



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 non-alcoholic fatty liver disease

Aiguo Han¹ · Yassin Labyed² · Ethan Z. Sy³ · Andrew S. Boehringer³ · Michael P. Andre^{4,5} · John W. Erdman Jr⁶ · Rohit Loomba⁷ · Claude B. Sirlin³ · William D. O'Brien Jr¹

Received: 20 March 2018 / Revised: 27 April 2018 / Accepted: 14 May 2018 / Published online: 4 June 2018

© European Society of Radiology 2018

Abstract

Objectives To assess inter-sonographer reproducibility of ultrasound attenuation coefficient (AC), backscatter coefficient (BSC) and shear wave speed (SWS) in adults with known/suspected non-alcoholic fatty liver disease (NAFLD).

Methods The institutional review board approved this HIPAA-compliant prospective study; informed consent was obtained. Participants with known/suspected NAFLD were recruited and underwent same-day liver examinations with a clinical scanner. Each participant was scanned by two of the six trained sonographers. Each sonographer performed multiple data acquisitions in the right liver lobe using a lateral intercostal approach. A data acquisition was a single operator button press that recorded a B-mode image, radio-frequency data, and the SWS value. AC and BSC were calculated from the radio-frequency data using the reference phantom method. SWS was calculated automatically using product software. Intraclass correlation coefficient (ICC) and within-subject coefficient of variation (wCV) were calculated for applicable metrics.

Results Sixty-one participants were recruited. Inter-sonographer ICC was 0.86 (95% confidence interval: 0.77–0.92) for AC and 0.87 (0.78–0.92) for log-transformed BSC ($\log_{10}BSC = 10\log_{10}BSC$) using one acquisition per sonographer. ICC was 0.88 (0.80–0.93) for both AC and logBSC averaging 5 acquisitions. ICC for SWS was 0.57 (0.29–0.74) using one acquisition per sonographer, and 0.84 (0.66–0.93) using 10 acquisitions. The wCV was ~7% for AC, and 19–43% for SWS, depending on number of acquisitions.

Conclusions Hepatic AC, BSC and SWS measures on a clinical scanner have good inter-sonographer reproducibility in adults with known or suspected NAFLD. Multiple acquisitions are required for SWS but not AC or BSC to achieve good inter-sonographer reproducibility.

Key Points

- AC, BSC and SWS measurements are reproducible in adults with NAFLD.
- Inter-sonographer reproducibility of SWS measurement improves with more acquisitions being averaged.
- Multiple acquisitions are required for SWS but not AC or BSC.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-018-5541-9>) contains supplementary material, which is available to authorized users.

✉ William D. O'Brien, Jr
wdo@uiuc.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 Siemens Healthineers USA, 22010 South East 51st Street, Issaquah, WA 98029, USA

³ Liver Imaging Group, Department of Radiology, University of California at San Diego, 9452 Medical Center Drive, La Jolla, CA 92037, USA

⁴ Department of Radiology, University of California at San Diego, 9500 Gilman Drive, San Diego, CA 92093, USA

⁵ San Diego VA Healthcare System, San Diego, CA, USA

⁶ Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, 905 South Goodwin Avenue, Urbana, IL 61801, USA

⁷ NAFLD Research Center, Division of Gastroenterology, Department of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Keywords Non-alcoholic fatty liver disease · Liver fibrosis · Reproducibility of results · Ultrasonography · Elastography

Abbreviations

AC	Attenuation coefficient
ANOVA	Analysis of variance
BMI	Body mass index
BSC	Backscatter coefficient
CAP	Controlled attenuation parameter
FOI	Field of interest
ICC	Intraclass correlation coefficient
logBSC	log-transformed backscatter coefficient
MRI	Magnetic resonance imaging
NAFLD	Non-alcoholic fatty liver disease
NASH CRN	Nonalcoholic Steatohepatitis Clinical Research Network
PDFF	Proton density fat fraction
QIB	Quantitative imagine biomarker
QIBA	Quantitative Imaging Biomarker Alliance
QUS	Quantitative ultrasound
RF	Radio-frequency
ROI	Region of interest
RSNA	Radiological Society of North America
SWS	Shear wave speed
USFDA	United States Food and Drug Administration
VTQ	Virtual touch quantification
wCV	within-subject coefficient of variation

Introduction

Advances in medical imaging technology have allowed quantitative information to be obtained from medical ultrasound for disease diagnosis and staging. Quantitative ultrasound (QUS) parameters [e.g., attenuation coefficient (AC, dB/cm-MHz) and backscatter coefficient (BSC, 1/cm-sr)] and the elasticity parameter shear wave speed (SWS, m/s) are emerging as clinically useful measurements derived from ultrasound [1–11]. AC is a measure of ultrasound energy loss in tissue, BSC is a measure of ultrasound energy returned from tissue, and SWS is an indicator of tissue stiffness.

