Abdominal Radiology (2019) 44:3285–3294
https://doi.org/10.1007/s00261-019-02183-0

HEPATOBILIARY

2D shear wave elastography: measurement acquisition and reliability criteria in noninvasive assessment of liver fibrosis

Maggie Chung[1] · Grayson L. Baird[2] · Krista E. Weiss[3] · Michael D. Beland[3]

Published online: 21 August 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Purpose The objective was to evaluate the accuracy of 2D shear wave elastography (SWE) in predicting stages of liver fibrosis using five individual versus grouped measurements and different reliability criteria.

Materials and methods This is a prospective study of 109 patients who underwent hepatic 2D SWE (Canon Aplio 500) prior to liver biopsy for varied indications. Liver fibrosis was staged using the METAVIR scoring system ($F=0$–4). Propagation mapping was used to guide ten SWE measurements from the liver parenchyma: five individual measurements and five grouped measurements. IQR/median, SD/median, and SD/mean were examined as quality criteria for patient inclusion at various thresholds (IQR/median $\leq$ 0.15, 0.2, 0.3, 0.4, 0.5; SD/median $\leq$ 0.15, 0.2, 0.3; SD/mean $\leq$ 0.2, 0.3, 0.5). Threshold for clinically significant fibrosis ($F \geq 2$) was determined with receiver operating characteristic (ROC) analysis.

Results There was high agreement between individual and grouped measurements without statistically significant differences (intraclass correlation coefficient = 0.82; $p=0.26$–0.96). When no quality criterion was used ($n=103$), the optimal threshold was 11.3 kPa [AUROC 0.78, 95% CI (0.69, 0.88)] with sensitivity and specificity of 80% and 66%, respectively. All quality criteria were associated with equal or higher AUROC ranging from 0.78 to 0.87. IQR/median $\leq$ 0.5 ($n=88$) achieved the highest sensitivity of 85% and only excluded a small subset of patients. The AUROC and specificity were 0.83 [95% CI (0.74, 0.92)] and 72%, respectively.

Significance Quality criterion IQR/median $\leq$ 0.5 increases sensitivity and specificity in prediction of clinically significant liver fibrosis while excluding only a small subset of patients. Grouped measurements are comparable to individual measurements and may help increase procedural efficiency.

Keywords Two-dimensional shear wave · Elastography · Liver cirrhosis · Ultrasonography

Introduction

Chronic liver disease is a major cause of morbidity and mortality in the United States [1, 2]. Accurate staging of fibrosis is essential for prognosis and treatment of chronic liver disease [3]. While liver biopsy is considered the reference standard in liver fibrosis assessment, it has several limitations. Liver biopsy is a costly and invasive test associated with complications including hospitalization (3%), bleeding (0.6%), and death (0.1%) [3, 4]. Liver biopsy is susceptible to sampling error as biopsy represents only 1/50,000th of the liver volume [5]. Furthermore, histologic assessment of liver fibrosis is variable with intra- and interobserver concordance of 75% and 78%, respectively [3]. Therefore, accurate and noninvasive methods for liver fibrosis staging are of great clinical value.

✉ Maggie Chung
Mchung2017@gmail.com

Grayson L. Baird
gbaird@lifespan.org

Krista E. Weiss
kristasuarezweiss@gmail.com

Michael D. Beland
mbeland@lifespan.org

[1] Department of Radiology and Biomedical Imaging, University of California, San Francisco, 505 Parnassus Ave., Room M391, Box 0628, San Francisco, CA 94143, USA

[2] Lifespan Biostatistics Core, Rhode Island Hospital, 593 Eddy St., Providence, RI 02903, USA

[3] Department of Diagnostic Imaging, Warren Alpert Medical School of Brown University, Rhode Island Hospital, 593 Eddy St., Providence, RI 02903, USA

2D SWE is a novel SWE technique that allows real-time visualization of shear wave propagation. Shear waves are detected at multiple lateral locations to produce multiple shear wave speed images that are then integrated into one 2D-SWE image. The real-time color elastogram is helpful for anatomic and tissue stiffness guidance. Multiple measurements can be acquired from a single image over a large field of view. Measurement approach and reliability criteria for 2D SWE are not currently well defined in the literature. While multiple measurements can be obtained from a single image, no study to date has compared the diagnostic accuracy of grouped versus individual measurements. For 2D SWE, a minimum of three to five measurements may be sufficient for accurate assessment of liver fibrosis if a quality criterion is used [6–13]. IQR/median ≤ 0.3 has been used in 2D SWE studies as a reliability criterion to exclude less reliable measurements with high variability [10, 14]. This criterion was adopted from transient elastography (TE) studies, and there is limited evidence for its use in 2D SWE [7]. Recent studies have also used SD/median ≤ 0.10–0.30 and SD/mean (coefficient of variance) ≤ 0.2 as reliability criteria for 2D SWE [8, 15, 16]. A comparison of these reliability criteria would help identify the ideal measurement variability parameter to improve diagnostic performance.

The aim of this study was to evaluate the diagnostic performance of 2D SWE in predicting stages of liver fibrosis among different reliability criteria and using five individual versus grouped measurements.

Materials and methods

Study design and patient selection

This was a prospective, single-center study approved by the Institutional Review Board. All patients provided written informed consent. Patients were referred for ultrasound-guided, nontargeted liver biopsy for known or suspected chronic liver disease. Patients with targeted biopsies were excluded from the study. For each patient, aspartate aminotransferase, alanine aminotransferase, and total bilirubin before liver biopsy were collected.

2D SWE examination

Two-dimensional SWE examinations were performed by one of five sonographers with an average of 19 years (range 12 to 27 years) of ultrasound experience. All sonographers received 2D SWE applications training and had at least 2 years of elastography experience prior to the start of the study. 2D SWE measurements were obtained using an Aplio 500 Platinum Series ultrasound machine and 6C1 (PVT-375BT) curved array ultrasound transducer (Canon Medical Systems Corporation, formerly Toshiba Medical Systems Corporation). All patients fasted overnight prior to 2D SWE examination. Sonographers obtained all measurements in accordance with the consensus statement from the Society of Radiologists in Ultrasound [6]. The examinations were performed using right intercostal approach with patients in supine position and right arm raised above the head. Measurements were obtained from the right hepatic lobe during expiratory breath hold. Acquisition box size was 2 to 4 cm in lateral direction and 3 to 5 cm in axial direction. All measurements were obtained using 10-mm-diameter circle ROIs. Circle ROIs were placed at least 1 cm below the liver capsule and within 5 cm of the skin surface. Circle ROIs were placed in areas of homogeneous portion of the speed image while avoiding vascular and biliary structures. Parallel propagation contours were observed in the region of homogenous color. A total of ten measurements were obtained: five measurements obtained individually (a single circle ROI per acquisition box) and five measurements obtained in groups (multiple circle ROIs per acquisition box) (Fig. 1). Grouped measurements were obtained in clusters of two to three measurements per acquisition. All ten measurements were obtained from similar location in the liver. All elasticity measurements were expressed in kilopascals (kPa).

