Appendix 1

The signal model in Eq. [1] has n+4 unknown variables, where n is the number of echoes and corresponding signal measurements. Determination of n+4 unknowns from an n-point Dixon dataset therefore requires more information than a single voxel's signal measurements. One constraint that can be imposed to account for M_0 is that the signal be normalized to the sum of its values, given by S_{norm},

$$S_{norm}\left(TE_i\right) = wS\left(TE_i\right) \quad [3]$$

where the normalization weight, w, is given by

$$w\sum_{i=1}^{n} S\left(TE_i\right) = 1 \quad [4]$$

$$w = \frac{1}{\sum_{i=1}^{n} S\left(TE_i\right)} \quad [5]$$

The values of S_{norm} will be independent of M_0 as the quantity M_0 cancels out in the w*S multiplication, so any value of M_0 can be used or the quantity can be omitted altogether. This normalization emphasizes that the shape of the signal is of interest rather than specific values and reduces the problem to n+3 unknowns. An additional assumption is made that more than one voxel can be selected with the same type of fat-water signal deviations (a_i) but allowing for these voxels to have different proportions of fat (f), or alternatively stated, that multiple voxels have signals with the same type of fat and the same B_0 field but varied amounts of fat. Voxels in the same tissue type within the same local region are selected to approximately satisfy this assumption. Last, an assumption is made that the scaling factor, k, is globally applied to all voxels. Note that each voxel still has independent T_2* and f values. With these assumptions, the number of unknowns becomes n+2m+1, where m is the number of voxels selected. For the 6-echo Dixon acquisition used in this work, n=6, and thus the problem has 7+2m unknowns and 6m measurements. As few as two voxels can be used to form a sufficient system of equations for determining the unknown variable values. Small regions of interest (ROIs) with approximately 5–10 voxels were used to determine corrections in this work.

Numerical determination of the unknown variables is approached as a parameter estimation problem using a regularized and constrained iterative least squares fitting approach. The objective function is given by

$$\Psi\left(\theta\right) = \left\| S_{meas,norm} - S_{model,norm}\left(\theta\right)\right\|_2^2 + \lambda R\left(S_{corr,norm}\right) \quad [6]$$

where θ is the vector of model parameters to be estimated, $R\left(S_{corr,norm}\right)$ is a penalty for deviations of $S_{corr,norm}$ from a pure normalized exponential decay, and λ is the weighting of the penalty term. The regularization function, R(), determines the sum of squared difference between the corrected signal and the best-fit normalized exponential decay function, which is a straightforward single-parameter fitting sub-problem. By selecting voxels with low fat content such as lean muscle tissue, the properly re-scaled data would be expected to have low deviation from a pure exponential decay and therefore parameter combinations should include a value of k that brings the corrected signal closer to a pure exponential decay. Note that voxels with low fat are preferred for this regularization as opposed to pure (or nearly pure) fat voxels given that pure fat can have multiple peaks and complicated dephasing that makes such a regularization difficult to reliably formulate rather than following an approximately pure exponential decay. The value $\lambda=10^{-6}$ was empirically chosen for use in phantom data and $\lambda=10^{-1}$ was chosen for *in vivo* data based on tuning the parameter using one phantom dataset and one traveling control dataset from each scanner. Constraints on model parameters were chosen as follows: T_2* >0, f ≥0, $-1 \leq a_i \leq 1$, $0.5 \leq k \leq 1.5$. Gradient descent optimization was used for parameter estimation (MATLAB R2023a, Mathworks, Inc.).

For validation experiments, the "Fat+Water" (FW) correction was compared to a "Water Only" (WO) correction. The WO correction used a simplified form of Eq. [1], where f=0 and a_i variables are omitted, yielding a model with only m+1 unknown parameters and 6m measurements. For the WO correction, voxels in an ROI containing no fat, e.g., agarose gel tubes, were averaged together, giving m=1, and used to estimate k and a single T_2* value. FW corrections were obtained using ROIs in fat-containing voxels, e.g., lean skeletal muscle or the phantom tubes with peanut oil mixtures (5% and 15% tubes).

