Thieme

Reliability Criteria for Liver Stiffness Measurements with Real-Time 2D Shear Wave Elastography in Different Clinical Scenarios of Chronic Liver Disease

Reliabilitätskriterien für die Messung der Lebersteifigkeit mittels Echtzeit-2D-Shearwave-Elastografie bei verschiedenen klinischen Szenarien der chronischen Lebererkrankung

Authors

Maja Thiele[1, 2, 3], Bjørn Stæhr Madsen[1, 2, 3], Bogdan Procopet[4], Janne Fuglsang Hansen[5], Linda Marie Sevelsted Møller[1, 2], Sönke Detlefsen[6], Annalisa Berzigotti[7], Aleksander Krag[1, 2]

Affiliations

1 Department of Gastroenterology and Hepatology, Odense University Hospital, Odense, Denmark
2 Institute of Clinical Research, Syddansk Universitet Det Sundhedsvidenskabelige Fakultet, Odense, Denmark
3 OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, Denmark
4 Department of Gastroenterology and Hepatology, Iuliu Hatieganu University of Medicine, Cluj Napoca, Romania
5 Department of Infectious Diseases, Odense University Hospital, Odense, Denmark
6 Department of Pathology, Odense University Hospital, Odense, Denmark
7 University Clinic for Visceral Surgery and Medicine (UVCM), Inselspital, University of Bern, Bern, Switzerland

Key words

elastography, liver, aixplorer, quality criteria, supersonic shear imaging

received 21.12.2015
accepted 26.04.2016

Bibliography

DOI https://doi.org/10.1055/s-0042-108431
Published online: June 07, 2016 | Ultraschall in Med 2017; 38: 648–654 © Georg Thieme Verlag KG, Stuttgart · New York, ISSN 0172-4614

Correspondence

Dr. Maja Thiele
Department of Gastroenterology and Hepatology, Odense University Hospital, Sdr. Boulevard 29, 5000 Odense, Denmark
Tel.: ++ 45/65 41 27 52
Fax: ++ 45/66 11 13 28
maja.thiele@rsyd.dk

ABSTRACT

Purpose Liver stiffness measurement by real-time 2-dimensional shear wave elastography (2D-SWE) lacks universal reliability criteria. We sought to assess whether previously published 2D-SWE reliability criteria for portal hypertension were applicable for the evaluation of liver fibrosis and cirrhosis, and to look for criteria that minimize the risk of misclassification in this setting.

Materials and Methods In a biopsy-controlled diagnostic study, we obtained five 2D-SWE measurements of optimal image quality. Correctly classified cases of fibrosis and cirrhosis were compared to misclassified cases. We compared reliability predictors (standard deviation (SD), SD/mean, size of region of interest (ROI) and difference between a single measurement and the patient's median) with those obtained in a prior study on clinically significant portal hypertension.

Results We obtained 678 2D-SWE measurements from 142 patients. Overall, the variability in liver stiffness within single 2D-SWE measurements was low (SD = 1.1 ± 1.5kPa; SD/mean = 12 ± 9 %). Intra-observer analysis showed almost perfect concordance (intraclass correlation coefficient = 0.95; 95 % CI 0.94 – 0.96; average difference from median = 0.4 ± 0.9kPa). For the diagnosis of cirrhosis, a smaller SD (optimally ≤ 1.75 kPa) and larger ROI size (optimally ≥ 18 mm) were associated with higher accuracy. Similarly, within the published cohort of patients assessed for portal hypertension, a low variability of measurements was associated with high reliability.

Conclusion A high quality 2D-SWE elastogram ensures low variability and high reliability, regardless of indication. We recommend aiming for a combination of low standard deviation and large ROI.

ZUSAMMENFASSUNG

Ziel Bei Messung der Lebersteifigkeit mittels Echtzeit-2D-Shearwave-Elastografie (2D-SWE) fehlen allgemein gültige Reliabilitätskriterien. Wir waren bestrebt herauszufinden,

inwieweit die zuvor publizierten 2D-SWE-Reliabilitätskriterien bei Pfortaderhochdruck auch bei der Bewertung der Leberfibrose und Zirrhose anwendbar sind und suchten nach Kriterien, das Risiko einer Falschdiagnose in diesem klinischen Umfeld zu minimieren.

