


# Accurate simultaneous quantification of liver steatosis and iron overload in diffuse liver diseases with MRI

Manuela França <sup>1</sup>, Ángel Alberich-Bayarri,<sup>2</sup> Luis Martí-Bonmatí,<sup>2</sup> Pedro Oliveira,<sup>3,4</sup> Francisca Emanuel Costa,<sup>5</sup> Graça Porto,<sup>6,7,8</sup> José Ramón Vizcaíno,<sup>5</sup> Javier Sanchez Gonzalez,<sup>9</sup> Eduardo Ribeiro,<sup>1</sup> João Oliveira,<sup>1</sup> Helena Pessequeiro Miranda<sup>4,10</sup>

<sup>1</sup>Imaging Department, Centro Hospitalar do Porto, Largo Prof Abel Salazar, 4099-001 Porto, Portugal

<sup>2</sup>Radiology Department, Hospital Universitario y Politécnico La Fe and Biomedical Imaging Research Group (GIBI230), Valencia, Spain

<sup>3</sup>Population Studies Department, Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal

<sup>4</sup>Epidemiology Research Unit (EPIUnit), Institute of Public Health of the University of Porto, Porto, Portugal

<sup>5</sup>Pathology Department, Centro Hospitalar do Porto, Porto, Portugal

<sup>6</sup>Hematology Department, Centro Hospitalar do Porto, Porto, Portugal

<sup>7</sup>I3S, Instituto de Investigação e Inovação em Saúde, Porto, Portugal

<sup>8</sup>IBMC, Institute for Molecular and Cell Biology, Porto, Portugal

<sup>9</sup>MR Clinical Science, Philips Healthcare, Madrid, Spain

<sup>10</sup>Liver and Pancreas Transplantation Unit and Medicine Department, Centro Hospitalar do Porto, Porto, Portugal

## Abstract

**Purpose:** To evaluate the diagnostic performances of 3 Tesla multi-echo chemical shift-encoded gradient echo magnetic resonance (MECSE-MR) imaging to simultaneously quantify liver steatosis and iron overload in a wide spectrum of diffuse liver diseases having biopsy as reference standard.

**Methods:** MECSE-MR-acquired images were used to calculate fat fraction and iron content in a single breath-hold in 109 adult patients. Proton density fat fraction (PDFF) was prospectively estimated using complex-based data reconstruction with multiplex fat modeling. Water  $R2^*$  was used to estimate iron content. Biopsy was obtained in all cases, grading liver steatosis, siderosis, inflammation, and fibrosis. Differences in PDFF and  $R2^*$  values across histopathological grades were analyzed, and ROC curves analyses evaluated the MR diagnostic performance.

**Results:** Calculated fat fraction measurements showed significant differences ( $p < 0.001$ ) among steatosis grades,

being unaffected by the presence of inflammation or fibrosis ( $p \geq 0.05$ ). A strong correlation was found between fat fraction and steatosis grade ( $R_S = 0.718$ ,  $p < 0.001$ ). Iron deposits did not affect fat fraction quantitation ( $p \geq 0.05$ ), except in cases with severe iron overload (grade 4). A strong positive correlation was also observed between  $R2^*$  measurements and iron grades ( $R_S = 0.704$ ,  $p < 0.001$ ). Calculated  $R2^*$  values were not different across grades of steatosis, inflammation, and fibrosis ( $p \geq 0.05$ ).

**Conclusion:** A MECSE-MR sequence simultaneously quantifies liver steatosis and siderosis, regardless coexisting liver inflammation or fibrosis, with high accuracy in a wide spectrum of diffuse liver disorders. This sequence can be acquired within a single breath-hold and can be implemented in the routine MR evaluation of the liver.

**Key words:** Fat quantification—Iron overload— $R2^*$  measurement—Chronic liver diseases—Quantitative imaging biomarkers

Correspondence to: Manuela França; email: mariamanuela.franca@gmail.com

Diffuse liver diseases result from a wide spectrum of disorders, with fat and iron being the most common

associated parenchymal deposits. Although fat accumulation is the histological hallmark of non-alcoholic fatty liver disease (NAFLD), steatosis itself is a common feature to many different conditions [1]. Hepatic iron overload is mainly found in hemochromatosis and transfusion-dependent anemia, although it is also commonly present in diverse chronic liver diseases [2].

Both fatty liver and iron overload are reversible conditions, imposing an oxidative parenchymal stress and interfering with each other in a cross-dependent fashion. In patients with NAFLD and non-alcoholic steatohepatitis, hepatic iron overload is frequently present and associated with disease severity and fibrosis development [2]. Also, coexisting steatosis in patients with iron overload acts as an important cofactor in the development of fibrosis and cirrhosis [3].

Core biopsy is still considered the gold standard technique to evaluate liver steatosis and iron overload [4, 5]. However, it is invasive, expensive, uncomfortable, and prone to complications. Biopsy has also high sampling inconsistency due to the spatial heterogeneous distribution of disease, the small size of the pathological specimen, and the large inter- and intra-observer variability [4]. Most pathological evaluations are categorized in descriptive subjective scores, without accurate quantitative measurements. These limiting aspects have promoted the development of *in vivo* imaging biomarkers able to accurately quantify fat and iron.

MR imaging has been used to detect and quantify liver steatosis and iron overload as independent entities. In-phase (IP) and opposed-phase (OP) dual-echo chemical shift T1-weighted gradient recalled echo imaging is used in most institutions for intracellular fat identification. Liver steatosis is recognized as a decrease in hepatic signal intensity on the OP images compared with the IP images [1, 6]. In iron-loaded livers, a decrease in hepatic signal intensity on the IP images compared with OP images occurs and confounds the detection of fat. The most widely used method for MR liver iron quantification uses the signal intensity ratio between the liver and the paraspinal muscles [7], but the coexistence of steatosis increases the liver signal intensity and underestimates iron quantification.

More recently, several studies have demonstrated the capability of multi-echo chemical shift-based encoded gradient echo MR (MECSE-MR) images for accurate quantification of liver steatosis, either using phantoms [8–11], MR spectroscopy [12–16], or liver biopsy [17–22] as reference standards. The MECSE-MR results must be adjusted for those factors biasing the voxel signal intensity of fat and water, such as  $T1$  relaxation time,  $T2^*$  decay effect, eddy currents, noise, and fat spectral complexity. MECSE-MR sequences independently estimate liver fat fraction and  $T2^*$ , allowing for the simultaneous assessment of liver iron [10, 14, 20, 22, 23].

In order to be clinically implemented as reliable non-invasive simultaneous quantitative imaging biomarker of liver steatosis and iron overload, MECSE-MR metrics must be validated in diverse clinical scenarios taking into account confounder factors, such as inflammation and/or fibrosis.

Our objective was to evaluate the diagnostic performances and confounding factors of 3 Tesla MECSE-MR sequence to simultaneously quantify liver steatosis and iron overload deposits, in a group of patients with different diffuse liver disorders, having histopathological validation as reference gold standard.

