10 (-0.17, 0.36) %; $p < 0.05$, MD ICC: 0.93 (0.85, 0.97), mean bias −0.003 (-0.035, 0.029) x10^{-3}mm^{2}s^{-1}; $p < 0.05$ and FA ICC 0.99 (0.97, 0.99), mean bias −0.001 (-0.012, 0.009); $p < 0.05$. Intra–rater variability analyses were also performed separately for the hamstrings and quadriceps muscle groups, and no substantial differences in variability were found (Table 2). There were two visible outliers in the intraobserver variability data. On investigation, these were both consistent with discrepancies in the ROIs for vastus lateralis and vastus intermedius between contours.

DISCUSSION

This study has demonstrated excellent test–retest variability of STEAM DTI and FF measurements as well as excellent inter- and intraobserver variability for ROI drawing. The study shows significant differences in FF and DTI measurements between the quadriceps and hamstrings muscles, with the hamstrings exhibiting greater fat content, reduced MD and increased FA values.

Mean values (Table 1) agreed well with previous studies using similar techniques to measure FF[33,34] and DTI values.[44] There was good agreement between measured FF values obtained using

Table 2. Variability ICC values for quadriceps, hamstrings and combined measures in both muscle groups

Test–retest	FF	MD	FA	λ1	λ 2	λ 3
Hamstrings	0.99 (0.96, 1)	0.92 (0.8, 0.97)	0.84 (0.63, 0.94)	0.72 (0.39, 0.89)	0.93 (0.83, 0.97)	0.85 (0.64, 0.94)
Quadriceps	0.98 (0.94, 0.99)	0.88 (0.70, 0.95)	0.88 (0.71, 0.95)	0.73 (0.40, 0.89)	0.94 (0.84, 0.98)	0.85 (0.64, 0.94)
Both	0.99 (0.98, 1)	0.94 (0.84, 0.97)	0.89 (0.74, 0.96)	0.85 (0.64, 0.94)	0.96 (0.89, 0.98)	0.88 (0.71, 0.95)
Mean bias	0.07 (−0.24, 0.38)	0.02 (−0.023, 0.026)	0.02 (−0.020, 0.025)	0.07 (−0.03, 0.04)	0.00 (−0.03, 0.03)	−0.02 (−0.04, 0.04)
Inter rater						
Hamstrings	0.96 (0.91, 0.99)	0.96 (0.90, 0.98)	0.98 (0.96, 1)	0.90 (0.76, 0.96)	0.99 (0.97, 1)	1 (0.98, 1)
Quadriceps	0.99 (0.98, 1)	1 (0.99, 1)	0.99 (0.98, 1)	1 (0.99, 1)	1 (0.99, 1)	0.98 (0.96, 0.99)
Both	0.97 (0.93, 0.99)	0.99 (0.97,1)	0.99 (0.98, 1)	0.98 (0.95,0.99)	1 (0.99,1)	0.99 (0.98, 1)
Mean bias	−0.09 (−0.60, 0.42)	−0.013 (−0.010, 0.007)	−0.001 (−0.010, 0.007	−0.005 (−0.03, 0.02)	−0.001 (−0.01, 0.01)	0.00 (−0.01, 0.01)
Intra-rater						
Hamstrings	0.99 (0.94, 1)	0.94 (0.84, 0.98)	0.99 (0.98, 1)	0.90 (0.76, 0.96)	0.98 (0.95, 0.99)	0.98 (0.95, 0.99)
Quadriceps	0.99 (0.98, 1)	0.88 (0.72, 0.96)	0.96 (0.90, 0.98)	0.90 (0.76, 0.96)	0.96 (0.90, 0.98)	0.98 (0.95, 0.99)
Both	0.99 (0.96, 1)	0.93 (0.85, 0.97)	0.99 (0.97, 0.99)	0.64 (0.29, 0.84)	0.98 (0.96, 0.99)	0.99 (0.98, 1)
Mean bias	0.10 (−0.17, 0.36)	−0.003 (−0.035, 0.029)	−0.001 (−0.012, 0.009)	0.01 (−0.16, 0.18)	−0.001 (−0.02, 0.02)	−0.001 (−0.01, 0.01)

FA, fractional anisotropy; FF, fat fraction; ICC, intraclass correlation coefficient; MD, mean diffusivity.

