Abdominal Radiology

Published online: 28 June 2018

Abdom Radiol (2019) 44:54–64
https://doi.org/10.1007/s00261-018-1683-0

Reader agreement and accuracy of ultrasound features for hepatic steatosis

Cheng William Hong,[1] Austin Marsh,[2] Tanya Wolfson,[3] Jeremy Paige,[4] Soudabeh Fazeli Dekhordy,[1] Alexandra N. Schlein,[1] Elise Housman,[1] Lisa H. Deiranieh,[1] Charles Q. Li,[1] Ashish P. Wasnik,[5] Hyun-Jung Jang,[6] Christoph F. Dietrich,[7] Fabio Piscaglia,[8] Giovanna Casola,[1] Mary O'Boyle,[1] Katherine M. Richman,[1] Mark A. Valasek,[9] Michael Andre,[1] Rohit Loomba,[10] and Claude B. Sirlin[1]

[1]Department of Radiology, University of California San Diego, 200 W. Arbor Drive #8756, San Diego, CA 92103-8756, USA
[2]School of Medicine, University of California San Diego, La Jolla, CA, USA
[3]Computational and Applied Statistics Laboratory, University of California San Diego, La Jolla, CA, USA
[4]Department of Radiology, University of Washington, Seattle, WA, USA
[5]Department of Radiology, University of Michigan, Ann Arbor, MI, USA
[6]Department of Radiology, University of Toronto, Toronto, ON, Canada
[7]Department of Internal Medicine 2, Caritas-Krankenhaus Bad Mergentheim, Bad Mergentheim, Germany
[8]Department of Internal Medicine, Università di Bologna, Bologna, Italy
[9]Department of Pathology, University of California San Diego, La Jolla, CA, USA
[10]NAFLD Research Center, Division of Gastroenterology, Department of Medicine, University of California San Diego, La Jolla, CA, USA

Abstract

Purpose: The purpose of the study is to assess the reader agreement and accuracy of eight ultrasound imaging features for classifying hepatic steatosis in adults with known or suspected hepatic steatosis.
Methods: This was an IRB-approved, HIPAA-compliant prospective study of adult patients with known or suspected hepatic steatosis. All patients signed written informed consent. Ultrasound images (Siemens S3000, 6C1HD, and 4C1 transducers) were acquired by experienced sonographers following a standard protocol. Eight readers independently graded eight features and their overall impression of hepatic steatosis on ordinal scales using an electronic case report form. Duplicated images from the 6C1HD transducer were read twice to assess intra-reader agreement. Intra-reader, inter-transducer, and inter-reader agreement were assessed using intraclass correlation coefficients (ICC). Features with the highest intra-reader agreement were selected as predictors for dichotomized histological steatosis using Classification and Regression Tree (CART) analysis, and the accuracy of the decision rule was compared to the accuracy of the radiologists' overall impression.
Results: 45 patients (18 males, 27 females; mean age 56 ± 12 years) scanned from September 2015 to July 2016 were included. Mean intra-reader ICCs ranged from 0.430 to 0.777, inter-transducer ICCs ranged from 0.228 to 0.640, and inter-reader ICCs ranged from 0.014 to 0.561. The CART decision rule selected only large hepatic vein blurring and achieved similar accuracy to the overall impression (74% to 75% and 68% to 72%, respectively).
Conclusions: Large hepatic vein blurring, liver–kidney contrast, and overall impression provided the highest reader agreement. Large hepatic vein blurring may provide the highest classification accuracy for dichotomized grading of hepatic steatosis.

Key words: Hepatic steatosis—Liver—Ultrasound features—Reader agreement—Hepatic vein blurring

Abbreviations

NAFLD	Nonalcoholic Fatty Liver Disease
NASH CRN	Nonalcoholic Steatohepatitis Clinical Research Network

Correspondence to: Claude B. Sirlin; *email:* csirlin@ucsd.edu

ICC	Intraclass correlation coefficients
CART	Classification and Regression Tree
MRI	Magnetic resonance imaging

Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease worldwide, with various studies estimating a global prevalence between 20% and 45% [1–5]. Liver ultrasound is inexpensive, noninvasive, and commonly used to make the qualitative diagnosis of hepatic steatosis [6]. Ultrasound can be limited by operator-dependence, interpretation subjectivity, and patient body habitus, factors that contribute to inter-observer variability [7] and may reduce sensitivity and accuracy [8, 9], especially at lower steatosis grades [10, 11].

On ultrasound, hepatic steatosis is assessed qualitatively using a combination of imaging features such as liver–kidney contrast, vessel blurring, posterior beam attenuation, focal fat sparing, and gallbladder and diaphragm visualization [12–15]. Although these features have been applied in clinical practice, the reader agreement of these individual features has not been assessed together in a single study, and there is little guidance on how radiologists should combine these features into a composite assessment of hepatic steatosis.

The purpose of this study is to assess the intra- and inter-reader agreement of these individual ultrasound features in a single study, and to provide guidance on applying the features with the highest agreement towards a composite assessment. No training was provided as it would inflate performance without capturing real-world performance of radiologists applying these features in current practice. As no prior study published an atlas for scoring, a secondary purpose was to create an atlas of representative B-mode images illustrating the severity spectrum for each ultrasound feature.

Subjects and methods

Study design

This was a prospectively designed cohort study of adult patients with known or suspected hepatic steatosis recruited consecutively from the institutional fatty liver clinic by the study hepatologist. The study was approved by an Institutional Review Board and was compliant with the Health Insurance Portability and Accountability Act. Informed consent was obtained from all individual participants included in the study.

Patients were not eligible for the study if they had moderate or higher alcohol consumption or chronic liver disease other than NAFLD, as determined by the study hepatologist. Demographic and anthropometric data were recorded and summarized descriptively. The majority of these patients had undergone liver biopsy for clinical care, and this subset of patients were included in the accuracy and decision tree component of the study.

Liver biopsy and histology

Nontargeted percutaneous biopsies of the right hepatic lobe were performed if needed for clinical care using 16G or 18G needles by hepatologists. For this research, clinically obtained histology slides were reviewed by an expert hepatopathologist blinded to clinical and radiologic data, who scored steatosis using a 4-point ordinal scale defined by the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) Histologic Scoring System [16, 17]. Steatosis scoring was based on the proportion of hepatocytes with macrovesicular steatosis: grade 0 ($< 5\%$), 1 (5% to 33%), 2 (33% to 66%), and 3 ($> 66\%$). Microvesicular steatosis was not analyzed for this study as it is a less characteristic feature of NAFLD and not used in the NASH CRN system to assess steatosis grade.