## Patients and methods

The Institutional Review Board approved this prospective study. Between February 2013 and January 2014, 116 consecutive adult patients with indication for liver biopsy were recruited after written informed consent. Seven patients were excluded due to claustrophobia ( $n = 1$ ), imaging artifacts ( $n = 3$ ; skin tattoos, movement, and radiofrequency zipper artifacts, respectively), unsatisfactory biopsy sample ( $n = 1$ ), and malignancy within the sample ( $n = 2$ ).

All patients had a percutaneous liver biopsy sample acquired under ultrasonography guidance (16–18 G needles). The biopsied liver segment was documented for further correlation with MR measurements. Patients with focal liver lesions also had a targeted biopsy for the lesion.

### *Histopathological evaluation*

Liver biopsy was the reference standard for grading steatosis, iron deposits, inflammation, and fibrosis. Two pathologists (JRV, more than 15 years' experience; FEC, 4th training year), blinded to the MR results, evaluated in consensus the biopsy samples using hematoxylin–eosin, Masson's trichrome, and Perls' Prussian stains.

Hepatic steatosis was scored as the proportion of hepatocytes containing fat vesicles: no steatosis (grade 0, less than 5%); mild steatosis (grade 1, 5–33%); moderate steatosis (grade 2, 33–66%); and severe steatosis (grade 3, more than 66%) [24]. Iron loading was graded considering the presence of granules within the hepatocytes as grade 0 (iron granules absent or barely discernible in high-power field  $\times 400$ ); grade 1 (granules easily confirmed at  $\times 400$  or barely discernible at  $\times 250$ ); grade 2 (granules resolved at  $\times 100$ ); grade 3 (granules resolved at  $\times 25$ ); and grade 4 (visible at low power  $\times 10$ ) [5]. The Ishak modified scale was used for grading inflammation (necro-inflammatory activity score range 0–18) and fibrosis (F0–F6) [25]. The necro-inflammatory activity score was grouped as no inflammation (score 0), mild inflammation (score 1–6), moderate inflammation (score 7–12), and severe inflammation (score 13–18); Ishak fibrosis staging was categorized into none or minimal

fibrosis (F0–F1), moderate fibrosis (F2–F3), and marked fibrosis/cirrhosis (F4–F6) [26].

### MR imaging examination

Liver MR imaging (3T-TX Achieva, Philips Healthcare, The Netherlands) with a sixteen-channel phased-array coil was performed within 30 days from liver biopsy. The 2D MECSE-MR sequence used 12 echoes (TE's = 0.99 to 8.69, short echo spacing = 0.7 ms; TR = 10 ms) with a 10° flip angle [23] to minimize T1 bias. The whole liver was covered under end-expiratory phase single breath-hold acquisition (34 slices; voxel dimensions, 3 × 3 mm; slice thickness, 7 mm; 0.3 mm gap; reconstruction voxel size, 2 × 2 mm; field of view, 375 × 302 mm; parallel imaging effective acceleration factor, 1.8; bandwidth, 2433 hertz per pixel). The total acquisition time ranged between 12 and 15 s. To avoid cross-talk between slices, excitation was independent for each slice.

### Image analysis

Images were exported as raw data to quantify proton density fat fraction (PDFFF) and iron-related  $R2^*$  measurements using QLiver software (QUIBIM, Valencia, Spain), which is based on least squares analysis by Levenberg–Marquardt algorithm. The magnitude and phase reconstructions allow two fitting procedures [23, 27]. The complex phase information estimates the resonance peak of the larger component, either water or fat, to generate frequency distribution maps. The algorithm performs the fitting of a sinusoid of water and fat signals modulated by exponential decays corresponding to  $R2^*$  of water and  $R2^*$  of fat. The  $R2^*$  water component was only considered for the  $R2^*$  measurements, the  $R2^*$  fat component being negligible. Spectral multipeak modeling of fat controlled the proton density of the multiple fat peaks (75%, 420 Hz; 17%, 318 Hz; 8%, –94 Hz) [9]. The pixel PDFFF was the ratio between the normalized fat proton density and the total (fat and water) proton density (PDFFF = PDF/[PDF + PDW]). The calculated  $R2^*$  was used to estimate iron content.

One radiologist (MF, 8 years' experience on MR imaging) reviewed the images blinded to histopathological results, knowing the biopsied Couinaud segment. Median PDFFF (%) and iron-related water  $T2^*$  shortening ( $R2^*$ ) ( $s^{-1}$ ) were determined with regions of interest (ROI, 4 mm radius) manually placed at the known biopsy segment (Figs. 1 and 2), avoiding vessels and lesions.

### Statistical analysis

As PDFFFs and  $R2^*$  values distributions were non-symmetrical, non-normally distributed variables were log-transformed. Groups' data are presented as mean and standard deviation or median and interquartile range.

The Spearman correlation coefficient was used to calculate the degree of association between continuous and/or ordinal variables. Differences of PDFFF (log-transformed) and  $R2^*$  values between histological groups of steatosis, iron, inflammation, and fibrosis were assessed with one-way analysis of variance, with post hoc Tukey HSD test, or by Kruskal–Wallis test with post hoc Mann–Whitney test and Bonferroni correction.

The influence of different independent ordinal variables (histological grades of steatosis, siderosis, necro-inflammatory activity, or fibrosis) on PDFFF and  $R2^*$  measurements was assessed by ANOVA tests. A Factorial Analysis of Variance was performed to evaluate the effect of the several histological variables on the PDFFF (log-transformed), all together, so that their joint effect could be assessed (dependent variable: log-transformed PDFFF; factors: histological categorical variables—steatosis, iron deposits, inflammation, and fibrosis). Only two-way interactions were considered to avoid empty cells. The iron grades were grouped for Factorial Analysis of Variance as none (grade 0), mild to moderate (grade 1–2), and marked iron overload (grade 3–4).

Receiver operating characteristic (ROC) analysis evaluated measurements' diagnostic performance. The cut-off values that provided the best possible specificity and sensitivity of MR imaging-derived PDFFF and  $R2^*$  measurements according to the histological grades were calculated.

