Title: Accuracy of magnitude- and complex-reconstruction chemical-shift-encoded magnetic resonance imaging proton density fat fraction for diagnosis of hepatic steatosis in subjects with severe obesity using histology as reference

Author List:
Tydus T. Thai, BS
Gavin Hamilton, PhD
Yesenia Covarrubias, MFS
Jonathan C. Hooker, BS
Alex N. Schlein, BS
Michael S. Middleton, MD, PhD
Michelle L. Ballanay, MEd
William M. Haufe, BS
Curtis N. Wiens, PhD
Nathan S. Artz, PhD
Luke M. Funk, MD, MPH
Alan McMillan, PhD
Rashmi Agni, MD
Michael Peterson, MD, PhD
Guilherme M. Campos, MD
Jacob Greenberg, MD, EdM
Santiago Horgan, MD
Garth Jacobson, MD
Tanya Wolfson, MA
Anthony Gamst, PhD
Jeffrey Schwimmer, MD
Scott B. Reeder, MD, PhD
Claude B. Sirlin, MD

Abstract

Purpose

To assess and compare the diagnostic performance of magnitude-reconstruction chemical-shift-encoded magnetic resonance imaging (MRI-M) and complex-reconstruction chemical-shift-encoded magnetic resonance imaging (MRI-C) for diagnosis of hepatic steatosis in subjects with severe obesity without known non-alcoholic fatty liver disease (NAFLD), using contemporaneous histology as reference.

Materials and Methods

This is an IRB-approved, HIPAA-compliant, two-center, cross-sectional study of a larger prospective trial that recruited patients without known NAFLD consecutively between October 2010 and March 2015 to undergo research MRI exams 1-2 days prior to clinical-care weight-loss surgery. Proton density fat fraction (PDFF) was estimated using MRI-M and MRI-C. Liver biopsies were obtained intraoperatively. Using histologically-determined presence of steatosis as the reference standard, receiver operating characteristics (ROC) analyses were used to identify MRI-M- and MRI-C-derived PDFF thresholds for diagnosing steatosis. Bootstrapped-based tests were used to compare their diagnostic performance.

Results

A total of 81 patients (67 female, 14 male, average age 48.2) were recruited for this study. MRI-M and MRI-C had areas under the ROC curve of 0.951 and 0.947, respectively, for diagnosing hepatic steatosis. For MRI-M, the Youden-index-based PDFF threshold of 6.5% provided 0.87 sensitivity (95% confidence internal: 0.75, 0.95), 0.96 specificity (0.81, 0.99), and 0.90 total accuracy (0.82, 0.96). For MRI-C, a PDFF threshold of 6.8%, provided 0.90 sensitivity

(0.77, 0.96), 0.96 specificity (0.81, 0.99), and 0.91 total accuracy (0.83, 0.97). Differences in performance parameters between MRI-M and MRI-C were not statistically significant.

Conclusion

MRI-M- and MRI-C-derived PDFF is accurate for non-invasive diagnosis of hepatic steatosis in subjects with severe obesity.

Introduction

Confounder-corrected chemical-shift-encoded magnetic resonance imaging (CSE-MRI) methods have been developed to estimate the proton density fat fraction (PDFF), an emerging quantitative biomarker of liver fat content defined as the ratio of fat proton density to total (fat and water) proton density (1-4). These methods minimize or correct for biases caused by T1, T2*, and the spectral complexity of fat. MRI-C also addresses noise-related bias and the effects of eddy currents (5-10). Studies have shown that PDFF correlates well with histologically-determined hepatic steatosis grades and suggest that it can diagnose the presence of steatosis, using contemporaneous histology as reference (11, 12-17).

A recently published MRI-M-based PDFF threshold of 6.4% had 86% sensitivity and 83% specificity to diagnose the presence of hepatic steatosis (17). However, that study enrolled subjects with known or clinically-suspected nonalcoholic fatty liver disease (NAFLD), resulting in the cohort being enriched in NAFLD. Since only 7% of enrolled subjects lacked steatosis, the specificity estimate had wide 95% confidence intervals (CIs) (0.36, 1.00). This may have also altered the actual specificity estimate and optimal diagnostic threshold. Additionally, the study did not include MRI-C. Since MRI-M is known to underestimate PDFF compared to MRI-C, a cutoff developed for MRI-M may be suboptimal for MRI-C (18), and their relative accuracy for diagnosing hepatic steatosis remains uncertain.

Therefore, the purpose of this study was to assess and compare the diagnostic performance of MRI-M- and MRI-C-derived PDFF for detecting hepatic steatosis in subjects with severe obesity, using contemporaneous histology as reference. These subjects are at risk for but not known a priori to have hepatic steatosis and so are expected to have a more balanced

distribution of hepatic steatosis, which may permit more robust estimation of diagnostic thresholds and their performance.