Prior studies have shown that QUS and SWS are complementary for noninvasive assessment of non-alcoholic fatty liver disease (NAFLD) [2–8]. Clinical studies have shown that AC and BSC can assess liver fat noninvasively in NAFLD [6–8], while SWS can detect advanced liver fibrosis [2–5]. These parameters can be acquired from a single ultrasound examination, thereby providing contemporary assessment of key disease components (steatosis and fibrosis). Moreover, they can be implemented on standard clinical ultrasound systems, with the potential for widespread clinical utilization for screening or serial assessment if validated. By comparison, the controlled attenuation parameter (CAP) measured by

Fibroscan (Echosens) is used clinically to assess liver steatosis [12, 13], but this method is proprietary, available only on a specialized device from a single manufacturer, and unavailable on most clinical ultrasound systems. Similarly, while magnetic resonance elastography and chemical-shift-encoded magnetic resonance imaging (MRI) are emerging as accurate and reproducible methods for assessing liver steatosis [14] and fibrosis [14, 15] in NAFLD, these methods are expensive, not widely available globally and less practical for screening.

For quantitative imaging biomarkers (QIBs) such as AC, BSC and SWS to play an important clinical role, rigorous assessment of their reliability and reproducibility is needed. In fact, the Quantitative Imaging Biomarker Alliance (QIBA) was organized by the Radiological Society of North America (RSNA) in 2007 “in response to the need for reliable and reproducible quantification of biomedical imaging data” [16].

Previous phantom and human studies have demonstrated excellent repeatability and inter-transducer reproducibility for QUS parameters AC and BSC [17, 18]. However, the inter-sonographer reproducibility for QUS parameters has not yet been evaluated in clinical studies. Previous studies have also assessed the reproducibility of SWS [3, 19, 20]. However, QUS and SWS have not been examined in the same population. We assess inter-sonographer reproducibility of QUS and SWS in a single 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 the same machine and transducer.

Materials and methods

Study design and participants

The institutional review board approved this HIPAA-compliant study. Research participants were prospectively recruited from the University of California at San Diego (UCSD) Research Center between March 2016 and November 2017. Participants aged ≥ 18 years old were recruited by the hepatology investigator (RL), a fellowship-trained hepatologist, if they had known or suspected NAFLD and were willing and able to participate. Patients were excluded if they had clinical, laboratory, or histology evidence of a liver disease other than NAFLD, if they consumed excess alcohol (≥ 14 (men) or ≥ 7 (women) drinks/week), or if they used steatogenic or hepatotoxic medications. Written informed consent was obtained. Demographic and anthropometric data were acquired. Data from contemporaneous hepatic MRI research studies and/or from clinical-care

liver biopsies were recorded if available to help characterize the participant cohort.

Ultrasound data acquisition

We used a clinical ultrasound system (Siemens S3000, Siemens Healthineers) that, under terms of a research agreement, allowed direct post-beamformed radio-frequency (RF) data acquisition. Six registered diagnostic medical sonographers (each with > 10 years of experience) were trained in the research protocol. Each participant was scanned by two of the six sonographers, the two being selected were based on scheduling availability. The 4C1 transducer (1–4 MHz nominal) was used. Between scans, participants were allowed to take a brief break (5–10 minutes), although few chose to get off the gurney; in every case, the participant was repositioned on the gurney for the next scan by the second sonographer.

Each sonographer performed the same protocol in the same location in the right liver lobe using a lateral intercostal approach. Consistent with standard clinical practice, each sonographer adjusted system settings in each participant to optimize right hepatic lobe visualization prior to the first RF/SWS acquisition; settings for that sonographer remained constant for all subsequent RF/SWS acquisitions. The sonographer manually placed a 0.6 (width) x 1-cm (height) rectangular SWS region of interest (ROI) overlaid on the B-mode image to a relative homogeneous region of the right liver lobe at least 2 cm below the liver capsule but not deeper than 7 cm. An acquisition was defined as a single operator button press that recorded a B-mode image, the RF data corresponding to the entire B-mode image, then the SWS value. Following standard SWS protocol, acquisitions were repeated during separate shallow expiration breath holds separated by about 15 seconds until 10 valid acquisitions were obtained, defined to be one in which the SWS value was displayed, or until 20 attempts were made, whichever came first. A calibrated reference phantom (CIRS, Inc.) was scanned to obtain RF data following completion of the repeated liver acquisitions without changing the settings.