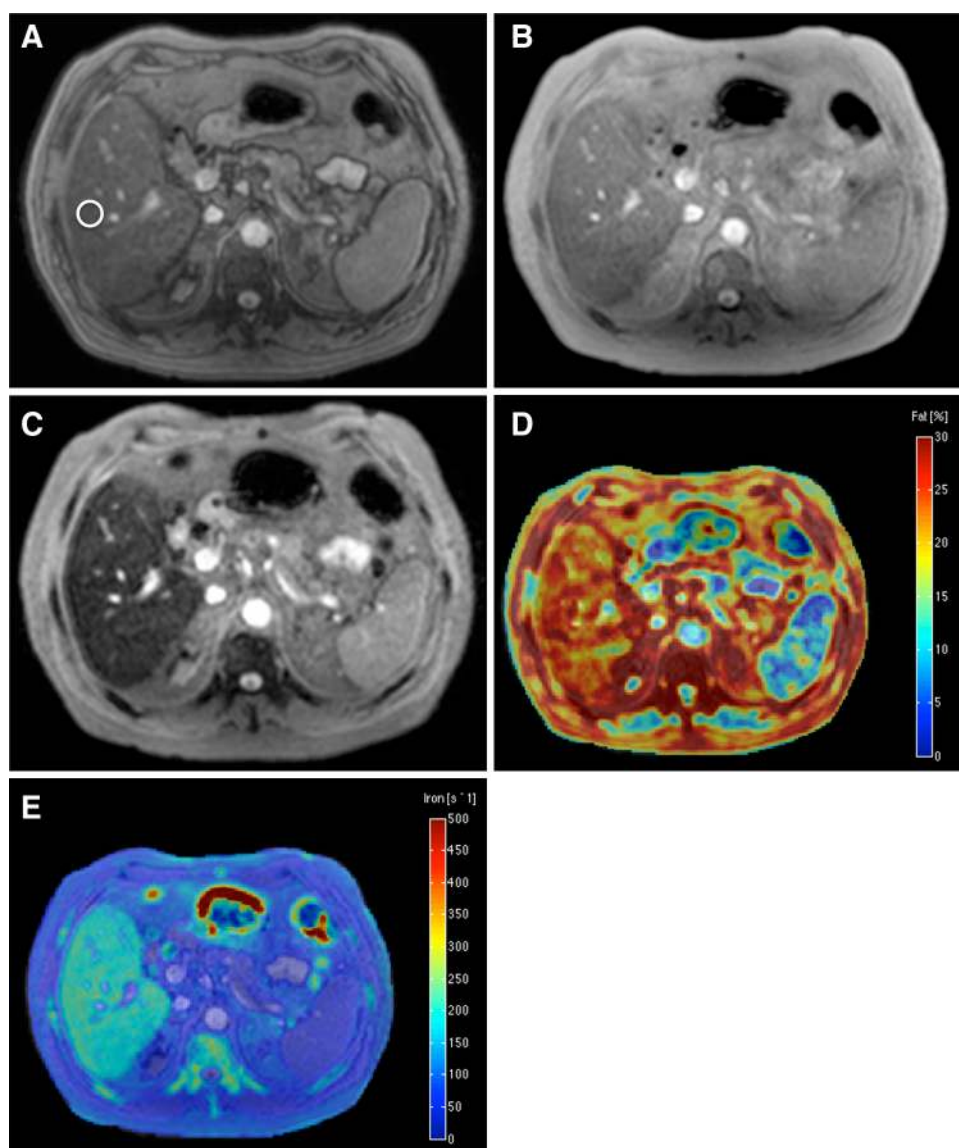
SPSS (version 22; SPSS, IBM, Chicago, Ill) was used for the analysis. For all tests, a two-tailed p value of less than 0.05 was considered as statistically significant.

## Results

The final study population ( $n = 109$ ) included 61 men and 48 women, with a mean (SD) age of 46.5 (13.5) years, range of 19–77 years (Table 1). The clinical indications for liver core biopsies were liver graft dysfunction ( $n = 32$ ), persistent elevation of liver enzymes ( $n = 21$ ), chronic HCV infection ( $n = 14$ ), hemochromatosis ( $n = 8$ ), alcoholic liver disease ( $n = 8$ ), assessment of liver parenchyma during investigation of focal liver lesion ( $n = 5$ ), auto-immune hepatitis ( $n = 5$ ), chronic HBV infection ( $n = 5$ ), toxic hepatitis ( $n = 4$ ), NAFLD ( $n = 4$ ), and Wilson disease ( $n = 3$ ). Mean time interval between biopsy and MR examination was 2 days (range 0–25). Most patients (58%) underwent abdominal MR imaging on the same day of liver biopsy.

### Histopathological results

Biopsy samples were scored for hepatic steatosis and iron as shown in Table 1. Forty-one patients had no steatosis or iron overload, while 21 patients had both steatosis and iron overload.



**Figure 1.** A–E PDFF and  $R2^*$  values estimated in a patient with severe steatosis and iron overload (histologic steatosis grade 3, iron grade 3). First echoes in Oppose Phase (A) and In Phase (B) and last echo In Phase (C) as representative images. Circular ROI was placed in the first image, in the same segment as liver biopsy, estimating PDFF of 26% and  $R2^*$  of  $209 \text{ s}^{-1}$ . PDFF (D) and  $R2^*$  (E) parametric maps demonstrate the distribution of fat and iron deposition in the liver parenchyma.

The distribution of iron deposits was registered in 41 patients: 28 (68% of them) had iron only within hepatocytes, 5 (12%) had reticuloendothelial (RES) iron, and 8 (20%) patients had iron both in hepatocytes and RES.

The Ishak modified classification system was scored in 95 patients (87%) for necro-inflammatory activity and fibrosis staging (Table 1). This classification could not be established in 14 cases (12.8%) due to sample degradation, and were considered “missing values.”

### PDFF measurements

Mean (SD), median, interquartile range, and range for MECSE-MR-derived PDFF were 5.5% (3.6%), 4.3%, 6.7%–3.5%, and 1.0%–20.0%, respectively.

