European Radiology (2019) 29:4699–4708
https://doi.org/10.1007/s00330-019-06035-9

ULTRASOUND

Inter-platform reproducibility of ultrasonic attenuation and backscatter coefficients in assessing NAFLD

Aiguo Han[1] · Yingzhen N. Zhang[2] · Andrew S. Boehringer[2] · Michael P. Andre[3] · John W. Erdman Jr.[4] · Rohit Loomba[5] · Claude B. Sirlin[2] · William D. O'Brien Jr.[1]

Received: 2 December 2018 / Revised: 24 December 2018 / Accepted: 22 January 2019 / Published online: 19 February 2019

Abstract

Objectives To assess inter-platform reproducibility of ultrasonic attenuation coefficient (AC) and backscatter coefficient (BSC) estimates in adults with known/suspected nonalcoholic fatty liver disease (NAFLD).

Methods This HIPAA-compliant prospective study was approved by an institutional review board; informed consent was obtained. Participants with known/suspected NAFLD were recruited and underwent same-day liver examinations with clinical ultrasound scanner platforms from two manufacturers. Each participant was scanned by the same trained sonographer who performed multiple data acquisitions in the right liver lobe using a lateral intercostal approach. Each data acquisition recorded a B-mode image and the underlying radio frequency (RF) data. AC and BSC were calculated using the reference phantom method. Inter-platform reproducibility was evaluated for AC and log-transformed BSC ($\log BSC = 10\log_{10}BSC$) by intraclass correlation coefficient (ICC), Pearson's correlation, Bland-Altman analysis with computation of limits of agreement (LOAs), and within-subject coefficient of variation (wCV; applicable to AC).

Results Sixty-four participants were enrolled. Mean AC values measured using the two platforms were 0.90 ± 0.13 and 0.94 ± 0.15 dB/cm/MHz while mean logBSC values were -30.6 ± 5.0 and -27.9 ± 5.6 dB, respectively. Inter-platform ICC was 0.77 for AC and 0.70 for log-transformed BSC in terms of absolute agreement. Pearson's correlation coefficient was 0.81 for AC and 0.80 for logBSC. Ninety-five percent LOAs were -0.21 to 0.13 dB/cm/MHz for AC, and -9.48 to 3.98 dB for logBSC. The wCV was 7% for AC.

Conclusions Hepatic AC and BSC are reproducible across two different ultrasound platforms in adults with known or suspected NAFLD.

Key Points

- *Ultrasonic attenuation coefficient and backscatter coefficient are reproducible between two different ultrasound platforms in adults with NAFLD.*
- *This inter-platform reproducibility may qualify quantitative ultrasound biomarkers for generalized clinical application in patients with suspected/known NAFLD.*

Keywords Nonalcoholic fatty liver disease · Reproducibility of results · Ultrasonography · Prospective studies · Phantoms, imaging

✉ Aiguo Han
han51@illinois.edu

[1] Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 306 North Wright Street, Urbana, IL 61801, USA

[2] Liver Imaging Group, Department of Radiology, University of California, San Diego, 9452 Medical Center Drive, La Jolla, CA 92037, USA

[3] Department of Radiology, University of California, San Diego, 9500 Gilman Drive, San Diego, CA 92093, USA

[4] Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, 905 South Goodwin Avenue, Urbana, IL 61801, USA

[5] NAFLD Research Center, Division of Gastroenterology, Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Abbreviations

AC	Attenuation coefficient
ANOVA	Analysis of variance
AUC	Area under the receiver operating characteristic curve
BMI	Body mass index
BSC	Backscatter coefficient
CAP	Controlled attenuation parameter
CUS	Conventional ultrasonography
FDA	Food and Drug Administration
FOI	Field of interest
HIPAA	Health Insurance Portability and Accountability Act
ICC	Intraclass correlation coefficient
logBSC	Log-transformed backscatter coefficient
LOA	Limit of agreement
MRI	Magnetic resonance imaging
NAFLD	Nonalcoholic fatty liver disease
NASH	Nonalcoholic steatohepatitis
NASH CRN	Nonalcoholic Steatohepatitis Clinical Research Network
PDFF	Proton density fat fraction
QIB	Quantitative imaging biomarker
QUS	Quantitative ultrasound
RF	Radio frequency
wCV	Within-subject coefficient of variation

Introduction

Nonalcoholic fatty liver disease (NAFLD) is emerging as the most common type of chronic liver disease in many parts of the world, affecting ~25% of the population globally [1]. NAFLD comprises a spectrum of liver pathologies ranging from simple steatosis to nonalcoholic steatohepatitis (NASH). Its mildest form, steatosis, is the accumulation of fat droplets within hepatocytes. The more severe form, NASH, is characterized by the presence of steatosis, ballooning degeneration, and lobular inflammation, typically accompanied by pericellular fibrosis [2]. NASH may progress to cirrhosis, liver failure, and hepatocellular carcinoma [1, 2]. Early detection and treatment may halt or reverse NAFLD disease progression [2]. Liver biopsy, costly and invasive, remains the current clinical reference standard to diagnose and grade hepatic steatosis in NAFLD.