Liver biopsy and histologic evaluation

Same day liver biopsies were performed subsequent to 2D SWE evaluation. Percutaneous core liver biopsies were performed using 18-gauge automated core needles. Samples were obtained from one to two passes through the right hepatic lobe in the region of 2D-SWE measurement acquisition. Tissue specimens were fixed in formalin and embedded in paraffin. Biopsy samples were reviewed by a pathologist with greater than 20 years of experience who was blinded to 2D SWE measurements. For all samples, fibrosis was staged using the METAVIR system. Samples with histologic evidence of nonalcoholic fatty liver disease were staged using both Brunt and METAVIR systems. Clinically significant fibrosis was defined as stage two and above.

Statistical analysis

Fibrosis stage was modeled by median elasticity (kPa) value by sampling approach (individual vs. grouped sampling) and for each reliability criteria (IQR/median ≤ 0.15, 0.2, 0.3, 0.4, 0.5; SD/median ≤ 0.15, 0.2, 0.3; SD/mean ≤ 0.2, 0.3, 0.5) using generalized linear modeling, assuming a binomial distribution (0–4) and a binary distribution (1/0), where a METAVIR fibrosis stage ≥ 2 was considered 1. Modeling between METAVIR and Brunt was accomplished using generalized estimating equations (GEE) with sandwich estimation, where observations

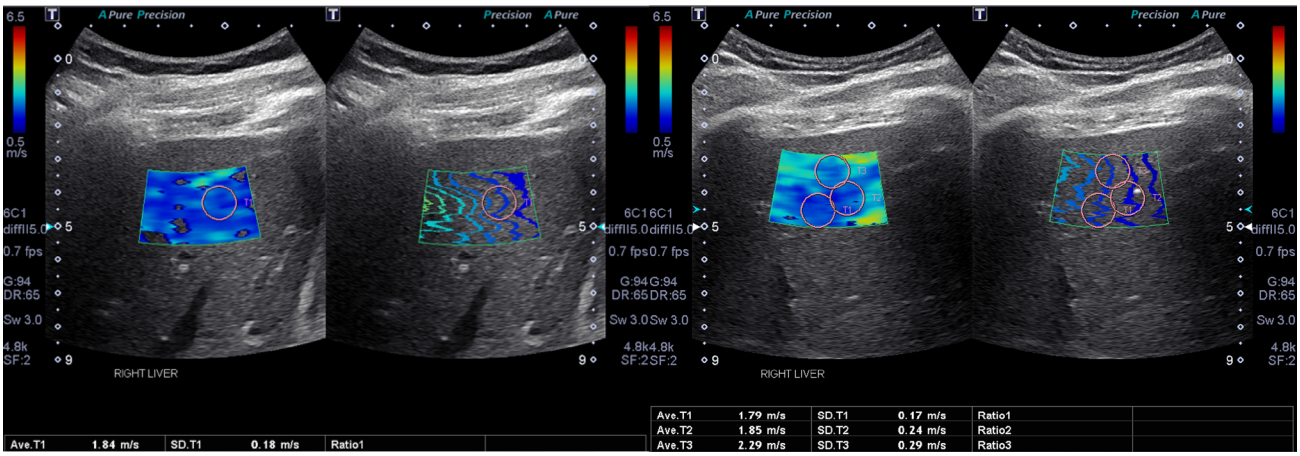

Fig. 1 A total of ten measurements were obtained from each patient: five measurements obtained individually (single ROI per acquisition) and five measurements obtained in groups (up to three ROIs per acquisition)

were nested within patient. Diagnostic performance metrics include odds ratios, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the curve (ROC) analyses. In addition, Youden's index was used to determine mathematically optimal thresholds. All analyses were modeled using SAS Software 9.4 (SAS Inc, Cary, NC) using the GLIMMIX and LOGISTIC procedures. Brunt and METAVIR fibrosis stage agreement was evaluated using weighted Kappa coefficient with the FREQ procedure. Alpha was established a priori at the 0.05 level and all interval estimates were calculated for 95% confidence.

Results

From November 2015 to December 2017, 109 patients underwent liver sonography with 2D SWE prior to random liver biopsy for various indications. Of the 109 patients, there were six excluded patients: one by targeted biopsy, two by technical failure, three by missing SWE measurements. Of the final 103 patients, the median age was 55 and range was 19–86 years old. There were 50 males (49%) and 53 females (51%). Primary causes of liver disease, determined by clinical history, laboratory results, and/or histological findings, included viral hepatitis, alcoholic liver disease, nonalcoholic fatty liver disease, hemochromatosis, Wilson's Disease, and autoimmune liver disease. Distribution of fibrosis stages was *F*0 49 (47%), *F*1 19 (18%), *F*2 13 (13%), *F*3 11 (11%), and *F*4 11 (11%). Median and mean elasticity were 11.2 kPa and 11.6 kPa, respectively. Elasticity values ranged from 3.4 to 50.5 kPa (Table 1).

METAVIR versus Brunt staging

Fifty-one patients had evidence of nonalcoholic fatty liver disease. In these patients, fibrosis was staged using both METAVIR and Brunt systems. There was very good agreement between the two staging systems [$\kappa = 0.85$ (95% CI 0.76, 0.97)]. Of the 51 patients, only seven patients had discordant fibrosis stages by the two systems. Of the seven patients, five were staged F0 by METAVIR and F1 by Brunt, one was staged F1 by METAVIR and F2 by Brunt, and one was staged F2 by METAVIR and F3 by Brunt (Fig. 2a). Likewise, similar relationships between elasticity (kPa) and fibrosis stage were observed for METAVIR and Brunt staging (Fig. 2b). For every 1-unit increase in median kPa, the odds of advancing one fibrosis stage increased 18.3% [OR 1.18, 95% CI (0.09, 1.39), $p = 0.039$] using Brunt staging and 20% [OR 1.20, 95% CI (1.02, 1.41), $p = 0.030$] using METAVIR staging; these slopes were not found to be statistically different (interaction effect, $p = 0.39$). Given that there was no evidence these two staging systems were significantly different, METAVIR fibrosis staging was used for all patients in subsequent regression analyses.

