

Automated Whole-Liver MRI Segmentation to Assess Steatosis and Iron Quantification in Chronic Liver Disease



David Martí-Aguado, MD, PhD • Ana Jiménez-Pastor, MSc • Ángel Alberich-Bayarri, PhD • Alejandro Rodríguez-Ortega, PhD • Clara Alfaro-Cervello, MD, PhD • Claudia Mestre-Alagarda, MD • Mónica Bauza, MD • Ana Gallén-Peris, MD • Elena Valero-Pérez, MD • María Pilar Ballester, MD, PhD • Marta Gimeno-Torres, MD • Alexandre Pérez-Girbés, MD • Salvador Benlloch, MD, PhD • Judith Pérez-Rojas, MD • Víctor Puglia, MD • Antonio Ferrández, MD, PhD • Victoria Aguilera, MD • Desamparados Escudero-García, MD • Miguel A. Serra, MD, PhD • Luis Martí-Bonmatí, MD, PhD

From the Departments of Digestive Diseases (D.M.A., M.P.B., D.E.G.), Pathology (C.A.C., C.M.A., A.F.), and Gastroenterology and Hepatology (D.M.A.), Clinic University Hospital, INCLIVA Health Research Institute, Avenida Blasco Ibáñez 17, 46010 Valencia, Spain; Biomedical Imaging Research Group (GIBI2³⁰), La Fe Health Research Institute, Valencia, Spain (D.M.A., A.R.O., L.M.B.); Quantitative Imaging Biomarkers in Medicine, Quibim SL, Valencia, Spain (A.J.P., Á.A.B.); University of Valencia, Faculty of Medicine, Valencia, Spain (C.A.C., A.F., D.E.G., M.A.S.); Departments of Pathology (M.B., J.P.R.), Digestive Diseases (E.V.P., M.G.T.), and Radiology (A.P.G., L.M.B.) and the Hepatology and Liver Transplantation Unit (V.A.), La Fe University and Polytechnic Hospital, Valencia, Spain; Departments of Digestive Diseases (A.G.P., S.B.) and Pathology (V.P.), Hospital Arnau de Vilanova, Valencia, Spain; CIBERehd (Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas), Instituto de Salud Carlos III, Madrid, Spain (S.B., V.A.); and Río Hortega, Instituto Salud Carlos III, Madrid, Spain (D.M.A.). Received April 22, 2021; revision requested June 15; revision received September 7; accepted September 17. Address correspondence to D.M.A. (e-mail: davidmmaa@gmail.com).

Supported by the Spanish Ministry of Science and Innovation, Instituto de Salud Carlos III (grant PI19/0380), and Gilead Sciences (grant GLD19/00050).

Conflicts of interest are listed at the end of this article.

See also the editorial by Moura Cunha and Fowler in this issue.

Radiology 2022; 302:345–354 • <https://doi.org/10.1148/radiol.2021211027> • Content codes:  

Background: Standardized manual region of interest (ROI) sampling strategies for hepatic MRI steatosis and iron quantification are time consuming, with variable results.

Purpose: To evaluate the performance of automatic MRI whole-liver segmentation (WLS) for proton density fat fraction (PDFF) and iron estimation (transverse relaxometry [R2*]) versus manual ROI, with liver biopsy as the reference standard.

Materials and Methods: This prospective, cross-sectional, multicenter study recruited participants with chronic liver disease who underwent liver biopsy and chemical shift–encoded 3.0-T MRI between January 2017 and January 2021. Biopsy evaluation included histologic grading and digital pathology. MRI liver sampling strategies included manual ROI (two observers) and automatic whole-liver (deep learning algorithm) segmentation for PDFF- and R2*-derived measurements. Agreements between segmentation methods were measured using intraclass correlation coefficients (ICCs), and biases were evaluated using Bland-Altman analyses. Linear regression analyses were performed to determine the correlation between measurements and digital pathology.

Results: A total of 165 participants were included (mean age \pm standard deviation, 55 years \pm 12; 96 women; 101 of 165 participants [61%] with nonalcoholic fatty liver disease). Agreements between mean measurements were excellent, with ICCs of 0.98 for both PDFF and R2*. The median bias was 0.5% (interquartile range, -0.4% to 1.2%) for PDFF and 2.7 sec^{-1} (interquartile range, 0.2 – 5.3 sec^{-1}) for R2* ($P < .001$ for both). Margins of error were lower for WLS than ROI-derived parameters (-0.03% for PDFF and -0.3 sec^{-1} for R2*). ROI and WLS showed similar performance for steatosis (ROI AUC, 0.96; WLS AUC, 0.97; $P = .53$) and iron overload (ROI AUC, 0.85; WLS AUC, 0.83; $P = .09$). Correlations with digital pathology were high ($P < .001$) between the fat ratio and PDFF (ROI $r = 0.89$; WLS $r = 0.90$) and moderate ($P < .001$) between the iron ratio and R2* (ROI $r = 0.65$; WLS $r = 0.64$).

Conclusion: Proton density fat fraction and transverse relaxometry measurements derived from MRI automatic whole-liver segmentation (WLS) were accurate for steatosis and iron grading in chronic liver disease and correlated with digital pathology. Automated WLS estimations were higher, with a lower margin of error than manual region of interest estimations.

© RSNA, 2021

Online supplemental material is available for this article.

Chronic diffuse liver disease is a major cause of morbidity, mortality, and health care resource use (1). Steatosis is the histopathologic hallmark of nonalcoholic fatty liver disease (NAFLD), the most prevalent liver disease, and is a common feature in other liver diseases. Abnormal liver iron deposition is also often recognized in patients with diffuse liver diseases and metabolic disorders. Despite the high prevalence of chronic liver disease, histologic evaluation is still considered the reference standard to assess disease severity. However, liver biopsy is an invasive procedure with limitations

due to sampling errors, interobserver variability, and low patient acceptance (2). Noninvasive imaging techniques are gaining utility in the diagnosis and monitoring of chronic liver diseases. MRI enables simultaneous evaluation of different histopathologic features. MRI-estimated proton density fat fraction (PDFF) and transverse relaxometry (R2*) with multiecho gradient-echo sequences have been shown to help accurately diagnose hepatic steatosis and iron overload (3–6). Parallel to the growing burden of metabolic disorders, the increasing frequency of NAFLD and dysmetabolic iron overload

This copy is for personal use only. To order printed copies, contact reprints@rsna.org

Abbreviations

AUC = area under the receiver operating characteristic curve, CNN = convolutional neural network, ICC = intraclass correlation coefficient, MOE = margin of error, NAFLD = nonalcoholic fatty liver disease, PDFF = proton density fat fraction, R2* = transverse relaxometry, ROI = region of interest, SD = standard deviation, WLS = whole-liver segmentation

Summary

For steatosis and iron quantification, automated whole-liver MRI segmentation measurements correlated with histologic data in chronic liver disease and showed similar diagnostic accuracy to manual region of interest assessment.

Key Results

- In a prospective multicenter study of 165 participants with chronic liver disease, automated whole-liver segmentation (WLS) compared with manual region of interest (ROI) segmentation had a high intraclass correlation coefficient (0.98 for both), with a median bias of 0.5% for proton density fat fraction (PDFF) and 2.7 sec⁻¹ for transverse relaxometry (R2*) ($P < .001$ for both).
- Correlations between digital pathology and both segmentation methods were high for PDFF ($r = 0.90$ for WLS and $r = 0.89$ for ROI) but moderate for R2* ($r = 0.64$ for WLS and $r = 0.65$ for ROI).

syndrome makes it necessary to facilitate efficient quantification of these markers in clinical routine (7).