imaging and spectroscopy (Figure 2), which has been observed previously.[28,34] Imaging slightly underestimated FF with a mean bias of 0.9% absolute FF, which was likely to be due to noise bias and the use of only a single spectral peak when quantifying the fat.[35,45,46]

Our test–retest variability scores for DTI values were high with ICC values of 0.94 & 0.89 for MD and FA respectively. In smaller studies using spin echo DTI Ponrartana et al[47] reported ICCs of 0.88 for MD and 0.75 for FA in the lower leg, and Froeling et al[19] reported ICCs of 0.89 for MD and 0.60 for FA in the forearm. For FF, our test–retest variability was 0.99 (0.98, 1), which again compares favourably with previous work, ICC 0.91.[31] Our higher variability scores may be due to the short interval between visits (30 min) in our study compared to 1–2 weeks.[19,31] Short-term factors such as exercise are known to affect DTI[7,8] and FF.[26] Our choice of a short between-scan interval minimized changes due to physiological variations between tests providing a measure of variability of the system alone. However, researchers undertaking longitudinal studies should expect poorer variability in the measurements than those presented here due to physiological variation.

Mean biases and confidence intervals in our study were small; MD: 0.02 (–0.023, 0.026) 0.03×10^{-3} mm^2s^{-1}, FA: 0.02 (-0.020, 0.025), FF: 0.07% (-0.24, 0.38) %. These results suggest that measured differences greater than MD: 0.03×10^{-3} mm^2s^{-1}, FA: 0.03 and FF: 0.4% should be detectable above system variability errors. The existing evidence shows that measurable changes due to diseases such as myositis[1,3] and diabetes[4,48] exceed these small variability biases, which suggests that diffusion and fat measurements should be reliably sensitive to changes due to disease in the muscle. This is an important early step in the validation of these measurements in muscle disease, as interest in these measurements as potential biomarkers for assessing muscle disease grows.

Our data showed that, in healthy volunteers, FF was higher in the hamstrings relative to the quadriceps, which is consistent with a previous study.[49] This may be because the hamstrings have a greater number of Type-I muscle fibres,[20] which have a higher lipid content than Type-II fibres.[50] Our results showed significant differences in measured diffusion parameters between the hamstrings and quadriceps. For FA, this finding was consistent across the individual muscles that make up these groups, so that any individual muscle in the hamstrings was significantly different in value to any individual muscle in the quadriceps and vice-versa. However, for MD this was not the case. This could imply that FA is more sensitive to the tissue microstructure differences between muscle groups than MD. The differences in diffusion parameters observed here are consistent with previous studies,[17] and this observation has been interpreted in different ways. Higher MD in the quadriceps could

Figure 6. Inter-rater variability Bland–Altman agreement plots for FF, MD and FA. FA, fractional anisotropy; ICC, intraclass correlation coefficient; MD, mean diffusivity.

