Practical Approaches to Bone Marrow Fat Fraction Quantification Across Magnetic Resonance Imaging Platforms

Abstract

Background: Proton density fat fraction (PDFF) measurements can objectively identify bone marrow oedema and fat metaplasia in spondyloarthritis and may be valuable for the quantification of inflammation in multi-center clinical trials and routine practice. However, many centers do not have access to specialist methods for PDFF measurement. This is a barrier to implementation.

Purpose/Hypothesis: To determine the agreement between fat fraction (FF) measurements derived from (1) basic vendor-supplied sequences (2) basic sequences with offline correction and (3) specialist vendor-supplied methods.

Study type: Prospective.

Population/subjects/phantom/specimen/animal model: Two substudies with ten and five healthy volunteers.

Field strength/sequence: Site A: mDixon Quant (Philips 3T Ingenia); Site B: IDEAL and FLEX (GE 1.5T Optima MR450W); Site C: DIXON, with additional 5-echo gradient echo acquisition for offline correction (Siemens 3T Skyra); Site D: DIXON, with additional VIBE acqusitions for offline correction (Siemens 1.5T Avanto). The specialist method at site A was used as a standard to compare to the basic methods at sites B, C and D.

Assessment: Regions of interest were placed on areas of subchondral bone on FF maps from the various methods in each volunteer.

Statistical tests: Relationships between FF measurements from the various sites and Dixon methods were assessed using Bland-Altman analysis and linear regression.

Results: Basic methods consisting of IDEAL, LAVA FLEX and DIXON produced FF values that were linearly related to reference FF values (P<0.0001), but produced mean biases of up to 10%. Offline correction produced a significant reduction in bias in both substudies (P<0.001).

Data Conclusion: FF measurements derived using basic vendor-supplied methods are strongly linearly related to those derived using specialist methods but produce a bias

of up to 10%. A simple offline correction that is accessible even when the scanner has only basic sequence options can significantly reduce bias.

Keywords

Spondyloarthritis

Chemical shift imaging

Inflammation

Imaging biomarker

Introduction

MRI is used to detect inflammation in spondyloarthritis and can also be used to monitor treatment [1]. However, standard visual interpretation of MRI scans is expertise-dependent and somewhat subjective. Therefore, there is a need for an objective, ideally quantitative, method for identifying and quantifying inflammation in patients with spondyloarthritis [2].

One potential approach is to measure fat fraction (FF) in the bone marrow using Dixon-based methods [3], [4]. It has been shown that proton density fat fraction (PDFF) measurements can distinguish regions of oedema and fat metaplasia from normal bone marrow in the subchondral bone marrow of patients with spondyloarthritis [5]. The potential to identify both active inflammation (oedema) and structural damage (fat metaplasia) with a single acquisition is a significant advantage, whilst the speed of the acquisition means that quantification could be applied to multiple joints. Furthermore, PDFF measurements have been shown to be accurate and precise in the liver [4], [6]–[8] and in bone marrow [5], [9]–[12]. The reproducibility of PDFF measurements across sites [8], [12] suggests potential for use in multi-center studies, and for the development of thresholds which could be used to define and quantify disease.

However, a potential limitation of PDFF methods is the necessity for offline processing or for purchasing vendor-supplied 'specialist' methods where this processing is performed in-line. Specialist PDFF imaging methods minimize confounds on the measurement by correcting for T2*, the spectral complexity of fat and by minimizing T1 bias [5], [7], [13], [14]. Although these methods are commercially available, they can be expensive and not all imaging centers will have purchased them as options. Processing can be performed offline, but this can be difficult to implement at multiple sites, since phase data are not consistently available and constructed in varying fashions. This issue presents a significant problem for multi-center studies aiming to use FF measurements as endpoints, and for the development of clinical criteria. Most centers do though have access to basic methods for separating fat and water signals, meaning that simple 'signal' fat fraction measurements can be obtained. If the bias

introduced by these simpler methods is sufficiently small, they may be of use for clinical studies. It may also be possible to ‘correct’ the FF measurements obtained from these methods by acquiring additional images in addition to the base Dixon acquisition.