At present, measurements of fat fraction and iron concentration require manual selection of regions of interest (ROIs) in the liver parenchyma. This strategy is time consuming, as multiple ROIs are required to obtain reproducible results (8,9). The diagnostic performance of PDFF manual whole-liver segmentation (WLS) has shown a strong correlation with results of different manual ROI sampling methods and spectroscopy (10–13). Under repeatability conditions, similar interobserver agreement is obtained for manual ROIs and manual WLSs, both contributing to variability in radiomic measurements (10,11,14–16). Therefore, automated convolutional neural network (CNN) solutions have been developed to obtain WLSs with similar diagnostic accuracy to manual segmentation but without variability errors (17). Once a CNN model is trained, as long as it is not retrained, it will produce the same output for the same input, without

variability. Although feasible, the performance of markers from automatic WLS has not been compared with that of the reference standard method of histologic evaluation (16–18).

We therefore hypothesized that automated WLS has the same accuracy as manual ROI to quantify liver steatosis and iron overload for the assessment of chronic liver diseases. The main aim of this study was to compare the diagnostic performance of MRI automatic WLS against that of standardized manual ROI sampling for the diagnosis of steatosis and iron overload by using histologic findings as a reference standard in a well-characterized sample of participants with chronic liver disease.

Materials and Methods

Study Design and Participants

We performed a cross-sectional, multicenter, prospective, and observational study with recruited consecutive participants with chronic liver disease and a clinical indication for liver biopsy at three medical centers (Valencia, Spain) between January 2017 and January 2021. Participants were invited to undergo a research MRI examination with a time interval less than 30 days from biopsy. Participant inclusion criteria were age of 18 years or older and having signed the informed consent. Exclusion criteria were contraindications to MRI, imaging artifacts, unsatisfactory biopsy sample (defined as <15 mm in length with fewer than six portal tracts), and malignancy within the sample (Fig 1). The study conforms to the ethical guidelines of the 1975 Declaration of Helsinki and had the a priori approval of the institutional review boards of the three participating hospitals (2016/209, 2017/0031/PI, and 13/2019). The Standards for Reporting Diagnostic Accuracy Studies 2015 guidelines were strictly followed (19). Part of the included participant data has been already published regarding technical developments of quantitative tools for digital pathology ($n = 101$) (20) and automatic liver segmentation ($n = 75$) (21). These methodologic studies did not evaluate the correlation between manual and deep learning results with pathologic findings as a reference standard. Our study evaluates the performance of CNN-automated WLS for steatosis and iron quantification, with liver biopsy and digital image analysis findings as the reference standards.

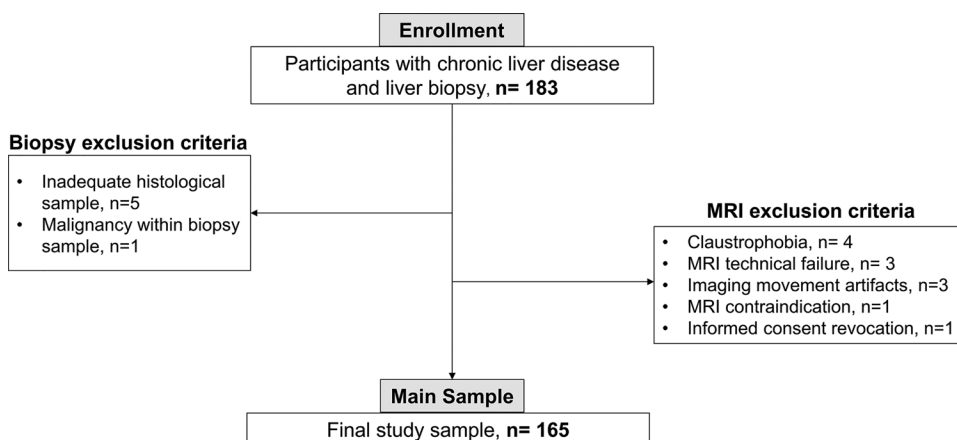


Figure 1: Participant flowchart.

MRI Examinations

MRI scans (3.0T TX Achieva, Philips Healthcare) were obtained with a 16-channel phased-array coil. Participants were asked to fast for a minimum of 4 hours. All participants had a standard nonenhanced MRI scan reviewed by a radiologist (A.P.G., with more than 8 years of experience in abdominal MRI) to exclude focal liver lesions before image analysis. A two-dimensional multiecho chemical shift–encoded gradient-echo sequence was performed in a

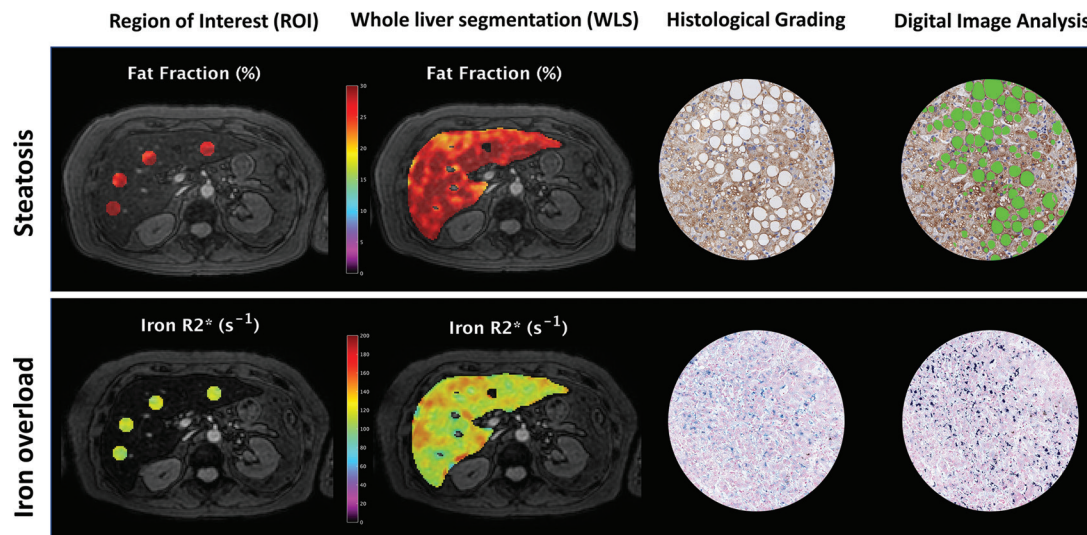


Figure 2: Schematic overview of the MRI (left two columns) and histologic (right two columns) workflow in a 60-year-old woman with nonalcoholic fatty liver disease. Top row shows an out-of-phase first echo (0.9 msec) gradient-recalled echo sequence with proton density fat fraction quantification (manual ROI, 30%; automated WLS, 28%) and adipophilin immunohistochemistry staining for steatosis grading (S3) and quantification (46% fat proportionate area). Bottom row shows the last echo (7.9 msec) of the same sequence with $R2^*$ estimation (manual ROI, 118 sec^{-1} ; automated WLS, 119 sec^{-1}) and Perls stain for iron deposits grading (Fe3) and iron proportionate area quantification (2% iron ratio). Images showing pathologic findings were obtained with a Ventana iScan HT slide scanner ($\times 40$ magnification).

single-breath-hold acquisition with 12 echoes (echo time, 0.9–7.9 msec; short echo spacing, 0.7 msec; repetition time, 9 msec) and a low flip angle (10°) to minimize T1 bias. Image postprocessing was performed with a fitting algorithm that corrects $T2^*$ effects and the spectral complexity of the fat signal (six-peak multifrequency) to calculate fat and iron content (Quibim Precision 2.8 Laboratory, Quibim) (22).