Materials and Methods

Study Design and Subjects

This Health Insurance Portability and Accountability Act (HIPAA)-compliant, two-center, cross-sectional study of a wider prospective study was approved by our institutional review board (IRB). Inclusion criteria included adults with severe obesity (BMI≥35kg/m^2, age≥18 years) undergoing weight loss surgery (laparoscopic sleeve gastrectomy or laparoscopic Roux-en-Y gastric bypass). Exclusion criteria included contraindications to MRI (e.g., claustrophobia, metallic implants), excessive alcohol consumption (>1.5 drinks per day), and type 1 diabetes.

Subjects provided written informed consent and underwent MR examinations at multiple time points as part of a wider longitudinal study examining changes in PDFF following weight loss surgery (WLS). PDFF estimates for this study were derived from MR examinations that took place after a two-to-four week, very low-calorie liquid diet and immediately prior to WLS. Subjects were excluded from this cross-sectional study if one or both of the MRI sequences methods was not acquired, or if intraoperative biopsy was not obtained. Subject demographics, time intervals from MRI to surgery, and reasons for exclusion were recorded.

Liver biopsy and histology

Surgeons performed wedge and/or core biopsies during WLS based on their judgment of the safest procedure in individual subjects. Wedge biopsies, typically about 1-4 cm^2 at the surface and 1-2 cm of depth, were taken from the anterior surface of the left lateral sector of the liver. Core biopsies, each weighing approximately 20-30 mg, were taken from the superficial

portions of the same sector to avoid blood vessels and major bile ducts. Histology slides were prepared with hematoxylin and eosin, Masson's trichome, and iron stains.

Blinded to the MRI results, two hepatopathologists (each with at least three years of experience) independently scored each biopsy specimen (wedge, core, or both) in each subject. Together they reviewed each specimen, adjudicated any disagreements, and assigned a final consensus score for each of several histology features.

Based on the estimated percentage of hepatocytes containing microscopically visible fat vacuoles, hepatic steatosis was scored in a granular fashion as 0%, 1%, 5%, 10%, 15%, 20%…, 100%. Whenever both core and wedge biopsies were performed, the scores were averaged to yield composite scores, which were then converted into the 4-point ordinal scoring system described by the NASH CRN (19): grade 0, <5%; grade 1, 5-33%, grade 2, 34-66%; grade 3, >66%). The hepatic steatosis score for each subject was further binarized as steatosis present (grade ≥ 1) or absent (grade 0). Lobular inflammation was scored according to the number of inflammatory foci per magnification field 200X as follows: grade 0, no inflammation; grade 1, less than 2 foci; grade 2, 2–4 foci; and grade 3, more than 4 foci. Portal inflammation was scored as follows: grade 0, none; grade 1, mild; grade 2, moderate; or grade 3, marked. Fibrosis was scored as follows: stage 0, none; stage 1, perisinusoidal or periportal; stage 2, perisinusoidal and portal or periportal; stage 3, bridging fibrosis; or stage 4, cirrhosis. Hepatocellular iron was scored on an ordinal scale from 0 to 4. The NAFLD Activity Score (NAS), defined as the unweighted sum of the scores for steatosis (0-3), lobular inflammation (0-3), and ballooning (0-2), was represented in Table 2 as scores of ≤ 3 (low possibility of steatohepatitis), 3.5 – 4.5, and ≥ 5 (high possibility of steatohepatitis) (19). Similar to the process for steatosis, individual scores

for other histologic features were averaged for subjects with both types of biopsy, which occasionally resulted in half-scores.

Subject preparation and positioning

Subjects were instructed to fast for at least four hours prior to MR examination. At one center, subjects underwent non-contrast MR examinations on a 3.0T GE twin-speed with an 8-element torso phased-array coil scanner (GE Signa, EXCITE HDxt, GE Healthcare, Waukesha, WI). At the other center, subjects underwent non-contrast MR examinations on a 3.0T GE 750 scanner with a 32-element torso phased-array coil; if the subject did not fit, a 1.5T 450W GE wide bore scanner with an 8-element torso phased-array coil was used instead. Coils were centered over the liver. At once center, dielectric pads were placed over the abdomen for 3.0T examinations to reduce B1 heterogeneity. Each MR examination lasted about 60 minutes.

MR sequences

A 2D spoiled gradient-recalled echo (SPGR) sequence was used for MRI-M while a 3D SPGR sequence was used for MRI-C. Axial images covering as much of the liver as possible within one breath hold were acquired with the parameters listed in Table 1. These parameters are similar to those used in prior MRI-M and MRI-C studies (5-10, 16-17). For both methods, PDFF maps were generated in-line on the scanner computer from the corresponding source data. Source images, PDFF maps, and fat and water images (for MRI-C) were transferred for off-line analysis.

MR analysis

Using the Osirix imaging software (Pixmeo, Geneva, Switzerland) and blinded to histologic results, trained image analysts (1-3 years experience) placed regions of interest (ROIs) on the fifth echo of the source images (MRI-M) or water images (MRI-C). ROIs were placed on

anatomic images rather PDFF maps to avoid feedback bias. These particular images were selected beforehand because they consistently provide adequate visualization of hepatic anatomy to guide proper ROI placement. For each sequence, one ROI was placed in each of the nine Couinaud liver segments, making sure to include only liver parenchyma and to exclude vessels, bile ducts, lesions, and artifacts. In some subjects, one or more of the nine planned ROIs could not be placed on one or both sequences, usually because the segment was not included in the scanned volume and sometimes due to artifacts. ROIs were then propagated to PDFF maps and averaged to obtain the mean PDFF values.