In this study, we aimed to investigate the agreement between basic Dixon methods and specialist methods, and to determine whether an offline ‘correction’ method making use of freely available sequences could improve this agreement.

Materials and Methods

This prospective cohort study received ethical approval (Queen Square Research Ethics Committee, London, UK, Ref 15/LO/1475). All volunteers gave their informed written consent prior to imaging.

Study Format

The results of two separate investigations are presented here:

Study 1: 10 participants were scanned at 3 sites for comparison of FF measurements using different methods.

Study 2: 5 participants were scanned at 3 sites in order to measure the repeatability and reproducibility of FF measurements using different methods. All scans were repeated 4 times. Acquisitions 1 and 2 were acquired sequentially with no change in participant setup. After acquisition 2, participants were removed from the scanner for a short time and participant setup in the scanner was started from the beginning. Acquisitions 3 and 4 were then acquired sequentially with no change in participant setup in-between. Acquisition 1 was used for inter-method comparisons.

Study Participants

Study 1:

Ten healthy participants (five male and five female) with no symptoms of sacroiliitis were recruited. Participants were selected to ensure a broad age range, and therefore a broad range of FF values, in the cohort (mean age 42 years, range 23 - 63 years). Imaging was performed between July 2017 and September 2017.

Study 2:

5 heathy participants (three male and two female; mean age 40 years, range 27 – 64 years) with no symptoms of sacroiliitis were recruited. A fat fraction phantom (see section – Phantoms) was additionally scanned at each site. Imaging was performed in September 2019.

Imaging Sites

All volunteers were scanned on three scanners in study 1 (each at a separate institution) over a four-month period, using proprietary and commercially available chemical-shift encoded (CSE) MRI sequences to separate fat and water signals. For study 2, Site B was not available. Instead a further site, Site D, was added. Acquisition details are provided in Table 1.

Site A (Studies 1 and 2): Scans were performed on a Philips (Best, Netherlands) 3T Ingenia scanner using the mDixon Quant commercial product (MDQ), yielding fat-only and water–only images and PDFF parameter maps. Note that this method has previously been validated in fat-water-bone phantoms and in patients with spondyloarthritis [5].

In study 2, an additional 3-D Gradient Echo acquisition, using 6 echo times, was acquired at this site; this acquisition yielded magnitude images only at each echo time.

Site B (Study 1 only): Scans were performed on a General Electric (Chicago, Illinois, USA) 1.5T Optima MR450W scanner using two separate commercial FF products: IDEAL and FLEX. IDEAL was acquired with three echo times and FLEX with two. Both methods yielded separated fat-only and water–only images.

Site C (Studies 1 and 2): Scans were performed on a Siemens (Erlangen, Germany) 3T Skyra scanner using the DIXON commercial product. Two echo times were used and separated fat-only and water–only images were produced. An additional 3-D Gradient Echo acquisition, using VIBE with 6 echo times, was acquired at this site; this acquisition yielded magnitude images only at each echo-time.

Site D (Study 2 only): Scans were performed on a Siemens (Erlangen, Germany) 1.5T Avanto scanner using the DIXON commercial product. Two echo times were used and separated fat-only and water–only images were produced. Three additional 3-D Gradient Echo acquisitions, using VIBE with 2 echo times, were acquired at this site; this acquisition yielded magnitude images only at each echo-time.

For all acquisitions, the flip angle was set to 3 degrees in order to minimize T1 bias. In all cases, images were acquired in a tilted coronal plane parallel to the long axis of the sacrum.