Several imaging modalities have been used to noninvasively diagnose and grade hepatic steatosis [3–6]. Conventional ultrasonography (CUS), the most commonly used imaging modality to diagnose and grade steatosis, is subjective, operator, and machine dependent, and inaccurate [3]. Quantitative ultrasound (QUS) techniques, which quantify the ultrasound signals to remove the operator and machine dependence, show promise for objective and accurate hepatic fat quantification [7–9]. Two fundamental quantitative imaging biomarkers (QIBs) derived by QUS techniques, the attenuation coefficient (AC, dB/cm/MHz) and backscatter coefficient (BSC, 1/cm/sr), have been shown to be correlated to hepatic fat fraction [7–9]. AC is an objective measure of the spatial rate of ultrasound energy loss in tissue, which is sensitive to the tissue composition, and BSC is an objective measure of the fraction of ultrasound energy returned from tissue, which is sensitive to the tissue microstructure [10]. Other quantitative ultrasound parameters, such as controlled attenuation parameter (CAP) [11, 12] and sound speed [13], have also been used for hepatic fat quantification.

The accuracy of AC and BSC for hepatic fat content assessment has been demonstrated in the literature [8, 9]. Lin et al showed that BSC correlated with magnetic resonance imaging proton density fat fraction (MRI-PDFF) (Spearman $\rho = 0.80$) in a prospective, cross-sectional study of a cohort of 204 adults with and without NAFLD [8]. The BSC achieved an area under the receiver operating characteristic curve (AUC) of 0.98 for identifying patients with NALFD, using MRI-PDFF as a reference [8]. Paige et al compared the diagnostic performance of CUS, QUS, and MRI-PDFF for predicting histology-confirmed steatosis grade in a cohort of 61 adults with NAFLD [9]. CUS achieved a grading accuracy of 51.7%, whereas AC, BSC, and MRI-PDFF had cross-validated grading accuracies of 55.0%, 68.3%, and 71.3%, respectively [9]. For differentiating between steatosis grades 1 and ≥ 2, the AUCs were 0.79, 0.85, and 0.96 for AC, BSC, and MRI-PDFF, respectively [9].

However, no prior studies have evaluated the reproducibility of the QUS biomarkers measured from multiple platforms from different scanner vendors. Cross-platform reproducibility is essential for the generalization of the QUS method. If the same patient examined with multiple platforms yields reproducible QUS biomarker values, then the published accuracy evaluated on single machines will be generalizable. If not, the applicability of the published accuracy values will be more limited. Indeed, repeatability and reproducibility are important precision measures for any QIB. Repeatability is "the measurement precision with conditions that remain unchanged between replicate measurements (repeatability conditions)" [14]. Reproducibility is "the measurement precision with conditions that vary between replicate measurements (reproducibility conditions)" [14]. Previous studies have assessed the repeatability, inter-sonographer reproducibility, and inter-transducer reproducibility for AC and BSC in phantoms [15] and in human subjects [16, 17].

The purpose of this study was to assess inter-platform reproducibility of AC and BSC in a cohort of prospectively recruited adults with known or suspected NAFLD and with variable degrees of liver steatosis and fibrosis, scanned by the same set of trained expert sonographers using two different ultrasound imaging platforms.

Materials and methods

Study design and participants

An institutional review board approved this Health Insurance Portability and Accountability Act (HIPAA)-compliant study at the University of California, San Diego (UCSD). Written informed consent was obtained. Research participants were prospectively recruited from the UCSD NAFLD Research Center between March 2017 and June 2018 by the hepatologist (RL). Inclusion criteria were age ≥ 18 years, known or suspected NAFLD, and willingness and ablility to participate. Exclusion criteria were clinical, laboratory, or histology evidence of a liver disease other than NAFLD, excessive alcohol consumption (≥ 14 (men) or ≥ 7 (women) drinks/week), and steatogenic or hepatoxic medication use. Demographic, anthropometric, and biochemical data were recorded.

To help characterize the participant cohort, data from contemporaneous hepatic MRI research studies and/or from clinical care liver biopsies were recorded if available. These data included the proton density fat fraction measured by confounder-corrected chemical-shift-encoded MRI (MRI-PDFF) [8] and the histological steatosis grade and fibrosis stage determined by an expert hepatopathologist according to the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) histological scoring system [18].

Ultrasound data acquisition

We used two clinical ultrasound systems, Siemens S3000 (Siemens Healthineers) with the 4C1 transducer (1–4 MHz nominal) and GE Logiq E9 (GE Healthcare) with the C1-6 transducer (1–6 MHz nominal), for direct post-beamformed radio frequency (RF) data acquisition provided under research agreements. Three registered diagnostic medical sonographers (each with > 10 years of experience overall) were trained and had several months to 2 years' experience performing the research protocol. Each participant was scanned by one of the three sonographers selected based on scheduling availability and underwent two same-day 15- to 20-min exams. Each exam was performed using a different platform in random order. Between exams, participants took a 5- to 10-min break, with many choosing to leave the table, then be repositioned on the gurney

Each sonographer performed a standardized research protocol in the right liver lobe using a lateral intercostal approach. For each platform, the system settings were adjusted for each participant to optimize right hepatic lobe visualization prior to the first RF acquisition, but remained constant for all subsequent RF acquisitions. An acquisition was a single operator button press that recorded a B-mode image and the RF data corresponding to the B-mode image. Acquisitions were repeated during separate shallow expiration breath holds separated by about 15 s until at least ten acquisitions were obtained. The skin-to-capsule distance was measured with the image caliper tool and recorded by the sonographer. Following completion of the repeated liver acquisitions on each platform, a calibrated reference phantom (CIRS, Inc.) with known AC and BSC was scanned to obtain RF without changing the system settings.