Ultrasound protocol

The ultrasound protocol was designed in consensus through an 8-week iterative process by a fellowship-trained abdominal faculty radiologist, an ultrasound physicist, and two experienced registered diagnostic medical sonographers. The protocol was electronically added to the scanner with a pre-set sequence designed to efficiently and consistently capture the ultrasound images and views needed for radiology scoring (discussed below). Conventional ultrasound images (Siemens S3000, 6C1HD and 4C1 transducers, Siemens Medical Solutions USA Inc., Malvern, PA) were acquired in each patient by one of the two medical sonographers. Sonographers were free to adjust scanner parameters including applying tissue harmonic imaging to capture the best quality images for each view, patient, and transducer. B-mode images with the following views were acquired:

- transverse plane: hepatic veins at the confluence with the inferior vena cava, main portal vein, right portal vein, right anterior portal vein branch, right posterior portal vein
- sagittal plane: middle/right hepatic vein, liver/kidney, liver depth including diaphragm

Additionally, if focal fat sparing was judged to be present by the sonographer, one or more images were acquired to show area(s) of focal fat sparing.

Exams from both transducers were acquired consecutively and included in the image bank as follows: the exam acquired with the 6C1HD transducer was included twice as duplicates for the assessment of intra-reader agreement as it is the transducer used more frequently at our institution during clinical practice by the sonographers, and the exam acquired with the 4C1 transducer was included only once. Images were stored in DICOM format with patient, operator, and institutional identifiers removed from both the file header and image overlay.

Assessment of hepatic steatosis features

We recruited eight readers for this study: seven readers were academic faculty with liver ultrasound expertise and one was a fourth-year radiology resident. These readers were from five different universities.

The eight readers independently graded the following eight individual features of hepatic steatosis on ordinal scales by applying the definitions given in Table 1 using an electronic case report form: large hepatic vein blurring (3-point scale), main right portal vein blurring (3-point), anterior and posterior right portal vein blurring (3-point), liver–kidney contrast (4-point), posterior beam attenuation (3-point), diaphragm definition (3-point), focal fat sparing (2-point), and liver echotexture (2-point). With regard to the 2-point echotexture score, the presence of coarse echoes in the parenchyma was considered to be abnormal and a sonographic feature of hepatic steatosis, in comparison to normal echoes which was considered a feature of healthy liver [7, 10]. These features and scores, as well as their definitions, were selected based on previous studies of imaging features for hepatic steatosis [7–13] and are summarized in Table 1. Readers displayed the images using a DICOM viewer with a complement of common workstation tools.

We deliberately did not create a training set for the readers because no prior study had published an atlas of ultrasound scoring to inform the training sessions. In the absence of a pre-existing evidence-based atlas, training sessions might introduce inadvertent biases into image interpretation unsubstantiated by evidence and may not be representative of how radiologists would apply published ultrasound criteria in their own practices. To avoid such biases and to more closely emulate how the features

Table 1. Ordinal scales and criteria for scoring each imaging feature

Feature	Grade	Criteria
Large hepatic vein blurring	0: No blurring	Large hepatic veins are clearly visualized and walls are sharply defined
	1: Mild–moderate blurring	Large hepatic veins are clearly visualized but walls are not sharply defined
	2: Severe blurring	Large hepatic veins are not clearly visualized
Main right portal vein blurring	0: No blurring	Main and right portal veins are clearly visualized and walls are sharply defined
	1: Mild–moderate blurring	Main and right portal veins are clearly visualized but walls are not sharply defined
	2: Severe blurring	Main and right portal veins are not clearly visualized
Anterior posterior right portal vein blurring	0: No blurring	Anterior and posterior portal veins are clearly visualized and walls are sharply defined
	1: Mild–moderate blurring	Anterior and posterior portal veins are clearly visualized but walls are not sharply defined
	2: Severe blurring	Anterior and posterior portal veins are not clearly visualized
Liver–kidney contrast	0: Isoechoic	Right liver lobe parenchyma is isoechoic compared to right kidney cortex
	1: Mildly hyperechoic	Right liver lobe parenchyma is mildly hyperechoic compared to right kidney cortex (requires careful inspection, not dramatically brighter)
	2: Moderately hyperechoic	Right liver lobe parenchyma is moderately hyperechoic compared to right kidney cortex (immediately apparent, but not dramatically brighter)
	3: Markedly hyperechoic	Right liver lobe parenchyma is markedly hyperechoic compared to right kidney cortex (immediately apparent and dramatically brighter)
	Not applicable	Right kidney not visualized or visible but clearly abnormal
Posterior beam attenuation	0: No attenuation	No definite posterior beam attenuation
	1: Mild-moderate attenuation	Definite posterior beam attenuation but not dramatic
	2: Markedly hyperechoic	Definite and dramatic posterior beam attenuation
Diaphragm definition	0: Clearly defined	Diaphragm visualized in its entirety as a sharp line
	1: Obscured	Diaphragm visualized as an interrupted or blurry line
	2: Obliterated	Diaphragm not visualized at all
Focal fat sparing	0: Absent	Focal fat sparing is absent
	1: Present	Focal geographic hypoechoic area(s) observed adjacent to gallbladder wall or portal vessel wall
Liver echotexture	0: Normal	Normal
	1: Abnormal	Coarse echoes
Overall impression	0: No hepatic steatosis	Based on subjective interpretation
	1: Mild hepatic steatosis	Based on subjective interpretation
	2: Moderate hepatic steatosis	Based on subjective interpretation
	3: Severe hepatic steatosis	Based on subjective interpretation

would be applied in clinical practice, we allowed the readers to score the features based on their interpretation of the verbal definitions. As described below, this approach allowed us to construct an unbiased atlas of representative ultrasound images for future use, based on the aggregated research readings.

Readers also provided an overall impression of steatosis severity for each exam on a 4-point scale, corresponding to the four histologic grades (none, mild, moderate, severe). Definitions were not provided for the overall impression scores because overall impression has not previously been studied as a predictor of steatosis severity and therefore there was no scientific evidence to inform definitions a priori. Instead, readers were allowed to apply their unconstrained subjective assessment.