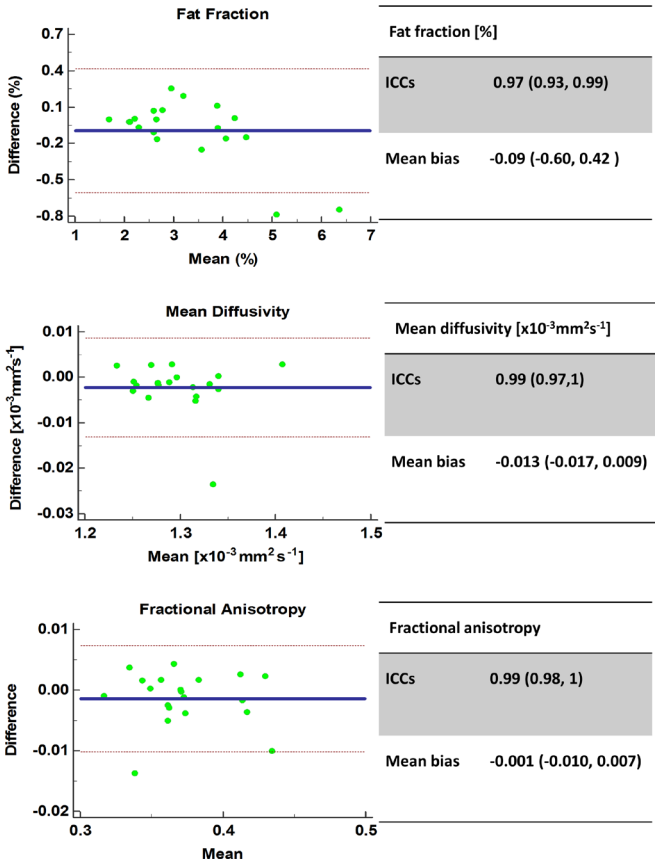
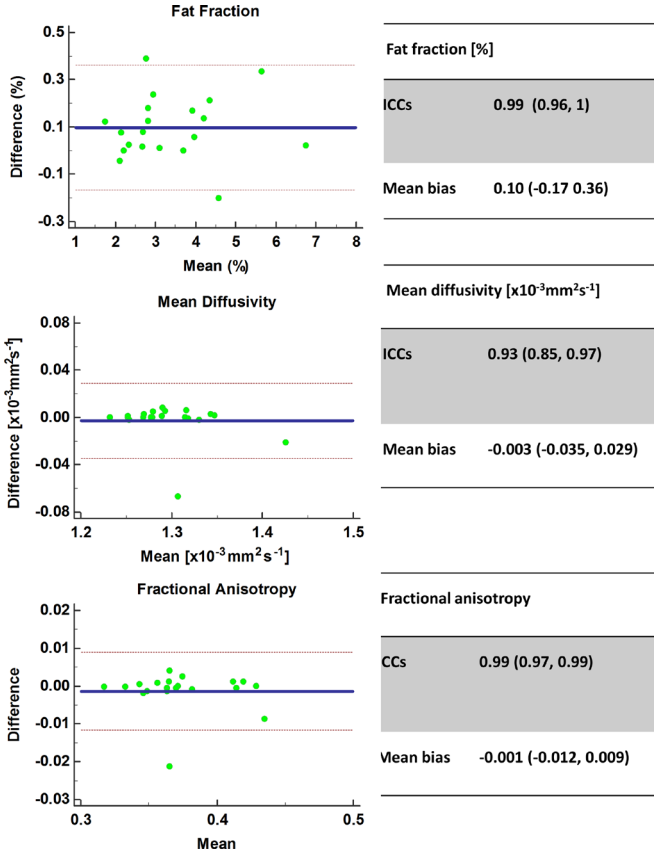


Figure 7. Intra-rater variability Bland–Altman agreement plots for FF, MD and FA. FA, fractional anisotropy; FF, fat fraction; ICC, intraclass correlation coefficient; MD, mean diffusivity.



be due to greater relative hydration (extracellular water) in these muscles,[25] whilst differences in FA could be due to differences in the uniformity of fibre tracks.[51] Alternatively, anatomical studies report a higher density of Type-I fibres in the hamstrings than the quadriceps,[20] so the reduced λ_2 and λ_3 and increased FA values that we have observed in the hamstrings could reflect the denser micro-structure in type I fibres.[22] The presence of fat is known to decrease MD measurements; however this is unlikely to be the reason we observed reduced MD in the hamstrings. The FF required to induce inaccuracies in diffusion measurements are estimated to be around 24%,[15] whereas the muscle FF in our study were below this (<10%). Furthermore, the use of STEAM with a long TM provides additional suppression of the olefinic fat peak so the limitations of spectral adiabatic inversion recovery-only fat suppression are addressed. Clinically, the fact that both FF and diffusion measurements differ between muscles is important. Firstly, it highlights the fact that these measures are sensitive to subtle differences between muscles that are not detectable visually from standard MR images. These differences may be of interest in studying healthy and diseased muscles. Secondly, researchers using these measures to investigate muscle disease should take these differences into account in their measurements so that their findings are not biased by normal differences between muscles.