AC and BSC computation

AC and BSC frequency spectra were computed offline on a desktop personal computer using custom software programmed in MATLAB (The MathWorks) [15]. The software implemented established AC and BSC methodologies designed to remove instrumentation/setting dependencies by comparing the RF data from the liver and the calibrated phantom [19]. AC and BSC were computed within a freehand field of interest (FOI) outlining the liver boundary, drawn on a B-mode image reconstructed from the RF data (Fig. 1). The FOIs were drawn under the supervision of an abdominal radiologist (CBS) by a research analyst (ASB) with 1-year experience in abdominal ultrasound research. A medical physicist (MPA) with career experience in medical ultrasound and MRI research and a research physician (YNZ) provided quality control checks on the work of the analyst by reviewing the FOIs and making corrections when necessary. To minimize analysis burden and in anticipation of possible future clinical applications of this technology, no effort was made to exclude hepatic vessels from the FOI. Five of the ten or more acquisitions were randomly chosen by the analyst for FOI drawing, excluding those that generate B-mode images showing blurring caused by participant breathing or artifacts from rib shadowing. The AC and BSC were computed automatically in the custom software by an investigator (AH, a biomedical engineer with 9 years' experience in QUS research) independent from the group who provided the FOIs. The AC values and separately the BSC values at all frequency points within 2.6–3.0 MHz were averaged to yield a single AC and a single BSC measure per data acquisition. This frequency range was chosen to be consistent with previous repeatability and reproducibility studies of AC and BSC [16, 17] at the center of the transducer bandwidth.

Statistical analysis

Statistical analysis was performed using R 3.4.2 (R Foundation for Statistical Computing). Participant characteristics were summarized descriptively. BSC was log-transformed ($\mathrm{logBSC} = 10\log_{10}\mathrm{BSC}$) to normalize the distribution.

The inter-platform reproducibility was assessed graphically using boxplots for AC and logBSC values obtained before intra-platform averaging of multiple acquisitions of the same

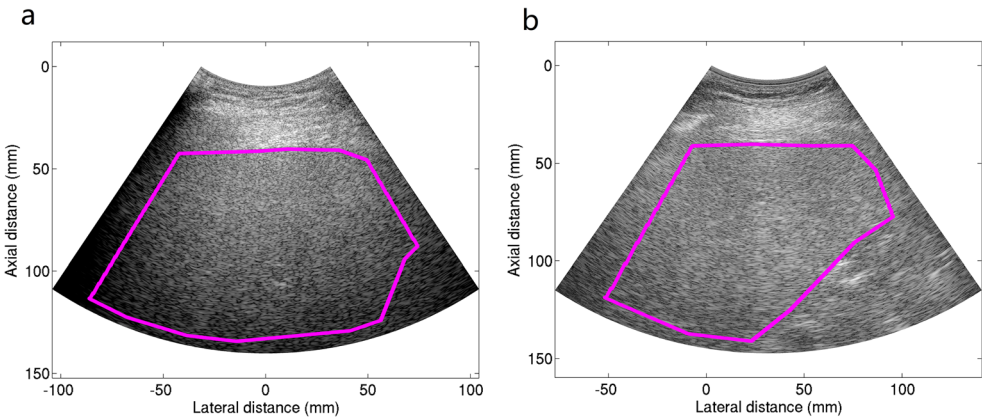

Fig. 1 Representative liver B-mode images reconstructed from the radio frequency data acquired from a 51-year-old male using (**a**) GE Logiq E9 and (**b**) Siemens S3000 clinical ultrasound scanners. The magenta field of interest lines were drawn on the reconstructed B-mode images to outline the liver boundary

participant. Scatter plots and Bland-Altman plots [20] were generated for AC and logBSC values obtained after intra-platform averaging of five acquisitions of the same participant. Pearson's correlation and limits of agreements (LOAs) were calculated accordingly. Inter-platform reproducibility was also assessed by using intraclass correlation coefficient (ICC) [21] and within-subject coefficient of variation (wCV; not applicable to logBSC because of negative values) [22] for AC and/or logBSC values obtained by averaging rising numbers (1 to 5) of repeated acquisitions. ICC values were calculated based on a single-unit two-way mixed effects analysis of variance (ANOVA) model where the participant was treated as a random effect while the platform was treated as a fixed effect. The ICC for absolute agreement was reported. Ninety-five percent confidence intervals (CIs) were computed when applicable. The following definitions were used to interpret the ICC estimates for absolute agreement: 0–0.39, poor; 0.40–0.59, fair; 0.60–0.74, good; and 0.75–1.0, excellent [23].

To test whether the inter-platform variability is affected by the participant body mass index (BMI) and subcutaneous fat, we calculate Pearson's correlation between the absolute between-platform difference in QUS biomarkers and the BMI, and separately the skin-to-capsule distance. For this analysis, the five repeated QUS measurements were averaged to yield a single measure for each platform.

The sample size was driven by feasibility and is typical for reproducibility studies [16, 17, 24–26].