AC and BSC computation

AC and BSC frequency spectra were derived using established methodologies that removed instrumentation/setting dependencies by comparing the liver and phantom RF data [17, 21]. Briefly, the RF data were processed offline on a desktop personal computer using custom software programmed in MATLAB (The MathWorks). The software first performed a quality control test by automatically checking if the system settings for the liver and phantom acquisitions were identical as intended. The liver

acquisitions that were found to have different system settings than the corresponding phantom acquisition were considered invalid for QUS analysis. Afterwards, a free-hand field of interest (FOI) outlining the liver boundary was drawn on each B-mode reconstructed from the RF data (Fig. 1). The FOIs were different from the SWS ROIs because of inherent technical differences between QUS and SWS techniques. The FOIs covered an area of 75 cm² on average, although the area ranged from 40 to 120 cm² depending on how much of the liver is visualized. FOIs were drawn under the supervision of an abdominal radiologist (CBS) by two research analysts (Ezs & ASB) with 2–3 years of experience in radiology body imaging research and by a medical physicist (MPA) with career experience in medical ultrasound and MRI research, who also provided quality control checks on the work of the two analysts by reviewing their FOIs and making corrections when necessary. To minimize analysis burden and in anticipation for possible future clinical applications of this technology, no effort was made to exclude hepatic vessels from the FOI, and only five of the 10–20 valid acquisitions were randomly chosen for B-mode reconstruction and FOI drawing. The AC and BSC from the FOI were computed automatically in the custom software using an established reference phantom technique [17, 21] by an investigator 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 image. This bandwidth was chosen because it has a narrow range around the 2.8-MHz center frequency with optimal signal-to-noise ratio [18].

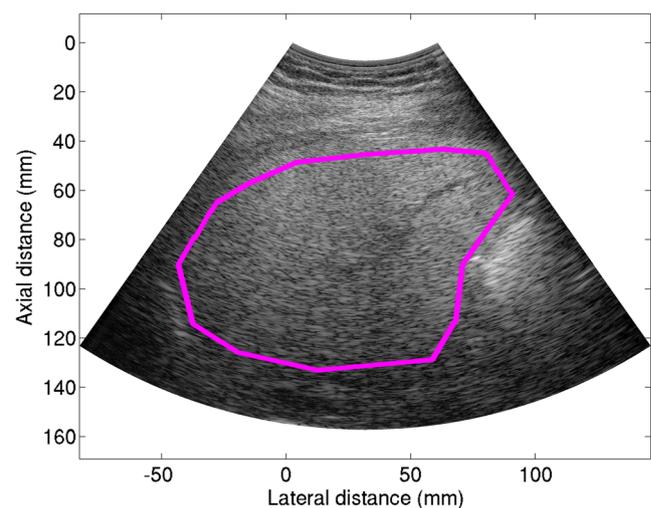


Fig. 1 (Reproduced from [18] with permission) A representative liver B-mode image reconstructed from the radio-frequency data acquired from a 32-year-old male. The pink field of interest line was drawn on the reconstructed B-mode image to outline the liver boundary

SWS computation

A United States Food and Drug Administration (USFDA)-cleared Siemens-developed Virtual Touch Quantification (VTQ)[®] algorithm was used for SWS calculation. The SWS values in m/sec, if valid by algorithm criteria, were displayed automatically by the scanner after each acquisition and recorded.

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 ($\log BSC = 10 \log_{10} BSC$) to normalize the distribution.

The inter-sonographer reproducibility was assessed using intraclass correlation coefficient (ICC) [22] and within-subject coefficient of variance (wCV; not applicable for logBSC because of negative values) [23]. ICC was calculated using the ‘irr’ package based on an absolute-agreement, one-way random effect analysis of variance (ANOVA) model. This model is appropriate for our study design wherein each participant was scanned by a (possibly) different set of two sonographers selected based on availability from a pool of six sonographers. Two ICC forms were reported: ICC(1,1) and ICC(1,k), estimated based on the ‘single’ and ‘average’ units specified in the ‘irr’ package, respectively. ICC(1,1) is defined as

$$ICC(1, 1) = \frac{\text{participant variability}}{\text{participant variability} + \text{sonographer variability} + \text{measurement error}},$$

which represents the measurement reliability if the measurement from a single sonographer is used as the basis of actual measurement. ICC(1,k) is defined as

$$ICC(1, k) = \frac{\text{participant variability}}{\text{participant variability} + (\text{sonographer variability} + \text{measurement error})/k},$$

which represents the measurement reliability if the mean value of k sonographers ($k=2$) is used as an assessment basis.

To compare QUS and SWS for reproducibility performance, the measurement from each sonographer on a participant was calculated by taking the mean of N acquisitions. To investigate if the ICC estimates were affected by the choice of taking the mean versus the median of N acquisitions, the SWS measurement of each sonographer was also determined separately by the median of N acquisitions. ICC values were calculated for $N = 1, 2, \dots, 10$ for SWS and for $N = 1, 2, \dots, 5$ for AC and logBSC to investigate the influence of N on ICC values. Confidence intervals of 95% (95% CI) were computed when applicable.