Material und Methoden In einer zytologisch kontrollierten diagnostischen Studie erzielten wir fünf 2D-SWE-Messungen mit optimaler Bildqualität und korrekt klassifizierten Fibrose- und Zirrhosefälle wurden mit falsch klassifizierten Fällen verglichen. Wir verglichen Prädiktoren für Verlässlichkeit (Standardabweichung (SD), SD/Mittelwert, Größe der „region of interest" (ROI) und Differenz zwischen Einzelmessung und Patientenmedian) mit denen in einer Vorgängerstudie bei klinisch signifikantem Pfortaderhochdruck erhobenen.

Ergebnisse Wir erhielten 678 2D-SWE-Messungen von 142 Patienten. Insgesamt war die Variabilität der Lebersteifigkeit innerhalb der einzelnen 2D-SWE-Messungen gering (SD = 1,1 ± 1,5 kPa; SD/Mittelwert = 12 ± 9 %); die Intraobserver-Analyse zeigte eine nahezu perfekte Übereinstimmung (Intraklasse-Korrelationskoeffizient = 0,95; 95 % CI 0,94 – 0,96; durchschnittliche Differenz zum Median = 0,4 ± 0,9 kPa). Bei der Diagnose einer Zirrhose war ein geringerer SD (optimal ≤ 1,75 kPa) und ein größerer ROI-Bereich (optimal ≥ 18 mm) mit einer höheren Genauigkeit assoziiert. Ähnlich wie in der publizierten Kohorte von Patienten mit Verdacht auf Pfortaderhochdruck bestand ein Zusammenhang zwischen geringerer Variabilität der Messungen und hoher Verlässlichkeit.

Schlussfolgerung Ein 2D-SWE-Elastogramm von hoher Qualität gewährleistet unabhängig von der Indikation niedrige Variabilität und hohe Verlässlichkeit. Wir empfehlen niedrige Standardabweichungen zusammen mit einer großen ROI anzustreben.

Cirrhosis is the eighth most common cause of premature death in the Western world [1] with 75 % of patients only being diagnosed after the development of complications [2]. Ultrasound elastography enables us to stage liver fibrosis in chronic liver disease and allows risk stratification of patients with compensated advanced chronic liver disease [3].

Real-time 2-dimensional shear wave elastography (2D-SWE) measures liver stiffness by combining B-mode ultrasound imaging with an elastogram of the liver. The elastogram is generated by repeated induction of shear waves in the liver tissue. Contrary to point shear wave elastography techniques, the elastogram in 2D-SWE is displayed continuously in real-time and is color-coded, and the circular region of interest (ROI) can be moved and its size can be changed (supplementary video https://youtu.be/0yb9Fw0lDXE). The result of a 2D-SWE measurement is displayed as the mean liver stiffness and its standard deviation (SD) within the single ROI. The result of a 2D-SWE exam can subsequently be reported as the mean or median of any number of single measurements.

2D-SWE diagnoses liver fibrosis and cirrhosis with high accuracy [4 – 10]. Three studies have also assessed the diagnostic accuracy of 2D-SWE for clinically significant portal hypertension (CSPH, portal pressure gradient ≥ 10 mmHg) [11 – 13]. However, due to a lack of universally agreed-upon reliability criteria for 2D-SWE, the quality appraisal of measurements in diagnostic studies ranges from no quality criteria described [6, 8, 9] to "when scanning conditions permit" [4, 5] and homogeneous color-coded elastogram required [7, 10]. As a consequence, failure rates range from 10 % [5] to 1.7 % [8].

Lack of objective reliability criteria may cause uncertainty regarding the validity of results and the comparability of studies, particularly in competition with other noninvasive markers [14, 15].

It has been suggested that low variability of measurements, expressed as SD/mean, could be used as a reliability criterion similar to that described for transient elastography [16]. In a study of 2D-SWE to diagnose CSPH in cirrhosis [13], an SD/mean less than 10 % resulted in higher diagnostic accuracy. We therefore aimed to validate this proposed reliability criteria in a cohort of patients investigated with 2D-SWE for the presence of liver fibrosis and early cirrhosis. We also aimed at evaluating whether other 2D-SWE characteristics predicted reliability, in order to explore other potential objective reliability criteria for use in different clinical scenarios of liver disease. Finally, we aimed to suggest how many measurements should be obtained for optimal reporting of results of a 2D-SWE exam and whether it is optimal to report the mean or the median of measurements.