Results

Participants

Sixty-four participants (38 females) were enrolled; the Siemens data from 36 of the 64 participants were published in a previous study [17]. The demographic, physical, biochemical, imaging, and histological characteristics of the study participants are summarized in Table 1. The mean age was 54 (F, 56; M, 50) years, and the age range was 26–74 (F, 26–74; M, 26–70) years. The mean BMI was 32.0 kg/m^2, and the BMI range was 21.7–43.8 kg/m^2. Fifty-three participants had MRI-PDFF within 0 to 110 days (mean, 15 days) of US; mean MRI-PDFF was 13.9%, and the MRI-PDFF range was 0.7–37.7%. Missing MRI scans were due to severe claustrophobia, exceeding bore diameter, refusing research MRI, or the scanner being nonoperational. Fifty-four participants had clinical care liver biopsy within 0 to 302 days (mean, 47 days) of US, 45 of whom also had MRI-PDFF. As indicated by the MRI-PDFF and histological data, the participant cohort of this reproducibility study covered a wide and relevant range of hepatic fat fractions and liver fibrosis ranges.

AC and BSC results

Five AC and five logBSC values were yielded per participant per platform. The statistics of the AC and logBSC results are summarized in Tables 2 and 3, respectively, for AC and logBSC values obtained by averaging rising numbers of intra-platform acquisitions ($N = 1$, 2, 3, 4, and 5). The inter-platform reproducibility was improved with increased number of acquisitions being averaged. When five acquisitions were used, the inter-platform ICC for AC (logBSC) was 0.77 (0.70) in terms of absolute agreement. The wCV was 7.3% for paired (GE vs Siemens) AC measurements. This metric is not applicable to logBSC, which can have negative measurement values.

Scatter plots and Bland-Altman plots are shown for AC averaged from five repeated intra-platform acquisitions (Fig. 2). Similar plots are also shown for logBSC (Fig. 3). There was a significant correlation between the AC values measured from the two platforms with Pearson's $r = 0.81$ (95% CI 0.71–0.88, $p = 0.29 \times 10^{-16}$). Similar results were found for logBSC, with Pearson's $r = 0.80$ (95% CI 0.69–0.87, $p = 0.29 \times 10^{-15}$). The Bland-Altman plots show a slight bias between the two platforms, with a −0.04 dB/cm/MHz mean difference between GE and Siemens AC values, and −2.75 dB mean difference between logBSC values of the

Table 1 Demographic, physical, biochemical, imaging, and histological characteristics of the study participants

	Summary statistics	Total number of participants with data available
Demographics		
Male, no. (%)	26 (40.6)	64
Age (year)[a]	54 ± 13	64
Height (cm)[a]	167.9 ± 11.0	63
Weight (kg)[a]	90.6 ± 17.8	63
BMI (kg/m^2)[a]	32.0 ± 4.7	63
Ethnic origin		
White, no. (%)	34 (53.1)	64
Hispanic, no. (%)	20 (31.3)	64
Asian, no. (%)	9 (14.1)	64
Black, no. (%)	1 (1.6)	64
Biochemical profile[a]		
Hemoglobin (g/dL)	14.3 ± 1.4	63
Hematocrit (%)	41.4 ± 6.2	63
Platelet count (× 10^3/μL)	254 ± 78	63
AST level (U/L)	43.7 ± 19.3	64
ALT level (U/L)	59.1 ± 34.8	64
Alkaline phosphatase level (U/L)	85.1 ± 23.5	63
Total bilirubin level (mg/dL)	0.6 ± 0.3	64
Albumin level (g/dL)	4.4 ± 0.6	64
Hemoglobin A1c (%)	6.2 ± 1.2	62
Triglycerides level (mg/dL)	153.2 ± 71.3	62
Total cholesterol level (mg/dL)	176.3 ± 36.9	62
HDL level (mg/dL)	46.3 ± 13.4	62
LDL level (mg/dL)	99.7 ± 30.0	62
INR	1.04 ± 0.07	64
Imaging[a]		
MRI-PDFF 5–8 (%)	13.9 ± 8.6	53
AC from GE Logiq E9 (dB/cm/MHz)	0.90 ± 0.13	64
AC from Siemens S3000 (dB/cm/MHz)	0.94 ± 0.15	64
logBSC from GE Logiq E9 (dB)	−30.6 ± 5.0	64
logBSC from Siemens S3000 (dB)	−27.9 ± 5.6	64
Biopsy		
Steatosis grade		
S0, no. (%)	5 (9.3)	54
S1, no. (%)	24 (44.4)	54
S2, no. (%)	21 (38.9)	54
S3, no. (%)	4 (7.4)	54
Fibrosis stage		
F0, no. (%)	20 (37.0)	54
F1, no. (%)	21 (38.9)	54
F2, no. (%)	3 (5.6)	54
F3, no. (%)	5 (9.3)	54
F4, no. (%)	5 (9.3)	54

All lipid labs were measured while patients were fasting

AC, attenuation coefficient; *AST*, aspartate aminotransferase; *ALT*, alanine aminotransferase; *logBSC*, log-transformed backscatter coefficient; *BMI*, body mass index; *GGT*, gamma-glutamyl transpeptidase; *HDL*, high-density lipoprotein; *INR*, international normalized ratio; *LDL*, low-density lipoprotein; *MRI*, magnetic resonance imaging; *PDFF*, proton density fat fraction, mean calculated from segments 5 to 8

[a] Mean value provided with standard deviations

two platforms. The 95% limits of agreement range from −0.21 to 0.13 dB/cm/MHz for AC, and from −9.48 to 3.98 dB for logBSC.