Statistical analysis

Statistical analyses were performed by a staff biostatistician under the supervision of a faculty statistician, both with more than 20 years of experience, using R software package version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria). Cohort anthropometric, laboratory, and histologic measures were summarized. MRI-M and MRI-C PDFF averages were computed for each subject using only the segmental PDFFs available for both sequences. Thus, in case of a missing segmental PDFF for either sequence, unpaired segments were excluded, and partial averages based only on paired segments were computed for both sequences.

Receiver operating characteristic curve (ROC) analysis was performed for each MR sequence using histologically-determined presence of steatosis (grade ≥ 1) as the reference standard. Areas under the ROCs (AUC) and their significances were calculated to assess diagnostic accuracy.

The diagnostic threshold for each sequence was selected from the PDFF estimates based on the Youden index, which provided the highest specificity and sensitivity combination. Performance parameters (sensitivity, specificity, total accuracy, negative predictive value, and

positive predictive value) and their confidence intervals were calculated for each threshold. Bootstrap-based tests were used to compare their performance. Bonferroni's correction for multiple comparisons was applied. Only p-values < 0.0125 would be considered significant to ensure a family-wise significance level of 0.05.

To examine the possible confounding impact of demographic and other liver histologic features (listed in Table 2) on diagnostic accuracy, a Bayesian Information Criterion (BIC)-based stepwise logistic regression was used.

Results

Subjects

The total number of recruited subjects who satisfied the inclusion criteria was 81. A summary of cohort characteristics at the time of MRI is presented in Table 2. Of the 81, ten subjects lost weight between enrollment and MRI; hence, BMIs measured at the MRI visit ranged below the inclusion requirement of ≥ 35 kg/m^2. Two thirds of subjects had steatosis, mainly grade 1, while the remaining third had no steatosis. Forty-four had both core and wedge biopsies, twenty-eight had only wedge biopsies, and nine had only core biopsies. Among the 44 subjects with both biopsies, they were concordant in 29 subjects for the presence of steatosis (≥ 5% fat-containing hepatocytes); they were concordant in 14 subjects for the absence of steatosis (<5% fat-containing hepatocytes). In the one subject with discordant scores, the wedge biopsy had 5% steatotic hepatocytes (grade 1) and the core biopsy had 1% steatotic hepatocytes (grade 0). Since the average of the two percentages (3%) was <5%, the final assigned steatosis grade was 0 and the outcome was binarized as steatosis absent. Relatively small proportions of subjects had advanced histologic features on at least one biopsy specimen as defined by bridging fibrosis

or cirrhosis: 6/81 (7%); lobular inflammation grade ≥ 2: 7/81 (9%); portal inflammation grade ≥ 2: 3/81 (4%); iron grade ≥ 2: 11/81 (8%), or any ballooning 16/81 (20%).

PDFF Data

PDFF was estimated at 3.0T in 77 (95%) of the 81 subjects, and at 1.5T in the remaining four (5%). Over the entire cohort, PDFF values ranged from 1.0% to 32.5% for MRI-M and 1.4% to 31.5% for MRI-C. PDFF values were slightly but consistently lower for MRI-M than MRI-C, 9.3% ± 6.8 vs 9.9% ± 6.7. The mean difference of 0.6% was statistically significant (p-value < 0.0001). Figure 1 shows representative MRI-PDFF maps at 3.0T of subjects with and without hepatic steatosis.

ROC Analysis

Figure 2 shows ROC curves for the diagnosis of hepatic steatosis. MRI-M and MRI-C had AUCs of 0.951 ($p < 0.0001$) and 0.947 ($p < 0.0001$), respectively. This difference was not statistically significant (bootstrap-based p-value = 0.70).

Youden-index PDFF Threshold Diagnostic Performance and Comparison

Table 3 summarizes performance metrics of the Youden-index-based thresholds for both sequences. The Youden-index-based PDFF threshold for MRI-M was 6.5%, which provided 0.87 sensitivity (95% CI: 0.75 – 0.95), 0.96 specificity (0.81 – 1.00), and 0.90 total accuracy (0.82 – 0.96). Seven subjects had false negative results; one had a false positive result. The Youden-index-based PDFF threshold for MRI-C was 6.8%, which provided 0.90 sensitivity (95% CI: 0.77 – 0.96), 0.96 specificity (0.81 – 1.00), and 0.91 total accuracy (0.83 – 0.97). Six subjects had false negative results; one had a false positive result. For the one subject with discordant results between wedge and core biopsies, the PDFF estimated by each MRI method was 5.9%; therefore, each method was classified as true negative for this subject.

Differences between sequences were not significant for sensitivity, specificity, or total accuracy, with bootstrap-based p-values of 0.855, 0.460, and 0.840, respectively (Table 4).