Acquisition method: individual versus grouped measurements

We examined the diagnostic accuracy of individual versus grouped measurements in predicting stages of liver fibrosis. There was excellent agreement between grouped and individual measurements for both median elasticity and mean elasticity with intraclass correlation (ICC) of 0.82 and 0.79, respectively [17]. Between individual and grouped measurements, there were overall similar numbers of patients excluded by each respective reliability criteria. Among all

Table 1 Patient characteristics, diagnoses, laboratory values, histologic fibrosis stages, and elasticity values

Patient characteristics		
	n	%
Sex		
Male	50	49
Female	53	51
	Median	Range
Age	55	19–86
AST (U/L)	49	15–180
ALT (U/L)	60	12–262
Total bilirubin (mg/dl)	0.9	0.3–8.2
Primary diagnoses	*n*	%
NAFLD/NASH	37	36
ALD	18	17
Viral Hepatitis	16	16
Autoimmune	8	8
Hemochromatosis	5	5
Primary sclerosing cholangitis	4	4
Drug induced	2	2
Unknown	13	13
METAVIR fibrosis stage	*n*	%
F0	49	48
F1	19	18
F2	13	13
F3	11	11
F4	11	11
Elasticity		kPa
Median		11.2
Mean		11.6
Range		3.4–50.5

AST aspartate aminotransferase, *ALT* alanine aminotransferase, *NAFLD* nonalcoholic fatty liver disease, *NASH* nonalcoholic steatohepatitis, *ALD* alcoholic liver disease

reliability criteria groups, no statistically significant differences were observed between the diagnostic accuracy of individual versus grouped measurements ($p = 0.26$–0.96) (Table 2). Given the high agreement and no statistically significant difference observed between the two measurement acquisition methods for predicting fibrosis stage, individual and grouped measurements were combined in subsequent reliability criteria analysis.

Reliability criteria

Various quality criteria for patient inclusion based on thresholds of IQR/median, SD/median, and SD/mean were examined (Table 3). For each reliability criterion, relationship between elasticity and fibrosis stage was analyzed using 10 measurements: five individual and five grouped measurements. When no reliability criterion was used ($n = 103$), the optimal threshold was 11.3 kPa with AUROC of 0.78 [OR 1.09, 95% CI (1.06, 1.12)] ($p < 0.01$) (Fig. 3). The sensitivity (SN) and specificity (SP) were 80% and 66%, respectively. The positive predictive value (PPV) and negative predictive value (NPV) were 55% and 86%, respectively. All reliability criteria were associated with equal or greater AUROC and ranged from 0.78 to 0.87. Using IQR/median ≤ 0.3, 48 patients were excluded ($n = 55$) (Fig. 4a). The optimal threshold was 11.5 kPa with AUROC of 0.81 [OR 1.12, 95% CI (1.06, 1.19)] ($p < 0.01$) (Fig. 4b). The SN and SP were 81% and 74%, respectively. The PPV and NPV were 57% and 91%, respectively. SD/median ≤ 0.15 ($n = 23$) achieved the