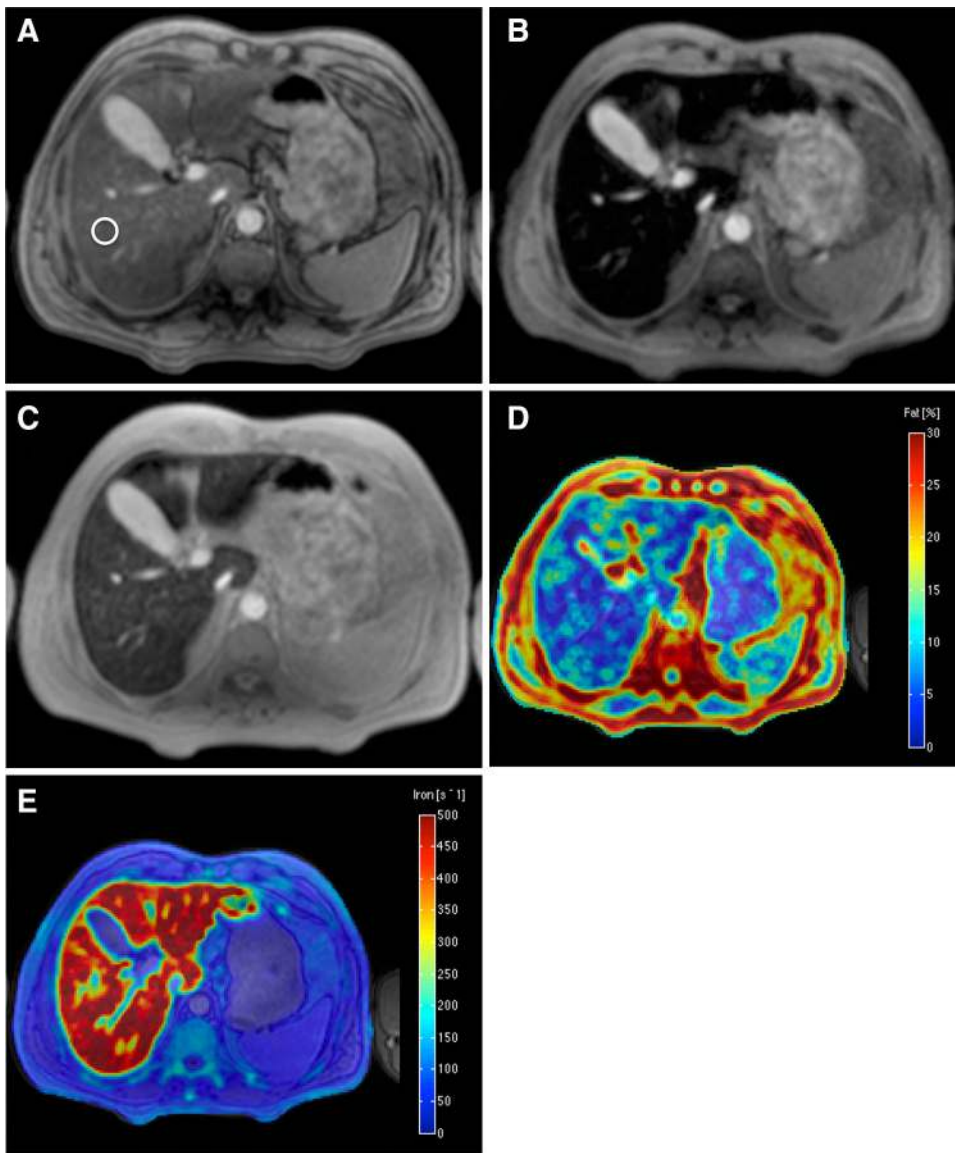
PDFF (log-transformed) values showed significant differences ( $p < 0.001$ ) among histological steatosis grade groups (Fig. 3; Table 2), with a strong relationship

between PDFF and steatosis grades ( $R_S = 0.718$ ,  $p < 0.001$ ). Using Tukey HSD test, PDFF measurements could differentiate between patients without steatosis (grade 0) and mild steatosis (grade 1), and between moderate (grade 2) and severe (grade 3) steatosis ( $p < 0.001$ ) (Fig. 3).

Threshold PDFF values were generated for histological steatosis grading (Table 3). To diagnose the presence of liver steatosis, distinguishing patients without steatosis (grade 0,  $n = 74$ ) from those with grade 1 or greater ( $n = 35$ ), PDFF had an area under the receiving operating characteristic curve (AUC) of 0.930 (95% confidence interval: 0.881–0.979;  $p < 0.001$ ).

A PDFF threshold of 4.8% identified any grade of steatosis (grade  $\geq 1$ , 88.6% sensitivity and 85.1% specificity). A PDFF threshold of 8.5% distinguished patients with none or mild steatosis (grade  $\leq 1$ ) from patients with moderate to severe steatosis (grade  $\geq 2$ , 81.3%





**Figure 2.** **A–E** PDFF and  $R2^*$  values estimated in a patient with severe iron overload (histologic grade 4), without steatosis. First echoes in Oppose Phase (**A**) and In Phase (**B**) and last echo In Phase (**C**) as representative images. Circular ROI was placed in the first image, in the same segment as liver biopsy, estimating  $R2^*$  of  $487\text{ s}^{-1}$ . PDFF was erroneously estimated as 9%. PDFF (**D**) and  $R2^*$  (**E**) parametric maps demonstrate the distribution of fat and iron deposition in the liver parenchyma.

sensitivity and 96.8% specificity). A PDFF threshold of 12.9% identified patients with severe steatosis (grade 3, 100% sensitivity and 100% specificity).

Overall in this study, univariate analysis of PDFF measurements (log-transformed) demonstrated that PDFF measurements were unaffected by liver inflammation ( $p = 0.738$ ), fibrosis ( $p = 0.610$ ), or iron deposits ( $p = 0.339$ ). In the Factorial Analysis of Variance, the only statistical significant main factor on log-transformed PDFF measurements was steatosis ( $p < 0.001$ ), in the presence of all two-way interactions (histological grades of steatosis, iron deposits, inflammation, and fibrosis) (Table 4). Nevertheless, we have observed that PDFF measurements overestimated liver steatosis in 3 of 5 patients with severe iron overload (iron histological grade 4) (Fig. 2), corresponding to extremes for histological steatosis grade 0, in Fig. 3. Those patients had  $R2^*$  measurements ranging from 487 to  $1299\text{ s}^{-1}$ . Fur-

thermore, in a subset analysis excluding the 5 patients with severe iron overload (histologic iron grade 4), a higher correlation was found between PDFF and steatosis grades ( $R_S = 0.744$ ,  $p < 0.001$ ).

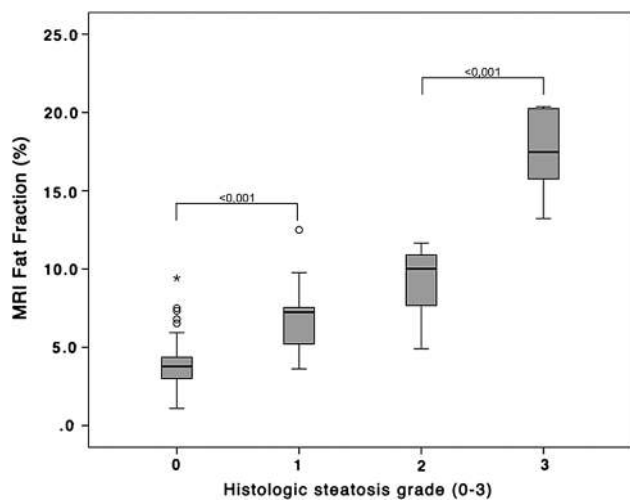
### *$R2^*$ measurements*

The mean  $R2^*$  values ( $\text{s}^{-1}$ ) were 86, median 42, IQR 66–35, range from 24 to 1299. A high significant correlation was observed between  $R2^*$  values and liver histological iron grades ( $R_S = 0.704$ ,  $p < 0.001$ ). Hepatic  $R2^*$  values increased with increasing iron grade ( $p < 0.001$ ) (Table 2; Fig. 4). Using Bonferroni post hoc correction, the  $R2^*$  measurements were significantly different ( $p < 0.001$ ) between levels of histological grade of iron deposits, except between grades 2 and 3, and grades 3 and 4. The  $R2^*$  values were not significantly different across different histological grades of steatosis

**Table 1.** Clinical and histological characteristics of patient population ( $N = 109$  patients). Data in parenthesis are percentage