Phantoms

Fat-water phantoms were constructed with FF values varying between 0% and 70% in 10% increments, based on previous designs [5], [15]. The phantom consisted of twelve 50ml centrifuge tubes with varying fat volume percentages (including four with hydroxyapatite – the mineral constituent of bone - added), with FF values chosen to reflect the range of values observed in both normal and pathological bone marrow. Peanut oil was used as a surrogate for human fat as its spectrum is similar to that of human adipose tissue [16]. We adopted the approach of Gee et al. and used fat volume percentages as reference FF values [11]. For each tube, the appropriate volume of peanut oil was dispensed by weight, assuming the density of peanut oil as 0.916g/cm^3. Sodium dodecyl sulphate (SDS) (surfactant; Sigma-Aldrich, St Louis, Missouri, USA) was added to the peanut oil and gently mixed to form an initial emulsion to achieve a final SDS concentration in each phantom of 28mM. The appropriate volume of 3.0% weight/volume agar solution was first heated to boiling and then added to each tube. Each tube was thoroughly agitated and then gently inverted for approximately two minutes. The tubes cooled at room temperature and all formed a solid gel. In four tubes, hydroxyapatite powder was also added to the oil and by agitation prior to adding surfactant, to achieve a bone mineral density of 200mg/cm3 in each of these tubes. The tubes containing hydroxyapatite had fat fraction values of 0, 20, 40 and 60%. The phantoms were scanned at room temperature.

Post-processing

All post-processing was performed with in-house code written using MATLAB (MathWorks, Natick, Massachusetts, USA). For the MDQ scans at Site A, the IDEAL and FLEX scans at site B and the DIXON scans at site C, the water-only (W) and fat-only

(F) images were used to calculate a fat-fraction image (FF) using the formula FF = F / (W+F).

The additional data collected at site C in study 1, and at sites A, C and D in study 2, were fitted offline using a magnitude-only signal model, given by

$$S(t_n) = \left|\left(\rho_W + \rho_F \sum_{m=1}^{M} r_m exp\left(i2\pi f_{F,m} t_n\right)\right)\right| exp(-t_n R_2^*)$$

where $S(t_n)$ is the magnitude signal acquired at the nth echo time, t_n. The quantities ρ_W and ρ_F are the amplitudes of the water and fat components respectively, r_m is the relative amplitude of the m^{th} fat component peak, $f_{F,m}$ is the frequency separation between water and the m^{th} fat component and R_2^* is common to both water and fat. The model included a single water peak and six fat peaks. The fat and water images derived from the product Dixon acquisitions were used to provide a first guess solution for the offline fit (specifically these were specified as start points for a trust-region solver, implemented by the MATLAB function *lsqcurvefit*). This yielded water-only and fat-only magnitude images, from which FF parameter maps were calculated.

RoI Analysis

Four freehand regions of interest (RoIs) were placed on the subchondral bone adjacent to the sacroiliac joint on a single representative slice for each FF dataset by an experienced MRI physicist (**), as shown in Figure 1, taking care to ensure that the selected regions on datasets from each site sampled the same part of the joint.

Statistical Analysis

Study 1:

Data from all regions of interest were combined for statistical analysis. Linear regression was used to demonstrate the relationships between the different measures of FF. The MDQ data was taken as the reference standard measurement. Bland Altman analyses were used to assess the bias between different measurement methods. Regression lines were also fitted to the Bland Altman data plots to

investigate whether the bias between methods was uniform with mean FF value. Paired t-tests were used to test for differences between measurement methods. In all cases significance was assumed at a level of $p < 0.05$.

Study 2:

Data from all regions of interest were combined for statistical analysis. The repeatability of the measurement using each method was tested by comparing data from acquisition 1 with acquisition 2, and from acquisition 3 with acquisition 4. The within-scanner reproducibility of the measurement using each method was tested by comparing data from acquisition 1 with acquisition 3, and from acquisition 2 with acquisition 4. Bland Altman analyses were used to assess the bias and limits of agreement between pairs of measurements.

The Study 2 data was analysed in the same way as for study 1, with the MDQ data was taken as the reference standard measurement. Linear regression was used to demonstrate the relationships between the different measures of FF. Regression lines were also fitted to the Bland Altman data plots to investigate whether the bias between methods was uniform with mean FF value. Paired t-tests were used to test for differences between measurement methods. In all cases significance was assumed at a level of $p < 0.05$.

Results

Representative FF parameter maps acquired using the various imaging methods are shown in Figure 1. In study 2, the offline-reconstructed FF maps from Site D were poor, due to problems with image scalings between the VIBE acquisitions, and were not included in any further statistical analysis. See Online Supplemental Figure 1 and the Limitations section in the DISCUSSION on this point.