The absolute between-platform difference in the AC was not statistically significantly correlated with the BMI (Fig. 4a; Pearson's $r = -0.11$, $p = 0.39$) or the skin-to-capsule distance (Fig. 4c; Pearson's $r = 0.13$, $p = 0.31$). Similarly, the absolute between-platform difference in the logBSC was not statistically significantly correlated with the BMI (Fig. 4b; Pearson's $r = -0.01$, $p = 0.91$)

Table 2 Attenuation coefficient inter-platform reproducibility for measurements based on a single acquisition ($N = 1$) and on the mean of multiple repeated intra-platform acquisitions ($N = 2$, 3, 4, and 5)

Number of acquisitions *N*	GE mean ± SD (dB/cm/MHz)	Siemens mean ± SD (dB/cm/MHz)	ICC (95% CI)	wCV
1	0.89 ± 0.13	0.94 ± 0.15	0.69 (0.50, 0.81)	0.09
2	0.89 ± 0.13	0.94 ± 0.15	0.76 (0.58, 0.86)	0.08
3	0.89 ± 0.13	0.94 ± 0.15	0.75 (0.57, 0.86)	0.08
4	0.90 ± 0.13	0.94 ± 0.15	0.77 (0.58, 0.86)	0.07
5	0.90 ± 0.13	0.94 ± 0.15	0.77 (0.59, 0.87)	0.07

ICC was calculated based on a single-unit, absolute agreement, two-way mixed effect analysis of variance (ANOVA) model

CI, confidence interval; *ICC*, intraclass correlation coefficient; *SD*, standard deviation; *wCV*, within-subject coefficient of variation

or the skin-to-capsule distance (Fig. 4d; Pearson's $r = 0.16$, $p = 0.21$).

Discussion

This study examines an essential aspect of QUS biomarker precision that was previously uninvestigated, the inter-platform reproducibility of AC and BSC. The inter-platform reproducibility is important if common cutoff values of AC and BSC are to be used for clinical diagnosis on different platforms. We demonstrated good to excellent absolute agreement (ICC 0.77 for AC and 0.70 for logBSC) between two clinical ultrasound imaging platforms, GE Logiq E9 and Siemens S3000, in measuring the hepatic AC and BSC values in NAFLD participants with a wide range of hepatic fat fraction and fibrosis stages. We also demonstrated that the inter-platform reproducibility was not affected by the participant BMI or the skin-to-capsule distance, but it was affected by the number of acquisitions. The inter-platform reproducibility was improved with increasing number of acquisitions being used for AC. The improvement was less for BSC, likely because the AC measurement is noisier than the BSC measurement with the current implementation of QUS techniques. Furthermore, small inter-platform biases were found. These biases were smaller than or comparable with the overall measurement standard deviation due to repeatability and inter-transducer reproducibility reported in a previous human study on NALFD participants (AC, ~ 0.07 dB/cm/MHz; logBSC, 2.5–2.9 dB) [16].

This study complemented previous AC and BSC repeatability and reproducibility studies in phantoms and humans, and also enabled us to compare QUS with other imaging modalities in terms of inter-platform reproducibility. A phantom-based study [15] reported excellent overall repeatability and reproducibility in AC and BSC, as demonstrated by the small measurement standard deviation due to repeatability and reproducibility (AC, < 0.02 dB/cm/MHz; logBSC, ~ 0.5 dB). Another phantom study [27] assessed the repeatability and reproducibility of the phantom power spectra (although not AC and BSC) using 11 transducers of the same model (Siemens 6C1) and five clinical ultrasound systems of the same model (Siemens Acuson S3000). It was determined that the data acquired from those transducers and systems produced equivalent phantom power spectrum estimates. A human study [16] on NAFLD participants demonstrated repeatable and inter-transducer reproducible AC and logBSC measures, with small overall measurement standard deviation due to repeatability and inter-transducer reproducibility (AC, ~ 0.07 dB/cm/MHz; logBSC, 2.5–2.9 dB). Another human

Table 3 Log-transformed backscatter coefficient inter-platform reproducibility for measurements based on a single acquisition ($N = 1$) and on the mean of multiple repeated intra-platform acquisitions ($N = 2$, 3, 4, and 5)

Number of acquisitions *N*	GE mean ± SD (dB)	Siemens mean ± SD (dB)	ICC (95% CI)
1	− 30.8 ± 4.9	− 27.8 ± 5.7	0.67 (0.19, 0.85)
2	− 30.7 ± 4.9	− 27.7 ± 5.7	0.68 (0.23, 0.85)
3	− 30.7 ± 5.0	− 27.8 ± 5.7	0.69 (0.24, 0.85)
4	− 30.6 ± 5.0	− 27.8 ± 5.7	0.70 (0.26, 0.86)
5	− 30.6 ± 5.0	− 27.9 ± 5.6	0.70 (0.28, 0.86)

ICC was calculated based on a single-unit, absolute agreement, two-way mixed effect analysis of variance (ANOVA) model

CI, confidence interval; *ICC*, intraclass correlation coefficient; *SD*, standard deviation

