




Impact of the Analytical Approach on the Reliability of MRI-Based Assessment of Hepatic Fat Content

Maggie S Burhans,^{1,2} Niranjana Balu,³ Kelsey A Schmidt,^{1,2} Gail Cromer,¹ Kristina M Utzschneider,^{4,5} Ellen A Schur,⁵ Sarah E Holte,¹ Timothy W Randolph,¹  and Mario Kratz^{1,2,5} 

¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; ²Department of Epidemiology, University of Washington, Seattle, WA, USA; ³Department of Radiology, University of Washington, Seattle, WA, USA; ⁴VA Puget Sound Health Care System, Seattle, Washington, WA, USA; and ⁵Department of Medicine, University of Washington, Seattle, WA, USA

ABSTRACT

MRI is a popular noninvasive method for the assessment of liver fat content. After MRI scan acquisition, there is currently no standardized image analysis procedure for the most accurate estimate of liver fat content. We determined intraindividual reliability of MRI-based liver fat measurement using 10 different MRI slice analysis methods in normal-weight, overweight, and obese individuals who underwent 2 same-day abdominal MRI scans. We also compared the agreement in liver fat content between analytical methods and assessed the variability in fat content across the entire liver. Our results indicate that liver fat content varies across the liver, with some slices averaging 54% lower and others 75% higher fat content than the mean of all slices (gold standard). Our data suggest that the entire liver should be contoured on at least every 10th slice to achieve close agreement with the gold standard. *Curr Dev Nutr* 2020;4:nzaa171.

Keywords: liver fat content, NAFLD, fatty liver, magnetic resonance imaging, proton density fat fraction

© The Author(s) 2020. Published by Oxford University Press on behalf of the American Society for Nutrition. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Manuscript received March 12, 2020. Initial review completed October 15, 2020. Revision accepted November 13, 2020. Published online December 18, 2020.

This work was supported by the University of Washington Nutrition and Obesity Research Center grant P30 DK035816. MSB and KAS were supported by the National Institutes of Health (NIH), MSB was supported by R25CA094880, T32DK007247, and T32HL007028. KAS was supported by R25CA094880 and T32CA094880. KMU is supported by the US Department of Veteran Affairs.

Author disclosures: The authors report no conflicts of interest.

Supplemental Tables 1–4 and Supplemental Figures 1–3 are available from the “Supplementary data” link in the online posting of the article and from the same link in the online table of contents at <https://academic.oup.com/cdn/>.

Address correspondence to MSB (e-mail: mburhans@fredhutch.org).

Abbreviations used: ICC, intraclass correlation coefficient; MRS, magnetic resonance spectroscopy; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; PDFF, proton density fat fraction; ROI, region of interest; 3D, 3-dimensional.

Introduction

The global prevalence of nonalcoholic fatty liver disease (NAFLD) is estimated to be 25% among the adult population (1, 2). NAFLD is a spectrum of diseases that includes nonalcoholic fatty liver, defined as $\geq 5\%$ liver fat content, and nonalcoholic steatohepatitis (NASH), which is characterized by excess fat accumulation accompanied by inflammation, cellular injury, and/or fibrosis (1). In the United States, NAFLD is recognized as an underlying cause for a significant proportion of chronic liver disease (1, 2). NASH may occur in up to 60% of individuals with NAFLD and significantly increases the risk for cirrhosis and hepatocellular carcinoma (1, 3).

Liver biopsy is considered the gold standard for measuring liver fat content, but there are risks associated with this invasive procedure (4, 5). There is also significant sampling variability among intraindividual liver specimens as a single biopsy is $\sim 1/50,000$ th of the total liver volume (6). Thus, noninvasive methods are regarded as essential tools to monitor liver fat content and are particularly useful in clinical trials given the limitations of biopsy. Imaging-based methods, including MRI and magnetic resonance spectroscopy (MRS), are available to estimate

liver fat content (7, 8). While MRS is widely accepted as a highly accurate quantitative method, it has several limitations that preclude it from widespread use, as it is more time consuming, more complex to conduct, and less widely accessible compared with MRI (8, 9). Modifications of the original Dixon MRI methodology that use newer water-fat image-acquisition protocols have improved separation of the water and fat signals for more accurate liver fat quantification (10, 11). Together, these characteristics of MRI have increased its use as a reliable, noninvasive method for the assessment of liver fat content (7, 9, 12, 13).

After MRI scan acquisition, there is currently no standardized, image-analysis procedure for the most accurate estimate of liver fat content, and the total contoured area analyzed for liver fat content determination differs among studies (14–18). The total area of liver (area per slice and number of slices) covered by the regions of interest (ROIs) might influence measurement variability, test-retest reliability, and the amount of time required for analysis. Furthermore, variability in fat content throughout the entire liver has not been systematically assessed. This is important because variability across an individual slice or the entire organ may influence the total liver area or ROI size required for a reliable estimate of fat content. We determined intraindividual

reliability of MRI-based liver fat measurement using 10 different MRI slice analysis methods in normal-weight, overweight, and obese individuals. We also compared the agreement in the determination of liver fat content between analytical methods. Finally, we assessed the variability in fat content across the entire liver. Taken together, these results may be used to inform the best analytical practice for the determination of human liver fat content using data from MRI scans.

Methods

Subjects

Study participants were recruited from the greater Seattle area. Subjects underwent telephone and in-person screening to determine eligibility to undergo MRI scanning. Eligibility criteria included BMI (kg/m^2) ≥ 20 and age between 18 and 75 y. Individuals were excluded if they had contraindications to undergo MRI scan (i.e., claustrophobia, imbedded or implanted metal objects, large abdominal tattoos) or excessive alcohol intake (> 14 drinks/wk). Written informed consent was obtained from all subjects. All scans were conducted between January 2016 and August 2017. This study was approved by the institutional review board at Fred Hutchinson Cancer Research Center.