MRI Analysis

PDFF and iron-related $R2^*$ values were measured after manual and automatic segmentations. In the ROI strategy, four large ROIs ($\geq 4 \text{ cm}^2$) were manually placed while avoiding major vessels, bile ducts, liver edges, and artifacts (Fig 2) (9). Two experienced and independent observers with more than 5 years of experience (D.M.A. and M.P.B.) manually located the ROIs on the first echo images with use of the open-source software ITK-SNAP (version 3.6.0, <http://www.itksnap.org>). The ROI segmentation was then automatically propagated onto the coregistered images acquired with the other echo times. Additionally, the whole liver parenchyma was automatically segmented, excluding large vessels, by means of a trained CNN (21). The segmentation algorithm architecture is an encoder-decoder CNN with four convolutional blocks on each branch (Fig E1 [online]). This automated algorithm has previously shown excellent agreement with manual WLS (Dice coefficient of 0.94) and has been externally validated in a cohort without patient overlap, different scanner manufacturers, imaging protocols, or magnetic fields, showing high reproducibility (Dice coefficient of 0.95) (21). However, its diagnostic performance with respect to biopsy is still unknown. The automated WLS analysis was run on a liver fat and iron multipoint suite (Quibim Precision 2.8, Quibim). The PDFF and $R2^*$ values across ROIs and WLSs were recorded with central tendency (mean for $R2^*$ and median for PDFF

values) and variability (standard deviation [SD]) metrics. The median statistic was used in PDFF estimation to mitigate noise-related bias (23). Image analysts were blinded to clinical and histologic data at the time of image analysis.

Clinical Evaluation

All participants underwent a standardized clinical evaluation. Collected data included age, sex, body mass index (kg/m^2), obesity (defined as body mass index $>30 \text{ kg}/\text{m}^2$), metabolic syndrome diagnosis based on the Adult Treatment Panel III criteria (24), history of alcohol intake, cause of chronic liver disease, and a complete biochemical profile.

Histologic Evaluation

Percutaneous biopsies were obtained with a semiautomatic 16-gauge two-step needle. After formalin fixation (10% buffered), tissue slices (thickness, $4 \mu\text{m}$) embedded in paraffin were immunohistochemically stained with adipophilin (Vitro Master Diagnóstica) for steatosis detection and Perls staining (Artisan Iron Staining Kit, Dako) for iron assessment. In our study, all biopsy samples were centrally evaluated by experienced liver pathologists (C.A.C. and A.F., each with over 10 years of experience) blinded to clinical and imaging data. Histologic scoring was performed with use of the Nonalcoholic Steatohepatitis Clinical Research Network system to grade steatosis from S0 to S3 (25). The grading of iron storage was assessed using the Scheuer scoring system (Fe0–Fe4) (26). Additionally, all stained biopsies were digitized with a Ventana iScan HT slide scanner (Roche, Ventana Medical Systems), capturing whole-slide digital images using a $\times 40$ magnification objective and a calibrated camera (4000×4000 pixels being 1 mm^2). Then, digital image analysis was performed to quantitatively obtain the proportionate area (as a percentage) of fat and iron (Fig 2). Semiautomatic

Table 1: Baseline Clinical and Laboratory Characteristics

Characteristic	Value
Sex	
F	96 (58)
M	69 (42)
Age (y)*	55 ± 12
Body mass index (kg/m ²)*	28.4 ± 5.1
Cause of liver disease	
Nonalcoholic fatty liver disease	101 (61)
Autoimmune hepatitis	29 (18)
Viral hepatitis	10 (6.1)
Alcohol consumption	9 (5.5)
Other	16 (10)
Metabolic traits	
Metabolic syndrome	63 (38)
Hypertension	74 (45)
Diabetes mellitus	54 (33)
Dyslipidemia	103 (62)
Platelet count (×10 ⁹ /L)*	235 ± 74
Creatinine (mg/dL)	0.7 ± 0.2
Alanine aminotransferase (U/L) [†]	39 (29–59)
Aspartate aminotransferase (U/L) [†]	39 (29–59)
γ-glutamyl transferase (U/L) [†]	49 (36–67)
Total bilirubin (mg/dL) [†]	82 (47–185)
Alkaline phosphatase (U/L) [†]	92 (70–128)
Albumin (g/dL) [†]	4.4 (4.2–4.6)
Triglycerides (mg/dL) [†]	107 (76–153)
Total cholesterol (mg/dL)*	192 ± 42
Low-density lipoprotein (mg/dL)*	117 ± 34
High-density lipoprotein (mg/dL)*	56 ± 17
International normalized ratio*	1.04 ± 0.09

Note.—Unless otherwise specified, data are numbers of participants ($n = 165$), with percentages in parentheses. Data are reported as means ± standard deviations when normally distributed and medians with interquartile ranges when the distribution is skewed.

* Data are means ± standard deviations.

† Data are medians, with interquartile ranges in parentheses.

commercially available software (MATLAB, version R2016a; MathWorks) was used for computerized analysis. Briefly, an algorithm based on enhanced color- and shape-based thresholds was used for histologic feature segmentation and quantification (20). Digital image analysis is an accurate and precise technique for histologic assessment and can also be considered the ground truth (20,27). Histologic scores graded by pathologists were considered categorical data, while proportionate areas obtained with digital image analysis were continuous measurements.

Statistical Analysis

Data were analyzed using MedCalc Statistical Software 19.4 (MedCalc Software) and the SPSS 25 software package (IBM) by one author (D.M.A., with more than 5 years of experience in statistics). Categorical data are expressed as frequencies with percentages, and quantitative data are expressed as means ± SDs or medians with interquartile ranges. For both the mean and SD

Table 2: Histopathologic Characteristic Distribution in the Study Sample

Characteristic	Value
Biopsy length (mm)*	20 (16–23)
No. of portal tracts [†]	11 ± 3
Steatosis grade	
S0	70 (42)
S1	33 (20)
S2	34 (21)
S3	28 (17)
Iron grade	
Fe0	129 (78)
Fe1	10 (6.1)
Fe2	8 (4.8)
Fe3	11 (6.7)
Fe4	7 (4.2)
Fat proportionate area (%)*	6.6 (2.7–13.6)
Iron proportionate area (%)*	1.7 (1.2–2.6)

Note.—Unless otherwise specified, data are numbers of participants ($n = 165$), with percentages in parentheses. Data are reported as means ± standard deviations when normally distributed and medians with interquartile ranges when the distribution is skewed.

* Data are medians, with interquartile ranges in parentheses.

† Number of portal tracts is presented as the mean ± standard deviation.

distributions of PDFF and R2*, their margins of error (MOEs) were defined as the half width of the 95% CI for the distribution, representing a measure of how precisely the study sample mean was estimated. It was computed as the standard error of the sample divided by \sqrt{n} , n being the sample size. A lower MOE indicates higher confidence levels in the produced results. Comparisons between paired distributions of manual ROI and automated WLS analyses for steatosis and iron quantification were assessed with the Wilcoxon matched pairs test stratified according to histologic grades. The agreement between segmentation methods (manual ROI and automated WLS) was measured using the intraclass correlation coefficient (ICC). Bias, defined as the average difference between automated WLS and manual ROI measurements per participant, was evaluated by using Bland-Altman analysis. The 95% limits of agreement were calculated. Linearity was evaluated using a regression model of automated WLS against manual ROI. From this regression, the coefficient of determination (R^2), slope, intercept, and standard errors were extracted. Agreement, linearity, and bias metrics were reported following the recommendations of the Radiological Society of North America Quantitative Imaging Biomarkers Alliance, or QIBA (28). Comparisons between histologic groups in terms of MRI quantitative data were performed using the Kruskal-Wallis post hoc Tukey range test. The area under the receiver operating characteristic curve (AUC) was calculated for each set of dichotomized groups of different histologic grades of steatosis and iron overload. For each receiver operating characteristic analysis, the optimal cutoff value was selected using the Youden index to maximize sensitivity and specificity. The DeLong test was used

to compare the AUCs of manual ROIs and automated WLSs for specific grades of steatosis and iron overload. Linear regression analysis was performed to determine the correlation (Pearson correlation coefficient [r] and Spearman rank correlation [ρ] coefficients) with the histologic proportionate areas of fat and iron. A t test was used to compare the correlation coefficients.