Our study was subject to a number of limitations. The lack of a body composition score, such as body mass index, in our inclusion criteria is a limitation of our study. However body mass index is limited as a measure of obesity and frailty and so participant exclusion based on this criterion might also have biased the results by incorrectly classifying healthy participants as obese.[52,53] Analysis of FF and the diffusion data was only made on a single slice, raising the possibility of sampling variation if the muscle is inhomogeneous.[54] The 2-point Dixon imaging technique did not correct for T_2* effects, eddy currents, noise-related bias, or the spectral complexity of fat. However, multiple studies have failed to show that the errors inherent to 2-point Dixon confound fat measurements in muscle in either *ex-vivo* or *in-vivo* analysis.[33–35] Furthermore, two point Dixon correlates strongly with confounder-corrected fat quantitation methods and with spectroscopy.[34,35] Therefore, although improved accuracy in FF measurements could have been achieved using confounder corrected fat quantitation methods, there is no evidence to suggest that the relative differences in FF observed in this study were not related to genuine differences in muscle fat.

CONCLUSION

This study has shown excellent test–retest variability in MR-based FF and STEAM-based diffusion measurements as well as demonstrating significant differences between the hamstrings and quadriceps with these measures in the healthy thigh.

ACKNOWLEDGMENT

JDB is funded by a National Institute for Health Research (NIHR) (and Health Education England) Clinical Lectureship. AJG was funded by the NIHR. The research is supported by the NIHR infrastructure at Leeds. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We are grateful to radiographers Rob Evans and Brian Chaka for carrying out the MR studies and Dr Elizabeth Hensor for statistical advice.