MRI scan

Subjects were nonfasted when they underwent 2 same-day abdominal scans (referred to as “A scan” and “B scan”) on a Philips 3T Ingenia CX (Philips Healthcare) whole-body scanner at the Bio-Molecular Imaging Center at the University of Washington in Seattle, Washington. After the first scan, the subject was removed from the scanner and out of the scanner bed. After a short break of ~ 10 min, the repeat scan (B scan) was started by placing the subject on the scanner bed, re-landmarking and rescanning the same sequences as the A scan. Images were acquired using the posterior coil integrated into the scanner table and a 32-channel anterior coil placed on the abdomen. After a 3-plane survey scan, a 3-dimensional (3D) 6-echo Dixon (mDixon) sequence with bipolar gradients was obtained with the following scan parameters: TR/TE1/delta TE (milliseconds), 5.8/1.00/0.7; flip angle 3° ; matrix, 152×122 ; resolution (millimeters), $2.5 \times 2.5 \times 6.0$; parallel imaging acceleration factor 2; and bandwidth, 2492 Hz, with a 13-s breath-hold. Seventy-five slices were obtained in each scan, capturing the entire liver. Proton density fat fraction (PDFFF) maps were automatically calculated and stored in dicom format. As reference for segmentation, a higher-resolution 3D 2-point Dixon sequence ($1.7 \times 1.7 \times 3.5$ mm interpolated to $1 \times 1 \times 1.75$ mm) was also obtained as a visual aid to be able to avoid including smaller liver vessels during segmentation, using a 17-s breath-hold.

Image processing

Dicom images were transferred to a custom segmentation and analysis program (CASCADE; University of Washington, Seattle, WA). Segmentation of the whole liver was conducted to set the gold standard for hepatic fat content in this study. The initial 5 and final 5 liver-containing slices were omitted due to minimal measurable area. Water images from both mDixon and higher-resolution anatomical reference Dixon sequence were displayed side-by-side. Contours were drawn on the mDixon sequence and projected to the reference Dixon sequence. The anatomical reference Dixon sequence was used as a visual aid

to avoid including areas of vasculature within the contours. Contours covering the liver parenchyma were drawn to avoid vasculature and areas where partial volume averaging might occur, such as liver margins. More than 1 contour was used when discontinuous areas of the liver were present in an image. Examples of contours are shown in **Supplemental Figure 1**. Effort was taken to avoid large hepatic vessels in segmentation. Mean, SD, maximum, minimum, and range of PDFFF corresponding to each contour were determined. Any contour that resulted in a negative value was considered to reflect an undetectable level of fat and a value of zero was input to replace negative values.

The mean PDFFF of each slice was weighted by the slice contour area (millimeters squared) as a percentage of the total contoured liver area. For each scan, the weighted PDFFFs of all liver slices were then summed. We set this analytical method as the study gold standard for percentage liver fat. To compare the study gold standard and smaller areas of contoured liver, we also determined PDFFF based on every other, every third, every fifth, every 10th, every 20th, and middle slice. We also assessed PDFFF based on a circular 3-cm^2 ROI on the middle image, middle 3 images, and middle 5 images. One reader (MSB) contoured all scans and was blinded to both clinical characteristics and scan pairs.

Statistical analyses

Statistical analyses were performed using the Statistical Package for the Social Sciences for Macintosh (version 26.0; IBM) and GraphPad Prism (version 8.1; GraphPad Software). Distribution of variables was assessed by histograms and the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. All data were log transformed prior to intraclass correlation coefficient (ICC) and Pearson's correlation coefficient analyses. In the 3-cm^2 ROI analyses, some participants had a PDFFF value of zero. For these analytical approaches, a value of 1 was added to the PDFFF values for all subjects prior to log transformation. A 2-way mixed effects model for ICC was performed for each image analysis method to determine the reliability of repeated abdominal MRI scans for the measurement of liver fat. ICCs were computed with all subjects together and also divided into 2 subgroups based on presence ($\geq 5\%$ fat) or absence ($< 5\%$ fat) of fatty liver. ICC values < 0.4 were considered poor, values between 0.4 and 0.75 were considered fair to good, and values > 0.75 were considered excellent (19). Pearson's correlation coefficients were computed for each contour method to determine the association between PDFFFs as determined by repeated MRI scans. Bland-Altman analyses were performed to assess the level of agreement between the gold-standard contour method and each of the alternative contour methods described above (20).

Results

Subject characteristics

Twenty individuals were enrolled into this cross-sectional study and all participants underwent 2 same-day abdominal MRI scans. Overall, the mean \pm SD age was 49.9 ± 16.5 and 53.8 ± 16.3 y for women ($n = 11$) and men ($n = 9$), respectively. BMI was 30.3 ± 6.8 and 28.9 ± 4.3 for women and men, respectively. Subject characteristics

TABLE 1 Characteristics of study participants, grouped by absence (<5% fat) or presence (\geq 5% fat) of fatty liver

	Liver fat <5% (n = 13)		Liver fat \geq 5% (n = 7)	
	Mean \pm SD	Range	Mean \pm SD	Range
Age, y				
Women (n = 11)	45.7 \pm 18.7	23–70	57.3 \pm 9.6	44–67
Men (n = 9)	50.5 \pm 16.7	21–65	60.3 \pm 16.5	42–74
BMI, kg/m ²				
Women	27.3 \pm 2.9	23.4–30.9	35.7 \pm 8.7	26.8–43.2
Men	27.8 \pm 3.3	24.1–32.3	31.2 \pm 5.9	27.6–38.0

by presence or absence of fatty liver disease are shown in [Table 1](#); the group with fatty liver (\geq 5% liver fat, $n = 7$) tended to be older and have a higher BMI than the group without fatty liver (<5% liver fat, $n = 13$).

