

A Pilot Comparative Study of Quantitative Ultrasound, Conventional Ultrasound, and MRI for Predicting Histology-Determined Steatosis Grade in Adult Nonalcoholic Fatty Liver Disease

Jeremy S. Paige¹
 Gregory S. Bernstein¹
 Elhamy Heba¹
 Eduardo A. C. Costa¹
 Marilia Ferreira¹
 Tanya Wolfson^{1,2}
 Anthony C. Gamst^{1,2}
 Mark A. Valasek³
 Grace Y. Lin³
 Aiguo Han⁴
 John W. Erdman, Jr.⁵
 William D. O'Brien, Jr.⁴
 Michael P. Andre^{1,6}
 Rohit Loomba^{7,8}
 Claude B. Sirlin¹

Keywords: attenuation, backscatter, grading accuracy, interobserver agreement, nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, proton density fat fraction, quantitative imaging biomarkers, quantitative ultrasound

DOI:10.2214/AJR.16.16726

Received May 17, 2016; accepted after revision October 24, 2016.

Based on a presentation at the Society of Abdominal Radiology 2015 annual meeting, San Diego, CA.

Supported in part by grants R01 DK088925, R01 DK106419, and R01 CA111289 from the National Institutes of Health and by a research grant from Siemens Healthcare.

WEB

This is a web exclusive article.

AJR2017; 208:W1–W10

0361–803X/17/2085–W1

© American Roentgen Ray Society

OBJECTIVE. The purpose of this study is to explore the diagnostic performance of two investigational quantitative ultrasound (QUS) parameters, attenuation coefficient and backscatter coefficient, in comparison with conventional ultrasound (CUS) and MRI-estimated proton density fat fraction (PDFF) for predicting histology-confirmed steatosis grade in adults with nonalcoholic fatty liver disease (NAFLD).

SUBJECTS AND METHODS. In this prospectively designed pilot study, 61 adults with histology-confirmed NAFLD were enrolled from September 2012 to February 2014. Subjects underwent QUS, CUS, and MRI examinations within 100 days of clinical-care liver biopsy. QUS parameters (attenuation coefficient and backscatter coefficient) were estimated using a reference phantom technique by two analysts independently. Three-point ordinal CUS scores intended to predict steatosis grade (1, 2, or 3) were generated independently by two radiologists on the basis of QUS features. PDFF was estimated using an advanced chemical shift–based MRI technique. Using histologic examination as the reference standard, ROC analysis was performed. Optimal attenuation coefficient, backscatter coefficient, and PDFF cutoff thresholds were identified, and the accuracy of attenuation coefficient, backscatter coefficient, PDFF, and CUS to predict steatosis grade was determined. Interobserver agreement for attenuation coefficient, backscatter coefficient, and CUS was analyzed.

RESULTS. CUS had 51.7% grading accuracy. The raw and cross-validated steatosis grading accuracies were 61.7% and 55.0%, respectively, for attenuation coefficient, 68.3% and 68.3% for backscatter coefficient, and 76.7% and 71.3% for MRI-estimated PDFF. Interobserver agreements were 53.3% for CUS ($\kappa = 0.61$), 90.0% for attenuation coefficient ($\kappa = 0.87$), and 71.7% for backscatter coefficient ($\kappa = 0.82$) ($p < 0.0001$ for all).

CONCLUSION. Preliminary observations suggest that QUS parameters may be more accurate and provide higher interobserver agreement than CUS for predicting hepatic steatosis grade in patients with NAFLD.

Quantification of hepatic steatosis is clinically important because of the increasing frequency of nonalcoholic fatty liver disease (NAFLD), the most common chronic liver disease worldwide, with an estimated 1 billion afflicted individuals [1]. Current methods to quantify liver fat are limited. Percutaneous

liver biopsy with histologic examination, the current reference standard, is invasive and has potential complications [2]. MRI-based techniques are accurate and precise for quantifying fat [3] but may be impractical for routine clinical use given the large number of affected patients. CT can assess liver fat, but radiation dose limits its suitability for this

¹Liver Imaging Group, Department of Radiology, University of California San Diego, 408 Dickinson St, San Diego, CA 92103-8226. Address correspondence to C. B. Sirlin (csirlin@ucsd.edu).

²Computational and Applied Statistics Laboratory, University of California San Diego, San Diego, CA.

³Department of Pathology, University of California San Diego, San Diego, CA.

⁴Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL.

⁵Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, IL.

⁶San Diego VA Healthcare System, San Diego, CA.

⁷NAFLD Translational Research Unit, Division of Gastroenterology, University of California San Diego, San Diego, CA.

⁸Division of Epidemiology, Department of Family and Preventive Medicine, University of California San Diego, San Diego, CA.

purpose [4]. Conventional ultrasound (CUS) is the most widely used imaging examination for assessing steatosis because of its accessibility and low cost [5, 6]. In the CUS procedure, radiologists review conventional B-mode sonographic images of the liver and assess liver fat on the basis of qualitative features, such as hyperechogenicity of liver parenchyma, intrahepatic vessel blurring, and diaphragm blurring [7]. Because CUS is assessed using qualitative features, interpretation is subjective, as well as both machine and operator dependent. Because, in part, of these limitations, CUS has limited accuracy [8–10] and high interobserver variability [7, 11, 12] for assessing hepatic steatosis.

Recently, Lin et al. [13] reported promising results in human subjects for a quantitative ultrasound (QUS) technique that assesses hepatic steatosis objectively. This technique estimates inherent ultrasonic tissue properties by using a calibrated reference phantom with known ultrasound properties [14]. Ultrasonic echoes acquired from the liver are compared with echoes acquired from the phantom, using identical instrument settings and from the same depths, thus accounting for machine and operator dependencies. Two fundamental liver tissue parameters, attenuation coefficient and backscatter coefficient, are estimated. The attenuation coefficient is a measure of

ultrasound energy loss in tissue and provides a quantitative parameter analogous to the obscuration of liver structures (vessel and diaphragm blurring) assessed qualitatively with CUS. The backscatter coefficient is a measure of ultrasound energy returned from tissue and provides a quantitative parameter analogous to the echogenicity assessed qualitatively with CUS. In the study by Lin et al. [13], backscatter coefficient was shown to accurately detect the presence of steatosis in patients with suspected NAFLD using MRI-estimated proton density fat fraction (PDFFF) as a reference standard. Although detection accuracy was shown, a histologic reference standard was not included, and so the ability of QUS to differentiate between different histologic grades of steatosis in patients with NAFLD was not evaluated. Also, a direct comparison with CUS was not made, and the performance of QUS versus CUS was not evaluated.

Therefore, the primary objective of this pilot study was to explore, in a cohort of adults with biopsy-proven NAFLD, the diagnostic performance and reliability of QUS parameters, attenuation coefficient and backscatter coefficient, in comparison with CUS for predicting histology-determined steatosis grades. In addition, because MRI-estimated PDFFF is emerging as the leading quantitative imaging biomarker for assessment of steatosis, the di-

agnostic performance of QUS and MRI-estimated PDFFF was compared, using histologic grade as the reference standard. A secondary objective of this study was to assess the correlations of QUS parameters versus MRI-estimated PDFFF and versus histologic grade.