Creation of Atlas

We created an atlas to illustrate the scoring spectrum using the 6C1HD transducer for each of the eight individual ultrasound features. Images acquired with the 6CIHD transducer were used to populate the atlas, because this transducer is used more often clinically for abdominal imaging at our institution and there were more data available for this transducer by having two sets of scores per reader. Analyzing the 16 separate reads (8 readers × 2 reads/reader), for each feature grade, the image that had the highest number of reads scoring it as that grade was selected as being the representative. The selected images for each grade were exported in uncompressed TIFF format for the atlas.

Statistical analysis

Statistical analysis was performed using R version 3.3.3 statistical software (R: A language and environment for statistical computing. 2016. R Foundation for Statistical Computing, Vienna, Austria).

Demographics of the study population were summarized descriptively. For each individual feature and for overall impression, three types of agreement were assessed using intraclass correlation coefficients (ICC):

- Intra-reader agreement for the 6C1HD transducer was assessed using the duplicate exams from the 6C1HD transducer.
- Inter-transducer agreement and inter-reader agreement for both transducers were assessed by randomly choosing one of the two reads from the duplicate exam from the 6C1HD transducer for each patient and reader to compare to the exam from the 4C1 transducer.

The ICC is analogous to Cohen's kappa statistic but generalizes to more than two readers. An ICC of 0–0.20 is typically characterized as slight agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–1 as near perfect agreement [18].

For each feature with three or four possible scores, the three types of agreement were recomputed after collapsing one or more scores together. This was done to explore whether reader agreement could be improved by reducing the number of score options post hoc.

Intra-reader agreement was compared informally between the resident and the academic faculty. For intra-reader agreement, ICC was computed for each reader with the mean ICC and the range of ICCs reported. For inter-transducer and inter-reader agreement, the ICCs and 95% confidence intervals (CI) were reported. 95% CIs were computed for ICCs using the exact confidence limit equation described by Searle [19].

Decision tree analysis

Classification and Regression Tree (CART) recursive models were used to select predictors of histological steatosis from features with the highest intra- and inter-reader agreement. Histological steatosis was dichotomized into grades 0 and 1 (no or mild steatosis) versus grades 2 and 3 (moderate or severe steatosis) for model stability. Models were fit for the 6C1HD and 4C1 transducers separately, with the same random 6C1HD read that was used for the inter-transducer agreement analysis. The classification accuracy and diagnostic performance characteristics of the resulting decision rules were informally compared to that of the radiologists' overall impressions, similarly dichotomized.

Results

Study population

Forty-five adult patients (18 male, 27 female) scanned from September 2015 to July 2016 were included in this analysis. Their mean age and BMI were 56 ± 12 years and 30.4 ± 5.6 kg/m^2, respectively. Histological data were available for 40 patients: 2 were Grade 0, 20 were Grade 1, 12 were Grade 2, and 6 were Grade 3 [16].

Reader agreement

Mean intra-reader ICCs computed from the repeat examinations of the 6C1HD transducer ranged from 0.430 to 0.777 for the various imaging features (Table 2). Four features had mean ICC point estimates ≥ 0.700 (large hepatic vein blurring: 0.760, liver–kidney contrast: 0.777, posterior beam attenuation: 0.706, and overall impression: 0.753). Because intra-reader agreement of the resident was comparable to that of the faculty, data were pooled for computing inter-transducer and inter-reader agreement.

Table 2. Intra-reader and inter-transducer agreement for each feature

Feature	Intra-reader agreement (Range)	Inter-transducer agreement (Range)
Large hepatic vein blurring	0.760 (0.653–0.884)	0.580 (0.331–0.838)
Main right portal vein blurring	0.667 (0.408–0.873)	0.391 (0.011–0.689)
Anterior posterior right portal vein blurring	0.611 (0.434–0.848)	0.432 (0.374–0.673)
Liver–kidney contrast	0.777 (0.630–0.894)	0.575 (0.374–0.673)
Posterior beam attenuation	0.706 (0.565–0.799)	0.538 (0.313–0.630)
Diaphragm definition	0.652 (0.538–0.858)	0.404 (0.192–0.671)
Focal fat sparing	0.644 (0.408–0.951)	0.525 (0.253–0.775)
Liver echotexture	0.430 (0.142–0.834)	0.228 (-0.108–0.705)
Overall impression	0.753 (0.508–0.911)	0.640 (0.529–0.744)

Intra-reader agreement was computed using repeated examinations from the 6C1HD transducer, and the mean ICC for all readers and the range of ICCs computed are shown. Inter-transducer agreement was computed using a random exam for the 6C1HD transducer compared to the exam for the 4C1 transducer, and the mean ICC for all readers and the range of ICCs computed are shown

Table 3. Inter-reader agreement for each feature by transducer

Feature	6C1HD transducer (95% CI)	4C1 transducer (95% CI)
Large hepatic vein blurring	0.493 (0.376–0.625)	0.414 (0.299–0.552)
Main right portal vein blurring	0.198 (0.107–0.325)	0.064 (0.003–0.159)
Anterior posterior right portal vein blurring	0.246 (0.148–0.380)	0.219 (0.125–0.349)
Liver–kidney contrast	0.524 (0.406–0.652)	0.561 (0.445–0.684)
Posterior beam attenuation	0.394 (0.279–0.533)	0.422 (0.306–0.560)
Diaphragm definition	0.318 (0.210–0.457)	0.317 (0.210–0.455)
Focal fat sparing	0.414 (0.294–0.555)	0.445 (0.324–0.584)
Liver echotexture	0.014 (− 0.035 to 0.093)	0.018 (− 0.031 to 0.099)
Overall impression	0.540 (0.423–0.666)	0.538 (0.422–0.664)

Inter-reader agreement for each feature by transducer, computed for each transducer over all seven readers. For the 6C1HD transducer, one of the two reads for each patient and reader was randomly selected. ICCs and 95% confidence intervals are shown

Mean inter-transducer ICCs between 6C1HD and 4C1 transducers ranged from 0.228 to 0.640 (Table 2). The four features with the highest intra-reader agreement also had the highest inter-transducer agreement, with mean ICC point estimates of 0.580, 0.575, 0.538, and 0.640 for large hepatic vein blurring, liver–kidney contrast, posterior beam attenuation, and overall impression, respectively.