References

52. Zhao YZ, Gan YG, Zhou JL, Liu JQ, Cao WG, Cheng SM, Bai DM, Wang MZ, Gao FQ, Zhou SM. Accuracy of multi-echo Dixon sequence in quantification of hepatic steatosis in Chinese children and adolescents. World J Gastroenterol 2019;25:1513-23.

https://dx.doi.org/10.21037/qims-24-287

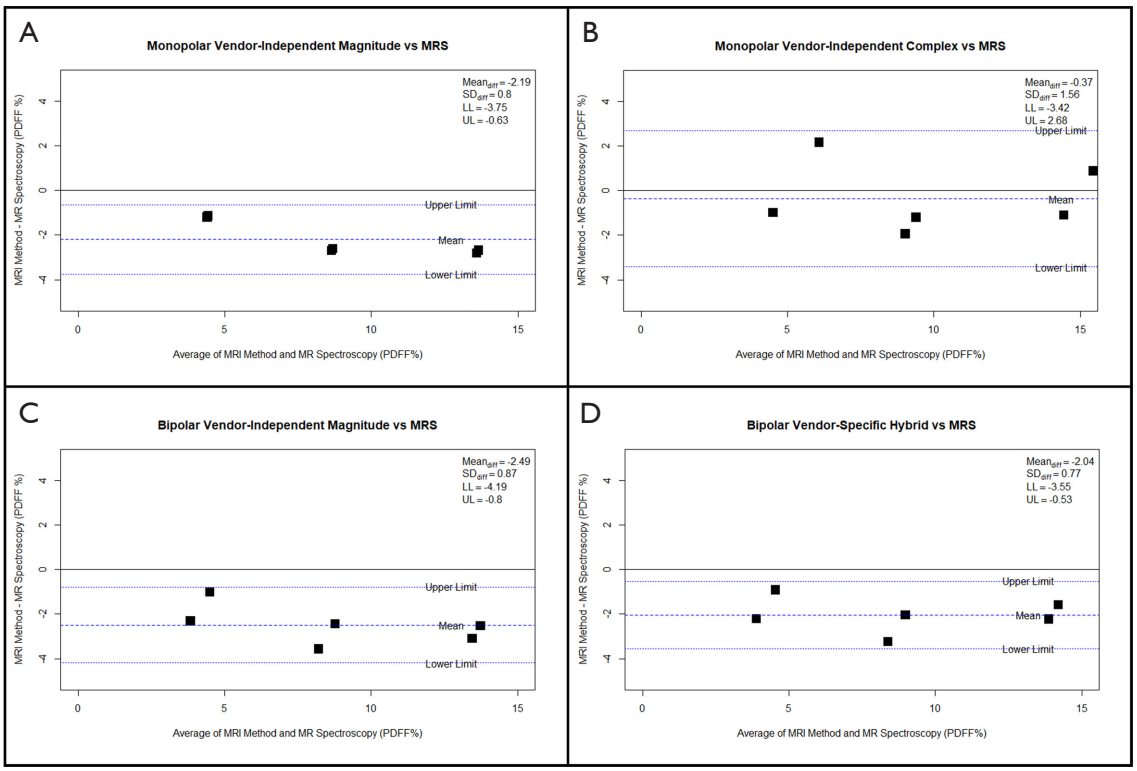

Figure S1 Bland-Altman plot for comparison of 6-point Dixon acquisition and processing methods to MRS in phantom for Site 1, scanner A. The MRS sequence used was a single voxel method, STEAM, similar to a previous report (52) with five TE values (12, 24, 36, 48, 72 ms) to support R2 correction, a TR of 3,000 ms, and a flip angle of 90 degrees. The voxel size was 30×30×40 mm^3 and was placed in each phantom tube individually (one acquisition for each tube). The vendor's inline processing was used to generate PDFF estimates. Values from left and right positions averaged over 15 scans were used. Note that the left and right positions provided similar values for (A), but with larger variations in (B), (C), (D) that could degrade reproducibility when obtaining PDFF from different limbs (e.g., operated limb could be left or right for various patients). Although the monopolar vendor-independent complex method had the lowest bias, there was notably greater variability between the two positions than other methods. LL, lower limit; UL, upper limit; SDdiff, standard deviation of differences between MRI-estimated PDFF and MRS-estimated PDFF; SD, standard deviation; MRI, magnetic resonance imaging; MRS, magnetic resonance spectroscopy; STEAM, stimulated echo acquisition mode; TE, echo time; PDFF, proton density fat fraction; TR, repetition time.