Subjects and Methods

Study Design and Subjects

This study was approved by the University of California, San Diego, institutional review board and is compliant with the HIPAA. This was a cross-sectional pilot analysis of a prospectively recruited cohort of 61 patients with biopsy-proven NAFLD. Because this was a pilot study, no a priori power analysis was performed. All patients underwent liver biopsy for clinical care and, within 100 days of biopsy, CUS, QUS, and MRI examinations for research. Because hepatic fat contents can vary over time [15, 16], the 100-day interval was chosen to limit possible changes in hepatic fat content during the interval between biopsy and imaging that might reduce the validity of the histology reference standard. Patients were recruited from the institutional fatty liver clinic (September 2012 to February 2014) by the study hepatologist. Patients were included if they were adults at least 18 years old and willing to undergo CUS, QUS, and MRI for research. Patients were excluded if any of the following criteria were met: moderate, heavy, or binge alcohol consumption, defined as three

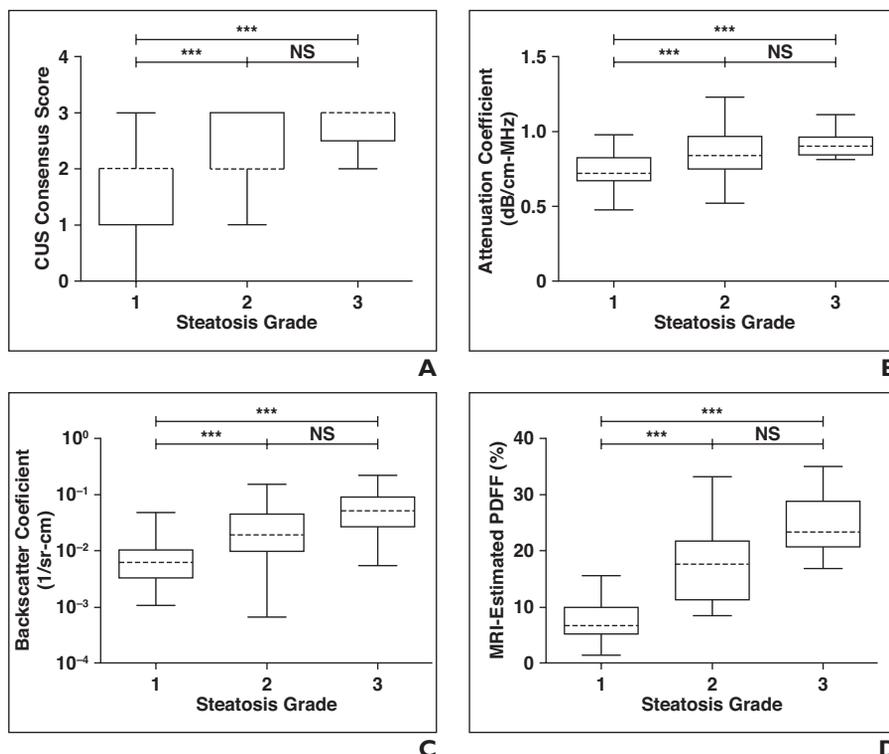


Fig. 1—Box-and-whisker plots of pairwise comparisons of all imaging measures. **A–D**, Graphs show comparisons between histology-determined steatosis grades and conventional ultrasound (CUS) consensus scores (**A**), two-analyst mean attenuation coefficient (**B**) and backscatter coefficient (**C**), and mean right liver lobe MRI-estimated proton density fat fraction (PDFFF) (**D**). All imaging-determined values were plotted as median (*dashed lines*), 25th to 75th percentile (*box*), and minimum and maximum values (*whiskers*) for each histology-determined steatosis grade. Horizontal brackets over plots indicate statistical significance as determined by Mann-Whitney test; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. NS = not significant.