According to previous studies (5,29,30), the expected proportion of steatosis and/or iron overload among patients with chronic liver disease undergoing a clinically indicated liver biopsy is approximately 50%. Sample size was calculated for estimating the diagnosis accuracy (AUC) of automated WLS compared with manual ROI by using a DeLong test for detecting an effect of δ of 0.05, with a high accuracy index (AUC > 0.90) (4,5,30,31). The level of statistically significant difference and power were set at $P < .05$ and .80, respectively, leading to an estimated necessary sample size of 172 participants.

Results

Participant Characteristics

We initially enrolled 183 participants with chronic liver disease who underwent liver biopsy (Fig 1); after exclusion, the final sample included 165 participants (mean age, 55 years \pm 12; 96 women). The demographic, laboratory, and histopathologic features of participants are summarized in Tables 1 and 2. In parallel with a high prevalence of obesity (69 of 165 participants [42%]) and metabolic syndrome (63 of 165 [38%]), the main chronic liver disease cause was NAFLD (101 of 165 [61%]). Only one participant had hemochromatosis. At recruitment, most participants (160 of 165 [97%]) had preserved liver function (median Model for End-Stage Liver Disease score, 7 [interquartile range, 6–7]). Histologic review of the 165 biopsy specimens showed that 33 participants (20%) had mild steatosis (S1), 34 (21%) had moderate steatosis (S2), and 28 (17%) had severe steatosis. Increased liver iron deposits ($\text{Fe} \geq 1$) were observed in 36 of 165 specimens (22%). The median time interval between biopsy and MRI was 18 days (range, 13–24 days). No adverse events related to biopsy or MRI were reported.

Quantitative Imaging of Marker Distribution

The mean \pm SD baseline liver PDFF and $R2^*$ values were $10.3\% \pm 5.3$ and $57.6 \text{ sec}^{-1} \pm 31.5$ for the ROI assessments and $10.8\% \pm 4.9$ and $59.6 \text{ sec}^{-1} \pm 27.3$ for the WLS assessments ($P < .001$).

The MOEs of the PDFF and $R2^*$ means were 0.41% and 2.45 sec^{-1} , respectively, for the ROI assessment and 0.38% and 2.12 sec^{-1} for the WLS assessment. MOEs were lower for WLS than ROI-derived parameters (-0.03% for PDFF and -0.3 sec^{-1} for $R2^*$). Agreement between central tendency measures of both methods was excellent, with an ICC of 0.98 (95% CI: 0.97, 0.99) for PDFF and 0.98 (95% CI: 0.97, 0.99) for $R2^*$.

Measures that described data dispersion were also analyzed. The means \pm SD for PDFF values were $6.5\% \pm 1.2$ for ROI and $8.3\% \pm 0.9$ for the WLS assessment ($P < .001$) (Fig 3). The median (IQR) for $R2^*$ values was 9.5 sec^{-1} (interquartile range, $7.4\text{--}11.3 \text{ sec}^{-1}$) for ROI and 22.5 sec^{-1} (interquartile range, $19.1\text{--}25.3 \text{ sec}^{-1}$) for WLS assessment ($P < .001$). The MOEs of the PDFF and $R2^*$ SDs were 0.09% and 2.4 sec^{-1} , respectively, for the ROI assessment and 0.07% and 0.6 sec^{-1} for the WLS assessment. Agreement of the variability measures (SD) between both methods was poor, with ICCs of 0.46 (95% CI: -0.16 , 0.79) for PDFF and 0.36 (95% CI: 0.14, 0.53) for $R2^*$.

Assessment of Linear Regression and Bias

Excellent correlation ($R^2 = 0.94$) was observed between the two segmentation strategies and their PDFF values (Fig E2A [online]). Both segmentations slightly disagreed as evidenced by a slope \pm SD of 1.05 ± 0.02 ($P < .001$) and an intercept of

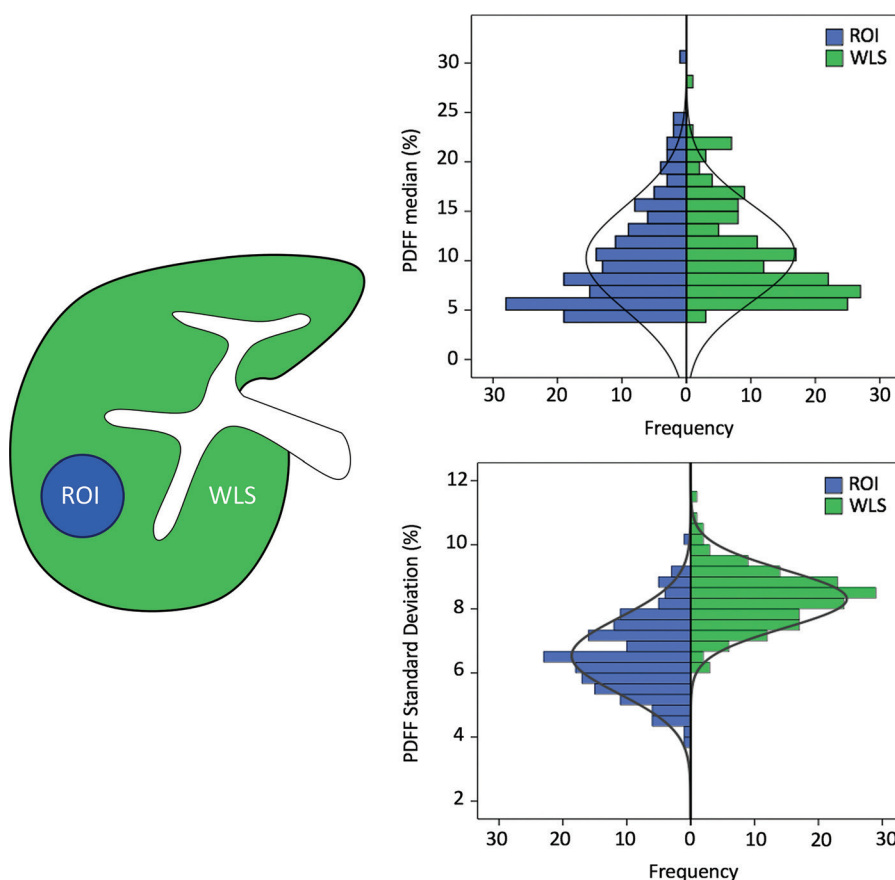


Figure 3: Histograms summarize proton density fat fraction (PDFF) (expressed as percentages) measures of central tendency (median, top image) and dispersion (standard deviation [SD], bottom image) values. The automated whole-liver segmentation (WLS) (green) strategy shows a higher median and SD, with a lower margin of error (ie, the degree of uncertainty due to sampling variation) than the manual region of interest (ROI) (blue) strategy. Frequency represents the number of observations.