REFERENCES

1. Ai T, Yu K, Gao L, Zhang P, Goerner F, Runge VM, et al. Diffusion tensor imaging in evaluation of thigh muscles in patients with polymyositis and dermatomyositis. *Br J Radiol* 2014; **87**: 20140261. doi: https://doi.org/10.1259/bjr.20140261
2. Qi J, Olsen NJ, Price RR, Winston JA, Park JH. Diffusion-Weighted imaging of inflammatory myopathies: polymyositis and dermatomyositis. *J. Magn. Reson. Imaging* 2008; **27**: 212–7. doi: https://doi.org/10.1002/jmri.21209
3. Sigmund EE, Baete SH, Luo T, Patel K, Wang D, Rossi I, et al. Mri assessment of the thigh musculature in dermatomyositis and healthy subjects using diffusion tensor imaging, intravoxel incoherent motion and dynamic DTI. *Eur Radiol* 2018; **28**: 5304–15. doi: https://doi.org/10.1007/s00330-018-5458-3
4. Edalati M, Hastings MK, Sorensen CJ, Zayed M, Mueller MJ, Hildebolt CF, et al. Diffusion tensor imaging of the calf muscles in subjects with and without diabetes mellitus. *J Magn Reson Imaging* 2018;: 1–11.
5. Zaraiskaya T, Kumbhare D, Noseworthy MD. Diffusion tensor imaging in evaluation of human skeletal muscle injury. *J. Magn. Reson. Imaging* 2006; **24**: 402–8. doi: https://doi.org/10.1002/jmri.20651
6. Giraudo C, Motyka S, Weber M, Karner M, Resinger C, Feiweier T, et al. Normalized STEAM-based diffusion tensor imaging provides a robust assessment of muscle tears in football players: preliminary results of a new approach to evaluate muscle injuries. *Eur Radiol* 2018; **28**: 2882–9. doi: https://doi.org/10.1007/s00330-017-5218-9
7. Okamoto Y, Mori S, Kujiraoka Y, Nasu K, Hirano Y, Minami M. Diffusion property differences of the lower leg musculature between athletes and non-athletes using 1.5T MRI. Magn Reson mater physics. *Biol Med* 2012; **25**: 277–84.
8. Okamoto Y, Kemp GJ, Isobe T, Sato E, Hirano Y, Shoda J, et al. Changes in diffusion tensor imaging (DTI) eigenvalues of skeletal muscle due to hybrid exercise training. *Magn Reson Imaging* 2014; **32**: 1297–300. doi: https://doi.org/10.1016/j.mri.2014.07.002
9. Froeling M, Oudeman J, Strijkers GJ, Maas M, Drost MR, Nicolay K, et al. Muscle changes detected with diffusion-tensor imaging after long-distance running. *Radiology* 2015; **274**: 548–62. doi: https://doi.org/10.1148/radiol.14140702
10. Yoon MA, Hong S-J, Ku MC, Kang CH, Ahn K-S, Kim BH. Multiparametric MR imaging of age-related changes in healthy thigh muscles. *Radiology* 2018; **287**: 235–46. doi: https://doi.org/10.1148/radiol.2017171316
11. Teruel JR, Cho GY, Moccaldi RTM, Goa PE, Bathen TF, Feiweier T, et al. Stimulated echo diffusion tensor imaging (STEAM-DTI) with varying diffusion times as a probe of breast tissue. *J Magn Reson Imaging* 2016;.
12. Noehren B, Andersen A, Feiweier T, Damon B, Hardy P. Comparison of twice refocused spin echo versus stimulated echo diffusion tensor imaging for tracking muscle fibers. *J. Magn. Reson. Imaging* 2015; **41**: 624–32. doi: https://doi.org/10.1002/jmri.24585
13. Mercier B, Granier P, Mercier J, Foucat L, Bielicki G, Pradere J, et al. Noninvasive skeletal muscle lactate detection between periods of intense exercise in humans. *Eur J Appl Physiol* 1998; **78**: 20–7. doi: https://doi.org/10.1007/s004210050382
14. Burakiewicz J, Hooijmans MT, Webb AG, Verschuuren J, Niks EH, Kan HE. Improved olefinic fat suppression in skeletal muscle DTI using a magnitude-based Dixon method. *Magn Reson Med* 2017; **159**: 152–9.
15. Williams SE, Heemskerk AM, Welch EB, Li K, Damon BM, Park JH. Quantitative effects of inclusion of fat on muscle diffusion tensor MRI measurements. *J. Magn. Reson. Imaging* 2013; **38**: 1292–7. doi: https://doi.org/10.1002/jmri.24045
16. Sigmund EE, Novikov DS, Sui D, Ukpebor O, Baete S, Babb JS, et al. Time-Dependent diffusion in skeletal muscle with the random permeable barrier model (RPBM): application to normal controls and chronic exertional compartment syndrome patients. *NMR Biomed* 2014; **27**: 519–28. doi: https://doi.org/10.1002/nbm.3087
17. Li K, Dortch RD, Welch EB, Bryant ND, Buck AKW, Towse TF, et al. Multi-parametric MRI characterization of healthy human thigh muscles at 3.0 T - relaxation, magnetization transfer, fat/water, and diffusion tensor imaging. *NMR Biomed* 2014; **27**: 1070–84. doi: https://doi.org/10.1002/nbm.3159
18. Fouré A, Properties D. And 3D architecture of human lower leg muscles assessed with and tractography: reproducibility and sensitivity to sex difference and intramuscular variability. *Radiology* 2018; **287**: 592–607.
19. Froeling M, Oudeman J, van den Berg S, Nicolay K, Maas M, Strijkers GJ, et al. Reproducibility of diffusion tensor imaging in human forearm muscles at 3.0 T in a clinical setting. *Magn. Reson. Med.* 2010; **64**: 1182–90. doi: https://doi.org/10.1002/mrm.22477
20. Johnson MA, Polgar J, Weightman D, Appleton D. Data on the distribution of fibre types in thirty-six human muscles. *J Neurol Sci* 1973; **18**: 111–29. doi: https://doi.org/10.1016/0022-510X(73)90023-3
21. Polgar J, Johnson MA, Weightman D, Appleton D. Data on fibre size in thirty-six human muscles. An autopsy study. *J Neurol Sci* 1973; **19**: 307–18.
22. Scheel M, von Roth P, Winkler T, Arampatzis A, Prokscha T, Hamm B, et al. Fiber type characterization in skeletal muscle by diffusion tensor imaging. *NMR Biomed* 2013; **26**: 1220–4. doi: https://doi.org/10.1002/nbm.2938
23. Schlaffke L, Rehmann R, Froeling M, Kley R, Tegenthoff M, Vorgerd M, et al. Diffusion tensor imaging of the human calf: variation of inter- and intramuscle-specific diffusion parameters. *J. Magn. Reson. Imaging* 2017; **46**: 1137–48. doi: https://doi.org/10.1002/jmri.25650
24. Yanagisawa O, Shimao D, Maruyama K, Nielsen M, Irie T, Niitsu M. Diffusion-Weighted magnetic resonance imaging of human skeletal muscles: gender-, age- and muscle-related differences in apparent diffusion coefficient. *Magn Reson Imaging* 2009; **27**: 69–78. doi: https://doi.org/10.1016/j.mri.2008.05.011
25. Kermarrec E, Budzik J-F, Khalil C, Le Thuc V, Hancart-Destee C, Cotten A. In vivo diffusion tensor imaging and tractography