The sample size was driven by feasibility. Our sample size is typical for reproducibility studies [3, 19, 24, 25].

Reference standards

Contemporaneous MRI-derived proton density fat fraction (MRI-PDFF) [6] and histological steatosis scores determined by a liver pathologist according to the Nonalcoholic Steatohepatitis Clinical Research Network (NASH CRN) histological scoring system [26] were used as the reference

standards to characterize the steatosis ranges of the participant cohort.

The histological fibrosis stages defined by the NASH CRN histological scoring system were used to characterize the fibrosis ranges of the participant cohort.

By definition, there is no reference standard for the reproducibility study.

Results

Participants

Sixty-one participants (43 females) were recruited. The mean age was 53 (F: 56; M: 47) years, and the age range was 25–74 (F: 26–72; M: 25–74) years. The mean body mass index (BMI) was 32.0 kg/m² and the BMI range was 23.9–43.3 kg/m². Forty-six participants had MRI-PDFF within 0 to 113 days (mean: 17 days) of US; mean MRI-PDFF was 15.8%, and the MRI-PDFF range was 0.7–41.1%. Some participants did not have MRI due to severe claustrophobia, exceeding bore diameter, refusing research MRI, or the scanner being down. Fifty-eight participants had clinical-care liver biopsy within 1 to 258



Fig. 2 Boxplots of measured attenuation coefficient values grouped by participant-sonographer pairs. Each participant is represented by two adjacent boxes (two sonographers) of the same colour. Left of the red vertical line are participants for which MRI-proton density fat fraction

(MRI-PDFF) are available and those participants are ordered by MRI-PDFF. Right of the red vertical line are participants for which MRI-PDFF are unavailable

days (mean: 55 days) of US and with the following distribution of histology-determined steatosis scores: 0: 2, 1: 25, 2: 26, and 3: 5, and the following distribution of histology-determined fibrosis stages: 0: 21, 1: 23, 2: 8, 3: 3, and 4: 3. The MRI-PDFF and histology data show that the participant cohort of this reproducibility study covered a wide range of hepatic fat fractions and liver fibrosis ranges (these data are not intended to be used to assess the diagnostic accuracy of QUS and SWS parameters).

AC and BSC results

Five AC values and five logBSC values were obtained per sonographer per participant. The AC and logBSC boxplots (Figs. 2 and 3, respectively) were generated such that the values were grouped by participant-sonographer pairs and

ordered by MRI-PDFF when available (left of the red vertical line in the boxplot) to provide an overview of the distribution and variability of the acquisition values without intra-sonographer averaging.

The ICC (and 95% CI) estimates of AC and logBSC are shown in Tables 1 and 2, respectively. ICC(1,1) was 0.76 for AC and 0.77 for logBSC using only one acquisition per sonographer, and 0.78 for AC and 0.79 for BSC using five acquisitions. The reliability of the AC and BSC measures was improved if the two sonographers' values were averaged, with ICC (1,2) reaching 0.86 for AC and 0.87 for logBSC using only one acquisition per sonographer, and 0.88 for AC and 0.88 for logBSC using five acquisitions per sonographer.

The wCV was in the range 6.9–7.5% for paired AC measurements. This metric is not applicable for logBSC, which can have negative measurement values.

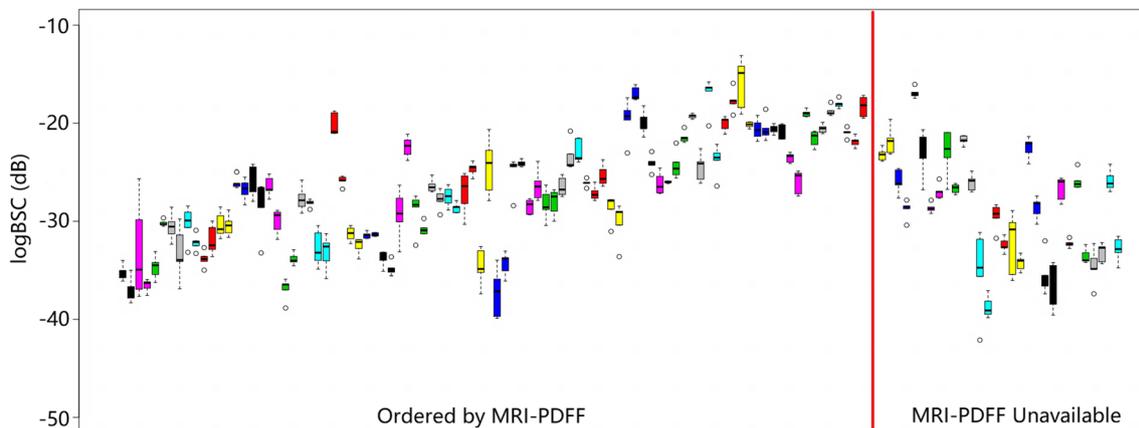


Fig. 3 Boxplots of measured log-transformed backscatter coefficient values grouped by participant-sonographer pairs. Each participant is represented by two adjacent boxes (two sonographers) of the same colour. Left of the red vertical line are participants for which MRI-

proton density fat fraction (MRI-PDFF) are available and those participants are ordered by MRI-PDFF. Right of the red vertical line are participants for which MRI-PDFF are unavailable