A BIC-based, regression-model search for confounders to classification accuracy revealed BMI to be the sole significant predictor of classification errors. Higher BMI was associated with lower odds of misclassification error (odds ratio = 0.8, chi-square p-value = 0.037). No histological measures had a significant confounding effect on the relationship between PDFF and steatosis grade.

Discussion

This study showed that both MRI-M and MRI-C are highly accurate for the diagnosis of hepatic steatosis in subjects with severe obesity. With one exception, core and wedge biopsies were concordant in all subjects in which both biopsies were obtained. The Youden-index, diagnostic PDFF thresholds for MRI-M and MRI-C were very similar and their diagnostic performance parameters were not significantly different.

Our PDFF thresholds for diagnosis of hepatic steatosis are within the range of those reported previously for each MR sequence (12-14, 16-17). Tang et al (17) reported that an MRI-M-based PDFF threshold of 6.4% had 0.86 sensitivity and 0.83 specificity to diagnose steatosis while an earlier study (16) reported an MRI-M-based PDFF diagnostic threshold of 6.4% with 0.97 sensitivity and 1.00 specificity. Kühn et al (11) found that an MRI-C-determined PDFF threshold of 5.1% had 0.86 sensitivity and 1.00 specificity to differentiate between no versus any steatosis. Importantly, the Tang and Kuhn thresholds were derived from different cohorts than ours. Tang's cohort comprised of subjects with known or clinically-suspected NAFLD, with hepatic steatosis absent in only 7% of subjects. Kuhn's cohort comprised of subjects with clinical indications for liver biopsy. Conversely, our study consisted of adults with severe obesity

undergoing WLS without known hepatic steatosis a priori. This higher proportion and absolute number of subjects without hepatic steatosis permitted more precise specificity estimation: the CIs for the specificity estimates in this study (0.81 – 0.99 for both sequences) are narrower than those reported by Tang et al (0.36 – 1.00) and Kuhn et al (0.48 – 1.00) (16-17).

A separate study reported that a lower PDFF threshold (~3%) is associated with the presence of metabolic syndrome in obese adolescent girls (20). The lower threshold associated with metabolic syndrome raises the possibility that small quantities of liver fat, currently considered within the normal range by conventional histological grading systems, may in fact be pathologic. Further research is needed to confirm the lower PDFF threshold associated with metabolic syndrome and to understand the biological relevance.

Our results suggest that MRI-M- and MRI-C-derived PDFFs have comparable diagnostic accuracy. As reported previously, we observed a slight PDFF underestimation by MRI-M relative to MRI-C (18). MRI-M-derived PDFFs in this study were consistently and significantly lower than MRI-C PDFF with a mean difference of 0.64% (p-value < 0.0001). However, an underestimation of < 1% point is not expected to be clinically meaningful. Another difference is that the range of measurable PDFF is up to 50% for MRI-M but up to 100% for MRI-C. This difference is of minor relevance in the liver, since hepatic PDFF rarely exceeds 50% (15-17). In our cohort, the maximum PDFF was 32.5% as measured by MRI-M and 31.5% as measured by MRI-C. Therefore, either method is likely to be suitable for assessing hepatic steatosis in the large majority of subjects.

Using multivariable analysis, we examined the impact of several demographic, anthropometric, and histologic variables on classification errors. Our analysis found that only BMI was a significant predictor of misclassification (p = 0.0377) with lower BMI associated

with more frequent classification errors. This is an unexpected finding, since we would have anticipated that larger habitus, not smaller, might have reduced accuracy by lessening image quality and/or promoting imaging artifacts. Further research is needed to confirm and elucidate the mechanism for this finding. Unlike Schwimmer et al (21) and Idilman et al (15), we did not find that hepatic fibrosis or any other histology feature confounded the performance of MRI for assessing hepatic steatosis, but the failure to detect a confounding effect may have reflected the relatively mild degree of fibrosis and other histological abnormalities in our population.

One limitation of our study is that it was performed at two academic centers with expertise in PDFF quantification using both MRI methods. Additionally, a single scanner manufacturer was utilized at both sites. At one of the centers, subjects who were too large for the 3.0T scanner had to undergo MR examination using a larger 1.5T scanner. Although studies have suggested excellent reproducibility of PDFF estimation across different field strengths, scanners, and reconstruction methods (22-27), agreement is not perfect, and it is conceivable that the cutoffs identified in our study may require modification in different settings, such as community centers with less expertise or different equipment available.

Since we enrolled only adults, the applicability of our study's PDFF thresholds to the pediatric population at risk for NAFLD needs to be verified. The pathophysiology and histology of adult versus pediatric NAFLD are not the same and still not fully understood (28-29).

Because safety was paramount, we obtained biopsies based on the surgeon's judgment rather than obtaining the same type of biopsy in all subjects. Additionally, as done in prior studies (11, 15-17), PDFF values in this study were averaged from PDFF measurements obtained from each of the nine liver segments to better reflect the liver's total fat content and reduce sampling variability. Conversely, biopsies sampled only the left lateral sector, potentially

introducing sampling variability (30-31). We also could not co-localize ROIs with the exact biopsy location, since the latter could not be recorded reliably intraoperatively. Given that PDFF and histology assess for steatosis using widely disparate volumes, it should be expected that they do not correlate perfectly, and this may account in part for the misclassifications in our study. Furthermore, MRI estimates the tissue proton density fat fraction on a continuous scale while histology estimates the proportion of hepatocytes containing microscopically resolvable triglyceride droplets on an ordinal scale.