Table S1 Phantom proton density fat fraction (PDFF) mean and standard deviation across acquisition and processing methods

Metric	Reference fat fraction	Monopolar vendor-independent magnitude (N=6)			Monopolar vendor-independent complex (N=6)			Bipolar vendor-independent magnitude (N=6)			Bipolar vendor-specific (N=6)		
		L	R	L & R	L	R	L & R	L	R	L & R	L	R	L & R
PDFF, mean (SD), %	5%	3.9 (0.7)	3.8 (0.3)	3.8 (0.5)	4.0 (0.7)	7.2 (0.2)	5.6 (1.7)	4.0 (0.7)	2.7 (0.3)	3.3 (0.8)	4.1 (0.1)	2.8 (0.3)	3.4 (0.7)
	10%	7.3 (0.3)	7.4 (0.1)	7.3 (0.2)	8.8 (0.4)	8.1 (0.2)	8.4 (0.5)	7.6 (0.2)	6.4 (0.1)	7.0 (0.6)	8.0 (0.2)	6.8 (0.2)	7.4 (0.6)
	15%	12.2 (0.5)	12.3 (0.3)	12.2 (0.4)	15.9 (0.5)	13.9 (0.3)	14.9 (1.1)	12.5 (0.6)	11.9 (0.5)	12.2 (0.6)	13.4 (0.2)	12.8 (0.1)	13.1 (0.4)
ICC	All	0.98	1.00	0.99	0.99	1.00	0.94	0.99	0.99	0.99	1.00	1.00	1.00

PDFF, proton density fat fraction; L, left; R, right; ICC, intraclass correlation coefficient; SD, standard deviation.

https://dx.doi.org/10.21037/qims-24-287

Table S2 Evaluation of phantom inter-site PDFF mean and reproducibility before and after scaling correction (N=6)

Metric	Reference fat fraction	Phantom (without correction)	Phantom (water only correction)	Phantom (fat and water correction)
PDFF, mean (SD), %	5%	3.1 (1.4)	3.2 (0.5)	3.2 (0.6)
	10%	5.4 (1.6)	6.4 (0.6)	6.5 (0.7)
	15%	8.8 (2.5)	10.2 (1.2)	10.3 (1.2)
ICC	All	0.80	0.98	0.98

Note that these data include the scan used for parameter tuning with consistent trends as observed in Table 4. PDFF, proton density fat fraction; SD, standard deviation; ICC, intraclass correlation coefficient.

Table S3 Phantom PDFF mean and intra-scanner repeatability

Metric	Reference fat fraction	Site 1, scanner A (N=15)			Site 2, scanner C (N=15)			Site 3, scanner D (N=14)			Site 3, scanner E (N=15)		
		L	R	L & R	L	R	L & R	L	R	L & R	L	R	L & R
PDFF, mean (SD), %	5%	4.0 (0.9)	4.1 (0.6)	4.0 (0.8)	2.8 (0.2)	2.7 (0.3)	2.7 (0.3)	2.8 (0.3)	3.1 (0.3)	2.9 (0.3)	3.1 (0.3)	3.0 (0.2)	3.1 (0.3)
	10%	7.3 (0.5)	7.3 (0.2)	7.3 (0.4)	5.8 (0.3)	5.7 (0.3)	5.7 (0.3)	5.8 (0.2)	6.3 (0.2)	6.1 (0.2)	6.5 (0.4)	6.3 (0.3)	6.4 (0.4)
	15%	12.2 (0.4)	12.1 (0.4)	12.2 (0.4)	9.1 (0.3)	9.0 (0.3)	9.1 (0.3)	8.8 (0.3)	9.5 (0.2)	9.1 (0.2)	9.6 (0.7)	9.3 (0.3)	9.5 (0.5)
ICC	All	0.98	0.99	0.98	1.00	1.00	0.99	1.00	1.00	0.99	0.99	1.00	0.99

PDFF, proton density fat fraction; ICC, intraclass correlation coefficient; SD, standard deviation; L, left; R, right.