Variability in liver fat content across the entire organ

The whole liver was contoured on every slice to determine the study gold standard for the PDFF of each scan for all participants. The mean number of contoured liver slices across all scans was 51, with a range of 35 to 67 slices. Liver fat content ranged from 0.69% to 18.5% among the A scans and 0.58% to 18.4% among the B scans based on the study gold-standard PDFF method. Assessment of PDFF on a slice by slice basis revealed variability in fat content across the liver for nearly all study participants ([Figure 1](#)). For each scan of all subjects, we calculated the degree to which the slices with the lowest and highest PDFF values of each individual deviated from the mean PDFF across all slices (denoted by the dashed lines in [Figure 1](#)). We then calculated the overall mean minimum and maximum deviations across all scans. We found that, on average, the lowest PDFF values were only 46% of the mean PDFF while the highest PDFF values were 175% of the mean ([Supplemental Figure 2](#)). Together, these results demonstrate the substantial variability in fat content throughout the liver in both the presence and absence of fatty liver.

Reliability of liver fat content by repeated MRI scans

ICCs for the 10 analytical methods for PDFF determination revealed excellent (>0.75) reliability when all subjects were included ([Table 2](#)). To assess whether reliability varies by high or low liver fat content, we also performed ICC analyses for the 10 methods after categorizing the subjects based on the absence (<5% liver fat) or presence (\geq 5% liver fat) of hepatic steatosis as assessed by the study gold-standard method. ICCs for the PDFF in the group without fatty liver were fair to excellent, ranging from 0.50 to 0.79 ([Supplemental Table 1](#)). ICCs were consistently higher for the methods that contoured the whole liver as compared with methods that relied only on 3-cm² ROIs ([Supplemental Table 1](#)). ICCs for all contour methods were excellent for the fatty liver group, with a range of 0.86 to 0.95 ([Supplemental Table 2](#)).

Pearson's correlation coefficients revealed strong relations between PDFFs from the 2 scans ([Figure 2](#)). The correlation coefficients for the analytical methods that included contouring of whole-liver slices ([Figure 2A–F](#)) were slightly greater (all ≥ 0.95) than those that contoured only 1 slice or relied on 3-cm² ROIs, which ranged from 0.88 to 0.90 ([Figure 2G–J](#)). We also conducted Pearson's correlations separately for subjects grouped according to absence or presence of fatty

liver ([Supplemental Tables 3 and 4](#)). Among subjects without fatty liver, correlation coefficients for methods that contoured whole-liver slices ranged from 0.74 to 0.79 ([Supplemental Table 3](#)), although the method that relied only on the middle slice had a correlation coefficient of 0.49. For methods that used 3-cm² ROIs, the correlations were also weaker, with r values between 0.51 to 0.55 ([Supplemental Table 3](#)). The correlation coefficients for the repeat scans were strong for all analytical methods among the subjects with fatty liver, ranging from 0.92 to 0.98 ([Supplemental Table 4](#)).

Agreement between analytical methods

Bland-Altman analyses were performed to assess the agreement between the study gold standard and the alternative analytical methods. There were no obvious patterns in the absolute differences between the gold standard and any method, with similar absolute deviation across the range of PDFFs within any given method comparison in the A scans ([Figure 3](#)). However, there was a clear trend for lower agreement with the gold standard as the total area of contoured liver in the comparison method declined (i.e., a greater spread of the data points and wider CIs). The level of bias (mean difference; blue dashed line in [Figure 3](#)) between the gold standard and the alternative methods ranged from -0.005 to 0.22. The absolute difference in PDFF between the gold standard and every other, every third, every fifth, and every 10th slice was <1% for all subjects ([Figure 3A–D](#)). However, the absolute difference increased up to 2% when agreement between every 20th slice, the middle slice, or a 3-cm² ROI on the middle slice, middle 3 slices, or middle 5 slices with the gold standard was assessed ([Figure 3E–I](#)). Overall, results were similar for the set of B scans (data not shown), with the bias for each of the method comparisons comparable to that of the A scans. However, the maximum absolute difference between the gold standard and each of the methods using ROIs was 4% for 1 participant in the B scans.

We then constructed Bland-Altman plots based on percentage difference between methods [e.g., $100 \times (\text{study gold standard \% fat} - \text{every other slice \% fat}) / \text{mean \% fat}$] to assess relative differences in agreement across the range of PDFFs. The level of bias differed among the analysis methods and ranged from -0.07% for the method that included every third slice up to $\sim 30\%$ for the methods that included only 3-cm² ROIs ([Supplemental Figure 3](#)).

Two main observations emerged from this analysis. First, the relative agreement with the study gold standard declined as the total area contoured declined. While the agreement between the gold standard and every other, every third, every fifth, and every 10th slice was good, the agreement declined noticeably starting with every 20th slice and