Study 1

There was a wide range of FF values measured within the volunteer cohort [range 39 - 85%; MDQ at Site A]. Figure 2 shows the MDQ FF values plotted against those from the base-level Dixon methods. Table 2 shows the results of performing a linear regression analysis using these data along with the mean bias and 95% limits of agreement from the Bland-Altman analysis. Figure 3 shows Bland Altman plots for comparison of the MDQ method at Site A with each of the other methods used at Sites B and C. FLEX and DIXON were similarly biased with mean differences of 0.09 (LAVA FLEX), 0.07 (IDEAL) and 0.071 (DIXON) when compared with MDQ. The offline fitted data had a mean difference of 0.024 compared with MDQ PDFF; this was significantly lower than for DIXON from the same machine ($P < 0.001$). Note that the IDEAL images from Site B were not useable for one participant because of artefacts.

The slopes of the regression lines fitted to the Bland Altman data are given in Table 2. The bias between MDQ and each of IDEAL, FLEX and DIXON was not uniform. In each case, the difference between the methods was larger at the lower end of the range of measured fat fraction values. The bias between the offline method and both FLEX and DIXON was also not uniform. The bias was not significantly correlated with the mean value for all comparisons between IDEAL, FLEX and DIXON and for the comparisons between the offline method and both MDQ and IDEAL.

In order to test whether the mean biases between pairs of methods were significant, the modulus of the difference between FF measurements was calculated for each datapoint and grouped by the pairwise comparison. Figure 4 shows a box-plot of the comparisons. The median of the difference between methods was significantly

different across all comparisons (P<0.001, Kruskal Wallis). Pairwise comparisons further showed that the median difference between MDQ and all of IDEAL, FLEX and DIXON was significantly larger than for the comparison between MDQ and the offline method (P < 0.001 for each comparison).

Study 2

Figure 5 shows Bland Altman plots for the repeatability measurements. Figure 6 shows Bland Altman plots for the reproducibility measurements. Table 3 shows the mean bias and limits of agreement from the Bland-Altman analyses for each comparison.

Online Supplemental Figure 1 shows the MDQ FF values plotted against those measured using the other methods. Online Supplemental Table 1 shows the results of linear regression and Bland-Altman analysis of these data. Online Supplemental Figure 2 shows Bland Altman plots for comparison of the MDQ method at Site A with the other methods. With similarity to the results of study 1, the offline methods compared well with MDQ method with mean differences of -0.0005 (Offline Site C), and 0.002 (Offline Site C). Again, with similarity to the results from study 1, the mean difference to MDQ for the offline fitted data at Site C was significantly lower than for DIXON at the same site (mean difference = 0.049), P<0.001. Online supplemental Figure 3 shows a box-plot of the mod differences for the comparisons to MDQ. The median of the difference between methods was significantly different across all comparisons (P<0.001, Kruskal Wallis). Pairwise comparisons further showed that the median difference between MDQ and DIXON at sites C and D was significantly larger than for the comparisons between MDQ and the offline methods (P < 0.001 for each comparison).

Phantoms

Data from the phantom experiments are shown in Online Supplemental Figures 4 and 5 for both inline (vendor-supplied) and offline Dixon processing methods. Online Supplemental Table 2 shows the results of linear regression and Bland-Altman analysis

of these data. In general, there was a small positive bias, of similar magnitude for both offline and offline methods. At Site A, the mean bias was similar for both online and offline methods. At Site B, FF values were accurate in the absence of bone, but in the presence of bone there was a bias of almost 20% in the 0% FF vial using the DIXON method. This was eliminated by the offline correction, although bias was marginally increased for intermediate FF values compared to the inline method. Bias at 0% FF was also observed at Site C, but at this site the offline correction performed poorly and increased bias.

Discussion

Fat fraction measurements potentially offer a simple and reproducible biomarker of inflammation in spondyloarthritis and could be of value in multi-center clinical trials as well as clinical practice. In order for FF measurements to be used in these settings, they need to be reproducible across a wide variety of MRI systems, preferably using freely available base-level sequences [17]. In this study, we showed strong positive relationships between FF values measured using base level methods and those using a previously-validated specialist method [5]. The data showed strong linear relationships between FF measurements from the various methods, although there was a degree of bias when comparing base level methods to the reference standard. Additionally, we found that the use of an offline 'correction' method – accessibly even when the scanner has only basic sequence options available - reduced bias compared to the reference standard. Data from the second study show that FF measurements using all of the methods included in this paper are highly repeatable and reproducible.