Methods

The regional ethics committees approved the study protocol (S-20 130 071, S-20 140 070; HCB/2014/0501). The study adhered to the declaration of Helsinki and all subjects gave oral and written consent before study inclusion.

Study populations

From April 2014 to June 2015, we consecutively recruited patients aged 18 – 80 years with alcoholic liver disease (ALD) or chronic viral hepatitis C (CHC) recruited from one municipal alcohol rehabilitation center and four hospital liver clinics in the region of Southern Denmark. The ALD patients described here form a subgroup of patients in a diagnostic test study of longer duration [17].

For comparison of reliability criteria of 2D-SWE between different indications, we used data from 69 patients with compensated (n = 55) or decompensated (n = 24) liver disease in whom the role of 2D-SWE to diagnose CSPH was assessed. The clinical characteristics of the included patients have been described in detail elsewhere [13]. CSPH was defined as hepatic venous pressure gradient (HVPG) ≥ 10 mmHg. In short, two criteria of liver stiffness by 2D-SWE were associated with misclassification of CSPH: SD/mean > 10 % and acquisition depth ≥ 5.6 cm [13]. The ROI was fixed at 15 mm.

Investigations at inclusion

Patients were assessed in fasting conditions, on a one-day visit, during which 2D-SWE (Aixplorer, Supersonic Imagine, France) and liver biopsy were performed.

One of two experienced ultrasonographers (MT and BSM) performed 2D-SWE according to previously described methods [17]. 2D-SWE measurements were considered valid when the elastogram was stable for at least three seconds before image acquisition and the ROI was homogeneously color-coded. We aimed to obtain five separate 2D-SWE measurements at a depth of less than 5.6 cm [13, 18].

We performed the liver biopsy in the same intercostal space as the elastography (Menghini method, 17G suction needle, Hepafix, Braun, Germany). Following the biopsy procedure, the samples were immediately stored in formalin 4 % and embedded in paraffin. Sections with a thickness of 4 µm were stained with sirius red for METAVIR grading by one experienced liver pathologist. The biopsy quality criterion was length of at least 10 mm or at least six portal tracts, other than in the case of regeneration nodules characteristic of cirrhosis.

Statistical analyses

We express summary statistics as medians with interquartile ranges (IQR) or counts and frequencies, with the Wilcoxon's rank sum test for group comparisons of descriptive statistics.

To label correct classifications and misclassification, we used optimal 2D-SWE cut-off values for METAVIR fibrosis stages ≥ F2, ≥ F3 and = F4, calculated according to the etiology of liver disease. The cut-off values were calculated from the mean of five measurements by maximizing the Youden Index using the non-parametric area under the receiver operating characteristics curve (AUC) with liver biopsy as the gold standard. For analysis of how reliability criteria affected the AUCs, we used single measurements with the DeLong test to test the equality of ROC areas. ROC curve comparison was also used to test whether AUCs significantly changed when reporting the mean or the median of 2D-SWE measurements, or when using three or five measurements for calculating the mean liver stiffness.

A 2D-SWE measurement was labelled as a misclassification if the measurement indicated a lower or higher fibrosis stage than what was revealed by liver histology or if it wrongfully indicated whether the patient had CSPH or not.

To test which 2D-SWE variables predicted correct classification of fibrosis stages, we used a uni- and multivariable mixed effects logistic regression with case identifier as a random effect modifier to adjust for multiple testing. The model included SD/mean, average absolute difference of a single measurement from the patient's median (in case of three or more valid measurements), SD and ROI size. After determining which predictors correlated with correct classifications, we determined the optimal cut-off for that predictor by optimizing the Youden Index from a receiver operating characteristics curve with classification status as the dependent variable.

Intraobserver variability was assessed by Bland-Altman plots and intraclass correlation coefficient using a two-way mixed effects, consistency of agreement, model. All statistical analyses were performed with the statistical software STATA 14 (Statacorp, TX, US).