Table 1 Intraclass correlation coefficients (ICC) and within-subject coefficients of variation (wCV) of the attenuation coefficient (AC) obtained from a single acquisition (N = 1) and of the mean AC obtained from multiple repeated intra-sonographer acquisitions (N = 2, 3, 4, and 5)

# Acquisitions (N)	ICC(1,1) (95% CI)	ICC(1,2) (95% CI)	wCV (%)
1	0.76 (0.63, 0.85)	0.86 (0.77, 0.92)	7.5
2	0.79 (0.67, 0.87)	0.88 (0.80, 0.93)	6.9
3	0.78 (0.66, 0.86)	0.87 (0.79, 0.92)	7.1
4	0.79 (0.67, 0.87)	0.88 (0.80, 0.93)	7.0
5	0.78 (0.67, 0.86)	0.88 (0.80, 0.93)	7.0

ICC(1,1) was calculated using the ‘irr’ package in R based on a single-unit, absolute-agreement, one-way random effect analysis of variance (ANOVA) model. ICC(1,2) was calculated using the same package based on an average-unit (i.e., average of two sonographers as the assessment basis), absolute-agreement, one-way random effect ANOVA model.

CI confidence interval, ICC intraclass correlation coefficient, wCV within-subject coefficient of variation

SWS Results

Three to ten valid SWS values were obtained per sonographer per participant; in 55 participants, 5 or more SWS values were obtained. The SWS boxplot (Fig. 4) groups the values by participant-sonographer pairs and ordered by fibrosis stage (when histology was available) to provide an overview of the distribution and variability of the acquisition values without intra-sonographer averaging.

When 10 acquisitions were averaged for each sonographer (28 participants), ICC(1,1) was 0.72 (Table 3). The reliability of the SWS measure was improved if the two sonographers were averaged, with ICC(1,2) = 0.84. However, if only one acquisition was used per sonographer, the ICC estimates

Table 2 Intraclass correlation coefficients (ICC) of the log-transformed backscatter coefficient (logBSC) obtained from a single acquisition (N = 1) and of the mean logBSC obtained from multiple repeated intra-sonographer acquisitions (N = 2, 3, 4, and 5)

# Acquisitions (N)	ICC(1,1) (95% CI)	ICC(1,2) (95% CI)
1	0.77 (0.64, 0.85)	0.87 (0.78, 0.92)
2	0.78 (0.67, 0.86)	0.88 (0.80, 0.93)
3	0.78 (0.66, 0.86)	0.88 (0.80, 0.93)
4	0.79 (0.68, 0.87)	0.88 (0.81, 0.93)
5	0.79 (0.67, 0.87)	0.88 (0.80, 0.93)

ICC(1,1) was calculated using the ‘irr’ package in R based on a single-unit, absolute-agreement, one-way random effect analysis of variance (ANOVA) model. ICC(1,2) was calculated using the same package based on an average-unit (i.e., average of two sonographers as the assessment basis), absolute-agreement, one-way random effect ANOVA model.

CI confidence interval, ICC intraclass correlation coefficient

dropped, with ICC(1,1) estimated to be 0.40, and ICC(1,2) estimated to be 0.57.

ICC estimates for SWS calculated using the median of multiple acquisitions are shown in Table 4. ICC point estimates computed using the median were statistically significantly lower than those computed using the mean ($p < 0.05$), and wCV estimate based on the median was statistically significantly higher than that computed based on the mean ($p < 0.05$). However, when 10 acquisitions were used, the absolute difference caused by the use of mean versus median was negligible for ICC and wCV estimates.

The wCV dropped from 43% using a single acquisition to 21% using 10 acquisitions for paired SWS measurements when the mean was used (Table 3). This metric dropped from 43% (single acquisition) to 22% (10 acquisitions) when the median was used (Table 4).