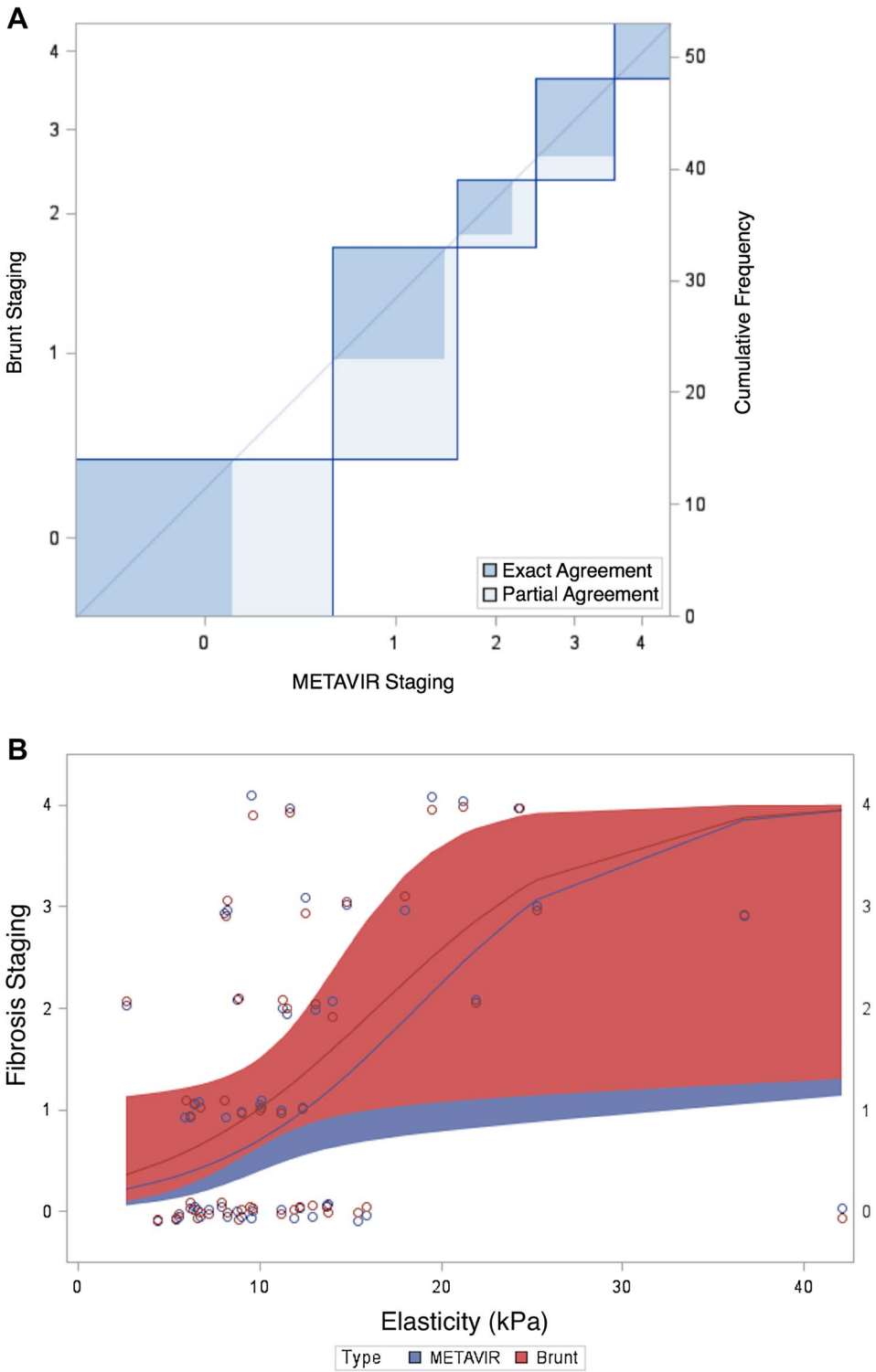

Fig. 2 **a** The dark blue areas represent exact agreement between Brunt and METAVIR staging. The light blue areas represent partial agreement between Brunt and METAVIR staging with difference of one stage. There was no statistically significant difference between Brunt and METAVIR fibrosis stages. **b** The data are shown for Brunt (red circles) and METAVIR (blue circles) fibrosis staging. The red (Brunt) and blue (METAVIR) lines are predicted from data modeling of likelihood of fibrosis stage versus elasticity; 95% CIs are shown in shaded areas. *CI* confidence interval

highest AUROC of 0.87 [OR 1.15, 95% CI (1.05, 1.27)] (Fig. 5a, b). The optimal threshold was 11.7 kPa. SN and SP were 83% and 88%, respectively. The PPV and NPV were 71% and 94%, respectively. Using IQR/median ≤ 0.5, 15 patients were excluded ($n = 88$) (Fig. 6a). The optimal threshold was 11.3 kPa with AUROC of 0.83 [OR 1.10, 95% CI (1.06, 1.14)] (Fig. 6b). The SN and SP were 85% and 72%, respectively. The PPV and NPV were 58% and 92%, respectively.

Table 2 Diagnostic accuracy of individual versus grouped measurements using different reliability criteria

Rule	Grouped							Individual							*p*
	n	OR	95% CI	ROC	Threshold	SN	SP	*n*	OR	95% CI	Threshold	ROC	SN	SP	
None	103	1.08	1.03, 1.13	0.79	12.1	0.81	0.75	103	1.09	1.03, 1.15	12.1	0.80	0.75	0.72	0.53
SD/median															
≤ 0.15	44	1.16	1.04, 1.29	0.78	11.7	0.79	0.73	66	1.62	1.06, 1.30	11.5	0.83	0.78	0.80	0.96
≤ 0.20	72	1.08	1.02, 1.15	0.76	12.4	0.74	0.76	62	1.07	0.99, 1.16	11.5	0.82	0.82	0.73	0.67
≤ 0.30	89	1.07	1.02, 1.12	0.80	12.4	0.80	0.76	88	1.10	1.02, 1.18	11.5	0.84	0.81	0.72	0.30
IQR/median															
≤ 0.15	6	NA	NA	NA	NA	NA	NA	8	NA	NA	NA	NA	NA		NA
≤ 0.20	24	1.07	0.85, 1.34	0.64	7.4	0.99	0.35	17	1.25	0.94, 1.66	15.5	0.92	0.80	0.92	0.26
≤ 0.30	45	1.17	1.05, 1.30	0.80	11.7	0.79	0.77	43	1.19	1.08, 1.32	11.5	0.84	0.77	0.83	0.75
≤ 0.40	73	1.09	1.02, 1.15	0.75	12.4	0.73	0.77	67	1.08	1.00, 1.16	11.5	0.83	0.80	0.75	0.77
≤ 0.50	88	1.07	1.02, 1.11	0.80	12.4	0.79	0.76	89	1.09	1.02, 1.16	11.5	0.83	0.79	0.71	0.37
SD/mean (coefficient of variance)															
≤ 0.20	77	1.09	1.03, 1.16	0.78	12.4	0.77	0.77	62	1.08	1.00, 1.17	11.5	0.84	0.83	0.75	0.63
≤ 0.30	92	1.07	1.02, 1.12	0.80	12.4	0.80	0.74	94	1.10	1.03, 1.17	9.7	0.82	0.91	0.57	0.34
≤ 0.50	102	1.07	1.02, 1.12	0.79	12.4	0.80	0.75	100	1.09	1.03, 1.16	12.1	0.81	0.74	0.73	0.34

OR odds ratio, *ROC* area under receiver operating characteristic curve, *CI* confidence interval, *SN* sensitivity, *SP* specificity, *IQR* interquartile range, *SD* standard deviation, *NA* not applicable

Table 3 Diagnostic accuracy of 10 measurements in predicting clinically significant fibrosis using different reliability criteria

Reliability criteria	*n*	OR	OR 95% CI	ROC	ROC 95% CI	Threshold	Sensitivity	Specificity	PPV	NPV
No rule	103	1.09	1.06, 1.12	0.78	0.69, 0.88	11.3 kPa	0.80	0.66	0.55	0.86
SD/median										
SD/median ≤ 0.15*	23	1.15	1.05, 1.27	0.87	0.69, 1.00	11.7 kPa	0.83	0.88	0.71	0.94
SD/median ≤ 0.2*	51	1.14	1.07, 1.21	0.79	0.66, 0.92	11.5 kPa	0.73	0.75	0.55	0.87
SD/median ≤ 0.3*	79	1.10	1.06, 1.13	0.82	0.72, 0.92	11.5 kPa	0.82	0.71	0.53	0.91
IQR/median										
IQR/median ≤ 0.15	17	1.07	0.94, 1.21	0.78	0.56, 1.00	9.6 kPa	1.00	0.615	0.44	1.00
IQR/median ≤ 0.2*	28	1.10	1.02, 1.18	0.81	0.65, 0.99	11.7 kPa	0.78	0.84	0.7	0.89
IQR/median ≤ 0.3*	55	1.12	1.06, 1.19	0.81	0.69, 0.93	11.5 kPa	0.81	0.74	0.57	0.91
IQR/median ≤ 0.4*	74	1.13	1.08, 1.18	0.82	0.72, 0.92	12.5 kPa	0.73	0.81	0.62	0.88
IQR/median ≤ 0.5*	88	1.10	1.06, 1.14	0.83	0.74, 0.92	11.3 kPa	0.85	0.72	0.58	0.92
SD/mean (coefficient of variance)										
SD/mean ≤ 0.2*	53	1.14	1.07, 1.22	0.80	0.68, 0.93	11.5 kPa	0.73	0.76	0.55	0.88
SD/mean ≤ 0.3*	83	1.10	1.06, 1.13	0.82	0.73, 0.92	11.3 kPa	0.83	0.73	0.56	0.92
SD/mean ≤ 0.5*	97	1.09	1.06, 1.12	0.81	0.72, 0.89	11.3 kPa	0.81	0.68	0.55	0.88