Use of Imaging to Predict Steatosis in Patients With NAFLD

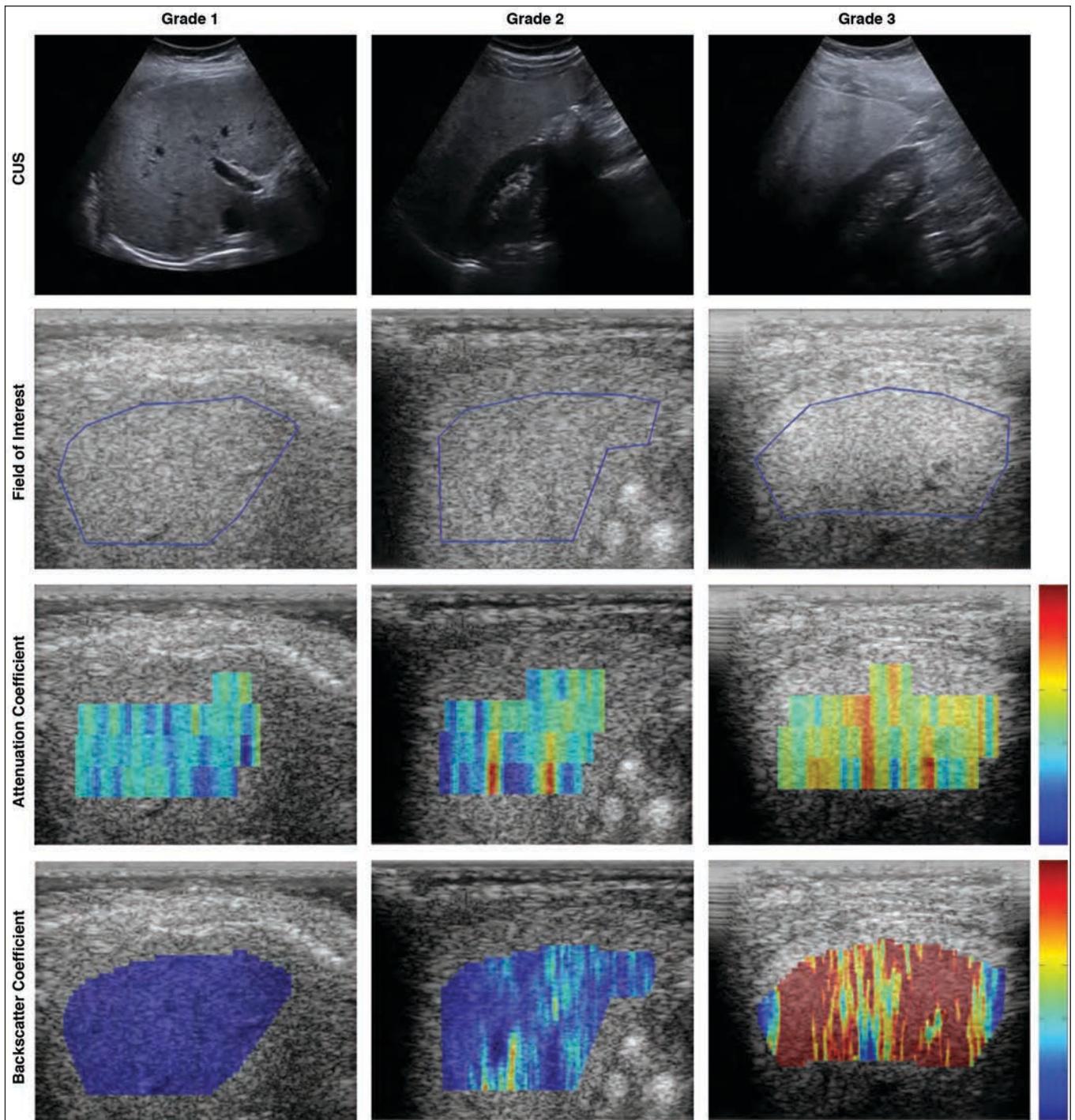


Fig. 2—Three patients with grade 1 (36-year-old woman), grade 2 (27-year-old man), and grade 3 (22-year-old man) steatosis as determined by histologic examination. Top row shows conventional ultrasound (CUS) B-mode images for each patient, which were used by radiologists for conventional scoring. Second row shows quantitative ultrasound images for same three patients with blue line outlining fields of interest as determined by one study image analyst. Third and fourth rows show parametric color-coded maps for attenuation coefficient (0–2 dB/cm-MHz) and backscatter coefficient (0–0.25 1/sr-cm), respectively. Red indicates higher values and blue indicates lower values.

drinks or more per day on average for men and two or more drinks per day on average for women; chronic liver disease other than NAFLD (e.g., viral, cholestatic, or autoimmune); use of steatogen-

ic drugs; exposure to known hepatotoxins; contraindications to MRI; and time interval exceeding 100 days between biopsy and ultrasound or MRI, or between MRI and ultrasound.

Research Clinic Visit

Patients underwent comprehensive clinical and laboratory phenotyping at the University of California, San Diego, NAFLD Translational Re-

search Unit. Demographic, anthropometric, and laboratory data were collected by trained research coordinators.

Liver Biopsy

Nontargeted percutaneous biopsies of the right liver lobe were performed for clinical care using a 16- or 18-gauge needle by hepatologists at University of California, San Diego.

Histologic Analysis

Although biopsies were performed for clinical care, histology slides subsequently were reviewed by a single expert hepatopathologist for this research. Blinded to clinical and radiologic data, this hepatopathologist scored steatosis at low-to-medium power using a 4-point ordinal score, as defined by the Nonalcoholic Steatohepatitis Clinical Research Network scoring system [17], according to the proportion of hepatocytes with macrovesicular steatosis: 0 (< 5%), 1 (5–33%), 2 (33–66%), and 3 (> 66%). Because our study included only patients with biopsy-proven NAFLD, only three steatosis grades (1, 2, and 3) were observed in the study cohort.

The pathologist also scored other features of NAFLD using the Nonalcoholic Steatohepatitis Clinical Research Network system [17]: lobular inflammation (4-point ordinal score), ballooning injury (3-point ordinal score), and fibrosis (5-point ordinal score). These data were collected to describe the study cohort.

Imaging

Patients were asked to fast for a minimum of 4 hours before imaging. Ultrasound and MRI examinations were performed on the same day if possible.

Ultrasound

Acquisition—Ultrasound examinations were performed by a research physician using a scanner (S3000, Siemens Healthcare) with a curved vector array transducer (4C1, Acuson). This transducer has a nominal frequency bandwidth of 1–4.5 MHz. Scanning was done with the patient in the dorsal decubitus position with the right arm at maximum abduction. The transducer was placed at 90° to the liver capsule through the right intercostal approach. With the patient in shallow suspended inspiration, multiple conventional B-mode images of the liver were obtained in the transverse and longitudinal planes to depict liver parenchyma, liver vessels, relative echogenicity of the liver to the kidney, and interface between the liver and diaphragm. These images were subsequently reviewed by radiologists for CUS scoring (described in the next subsection). In addition, a single QUS acquisition was obtained using the scanner's re-

		CUS			Attenuation Coefficient				
		Score Grade	1	2	3	Score Grade	1	2	3
Histology	1	12	13	2	19	0	8		
	2	1	4	7	6	1	9		
	3	0	1	13	0	0	17		
		Total Accuracy = 55%			Total Accuracy = 62%				
		Backscatter Coefficient			MRI-Estimated PDFF				
		Score Grade	1	2	3	Score Grade	1	2	3
Histology	1	22	2	3	26	1	0		
	2	4	4	8	5	3	8		
	3	1	1	15	0	0	17		
		Total Accuracy = 68%			Total Accuracy = 77%				

Fig. 3—Classification tables showing number of correctly classified (*shaded boxes*) and incorrectly classified (*unshaded boxes*) patients for each histologic steatosis grade for each imaging measure. Tables represent classification data based on two-radiologist consensus conventional ultrasound (CUS) scores, two-analyst mean quantitative ultrasound scores (attenuation and backscatter coefficients), or MRI-estimated proton density fat fraction (PDFF) scores.

search mode in a representative portion of the right lobe of the liver, while avoiding major vessels and with the patient suspending breathing [18]. This QUS acquisition captured the raw (unprocessed) radiofrequency data received by the transducer. Then without changing any scanner settings, the transducer was placed on a calibrated reference phantom and the QUS acquisition was repeated immediately beside the patient [14]. The QUS data from the patient and the phantom were exported for offline analysis. The reference phantom has known attenuation and backscatter coefficient properties, each independently calibrated.

CUS scoring—B-mode images of the liver were scored qualitatively on a 3-point ordinal scale adapted from Ballestri et al. [19] and intended to match the three histologic steatosis grades observed in our study cohort. The scoring incorporates several criteria, including increased liver echogenicity in comparison with the kidney (positive liver-to-kidney contrast), loss of clearly demarcated diaphragm border, and loss of echoes from the walls of the portal vein. Mild steatosis (CUS score 1) was defined by positive liver-to-kidney contrast with clearly demarcated diaphragm

border and with preservation of echoes from the walls of the portal vein. Moderate steatosis (CUS score 2) was defined by positive liver-to-kidney contrast with accompanying blurring of the diaphragm and loss of echoes from the walls of the peripheral branches of the portal vein. Severe steatosis (CUS score 3) was defined by a marked reduction in beam penetration, loss of diaphragm visualization, and loss of echoes from most of the portal vein wall, including the main branches.