Inter-reader ICC point estimates ranged from 0.014 to 0.540 for the 6C1HD transducer and 0.018–0.561 for the 4C1 transducer. As shown in Table 3, three features were among the four highest ranked features for inter-reader ICC for each transducer: large hepatic vein blurring (6C1HD: 0.493; 4C1: 0.414), liver–kidney contrast (0.524; 0.561), and overall impression (0.540; 0.538). Focal fat sparing (0.414) was among the four highest ranked features for the 6C1HD transducer, and posterior beam attenuation (0.422) was among the four highest ranked features for the 4C1 transducer.

Overall, large hepatic vein blurring, liver–kidney contrast, and overall impression were the three features with the highest intra-reader, inter-transducer, and inter-reader agreement. Liver echotexture had the lowest intra-reader (ICC 0.430), inter-transducer (ICC 0.228), and inter-reader (6C1HD: 0.014; 4C1: 0.018) ICC point estimates. Intra-reader, inter-transducer, and inter-reader agreement for each imaging feature are illustrated in Fig. 1.

When these analyses were repeated after collapsing scores post hoc, there was no set of collapsed scores that improved the ICC for any reader agreement (intra-reader, inter-transducer, or inter-reader) by more than 0.035; 81 of the 88 possible collapsed feature scores (22 possible collapsed feature scores for intra-reader, 22 for inter-transducer, and 22 for inter-reader using the 6C1HD transducer, and 22 for inter-reader using the 4C1 transducer) actually lowered the ICC point estimate.

Accuracy for hepatic steatosis

The four individual features with the highest inter-reader ICC point estimates using the 6C1HD transducer (large hepatic vein blurring, liver–kidney contrast, posterior beam attenuation, and focal fat sparing) were entered into the CART regression, separately for each transducer. The final CART rule was the same for each transducer and selected only a large hepatic vein blurring grade of 2 to predict grade 2 or 3 steatosis on histology. In the subset of patients with a histological reference standard, this decision rule achieved 74% accuracy (59%

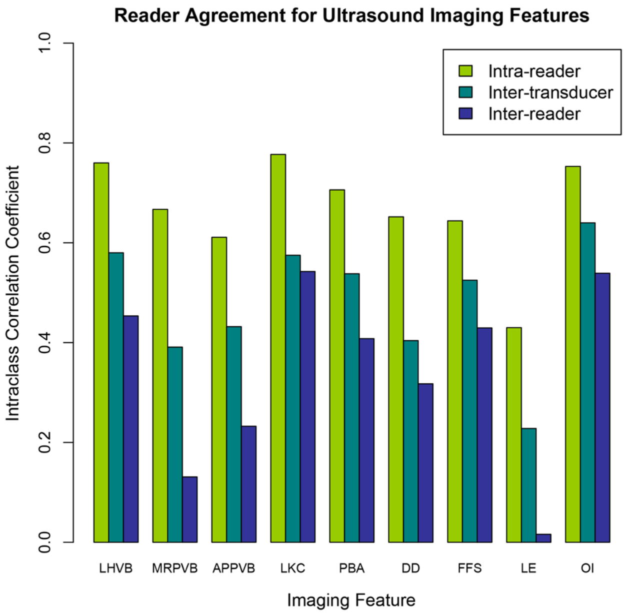


Fig. 1. Intra-reader (light green), inter-transducer (teal), and inter-reader (blue) agreement for each imaging feature are shown. For each feature, intra-reader and inter-transducer agreement are both averaged over all readers, and inter-reader agreement is averaged over the two transducers. *LHVB* large hepatic vein blurring; *MRPVB* main and right portal vein blurring; *APPVB* anterior and posterior portal vein blurring; *LKC* liver–kidney contrast; *PBA* posterior beam attenuation; *DD* diaphragm definition; *FFS* focal fat sparing; *LE* liver echotexture; *OI* overall impression.