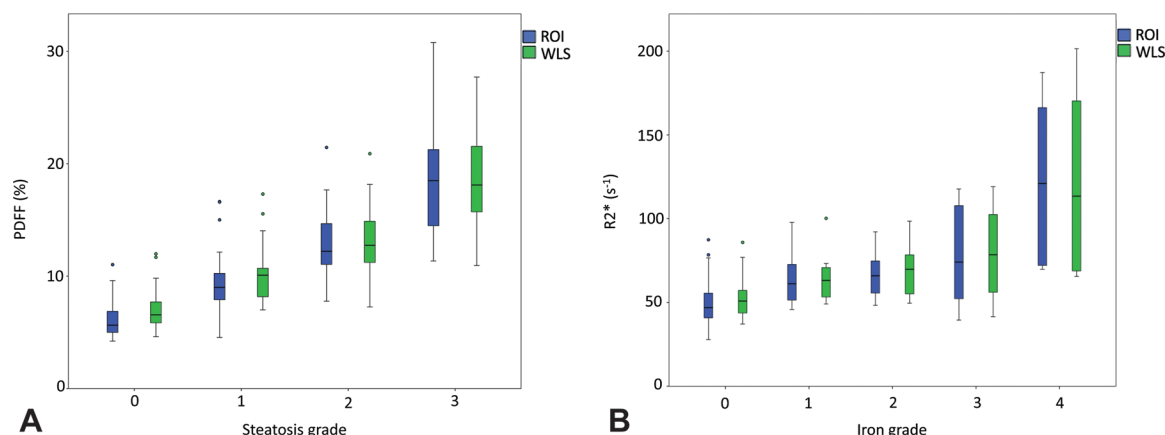


Figure 4: Box and whisker plots of MRI markers of **(A)** proton density fat fraction (PDFF) (expressed as percentages) and **(B)** transverse relaxometry ($R2^*$) (expressed as sec^{-1}) versus histologic grades. Blue boxes correspond to manual region of interest (ROI) and green boxes to automated whole-liver segmentation (WLS). Midlines indicate medians, boxes represent upper and lower quartiles, whiskers denote upper and lower extremes, and dots indicate outliers.

-0.99 ± 0.25 ($P < .001$). The median bias was 0.5% (interquartile range, -0.4% to 1.2%), which carried the most weight with lower PDFF values ($P < .001$) (Fig E3A [online]).

For iron concentration, excellent correlation ($R^2 = 0.96$) was also observed between the two segmentation strategies (Fig E2B [online]), although a slight disagreement between methods was evidenced by a slope \pm SD of 1.02 ± 0.02 ($P < .001$) and an intercept of -3.39 ± 1.11 ($P = .003$). The median 2.7 sec^{-1} (interquartile range, 0.2 – 5.3 sec^{-1}) bias was small and carried the most weight with lower $R2^*$ values ($P < .001$) (Fig E3B [online]).

Performance Analysis of MRI against Histologic Grades

Both ROI- and WLS-derived PDFF values were higher with steatosis (Fig 4A) and were able to discriminate between all individual grades of steatosis (S0, $6.8\% \pm 1.2$; S1, $9.9\% \pm 2.3$; S2, $13.2\% \pm 2.9$; S3, $18.3\% \pm 3.9$; $P < .001$ for WLS). There were differences between median PDFF values quantified with ROI and WLS assessments among individual grades of no (S0) and moderate (S2) steatosis (Table 3). For detecting steatosis (S0 vs S1–S3), the AUC was 0.96 (95% CI: 0.93, 0.98) for ROI and 0.97 (95% CI: 0.94, 0.99) for WLS (Fig 5A). The DeLong test showed no difference when comparing the AUCs of manual ROI and automated WLS for detecting steatosis ($P = .53$). Table 4 shows the receiver operating characteristic analysis for each set of dichotomized groups of different steatosis grades.

Additionally, both ROI- and WLS-derived $R2^*$ values were higher with iron overload (Fe0, $52.4 \text{ sec}^{-1} \pm 10.7$; Fe1, $65.4 \text{ sec}^{-1} \pm 14.8$; Fe2, $69.3 \text{ sec}^{-1} \pm 16.4$; Fe3, $79.0 \text{ sec}^{-1} \pm 26.5$; Fe4, $122.2 \text{ sec}^{-1} \pm 56.5$; $P < .001$ for WLS) but able to discriminate only Fe0 and Fe4 among the iron grades (Fig 4B). There were differences between mean $R2^*$ values quantified with ROI and WLS only among the individual grade of no iron granules (Table 3). For detecting iron deposits (Fe0 vs Fe1–Fe4), the AUC was 0.85 (95% CI: 0.77, 0.92) for ROI and 0.83 (95% CI: 0.75, 0.91) for WLS (Fig 5B). Table 4 shows the receiver operating characteristic analysis for each set of dichotomized groups of different iron grades.

Table 3: Results of Manual ROI and Automated WLS Measurements of PDFF and $R2^*$ for Each Grade of Steatosis and Iron Deposits

MRI Marker and Histologic Grade	Manual ROI	Automated WLS	P Value
PDFF, steatosis (%)			
S0	6.1 ± 1.5	6.8 ± 1.2	$<.001$
S1	9.6 ± 2.7	9.9 ± 2.3	.17
S2	12.7 ± 3.0	13.2 ± 2.9	.03
S3	18.4 ± 4.6	18.3 ± 3.9	.80
$R2^*$, iron (sec^{-1})			
Fe0	48.6 ± 11.6	52.4 ± 10.7	$<.001$
Fe1	64.4 ± 16.5	65.4 ± 14.8	.20
Fe2	66.6 ± 14.6	69.3 ± 16.4	.07
Fe3	79.2 ± 29.1	79.0 ± 26.5	.66
Fe4	122.9 ± 52.7	122.2 ± 56.5	.25

Note.—Unless otherwise specified, data are means \pm standard deviations. PDFF = proton density fat fraction, $R2^*$ = transverse relaxometry, ROI = region of interest, WLS = whole-liver segmentation.

The DeLong test showed no difference when comparing the AUCs of manual ROI and automated WLS for detecting iron deposits ($P = .09$).

Correlation Analysis of MRI against Digital Image Analysis

A strong correlation was seen between PDFF values and digital pathology fat percentages ($r = 0.89$ and $\rho = 0.84$ for ROI [$P < .001$]; $r = 0.90$ and $\rho = 0.87$ for WLS [$P < .001$]). No difference was found between these correlation coefficients ($P = .24$) (Fig 6A).

A weak correlation was seen between $R2^*$ values and the percentages of iron quantified with digital pathology ($r = 0.65$ and $\rho = 0.24$ for ROI [$P < .002$]; $r = 0.64$ and $\rho = 0.26$ for WLS [$P < .001$]), and we found no evidence of a difference between these correlation coefficients ($P = .35$) (Fig 6B).