To become familiar with and achieve consistency using the CUS scoring system, two fellowship-trained abdominal radiologists (with 5 and 6 years' experience in interpreting abdominal sonography, respectively) each underwent two separate training sessions consisting of 10 B-mode examinations of the liver acquired of patients not included in the current study. In these sessions, the radiologists scored the B-mode images independently and then in consensus; this was done to achieve uniformity in how the scoring criteria were applied. After completing this training and while blinded to each other's scores as well as all clinical, histologic, and other imaging data, the radiologists independently scored the conventional B-mode images

Use of Imaging to Predict Steatosis in Patients With NAFLD

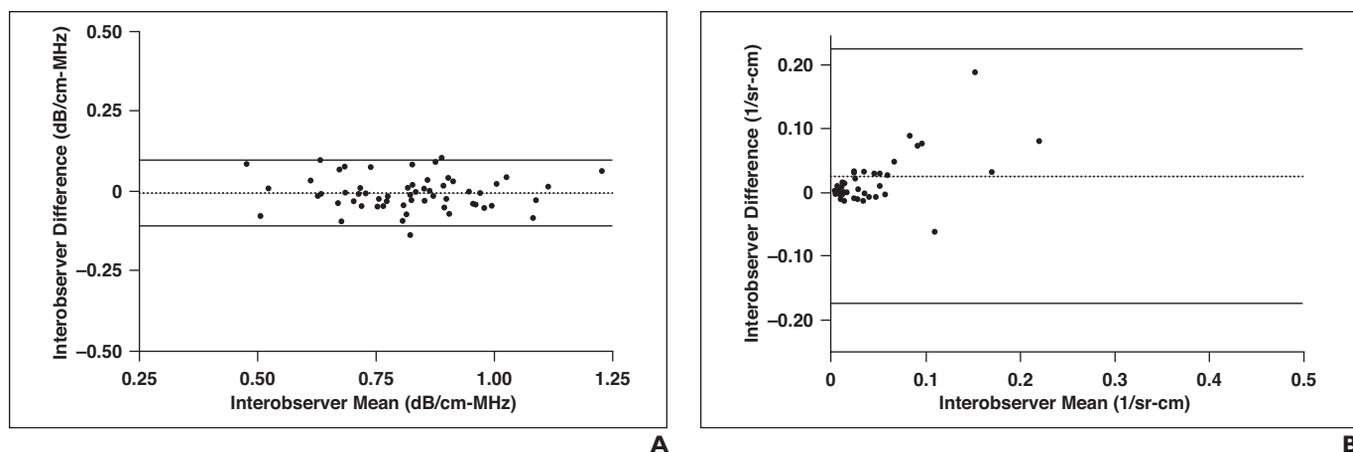


Fig. 4—Bland-Altman limits of agreement.

A and **B**, Graphs show mean interobserver agreement and differences for attenuation coefficients (**A**) and backscatter coefficients (**B**) calculated by each research analyst. One substantial outlier is not shown (interobserver mean = 0.40 1/sr-cm; interobserver difference = 0.76 1/sr-cm) for backscatter coefficient plot (**B**) for clarity. Black dotted lines represent mean difference, and black solid lines represent 95% Bland-Altman limits of agreement.

of the study patients. In addition to the two independent scores, a consensus score was assigned for each CUS examination. The consensus score was defined as the independent score if the independent scores were identical. If the independent scores were not identical, the two radiologists reviewed the images together, adjudicated their disagreements, and assigned a consensus score.

Quantitative ultrasound field of interest selection—Using custom software, a single B-mode image was reconstructed from the raw QUS data for each study patient [13]. Two image analysts (each with 4 weeks of training in the analysis procedures) who were blinded to each other's analysis as well as all clinical, histologic, and MRI data independently reviewed the reconstructed B-mode images and manually delineated polygonal fields of interest that encompassed as much liver tissue as possible, while avoiding artifacts and the most superficial 1 cm of hepatic parenchyma. If the borders of the liver parenchyma could not be adequately identified within the reconstructed B-mode image, the examination was scored as a technical failure.

Computation of attenuation coefficient and backscatter coefficient—The software automatically divided each manually traced polygonal field of interest into small (≈ 20 mm axial \times 46 A-lines lateral) overlapping (50%) rectangular ROIs. The attenuation coefficient (in units of decibels per centimeter-megahertz) within each ROI (i.e., the local attenuation coefficient) and the backscatter coefficient (in units per steradian-centimeter) within each ROI (i.e., the local backscatter coefficient) were computed as functions of ultrasound frequency, as described elsewhere [13]; local backscatter coefficient computation required correction for attenuation coefficient effects [13].

To remove operator and instrument-setting dependencies, the software automatically normalized the power spectrum of the echo signals from the liver by the corresponding power spectrum of the echo signals from the same depths in the calibrated reference phantom [13, 14]. Local attenuation and backscatter coefficient values were calculated using the normalized power spectrum of the echo signals and the known phantom attenuation and backscatter coefficient values. The mean attenuation and backscatter coefficient values of the entire hepatic field of interest were then calculated from the individual local attenuation and backscatter coefficients, respectively. Although data were collected over the entire 1–4.5 MHz frequency range of the transducer, a frequency window from 2.5 to 3.5 MHz was selected for subsequent analyses because it had the optimal signal-to-noise ratio and because the frequency window included the transducer center frequency. The mean attenuation and backscatter coefficients across this selected frequency range were calculated and used in the analyses described later in this article. Custom software was used to integrate the process of manual delineation of fields of interest with computation of attenuation and backscatter coefficient values; postprocessing time was approximately 5 minutes per patient. For illustrative purposes, parametric maps were generated to show the spatial distribution of attenuation and backscatter coefficient values in the frequency of 2.5–3.5 MHz within the analyst-defined fields of interest.

MRI Examination and Analysis

Patients were examined in the supine position with a standard torso phased-array coil centered over the liver at 3 T (Signa Excite HD, GE Healthcare; eight-channel coil). PDFF was es-

timated with a magnitude-based gradient recalled-echo technique using a low flip angle to minimize T1 bias and six gradient-recalled echoes to calculate and correct for T2* signal decay [3, 20]. Parametric liver PDFF maps were then computed pixel by pixel from source images using custom-developed software that models observed signal as a function of TE, taking into account the multiple frequency components of proton signals from triglyceride [3, 20, 21]. A single trained image analyst (with 2 years of experience), who was blinded to other QUS and CUS results, as well as to clinical and histologic data, manually outlined each liver segment on the MRI-estimated PDFF maps. The mean PDFF in the right-lobe segments (segments 5–8) was calculated and used in this study as the per-patient right-lobe PDFF value.

Statistical Analysis

Cohort characteristics were summarized descriptively.

Pairwise comparison of imaging measures and histologic steatosis grades—Imaging measures (consensus CUS scores, two-analyst mean attenuation coefficient values, two-analyst mean backscatter coefficient values, and MRI-estimated PDFF values) for different steatosis grades were compared pairwise (Mann-Whitney test).