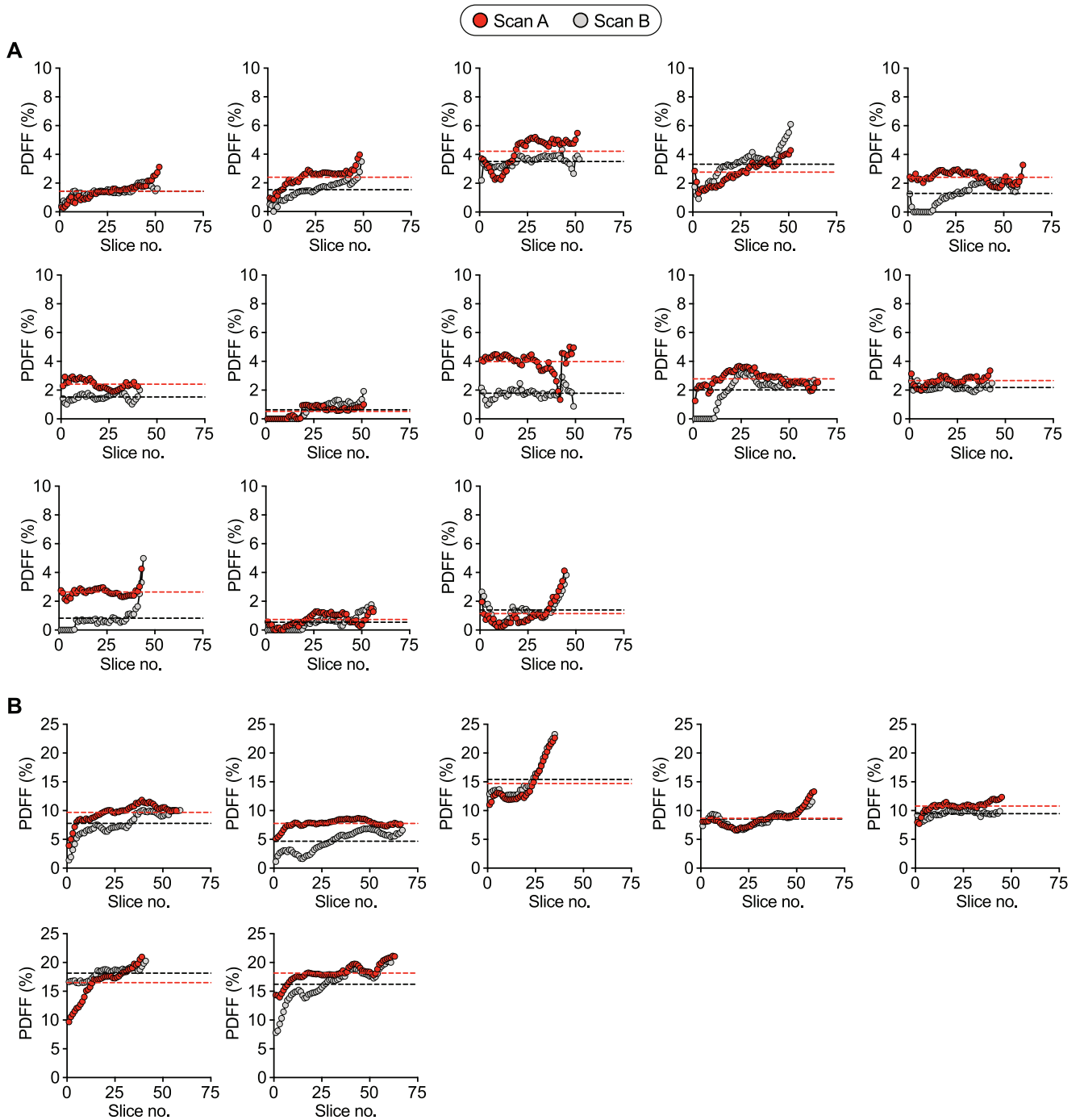


FIGURE 1 Individual variation in hepatic fat content. Abdominal MRI scans were conducted in nonobese and obese adults to assess liver fat content. Participants are grouped by absence (<5% liver fat) (A) or presence ($\geq 5\%$ liver fat) (B) of fatty liver disease. Note the different scales of the y-axes in A and B panels; this is to allow for better visualization of variability in fat across the liver among subjects with low liver fat content. The whole liver was contoured on every slice. Slice number refers to transverse liver sections going in the caudal to cranial direction. Dashed lines represent mean percentage liver fat for scan A (shown in red) and scan B (shown in black). PDFF, proton density fat fraction.

declined further when only 3-cm² ROIs were contoured (Supplemental Figure 3E-I). The bias (denoted by the blue dashed line in each panel) of $\sim 30\%$ for the 3 methods that relied on ROIs (Supplemental Figure 3G-I) reveals that these methods tend to underestimate fat

content relative to the study gold standard, although this problem occurred predominantly among the subjects with low liver fat. Indeed, the second main observation from these analyses revealed that the relative differences between methods among measurements in subjects without

TABLE 2 ICCs for PDFFs between 2 successive same-day MRI scans, all participants included¹

Contour method	ICC (95% CI)
All slices	0.95 (0.88–0.98)
Every other slice	0.95 (0.87–0.98)
Every third slice	0.95 (0.87–0.98)
Every fifth slice	0.94 (0.86–0.98)
Every 10th slice	0.95 (0.87–0.98)
Every 20th slice	0.95 (0.88–0.98)
Middle slice	0.88 (0.72–0.95)
Middle 5 slices ROI	0.90 (0.76–0.96)
Middle 3 slices ROI	0.90 (0.76–0.96)
Middle slice ROI	0.89 (0.74–0.95)

¹*n* = 20. ICC, intraclass correlation coefficient; PDFF, proton density fat fraction; ROI, region of interest.

fatty liver (i.e., when liver fat content was <5%) had the greatest impact on the bias. This was especially problematic among the methods that relied on the 3-cm² ROIs, with >100% deviation from the mean PDFF for several individual subjects (Supplemental Figure 3G–I). Results for the set of B scans were nearly identical (data not shown).

Discussion

We conducted this study to assess reliability of liver fat quantification by the Philips 3T Ingenia whole-body MRI scanner and to determine whether reliability is impacted by the analytical approach of the MRI images. Manual contouring of MRI slices for quantification of liver fat

is a time-intensive endeavor. This may be especially true for longitudinal research studies in which subjects undergo MRI scans at multiple time points to determine the effect of a given intervention on liver fat content. In addition, with several dietary components contributing to the development of NAFLD, the assessment of liver fat content is central to clinical nutrition research studies or nutritional therapy related to the prevention or management of fatty liver disease (21, 22). For example, the effect of various therapies that target the gut microbiome as well as a wide range of dietary supplements on liver fat content using MRI-PDFF measurements have been reported (23, 24). Currently, the field lacks a standardized method for the analysis of MRI slices for liver fat quantification. We questioned whether image analysis methods that do not require contouring the entire liver on every MRI slice would agree with fat content estimates by contouring the entire liver on every slice, which we defined as the study gold standard. To address this, we assessed the performance of 9 different post-scan image analysis methods relative to the gold standard. Our analyses revealed that the degree of agreement between the gold standard and the alternative methods declined as the total area of contoured liver declined. We found that contouring the whole liver on every other, every third, every fifth, or every 10th slice provided close agreement with PDFFs based on contouring every slice. However, the agreement with the gold standard declined when fewer than every 10th slice was contoured or when 5, 3, or 1 3-cm² ROIs were used.