Characteristic	Result
Sex	
Man	61 (56)
Woman	48 (44)
Age (y)—mean $\pm$ SD, range	46.5 $\pm$ 13.5, 19–77
Man	48.0 $\pm$ 11.6, 20–76
Woman	44.5 $\pm$ 15.5, 19–77
Steatosis	
0; <5% hepatocytes	74 (67.9)
1; >5–33% hepatocytes	19 (17.4)
2; >33–66% hepatocytes	11 (10.1)
3; >66% hepatocytes	5 (4.6)
Iron	
0 (iron granules absent or barely discernible in high-power field $\times$ 400)	55 (50.4)
1 (iron granules easily confirmed at $\times$ 400 or barely discernible at $\times$ 250)	27 (24.8)
2 (iron granules resolved at $\times$ 100)	16 (14.7)
3 (iron granules resolved at $\times$ 25)	6 (5.5)
4 (iron granules visible at low power $\times$ 10)	5 (4.6)
Ishak necro-inflammatory activity	
None (0)	7 (6.4)
Mild (1–6)	72 (66.1)
Moderate (7–12)	16 (14.7)
Severe (13–18)	0 (0)
N/A	14 (12.8)
Ishak fibrosis stage	
None or minimal (F0–F1)	52 (47.7)
Moderate (F2–F3)	24 (22.0)
Marked fibrosis/cirrhosis (F4–F6)	19 (17.4)
N/A	14 (12.8)

N/A not applicable



**Figure 3.** Box and whisker plots of PDFF (%) measured in all patients ( $n = 109$ ), compared with histologic steatosis grading. Extremes in grade 0 corresponded to patients with hemochromatosis and histologic iron grade 4 ( $R2^*$  measurements ranging from 487 to 1299  $s^{-1}$ ). Horizontal lines indicate significant differences ( $p < 0.05$ ) between groups, evaluated by post hoc analysis.

( $p = 0.947$ ), inflammation ( $p = 0.530$ ), and fibrosis ( $p = 0.101$ ). Unfortunately, a Factor Analysis of Variance for the  $R2^*$  measurements with multiple factors

could not be performed due to the asymmetric distribution of the  $R2^*$  data and heterogeneity of variances.

Liver  $R2^*$  had an AUC of 0.848 (95% CI 0.776–0.919,  $p < 0.001$ ) for distinguishing patients with iron at liver biopsy (grade  $\geq 1$ ) from those without, and an AUC of 0.976 (95% CI 0.949–1.000;  $p < 0.001$ ) for distinguishing patients with none or mild iron overload (grade  $\leq 2$ ) from patients with moderate to severe iron overload (grade  $\geq 3$ ) (Table 3). For the diagnosis of any grade of siderosis, a  $R2^*$  threshold of 42  $s^{-1}$  provided a sensitivity of 80.0% and a specificity of 75.9%. For differentiation of histological iron grade 3 or greater from iron grade 2 or less, a  $R2^*$  threshold of 91  $s^{-1}$  had a sensitivity of 90.9% and a specificity of 94.9%.

## Discussion

MECSE-MR imaging-derived fat and iron biomarkers must be validated in wide normal and pathological liver scenarios. In this prospective study, simultaneous quantification of liver PDFF and  $R2^*$  was performed with high accuracy, using a 3 T magnet, in a large group of patients with diverse diffuse liver disorders, and taking into account the main potential confounders such as liver inflammation and fibrosis.

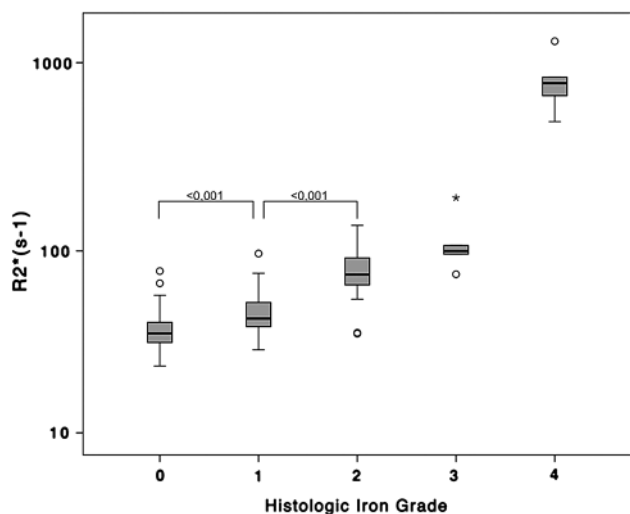
PDFF measurements had an excellent correlation with histological steatosis grade, regardless iron content, inflammation, or fibrosis, and distinguishing between dichotomized steatosis grades with great accuracy. Studies presenting PDFF threshold values to diagnose hepatic steatosis reported values ranging from 2.9% to 7.5% [18, 19, 21, 28, 29], similar to our 4.8% threshold. Our thresholds differentiating steatosis grades are lower than previously published [18, 19, 29], the differences being related to different methodologies and patients' cohorts. Multipeak reconstruction models rely on "a priori" known fat spectra [9, 30]. In contrast to previous studies [12–14, 18, 22, 29, 30], our study population had predominantly low liver fat content, with the risk of over-fitting the curve model. The three fat peaks model (420, 318 and  $-94$  Hz, at 3 T) was shown to provide good results [9]. However, no specific choice of spectral model seems to be significantly superior [31] for PDFF measurements.

PDFF quantification was not confounded by the coexistence of inflammation or fibrosis. Although there were discrepancies regarding fibrosis results [19, 21], most authors have also reported no significant confounding effects of fibrosis [18, 28, 29] or inflammation [28]. Iron overload was neither a confounding factor for fat quantification. Nevertheless, PDFF measurements overestimated the steatosis grade in patients with high-grade hemosiderosis (liver  $R2^*$  ranging 487–1299  $s^{-1}$ ). As expected, the correlation between PDFF and steatosis grades improved when patients with severe iron overload (histologic iron grade 4) were excluded. A previous report

**Table 2.** PDFF and  $R2^*$  results stratified by histologic grading of hepatic steatosis, iron deposits, necro-inflammatory activity and fibrosis

	Histologic classification (n. of patients)	PDFF (%)	$R2^*$ ( $s^{-1}$ )
		Mean (SD), median	Mean, median, IQR
Steatosis grade ( $n = 109$ )	0 (64)	3.9 (1.4), 3.8	81, 38, 48–34
	1 (15)	6.7 (2.1), 7.2	101, 44, 65–37
	2 (11)	9.1 (2.5), 10.0	67, 58, 96–51
	3 (5)	17.4 (3.0), 17.5	93, 67, 144–55
Iron grade ( $n = 109$ )	0 (48)	5.2 (3.6), 4.2	36, 41–32
	1 (24)	5.5 (2.6), 4.5	43, 53–39
	2 (14)	4.7 (2.6), 4.4	75, 95–65
	3 (5)	11.0 (7.7), 9.7	100, 128–91
ISHAK necro-inflammatory activity score ( $n = 95$ )	4 (4)	7.8 (1.0), 7.4	779, 1069–578
	0 (7)	7.5 (6.0), 6.3	70, 48, 72–43
	1–6 (72)	5.7 (3.7), 4.4	94, 42, 69–36
	7–12 (16)	4.6 (2.3), 4.0	42, 39, 46–29
ISHAK fibrosis score ( $n = 95$ )	F0–F1 (52)	5.6 (3.7), 4.3	49, 40, 52–34
	F2–F3 (24)	6.1 (4.4), 4.6	103, 49, 77–37
	F4–F6 (19)	5.1 (2.3), 4.2	151, 45, 97–35