Classification of steatosis grades—ROC curve analysis was performed on each of the quantitative imaging measures of interest: attenuation coefficient derived by each of the two analysts and the two analysts' mean, backscatter coefficient derived by each analyst and the two analysts' mean, and MRI-estimated PDFF. For each quantitative measure, two ROC curves were generated: one for separating histologic steatosis grade 1 versus grades 2 or higher and another for separating his-

TABLE 1: Patient Clinical Characteristics and Laboratory Data

Characteristic	Value
No. of patients	60
Sex	
Female	30 (50.0)
Male	30 (50.0)
Age (y), mean \pm SD (range)	50 \pm 14 (22–76)
Body mass index, mean \pm SD (range)	32.6 \pm 6.9 (23.8–68.7)
Waist circumference (cm), mean \pm SD (range)	104 \pm 12 (82–133)
Steatosis grade	
0 (< 5% hepatocytes involved)	0 (0.0)
1 (5–33% hepatocytes involved)	27 (45.0)
2 (33–66% hepatocytes involved)	16 (26.7)
3 (> 66% hepatocytes involved)	17 (28.3)
Lobular inflammation	
0 (no foci)	0 (0.0)
1 (< 2 foci per 200 \times field)	28 (46.7)
2 (2–4 foci per 200 \times field)	31 (51.7)
3 (> 4 foci per 200 \times field)	1 (1.7)
Hepatocellular ballooning	
0 (none)	17 (28.3)
1 (few balloon cells)	31 (51.7)
2 (many cells or prominent)	12 (20.0)
Fibrosis stage	
0 (none)	26 (43.3)
1 (perisinusoidal or periportal)	18 (30.0)
2 (perisinusoidal and periportal)	6 (10.0)
3 (bridging fibrosis)	6 (10.0)
4 (cirrhosis)	4 (6.7)
Nonalcoholic steatohepatitis diagnosis	
0 (not steatohepatitis)	14 (23.3)
1 (possible or borderline)	9 (15.0)
2 (definite steatohepatitis)	37 (61.7)
Attenuation coefficient (dB/cm-MHz), mean \pm SD (range) ^a	0.82 \pm 0.15 (0.48–1.23)
Backscatter coefficient (1/sr-cm), mean \pm SD (range) ^a	0.038 \pm 0.064 (0.001–0.400)
MRI-estimated proton density fat fraction (%), mean \pm SD (range) ^a	15.0 \pm 9.0 (1.4–35.0)

Note—Unless otherwise indicated, data are number (%) of patients. Body mass index is weight in kilograms divided by the square of height in meters.

^aMean calculated for right lobe of liver.

tologic steatosis grades 2 or less versus grade 3. AUC values were computed, and DeLong 95% CIs were constructed around each AUC. Optimal thresholds for separating histologic steatosis grade 1 versus grades 2 or higher and grades 2 or less versus grade 3 were selected on the basis of the Youden index, to provide the maximum possible sum of sensitivity and specificity. On the basis of these thresholds, the quantitative imaging mea-

sures (attenuation coefficient, backscatter coefficient, and MRI-estimated PDFF) were converted into 3-point ordinal scores intended to match the ordinal histologic steatosis grade. Consensus CUS score was already defined to match the steatosis grades and, thus, no conversion was necessary.

Accuracy and cross-validation—Raw and eight-fold cross-validated accuracies (percentage of correctly classified histologic grades) were estimated

for the attenuation coefficient–based, backscatter coefficient–based, and MRI-estimated PDFF–based ordinal scores along with exact binomial 95% CIs. Accuracy and binomial 95% CIs were also computed for the CUS score assigned independently by each of the two radiologists, as well as for the two radiologists' consensus score. Cross-validation was not applicable to the CUS scoring, because the scores were based on criteria established a priori and not on cohort-derived thresholds.

Interobserver agreement—In a secondary analysis, we examined interobserver agreement for the continuous attenuation and backscatter coefficient values using Bland-Altman analysis. We also assessed interobserver percentage agreement and corresponding Cohen kappa values for ordinal steatosis scores derived from CUS, attenuation coefficient, and backscatter coefficient; 95% CIs and *p* values were computed for each kappa value.

Correlation Analysis of Quantitative Ultrasound Versus MRI-Estimated Proton Density Fat Fraction and Versus Histologic Steatosis Grade

Attenuation and backscatter coefficient values, for each analyst individually as well as for the two-analyst means, were correlated on a continuous scale with either MRI-estimated PDFF values (continuous scale) or histologic steatosis grades (ordinal scale). Spearman correlation coefficients were computed, and the significance of each correlation was assessed.

Results

Patients

Sixty-one patients with histology-confirmed NAFLD (steatosis grade \geq 1), for whom other causes of liver disease were excluded clinically and by laboratory testing, underwent research CUS, QUS, and MRI examinations within 100 days of right-liver-lobe biopsy performed for clinical care. CUS and QUS examinations were performed on the same day for all patients. The time intervals were 1–100 days (median, 32 days) between CUS and QUS and biopsy, 0–31 days (median, 0 days) between CUS and QUS and MRI, and 1–100 days (median, 35 days) between MRI and biopsy. For one patient (24-year-old woman; body mass index [BMI; weight in kilograms divided by the square of height in meters], 37.1), the QUS acquisition was considered a technical failure by both analysts, likely due to obscuring of the echoes by the ribs or patient motion, so this patient was excluded from subsequent analyses.

Use of Imaging to Predict Steatosis in Patients With NAFLD

TABLE 2: Raw Quantitative Ultrasound Estimates

Coefficient	All Patients (n = 60)	Steatosis Grade 1 (n = 27)	Steatosis Grade 2 (n = 16)	Steatosis Grade 3 (n = 17)
Attenuation coefficient (dB/cm-MHz)				
Analyst 1	0.82 ± 0.15 (0.43–1.20)	0.75 ± 0.13 (0.43–1.01)	0.84 ± 0.18 (0.52–1.20)	0.92 ± 0.09 (0.81–1.12)
Analyst 2	0.81 ± 0.15 (0.47–1.26)	0.73 ± 0.13 (0.47–0.95)	0.85 ± 0.17 (0.53–1.26)	0.91 ± 0.09 (0.78–1.12)
Mean for both analysts	0.82 ± 0.15 (0.48–1.23)	0.74 ± 0.12 (0.48–0.98)	0.85 ± 0.17 (0.52–1.23)	0.92 ± 0.09 (0.81–1.11)
Backscatter coefficient (1/sr-cm)				
Analyst 1	0.0250 ± 0.0357 (0.0004–0.1799)	0.0081 ± 0.0103 (0.0004–0.0508)	0.0214 ± 0.0193 (0.0005–0.0583)	0.0551 ± 0.0521 (0.0051–0.1799)
Analyst 2	0.0505 ± 0.1105 (0.0009–0.7782)	0.0377 ± 0.1484 (0.0016–0.7782)	0.0485 ± 0.0629 (0.0009–0.2466)	0.0726 ± 0.0687 (0.0058–0.2607)
Mean for both analysts	0.0377 ± 0.0645 (0.0007–0.4003)	0.0229 ± 0.0760 (0.0011–0.4003)	0.0350 ± 0.0398 (0.0007–0.1524)	0.0639 ± 0.0580 (0.0054–0.2203)

Note—Data are mean ± SD (range).

QUS was considered technically successful for the remaining 60 patients. These 60 patients included 30 men and 30 women with a mean age of 50 years (range, 22–76 years). The mean BMI was 32.6. Of the 60 patients, 22 (37%) had BMI less than 30.0, 18 (30%) had BMI 30.1–35.0, 15 (25%) had BMI 35.1–40.0, and five (8%) had BMI greater than 40.0; 27 (45.0%) had histologic grade 1 steatosis, 16 (26.7%) had grade 2, and 17 (28.3%) had grade 3. Detailed cohort characteristics are presented in Table 1.