Choice of Method

Key to the performance of FF as a biomarker is dealing effectively with the major confounds on the measurement, namely T1- and T2*-relaxation and the spectral complexity of the fat signal [18]. Methods that meet these requirements – such as MDQ at site A - can be described as specialist methods. The signal model used for MDQ includes a 7-peak fat model and an R2* correction factor that is common to both the water and fat signals. Broadly equivalent specialist methods are available from the manufacturers of the other platforms used in this study (GE offers IDEAL IQ and Siemens offers qDIXON within their LiverLab package). Specialist methods have been demonstrated to have good linearity and negligible bias in the context of clinical assessment of liver fat [7]. The focus of this work, though, was to consider the situation of an imaging center that may wish to introduce quantitative methods into imaging protocol, but without such a specialist method being available to them. This is a common situation: in a current multi-site imaging study of bone marrow in spondyloarthritis patients (DyNAMISM, Research Registry identifier 2783) only 1 out

of 8 sites surveyed had a specialist FF method available on their imaging equipment. Specialist options come with significant cost, which may be unacceptable in multi-center studies where a number of different scanners are used.

If a specialist FF measurement method is not available, then it may be necessary to rely on a basic, and more generally available, method for FF measurement. The inline methods used at sites B, C and D use fewer echo times and simpler signal models that do not account for all of the confounds stated above. IDEAL, based on the work of Reeder *et al.* [18][19], and FLEX both model fat as a single peak with no R2* correction in the signal model. The DIXON method includes a multi-peak fat spectrum, but no R2* correction.

Comparisons between methods

The basic methods used here yielded FF measurements with a mean bias of up to approximately 10%. In the context of assessment of bone marrow disease, this may be tolerable: Bray et al showed that the median FF in areas of normal bone marrow in spondyloarthritis patients was 47% compared to 27% in areas of oedema and 82% in areas of fat metaplasia. However, bias between the different methods makes it more difficult to apply global quantitative FF thresholds to identify regions of disease, and thus reductions in bias are desirable. The offline method used at site C yielded FF maps that were less biased when compared with MDQ. Further, in the second study, this offline methods at both sites A and C were also less biased when compared with MDQ than were the basic methods. The offline approach was designed to use simpe methods that should be accessible at a large number of sites even with access to only basic scanner functionality. As a result we chose to use only magnitude data and to use whatever inline FF measurement was available as a first guess solution in order to minimize fat-water swap artefacts in the resulting FF maps. The results suggest that using an offline analysis may be an accessible approach where specialist methods are not available.

Repeatability and Reproducibility

The repeatability of all methods was excellent with a mean bias of less than 0.5 percentage points FF and the limits of agreement giving a range of 2-3 percentage points FF. the repeatability analysis re-used the same RoIs for the comparisons between repeated measures. The reproducibility was also good. Mean bias was largest for the MDQ method at 0.7 percentage points FF. The limits of agreements generally had larger ranges than for the repeatability measure (7 -10 percentage points FF). RoIs were re-drawn for the reproducibility comparisons and the larger LoA range is likely to be largely due to errors in accurately reproducing RoIs in the same region of bone marrow. The strong linear correlations between methods, and the good repeatability and reproducibility demonstrated in study 2, suggest that the basic methods should be capable of showing differences between normal and abnormal bone marrow, and monitoring disease progression within subjects.

Phantom data

The use of phantoms scanned across multiple scanners (Study 2) enabled assessment of the size and direction of bias at each of the sites. In general, both inline and offline processing methods performed well, although there was a small positive bias compared to reference FF values, and the offline method failed at Site D. Interestingly, the use of inline (vendor-supplied) methods generated a substantial bias at 0% FF values at two sites, which could be corrected using the offline method. This bias likely arises because signal loss due to T2* decay is greater in the presence of bone and incorrectly ascribed to signal loss due to chemical shift by simple two-point techniques. However, care should be taken in the interpretation of these phantom data. The phantoms were not temperature regulated and thus were scanned at room temperature. The shift in the frequency of the water peak at room temperature compared with body temperature means that the model functions will be incorrect for room temperature acquisitions and this is likely to affect the returned FF values [20].