Table S4 Phantom PDFF mean, cross-phantom variation, and reproducibility: inter-vendor and inter-site

Metric	Reference fat fraction	Intra-site (Site 1), same vendor (scanner A), cross-phantom (N=3)	Intra-site (Site 1), same vendor (scanner B), cross-phantom (N=3)	Intra-site (Site 1), inter-vendor, same phantom (N=6)	Inter-site (Sites 1, 2), same vendor, same phantom (phantom 2) (N=2)	Inter-site (Sites 1, 3), same vendor, same phantom (phantom 3) (N=2)	Inter-site (Sites 1, 2, 3), inter-vendor (N=3)
PDFF, mean (SD), %	5%	4.4 (0.7)	3.4 (0.3)	3.9 (0.7)	3.1 (0.5)	3.4 (0.5)	3.3 (0.7)
	10%	7.3 (0.1)	6.7 (0.5)	7.0 (0.5)	6.2 (0.7)	6.7 (0.5)	6.4 (0.8)
	15%	10.4 (0.9)	10.4 (0.6)	10.4 (0.5)	9.4 (0.8)	10.1 (1.0)	10.2 (1.7)
ICC	All	0.94	0.98	0.96	1.00	0.99	0.97

Variability in PDFF across phantoms on the same scanner at Site 1 measured by SD was between 0.1–0.9% PDFF and ICC values for the two scanners were 0.94 (scanner A) and 0.98 (scanner B). PDFF, proton density fat fraction; SD, standard deviation; ICC, intraclass correlation coefficient.

https://dx.doi.org/10.21037/qims-24-287

Table S5 Evaluation of traveling control PDFF mean and reproducibility before and after scaling correction (N=4)

Metric	Muscle group	Traveling control (without correction)	Traveling control (fat and water correction)
PDFF, mean (wSD), %	HL	7.9 (1.5)	6.6 (1.2)
	HR	7.8 (1.6)	6.5 (1.0)
	QL	6.6 (1.5)	5.5 (0.9)
	QR	6.7 (1.5)	5.6 (0.7)
	ML	9.0 (1.4)	7.9 (0.9)
	MR	8.3 (1.5)	7.2 (0.8)
ICC	All	0.70	0.92

Note that these data include the scan used for parameter tuning with consistent trends as observed in Table 5. HL, hamstrings left; HR, hamstrings right; QL, quadriceps left; QR, quadriceps right; ML, medial left; MR, medial right; PDFF, proton density fat fraction; ICC, intraclass correlation coefficient; wSD, within-subject standard deviation.

Table S6 *In vivo* PDFF mean, intra-site repeatability, and inter-site reproducibility

Metric	Muscle Group	Site 1 (N=8)	Site 2 (N=11)	Site 3 (N=16)	Inter-site (N=5)
PDFF, mean (wSD), %	HL	7.8 (0.9)	5.7 (0.4)	5.9 (0.6)	6.6 (1.2)
	HR	7.3 (0.3)	5.4 (0.4)	5.8 (0.6)	6.5 (1.0)
	QL	6.5 (0.9)	4.9 (0.2)	5.2 (0.5)	5.5 (0.9)
	QR	6.0 (0.4)	4.8 (0.2)	5.3 (0.3)	5.6 (0.8)
	ML	8.3 (0.8)	7.0 (0.3)	8.2 (1.0)	7.9 (0.9)
	MR	7.1 (0.2)	6.1 (0.2)	7.2 (0.7)	7.2 (0.7)
ICC	All	0.96	0.98	0.95	0.92

wSD, within-subject standard deviation; PDFF, proton density fat fraction; ICC, intraclass correlation coefficient; HL, hamstrings left; HR, hamstrings right; QL, quadriceps left; QR, quadriceps right; ML, medial left; MR, medial right.

https://dx.doi.org/10.21037/qims-24-287