Discussion

This study assesses an important clinical aspect of QIB precision: inter-sonographer reproducibility, for three QIBs (AC, BSC, and SWS) that are measurable from the same clinical scanner. Both accuracy and precision must be demonstrated for a QIB to play important roles in clinical practice. This study deals with precision (reproducibility); diagnostic accuracy (of AC and BSC for assessing hepatic steatosis [6–8], and of SWS for detecting advanced liver fibrosis [2–5]) has been evaluated in prior studies using biopsy and MRI-PDFF as reference standards. High ICC values were obtained for the three QIBs in NAFLD participants with a wide range of hepatic fat fraction and fibrosis stages. These values demonstrate good inter-sonographer reproducibility of AC, BSC, and SWS in NAFLD participants, and represent a necessary step toward the eventual clinical application of these QIBs.

We also investigate how the inter-sonographer reproducibility is affected by the number of acquisitions. It is a common practice to perform multiple acquisitions and use the mean or median as the final measurement of a QIB. For example, 6 to 10 valid acquisitions are commonly used clinically to yield good inter-sonographer SWS reproducibility [19]. The use of multiple SWS acquisitions is supported by our observation of increased ICC and decreased wCV as the number of acquisitions increased. In contrast, our comparison between QUS and SWS shows that a large number of acquisitions is not necessary for QUS (AC and BSC). Increasing the number of acquisitions only slightly improved the ICC estimates for AC and logBSC and using only one acquisition per sonographer already yielded good inter-sonographer reproducibility. QUS’s ability to achieve good cross-sonographer reproducibility using a single acquisition may be an advantage.

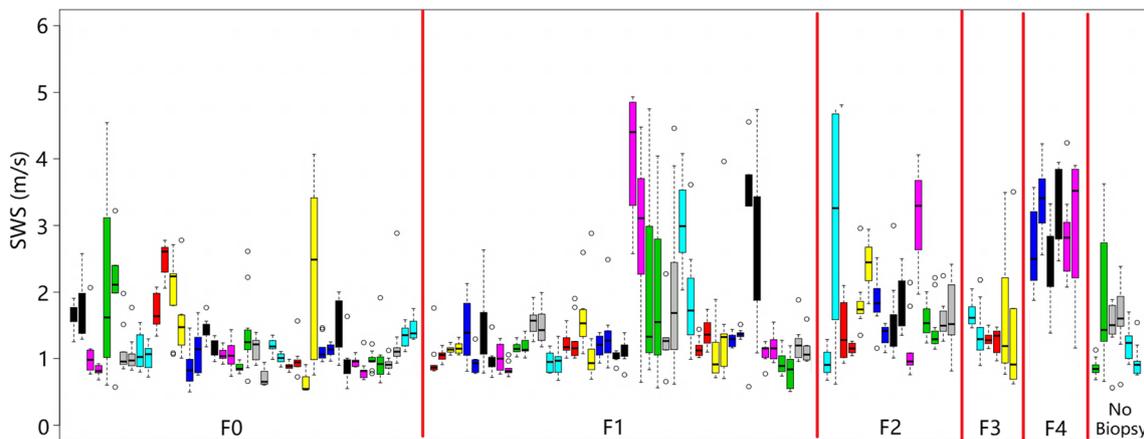


Fig. 4 Boxplots of measured shear wave speed values grouped by participant-sonographer pairs. Each participant is represented by two adjacent boxes (two sonographers) of the same colour. Histology-

determined fibrosis stages are indicated by F0 through F4 when available. The red vertical lines mark the boundaries between adjacent fibrosis stages

The good inter-sonographer reproducibility of AC and BSC demonstrated herein provides important confirmation of this aspect of QUS precision. The repeatability and reproducibility are two metrics that address QIB precision. Repeatability is “the measurement of precision with conditions that remain unchanged between replicate measurements (repeatability conditions)” [16], while reproducibility is “the measurement of precision with conditions that vary between replicate measurements (reproducibility conditions)” [16]. There are a variety of reproducibility conditions, and it is not feasible to assess all reproducibility conditions in a single clinical study. A previous clinical study has assessed the reproducibility between FOI analysts [8]. A previous phantom study has demonstrated that the transducer and sonographer effects are negligible on the overall variability of QUS parameters [17]. The transducer effect was also shown to be negligible on QUS variability in a previous clinical study in adults

with known or suspected NAFLD [18]. However, the more clinically relevant inter-sonographer reproducibility has not been evaluated for QUS in clinical studies. The current study on NAFLD participants addresses this need and shows good inter-sonographer reproducibility.