For each reliability criterion, the relationship between elasticity and fibrosis stage was analyzed using 10 measurements: five individual and five grouped measurements

OR odds ratio, *CI* confidence interval, *ROC* area under receiver operating characteristic curve, *SN* sensitivity, *SP* specificity, *IQR* interquartile range, *SD* standard deviation

*$p < 0.01$

Discussion

Identifying clinically significant liver fibrosis is necessary for treatment of early disease before progression to irreversible cirrhosis. Accurate diagnosis of clinically significant fibrosis using 2D SWE can reduce the number of invasive liver biopsies in the management of chronic liver disease. Several factors including patient obesity and operator variability may lead to unreliable measurements

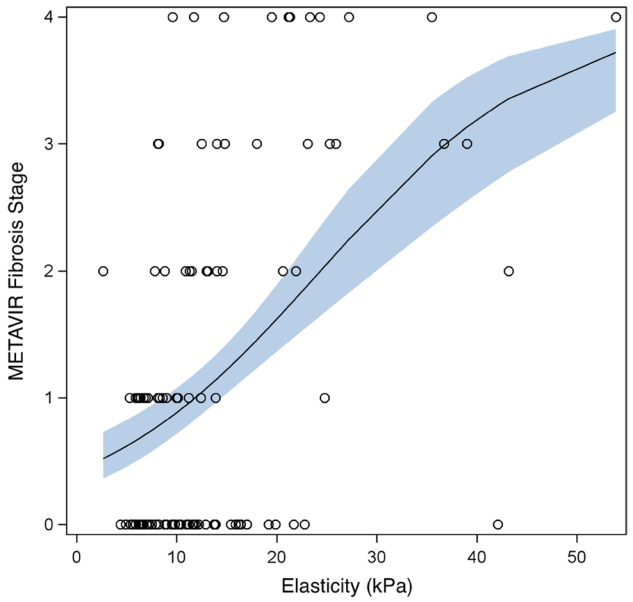

Fig. 3 The circles represent median elasticity value (kPa) for each patient. The line is predicted from data modeling of likelihood of METAVIR fibrosis stage versus elasticity; 95% CI is shown in shaded area. *CI* confidence interval

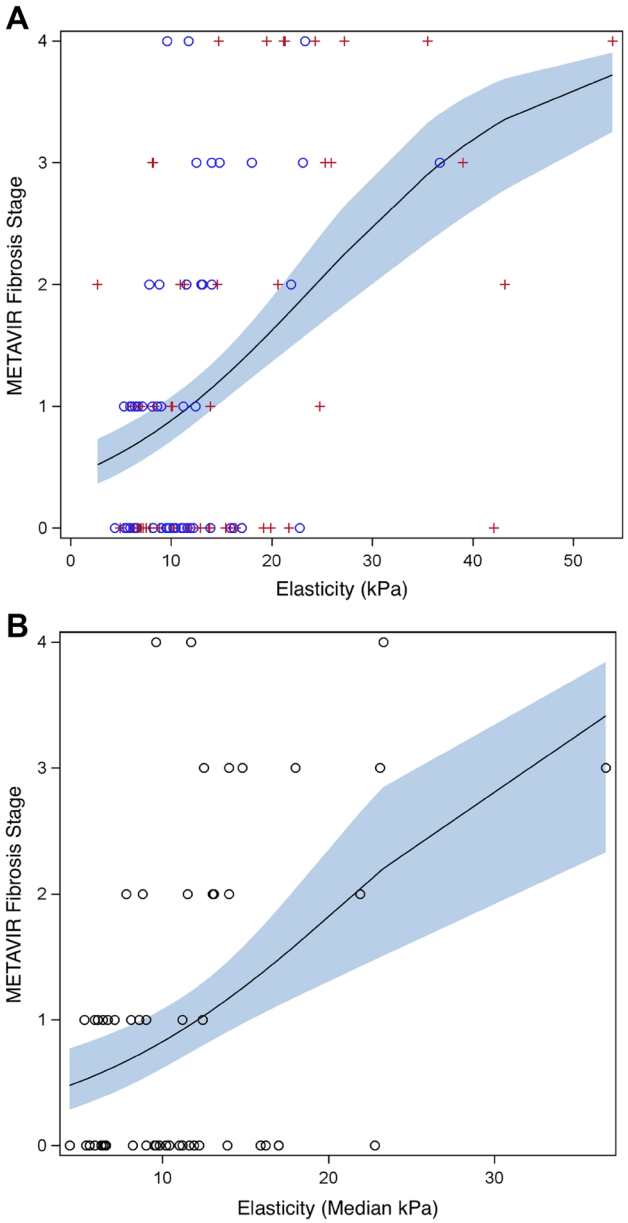

Fig. 4 **a** The circles represent patients with IQR/median ≤ 0.3. The red crosses represent excluded patients with IQR/median > 0.3. The line is predicted from data modeling of likelihood of METAVIR fibrosis stage versus median elasticity for all patients; 95% CI is shown in shaded area. *IQR* interquartile range, *CI* confidence interval. **b** Correlation between elasticity values and METAVIR fibrosis stage for patients with IQR/median ≤ 0.3. The circles represent median elasticity value (kPa) for each patient. The line is predicted from data modeling of likelihood of METAVIR fibrosis stage versus elasticity; 95% CI is shown in shaded area. *IQR* interquartile range, *CI* confidence interval

which may provide inaccurate assessment of the liver parenchyma. Quality criteria can improve diagnostic accuracy by excluding measurements with high variability.