In summary, this study further supports the use of MRI-M- and MRI-C-derived PDFF as an accurate, quantitative, non-invasive alternative to histopathologic analysis for the diagnosis of hepatic steatosis in subjects with severe obesity at risk of developing NAFLD. Future studies are needed to confirm our results in scanners of other manufacturers, in different study populations including children, and in community centers.

References

1. Yokoo T, Bydder M, Hamilton G, et al. Nonalcoholic fatty liver disease: diagnostic and fat-grading accuracy of low-flip-angle multiecho gradient-recalled-echo MR imaging at 1.5 T. Radiology. 2009;251:67–76.

2. Yokoo T, Shiehmorteza M, Hamilton G, et al. Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. Radiology. 2011;258:749–759.

3. Reeder SB, McKenzie CA, Pineda AR, et al. Water-fat separation with IDEAL gradient-echo imaging. J Magn Reson Imaging. 2007;25: 644–652.

4. Reeder SB, Hu HH, Sirlin CB. Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. J Magn Reson Imaging. 2012;36(5):1011-4.

5. Liu CY, McKenzie CA, Yu H, et al. Fat quantification with IDEAL gradient echo imaging: correction of bias from T(1) and noise. Magn Reson Med. 2007;58:354–64.

6. Yu H, McKenzie CA, Shimakawa A, et al. Multiecho reconstruction for simultaneous water-fat decomposition and T2* estimation. J Magn Reson Imaging. 2007;26:1153–61.

7. Bydder M, Yokoo T, Hamilton G, et al. Relaxation effects in the quantification of fat using gradient echo imaging. Magn Reson Imaging. 2008 Apr;26(3):347-59. Epub 2008 Feb 21.

8. Yu H, Shimakawa A, McKenzie CA, et al. Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. Magn Reson Med. 2008;60:1122–34. doi: 10.1002/mrm.21737.

9. Hernando D, Hines CD, Yu H, Reeder SB. Addressing phase errors in fat-water imaging using a mixed magnitude/complex fitting method. Magn Reson Med. 2012 Mar;67(3):638-44. doi: 10.1002/mrm.23044.

10. Yu H, Shimakawa A, Hines CD, et al. Combination of complex-based and magnitude-based multiecho water-fat separation for accurate quantification of fat-fraction. Magn Reson Med. 2011;66:199–206.

11. Permutt Z, Le TA, Peterson MR, et al. 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. 2012 Jul;36(1):22-29.

12. Kühn JP, Hernando D, Muñoz del Rio A, et al. Effect of multipeak spectral modeling of fat for liver iron and fat quantification: correlation of biopsy with MR imaging results. Radiology. 2012;265(1):133–142.

13. Kang BK, Yu ES, Lee SS, et al. 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. 2012;47(6):368–75.

14. Wu CH1, Ho MC, Jeng YM, et al. Quantification of hepatic steatosis: a comparison of the accuracy among multiple magnetic resonance techniques. J Gastroenterol Hepatol. 2014; 29(4):807-13

15. Idilman IS, Aniktar H, Idilman R, et al. Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. Radiology. 2013;267(3):767–75.

16. Tang A, Tan J, Sun M, et al. Nonalcoholic fatty liver disease: MR imaging of liver proton density fat fraction to assess hepatic steatosis. Radiology. 2013 May;267(2):422-31.

17. 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 Feb;274(2):416-25.

18. Haufe WM, Wolfson T, Hooker CA, et al. Accuracy of PDFF Estimation by Magnitude-Based and Complex-Based MRI in Children with MR Spectroscopy as a Reference. J Magn Reson Imaging. 2017 Dec;46(6):1641-47.

19. Kleiner DE, Brunt EM, Natta MV, et al. Design and Validation of a Histological Scoring System for Nonalcoholic Fatty Liver Disease. Hepatology. 2005;41(6):1313–21.

20. Rehm JL, Wolfgram PM, Hernando D, Eickhoff JC, Allen DB, Reeder SB. Proton density fat-fraction is an accurate biomarker of hepatic steatosis in adolescent girls and young women. Eur Radiol. 2015 Oct;25(10):2921-30.

21. Schwimmer JB, Middleton MS, Behling C, et al. Magnetic resonance imaging and liver histology as biomarkers of hepatic steatosis in children with nonalcoholic fatty liver disease. Hepatology. 2015 Jun;61(6): 1887-95.

22. Artz NS, Haufe WM, Hooker CA, et al. Reproducibility of MR-based liver fat quantification across field strength: Same-day comparison between 1.5T and 3.0T in obese subjects. J Magn Reson Imaging. 2015;42(3):811-17.