▶ **Table 1** Patient characteristics.

male/female	92 (65 %)/50 (35 %)
age	53 ± 12 years
etiology	111 (78 %) alcohol 31 (22 %) hepatitis C
METAVIR fibrosis stages	F0 = 15 (11 %) F1 = 67 (47 %) F2 = 20 (14 %) F3 = 11 (8 %) F4 = 29 (20 %)
child-pugh class in patients with cirrhosis	19 (66 %) child-pugh A 10 (34 %) child-pugh B
BMI	26 ± 5 kg/m^2
mean 2D-SWE	8.9 ± 9.4 kPa
drinking pattern	
abstinent	76 (54 %)
ongoing drinking	66 (46 %)
standard deviation/mean	
< 10 %	41 (29 %)
11 – 20 %	74 (52 %)
21 – 30 %	25 (18 %)
> 30 %	2 (1 %)
standard deviation	
0.25 – 1.0 kPa	55 (39 %)
1.1 – 1.75 kPa	42 (29 %)
> 1.75 kPa	45 (32 %)
diameter of region of interest	
13 – 17 mm	65 (46 %)
18 – 23 mm	77 (54 %)

2D-SWE: real-time 2-dimensional shear wave elastography; BMI: body mass index.

Results

Patient characteristics

144 of 234 screened patients agreed to participate in the study. 2 had to be excluded after liver biopsy due to insufficient biopsy material and failure to obtain any valid 2D-SWE measurements. Of the final 142 patients (111 ALD, 31 CHC), 92 were men (65 %). The mean age was 53 ± 12 years and the mean BMI was 26 ± 5 kg/m^2 (▶ **Table 1**). ALD patients had lower levels of alanine transaminase (41 vs. 88 U/L, P = 0.05) and milder disease than CHC patients (METAVIR fibrosis stages 0 – 4 = 15/57/13/7/19 in ALD vs. 0/10/7/4/10 in CHC, P< 0.001).

Characteristics of real-time 2-dimensional shear wave elastography

We acquired a total of 678 2D-SWE measurements: 5 2D-SWE measurements were obtained from 130 patients while 4 or less valid measurements were obtained from 12 patients due to obesity, which led to an insufficient acoustic signal to meet the subjective quality criteria for 2D-SWE.

The SD/mean was 12 ± 9 % and the median SD was 1.1 ± 1.5 kPa (▶ **Table 1**). The SD increased in a step-wise manner from METAVIR F0 – 1 to F2, F3 and F4. In contrast, the SD/mean was largely independent of the fibrosis stage (▶ **Fig. 1**). The ROI size averaged 19 ± 4 mm, independent of the fibrosis stages.

Liver stiffness measurements varied little within patients: On average, there was a 0.5 ± 0.9 kPa within-patient difference between single measurements and the median of 5 measurements. In accordance with the low level of variance, intraobserver agreement was high with an intraclass correlation coefficient of 0.95 (95 % CI 0.94 – 0.96) (▶ **Fig. 2**).

Diagnostic accuracy and optimal reporting of 2D-SWE

2D-SWE had excellent accuracy for the diagnosis of significant fibrosis (≥F2), severe fibrosis (≥F3) and cirrhosis (= F4) (AUC_{F2} = 0.93, 0.90 – 0.97; AUC_{F3} = 0.93, 0.88 – 0.97; AUC_{F4} = 0.94, 0.90 – 0.98, respectively).

2D-SWE correctly classified METAVIR stage ≥F2 in 591 measurements and METAVIR ≥F3 in 586 out of 678 measurements (▶ **Table 2**). All 5 measurements correctly classified the fibrosis stage in 40 patients, while all measurements misclassified the fibrosis stage in another 40 patients. The remaining 62 patients had a mixture of correct classifications and misclassifications.

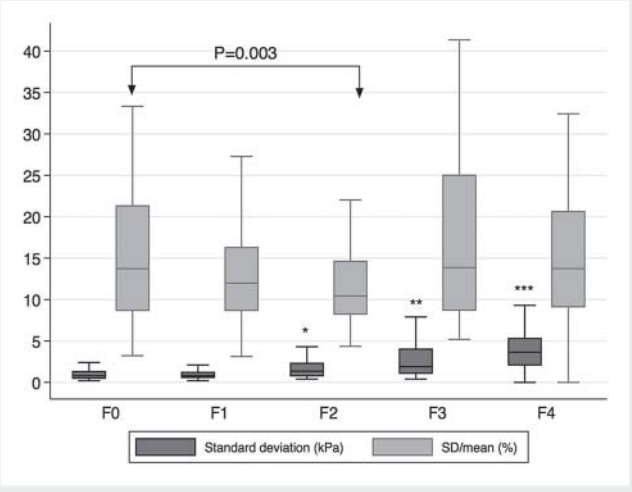
For significant and severe fibrosis, but not for cirrhosis, reporting the mean of the measurements yielded a slightly higher diagnostic accuracy than reporting the median (AUC_{F2} difference 0.008, P = 0.024; AUC_{F3} difference 0.006, P = 0.038; AUC_{F4} difference 0.004, P = 0.225). There was no difference in diagnostic accuracy between reporting the mean of all 5 measurements or the mean of only the first 3 measurements (P≥ 0.4).