MRI image analysis methods currently used vary by total liver area contoured, number of slices contoured, location of slices within the liver, and regional placement of ROIs on liver slices. Examples from the literature confirm such variability in methodology performed and

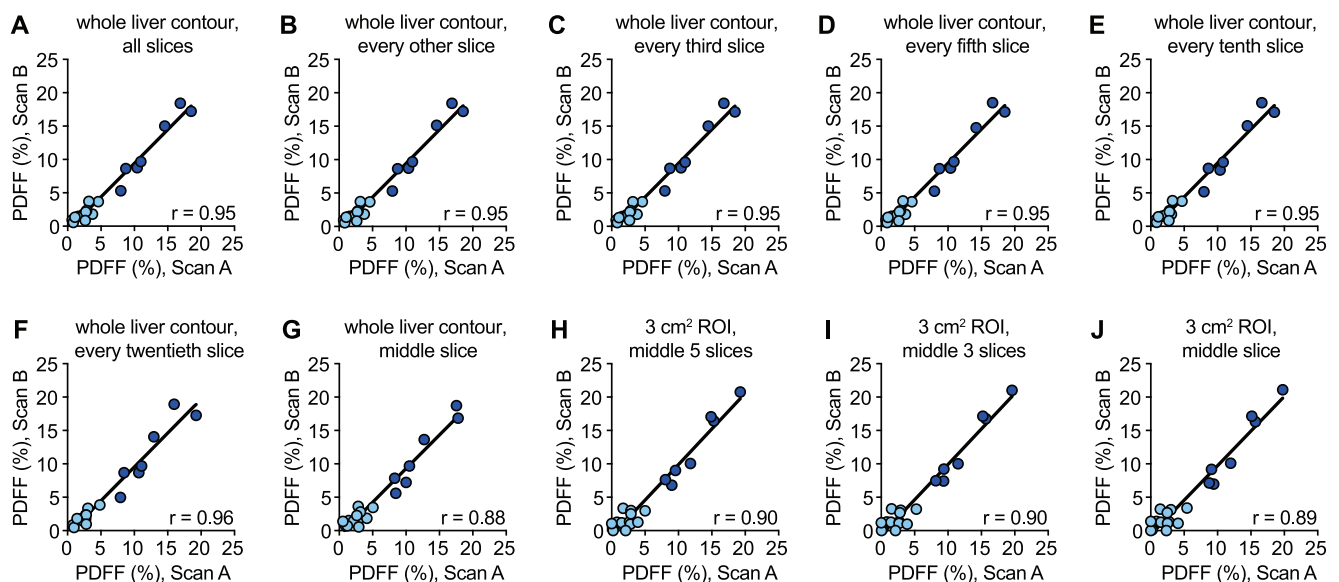


FIGURE 2 Associations between hepatic fat content estimated by same-day abdominal MRI scans (scans A and B) using 10 different liver slice analysis methods. Pearson's correlation coefficients between the 2 scans were calculated for analyses when the entire liver was contoured on all slices (A), every other slice (B), every third slice (C), every fifth slice (D), every 10th slice (E), every 20th slice (F), and middle slice only (G). Correlations were also conducted between the 2 scans when hepatic fat content was assessed using a 3-cm² ROI on the middle 5 slices (H), middle 3 slices (I), and middle slice only (J). Light blue indicates participants without fatty liver (<5% fat), dark blue indicates participants with fatty liver (≥5% fat). All data were log transformed prior to analyses. *P* < 0.0001 for all correlation coefficients. PDFF, proton density fat fraction; ROI, region of interest.