Pairwise Comparison of Imaging Measures and Histologic Steatosis Grades

Detailed information about attenuation and backscatter coefficient estimates for each analyst across different steatosis grades is provided in Table 2. For patients with steatosis grade 1 (n = 27), the median CUS consensus score was 2, the mean (± SD) attenuation coefficient was 0.74 ± 0.12 dB/cm-MHz, the mean backscatter coefficient was 0.0229 ± 0.0760 1/sr-cm, and the mean MRI-estimated PDFF was 7.3% ± 3.2%. For patients with steatosis grade 2 (n = 16), the median CUS consensus score was 2, the mean attenuation coefficient was 0.85 ± 0.17 dB/cm-MHz, the mean backscatter coefficient was 0.0350 ± 0.0398 1/sr-cm, and the mean MRI-estimated PDFF was 17.5% ± 7.2%. Finally, for patients with steatosis grade 3 (n = 17), the median CUS consensus score was 3, the mean attenuation coefficient was 0.92 ± 0.09 dB/cm-MHz, the mean backscatter coefficient was 0.0639 ± 0.0580 1/sr-cm, and mean the MRI-estimated PDFF was 24.7% ± 5.4%.

In pairwise comparisons, all imaging measures were significantly different (p < 0.05) in patients with steatosis grades 1 ver-

sus 2 and in patients with steatosis grades 1 versus 3 (Figs. 1A–1D). However, only backscatter coefficient and MRI-estimated PDFF were significantly different in patients with steatosis grades 2 versus 3 (Figs. 1C and 1D). Parametric color-coded maps of attenuation and backscatter coefficient values reflected these differences in patients with different steatosis grades (Fig. 2).

Classification of Steatosis Grades

Table 3 summarizes the results of the ROC analyses of the quantitative imaging measures for differentiating between dichotomized steatosis grades. Depending on the classification, AUCs were 0.779–0.804 for attenuation coefficient, 0.811–0.860 for backscatter coefficient, and 0.929–0.962 for MRI-estimated PDFF. Optimal thresholds based on the Youden index are listed in Table 3, with the corresponding raw sensitivities and specificities. The two-analyst mean QUS thresholds and MRI-estimated PDFF thresholds for separating steatosis grades 1 versus 2 or higher and grades 2 or less versus 3 were 0.809 and 0.815 dB/cm-MHz for attenuation coefficient, 0.0112 and 0.0166 1/sr-cm for backscatter coefficient, and 13.4% and 16.8% for MRI-estimated PDFF. Using these thresholds, quantitative imaging measures were converted into ordinal imaging-based steatosis scores to predict the ordinal steatosis grades.

Accuracy and Cross-Validation

Table 4 summarizes the raw steatosis grading accuracies of CUS scores and of the attenuation coefficient–derived, backscatter coefficient–derived, and MRI-estimated PDFF–derived scores. Cross-validated grading accuracies are also provided for two-ana-

lyst mean attenuation coefficient– and backscatter coefficient–derived scores as well as for MRI-estimated PDFF–derived scores. Grading accuracies tended to be higher for the attenuation coefficient– and backscatter coefficient–derived scores than for the CUS scores, although the highest accuracies were provided by the MRI-estimated PDFF–derived scores. Imaging modalities varied with regard to the number of correctly and incorrectly classified patients within each steatosis grade (Fig. 3). The QUS-based scores showed fewer misclassifications than did CUS-based scores when the histologic steatosis grade was 1 or 3, whereas CUS showed fewer misclassifications when the histologic steatosis grade was 2.

Interobserver Agreement

For attenuation coefficient, the mean difference in the values calculated between the two analysts was –0.01 dB/cm-MHz, and the Bland-Altman 95% adjusted limits of agreement were ± 0.199 across a range of mean values from 0.48 to 1.23 dB/cm-MHz (Fig. 4A). For backscatter coefficient, the mean difference in the values calculated between the two analysts was 0.025 1/sr-cm and the Bland-Altman 95% adjusted limits of agreement were ± 0.102 across a range of mean values from 0.001 to 0.400 1/sr-cm (Fig. 4B). There was one significant outlier for backscatter coefficient (0.400 1/sr-cm; not shown in Fig. 4B), in which the mean difference was large between the two analysts. This difference was due to inclusion of the liver capsule in the field of interest by one of the analysts.

The interobserver percentage agreement for predicting steatosis grade was 53.3% (κ = 0.61) for CUS, 90.0% (κ = 0.87) for at-

tenuation coefficient, and 71.7% ($\kappa = 0.82$) for backscatter coefficient. All kappa values were significant ($p < 0.0001$). These findings are summarized in Table 5.

Correlation Analysis of Quantitative Ultrasound Versus MRI-Estimated Proton Density Fat Fraction and Versus Histologic Steatosis Grade

As summarized in Table 6, the Spearman correlation coefficients between both attenuation coefficient and backscatter coefficient versus MRI-estimated PDFF or versus histologic steatosis grade were positive and statistically significant ($p < 0.001$) for each analyst individually as well as for the two-analyst means. However, for both attenuation coefficient and backscatter coefficient, the correlations versus MRI-estimated PDFF were nominally higher than those versus histologic steatosis grade. In addition, the correlation between backscat-

ter coefficient versus MRI-estimated PDFF and versus histologic grade were nominally higher than the corresponding correlations for attenuation coefficient.

Discussion

In this prospectively designed pilot study, we found that the QUS-derived parameters attenuation coefficient and backscatter coefficient increased progressively with greater histology-determined steatosis grades. Using thresholds derived from dichotomized steatosis grades, our preliminary results suggest that attenuation coefficient and backscatter coefficient values may be more accurate for classifying steatosis grade than CUS scores, although they were less accurate than MRI-estimated PDFF. QUS provided higher interobserver agreement for predicting steatosis grade than did CUS. Moreover, as a secondary analysis, we found that both attenuation coefficient and backscatter coefficient

correlate nominally more with MRI-estimated PDFF than with histologic steatosis grade.

These preliminary results warrant further study of these investigational QUS parameters for the objective assessment of hepatic steatosis. The backscatter coefficient is a measure of ultrasound energy returned from tissue and provides a quantitative parameter analogous to the echogenicity assessed qualitatively on the B-mode image. The attenuation coefficient is a measure of ultrasound energy loss in tissue and provides a numeric parameter analogous to posterior beam attenuation. Both attenuation and backscatter coefficients are dependent on tissue structure and composition. Previous studies of CUS have suggested that increased liver echogenicity is a more accurate qualitative marker of hepatic steatosis than is posterior beam attenuation [9]. Our observation that backscatter coefficient is a more accurate quantitative marker of steatosis grade than is at-

TABLE 3: ROC Analysis for Attenuation Coefficient, Backscatter Coefficient, and MRI-Estimated Proton Density Fat Fraction (PDFF)