Reliability criteria for 2 D shear wave elastography to diagnose liver fibrosis and cirrhosis

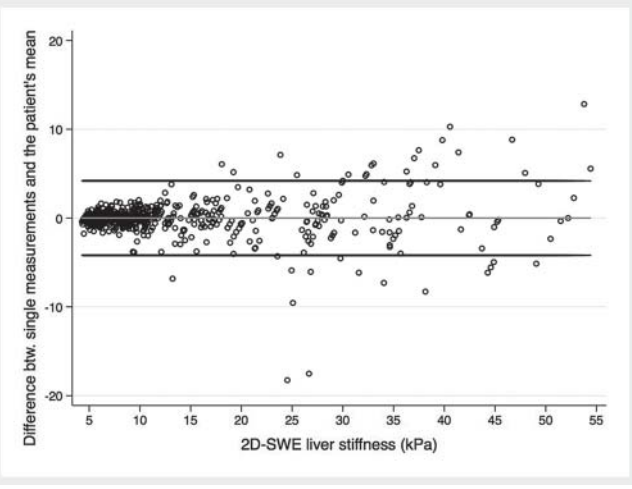
A lower SD and larger ROI independently correlated with correct classifications of severe fibrosis and cirrhosis in a mixed regression analysis adjusting for multiple measurements per patient (all P< 0.045), while no reliability predictors correlated with correct classification of significant fibrosis.

When using the mean of all 5 measurements, a lower SD still correlated with correct classifications of cirrhosis (P = 0.002), while the result was only borderline significant for a larger ROI (P = 0.072) and there was no correlation with correct classifications of severe fibrosis.

For cirrhosis, correctly classified measurements on average diverged 0.3 kPa from the median, while misclassifications diverged 1.3 kPa (P< 0.001) (▶ **Table 2**). However, this finding was not robust with respect to adjustment for multiple measurements



▶ **Fig. 1** 2D-SWE standard deviation and SD/mean ratio according to METAVIR fibrosis grades. Standard deviation increased in a step-wise manner from F0 – 1 to F2, F3 and F4 (symbolizes a significant difference between groups). SD/mean was independent of METAVIR grade, except for a higher SD/mean in F0 compared to F2.



▶ **Fig. 2** Bland-Altman plot of differences between individual 2D-SWE measurements and the patient's mean 2D-SWE

(P = 0.281). We did not find any association between reliability and SD/mean below 10 % in the present cohort.

Despite their statistical significance, the associations between reliable measurements and SD or ROI size had only moderate clinical implications: The risk of misclassifying severe fibrosis increased 2 % for every 1 kPa increase in SD and 1 % for every 1 mm decrease in ROI. The risk of misclassifying cirrhosis increased 3 % and 0.6 % for every 1 kPa increase in SD and 1 mm decrease in ROI, respectively.

SD was the strongest predictor of reliability (accuracy of SD to predict correct classification of cirrhosis, AUC = 0.80, compared to ROI size, AUC = 0.66; P = 0.002).

The optimal SD cut-off value for correct classification of cirrhosis was 1.75 kPa (1.35 kPa for severe fibrosis). In addition, 21 % of

▶ **Table 2** Characteristics of 2D-SWE measurements that correctly classify or misclassify METAVIR fibrosis stages ≥ F2, ≥ F3 and = F4.