The evaluation of SWS in the same cohort not only places the QUS results into context (e.g., the number of acquisitions discussion), but also contributes to the SWS literature. The inter-sonographer SWS reproducibility has not been evaluated previously in participants with known or suspected NAFLD and with variable degrees of steatosis and fibrosis. This study found an inter-sonographer ICC(1,1) of 0.72 and ICC(1,2) of 0.84 when 10 SWS acquisitions were averaged for each sonographer. In comparison, a previous study [19] on the inter-operator reproducibility of SWS in a mostly cirrhotic population (58 participants = 41 cirrhotic + 17 noncirrhotic) yielded an overall ICC of 0.81, an ICC of 0.83 for the cirrhotic sub-

Table 3 Intraclass correlation coefficients (ICC) and within-subject coefficients of variation (wCV) of the shear wave speed (SWS) obtained from a single acquisition (N = 1) and of the mean SWS obtained from multiple repeated intra-sonographer acquisitions (N = 2, 3, ..., 10)

# Acquisitions (N)	ICC(1,1) (95% CI)	ICC(1,2) (95% CI)	wCV (%)	# Participants
1	0.40 (0.17, 0.59)	0.57 (0.29, 0.74)	43.0	61
2	0.48 (0.27, 0.65)	0.65 (0.42, 0.79)	34.9	61
3	0.50 (0.29, 0.67)	0.67 (0.45, 0.80)	33.9	61
4	0.77 (0.63, 0.85)	0.87 (0.78, 0.92)	21.5	57
5	0.77 (0.63, 0.86)	0.87 (0.78, 0.92)	19.2	55
6	0.67 (0.49, 0.80)	0.80 (0.65, 0.89)	20.5	51
7	0.67 (0.48, 0.80)	0.81 (0.65, 0.89)	20.6	47
8	0.69 (0.50, 0.82)	0.82 (0.67, 0.90)	20.4	44
9	0.69 (0.48, 0.83)	0.82 (0.65, 0.90)	21.0	38
10	0.72 (0.49, 0.86)	0.84 (0.66, 0.93)	21.3	28

ICC(1,1) was calculated using the ‘irr’ package in R based on a single-unit, absolute-agreement, one-way random effect analysis of variance (ANOVA) model. ICC(1,2) was calculated using the same package based on an average-unit (i.e., average of two sonographers as the assessment basis), absolute-agreement, one-way random effect ANOVA model.

CI confidence interval, ICC intraclass correlation coefficient, wCV within-subject coefficient of variation

Table 4 Intraclass correlation coefficients (ICC) and within-subject coefficients of variation (wCV) of the shear wave speed (SWS) obtained from a single acquisition ($N = 1$) and of the median SWS obtained from multiple repeated intra-sonographer acquisitions ($N = 2, 3, \dots, 10$)

# Acquisitions (N)	ICC(1,1) (95% CI)	ICC(1,2) (95% CI)	wCV (%)	# Participants
1	0.40 (0.17, 0.59)	0.57 (0.29, 0.74)	43.0	61
2	0.48 (0.27, 0.65)	0.65 (0.42, 0.79)	34.9	61
3	0.42 (0.19, 0.61)	0.59 (0.32, 0.76)	39.9	61
4	0.67 (0.50, 0.79)	0.80 (0.67, 0.88)	26.7	57
5	0.67 (0.50, 0.80)	0.81 (0.67, 0.89)	24.1	55
6	0.62 (0.42, 0.76)	0.76 (0.59, 0.86)	23.1	51
7	0.63 (0.42, 0.77)	0.77 (0.59, 0.87)	23.1	47
8	0.67 (0.47, 0.80)	0.80 (0.64, 0.89)	21.8	44
9	0.67 (0.45, 0.81)	0.80 (0.62, 0.90)	22.8	38
10	0.72 (0.48, 0.86)	0.84 (0.65, 0.92)	22.0	28

ICC(1,1) was calculated using the ‘irr’ package in R based on a single-unit, absolute-agreement, one-way random effect analysis of variance (ANOVA) model. ICC(1,2) was calculated using the same package based on an average-unit (i.e., average of two sonographers as the assessment basis), absolute-agreement, one-way random effect ANOVA model.

CI confidence interval, ICC intraclass correlation coefficient, wCV within-subject coefficient of variation

population, and an ICC of 0.70 for the noncirrhotic sub-population. It was not clear if ICC(1,1) or ICC(1,k) were estimated in [19]. However, our ICC(1,2) estimates appear to be consistent with the ICC estimates reported in [19].

The QUS and SWS reproducibility results are comparable to or better than other imaging modalities for liver assessment. For example, an overall ICC of 0.68 was reported for MR elastography, which is used for assessing liver fibrosis [24]. Transient elastography (Fibroscan) has an excellent overall inter-observer reproducibility with a reported ICC up to 0.96 [25]. However, Fibroscan’s reproducibility depends on the liver fibrosis stage (ICC = 0.6 for fibrosis stage ≤ 1 ; ICC = 0.99 for fibrosis stage ≥ 2) and BMI [27].