23. Mashhood A, Railkar R, Yokoo T, et al. Reproducibility of hepatic fat fraction measurement by magnetic resonance imaging. J Magn Reson Imaging. 2013;37(6):1359-70.

24. Kang GH, Cruite I, Shiehmorteza M, et al. Reproducibility of MRI-determined proton density fat fraction across two different MR scanner platforms. J Magn Reson Imaging. 2011;34(4):928-34.

25. Serai SD, Dillman JR, Trout AT. Proton Density Fat Fraction Measurements at 1.5- and 3-T Hepatic MR Imaging: Same-Day Agreement among Readers and across Two Imager Manufacturers. Radiology. 2017:284(1):244-54.

26. Wu B, Han W, Li Z, et al. Reproducibility of Intra- and Inter-scanner Measurements of Liver Fat Using Complex Confounder-corrected Chemical Shift Encoded MRI at 3.0 Tesla. Sci Rep. 2016;6:19339.

27. [in press manuscript, bibliographic information withheld to maintain blinding]

28. Awai HI, Newton KP, Sirlin CB, et al. Evidence and Recommendations for Imaging Liver Fat in Children, Based on Systematic Review. Clin Gastroenterol Hepatol. 2014 May;12(5):765-773.

29. Roberts, EA. Pediatric Non-Alcoholic Fatty Liver Disease: A “growing” problem? Journal of Hepatology. 2007 June;46(6):1133-42.

30. El-Badry AM, Breitenstein S, Jochum W, et al. Assessment of hepatic steatosis by expert pathologists: the end of a gold standard. Ann Surg. 2009;250(5):691–7.

31. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonal-coholic fatty liver disease. Gastroenterology. 2005;128(7):1898–1906.

Table 1: Key acquisition parameters for magnitude-reconstruction chemical-shift-encoded magnetic resonance imaging (MRI-M) and complex-reconstruction chemical-shift-encoded magnetic resonance imaging (MRI-C) at 3.0T and 1.5T. TE = echo time; TR = repetition time.

	Field strength	**TE (ms)**	**TR (ms)**	**Flip angle (°)**	**Bandwidth (kHz)**	**Slice thickness (mm)**	**Matrix**	**Number of slices**	**Length of single breath hold (sec)**
MRI-M	3.0T	1.15, 2.3, 3.45, 4.6, 5.75, 6.9	120-170	10	±125	8	192-256 X 160-256	8-20	12-34
	1.5T	2.3, 4.6, 6.9, 9.2, 11.5, 13.8	170	10	±83	8	192-256 X 160-256	8-20	12-34
MRI-C	3.0T	1.2, 2.2, 3.2, 4.2, 5.2, 6.2	8.6	3	±125	8	256 X 128	32	20
	1.5T	1.2, 3.2, 5.2, 7.2, 9.2, 11.2	13.4	5	±125	8	256 x 160	32	20

Table 2: Cohort (n=81) demographic and liver histology characteristics.

Characteristic	Value
Sex	
Female	67 (82.7%)
Male	14 (17.3%)
Race	
Black	2 (2.5%)
Other	8 (9.9%)
White	71 (87.7%)
Ethnicity	
Hispanic/Latino	12 (14.8%)
Non-Hispanic/Latino	69 (85.2%)
Age (years)	48.2 ± 12.5 (23.7 – 70.6)*
Mean body mass index (kg/m^2)	41.7 ± 5.5 (32.5 – 56.8)*
Waist-to-Hip Ratio	0.89 ± 0.10 (0.74 – 1.19)*
Steatosis grade**	
0; <5% hepatocytes	27 (33.3%)
1; 5%-33% hepatocytes	39 (48.1%)
2; 33%-66% hepatocytes	11 (13.6%)
3; >66% hepatocytes	4 (4.9%)
Lobular inflammation***	
0; no foci	48 (59.3%)
0.5	7 (8.6%)
1; <2 foci per 200x field	19 (23.5%)
1.5	3 (3.7%)
2; 2-4 foci per 200x field	4 (4.9%)
3; >4 foci per 200x field	0 (0%)
Portal Inflammation***	
0; none	47 (58%)
0.5	13 (16%)
1; mild	18 (22.2%)
1.5	3 (3.7%)
2; moderate	0 (0%)
3; marked	0 (0%)
Hepatocellular ballooning***	
0; none	65 (80.2%)
0.5	1 (1.2%)
1; few balloon cells	10 (12.3%)
1.5	1 (1.2%)
2; many cells/prominent ballooning	4 (4.9%)
Iron grade***	
0	44 (54%)
0.5	8 (10%)
1	13 (16%)

1.5	4 (5%)
2	5 (6%)
3	2 (3%)
N/A†	5 (6%)
Fibrosis stage***	
0; no fibrosis	33 (41.2%)
0.5	7 (8.8%)
1; perisinusoidal or periportal	22 (27.5%)
1.5	10 (12.5%)
2; perisinusoidal and periporal	2 (2.5%)
2.5	1 (1.2%)
3; bridging fibrosis	4 (5.0%)
4; cirrhosis	1 (1.2%)
NA†	1 (1.2%)
NAS***‡	
≤ 3	68 (84%)
3.5 – 4.5	9 (11%)
≥ 5	4 (5%)
Mean MR-determined PDFF value (%)	
MRI-M	9.3 ± 6.8 (1.0 —32.5)*
MRI-C	9.9 ± 6.7 (1.4—31.5)*

Note— *Data are averages ± standard deviations, with ranges in parentheses.