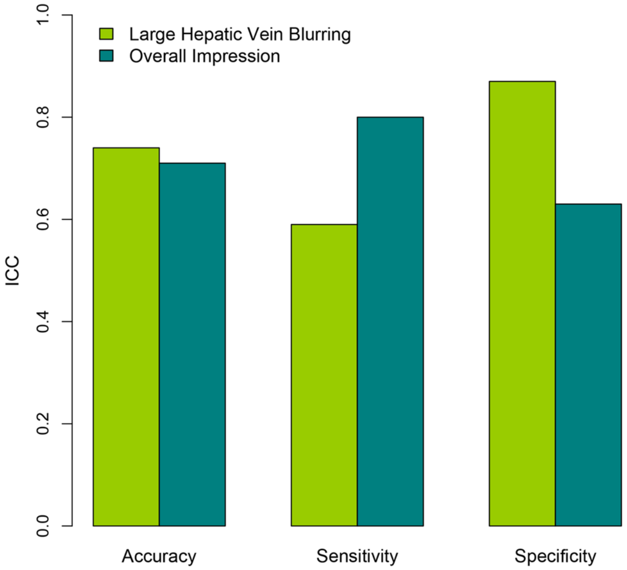


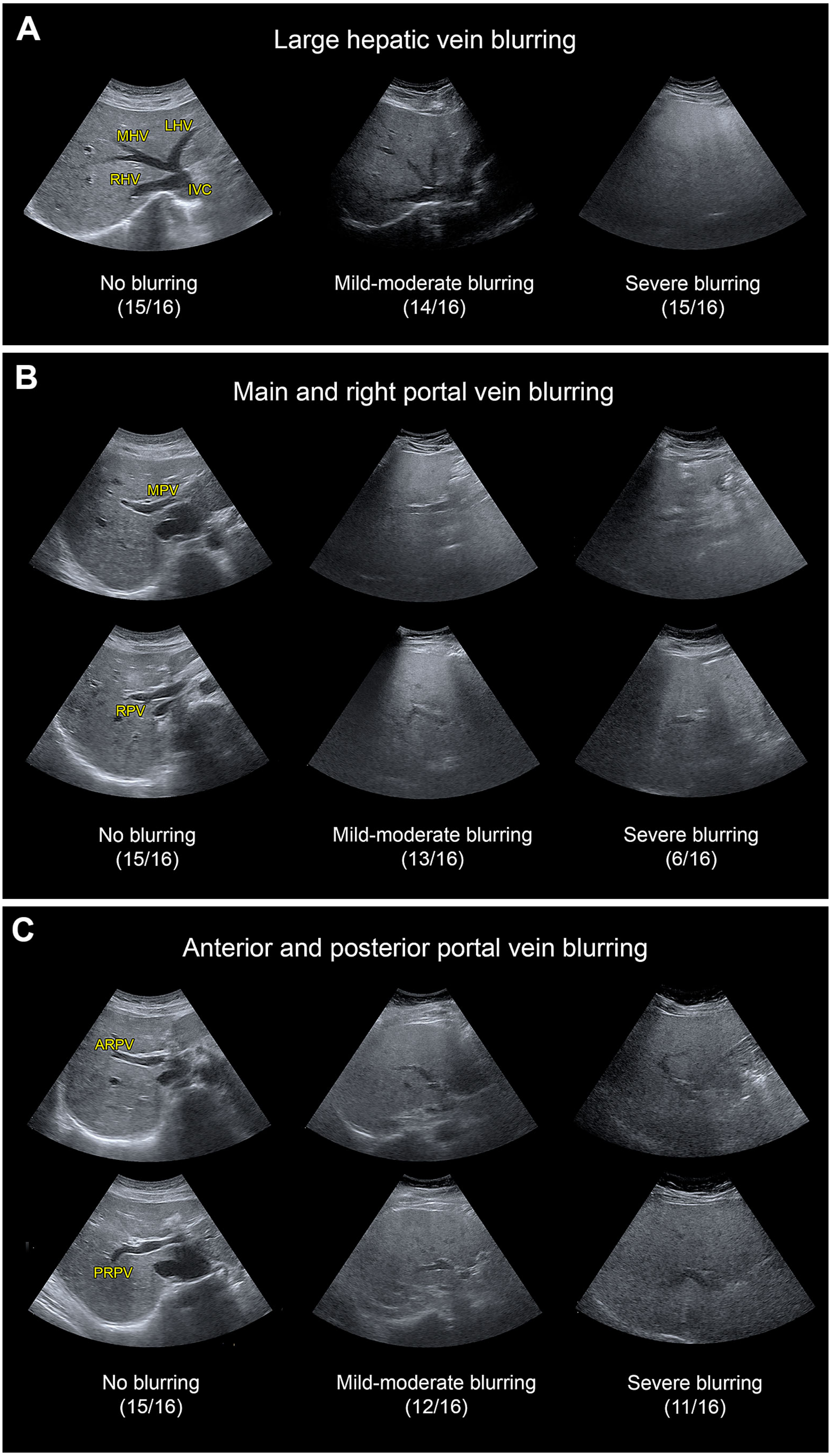
Fig. 2. Accuracy, sensitivity, and specificity of the 6C1HD and 4C1 decision rules (light green) applying large hepatic vein blurring for dichotomized hepatic steatosis compared to the radiologists' overall impression (teal). The decision rules achieved similar or slightly higher accuracy than the radiologists' overall impression, with lower sensitivity but higher specificity.

Fig. 3. **A** Scoring spectrum for large hepatic vein blurring.▶ For each scoring grade, the examination with the highest proportion of reads selecting it as that particular score was chosen. Proportions of the 16 reads (2 from each reader) identifying each exam as the corresponding score are shown. *IVC* inferior vena cava; *RHV* right hepatic vein; *MHV* middle hepatic vein; *LHV* left hepatic vein. **B** Scoring spectrum for main and right portal vein blurring. For each scoring grade, the examination with the highest proportion of reads selecting it as that particular score was chosen. Proportions of the 16 reads (2 from each reader) identifying each exam as the corresponding score are shown. *MPV* main portal vein; *RPV* right portal vein. **C** Scoring spectrum for anterior and posterior portal vein blurring. For each scoring grade, the examination with the highest proportion of reads selecting it as that particular score was chosen. Proportions of the 16 reads (2 from each reader) identifying each exam as the corresponding score are shown. *ARPV* anterior right portal vein; *PRPV* posterior right portal vein. **D** Scoring spectrum for liver–kidney contrast. For each scoring grade, the examination with the highest proportion of reads selecting it as that particular score was chosen. Proportions of the 16 reads (2 from each reader) identifying each exam as the corresponding score are shown. *L* liver; *K* kidney. **E** Scoring spectrum for posterior beam attenuation. For each scoring grade, the examination with the highest proportion of reads selecting it as that particular score was chosen. Proportions of the 16 reads (2 from each reader) identifying each exam as the corresponding score are shown. **F** Scoring spectrum for diaphragm definition. For each scoring grade, the examination with the highest proportion of reads selecting it as that particular score was chosen. Proportions of the 16 reads (2 from each reader) identifying each exam as the corresponding score are shown. L liver; *K* kidney; *D* diaphragm. **G** Scoring spectrum for focal fat sparing. For each scoring grade, the examination with the highest proportion of reads selecting it as that particular score was chosen. Proportions of the 16 reads (2 from each reader) identifying each exam as the corresponding score are shown. *MPV* main portal vein; *RPV* right portal vein. Arrow denotes the area of focal fat sparing. **H** Scoring spectrum for liver echotexture. For each scoring grade, the examination with the highest proportion of reads selecting it as that particular score was chosen. Proportions of the 16 reads (2 from each reader) identifying each exam as the corresponding score are shown.

sensitivity, 86% specificity) using the 6C1HD transducer and 75% accuracy (57% sensitivity, 90% specificity) using the 4C1 transducer for identifying patients with grade 2 or 3 steatosis (Fig. 2). By comparison, the radiologists' overall impression (grade 2 or 3) for the 6C1HD transducer achieved 68% accuracy (83% sensitivity, 57% specificity) and the overall impression for the 4C1 transducer achieved 72% accuracy (77% sensitivity, 68% specificity) for the same classification.