There is limited evidence for the use of quality criteria in 2D SWE. Based on evidence from TE studies, quality criterion IQR/median ≤ 0.3 has been used in 2D SWE studies to identify reliable measurements [10, 14]. Alternative thresholds using SD/median and SD/mean (CV: coefficient of variance) have also been proposed for 2D SWE. Propocet et al. identified SD/median ≤ 0.10 as the optimal quality criterion in the diagnosis of portal hypertension [15]. Thiele et al. used SD/median ≤ 0.30 in the assessment of liver fibrosis in patients with alcohol abuse [8]. Lim et al. identified higher diagnostic performance using CV ≤ 0.20 compared to IQR/median ≤ 0.20 [16]. To the best of the authors' knowledge, this is the first study to evaluate diagnostic performance of 2D SWE among all three reliability criteria. Our study shows that using quality criteria for measurement inclusion improves the prediction of clinically significant liver fibrosis with high sensitivity. In this study, all quality criteria using IQR/median, SD/median, or SD/mean achieved equal or higher diagnostic accuracy compared to baseline (0.78–0.87 vs. 0.78). While IQR/median ≤ 0.3 increased the SN and SP in prediction of clinically significant liver fibrosis, it excluded approximately half of the patients (48/103) in our study. Similarly, while SD/median ≤ 0.15 was the most accurate (AUROC = 0.87), it was exceedingly restrictive and excluded 78% of patients (80/103). We identified IQR/median ≤ 0.5 as the optimal quality criterion as it increased the AUROC from 0.78 to 0.83, SN from 80 to 85%, and SP from 66 to 72% while only excluding 15% of patients (15/103). Furthermore, IQR/median ≤ 0.5 was associated with a high NPV of 92%, which is advantageous in selecting patients for liver biopsy.

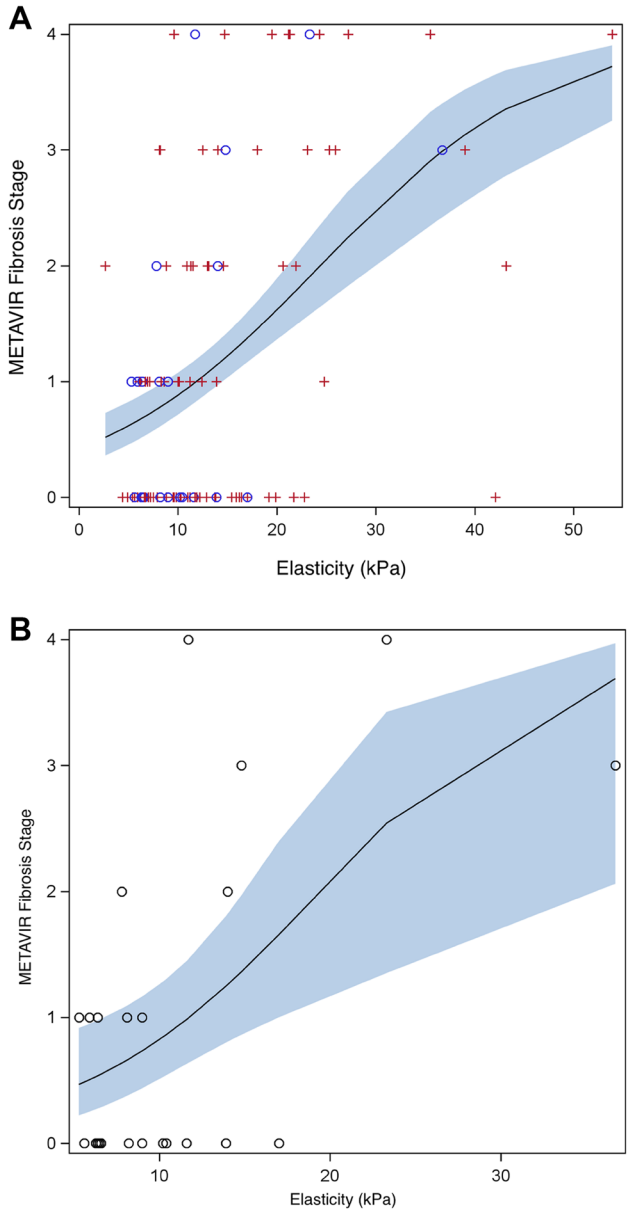

Fig. 5 **a** The circles represent patients with SD/median ≤ 0.15. The red crosses represent excluded patients with SD/median > 0.15. The line is predicted from data modeling of likelihood of METAVIR fibrosis stage versus median elasticity for all patients; 95% CI is shown in shaded area. *SD* standard deviation, *CI* confidence interval. **b** Correlation between elasticity values and METAVIR fibrosis stage for patients with SD/median ≤ 0.15. The circles represent median elasticity value (kPa) for each patient. The line is predicted from data modeling of likelihood of METAVIR fibrosis stage versus elasticity; 95% CI is shown in shaded area. *SD* standard deviation, *CI* confidence interval