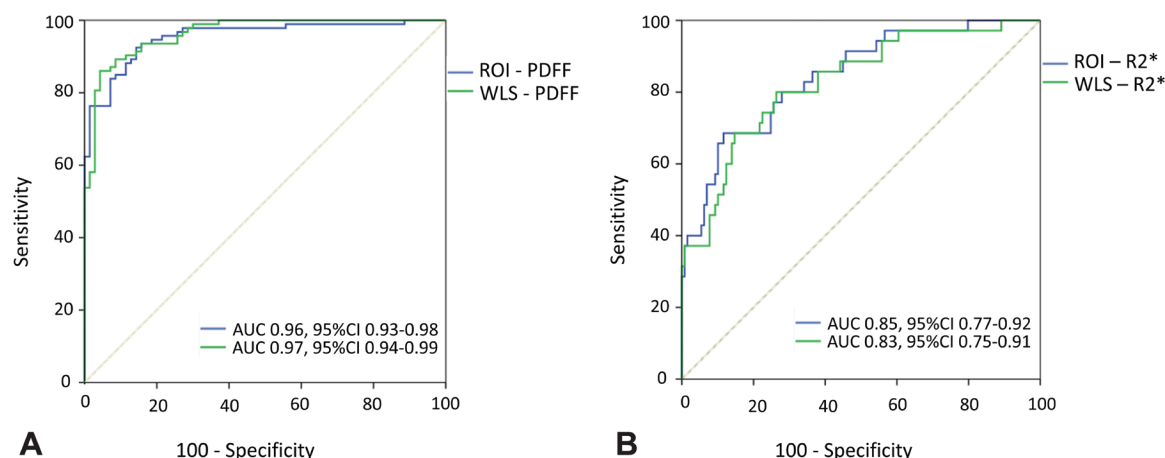


Figure 5: Receiver operating characteristic curves for MRI markers for discriminating the presence of histologic features. **(A)** Proton density fat fraction (PDFF)-based diagnosis of steatosis (S0 vs S1–S3) and **(B)** transverse relaxometry (R2*)-based diagnosis of iron overload (Fe0 vs Fe1–Fe4). The blue line corresponds to manually drawn regions of interest (ROIs) and the green line to automated whole-liver segmentations (WLSs). The diagonal dotted gray line represents the random classifier. AUC = area under the receiver operating characteristic curve.

Table 4: Diagnostic Performance of MRI PDFF and R2* Thresholds for Steatosis and Iron Grading: Raw Performance Parameters for Different Segmentation Strategies

MRI Marker and Histologic Grade	Manual ROI Cutoff*	Automated WLS Cutoff*	Manual ROI Sensitivity and Specificity†	Automated WLS Sensitivity and Specificity†	DeLong P Value
PDFF, steatosis (%)					
≥S1	7.7 (0.96 [0.93, 0.98])	7.8 (0.97 [0.94, 0.99])	94 (88/94) 84 (59/70)	94 (88/94) 84 (59/70)	.53
≥S2	10.2 (0.95 [0.92, 0.98])	10.8 (0.96 [0.93, 0.99])	92 (56/61) 90 (93/103)	89 (54/61) 91 (94/103)	.66
≥S3	12.8 (0.96 [0.93, 0.99])	14.1 (0.96 [0.93, 0.99])	93 (26/28) 87 (118/136)	89 (25/28) 91 (124/136)	.76
R2*, iron (sec⁻¹)					
≥Fe1	54 (0.85 [0.77, 0.92])	55 (0.83 [0.75, 0.91])	80 (28/35) 72 (93/129)	80 (28/35) 74 (95/129)	.09
≥Fe2	55 (0.85 [0.77, 0.94])	56 (0.84 [0.74, 0.93])	84 (21/25) 74 (103/139)	84 (21/25) 71 (99/139)	.14
≥Fe3	66 (0.86 [0.75, 0.98])	64 (0.85 [0.73, 0.97])	82 (14/17) 86 (126/147)	83 (14/17) 81 (119/147)	.22
≥Fe4	72 (0.96 [0.91, 0.99])	69 (0.94 [0.87, 0.99])	86 (6/7) 89 (139/157)	86 (6/7) 82 (128/157)	.19

Note.—Performance is presented as the area under the receiver operating characteristic curve (AUC), with 95% CIs and sensitivity and specificity values according to the optimal selected cutoff. PDFF = proton density fat fraction, R2* = transverse relaxometry, ROI = region of interest, WLS = whole-liver segmentation.

* Data in parentheses are AUCs, with 95% CIs in brackets.

† Data in parentheses are numbers of participants.

Discussion

In clinical practice, MRI measurements of liver fat and iron are pivotal in the growing number of patients with metabolic disorders, paving the need for computerized workflows for analysis. This study evaluated an automatic whole-liver volume segmentation model for multiecho chemical shift–encoded gradient-echo MRI quantification of fat (proton density fat fraction) and iron (transverse relaxometry) in the assessment of chronic liver diseases. Our findings demonstrate that measurements derived from whole-

liver segmentation are accurate for steatosis grading (area under the receiver operating characteristic curve, 0.97) and strongly correlate with digital pathology fat ratio ($r = 0.90$), showing higher mean ($+0.5\%$ mean difference, $P < .001$) and dispersion ($+1.8\%$ mean difference, $P < .001$) values with a lower degree of uncertainty (-0.03%) than manual region of interest sampling.

The CNN model for WLS was previously developed in patients with suspected liver disease but normal PDFF and R2* values (median, 5.6% [range, $0\%–28.4\%$] for PDFF and 47.7

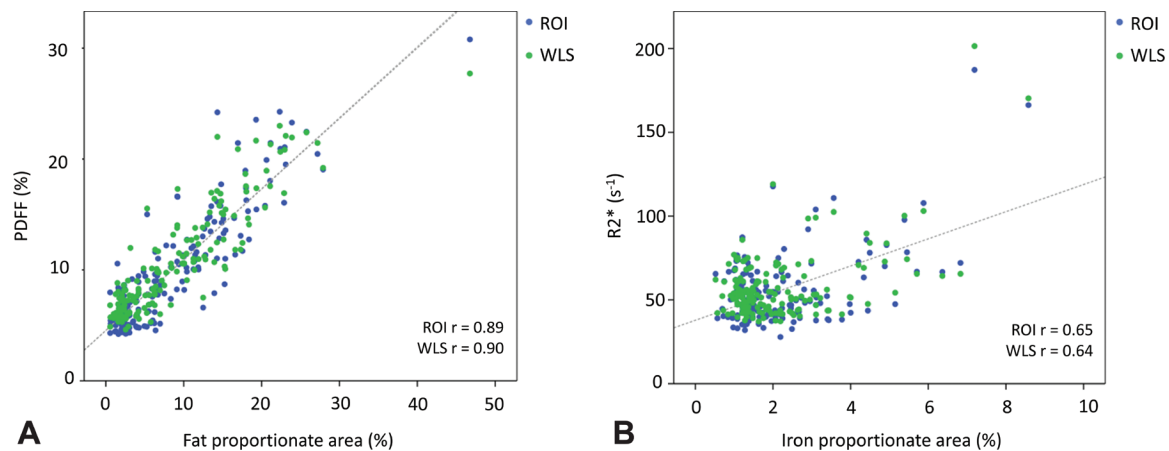


Figure 6: Scatterplots of MRI parameters and digital image analysis, categorized according to sampling strategy. **(A)** Proton density fat fraction (PDFF) versus the fat ratio and **(B)** transverse relaxometry ($R2^*$) versus the iron ratio. Blue dots correspond to manually drawn regions of interest (ROIs) and green dots to automated whole-liver segmentations (WLSs). The gray dotted line represents the linear regression fit of WLS data.

sec^{-1} [range, 28.6–200.2 sec^{-1}] for $R2^*$) (21). The present study was performed in a clinically well-characterized sample of 165 patients with chronic liver disease with greater degrees of liver injury (mean PDFF, 11% and mean $R2^*$, 60 sec^{-1}). ICCs for agreement between manual WLS and manual ROI for PDFF measurements ranged from 0.87 to 0.97, with an at-most bias of 2.8% (10,13). Our CNN algorithm achieved higher ICCs and lower bias, probably related to the automatic methods with no variability. Agreement between automated WLS and manual WLS for PDFF and $R2^*$ measurements has been previously assessed, with excellent correlations (≥ 0.98) between both parameters being reported (16–18,21,32). Among these segmentation accuracy studies, mean bias ranged from -0.4% to $+0.6\%$ for PDFF and $+3.93\%$ to $+4.51\%$ for $R2^*$ (16–18,21,32). Similar metrics were obtained in our study, with marginal overestimation of PDFF and $R2^*$. Although statistically significant, differences are unlikely to be clinically relevant, given that they are smaller than reported inpatient variability (14). As similar diagnostic accuracy to manual segmentation is obtained, the main advantage of the automated segmentation in clinical practice is the absence of variability and procedure time savings.