Atlas

Figure 3a–h shows the B-mode images acquired with the 6C1HD transducer and selected for the atlas to illustrate


A
Large hepatic vein blurring
MHV
LHV
RHV
IVC
No blurring
(15/16)
Mild-moderate blurring
(14/16)
Severe blurring
(15/16)
B
Main and right portal vein blurring
MPV
RPV
No blurring
(15/16)
Mild-moderate blurring
(13/16)
Severe blurring
(6/16)
C
Anterior and posterior portal vein blurring
ARPV
PRPV
No blurring
(15/16)
Mild-moderate blurring
(12/16)
Severe blurring
(11/16)

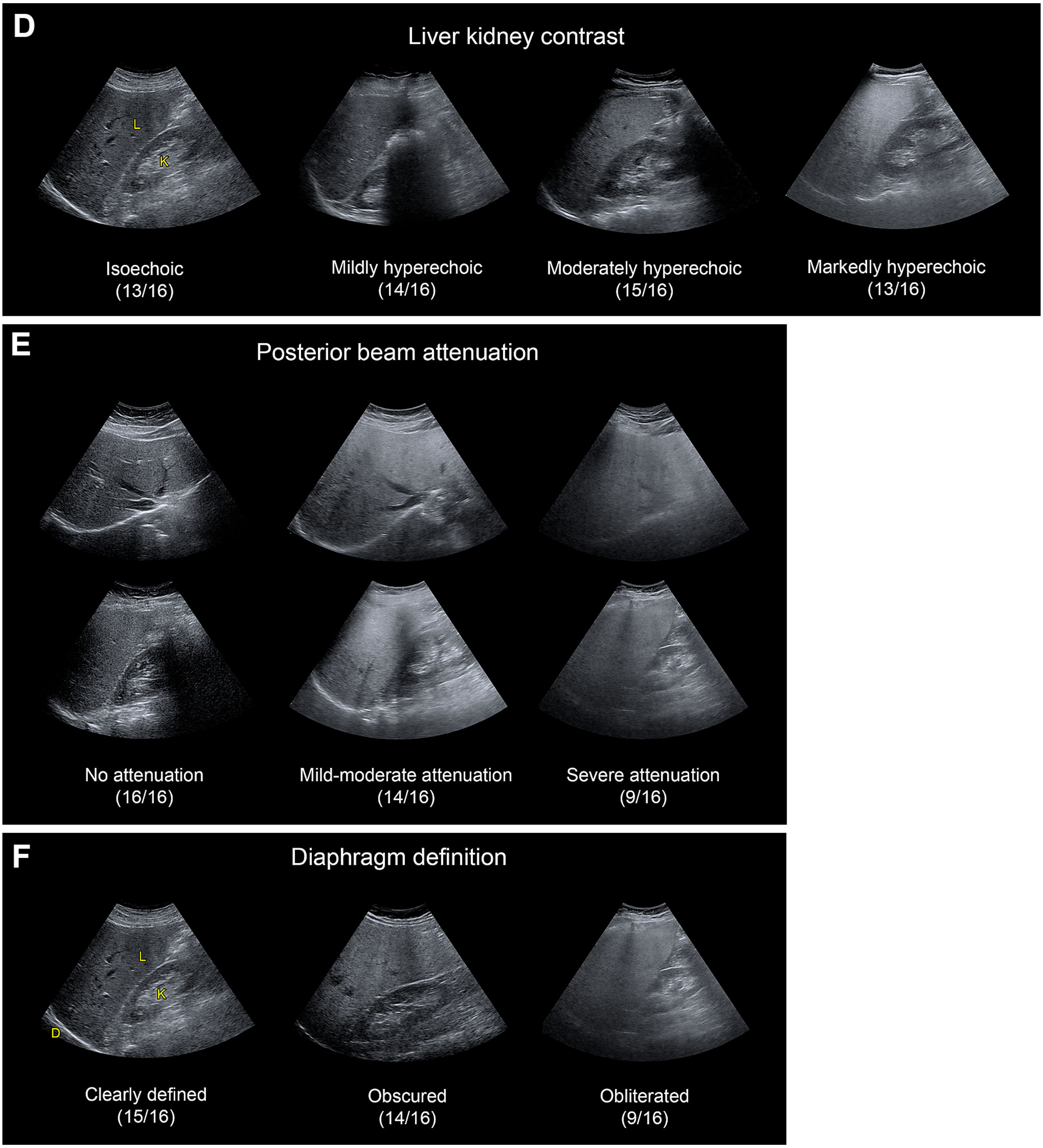


Fig. 3. continued.

the scoring spectrum for each of the 8 individual features. For each feature and score, the proportion of reads selecting that score is indicated; as shown, the proportion ranged from 6/16 to 16/16, depending on the feature and score. Of the 45 patients, 16 had at least one image selected for the atlas.

Discussion

In this study, intra-reader agreement ranged from 0.430 to 0.777, inter-transducer agreement ranged from 0.228 to 0.640, and inter-reader agreement ranged from 0.014 to 0.561, depending on the feature. Large hepatic vein

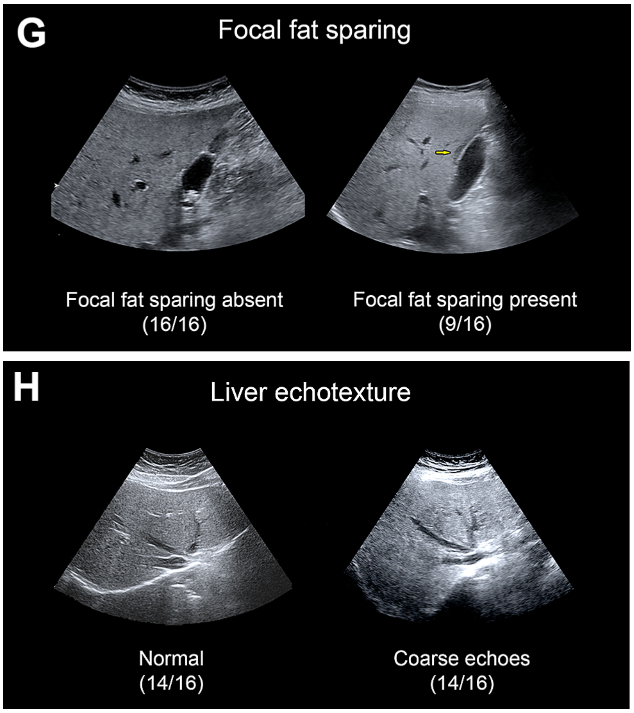


Fig. 3. continued.

blurring, liver–kidney contrast, focal fat sparing, and overall impression had the highest reader agreement. Our study used a standardized imaging protocol acquired by experienced sonographers and image interpretation by expert radiologists for published features of hepatic steatosis, thus it is unlikely that reader agreement could be further improved with increased expertise with image acquisition or interpretation. As such, if more reproducible measurements of hepatic steatosis are needed, less subjective measures such as the hepatorenal index that can be applied to conventional ultrasound images may be used [20, 21]. Also, quantitative ultrasound-based approaches such as the controlled attenuation parameter measured as part of vibration-controlled transient elastography [22–25] or more recent quantitative measurements of ultrasound backscatter and attenuation coefficients [26, 27] may also be used. Other modalities such as advanced chemical-shift-encoded MRI may be required to quantify proton density fat fraction [27–32]. If conventional ultrasound is to be used for qualitative assessment, our findings suggest that radiologists should mainly focus on features with the highest reader agreement (e.g., large hepatic vein blurring, liver–kidney contrast, and focal fat sparing). Correspondingly, the poor reader agreement of liver echotexture may limit its application in clinical practice, especially since it is known that coarse echoes can also be seen in the setting of liver cirrhosis [33, 34] and so are not specific for steatosis.