The histological classification of liver fibrosis and inflammation was performed in 95 of 109 patients. Histological evaluation of liver fibrosis and inflammatory activity could not be evaluated in 14 cases (12.8% of patients) due to sample degradation



**Figure 4.** Box and whisker plots of  $R2^*$  (log-transformed) measured in all patients ( $n = 109$ ) compared with histologic iron grading. Horizontal lines indicate significant differences ( $p < 0.05$ ) between groups, evaluated by post hoc analysis.

found no influence of iron overload [22], if  $T2^*$  correction was used. However, they did not have patients with  $R2^*$  over  $200 s^{-1}$  [22]. As large iron content lead to extremely low MR signal in gradient echo sequences, PDFF measurements are corrupted by severe iron overload. Therefore, PDFF quantitation is hampered if  $R2^*$  measurements are higher than  $487 s^{-1}$ . As this  $R2^*$  cut-off value is based on our study population, it should be validated with larger populations and using the same quantitative approach. Phantom studies employing fat and iron solutions will also help in defining this threshold value.

Regarding iron evaluation,  $R2^*$  values were positively correlated with histological iron grades without being influenced by steatosis, inflammation, or fibrosis. Most previous clinical studies addressing hepatic iron quantification in diffuse liver diseases were performed with 1.5 T magnets [22, 32]. A recent study with 3 T magnet, like in our series, also showed good estimation of iron overload [33], however, without evaluation of possible confounders. Our correlation (0.704) between  $R2^*$  mea-

**Table 3.** PDFF and  $R2^*$  thresholds compared with histologically determined liver fat and iron grades. Numbers in parentheses are 95% confidence interval

Liver steatosis grade	Number of patients	PDFF threshold <sup>a</sup>	AUC <sup>a</sup>	Sensitivity <sup>a</sup>	Specificity <sup>a</sup>	Accuracy <sup>a</sup>	PPV <sup>a</sup>	NPV <sup>a</sup>
0 vs. $\geq 1$	35	4.8	0.930 (0.881–0.979)	88.6 (73.2–96.8)	85.1 (75.0–92.3)	86.2	73.8 (58.0–86.1)	94.0 (85.4–98.4)
$\leq 1$ vs. $\geq 2$	16	8.5	0.950 (0.901–0.998)	81.3 (54.4–96.0)	96.8 (90.0–99.3)	94.5	81.3 (54.4–96.0)	96.8 (90.9–99.3)
$\leq 2$ vs. 3	5	12.9	1.000 (1.000–1.000)	100 (47.8–100)	100 (96.5–100)	100.0	100.0 (47.8–100)	100.0 (96.5–100)
Liver iron grade	Number of patients	$R2^*$ threshold <sup>b</sup>	AUC	Sensitivity	Specificity <sup>a</sup>	Accuracy <sup>a</sup>	PPV <sup>a</sup>	NPV <sup>a</sup>
0 vs. $\geq 1$	55	42	0.848 (0.776–0.919)	80.0 (67.0–89.6)	75.9 (62.3–86.5)	78.0	77.2 (64.1–87.3)	78.8 (65.3–88.9)
$\leq 2$ vs. $\geq 3$	11	91	0.976 (0.949–1.000)	90.9 (58.7–99.8)	94.9 (88.5–98.3)	94.5	66.7 (38.4–88.2)	98.9 (94.2–100)

PPV, Positive predictive values; NPV, negative predictive values

<sup>a</sup> Data are shown in percentage

<sup>b</sup> Data are  $s^{-1}$

**Table 4.** Factorial analysis of variance for PDFF measurements, considering PDFF (log transformed) as the dependent variable, and the histological categorical variables as multiple factors (steatosis, iron deposits, inflammation and fibrosis)

Factor	<i>p</i> value
Steatosis grade	<0.001
Iron deposits	0.339
Inflammation	0.738
Fibrosis	0.610
Interaction steatosis*fibrosis	0.669
Interaction steatosis*iron deposits	0.590
Interaction steatosis*inflammation	0.380
Interaction iron deposits*fibrosis	0.152
Interaction inflammation*fibrosis	0.950
Interaction iron deposits*inflammation	0.522
$R^2 = 0.746$	

measurements and histological iron grading is similar to others (0.7–0.8) [32, 33]. This correlation might be underestimated by coexistence of iron within hepatocytes and RES in diffuse liver diseases [2] as  $R2^*$  reflects liver iron without discriminating its cellular location. Although most of our patients with overload had iron distributed only within hepatocytes, this might explain some inaccuracies in patients with iron also within the RES, not considered for the histological iron score.

Our study has some limitations. The study population is from a real clinical scenario with a relatively small number of patients having severe steatosis and severe iron overload. The correlation is therefore biased towards lower deposits. Nevertheless, our findings provide meaningful information on the accuracies of PDFF and  $R2^*$  measurements and the absence of confounding effects when a MECSE approach is used for fat and iron quantification in clinical practice. Although histological grading was obtained by consensus, biopsy as reference standard has large inter-observer and intra-observer variability [4, 34]. In order to minimize the sampling bias, our MR measurements were estimated from the biopsied liver segment. The use of visual categorical scores for the histological assessment of hepatic overload might explain the absence of significant differences in PDFF measurements between histologic steatosis grade 1 and 2, and it could also have underestimated the correlation and diagnostic accuracy of PDFF and  $R2^*$  quantitative measurements. Visual interpretation is observer-dependent, not reproducible, and ignores the lipid composition [35]. Computerized digital image analysis of liver tissue will probably eliminate subjective imprecisions [35]. Image processing tools for histological evaluation will, therefore, probably improve the correlation between histology and MR-derived PDFF quantification when fat accumulation is mild or moderate. Furthermore, more precise quantitative analytical essays for liver fat and iron assessment will require a different liver sample and they are usually not performed in clinical practice. Unfortunately, digital and analytical quantitative methods were not available in our study.