significant fibrosis METAVIR ≥F2	**correct classifications**	**misclassifications**	**p-value**
number of measurements	591 (87 %)	87 (13 %)	
SD/mean (%)	12 ± 9	11 ± 9	0.420
standard deviation (kPa)	1.1 ± 1.7	1.1 ± 1.1	0.393
region of interest (mm)	19 ± 4	18 ± 4	0.338
deviation from patient's median[1] (kPa)	0.4 ± 1.0	0.2 ± 0.8	0.062
deviation/mean[2] (%)	4 ± 9	3 ± 8	0.078
severe fibrosis METAVIR ≥F3	**correct classifications**	**misclassifications**	**p-value**
number of measurements	586 (86 %)	92 (14 %)	
SD/mean (%)	12 ± 9 %	10 ± 7 %	0.024
standard deviation (kPa)	1.1 ± 1.4 kPa	1.7 ± 1.4 kPa	< 0.001
region of interest (mm)	19 ± 4 mm	18 ± 4 mm	0.012
deviation from patient's median[1] (kPa)	0.4 ± 0.8 kPa	0.5 ± 1.2 kPa	0.164
deviation/mean[2] (%)	4 ± 9 %	4 ± 10 %	0.660
cirrhosis METAVIR = F4	**correct classifications**	**misclassifications**	**p-value**
number of measurements	614 (91 %)	62 (9 %)	
SD/mean (%)	12 ± 9 %	12 ± 15 %	0.371
standard deviation (kPa)	1.1 ± 1.2 kPa	2.9 ± 2.4 kPa	< 0.001
region of interest (mm)	19 ± 4 mm	17 ± 3 mm	< 0.001
deviation from patient's median[1] (kPa)	0.3 ± 0.8 kPa	1.3 ± 3.3 kPa	< 0.001
deviation/mean[2] (%)	4 ± 9 %	5 ± 13 %	0.146

Results are given as median ± IQR. P-values from the Wilcoxon rank sum test denote significant between-group differences. Based on optimal cut-offs: ≥ F2 = 10.2kPa for CHC and 10.7kPa for ALD; ≥ F3 = 11.1kPa for CHC and 10.7kPa for ALD; for F4 = 14.5kPa for CHC and 16.8 kPa for ALD. 2D-SWE: real-time 2-dimensional shear wave elastography; SD: standard deviation.

[1] Difference between a single liver stiffness measurement and the patient's median (five measurements).

[2] Ratio of highest deviation over patient's overall mean.

measurements with SD > 3.2 kPa misclassified patients (1/108 incorrectly classified as not having cirrhosis, and 22/108 incorrectly classified as having cirrhosis). An ROI above or equal to 18 mm was the optimal cut-off for correct classifications of both severe fibrosis and cirrhosis.

The diagnostic accuracy of 2D-SWE increased in a step-wise manner when using an SD below 1.75 kPa and an ROI above 18 mm as reliability criteria (▶ **Table 3**). With both criteria met, the AUC of 2D-SWE to diagnose cirrhosis was 0.99 compared to an AUC of 0.75 when none of the reliability criteria were met. Since SD was the strongest reliability predictor, the diagnostic accuracy for cirrhosis decreased more if the SD was high than if the ROI size was small. The same pattern was observed for severe fibrosis (data not shown).

Comparison with reliability criteria for 2D-SWE to diagnose clinically significant portal hypertension

In the 69 patients belonging to the previously published study [13], 2D-SWE correctly classified 57 patients with (n = 34) or without (n = 23) CSPH, while 12 patients were misclassified (8 as not having CSPH when they did, and 4 as having CSPH when they did not) by 2D-SWE.

In comparison to the present cohort, patients included in the CSPH study had higher liver stiffness (17.2 ± 9.7 kPa), higher SD (1.8 ± 1.2 kPa) and higher within-patient difference of individual measurements (1.8 ± 1.0 kPa) due to the different clinical scenario. As such, the reliability criteria for fibrosis and cirrhosis diagnosis described above had a very low applicability and were not suitable for the population assessed for CSPH.

However, similarly to what we observed in the present cohort, misclassified patients had 2D-SWE measurements with a larger SD compared to well classified patients (SD = 2.73 ± 1.19 kPa in correctly classified measurements vs. SD = 2.43 ± 1.43 kPa in misclassified measurements, P = 0.08). 90 % of misclassified patients had an SD above 3.2 kPa. As such, an SD ≥ 3.2 kPa was almost invariably associated with misclassifications both for the diagnosis of cirrhosis and for the diagnosis of CSPH.

▶ **Table 3** Accuracy of 2D-SWE for diagnosing cirrhosis according to which reliability criteria are met.