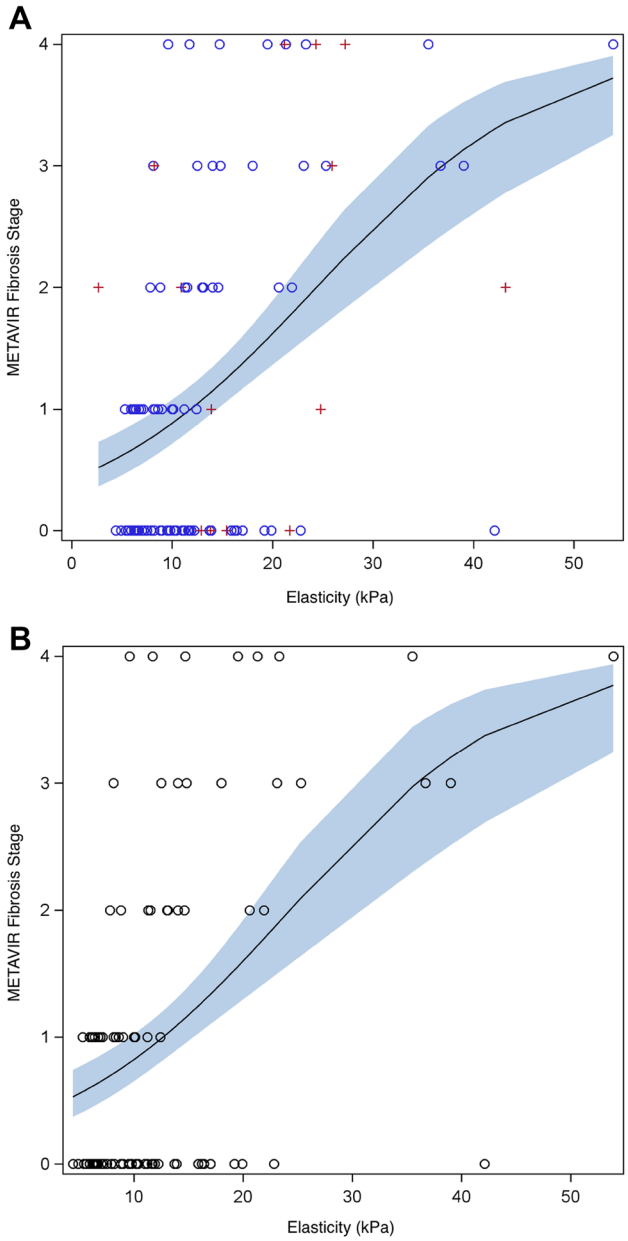

Fig. 6 **a** The circles represent patients with IQR/median ≤ 0.5. The red crosses represent excluded patients with IQR/median > 0.5. The line is predicted from data modeling of likelihood of METAVIR fibrosis stage versus median elasticity for all patients; 95% CI is shown in shaded area. *IQR* interquartile range, *CI* confidence interval. **b** Correlation between elasticity values and METAVIR fibrosis stage for patients with IQR/median ≤ 0.5. The circles represent median elasticity value (kPa) for each patient. The line is predicted from data modeling of likelihood of METAVIR fibrosis stage versus elasticity; 95% CI is shown in shaded area. *IQR* interquartile range, *CI* confidence interval

Patient-related, operator-related, and/or manufacturer model-related factors may have contributed to measurement variability and patient exclusion. In our study, 51/103 patients had primary diagnosis of nonalcoholic fatty liver disease and possible nonalcoholic steatohepatitis. Nonalcoholic steatohepatitis is associated with heterogenous distribution of liver fibrosis [18]. Consequently, quality criteria may exclude some patients with high measurement variability due to areas with different degrees of fibrosis

rather than technical inaccuracies. In 2D SWE, a learning curve has been observed with higher reproducibility from more experienced operators [13]. For patients imaged earlier in the study, comparatively less sonographer experience with the 2D SWE system may have resulted in more variable measurements. Also, manufacturer differences may have contributed to high IQR/median ratios as measurement variability can differ among shear wave elastography manufacturer models [19].

Using IQR/median ≤ 0.5 as the quality criterion, this study identified 11.3 kPa as the mathematically optimal threshold for predicting clinically significant fibrosis with SN and SP of 85% and 72%, respectively. Other studies have reported 2D SWE thresholds for $F \geq 2$ ranging from 6.3 to 12.4 kPa [20, 21]. Our threshold is at the higher end of the range, and may be attributable to manufacturer differences between our study (Aplio 500, Canon, formerly Toshiba) and previous studies (Aixplorer, Supersonic Imagine). Another potential factor is that our study allowed up to three ROI measurements to be obtained within a single color box (grouped measurements); however, no significant differences in diagnostic accuracy were identified between individual and grouped measurements. The difference in study populations may also account for our higher threshold. Many previous 2D SWE studies focused on patients with one specific etiology (e.g., viral, alcoholic, NAFLD) while our study included patients with a variety of liver pathologies including less commonly studied etiologies (e.g., autoimmune, hemochromatosis, and primary sclerosing cholangitis). This heterogeneous population introduces inherent variability, which makes identifying a uniform threshold more challenging.

To the best of the authors' knowledge, this is the first study to compare individual versus grouped measurements. While 2D SWE allows multiple ROIs to be placed within a single elastogram, there is a theoretical concern for less reliable measurements because an error in one image may be reproduced in all measurements from that group [13]. This study found no significant difference between thresholds and diagnostic accuracies between individual and grouped measurements. Compared to pSWE, 2D SWE examination times are approximately twice as long because the sonographer must wait for stabilization of the elastogram and adjust ROI placement until parallel dispersion lines are observed [22]. Given similar diagnostic accuracies between 2D SWE and pSWE, longer examination times may discourage the use of 2D SWE over pSWE in clinical settings [23, 24]. Using grouped measurements may increase procedural efficiency to offset this difference. Further study is needed to assess the average time saved by using grouped measurements.

One potential limitation of this study is the use of METAVIR fibrosis staging for all patients, including those with nonalcoholic steatohepatitis (NASH), which may affect threshold analysis and diagnostic accuracy. One staging system was used to allow for regression analysis with all patients. To assess potential differences in fibrosis stages using other staging systems, NAFLD/NASH patients were staged using both METAVIR and Brunt staging criteria. There was high agreement between the two staging systems. Of the seven patients with discordant fibrosis stages, only one patient had *F*1–*F*2 disagreement that would affect threshold analysis and diagnostic accuracy. Another limitation of the study is the relatively small proportion of *F*3 and *F*4 patients (22%). As this is a prospective study, this reflects the proportion of patients who were referred to biopsy and found to have severe fibrosis and cirrhosis. SWE measurements are susceptible to interobserver variability as patients were imaged by one of five sonographers. Lastly, this was a single institution study using one manufacturer model (Aplio 500, Canon, formerly Toshiba), and results may not be generalizable to all SWE models. A larger multi-institutional study using multiple manufacturer models may be required to confirm the optimal reliability criterion.

In conclusion, the use of IQR/median ≤ 0.5 as the reliability criterion can improve the diagnostic performance of 2D SWE in predicting clinically significant fibrosis. IQR/median ≤ 0.5 increases the sensitivity and specificity while excluding only a small subset of patients. Grouped measurements are comparable to individual measurements, and may help increase procedural efficiency.

Compliance with ethical standards

Conflict of interest Michael Beland: Received an equipment grant from Canon Medical Systems Corporation. All the other authors declare that they have no conflict of interest.

Ethical approval IRB: This study was a prospective, single-center study approved by the Institutional Review Board.

References

1. Asrani SK, Larson JJ, Yawn B, Therneau TM, Kim WR. Underestimation of Liver-Related Mortality in the United States. *Gastroenterology*. 2013;145(2):375-382.e2. https://doi.org/10.1053/j.gastro.2013.04.005
2. Bell BP, Manos MM, Zaman A, et al. The Epidemiology of Newly Diagnosed Chronic Liver Disease in Gastroenterology Practices in the United States: Results From Population-Based Surveillance. *Am J Gastroenterol*. 2008;103(11):2727-2736. doi:10.1111/j.1572-0241.2008.02071.x
3. Tapper EB, Lok AS-F. Use of Liver Imaging and Biopsy in Clinical Practice. Longo DL, ed. *N Engl J Med*. 2017;377(8):756-768. https://doi.org/10.1056/nejmra1610570
4. Janes CH, Lindor KD. Outcome of patients hospitalized for complications after outpatient liver biopsy. *Ann Intern Med*. 1993;118(2):96-98. http://www.ncbi.nlm.nih.gov/pubmed/8416324. Accessed May 6, 2018.