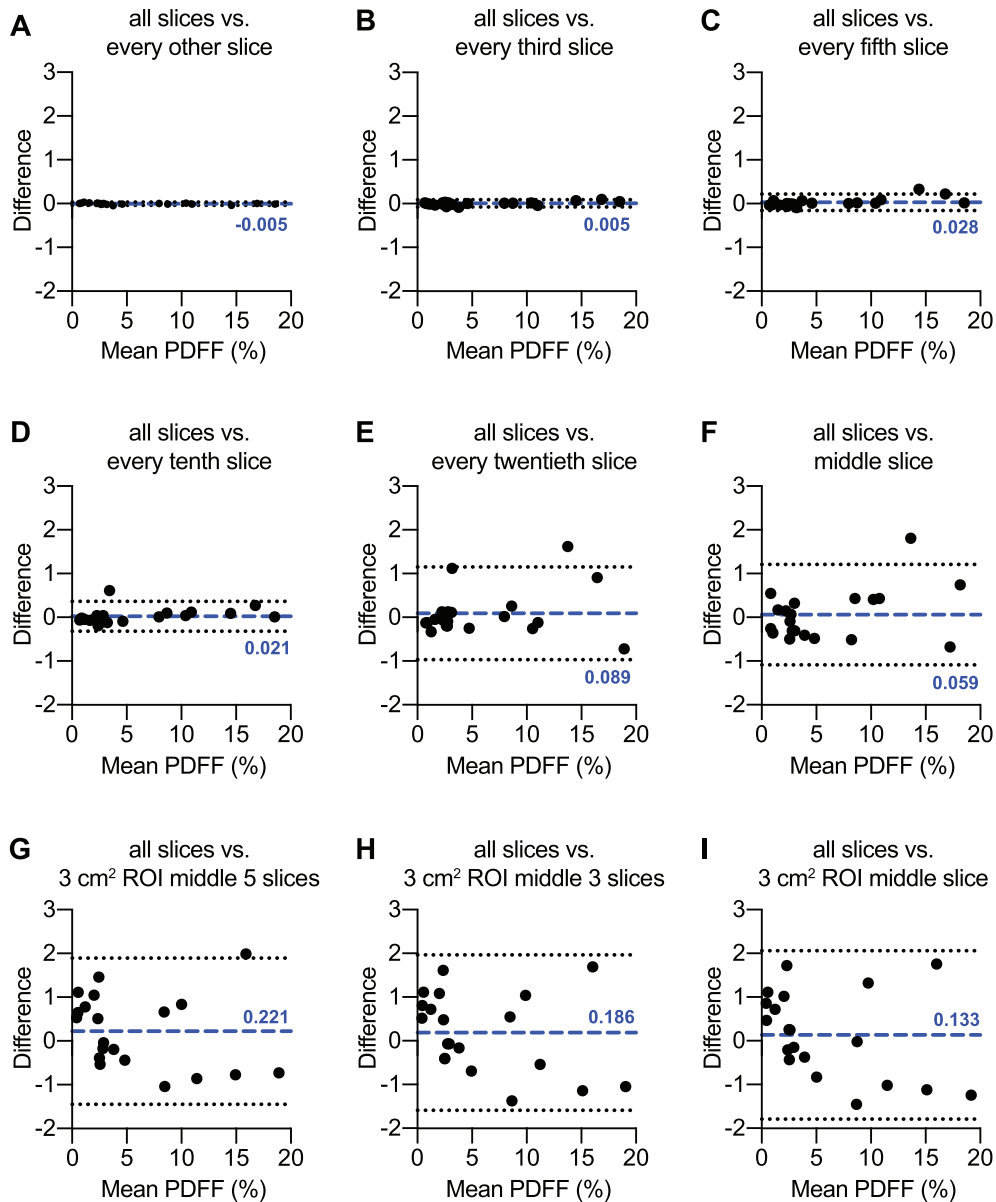


FIGURE 3 Bland-Altman plots of agreement between different analytical methods of abdominal MRI images to estimate hepatic fat content. The whole liver was contoured on every slice to set the study gold standard for liver fat content. Plots show the difference between the study gold standard (all slices) and whole-liver contours on every other slice (A), every third slice (B), every fifth slice (C), every 10th slice (D), every 20th slice (E), and middle slice (F). Plots of agreement are also shown for the study gold standard and a 3-cm² ROI on the middle 5 slices (G), the middle 3 slices (H), and the middle slice (I). The blue dashed line represents the bias (mean difference) between methods; the numeric value of the bias is shown in blue text. The black dotted lines represent the 95% CI of the limits of agreement. The x-axis mean PDFF % is the mean of the 2 methods compared in each plot. All results shown were generated from the A scans. PDFF, proton density fat fraction; ROI, region of interest.

include the following: single ROIs on each of the 9 liver segments on an image (25, 26), 5 ROIs on an image (14), 3 ROIs on each of the 9 liver segments (27), a single ROI on 3 slices (28), and an ROI on the right and left liver lobes (29). Few studies report which liver slice(s) are contoured, or whether the contoured liver slice was chosen using a standardized approach. There are a limited number of reports comparing MRI analysis methods for the best estimate of

liver fat content. Vu et al. (30) compared the reliability of 17 ROI sampling methods that ranged in total contoured area from 0.8 to 45 cm² with varying coverage of the liver lobes. Although most sampling techniques yielded excellent correlation against whole-liver contouring, the authors found that it was attenuated as the total area contoured declined. In addition, Vu et al. (30) reported that reliance predominantly on the right liver lobe for ROI contouring resulted in a significantly

greater liver fat content estimation as compared with whole-liver segmentation.

ICCs of liver fat quantification showed excellent reliability when all subjects, representing a wide range of liver fat content from ~1% to 18%, were included in the analysis. Our results are consistent with others that have also reported high reliability of MRI-based measurement of liver fat content (17, 26, 28, 31). Excellent reliability was reported in a study of obese children and adults, with a liver fat range of ~1% to 35%, who underwent 3 same-day MRI scans. In this study, 1 slice was analyzed per MRI scan in which liver fat was assessed in the 9 Couinaud liver segments and then averaged to determine right lobe, left lobe, and whole-liver fat content; ICCs for all regions assessed were >0.99 (31). Wu et al. (17) conducted a study of 15 individuals who underwent a total of 4 MRI scans, 2 scans each in 2 different scanners. Excellent reliability was reported for both intra- and interscanner measurements, with ICCs >0.83 across both hepatic lobes. Using a single ROI on 3 consecutive slices, Tyagi et al. (28) reported excellent inter-examination reliability, with an ICC of >0.99 in a cohort of overweight and obese subjects who underwent 3 same-day MRI scans with liver fat content ranging from 0.4% to 34%.

However, none of these studies assessed the reliability of liver fat measurements separately for individuals grouped by the presence or absence of fatty liver. Our results address this gap and demonstrate that, although reliability is excellent when assessed across a wide range of liver fat percentages, the low end of the range suffers from reduced test-retest reliability when a limited area of the liver is contoured. Despite this, MRI-based quantification of liver fat sufficiently distinguishes between absence and presence of fatty liver. Furthermore, given the very wide range of liver fat content that would likely be seen in most Western populations, particularly those with obesity or metabolic disease, the relative inaccuracies in individuals with low liver fat may be less concerning than they seem.

Heterogeneity in liver fat distribution is a well-known feature and may take various forms among individuals (32). The most common pattern for liver fat is diffuse distribution, but in some cases, it does not occur homogeneously (32, 33). For example, other distribution patterns may include localized regions of accumulation—for example, near the gallbladder or specific liver segments, localized regions of fat sparing, or several distinct fatty lesions spread throughout the organ (32, 33). It is possible that at least some of the variability in fat content observed by liver biopsy may be explained by these heterogeneous fat distribution patterns (6). In *ex vivo* livers, Bannas et al. (26) demonstrated lower agreement between repeated liver fat measurements using histological assessment of biopsies when compared with quantification by MRI across a wide range of liver fat content.