Overall, intra-reader agreement was highest, followed by inter-transducer agreement, and inter-reader agreement was lowest. One plausible explanation is that readers are able to develop and apply their own internally consistent criteria for assessing some of these features, leading to reasonably high intra-reader agreement for those features. Even when two different transducers are used, the application of internally consistent criteria by the same reader results in agreement that is higher than between different readers using the same transducer. Because the internally applied criteria may differ across readers, inter-reader agreement is consistently lower than intra-reader and inter-transducer agreement. Further research is needed to assess whether the utilization of the standardized atlas of ultrasound features developed in this study will improve reader agreement in future studies and in clinical care.

Large hepatic vein blurring was the only feature for each transducer that was selected by the CART for predicting hepatic steatosis. For each transducer, it achieved similar accuracy as the radiologists' overall impression and achieved higher specificity at the cost of lower sensitivity. Although this finding requires further validation in future studies, large hepatic vein blurring may be the imaging feature most indicative of hepatic steatosis on histology, which would suggest placing additional emphasis on this feature during clinical practice. Importantly, the same feature was selected by the CART for each transducer, suggesting the feature may be robust to the choice of transducer. Nevertheless, vessel blurring as an imaging feature of hepatic steatosis has limitations, as it potentially also may be caused by inadequate penetration in obese patients, by thickness of overlying fat on beam focus and it can depend on the type of harmonic imaging used, which may vary with manufacturers [35–37].

Limitations of this study include its relatively small sample size, which may reduce its generalizability to other cohorts. This limitation also precluded subgroup analyses such as the effect of increasing BMI on imaging features. We acquired images with only Siemens ultrasound scanners, so research is needed on scanners from other manufacturers. Within the group of patients where histological data are available, only 2 of 40 patients had steatosis grade 0. Thus, our research cohort likely has more severe hepatic steatosis than the general patient population undergoing liver ultrasound, which is consistent with the selection bias of patients where liver biopsy is clinically indicated. The use of liver histology as a reference standard is also a limitation as it only samples a small part of the liver. In addition, histologic evidence of lobular inflammation or ballooning injury was not analyzed for this study, although it is possible these tissue factors could influence appearance of steatosis on ultrasound. Even so, our study adds to the literature by assessing agreement among many expert readers and by assessing all published imaging features in a single study allowing direct comparison of the reader agreement of these features.

Although our preliminary results need external confirmation, intra- and inter-reader agreement of ultrasound imaging features for hepatic steatosis ranged from 0.430 to 0.777 and from 0.014 to 0.561, respectively, for the different imaging features. Since reader agreement varies considerably for published imaging features of hepatic steatosis on ultrasound, radiologists should perhaps focus on assessing large hepatic vein blurring, liver–kidney contrast, and overall impression, which have the highest reader agreement in this study. These results suggest large hepatic vein blurring in particular may be the most indicative of hepatic steatosis on histology. Future research is needed to assess the impact of a conventional ultrasound atlas on reader reliability and accuracy for assessing hepatic steatosis.

Compliance with ethical standards

Funding The authors acknowledge Grant support from the National Institutes of Health T32 EB005970-09 and R01 DK106419-03. The study was also supported in part by an investigator-initiated research grant from Siemens to the senior and second senior author. No author of this study is an employee or consultant to Siemens. All data belonged to the authors who had complete control of the manuscript content. The senior author has research grants from Siemens, GE, Philips, and Bayer.

Conflict of interest C.W.H., A.W., T.W., J.P., S.F.D., A.S.N., E.H., L.D., C.Q.L., A.W, H.J.J., C.F.D., G.C., M.O.B., K.M.R., M.A.V., and M.A. declare that they have no conflicts of interest. F.P. is on the advisory board and speaker bureau of Bayer, on the advisory board of EISAI, on the speaker bureau of Bracco, and has a research contract with ESAOTE. R.L. receives grant funding from Adheron, Arora, BMS, Daiichi-Sankyo Inc., Galectin, Galmed, GE, Genfit, Gilead, Immuron, Intercept, Janssen Inc, KineMed, Madrigal, Merck, NGM, Promedior, Prometheus, Siemens, Sirius, and Tobira and is a consultant or on the advisory committee for Bird Rock Bio, BMS, Boehringer Ingelheim, Celgene, Conatus, Eli Lily, Enanta, Gilead, GRI Bio, Madrigal, Metacrine, NGM, Pfizer, Receptos, Sanofi, Scholar Rock, Zafgen, Arrowhead Research, Conatus, Galmed, Gilead, Intercept, NGM, Nimbus, Octeta, Tobira, GIR, Inc., and Metacrine, Inc. C.S. receives grant funding from Gilead, GE Healthcare, Siemens, GE MRI, Bayer AMRI, GE Digital, GE US, ACR Innovation, performs contracted research for ICON Medical Imaging/Enanta, Philips, Gilead, Shire, VirtualScopics, Intercept, Synageva, is a consultant for Boehringer Ingelheim, is a member of the external advisory board of AMRA Medical and Guerbet (fee paid to the University of California Regents), and a speaker for GE Healthcare and Bayer (fee paid to the University of California Regents).