of human thigh muscles in healthy subjects. *American Journal of Roentgenology* 2010; **195**: W352–W356. doi: https://doi.org/10.2214/AJR.09.3368
26. Fischmann A, Kaspar S, Reinhardt J, Gloor M, Stippich C, Fischer D. Exercise might bias skeletal-muscle fat fraction calculation from Dixon images. *Neuromuscular Disorders* 2012; **22**(Suppl 2): S107–S110. doi: https://doi.org/10.1016/j.nmd.2012.05.014
27. Fischmann A, Hafner P, Fasler S, Gloor M, Bieri O, Studler U, et al. Quantitative MRI can detect subclinical disease progression in muscular dystrophy. *J Neurol* 2012; **259**: 1648–54. doi: https://doi.org/10.1007/s00415-011-6393-2
28. Triplett WT, Baligand C, Forbes SC, Willcocks RJ, Lott DJ, DeVos S, et al. Chemical shift-based MRI to measure fat fractions in dystrophic skeletal muscle. *Magn. Reson. Med.* 2014; **72**: 8–19. doi: https://doi.org/10.1002/mrm.24917
29. Yao L, Yip AL, Shrader JA, Mesdaghinia S, Volochayev R, Jansen A V, et al. Magnetic resonance measurement of muscle T2, fat-corrected T2 and fat fraction in the assessment of idiopathic inflammatory myopathies. *Rheumatology* 2016; **55**: 441–9.
30. Grimm A, Meyer H, Nickel MD, Nittka M, Raithel E, Chaudry O, et al. Repeatability of Dixon magnetic resonance imaging and magnetic resonance spectroscopy for quantitative muscle fat assessments in the thigh. *J Cachexia Sarcopenia Muscle* 2018;.
31. Morrow JM, Sinclair CDJ, Fischmann A, Reilly MM, Hanna MG, Yousry TA, et al. Reproducibility, and age, body-weight and gender dependency of candidate skeletal muscle MRI outcome measures in healthy volunteers. *Eur Radiol* 2014; **24**: 1610–20. doi: https://doi.org/10.1007/s00330-014-3145-6
32. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract* 2004; **10**: 307–12. doi: https://doi.org/10.1111/j..2002.384.doc.x
33. Noble JJ, Keevil SF, Totman J, Charles-Edwards GD. *In vitro* and *in vivo* comparison of two-, three- and four-point Dixon techniques for clinical intramuscular fat quantification at 3 T. *Br J Radiol* 2014; **87**: 20130761. doi: https://doi.org/10.1259/bjr.20130761
34. Fischer MA, Nanz D, Shimakawa A, Schirmer T, Guggenberger R, Chhabra A, et al. Quantification of muscle fat in patients with low back pain: comparison of multi-echo MR imaging with single-voxel MR spectroscopy. *Radiology* 2013; **266**: 555–63. doi: https://doi.org/10.1148/radiol.12120399
35. Fischer MA, Pfirrmann CWA, Espinosa N, Raptis DA, Buck FM. Dixon-based MRI for assessment of muscle-fat content in phantoms, healthy volunteers and patients with achillodynia: comparison to visual assessment of calf muscle quality. *Eur Radiol* 2014; **24**: 1366–75. doi: https://doi.org/10.1007/s00330-014-3121-1
36. Gaeta M, Messina S, Mileto A, Vita GL, Ascenti G, Vinci S, et al. Muscle fat-fraction and mapping in Duchenne muscular dystrophy: evaluation of disease distribution and correlation with clinical assessments. *Skeletal Radiol* 2012; **41**: 955–61. doi: https://doi.org/10.1007/s00256-011-1301-5
37. Gaeta M, Scribano E, Mileto A, Mazziotti S, Rodolico C, Toscano A, et al. Muscle fat fraction in neuromuscular disorders: Dual-Echo Dual-Flip-Angle spoiled gradient-recalled MR imaging technique for Quantification—A feasibility study. *Radiology* 2011; **259**: 487–94. doi: https://doi.org/10.1148/radiol.10101108