**A 4-point ordinal steatosis score was derived from the granular steatosis score or, for subjects with both wedge and core biopsies, from the average granular steatosis score.

***Half grades/stages are due to averaging results from wedge and core biopsies.

† No data available for these subjects.

‡ NAS stands for the NAFLD Activity Score, which is the unweighted sum of the scores for steatosis (0-3), lobular inflammation (0-3), and ballooning (0-2). Scores of ≤ 3 (low possibility of steatohepatitis), 3.5 – 4.5 (moderate possibility of steatohepatitis), and ≥ 5 (high possibility of steatohepatitis).

Table 3: Summary of performance metrics of the magnitude-based magnetic resonance imaging (MRI-M-) and complex-based magnetic resonance imaging (MRI-C-) derived, Youden-index-based proton density fat fraction (PDFF) thresholds for diagnosing hepatic steatosis in subjects with severe obesity undergoing weight-loss surgery, using histology as reference.

Sequence	Youden-index-based PDFF threshold (%)	Sensitivity	Specificity	PPV	NPV	Total accuracy
MRI-M	6.5	0.87 (47/54) [0.75 – 0.95]	0.96 (26/27) [0.81 – 1.00]	0.98 (47/48) [0.89 – 1.00]	0.79 (26/33) [0.61 – 0.91]	0.90 (73/81) [0.82 – 0.96]
MRI-C	6.8	0.89 (48/54) [0.77 – 0.96]	0.96 (26/27) [0.81 – 1.00]	0.98 (48/49) [0.89 – 1.00]	0.81 (26/32) [0.64 – 0.93]	0.91 (74/81) [0.83 – 0.97]

Note—numerators and denominators of percentages are in parentheses. Numbers in brackets are 95% confidence intervals. PPV = positive predictive value; NPV = negative predictive value.

Table 4: Difference between performance metrics of magnitude-based magnetic resonance imaging (MRI-M-) and complex-based magnetic resonance imaging (MRI-C-) derived proton density fat fraction (PDFF) thresholds and their statistical significance (95% confidence interval and p-values).

Performance Parameter	Estimate (MRI-M – MRI-C)	95% CI lower limit	95% CI upper limit	Bootstrap-based p-value
AUC	0.004	-0.014	0.021	0.700
Sensitivity	-0.02	-0.06	0.10	0.855
Specificity	0.00	-0.12	0.00	0.460
Total accuracy	-0.01	-0.04	0.05	0.840

Note: Each estimate was calculated by subtracting MRI-C value from MRI-M value.