Coefficient, Analyst, and Steatosis Grade Dichotomization	Mean AUC Value	Threshold ^a	Sensitivity (%)	Specificity (%)
Attenuation coefficient				
Analyst 1				
Grade 1 vs ≥ 2	0.779 (0.655–0.903)	0.790	0.848 (0.681–0.949)	0.704 (0.498–0.862)
Grade ≤ 2 vs 3	0.796 (0.655–0.903)	0.813	1.000 (0.805–1.000)	0.581 (0.421–0.730)
Analyst 2				
Grade 1 vs ≥ 2	0.799 (0.680–0.918)	0.766	0.879 (0.718–0.966)	0.741 (0.537–0.889)
Grade ≤ 2 vs 3	0.792 (0.681–0.903)	0.779	1.000 (0.805–1.000)	0.605 (0.444–0.750)
Mean for both analysts				
Grade 1 vs ≥ 2	0.793 (0.676–0.911)	0.809	0.818 (0.645–0.930)	0.704 (0.498–0.862)
Grade ≤ 2 vs 3	0.804 (0.696–0.913)	0.815	1.000 (0.805–1.000)	0.605 (0.444–0.750)
Backscatter coefficient				
Analyst 1				
Grade 1 vs ≥ 2	0.838 (0.736–0.941)	0.0067	0.879 (0.718–0.966)	0.667 (0.460–0.835)
Grade ≤ 2 vs 3	0.855 (0.755–0.955)	0.0260	0.765 (0.501–0.932)	0.837 (0.693–0.932)
Analyst 2				
Grade 1 vs ≥ 2	0.860 (0.753–0.966)	0.0160	0.788 (0.611–0.910)	0.889 (0.708–0.976)
Grade ≤ 2 vs 3	0.811 (0.698–0.924)	0.0236	0.824 (0.566–0.962)	0.767 (0.614–0.882)
Mean for both analysts				
Grade 1 vs ≥ 2	0.854 (0.748–0.961)	0.0112	0.848 (0.681–0.949)	0.815 (0.619–0.937)
Grade ≤ 2 vs 3	0.830 (0.719–0.942)	0.0166	0.882 (0.636–0.985)	0.744 (0.588–0.865)
MRI-estimated PDFF segments analyzed, segments 5–8				
Grade 1 vs ≥ 2	0.962 (0.922–1.000)	13.45	0.848 (0.681–0.949)	0.963 (0.810–0.999)
Grade ≤ 2 vs 3	0.929 (0.865–0.993)	16.83	1.000 (0.805–1.000)	0.814 (0.666–0.916)

Note—Data in parentheses are 95% CIs.

^aThreshold for attenuation coefficient is based on Youden index (units = dB/cm-MHz), threshold for backscatter coefficient is based on Youden index (units = 1/sr-cm), and threshold for MRI-estimated PDFF segment analysis is based on Youden index (units = %).

Use of Imaging to Predict Steatosis in Patients With NAFLD

TABLE 4: Accuracy of Imaging-Based Scores for Predicting Steatosis Grade in 60 Patients

Imaging Modality	Accuracy	Cross-Validated Accuracy
Conventional ultrasound score		
Radiologist 1	36 (60.0)	
Radiologist 2	23 (38.3)	
Consensus	31 (51.7)	
Attenuation coefficient–derived score		
Analyst 1	37 (61.7)	33 (55.0)
Analyst 2	39 (65.0)	
Mean for both analysts	37 (61.7)	
Backscatter coefficient–derived score		
Analyst 1	38 (63.3)	41 (68.3)
Analyst 2	42 (70.0)	
Mean for both analysts	41 (68.3)	
MRI-estimated proton density fat fraction–derived score, right lobe (segments 5–8) mean	46 (76.7)	43 (71.3)

Note—Data are number (%) of patients accurately predicted.

tenation coefficient is consistent with these prior results. By comparison, prior studies have found that posterior beam attenuation is sensitive for diagnosing severe steatosis [9], and we found that attenuation coefficient showed 100% accuracy for predicting the presence of grade 3 steatosis (Fig. 3).

Although only limited data exist on the topic, our measurements of the accuracy of CUS and MRI-estimated PDFF for predicting steatosis grade are in agreement with prior literature findings. Previous studies have reported the accuracy of CUS for predicting histology-determined steatosis grade as 53–57% [10], which is in line with the consensus CUS accuracy of 51.7% in this study. Our finding that MRI-estimated PDFF can classify steatosis grade with an accuracy of 76.7% is also in agreement with previous studies [3]. Several studies have been published on the interobserver agreement of CUS for grading steatosis [7, 11, 12]. Our finding that CUS observers were in agreement 53.3% of the time is within range of previous studies reporting 40–64% interobserver agreement [11, 12]; however, our interobserver CUS kappa value of 0.61 was higher than the kappa value range of 0.20–0.54 reported previously [7, 11, 12].

An alternative ultrasound-based parameter, the controlled attenuation parameter, has also been developed to quantify steatosis in patients with NAFLD. The controlled attenuation parameter quantifies the degree

of ultrasound attenuation in a region of tissue examined by vibration control transient elastography [22]. The controlled attenuation parameter can accurately diagnose hepatic steatosis [22] but has low accuracy for

grading hepatic steatosis because of the extensive overlap in controlled attenuation parameter values between steatosis grades [23–25]. In comparison, we found backscatter coefficient values to be significantly different in pairwise comparisons of grades 1–3, and the attenuation coefficient was significantly different between all grades except 2 versus 3. Second, the accuracy of the controlled attenuation parameter decreases in obese subjects when compared with nonobese subjects [23], and the technical failure rate associated with obesity can be as high as 33% [26]. In comparison, although most patients included in our QUS study were obese, the technical failure rate was only 1.6% (1/61), suggesting that QUS may be more robust for obese subjects than the controlled attenuation parameter. Third, the controlled attenuation parameter does not provide anatomic images and the location in the liver from which the measurements are made cannot be recorded, which may complicate longitudinal monitoring. Finally, the controlled attenuation parameter is available on only one instrument type, whereas QUS in principle could be available across multiple instruments and manufacturers.

With over 1 billion patients diagnosed with NAFLD worldwide, it is important for

TABLE 5: Interobserver Agreement of Conventional Ultrasound– and Quantitative Ultrasound–Based Scores for Predicting Steatosis Grade

Imaging Modality	Agreement (%)	<i>p</i>	κ	<i>p</i>
Conventional ultrasound	53.3 (40.0–66.3)	0.6985	0.61 (0.46–0.76)	< 0.0001
Attenuation coefficient	90.0 (79.5–96.2)	< 0.0001	0.87 (0.75–0.99)	< 0.0001
Backscatter coefficient	71.7 (58.6–82.5)	0.0012	0.82 (0.74–0.91)	< 0.0001

Note—Data in parentheses are 95% CIs.

TABLE 6: Correlation of Attenuation Coefficient and Backscatter Coefficient With MRI-Estimated Proton Density Fat Fraction (PDFF) and Histology-Determined Steatosis Grade

Coefficient	Spearman Correlation Coefficient (<i>p</i>) ^a	
	MRI-Estimated PDFF Value	Histology-Determined Steatosis Grade
Attenuation coefficient		
Analyst 1	0.64 (0.45–0.77)	0.53 (0.31–0.70)
Analyst 2	0.69 (0.52–0.81)	0.55 (0.34–0.71)
Mean for both analysts	0.69 (0.53–0.81)	0.55 (0.34–0.71)
Backscatter coefficient		
Analyst 1	0.70 (0.53–0.81)	0.64 (0.46–0.77)
Analyst 2	0.73 (0.58–0.83)	0.65 (0.47–0.78)
Mean for both analysts	0.72 (0.57–0.83)	0.67 (0.49–0.79)

Note—Data in parentheses are 95% CIs.