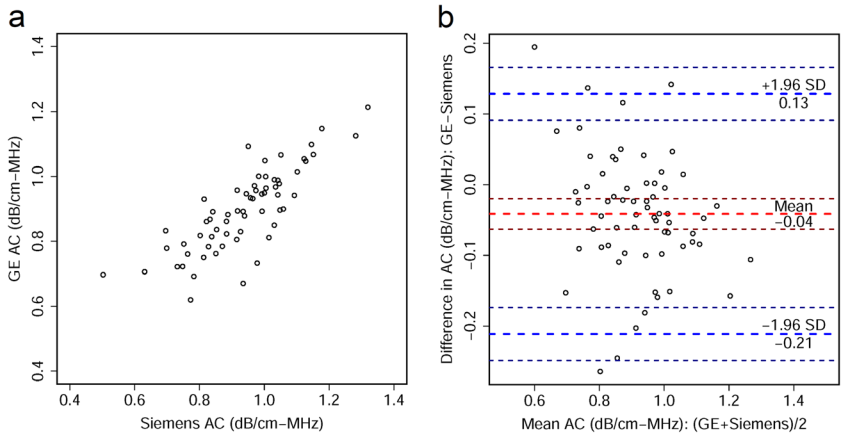

Fig. 2 **a** Scatter plot shows relationship between attenuation coefficient (AC) values measured from GE Logiq E9 and Siemens S3000. **b** Bland-Altman plot shows agreement between AC values measured using the two platforms. Thick red dashed line shows the mean difference in AC values between the two platforms and thick blue dashed lines demarcate ± 1.96 standard deviations (SD), with associated 95% confidence intervals indicated by thin dashed lines. The AC values in both plots were obtained by averaging five intra-platform repeated acquisitions per participant

study on NAFLD participants [17] found AC and logBSC to be reproducible between sonographers, with a single-unit (measurement made by a single sonographer as the final result) absolute agreement inter-sonographer ICC of 0.78 for AC and 0.79 for logBSC, and a double-unit (mean of measurements made by two sonographers as the final result) absolute agreement ICC of 0.88 for both AC and logBSC. Comparing the single-unit absolute agreement ICC values in [17] with those reported in the current study, we found similar ICC values for the AC (inter-sonographer, 0.78; inter-platform, 0.77), but differing ICC values for the logBSC (inter-sonographer, 0.79; inter-platform, 0.70). The current paper did not report the double-unit ICC values because it was unlikely in future clinical practices that two platforms would be used to yield average AC and BSC measures as the final results for a patient. Comparing QUS with other modalities for liver assessment, the reproducibility of QUS biomarkers AC and BSC appeared to be better than that of ultrasound elastography, similar to that of MR elastography, and less than that of MRI-PDFF. In ultrasound elastography, the shear wave speed or liver stiffness measures were not reproducible across manufacturers because the techniques vary between manufacturers, leading to manufacturer-dependent cutoff values for liver fibrosis assessment [28]. In terms of MR elastography, Trout et al assessed the inter-manufacturer (GE and Philips) reproducibility in liver stiffness measures on 24 volunteer adult participants, reporting

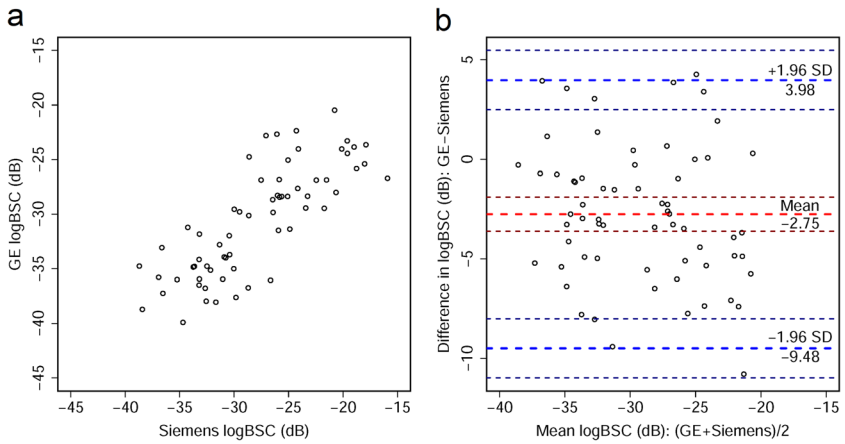

Fig. 3 **a** Scatter plot shows relationship between log-transformed backscatter coefficient (logBSC) values measured from GE Logiq E9 and Siemens S3000. **b** Bland-Altman plot shows agreement between logBSC values measured using the two platforms. Thick red dashed line shows the mean difference in logBSC values between the two platforms and thick blue dashed lines demarcate ± 1.96 standard deviations (SD), with associated 95% confidence intervals indicated by thin dashed lines. The logBSC values in both plots were obtained by averaging five intra-platform repeated acquisitions per participant