	AUC (95 % confidence interval)	number of reliable/unreliable measurements that meet the criteria	proportion of measurements that meet the criteria	P[1]
SD≤ 1.75 kPa and ROI≥ 18 mm	0.99 (0.97 – 1.00)	298/3	301/678 (44 %)	–
SD≤ 1.75 kPa and ROI< 18 mm	0.94 (0.90 – 0.98)	152/11	163/678 (24 %)	0.032
SD> 1.75 kPa and ROI≥ 18 mm	0.83 (0.75 – 0.92)	102/22	124/678 (18 %)	0.024
SD> 1.75 kPa and ROI< 18 mm	0.75 (0.65 – 0.85)	62/28	90/678 (13 %)	0.215

AUC: area under the receiver operating characteristics curve; ROI: region of interest; SD: standard deviation.
[1] ROC curve comparison between groups.

Discussion

In this study of 2D-SWE reliability criteria, we found that a homogeneous elastogram with color stability for 2 – 4 seconds ensures low variation in measurements and high reliability, regardless of whether the clinical scenario is diagnosis of liver fibrosis or diagnosis of portal hypertension.

In the present cohort as well as in the previously published study [13], markers of low variance were associated with reliable measurements: An SD below 1.75 kPa and an ROI above 18 mm in diameter significantly increased the accuracy of a 2D-SWE measurement for correctly classifying cirrhosis, while an SD/mean below 10 % and a depth of measurement < 5.6 cm were important for high reliability when using 2D-SWE to assess patients for portal hypertension [13]. We confirmed that a larger SD was also associated with misclassifications for CSPH evaluation, and that measurements with an SD above 3.2 kPa were likely to be unreliable for both liver fibrosis staging and CSPH assessment. The role of ROI size and measurement depth could not be compared across the two cohorts, because the present study measured at a depth above 5.6 cm and because the previously published study used a fixed ROI of 15 mm.

Our results do not support the use of strict binary reliability criteria, as an SD slightly higher than 1.75 kPa and an ROI slightly lower than 18 mm were still associated with a high rate of correct classifications, and only 44 % of measurements fulfilled both criteria. However, the higher the SD and the lower the ROI, the higher the risk of misclassifications was.

Differences in liver disease severity between the two cohorts and differences in the inclusion criteria between the present study and the previously published study [13] might explain the difference in the strongest predictors of reliability. However, 2D-SWE is innately less reliable for staging fibrosis than for diagnosing CSPH, since liver biopsy is a worse gold standard for fibrosis than HVPG for CSPH [19]. Biopsy sampling error in addition to false-positive 2D-SWE measurements may consequently have caused some of the misclassifications, irrespective of reliability predictors. A larger cohort may therefore be needed to detect robust reliability criteria for 2D-SWE to stage fibrosis. In a larger cohort it would also be possible to adjust analyses more accurately for a potential correlation bias from multiple measurements in the same patient.

Similarly to what was suggested in the CSPH cohort, there was no difference between obtaining three or five 2D-SWE measurements in our series. This is likely explained by a very low intraobserver variance, which is comparable to other studies in healthy controls [20] and CHC patients [8]. Reporting the mean rather than the median resulted in marginally better diagnostic accuracy, which is in line with another study using transient elastography as a reference [18].

In conclusion, our study highlights the importance of selecting an elastogram with the highest possible image quality for 2D-SWE measurements. This optimizes reliability, ensures low variance and yields the highest possible diagnostic accuracy for the evaluation of liver fibrosis as well as portal hypertension. By combining the results obtained in the two analyzed series, low variability of the measurements (absolute SD or SD/mean) with a larger ROI and optimal measurement depths is associated with more reliable results in different clinical scenarios. We suggest the use of these reliability criteria as a guide for the appraisal of 2D-SWE measurements in the setting of chronic liver diseases.

Grant support

The study was investigator-initiated and partly funded by the Danish National Advanced Technology Foundation and Innovation Fund Denmark. The Supersonic Aixplorer and FibroScan XL-probe were acquired with grants from the A. P. Moeller Foundation and Toyota Foundation. Working grants from the University of Southern Denmark, Odense University Hospital and Region of Southern Denmark support doctors MT, BSM and JFH.

Author contributions

MT, BSM and AK conceptualized and designed the study. All authors acquired data for the study. MT and AB analyzed the data. MT, AB and AK interpreted the data. MT drafted the manuscript. All authors revised the work for important intellectual content. All authors have approved the final manuscript.