38. Nozaki T, Tasaki A, Horiuchi S, Ochi J, Starkey J, Hara T, et al. Predicting Retear after repair of full-thickness rotator cuff tear: two-point Dixon MR imaging quantification of fatty muscle Degeneration—Initial experience with 1-year follow-up. *Radiology* 2016; **280**: 500–9. doi: https://doi.org/10.1148/radiol.2016151789
39. Merboldt K-D, Hanicke W, Frahm J. Self-Diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance* 1985; **64**: 479–86. doi: https://doi.org/10.1016/0022-2364(85)90111-8
40. Jellus V. Phase correction method. 2011; **7**: 936–B2.
41. Gold GE, Han E, Stainsby J, Wright G, Brittain J, Beaulieu C, Suh B, Sawyer-glover A. Musculoskeletal MRI at 3.0 T: relaxation times and image contrast. *American Journal of Roentgenology* 2004; **183**: 343–51. doi: https://doi.org/10.2214/ajr.183.2.1830343
42. Liu C-Y, McKenzie CA, Yu H, Brittain JH, Reeder SB. Fat quantification with ideal gradient echo imaging: correction of bias fromT1 and noise. *Magn. Reson. Med.* 2007; **58**: 354–64. doi: https://doi.org/10.1002/mrm.21301
43. Guidlines CDV. Criteria, and rules of thumb for Evalauting Normed and standardized assessment instruments in psychology. *Psychol Assess* 1994; **6**: 284–90.
44. Sigmund EE, Sui D, Ukpebor O, Baete S, Fieremans E, Babb JS, et al. Stimulated echo diffusion tensor imaging and SPAIR T 2 -weighted imaging in chronic exertional compartment syndrome of the lower leg muscles. *J. Magn. Reson. Imaging* 2013; **38**: 1073–82. doi: https://doi.org/10.1002/jmri.24060
45. Ma J. Dixon techniques for water and fat imaging. *J. Magn. Reson. Imaging* 2008; **28**: 543–58. doi: https://doi.org/10.1002/jmri.21492
46. Yu H, Shimakawa A, a MC, Brodsky E, Brittain JH, Reeder SB, et al. And simultaneous R2* estimation with Multi-frequency fat spectrum modeling. *Magn Reson Med* 2011; **60**: 1122–34.
47. Ponrartana S, Andrade K, Wren T. Repeatability of diffusion tensor imaging for the evaluation of lower extremity skeletal muscle. *AJR Am J Roentgenol* 2014; **202**: W567–74.
48. Karampinos DC, Baum T, Nardo L, Alizai H, Yu H, Carballido-Gamio J, et al. Characterization of the regional distribution of skeletal muscle adipose tissue in type 2 diabetes using chemical shift-based water/fat separation. *J Magn* 2012; **35**: 899–907.
49. Kumar D, Karampinos DC, MacLeod TD, Lin W, Nardo L, Li X, et al. Quadriceps intramuscular fat fraction rather than muscle size is associated with knee osteoarthritis. *Osteoarthritis and Cartilage* 2014; **22**: 226–34. doi: https://doi.org/10.1016/j.joca.2013.12.005
50. Schick F, Machann J, Brechtel K, Strempfer A, Klumpp B, Stein DT, et al. Mri of muscular fat. *Magn Reson Med* 2002; **47**: 720–7. doi: https://doi.org/10.1002/mrm.10107
51. Sinha S, Sinha U, Edgerton VR. In vivo diffusion tensor imaging of the human calf muscle. *J. Magn. Reson. Imaging* 2006; **24**: 182–90. doi: https://doi.org/10.1002/jmri.20593
52. Nevill AM, Stewart AD, Olds T, Holder R. Relationship between adiposity and body size reveals limitations of BMI. *Am J Phys Anthropol* 2006; **129**: 151–6. doi: https://doi.org/10.1002/ajpa.20262
53. Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. *Nutr Today* 2015; **50**: 117–28.
54. Miokovic T, Armbrecht G, Felsenberg D, Belavý DL. Heterogeneous atrophy occurs within individual lower limb muscles during 60 days of bed rest. *J Appl Physiol* 2012; **113**: 1545–59. doi: https://doi.org/10.1152/japplphysiol.00611.2012