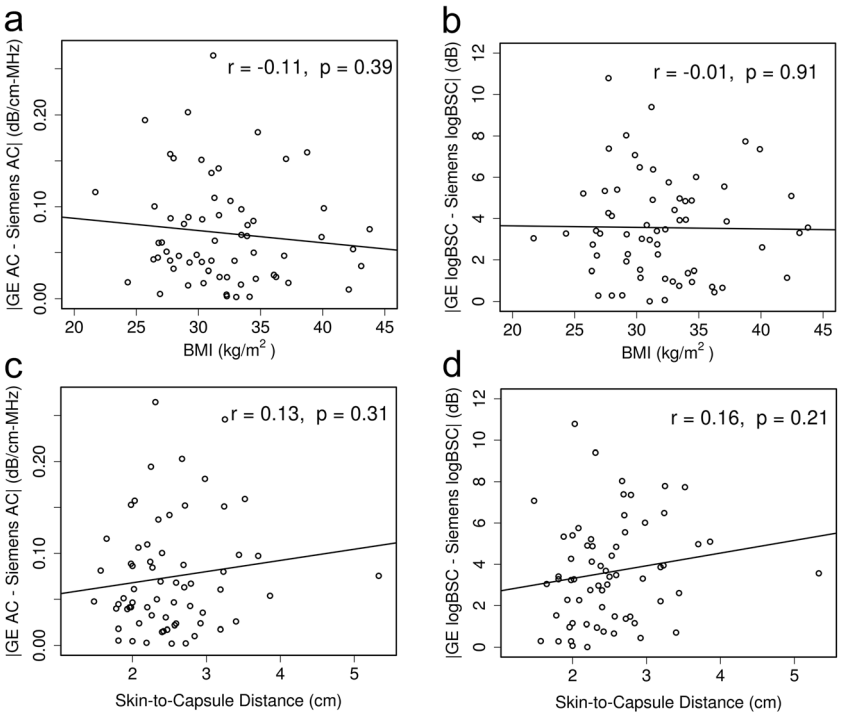

Fig. 4 Absolute between-platform difference in QUS biomarker versus participant condition plots (**a** AC absolute difference versus BMI, **b** logBSC absolute difference versus BMI, **c** AC absolute difference versus skin-to-capsule distance, **d** logBSC absolute difference versus skin-to-capsule distance) show that the QUS inter-platform variability is not significantly affected by the participant BMI or subcutaneous fat. The corresponding Pearson's correlation and *p* value are displayed in each subfigure

absolute agreement inter-manufacturer ICC values of 0.82, 0.71, 0.67, and 0.69, respectively, for four imaging conditions: 1.5-T 2D gradient-echo, 3.0-T 2D gradient-echo, 1.5-T 2D spin-echo echo-planar imaging, and 3.0-T 2D spin-echo echo-planar imaging [24]. The inter-platform reproducibility of MRE in patients with liver disease is not yet well understood, however. For MRI-PDFF, Serai et al reported absolute agreement inter-manufacturer ICC values ranging from 0.91 to 0.95 depending on the readers, derived from a study of 24 adult volunteers scanned with MR imagers from two vendors, GE and Philips [29]. Mashhood et al reported ICC of 0.85–0.91 for reproducibility between five imaging centers using three different manufacturers and three different FF methods [30]. Kang et al assessed the reproducibility of MRI-determined PDFF across two different MR scanner platforms, Siemens at 1.5 T and GE at 3 T, and reported that the MRI-determined PDFF differences at 1.5 T compared with 3 T ranged from − 3.2 to + 4.6%, while the limits for the interval defined by the mean difference ± 1.96 standard deviations were − 1.9 and + 3.7% [31]. In a multi-site, multi-vendor fat-water phantom study, Hernando et al reported significant effect of vendor with a bias of − 0.37% (Philips) and − 1.22% (Siemens) relative to GE Healthcare [32]. Overall ICC across sites, vendors, protocols, and field strengths was 0.999 in that phantom study [32].

There were a few limitations in the current study. This was a single-site study, and we only evaluated two vendor platforms. Also, inter-platform reproducibility was assessed only on the AC and BSC values, and not on the clinical diagnostic outcomes (e.g., predicted fat fraction and NAFLD diagnosis). Future studies may evaluate the reproducibility using more platforms at multiple sites and evaluate the reproducibility of clinical diagnostic outcomes. Furthermore, future efforts may be directed to develop QUS techniques for better repeatability and reproducibility.

In conclusion, our study showed good to excellent inter-platform reproducibility of AC and BSC in a cohort of adult participants spanning a wide and relevant range of hepatic steatosis and fibrosis and scanned by the same set of trained expert sonographers using the ultrasound scanners from two manufacturers. Additional research is needed to evaluate the impact of different imaging platforms on diagnostic performances.

Acknowledgements The authors thank the research participants for making this study possible, the sonographers, Elise Housman, Susan Lynch, and Minaxi Trivedi, for the dedicated contributions and expertise, the clinical coordinator Vivian Montes for her outstanding organization of the many moving parts, and the pathologist Mark A. Valasek, MD, PhD, for reading the histology and determining the steatosis grade and fibrosis stage.

Funding This study has received funding by the National Institutes of Health (R01DK106419), Siemens Healthineers USA, and GE Healthcare.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Claude B. Sirlin, MD (University of California, San Diego).

Conflict of interest The authors of this manuscript declare relationships with the following companies:

The work is supported in part by research grants from Siemens Healthineers USA and GE Healthcare. The use of the Siemens S3000 scanner was loaned to the University of California, San Diego under a research agreement with Siemens Healthineers, USA. The use of the GE Logiq E9 scanner was loaned to the University of California, San Diego under a research agreement with GE Healthcare.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohorts overlap Some study subjects or cohorts have been previously reported in [17].