Limitations

One limitation of this work is that in Study 1 the offline correction was performed at Site C but not Site B due to difficulties with performing an acquisition with multiple echo times at the latter site. In Study 2, a similar problem was faced at site D where it was only possible to set up an acquisition with 2 echo times. In order to address this, 3 separate scans were acquired to yield 6 echo times in total. However, on processing the data we found that the image scaling between acquisition was not identical and so the data were not suitable for fitting to the offline model (fitting produced essentially nonsense values). This problem could have been be resolved by more careful optimisation of the acquisition to ensure that the image scaling was identical for each acquisition. However, for this work we aimed to use methods which were included as standard on each of the scanners and could be easily applied. So the problems with acquiring and fitting to offline data at sites B and D represent a limitation in flexibility.

It would also have been desirable to have performed the scans on different field strengths for each of the three vendors, but this would have presented an unacceptable burden for the volunteers involved.

Conclusions

In conclusion, FF measurements derived using basic vendor-supplied methods are strongly linearly related to those derived using specialist methods, but produce a bias of up to 10%. A simple offline correction that can be applied to multi-echo gradient echo data, and where all echos are acquired in a single acquisition, may be accessible even when the scanner has only basic sequence options can significantly reduce bias.

References

[1] R. G. W. Lambert *et al.*, “Defining active sacroiliitis on MRI for classification of axial spondyloarthritis: update by the ASAS MRI working group.,” *Ann. Rheum. Dis.*, vol. 75, no. 11, pp. 1958–1963, Nov. 2016.

[2] M. A. Hall-Craggs, T. J. Bray, and A. P. Bainbridge, “Quantitative imaging of inflammatory disease: are we missing a trick?,” *Ann. Rheum. Dis.*, vol. 77, no. 11, pp. 1689–1691, 2018.

[3] W. T. Dixon, “Simple proton spectroscopic imaging.,” *Radiology*, vol. 153, no. 1, pp. 189–194, 1984.

[4] H. H. Hu *et al.*, “ISMRM workshop on fat-water separation: Insights, applications and progress in MRI,” *Magn. Reson. Med.*, vol. 68, no. 2, pp. 378–388, 2012.

[5] T. J. P. Bray, A. Bainbridge, S. Punwani, Y. Ioannou, and M. A. Hall-Craggs, “Simultaneous Quantification of Bone Edema/Adiposity and Structure in Inflamed Bone Using Chemical Shift-Encoded MRI in Spondyloarthritis,” *Magn. Reson. Med.*, vol. 79, no. 2, pp. 1031–1042, 2018.

[6] S. B. Reeder, H. H. Hu, and C. B. Sirlin, “Proton density fat-fraction: A standardized mr-based biomarker of tissue fat concentration,” *J. Magn. Reson. Imaging*, vol. 36, no. 5, pp. 1011–1014, 2012.

[7] T. Yokoo *et al.*, “Linearity, Bias, and Precision of Hepatic Proton Density Fat Fraction Measurements by Using MR Imaging: A Meta-Analysis,” *Radiology*, 2018.

[8] D. Hernando *et al.*, “Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5T and 3T using a fat–water phantom,” *Magn. Reson. Med.*, vol. 77, no. 4, pp. 1516–1524, 2017.

[9] S. Ruschke *et al.*, “Measurement of vertebral bone marrow proton density fat fraction in children using quantitative water–fat MRI,” *Magn. Reson. Mater. Phys. Biol. Med.*, 2017.

[10] T. Baum *et al.*, “Anatomical variation of age-related changes in vertebral bone marrow composition using chemical shift encoding-based water-fat magnetic resonance imaging,” *Front. Endocrinol.*, 2018.