5. Bravo AA, Sheth SG, Chopra S. Liver Biopsy. *N Engl J Med*. 2001;344(7):495-500. https://doi.org/10.1056/nejm200102153440706
6. Barr RG, Ferraioli G, Palmeri ML, et al. Elastography Assessment of Liver Fibrosis: Society of Radiologists in Ultrasound Consensus Conference Statement. *Radiology*. 2015;276(3):845-861. https://doi.org/10.1148/radiol.2015150619
7. Dietrich CF, Bamber J, Berzigotti A, et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Liver Ultrasound Elastography, Update 2017 (Long Version) EFSUMB-Leitlinien und Empfehlungen zur klinischen Anwendung der Leberelastographie, Update 2017 (Langversion) Guidelines & Recommendations Dietrich CF et al. EFSUMB Guidelines and… Ultraschall in Med Basic Principles and Technology. https://doi.org/10.1055/s-0043-103952
8. Thiele M, Madsen B, Procopet B, et al. Reliability Criteria for Liver Stiffness Measurements with Real-Time 2D Shear Wave Elastography in Different Clinical Scenarios of Chronic Liver Disease. *Ultraschall der Medizin - Eur J Ultrasound*. 2017;38(06):648-654. https://doi.org/10.1055/s-0042-108431
9. Zeng J, Liu G-J, Huang Z-P, et al. Diagnostic accuracy of two-dimensional shear wave elastography for the non-invasive staging of hepatic fibrosis in chronic hepatitis B: a cohort study with internal validation. *Eur Radiol*. 2014;24(10):2572-2581. https://doi.org/10.1007/s00330-014-3292-9
10. Yoon JH, Lee JM, Han JK, Cho BI. Shear Wave Elastography for Liver Stiffness Measurement in Clinical Sonographic Examinations. *J Ultrasound Med*. 2014;33(3):437-447. https://doi.org/10.7863/ultra.33.3.437
11. Sporea I, Grădinaru-Taşcău O, Bota S, et al. How many measurements are needed for liver stiffness assessment by 2D-Shear Wave Elastography (2D-SWE) and which value should be used: the mean or median? *Med Ultrason*. 2013;15(4):268-272. http://www.ncbi.nlm.nih.gov/pubmed/24286089. Accessed May 6, 2018.
12. Nadebaum DP, Nicoll AJ, Sood S, Gorelik A, Gibson RN. Variability of Liver Shear Wave Measurements Using a New Ultrasound Elastographic Technique. *J Ultrasound Med*. 2017. https://doi.org/10.1002/jum.14375
13. Ferraioli G, Wong VW-S, Castera L, et al. Liver Ultrasound Elastography: An Update to the World Federation for Ultrasound in Medicine and Biology Guidelines and Recommendations. *Ultrasound Med Biol*. 2018;44(12):2419-2440. https://doi.org/10.1016/j.ultrasmedbio.2018.07.008
14. Song P, Mellema DC, Sheedy SP, et al. Performance of 2-Dimensional Ultrasound Shear Wave Elastography in Liver Fibrosis Detection Using Magnetic Resonance Elastography as the Reference Standard: A Pilot Study. *J Ultrasound Med*. 2016;35(2):401-412. https://doi.org/10.7863/ultra.15.03036
15. Procopet B, Berzigotti A, Abraldes JG, et al. Real-time shear-wave elastography: applicability, reliability and accuracy for clinically significant portal hypertension. *J Hepatol*. 2015;62(5):1068-1075. https://doi.org/10.1016/j.jhep.2014.12.007
16. Lim S, Kim SH, Kim Y, et al. Coefficient of Variance as Quality Criterion for Evaluation of Advanced Hepatic Fibrosis Using 2D Shear-Wave Elastography. *J Ultrasound Med*. 2018;37(2):355-362. https://doi.org/10.1002/jum.14341
17. Cicchetti D V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284-290. https://doi.org/10.1037/1040-3590.6.4.284
18. Goldstein NS, Hastah F, Galan M V., Gordon SC. Fibrosis Heterogeneity in Nonalcoholic Steatohepatitis and Hepatitis C Virus Needle Core Biopsy Specimens. *Am J Clin Pathol*. 2005;123(3):382-387. https://doi.org/10.1309/ey72-f1en-9xcb-1kxx
19. Mulabecirovic A, Mjelle AB, Gilja OH, Vesterhus M, Havre RF. Liver elasticity in healthy individuals by two novel shear-wave elastography systems—Comparison by age, gender, BMI and number of measurements. Strnad P, ed. *PLoS One*. 2018;13(9):e0203486. https://doi.org/10.1371/journal.pone.0203486
20. Xie L-T, Yan C-H, Zhao Q-Y, He M-N, Jiang T-A. Quantitative and noninvasive assessment of chronic liver diseases using two-dimensional shear wave elastography. *World J Gastroenterol*. 2018;24(9):957-970. https://doi.org/10.3748/wjg.v24.i9.957
21. Sigrist RMS, Liau J, Kaffas A El, Chammas MC, Willmann JK. Ultrasound Elastography: Review of Techniques and Clinical Applications. *Theranostics*. 2017;7(5):1303-1329. https://doi.org/10.7150/thno.18650
22. Schellhaas B, Strobel D, Wildner D, Goertz RS, Neurath MF, Pfeifer L. Two-dimensional shear-wave elastography: A new method comparable to acoustic radiation force impulse imaging? *Eur J Gastroenterol Hepatol*. 2017;29(6):723-729. https://doi.org/10.1097/meg.0000000000000846
23. Cassinotto C, Boursier J, de Lédinghen V, et al. Liver stiffness in nonalcoholic fatty liver disease: A comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. *Hepatology*. 2016;63(6):1817-1827. https://doi.org/10.1002/hep.28394
24. Cassinotto C, Lapuyade B, Mouries A, et al. Non-invasive assessment of liver fibrosis with impulse elastography: Comparison of Supersonic Shear Imaging with ARFI and FibroScan®. *J Hepatol*. 2014;61(3):550-557. https://doi.org/10.1016/j.jhep.2014.04.044

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