The comparison of WLS evaluations with histologic data, to our knowledge, has not previously been well established in the literature. Our study adds evidence by applying the developed automated WLS technique in a large sample of participants with chronic liver disease having a wide spectrum of liver injuries as evaluated with use of liver biopsy. As a variety of liver disease causes were included, the obtained PDFF and $R2^*$ cutoff values for different histologic grades are lower than those in studies that investigated NAFLD or iron overload diseases (3,5,6). Although obtained PDFF cutoffs for grade S1 steatosis are in line with proposed intervals, the slightly high value probably relates to simultaneous accumulation of lipid droplets with other histologic features of liver disease and a weak but positive influence from inflammation and fibrosis (33–35). The automated WLS AUC values for each set of dichotomized groups of different histologic grades were high for both steatosis and iron deposits, with no statistical

difference when comparing the AUCs of both segmentation methods.

To be useful diagnostic markers, MRI-derived parameters must demonstrate strong performance in quantifying hepatic changes throughout the liver in all subsamples compared with the currently accepted histologic reference standard (36). Strong correlations between histologic features and PDFF or $R2^*$ have already been demonstrated and validated using manual ROI selection (3–6,30). Our study confirms the performance of the proposed WLS method, with higher mean and dispersion (SD) values. These higher values may be related to WLS being a more representative sampling that includes the subcapsular peripheral liver and areas close to the falciform ligament and hilum, which usually have higher fat deposits. Furthermore, MOE analysis shows that WLS sampling distribution of the mean has a lower degree of uncertainty and discrepancy, with higher CIs in the produced results. A higher MOE with manual ROI assessment might suggest a higher sampling error or variation (37).

Linear association was assessed against the reference standard of digital image analysis. Biopsy digital image analysis shows a strong correlation with steatosis and iron grades in patients with NAFLD and chronic viral hepatitis (20,38,39). Automated analysis of digitized whole-slide images allows an objective, precise, and quantitative assessment of histologic features as ground truth (27). PDFF has been previously compared with pathologists' visual estimations of the percentage of hepatocytes containing fat, considered a continuous scale, with a good correlation between MRI and subjective readings ($r = 0.85$ – 0.87) (3,30). Only one study (29) used digital pathology, showing an excellent correlation with PDFF ($r = 0.97$), but it included only 17 patients and lacked an appropriate distribution of steatosis severity (41% of patients had $<5\%$ fat, compared with 30% with $>22\%$ fat) or reliable digital image analysis methods. Our study, which evaluated a large sample with use of adequate objective digital image analysis methods, confirmed a high PDFF–fat proportionate area correlation ($r = 0.90$). In contrast, only a moderate correlation was seen for the $R2^*$ –iron proportionate area correlation ($r = 0.64$). This finding might be due to the small prevalence and low spectrum of iron overload and to the lower performance of

R2* in patients with NAFLD, attributable to the influence of fat (40). Pathologic digital analysis and deep learning MRI organ segmentation offer a robust method of evaluating liver markers. The automated segmentation reduces the radiologist interpretation time and operator dependence on MRI in the assessment of hepatic fat and iron.

Some limitations must be highlighted. First, calculated sample size was slightly higher than final recruitment (172 vs 165 participants) due to more exclusions than expected. This small difference might have a minor influence on the results. Second, participants with iron overload were underrepresented because liver iron deposition is relatively rare and of low grade in patients with NAFLD. Third, the proposed segmentation algorithm was not able to exclude small vessels within the segmented volume, which may have introduced a small systematic bias for the estimation of the true liver PDFF and R2* values. Finally, accuracy of CNN WLS, taking liver biopsy as a reference, was tested with a single-scanner data, requiring validation under different conditions (pathologists and image data).

In conclusion, this prospective study established the performance of a convolutional neural network (CNN) whole-liver segmentation (WLS) algorithm for MRI proton density fat fraction and transverse relaxometry estimations in participants with chronic liver disease in whom histopathologic evaluation was performed. This automatic sampling technique has implications in clinical practice for the following reasons: (a) WLS assessments have a similar diagnostic accuracy to manual region of interest (ROI) assessments, stratified according to histologic grades and correlated with digital image analysis; (b) marginal overestimation of WLS-derived measurements has no clinical relevance and reveals better representability of liver injury; and (c) lower variability errors and degree of uncertainty than estimations derived from manually outlined ROIs show the outperformance of deep learning methods. The accurate MRI automatic WLS approach can be extended and seamlessly integrated into the clinical practice workflow to streamline liver fat and iron quantification. In future studies, we plan to improve MRI accuracy for liver fat and iron characterization by using CNN solutions in a larger sample of diseases and extending the automatic segmentation algorithm and analysis to other organs, such as the pancreas and muscles, for better disease management.

Author contributions: Guarantors of integrity of entire study, D.M.A., A.R.O., A.G.P., E.V.P., J.P.R.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, D.M.A., Á.A.B., A.R.O., A.G.P., J.P.R., V.P., D.E.G., M.A.S., L.M.B.; clinical studies, D.M.A., C.A.C., M.B., A.G.P., E.V.P., M.P.B., M.G.T., A.P.G., S.B., J.P.R., V.P., A.F., V.A., D.E.G., M.A.S.; experimental studies, A.J.P., Á.A.B., A.R.O., C.M.A., M.B., A.P.G., J.P.R., D.E.G., M.A.S.; statistical analysis, D.M.A., Á.A.B., J.P.R., D.E.G., M.A.S.; and manuscript editing, D.M.A., Á.A.B., A.R.O., C.A.C., C.M.A., M.B., M.P.B., M.G.T., A.P.G., J.P.R., A.F., V.A., D.E.G., M.A.S., L.M.B.

Disclosures of conflicts of interest: D.M.A. No relevant relationships. A.J.P. Employee of Quibim. Á.A.B. Chief executive officer of Quibim; patents or patents pending for chest radiograph classification and white matter lesion detection; shareholder at Quibim. A.R.O. No relevant relationships. C.A.C. No relevant relationships. C.M.A. No relevant relationships. M.B. No relevant relationships. A.G.P. No relevant relationships. E.V.P. No relevant relationships. M.P.B. No relevant relationships. M.G.T. No relevant relationships. A.P.G. No relevant relationships.

S.B. No relevant relationships. J.P.R. No relevant relationships. V.P. No relevant relationships. A.F. No relevant relationships. V.A. No relevant relationships. D.E.G. No relevant relationships. M.A.S. No relevant relationships. L.M.B. Member of the Quibim scientific advisory board.

Data sharing: Data generated or analyzed during the study are available from the corresponding author by request.