[11] R. Gee, S, Nguyen, J, Marquez, C, Heunis, J, Lai, A, Wyatt, C, Han, M, Kazakia, G, Burghardt, A, Karampinos, D, Carballido-Gamio, J, Krug, “Validation of Bone Marrow Fat Quantification in the Presence of Trabecular Bone Using MRI,” *J Magn Reson Imaging*, vol. 42, no. 2, pp. 539–544, 2014.

[12] F. C. Schmeel *et al.*, “Proton density fat fraction MRI of vertebral bone marrow: Accuracy, repeatability, and reproducibility among readers, field strengths, and imaging platforms,” *J. Magn. Reson. Imaging*, 2019.

[13] C. D. G. Hines, H. Yu, A. Shimakawa, C. A. McKenzie, J. H. Brittain, and S. B. Reeder, “T1 independent, T2* corrected MRI with accurate spectral modeling for quantification of fat: Validation in a fat-water-SPIO phantom,” *J. Magn. Reson. Imaging*, vol. 30, no. 5, pp. 1215–1222, 2009.

[14] I. S. Idilman *et al.*, “Quantification of liver, pancreas, kidney, and vertebral body MRI-PDFF in non-alcoholic fatty liver disease,” *Abdom. Imaging*, vol. 40, no. 6, 2015.

[15] C. D. G. Hines, H. Yu, A. Shimakawa, C. A. McKenzie, J. H. Brittain, and S. B. Reeder, “T1 independent, T2* corrected MRI with accurate spectral modeling for quantification of fat: Validation in a fat-water-SPIO phantom,” *J. Magn. Reson. Imaging*, vol. 30, no. 5, pp. 1215–1222, 2009.

[16] S. Ozcan, M, Seven, “Physical and chemical analysis and fatty acid composition of peanut, peanut oil, and peanut butter from COM and NC-7 cultivars.,” *Grasas Aceitis*, vol. 54, pp. 12–18, 2003.

[17] J. P. B. O’Connor *et al.*, “Imaging biomarker roadmap for cancer studies,” *Nat. Rev. Clin. Oncol.*, vol. 14, no. 3, pp. 169–186, 2017.

[18] S. B. Reeder, I. Cruite, G. Hamilton, and C. B. Sirlin, “Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy,” *Journal of Magnetic Resonance Imaging*, vol. 34, no. 4. pp. 729–749, 2011.

[19] S. B. Reeder *et al.*, “Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): Application with fast spin-echo imaging,” *Magn. Reson. Med.*, vol. 54, no. 3, pp. 636–644, 2005.

[20] D. Hernando, S. D. Sharma, H. Kramer, and S. B. Reeder, “On the Confounding Effect of Temperature on Chemical Shift-Encoded Fat Quantification,” *Magn. Reson. Med. Off. J. Soc. Magn. Reson. Med. Soc. Magn. Reson. Med.*, vol. 72, no. 2, pp. 464–470, Aug. 2014.

Tables

Table 1: Acquisition parameters for each site. mDixon Quant, IDEAL, FLEX and DIXON are all commercial imaging products that yield fat-only and water-only images. The 3D VIBE acquisition was used for offline reconstruction of fat-only and water-only images.

Site and Scanner	Study	Sequence / Product	TR (ms)	Flip Angle (°)	Number of Echoes	TE (ms)
A: 3T Philips Ingenia	1 and 2	mDixon Quant	6.9	3	6	1.67 + n.0.9
	2	ME FFE	7.4	3	6	1.33 + n.1.0
B: 1.5T GE Optima MR450W	1	IDEAL	8.34	3	3	Minimum – not user definable
	1	FLEX	6.41	3	2	In-phase / out-phase
C: 3T Siemens Skyra	1 and 2	DIXON	8.0	3	2	In-phase / out-phase
	1 and 2	3D VIBE	10.0	3	6	1.23 + n.1.37
D: 1.5T Siemens Avanto	2	DIXON	7.0	3	2	In-Phase / out phase
	2	3D VIBE	11.8	3	2	1.89, 3.28
	2	3D VIBE	11.8	3	2	4.67, 6.06
	2	3D VIBE	11.8	3	2	7.45, 9.34

Table 2: Results of linear correlations and Bland-Altman Analyses. The data from each imaging method was used to calculate fat fraction (FF). Linear regression and Bland-Altman analyses were performed for each pairwise comparison of methods. The data points in each Bland-Altman plot were fitted with a linear regression. For each linear regression, the slope, intercept, r^2 and p-value for the significance of the correlation are quoted. For the Bland-Altman analyses, the mean bias and limits of agreement (LoA) are quoted.