^a*p* < 0.001 for all values.

health care providers to have an accurate, reproducible, and cost-effective technique to assess the degree of hepatic steatosis. The QUS technology described in this article shows promise as a potentially more accurate and less observer-dependent alternative to CUS for objective noninvasive assessment of steatosis. To our knowledge, our study is the first to identify QUS thresholds for predicting histology-determined steatosis grade in patients with NAFLD. This is an important advance because QUS has the potential to address many of the machine- and observer-related dependencies of CUS that limit the accuracy and repeatability of sonography. It is possible that with further optimization, QUS may provide a safer more practical alternative to liver biopsy and MRI-estimated PDFF for quantifying fat and monitoring response to interventions.

Although statistically significant differences were found for mean attenuation coefficient and backscatter coefficient values among groups of patients with different steatosis grades, there was considerable overlap among the groups for all imaging modalities in this study (Fig. 1). As a result, the differences between threshold values used for predicting steatosis grade were small for each QUS analyst and for the two-analyst mean. Although the two-analyst mean values resulted in an ordinal score-based model that could predict steatosis grade with reasonable accuracy, interobserver agreement was modest and the mean interobserver bias for measurements of attenuation coefficient and backscatter coefficient were greater than the difference between the individual threshold values for attenuation coefficient and backscatter coefficient. Therefore, QUS may eventually provide a safer more practical alternative to liver biopsy; however, before this happens, applied precision will need to be improved significantly to ensure that standardization across multiple clinical settings is possible.

Our study had several additional limitations. Because this was a pilot proof-of-principle study of an investigational technology, we included only a small cohort of patients at a single center and focused on a single ultrasound scanner. Another limitation was the absence of patients without steatosis (histology-confirmed steatosis grade 0). Our reference standard for this study was histologic examination, and because patients with non-clinical steatosis do not regularly undergo biopsy, it would have been unethical to include additional healthy volunteers. However, a recent study of a cohort of 204 patients showed that backscatter coefficient accurately differ-

entiates patients with and without steatosis using MRI-estimated PDFF as a reference [13].

In summary, our preliminary results suggest that the QUS parameters attenuation coefficient and backscatter coefficient may be more accurate and provide higher interobserver agreement than CUS for predicting the histology-confirmed steatosis grade of adults with NAFLD. Larger prospective studies with scanners of different manufacturers and different scanner operators are needed to confirm our findings and to establish the accuracy, reproducibility, and repeatability of QUS for grading hepatic steatosis in clinical practice.

References

- Loomba R, Sanyal AJ. The global NAFLD epidemic. *Nat Rev Gastroenterol Hepatol* 2013; 10:686–690
- Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD; American Association for the Study of Liver Diseases. Liver biopsy. *Hepatology* 2009; 49:1017–1044
- Tang A, Desai A, Hamilton G, et al. Accuracy of MR imaging-estimated proton density fat fraction for classification of dichotomized histologic steatosis grades in nonalcoholic fatty liver disease. *Radiology* 2015; 274:416–425
- Shah PK, Mudaliar S, Chang AR, et al. Effects of intensive insulin therapy alone and in combination with pioglitazone on body weight, composition, distribution and liver fat content in patients with type 2 diabetes. *Diabetes Obes Metab* 2011; 13:505–510
- Khov N, Sharma A, Riley TR. Bedside ultrasound in the diagnosis of nonalcoholic fatty liver disease. *World J Gastroenterol* 2014; 20:6821–6825
- Ratziu V, Bellentani S, Cortez-Pinto H, Day C, Marchesini G. A position statement on NAFLD/NASH based on the EASL 2009 special conference. *J Hepatol* 2010; 53:372–384
- Saadeh S, Younossi ZM, Remer EM, et al. The utility of radiological imaging in nonalcoholic fatty liver disease. *Gastroenterology* 2002; 123:745–750
- Bohte AE, van Werven JR, Bipat S, Stoker J. 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* 2011; 21:87–97
- Dasarathy S, Dasarathy J, Khiyami A, Joseph R, Lopez R, McCullough AJ. Validity of real time ultrasound in the diagnosis of hepatic steatosis: a prospective study. *J Hepatol* 2009; 51:1061–1067
- Kim SH, Lee JM, Kim JH, et al. Appropriateness of a donor liver with respect to macrosteatosis: application of artificial neural networks to US images—initial experience. *Radiology* 2005; 234:793–803
- Strauss S, Gavish E, Gottlieb P, Katsnelson L. Interobserver and intraobserver variability in the sonographic assessment of fatty liver. *AJR* 2007; 189:[web]W320–W323
- Williamson RM, Perry E, Glancy S, et al. The use of ultrasound to diagnose hepatic steatosis in type 2 diabetes: intra- and interobserver variability and comparison with magnetic resonance spectroscopy. *Clin Radiol* 2011; 66:434–439
- Lin SC, Heba E, Wolfson T, et al. Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat using a new quantitative ultrasound technique. *Clin Gastroenterol Hepatol* 2015; 13:1337–1345
- Yao LX, Zagzebski JA, Madsen EL. Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors. *Ultrason Imaging* 1990; 12:58–70
- Adams LA, Lymp JF, St Sauver J, et al. The natural history of nonalcoholic fatty liver disease: a population-based cohort study. *Gastroenterology* 2005; 129:113–121
- 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; 267:422–431
- Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005; 41:1313–1321
- Brunke SS, Insana MF, Dahl JJ, Hansen C, Ashfaq M, Ermert H. An ultrasound research interface for a clinical system. *IEEE Trans Ultrason Ferroelectr Freq Control* 2007; 54:198–210
- Ballestri S, Lonardo A, Romagnoli D, et al. Ultrasonographic fatty liver indicator, a novel score which rules out NASH and is correlated with metabolic parameters in NAFLD. *Liver Int* 2012; 32:1242–1252
- Yokoo T, Shieh-morteza 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
- Hamilton G, Yokoo T, Bydder M, et al. In vivo characterization of the liver fat ¹H MR spectrum. *NMR Biomed* 2011; 24:784–790
- Ricci C, Longo R, Gioulis E, et al. Noninvasive in vivo quantitative assessment of fat content in human liver. *J Hepatol* 1997; 27:108–113
- Chan WK, Nik Mustapha NR, Mahadeva S. Controlled attenuation parameter for the detection and quantification of hepatic steatosis in nonalcoholic fatty liver disease. *J Gastroenterol Hepatol* 2014; 29:1470–1476
- Chon YE, Jung KS, Kim SU, et al. Controlled attenuation parameter (CAP) for detection of hepatic steatosis in patients with chronic liver diseases: a prospective study of a native Korean population. *Liver Int* 2014; 34:102–109
- Kumar M, Rastogi A, Singh T, et al. Controlled attenuation parameter for non-invasive assessment of hepatic steatosis: does etiology affect performance? *J Gastroenterol Hepatol* 2013; 28:1194–1201
- Shen F, Zheng RD, Shi JP, et al. Impact of skin capsular distance on the performance of controlled attenuation parameter in patients with chronic liver disease. *Liver Int* 2015; 35:2392–2400