Although we did not investigate heterogeneity within slices to describe fat-distribution patterns, our slice by slice PDFF assessment revealed a heterogeneous distribution throughout the liver for nearly all subjects. Others have reported variability in hepatic fat content by MRI across the 9 Couinaud segments that make up the liver, with significant interindividual variability in the degree of liver fat heterogeneity (26, 30, 34). Hui et al. (35) conducted a study to assess heterogeneity of liver fat distribution in adults and children with or without NAFLD. The authors placed an ROI on the superior right lobe, superior left lobe, and the inferior right lobe to assess fat within different regions. Fat content differed significantly among the 3 regions in both NAFLD

groups but not in the groups without NAFLD, with the highest fat content in the superior right lobe and the lowest content in the superior left lobe. Regional variation in fat content has been hypothesized to be at least partially due to differential insulin exposure or contribution of blood sources (mesenteric vs. splenic) across the hepatic lobes (32, 34, 35).

Our study had several limitations. First, as we had only 1 reader contour the MRI slices, we do not have an assessment of interobserver reliability. Whether our findings of strong agreement between analytical methods and high variability in fat content throughout the liver would be consistent with other observers is not known. In addition, because all of our scans were conducted on a single-vendor MRI scanner and in a single research facility, our results may not be broadly generalizable to other scanners and techniques. Our study did not include any participants with known liver disease, which may limit the generalizability of the results to a broader patient population. Another limitation is that we did not assess the reliability of the panel of analytical approaches using currently available automatic or semi-automatic segmentation software. With increased availability of such automatic software programs, the reliance on manual segmentation for liver fat assessment may be declining.

The prevalence of fatty liver disease has reached epidemic proportions in adults and children (21, 36). Thus, the availability of reliable, noninvasive methods for liver fat quantification is critically important, not only to assess liver fat content but also to assess responses to interventions (37). However, it is also important to note that it remains unclear whether changes in liver fat are informative of potential changes in histological characteristics of NASH, including lobular inflammation, hepatocyte ballooning, and fibrosis. A recent longitudinal study of NASH patients demonstrated no association between improvements in liver fat content and histological changes (38), suggesting that the assessment of liver fat content alone may not be a robust biomarker for NASH histological features.

Overall, our results are consistent with the existing literature that demonstrates excellent reliability of liver fat quantification by MRI and the ability to distinguish between the absence or presence of fatty liver. It should be emphasized, however, that the greatest reliability occurred among individuals with hepatic fat content $\geq 5\%$. Together with the evidence discussed earlier, our study results demonstrate that total area and location of contour sampling affect the hepatic fat estimate as determined by MRI. These findings may be useful in clinical trials and the clinical assessment of liver fat content that rely on MRI-PDFF and lack a standardized post-scan processing method. The results from our assessment of different analytical methods indicate that, for the most reliable estimate of liver fat content, it is beneficial to contour larger areas, which may be especially important in longitudinal studies or studies that include a low-liver-fat group. Specifically, our data suggest that the entire liver should be contoured on at least every 10th slice to achieve close agreement with the gold standard.

Acknowledgments

The authors' responsibilities were as follows—MSB, NB, KAS, GC, KMU, EAS, SEH, TWR, and MK: designed the research and wrote the manuscript; MSB, NB, KAS, and GC: conducted the research; MSB and MK: analyzed data; MSB: had primary responsibility for final content; and all authors: read and approved the final manuscript.