PDFF measurements and thresholds may be affected by the MECSE-MR sequence parameters and reconstruction model. The TR was selected to minimize acquisition time minimizing  $T1$  influence. The combination of low flip angle and TR was also settled to optimize signal-to-noise ratio (close to Ernst angle) and contrast. Although the used flip angle might introduce some measurement uncertainty due to residual  $T1$  bias, this bias was estimated to introduce a maximum deviation of 5% for PDFF values around 50% [36]. Acquisition and postprocessing protocols might need to be adjusted when simultaneous fat and iron quantification is intended. For example, while a low flip angle is needed to reduce  $T1$  bias in PDFF quantification, a too low flip angle sequence will result in decreased SNR, harming iron quantification [6]. Our MECSE-MR sequence used 12 echoes, with a short echo spacing (0.7 ms). Although the best curve fitting and number of echoes strategy have yet to be defined and the number of echoes might be smaller for PDFF quantification without decreasing accuracy [37], our 12-echo approach improves  $R2^*$  quantification. Of importance, the first echo and the echo spacing should be as short as possible to better capture the signal decay in cases of severe iron overload [38]. 3 T MR magnets have higher signal-to-noise ratio and overall quality than 1.5 T units, being advantageous for abdominal imaging. The  $R2^*$  values given in this study cannot be extrapolated to 1.5 T magnets. Higher susceptibility artifacts on 3 T magnets [39] result in lower thresholds for the maximum iron burden quantification at this field strength, due to  $T2^*$ -shortening effect [2, 40]. Although the repeatability of our MECSE-MR sequence was not evaluated, similar approaches have been found to have extremely high intra- and inter-examination PDFF estimation repeatability [41, 42]. Finally, we did not compare our results of PDFF and  $R2^*$  quantification against other MRI methods or tools. However, it has been demonstrated that estimation of hepatic PDFF is reproducible across imaging methods, magnetic field strengths, and different vendors [41, 42]. Iron-related  $R2^*$  quantification is also dependent on the magnetic field strength and the quantitative MR protocol [6]. Nevertheless, hepatic  $R2^*$  quantification seems to be reproducible at 1.5 T and 3 T [43] and using fat-corrected models might be more advantageous [6, 22, 44]. Further calibration studies with phantoms are necessary to guarantee robustness, precision, and reproducibility of the measurements.

In conclusion, 3 T MECSE  $T1$ - and  $T2^*$ -corrected MR sequence simultaneously identifies and quantifies liver steatosis and siderosis, with high accuracy, and in a wide spectrum of diffuse liver disorders, regardless of coexisting liver inflammation or fibrosis. Because it can be acquired within a single breath-hold, these parametric images could be easily implemented in the routine clinical MR evaluation of the liver.



### Compliance with ethical standards

**Funding** This work was partially funded by a research grant from the Teaching and Research Department of Centro Hospitalar do Porto (DEFI:309/12(213-DEFI/251-CES)) and from a Spanish Ministry of Health and Carlos III Health Institute funding grant (PI12/01262). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Conflict of interest** Javier Sanchez Gonzalez is employee at Philips Healthcare Iberia. Angel Alberich Bayarri and Luis Marti-Bonmati are co-founders of QUIBIM SME. The other authors declare that they have no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

### References

- Reeder SB, Sirlin CB (2010) Quantification of liver fat with magnetic resonance imaging. *Magn Reson Imaging Clin N Am* 18:337–357. doi:10.1016/j.mric.2010.08.013
- Sirlin CB, Reeder SB (2010) Magnetic resonance imaging quantification of liver iron. *Magn Reson Imaging Clin N Am* 18:359–381. doi:10.1016/j.mric.2010.08.014
- Powell EE, Ali A, Clouston AD, et al. (2005) Steatosis is a cofactor in liver injury in hemochromatosis. *Gastroenterology* 129:1937–1943. doi:10.1053/j.gastro.2005.09.015
- Ratzu V, Charlotte F, Heurtier A, et al. (2005) Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 128:1898–1906
- Deugnier Y, Turlin B (2007) Pathology of hepatic iron overload. *WJG* 13:4755–4760
- Yokoo T, Browning JD (2014) Fat and iron quantification in the liver: past, present, and future. *Top Magn Reson Imaging* 23:73–94. doi:10.1097/RMR.0000000000000016
- Gandon Y, Olivie D, Guyader D, et al. (2004) Non-invasive assessment of hepatic iron stores by MRI. *Lancet* 363:357–362. doi:10.1016/S0140-6736(04)15436-6
- Yu H, McKenzie CA, Shimakawa A, et al. (2007) Multiecho reconstruction for simultaneous water-fat decomposition and  $T2^*$  estimation. *J Magn Reson Imaging* 26:1153–1161. doi:10.1002/jmri.21090
- Yu H, Shimakawa A, McKenzie CA, et al. (2008) Multiecho water-fat separation and simultaneous  $R2^*$  estimation with multifrequency fat spectrum modeling. *Magn Reson Med* 60:1122–1134. doi:10.1002/mrm.21737
- O'Regan DP, Callaghan MF, Wylezinska-Arridge M, et al. (2008) Liver fat content and  $T2^*$ : simultaneous measurement by using breath-hold multiecho MR imaging at 3.0 T—feasibility. *Radiology* 247:550–557. doi:10.1148/radiol.2472070880
- Hines CDG, Yu H, Shimakawa A, et al. (2009)  $T1$  independent,  $T2^*$  corrected MRI with accurate spectral modeling for quantification of fat: validation in a fat-water-SPIO phantom. *J Magn Reson Imaging* 30:1215–1222. doi:10.1002/jmri.21957
- Meisamy S, Hines CDG, Hamilton G, et al. (2011) Quantification of hepatic steatosis with  $T1$ -independent,  $T2^*$ -corrected MR imaging with spectral modeling of fat: blinded comparison with MR spectroscopy. *Radiology* 258:767–775. doi:10.1148/radiol.10100708
- Yokoo T, Bydder M, Hamilton G, et al. (2009) Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T. *Radiology* 251:67–76. doi:10.1148/radiol.2511080666
- Yokoo T, Shiehorteza M, Hamilton G, et al. (2011) Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. *Radiology* 258:749–759. doi:10.1148/radiol.10100659
- Hines CDG, Frydrychowicz A, Hamilton G, et al. (2011)  $T1$  independent,  $T2^*$  corrected chemical shift based fat-water separation with multi-peak fat spectral modeling is an accurate and precise measure of hepatic steatosis. *J Magn Reson Imaging* 33:873–881. doi:10.1002/jmri.22514
- Kühn J-P, Hernando D, Mensel B, et al. (2014) Quantitative chemical shift-encoded MRI is an accurate method to quantify hepatic steatosis. *J Magn Reson Imaging* 39:1494–1501. doi:10.1002/jmri.24289
- Permutt Z, Le T-A, Peterson MR, et al. (2012) Correlation between liver histology and novel magnetic resonance imaging in adult patients with non-alcoholic fatty liver disease—MRI accurately quantifies hepatic steatosis in NAFLD. *Aliment Pharmacol Ther* 36:22–29. doi:10.1111/j.1365-2036.2012.05121.x
- Tang A, Tan J, Sun M, et al. (2013) Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. *Radiology* 267:422–431. doi:10.1148/radiol.12120896
- Idilman IS, Aniktar H, Idilman R, et al. (2013) Hepatic steatosis: quantification by proton density fat fraction with MR imaging vs. liver biopsy. *Radiology* 267:767–775. doi:10.1148/radiol.13121360
- McPherson S, Jonsson JR, Cowin GJ, et al. (2009) Magnetic resonance imaging and spectroscopy accurately estimate the severity of steatosis provided the stage of fibrosis is considered. *J Hepatol* 51:389–397. doi:10.1016/j.jhep.2009.04.012
- Parente DB, Rodrigues RS, Paiva FF, et al. (2014) Is MR spectroscopy really the best MR-based method for the evaluation of fatty liver in diabetic patients in clinical practice? *PLoS ONE* 9:e112574. doi:10.1371/journal.pone.0112574.t005
- Kühn J-P, Hernando D, del Rio AM, et al. (2012) Effect of multi-peak spectral modeling of fat for liver iron and fat quantification: correlation of biopsy with MR imaging results. *Radiology* 265:133–142. doi:10.1148/radiol.12112520-/-DC1
- Marti-Bonmati L, Alberich-Bayarri A, Sánchez-González J (2011) Overload hepatitides: quanti-qualitative analysis. *Abdom Imaging* 37:180–187. doi:10.1007/s00261-011-9762-5
- Kleiner DE, Brunt EM, Van Natta M, et al. (2005) Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 41:1313–1321. doi:10.1002/hep.20701
- Ishak K, Baptista A, Bianchi L, et al. (1995) Histological grading and staging of chronic hepatitis. *J Hepatol* 22:696–699
- Mannan R (2014) A comparative evaluation of scoring systems for assessing necro-inflammatory activity and fibrosis in liver biopsies of patients with chronic viral hepatitis. *J Clin Diagn Res* 8:F08-12. doi:10.7860/JCDR/2014/8704.4718
- Hijona E, Sánchez-González J, Alústiza JM, et al. (2012) Accurate fat fraction quantification by multiecho gradient-recalled-echo magnetic resonance at 1.5 T in rats with nonalcoholic fatty liver disease. *Eur J Radiol* 81:1122–1127. doi:10.1016/j.ejrad.2011.02.065
- Kang B-K, Yu ES, Lee SS, et al. (2012) Hepatic fat quantification: a prospective comparison of magnetic resonance spectroscopy and analysis methods for chemical-shift gradient echo magnetic resonance imaging with histologic assessment as the reference standard. *Invest Radiol* 47:368–375. doi:10.1097/RLI.0b013e31824baff3
- Tang A, Desai A, Hamilton G, et al. (2015) Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. *Radiology* 274:416–425. doi:10.1148/radiol.14140754
- Hamilton G, Yokoo T, Bydder M, et al. (2010) In vivo characterization of the liver fat 1H MR spectrum. *NMR Biomed* 24:784–790. doi:10.1002/nbm.1622
- Wang X, Hernando D, Reeder SB (2016) Sensitivity of chemical shift-encoded fat quantification to calibration of fat MR spectrum. *Magn Reson Med* 75(2):845–851. doi:10.1002/mrm.25681
- Chandarana H, Lim RP, Jensen JH, et al. (2009) Hepatic iron deposition in patients with liver disease: preliminary experience with breath-hold multiecho  $T2^*$ -weighted sequence. *Am J Roentgenol* 193:1261–1267. doi:10.2214/AJR.08.1996
- Banerjee R, Pavlides M, Tunnicliffe EM, et al. (2014) Multiparametric magnetic resonance for the non-invasive diagnosis of liver disease. *J Hepatol* 60:69–77. doi:10.1016/j.jhep.2013.09.002
- Wood JC (2014) Use of magnetic resonance imaging to monitor iron overload. *Hematol Oncol Clin N Am* 28:747–764. doi:10.1016/j.hoc.2014.04.002