References

1. GBD 2017 Cirrhosis Collaborators. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* 2020;5(3):245–266.
2. Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J Hepatol* 2020;73(6):1322–1332.
3. Tang A, Desai A, Hamilton G, et al. Accuracy of MR imaging–estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. *Radiology* 2015;274(2):416–425.
4. Gu J, Liu S, Du S, et al. Diagnostic value of MRI-PDFF for hepatic steatosis in patients with non-alcoholic fatty liver disease: a meta-analysis. *Eur Radiol* 2019;29(7):3564–3573.
5. França M, Alberich-Bayarri Á, Martí-Bonmati L, et al. Accurate simultaneous quantification of liver steatosis and iron overload in diffuse liver diseases with MRI. *Abdom Radiol (NY)* 2017;42(5):1434–1443.
6. Eddowes PJ, McDonald N, Davies N, et al. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. *Aliment Pharmacol Ther* 2018;47(5):631–644.
7. França M, Carvalho JG. MR imaging assessment and quantification of liver iron. *Abdom Radiol (NY)* 2020;45(11):3400–3412.
8. Hong CW, Wolfson T, Sy EZ, et al. Optimization of region-of-interest sampling strategies for hepatic MRI proton density fat fraction quantification. *J Magn Reson Imaging* 2018;47(4):988–994.
9. Campo CA, Hernando D, Schubert T, Bookwalter CA, Pay AJV, Reeder SB. Standardized approach for ROI-based measurements of proton density fat fraction and R2* in the liver. *AJR Am J Roentgenol* 2017;209(3):592–603.
10. Procter AJ, Sun JY, Malcolm PN, Toms AP. Measuring liver fat fraction with complex-based chemical shift MRI: the effect of simplified sampling protocols on accuracy. *BMC Med Imaging* 2019;19(1):14.
11. Song J, Yu X, Song W, et al. MRI-based radiomics models developed with features of the whole liver and right liver lobe: assessment of hepatic inflammatory activity in chronic hepatic disease. *J Magn Reson Imaging* 2020;52(6):1668–1678.
12. Kang BK, Kim M, Song SY, Jun DW, Jang K. Feasibility of modified Dixon MRI techniques for hepatic fat quantification in hepatic disorders: validation with MRS and histology. *Br J Radiol* 2018;91(1089):20170378.
13. Vu KN, Gilbert G, Chalut M, Chagnon M, Chartrand G, Tang A. MRI-determined liver proton density fat fraction, with MRS validation: comparison of regions of interest sampling methods in patients with type 2 diabetes. *J Magn Reson Imaging* 2016;43(5):1090–1099.
14. Yokoo T, Serai SD, Pirasteh A, et al. Linearity, bias, and precision of hepatic proton density fat fraction measurements by using MR imaging: a meta-analysis. *Radiology* 2018;286(2):486–498.
15. Hooker JC, Hamilton G, Park CC, et al. Inter-reader agreement of magnetic resonance imaging proton density fat fraction and its longitudinal change in a clinical trial of adults with nonalcoholic steatohepatitis. *Abdom Radiol (NY)* 2019;44(2):482–492.
16. Tipirneni-Sajja A, Song R, McCarville MB, Loeffler RB, Hankins JS, Hillenbrand CM. Automated vessel exclusion technique for quantitative assessment of hepatic iron overload by R2*-MRI. *J Magn Reson Imaging* 2018;47(6):1542–1551.
17. Wang K, Mamidipalli A, Retson T, et al. Automated CT and MRI liver segmentation and biometry using a generalized convolutional neural network. *Radiol Artif Intell* 2019;1(2):180022.
18. Stocker D, Bashir MR, Kannengiesser SAR, Reiner CS. Accuracy of automated liver contouring, fat fraction, and R2* measurement on gradient multiecho magnetic resonance images. *J Comput Assist Tomogr* 2018;42(5):697–706.
19. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
20. Martí-Aguado D, Rodríguez-Ortega A, Mestre-Alagarda C, et al. Digital pathology: accurate technique for quantitative assessment of histological

- features in metabolic-associated fatty liver disease. *Aliment Pharmacol Ther* 2021;53(1):160–171.
21. Jimenez-Pastor A, Alberich-Bayarri A, Lopez-Gonzalez R, et al. Precise whole liver automatic segmentation and quantification of PDFF and R2* on MR images. *Eur Radiol* 2021;31(10):7876–7887.
 22. Martí-Aguado D, Alberich-Bayarri A, Martín-Rodríguez JL, et al. Differences in multi-echo chemical shift encoded MRI proton density fat fraction estimation based on multifrequency fat peaks selection in non-alcoholic fatty liver disease patients. *Clin Radiol* 2020;75(11):880.e5–880.e12.
 23. Roberts NT, Hernando D, Holmes JH, Wiens CN, Reeder SB. Noise properties of proton density fat fraction estimated using chemical shift-encoded MRI. *Magn Reson Med* 2018;80(2):685–695.
 24. Grundy SM, Brewer HB Jr, Cleeman JI, et al. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation* 2004;109(3):433–438.
 25. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;41(6):1313–1321.
 26. Deugnier Y, Turlin B. Pathology of hepatic iron overload. *World J Gastroenterol* 2007;13(35):4755–4760.
 27. Melo RCN, Raas MWD, Palazzi C, Neves VH, Malta KK, Silva TP. Whole slide imaging and its applications to histopathological studies of liver disorders. *Front Med (Lausanne)* 2020;6:310.
 28. Raunig DL, McShane LM, Pennello G, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Stat Methods Med Res* 2015;24(1):27–67.
 29. Kukuk GM, Hittatiya K, Sprinkart AM, et al. Comparison between modified Dixon MRI techniques, MR spectroscopic relaxometry, and different histologic quantification methods in the assessment of hepatic steatosis. *Eur Radiol* 2015;25(10):2869–2879.
 30. Kühn JP, Hernando D, Muñoz del Río A, et al. Effect of multipoint spectral modeling of fat for liver iron and fat quantification: correlation of biopsy with MR imaging results. *Radiology* 2012;265(1):133–142.
 31. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform* 2014;48:193–204.
 32. Dzyubak B, Li J, Chen J, et al. Automated analysis of multiparametric magnetic resonance imaging/magnetic resonance elastography exams for prediction of nonalcoholic steatohepatitis. *J Magn Reson Imaging* 2021;54(1):122–131.
 33. Kinner S, Reeder SB, Yokoo T. Quantitative imaging biomarkers of NAFLD. *Dig Dis Sci* 2016;61(5):1337–1347.
 34. Khalili M, Kleiner DE, King WC, et al. Hepatic steatosis and steatohepatitis in a large North American cohort of adults with chronic hepatitis B. *Am J Gastroenterol* 2021. 10.14309/ajg.0000000000001257. Published online April 8, 2021.
 35. Dennis A, Kelly MD, Fernandes C, et al. Correlations between MRI biomarkers PDFF and cT1 with histopathological features of non-alcoholic steatohepatitis. *Front Endocrinol (Lausanne)* 2021;11:575843.
 36. Dulai PS, Sirlin CB, Loomba R. MRI and MRE for non-invasive quantitative assessment of hepatic steatosis and fibrosis in NAFLD and NASH: clinical trials to clinical practice. *J Hepatol* 2016;65(5):1006–1016.
 37. Caussy C, Reeder SB, Sirlin CB, Loomba R. Noninvasive, quantitative assessment of liver fat by MRI-PDFF as an endpoint in NASH trials. *Hepatology* 2018;68(2):763–772.
 38. Forlano R, Mullish BH, Giannakeas N, et al. High-throughput, machine learning-based quantification of steatosis, inflammation, ballooning, and fibrosis in biopsies from patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2020;18(9):2081–2090.e9.
 39. Mendes LC, Ferreira PA, Miotto N, et al. Controlled attenuation parameter for steatosis grading in chronic hepatitis C compared with digital morphometric analysis of liver biopsy: impact of individual elastography measurement quality. *Eur J Gastroenterol Hepatol* 2018;30(8):959–966.
 40. Bashir MR, Wolfson T, Gamst AC, et al. Hepatic R2* is more strongly associated with proton density fat fraction than histologic liver iron scores in patients with nonalcoholic fatty liver disease. *J Magn Reson Imaging* 2019;49(5):1456–1466.