There are a few limitations in the current form of QUS techniques. First, an external phantom is needed, although phantom stability over time does not appear to be a factor in QUS reproducibility (Appendix). This limitation may be resolved by pre-calibrating the machine settings using a phantom and building the calibration results into the system. Another limitation is that RF data are required. Although every ultrasound system acquires RF data, not all systems provide a means to record or export it. This limitation is alleviated as more manufacturers provide RF output capabilities. This limitation can be eliminated in the future as it is feasible to perform the entire QUS analysis internally in the scanner without the need for external processing.

There are also a few limitations of the current reproducibility study. We did not assess all components of reproducibility. Future research is needed to assess more components such as the inter-platform reproducibility and multiple-site validation. The reproducibility evaluated herein is an aggregate effect of repeatability and sonographers. Our study design does not allow us to separate the sonographer effect from the aggregate effect.

In conclusion, our study shows good inter-sonographer reproducibility of AC, BSC, and SWS in the same cohort of participants scanned in the same instance by the same set of trained expert sonographers using the same machine and transducer. Multiple acquisitions are required for SWS, but not AC or BSC, to achieve good inter-sonographer reproducibility. Further research may be needed to separate the contributions of sonographers and repeatability on the observed reproducibility.

Acknowledgements We are grateful for the dedicated contributions and expertise of the six sonographers who participated in this study, Lara Callahan, Lisa Deiranieh, Elise Housman, Christopher Lucas, Susan Lynch, and Minaxi Trivedi, without whom this work could not be completed.

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

Compliance with ethical standards

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

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

One of the authors (YL) is an employee of Siemens Healthineers USA. The work is supported in part by a research grant from Siemens Healthineers USA. The use of the Siemens S3000 scanner was loaned to the University of California San Diego under a research agreement with Siemens Healthineers USA. At all times, the study data was under the control of the other authors, none of whom are employees of Siemens.

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.

Methodology

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

References

- Shiina T, Nightingale KR, Palmeri ML et al (2015) WFUMB guidelines and recommendations for clinical use of ultrasound elastography: Part 1: basic principles and terminology. *Ultrasound Med Biol* 41:1126–1147
- Palmeri ML, Wang MH, Rouze NC et al (2011) Noninvasive evaluation of hepatic fibrosis using acoustic radiation force-based shear stiffness in patients with nonalcoholic fatty liver disease. *J Hepatol* 55:666–672
- 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
- Ferraioli G, Parekh P, Levitov AB, Filice C (2014) Shear wave elastography for evaluation of liver fibrosis. *J Ultrasound Med* 33:197–203
- Bavu E, Genisson JL, Couade M et al (2011) Noninvasive in vivo liver fibrosis evaluation using supersonic shear imaging: a clinical study on 113 hepatitis C virus patients. *Ultrasound Med Biol* 37:1361–1373
- 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
- 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
- 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. *Am J Roentgenol* 208:W1–W10
- McFarlin BL, Balash J, Kumar V et al (2015) Development of an ultrasonic method to detect cervical remodeling in vivo in full-term pregnant women. *Ultrasound Med Biol* 41:2533–2539
- McFarlin BL, Kumar V, Bigelow TA et al (2015) Beyond cervical length: a pilot study of ultrasonic attenuation for early detection of preterm birth risk. *Ultrasound Med Biol* 41:3023–3029
- Sadeghi-Naini A, Papanicolaou N, Falou O et al (2013) Quantitative ultrasound evaluation of tumor cell death response in locally advanced breast cancer patients receiving chemotherapy. *Clin Cancer Res* 19:2163–2173
- 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
- Caussy C, Alquiraish MH, Nguyen P et al (2017) Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. *Hepatology*. <https://doi.org/10.1002/hep.29639>
- 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
- Cui J, Heba E, Hernandez C et al (2016) Magnetic resonance elastography is superior to acoustic radiation force impulse for the diagnosis of fibrosis in patients with biopsy-proven nonalcoholic fatty liver disease: A prospective study. *Hepatology* 63:453–461
- Sullivan DC, Obuchowski NA, Kessler LG et al (2015) Metrology standards for quantitative imaging biomarkers. *Radiology* 277:813–825
- 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
- Han A, Andre MP, Erdman JW, Loomba R, Claude SB, O'Brien WD (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*. <https://doi.org/10.1002/jum.14537>
- Bota S, Sporea I, Sirlu 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
- Sporea I, Bota S, Jurchis A et al (2013) Acoustic radiation force impulse and supersonic shear imaging versus transient elastography for liver fibrosis assessment. *Ultrasound Med Biol* 39:1933–1941
- Yao LX, Zagzebski JA, Madsen EL (1990) Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors. *Ultrasound Imaging* 12:58–70
- Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86:420–428
- 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
- 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
- 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
- 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
- Fraquelli M, Rigamonti C, Casazza G et al (2007) Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. *Gut* 56:968–973