35. El-Badry AM, Breitenstein S, Jochum W, et al. (2009) Assessment of hepatic steatosis by expert pathologists: the end of a gold standard. *Ann Surg* 250:691–697. doi:[10.1097/SLA.0b013e3181bcd6dd](https://doi.org/10.1097/SLA.0b013e3181bcd6dd)
36. Liu C-Y, McKenzie CA, Yu H, Brittain JH, Reeder SB (2007) Fat quantification with IDEAL gradient echo imaging: correction of bias from  $T_1$  and noise. *Magn Reson Med* 58:354–364. doi:[10.1002/mrm.21301](https://doi.org/10.1002/mrm.21301)
37. Heba ER, Desai A, Zand KA, et al. (2016) Accuracy and the effect of possible subject-based confounders of magnitude-based MRI for estimating hepatic proton density fat fraction in adults, using MR spectroscopy as reference. *J Magn Reson Imaging* 43:398–406. doi:[10.1002/jmri.25006](https://doi.org/10.1002/jmri.25006)
38. Hernando D, Levin YS, Sirlin CB, Reeder SB (2014) Quantification of liver iron with MRI: state of the art and remaining challenges. *J Magn Reson Imaging* 40:1003–1021. doi:[10.1002/jmri.24584](https://doi.org/10.1002/jmri.24584)
39. Storey P, Thompson AA, Carqueville CL, et al. (2007)  $R_2^*$  imaging of transfusional iron burden at 3 T and comparison with 1.5 T. *J Magn Reson Imaging* 25:540–547. doi:[10.1002/jmri.20816](https://doi.org/10.1002/jmri.20816)
40. Meloni A, Positano V, Keilberg P, et al. (2012) Feasibility, reproducibility, and reliability for the  $T_2^*$  iron evaluation at 3 T in comparison with 1.5 T. *Magn Reson Med* 68:543–551. doi:[10.1002/mrm.23236](https://doi.org/10.1002/mrm.23236)
41. Tyagi A, Yeganeh O, Levin Y, et al. (2015) Intra- and inter-examination repeatability of magnetic resonance spectroscopy, magnitude-based MRI, and complex-based MRI for estimation of hepatic proton density fat fraction in overweight and obese children and adults. *Abdom Imaging* 40:3070–3077. doi:[10.1007/s00261-015-0542-5](https://doi.org/10.1007/s00261-015-0542-5)
42. Hernando D, Sharma SD, Aliyari Ghasabeh M et al (2016) Multisite, multivendor validation of the accuracy and reproducibility of proton-density fat-fraction quantification at 1.5 T and 3 T using a fat-water phantom. *Magn Reson Med*. doi: [10.1002/mrm.26228](https://doi.org/10.1002/mrm.26228).
43. Alam MH, Auger D, McGill LA, et al. (2016) Comparison of 3 T and 1.5 T for  $T_2^*$  magnetic resonance of tissue iron. *J Cardiovasc Magn Reson* 18:40. doi:[10.1186/s12968-016-0259-9](https://doi.org/10.1186/s12968-016-0259-9)
44. Hernando D, Kramer JH, Reeder SB (2013) Multipeak fat-corrected complex  $R_2^*$  relaxometry: theory, optimization, and clinical validation. *Magn Reson Med* 70:1319–1331. doi:[10.1002/mrm.24593](https://doi.org/10.1002/mrm.24593)