Methodology
- prospective
- cross-sectional study
- performed at one institution

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Loomba R, Sanyal AJ (2013) The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol 10:686–690
2. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ (2018) Mechanisms of NAFLD development and therapeutic strategies. Nat Med 24:908–922
3. Machado MV, Cortez-Pinto H (2013) Non-invasive diagnosis of non-alcoholic fatty liver disease. A critical appraisal. J Hepatol 58:1007–1019
4. Park CC, Nguyen P, Hernandez C et al (2017) 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
5. Reeder SB, Cruite I, Hamilton G, Sirlin CB (2011) Quantitative assessment of liver fat with magnetic resonance imaging and spectroscopy. J Magn Reson Imaging 34:729–749
6. Artz NS, Hines CD, Brunner ST et al (2012) Quantification of hepatic steatosis with dual-energy computed tomography: comparison with tissue reference standards and quantitative magnetic resonance imaging in the ob/ob mouse. Invest Radiol 47:603–610
7. Andre MP, Han A, Heba E et al (2014) Accurate diagnosis of nonalcoholic fatty liver disease in human participants via quantitative ultrasound. In: 2014 IEEE International Ultrasonics Symposium. pp 2375–2377
8. 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 13:1337–1345.e6
9. Paige JS, Bernstein GS, Heba E et al (2017) A pilot comparative study of quantitative ultrasound, conventional ultrasonography, and magnetic resonance imaging for predicting histology-determined steatosis grade in adult nonalcoholic fatty liver disease. AJR Am J Roentgenol 208:W1–W10
10. Oelze ML, Mamou J (2016) Review of quantitative ultrasound: envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound. IEEE Trans Ultrason Ferroelectr Freq Control 63:336–351
11. Sasso M, Beaugrand M, de Ledinghen V et al (2010) Controlled attenuation parameter (CAP): a novel VCTE™ guided ultrasonic attenuation measurement for the evaluation of hepatic steatosis: preliminary study and validation in a cohort of patients with chronic liver disease from various causes. Ultrasound Med Biol 36:1825–1835
12. Caussy C, Alquiraish MH, Nguyen P et al (2018) Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. Hepatology 67: 1348–1359
13. Imbault M, Faccinetto A, Osmanski BF et al (2017) Robust sound speed estimation for ultrasound-based hepatic steatosis assessment. Phys Med Biol 62:3582–3598
14. Sullivan DC, Obuchowski NA, Kessler LG et al (2015) Metrology standards for quantitative imaging biomarkers. Radiology 277: 813–825
15. Han A, Andre MP, Erdman JW, Loomba R, Sirlin CB, O'Brien WD (2017) Repeatability and reproducibility of a clinically based QUS phantom study and methodologies. IEEE Trans Ultrason Ferroelectr Freq Control 64:218–231
16. Han A, Andre MP, Deiranieh L et al (2018) Repeatability and reproducibility of ultrasonic attenuation coefficient and backscatter coefficient measured in the right lobe of the liver in adults with known or suspected nonalcoholic fatty liver disease. J Ultrasound Med 37:1913–1927
17. Han A, Labyed Y, Sy EZ et al (2018) Inter-sonographer reproducibility of quantitative ultrasound outcomes and shear wave speed measured in the right lobe of the liver in adults with known or suspected nonalcoholic fatty liver disease. Eur Radiol 28:4992–5000
18. Kleiner DE, Brunt EM, Van Natta M et al (2005) Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 41:1313–1321
19. Yao LX, Zagzebski JA, Madsen EL (1990) Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors. Ultrason Imaging 12:58–70
20. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310
21. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428
22. Raunig DL, McShane LM, Pennello G et al (2015) Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. Stat Methods Med Res 24:27–67
23. Fleiss JL (1999) The design and analysis of clinical experiments. Wiley, New York, pp 1–32
24. Trout AT, Serai S, Mahley AD et al (2016) Liver stiffness measurements with MR elastography: agreement and repeatability across imaging systems, field strengths, and pulse sequences. Radiology 281:793–804
25. Nobili V, Vizzutti F, Arena U et al (2008) Accuracy and reproducibility of transient elastography for the diagnosis of fibrosis in pediatric nonalcoholic steatohepatitis. Hepatology 48:442–448
26. Bota S, Sporea I, Sirli R, Popescu A, Danila M, Costachescu D (2012) Intra- and interoperator reproducibility of acoustic radiation force impulse (ARFI) elastography–preliminary results. Ultrasound Med Biol 38:1103–1108
27. Guerrero QW, Fan L, Brunke S, Milkowski A, Rosado-Mendez IM, Hall TJ (2018) Power spectrum consistency among systems and

transducers. Ultrasound Med Biol, online. https://doi.org/10.1016/j.ultrasmedbio.2018.05.013
28. Ferraioli G, Filice C, Castera L et al (2015) WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 3: liver. Ultrasound Med Biol 41:1161–1179
29. Serai SD, Dillman JR, Trout AT (2017) Proton density fat fraction measurements at 1.5- and 3-T hepatic MR imaging: same-day agreement among readers and across two imager manufacturers. Radiology 284:244–254
30. Mashhood A, Railkar R, Yokoo T et al (2013) Reproducibility of hepatic fat fraction measurement by magnetic resonance imaging. J Magn Reson Imaging 37:1359–1370
31. Kang GH, Cruite I, Shiehmorteza M et al (2011) Reproducibility of MRI-determined proton density fat fraction across two different MR scanner platforms. J Magn Reson Imaging 34:928–934
32. Hernando D, Sharma SD, Aliyari Ghasabeh M et al (2017) Multi-site, multi-vendor validation of the accuracy and reproducibility of proton-density fat- fraction quantification at 1.5T and 3T using a fat-water phantom. Magn Reson Med 77:1516–1524