Comparison	**Linear Regression**				**Bland Altman**			
	Slope	**Intercept**	**R^2**	***P*-value**	**Mean Bias**	**LoA**	**Slope**	***P*-value**
MDQ vs IDEAL	1.14	-0.15	0.96	< 0.0001	0.07	[0.00, 0.14]	-0.15	< 0.05
MDQ vs FLEX	1.25	-0.24	0.92	< 0.0001	0.09	[-0.02, 0.20]	-0.27	< 0.005
MDQ vs DIXON	1.15	-0.16	0.97	< 0.0001	0.07	[0.01, 0.14]	-0.14	< 0.01
MDQ vs Offline	0.99	0.02	0.95	< 0.0001	0.02	[-0.03, 0.08]	-0.004	> 0.99
IDEAL vs FLEX	1.10	0.07	0.96	< 0.0001	-0.03	[-0.33, 0.27]	-0.12	> 0.1
IDEAL vs DIXON	0.96	0.02	0.94	< 0.0001	-0.04	[-0.31, 0.22]	0.002	> 0.99
IDEAL vs Offline	0.84	0.13	0.93	< 0.0001	-0.09	[-0.37, 0.19]	0.13	> 0.2
FLEX vs DIXON	0.85	0.09	0.90	< 0.0001	-0.02	[-0.12, 0.09]	0.12	> 0.4
FLEX vs Offline	0.73	0.20	0.89	< 0.0001	-0.06	[-0.19, 0.06]	0.25	< 0.05
DIXON vs Offline	0.73	0.20	0.89	< 0.0001	-0.05	[-0.10, 0.01]	0.15	< 0.0005

Table 3: Results of linear correlations and Bland-Altman Analyses for Repeatability and Reproducibility of the measurements in Study 2. The data from each imaging method was used to calculate fat fraction (FF). Bland-Altman analyses were performed for each pairwise comparison of data. The mean bias and limits of agreement (LoA) are quoted.

Bland – Altman FF analyses	**Repeatability**		**Reproducibility**	
	Mean Bias	**LoA**	**Mean Bias**	**LoA**
Site A: 3T MDQ	0.0004	[0.008, -0.007]	-0.007	[0.032, -0.046]
Site C: 3T DIXON	0.0003	[0.010, -0.010]	0.0012	[0.024, 0.049]
Site D: 1.5T DIXON	0.003	[0.019, -0.012]	0.004	[0.035, -0.027]
Site A: Offline	0.003	[0.019, -0.013]	-0.006	[0.044, -0.055]
Site C: Offline	0.003	[0.016, -0.011]	0.003	[0.045, -0.038]

Figure Legends

Figure 1: Representative Fat fraction (FF) maps (computed from each of the imaging methods used in the study. FF maps: a) Site A, mDixon Quant; (b) Site B, IDEAL; (c) Site B, FLEX; (d) Site C, DIXON; (e) Site C, Offline reconstruction; f) Site A, offline reconstruction; g) Site D, DIXON; h) Site D, offline reconstruction.

Figure 2: Scatter plot of fat fraction values measured in Study 1 and computed using the basic methods from sites B and C and the offline method from site C plotted against values from the specialist method at site A. All data from all subjects are included.

Figure 3: Bland Altman plots for the comparisons of FF measurement using mDixon Quant (MDQ) with each other method used in Study 1. The mean difference and limits of agreement are shown for each comparison.

Figure 4: Box and whisker plots showing the modulus of the difference between fat fraction measured using pairs of methods from Study 1.

Figure 5: Bland-Altman plots showing the Repeatability of the methods in Study 2

Figure 6: Bland-Altman plots showing the Reproducibility of the methods in Study 2