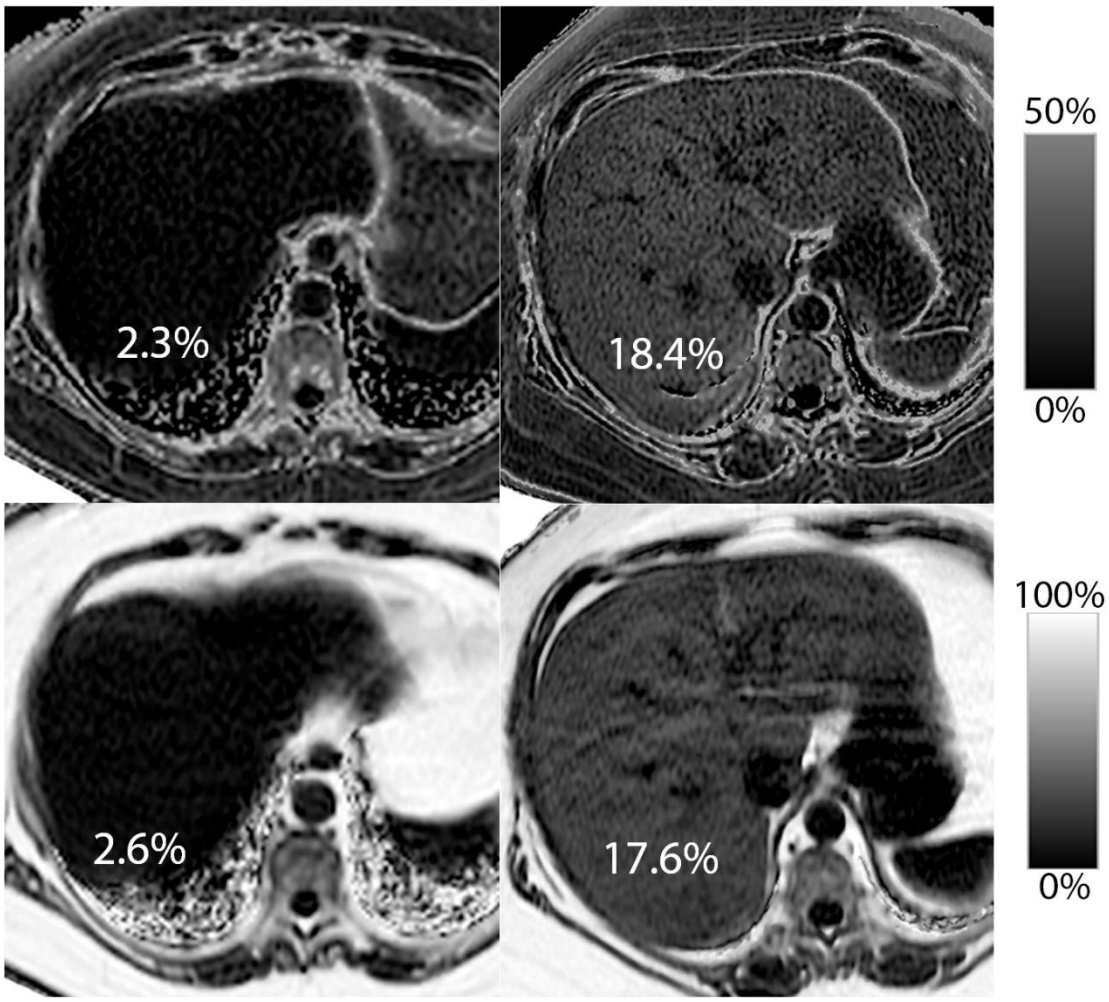

Figure 1: Magnetic resonance imaging (MRI) proton density fat fraction (PDFF) maps generated by magnitude-based magnetic resonance imaging (MRI-M, top row) and complex-based magnetic resonance imaging (MRI-C, bottom row) for subjects without (left column: 55-year old woman, body mass index 33.2 kg/m^2) versus with (right column: 58-year old woman, body mass index 44.7 kg/m^2) hepatic steatosis. Notice that MRI-M estimates PDFF from 0-50% whereas MRI-C estimates PDFF from 0 to 100%, as reflected in the scale bars. As a result, adipose tissue appears dark on MRI-M maps but bright on MRI-C maps. Despite these differences, close PDFF agreement is observed in the liver.

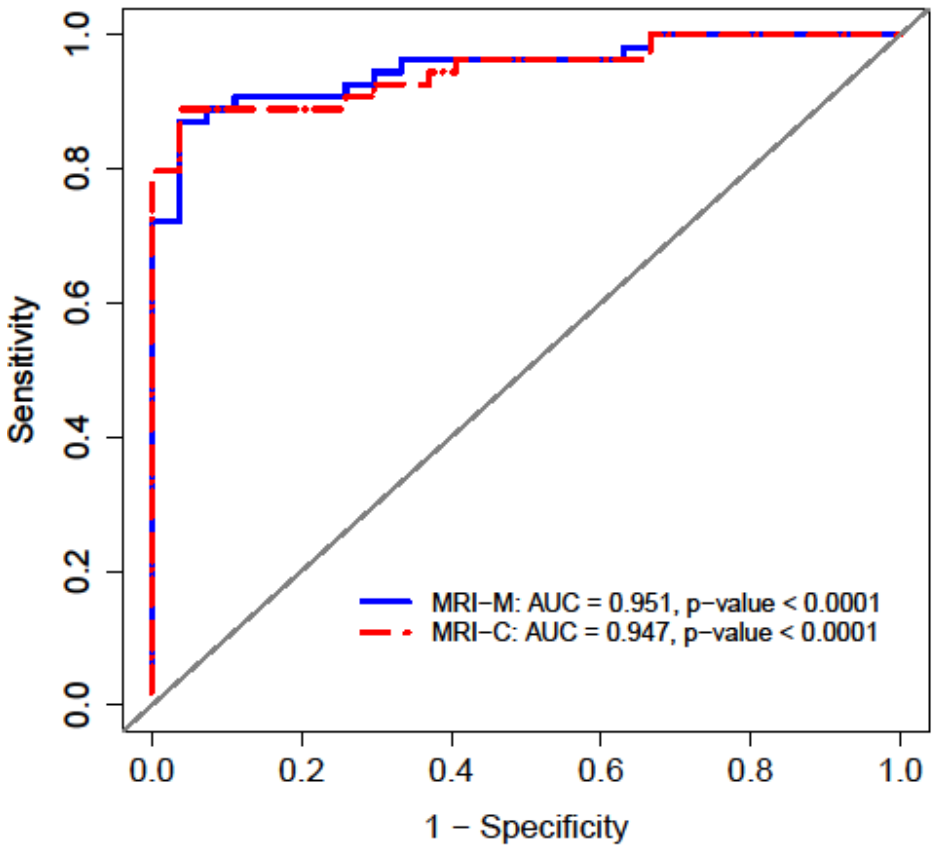

Figure 2: Overlaid magnitude-based magnetic resonance imaging (MRI-M) and complex-based magnetic resonance imaging (MRI-C) receiver operating characteristic (ROC) curves with their respective area under the ROC curve (AUC) and p-values for diagnosing hepatic steatosis.