References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64:73–84.
2. Younossi ZM, Stepanova M, Afendy M, Fang Y, Younossi Y, Mir H, Srishord M. Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008. *Clin Gastroenterol Hepatol* 2011;9:524–30.
3. Starley BQ, Calcagno CJ, Harrison SA. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. *Hepatology* 2010;51:1820–32.
4. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, Brunt EM, Sanyal AJ. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 2018;67:328–57.
5. Bravo AA, Sheth SG, Chopra S. Liver biopsy. *N Engl J Med* 2001;344:495–500.
6. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, Grimaldi A, Capron F, Poynard T; for the LIDO Study Group. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005;128:1898–906.
7. Reeder SB, Sirlin CB. Quantification of liver fat with magnetic resonance imaging. *Magn Reson Imaging Clin N Am* 2010;18:337–7, ix.
8. Springer F, Machann J, Claussen CD, Schick F, Schweser NF. Liver fat content determined by magnetic resonance imaging and spectroscopy. *World J Gastroenterol* 2010;16:1560–6.
9. Cowin GJ, Jonsson JR, Bauer JD, Ash S, Ali A, Osland EJ, Purdie DM, Clouston AD, Powell EE, Galloway GJ. Magnetic resonance imaging and spectroscopy for monitoring liver steatosis. *J Magn Reson Imaging* 2008;28:937–45.
10. Dixon WT. Simple proton spectroscopic imaging. *Radiology* 1984;153:189–94.
11. Ma J. Dixon techniques for water and fat imaging. *J Magn Reson Imaging* 2008;28:543–58.
12. Hu HH, Kim HW, Nayak KS, Goran MI. Comparison of fat-water MRI and single-voxel MRS in the assessment of hepatic and pancreatic fat fractions in humans. *Obesity (Silver Spring)* 2010;18:841–7.
13. Imajo K, Kessoku T, Honda Y, Tomeno W, Ogawa Y, Mawatari H, Fujita K, Yoneda M, Taguri M, Hyogo H, et al. Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. *Gastroenterology* 2016;150:626–37, e627.
14. Burgert TS, Taksali SE, Dziura J, Goodman TR, Yeckel CW, Papademetris X, Constable RT, Weiss R, Tamborlane WV, Savoye M, et al. Alanine aminotransferase levels and fatty liver in childhood obesity: associations with insulin resistance, adiponectin, and visceral fat. *J Clin Endocrinol Metab* 2006;91:4287–94.
15. House MJ, Gan EK, Adams LA, Ayonrinde OT, Bangma SJ, Bhathal PS, Olynyk JK, St Pierre TG. Diagnostic performance of a rapid magnetic resonance imaging method of measuring hepatic steatosis. *PLoS One* 2013;8:e59287.
16. Thomas EL, Hamilton G, Patel N, O'Dwyer R, Dore CJ, Goldin RD, Bell JD, Taylor-Robinson SD. Hepatic triglyceride content and its relation to body adiposity: a magnetic resonance imaging and proton magnetic resonance spectroscopy study. *Gut* 2005;54:122–7.
17. Wu B, Han W, Li Z, Zhao Y, Ge M, Guo X, Wu X. 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.
18. Loomba R, Wolfson T, Ang B, Hooker J, Behling C, Peterson M, Valasek M, Lin G, Brenner D, Gamst A, et al. Magnetic resonance elastography predicts advanced fibrosis in patients with nonalcoholic fatty liver disease: a prospective study. *Hepatology* 2014;60:1920–8.
19. Rosner B. *Fundamentals of biostatistics*. Boston (MA): Cengage Learning; 2016.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
21. Brunt EM, Wong VW, Nobili V, Day CP, Sookoian S, Maher JJ, Bugianesi E, Sirlin CB, Neuschwander-Tetri BA, Rinella ME. Nonalcoholic fatty liver disease. *Nat Rev Dis Primers* 2015;1:15080.
22. Zivkovic AM, German JB, Sanyal AJ. Comparative review of diets for the metabolic syndrome: implications for nonalcoholic fatty liver disease. *Am J Clin Nutr* 2007;86:285–300.
23. Sharpton SR, Maraj B, Harding-Theobald E, Vittinghoff E, Terrault NA. Gut microbiome-targeted therapies in nonalcoholic fatty liver disease: a systematic review, meta-analysis, and meta-regression. *Am J Clin Nutr* 2019;110:139–49.
24. Kilchoer B, Vils A, Minder B, Muka T, Glisic M, Bally L. Efficacy of dietary supplements to reduce liver fat. *Nutrients* 2020;12:2302.
25. Artz NS, Haufe WM, Hooker CA, Hamilton G, Wolfson T, Campos GM, Gamst AC, Schwimmer JB, Sirlin CB, Reeder SB. Reproducibility of MR-based liver fat quantification across field strength: same-day comparison between 1.5T and 3T in obese subjects. *J Magn Reson Imaging* 2015;42:811–7.
26. Bannas P, Kramer H, Hernando D, Agni R, Cunningham AM, Mandal R, Motosugi U, Sharma SD, Munoz del Rio A, Fernandez L, et al. Quantitative magnetic resonance imaging of hepatic steatosis: validation in ex vivo human livers. *Hepatology* 2015;62:1444–55.
27. Mashhood A, Railkar R, Yokoo T, Levin Y, Clark L, Fox-Bosetti S, Middleton MS, Riek J, Kauh E, Dardzinski BJ, et al. Reproducibility of hepatic fat fraction measurement by magnetic resonance imaging. *J Magn Reson Imaging* 2013;37:1359–70.
28. Tyagi A, Yeganeh O, Levin Y, Hooker JC, Hamilton GC, Wolfson T, Gamst A, Zand AK, Heba E, Loomba R, et al. 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* 2015;40:3070–7.
29. Saadeh S, Younossi ZM, Remer EM, Gramlich T, Ong JP, Hurley M, Mullen KD, Cooper JN, Sheridan MJ. The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology* 2002;123:745–50.
30. 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;43:1090–9.
31. Negrete LM, Middleton MS, Clark L, Wolfson T, Gamst AC, Lam J, Changchien C, Deyoung-Dominguez IM, Hamilton G, Loomba R, et al. Inter-examination precision of magnitude-based MRI for estimation of segmental hepatic proton density fat fraction in obese subjects. *J Magn Reson Imaging* 2014;39:1265–71.
32. Vilgrain V, Ronot M, Abdel-Rehim M, Zappa M, d'Assignies G, Bruno O, Vullierme MP. Hepatic steatosis: a major trap in liver imaging. *Diagn Interv Imaging* 2013;94:713–27.
33. Hamer OW, Aguirre DA, Casola G, Lavine JE, Woenckhaus M, Sirlin CB. Fatty liver: imaging patterns and pitfalls. *Radiographics* 2006;26:1637–53.
34. Bonekamp S, Tang A, Mashhood A, Wolfson T, Changchien C, Middleton MS, Clark L, Gamst A, Loomba R, Sirlin CB. Spatial distribution of MRI-determined hepatic proton density fat fraction in adults with nonalcoholic fatty liver disease. *J Magn Reson Imaging* 2014;39:1525–32.
35. Hui SCN, So HK, Chan DFY, Wong SKH, Yeung DKW, Ng EKW, Chu WCW. Validation of water-fat MRI and proton MRS in assessment of hepatic fat and the heterogeneous distribution of hepatic fat and iron in subjects with non-alcoholic fatty liver disease. *Eur J Radiol* 2018;107:7–13.
36. Nobili V, Alisi A, Newton KP, Schwimmer JB. Comparison of the phenotype and approach to pediatric vs adult patients with nonalcoholic fatty liver disease. *Gastroenterology* 2016;150:1798–810.
37. Wong VW, Adams LA, de Ledinghen V, Wong GL, Sookoian S. Noninvasive biomarkers in NAFLD and NASH—current progress and future promise. *Nat Rev Gastroenterol Hepatol* 2018;15:461–78.
38. Bril F, Barb D, Lomonaco R, Lai J, Cusi K. Change in hepatic fat content measured by MRI does not predict treatment-induced histological improvement of steatohepatitis. *J Hepatol* 2020;72:401–10.