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

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

References

1. Rinella ME (2015) Nonalcoholic fatty liver disease: a systematic review. JAMA. 313(22):2263–2273
2. Lazo M, Hernaez R, Eberhardt MS, et al. (2013) Prevalence of nonalcoholic fatty liver disease in the United States: the Third National Health and Nutrition Examination Survey, 1988–1994. Am J Epidemiol 178(1):38–45
3. Williams CD, Stengel J, Asike MI, et al. (2011) Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology 140(1):124–131
4. Chalasani N, Younossi Z, Lavine JE, et al. (2012) The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. Gastroenterology 142(7):1592–1609
5. Loomba R, Sanyal AJ (2013) The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol 10(11):686–690
6. Spengler EK, Loomba R (2015) Recommendations for diagnosis, referral for liver biopsy, and treatment of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Mayo Clin Proc 90(9):1233–1246
7. Strauss S, Gavish E, Gottlieb P, Katsnelson L (2007) Interobserver and intraobserver variability in the sonographic assessment of fatty liver. AJR Am J Roentgenol 189(6):W320–W323
8. Hamaguchi M, Kojima T, Itoh Y, et al. (2007) The severity of ultrasonographic findings in nonalcoholic fatty liver disease reflects the metabolic syndrome and visceral fat accumulation. Am J Gastroenterol 102(12):2708–2715
9. Fishbein M, Castro F, Cheruku S, et al. (2005) Hepatic MRI for fat quantitation: its relationship to fat morphology, diagnosis, and ultrasound. J Clin Gastroenterol 39(7):619–625
10. Dasarathy S, Dasarathy J, Khiyami A, et al. (2009) Validity of real time ultrasound in the diagnosis of hepatic steatosis: a prospective study. J Hepatol 51(6):1061–1067
11. Saadeh S, Younossi ZM, Remer EM, et al. (2002) The utility of radiological imaging in nonalcoholic fatty liver disease. Gastroenterology 123(3):745–750
12. Yajima Y, Ohta K, Narui T, et al. (1983) Ultrasonographical diagnosis of fatty liver: significance of the liver-kidney contrast. Tohoku J Exp Med 139(1):43–50
13. Ballestri S, Lonardo A, Romagnoli D, et al. (2012) Ultrasonographic fatty liver indicator, a novel score which rules out NASH and is correlated with metabolic parameters in NAFLD. Liver Int 32(8):1242–1252
14. Hirche TO, Ignee A, Hirche H, Schneider A, Dietrich CF (2007) Evaluation of hepatic steatosis by ultrasound in patients with chronic hepatitis C virus infection. Liver Int. 27(6):748–757
15. Caturelli E, Squillante MM, Andriulli A, et al. (1992) Hypoechoic lesions in the "bright liver": a reliable indicator of fatty change. A prospective study. J Gastroenterol Hepatol 7(5):469–472
16. Kleiner DE, Brunt EM, Van Natta M, et al. (2005) Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 41(6):1313–1321
17. Noureddin M, Lam J, Peterson MR, et al. (2013) Utility of magnetic resonance imaging versus histology for quantifying changes in liver fat in nonalcoholic fatty liver disease trials. Hepatology 58(6):1930–1940
18. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biom Int Biom Soc 33(1):159
19. Searle SR (1971) A biometrics invited paper. Topics in variance component estimation. Biom Int Biom Soc 27(1):1
20. Marshall RH, Eissa M, Bluth EI, Gulotta PM, Davis NK (2012) Hepatorenal index as an accurate, simple, and effective tool in screening for steatosis. Am J Roentgenol 199(5):997–1002
21. Shiralkar K, Johnson S, Bluth EI, et al. (2015) Improved method for calculating hepatic steatosis using the hepatorenal index. J Ultrasound Med 34(6):1051–1059
22. Lédinghen V, Vergniol J, Foucher J, Merrouche W, Bail B (2012) Non-invasive diagnosis of liver steatosis using controlled attenuation parameter (CAP) and transient elastography. Liver Int 32(6):911–918
23. Myers RP, Pollett A, Kirsch R, et al. (2012) Controlled attenuation parameter (CAP): a noninvasive method for the detection of hepatic steatosis based on transient elastography. Liver Int 32(6):902–910

24. Park CC, Nguyen P, Hernandez C, et al. (2016) Magnetic resonance elastography vs transient elastography in detection of fibrosis and noninvasive measurement of steatosis in patients with biopsy-proven nonalcoholic fatty liver disease. Gastroenterology 152:598–607
25. Berzigotti A, Ferraioli G, Bota S, Gilja OH, Dietrich CF (2018) Novel ultrasound-based methods to assess liver disease: the game has just begun. Dig Liver Dis 50(2):107–112
26. Lin SC, Heba E, Wolfson T, et al. (2015) Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat using a new quantitative ultrasound technique. Clin Gastroenterol Hepatol 13(7):1337–1345.e6
27. Paige JS, Bernstein GS, Heba E, et al. (2017) A pilot comparative study of quantitative ultrasound, conventional ultrasound, and MRI for predicting histology-determined steatosis grade in adult nonalcoholic fatty liver disease. Am J Roentgenol 208(5):W168–W177
28. Bohte AE, van Werven JR, Bipat S, Stoker J (2011) The diagnostic accuracy of US, CT, MRI and 1H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: a meta-analysis. Eur Radiol 21(1):87–97
29. Reeder SB, Hu HH, Sirlin CB (2012) Proton density fat-fraction: a standardized MR-based biomarker of tissue fat concentration. J Magn Reson Imaging 36(5):1011–1014
30. Yokoo T, Shiehmorteza M, Hamilton G, et al. (2011) Estimation of hepatic proton-density fat fraction by using MR imaging at 3.0 T. Radiology 258(3):749–759
31. Idilman IS, Aniktar H, Idilman R, et al. (2013) Hepatic steatosis: quantification by proton density fat fraction with MR imaging versus liver biopsy. Radiology 267(3):767–775
32. Dulai PS, Sirlin CB, Loomba R (2016) MRI and MRE for non-invasive quantitative assessment of hepatic steatosis and fibrosis in NAFLD and NASH: clinical trials to clinical practice. J Hepatol 65(5):1006–1016
33. Heller MT, Tublin ME (2014) The role of ultrasonography in the evaluation of diffuse liver disease. Radiol Clin North Am 52(6):1163–1175
34. Bonekamp S, Kamel I, Solga S, Clark J (2009) Can imaging modalities diagnose and stage hepatic fibrosis and cirrhosis accurately? J Hepatol 50(1):17–35
35. Anvari A, Forsberg F, Samir AE (2015) A primer on the physical principles of tissue harmonic imaging. RadioGraphics 35(7):1955–1964
36. Whittingham TA (1999) Tissue harmonic imaging. Eur Radiol 9(Suppl 3):S323–S326
37. Shapiro RS, Wagreich J, Parsons RB, et al. (1998) Tissue harmonic imaging sonography: evaluation of image quality compared with conventional sonography. AJR Am J Roentgenol 171(5):1203–1206