ABBREVIATIONS USED IN THIS PAPER

2D-SWE	real-time 2-dimensional shear wave elastography
AUC	area under the receiver operating characteristics curve
BMI	body mass index
CSPH	clinically significant portal hypertension
GGT	gamma-glutamyltransferase
HVPG	hepatic venous pressure gradient
IQR	interquartile range
kPa	kilopascal
ROI	region of interest
SD	standard deviation
TE	transient elastography

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. The Lancet 2015; 385: 117 – 171

[2] Fialla AD, de Muckadell OBS, Touborg Lassen A. Incidence, etiology and mortality of cirrhosis: a population-based cohort study. Scandinavian journal of gastroenterology 2012; 47: 702 – 709

[3] de Franchis R. Expanding Consensus in Portal Hypertension – Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. J Hepatol 2015; 63: 743 – 752

[4] Ferraioli G, Tinelli C, Dal Bello B et al. Accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. Hepatology 2012; 56: 2125 – 2133

[5] Poynard T, Munteanu M, Luckina E et al. Liver fibrosis evaluation using real-time shear wave elastography: applicability and diagnostic performance using methods without a gold standard. J Hepatol 2013; 58: 928 – 935

[6] Zheng J, Guo H, Zeng J et al. Two-dimensional Shear-Wave Elastography and Conventional US: The Optimal Evaluation of Liver Fibrosis and Cirrhosis. Radiology 2015; 275: 290 – 300

[7] Bota S, Paternostro R, Etschmaier A et al. Performance of 2-D Shear Wave Elastography in Liver Fibrosis Assessment Compared with Serologic Tests and Transient Elastography in Clinical Routine. Ultrasound Med Biol 2015; 41: 2340 – 2349

[8] Deffieux T, Gennisson JL, Bousquet L et al. Investigating liver stiffness and viscosity for fibrosis, steatosis and activity staging using shear wave elastography. J Hepatol 2015; 62: 317 – 324

[9] Tada T, Kumada T, Toyoda H et al. Utility of real-time shear wave elastography for assessing liver fibrosis in patients with chronic hepatitis C infection without cirrhosis: Comparison of liver fibrosis indices. Hepatology research. 2015; 45: 122 – 129

[10] Yoneda M, Thomas E, Sclair SN et al. Supersonic Shear Imaging and Transient Elastography With the XL Probe Accurately Detect Fibrosis in Overweight or Obese Patients With Chronic Liver Disease. Clin Gastroenterol Hepatol 2015; 13: 1502 – 1509

[11] Elkrief L, Rautou PE, Ronot M et al. Prospective Comparison of Spleen and Liver Stiffness by Using Shear-Wave and Transient Elastography for Detection of Portal Hypertension in Cirrhosis. Radiology 2015; 275: 589 – 598

[12] Kim TY, Jeong WK, Sohn JH et al. Evaluation of portal hypertension by real-time shear wave elastography in cirrhotic patients. Liver international: official journal of the International Association for the Study of the Liver 2015; 35: 2416 – 2424

[13] 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: 1068 – 1075

[14] Cassinotto C, de Lédinghen V. Reply to: "New imaging assisted methods for liver fibrosis quantification: Is it really favorable to classical transient elastography?". J Hepatol 2015; 63: 767

[15] Cassinotto C, Lapuyade B, Mouries A et al. Noninvasive assessment of liver fibrosis with impulse elastography: comparison of Supersonic Shear Imaging with ARFI and Fibroscan. J Hepatol 2014; 61: 550 – 557

[16] Boursier J, Zarski JP, de Ledinghen V et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. Hepatology 2013; 57: 1182 – 1191

[17] Thiele M, Detlefsen S, Møller L et al. Transient and 2-dimensional shear-wave elastography provide comparable assessment of alcoholic liver fibrosis and cirrhosis. Gastroenterology 2016; 150: 123 – 133

[18] Sporea I, Gradinaru-Tascau 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? Medical ultrasonography 2013; 15: 268 – 272

[19] Gluud C, Brok J, Gong Y et al. Hepatology may have problems with putative surrogate outcome measures. J Hepatol 2007; 46: 734 – 742

[20] Ferraioli G, Tinelli C, Zicchetti M et al. Reproducibility of real-time shear wave elastography in the evaluation of liver elasticity. European journal of radiology 2012; 81: 3102 – 3106