European Radiology
https://doi.org/10.1007/s00330-020-07212-x

GASTROINTESTINAL

Comparison of liver stiffness measurements by a 2D-shear wave technique and transient elastography: results from a European prospective multi-centre study

Maxime Ronot[1,2] · Giovanna Ferraioli[3,4] · Hans-Peter Müller[5] · Mireen Friedrich-Rust[6] · Carlo Filice[3,4] · Valérie Vilgrain[1,2] · David Cosgrove[7] · Adrian K. Lim[7]

Received: 24 April 2020 / Revised: 13 July 2020 / Accepted: 19 August 2020

Abstract

Objectives To compare liver stiffness measurement (LSM) provided by Canon 2D-shear wave elastography (2D-SWE) and transient elastography (TE), the latter being the reference method.

Methods Prospective study conducted in four European centres from 2015 to 2016 including patients with various chronic liver diseases who had LSMs with both 2D-SWE and TE on the same day. Median of 10 valid measurements (in kPa) was used for comparison using paired *t* test, Pearson correlation, intraclass correlation coefficient (ICC) and Bland-Altman plot. The ability of 2D-SWE to stratify patient according to recognised LSM-TE thresholds was assessed by ROC curve analysis.

Results Six hundred forty patients were scanned, where 593 (92.7%), 572 (89.4%) and 537 (83.9%) had reliable LSMs by TE, 2D-SWE and both combined, respectively. In the latter (n = 537, 310 [57.7%] male, mean 55.3 ± 14.8 years), median LSM-TE and LSM-2D-SWE had a mean of 10.1 ± 9.4 kPa (range 2.4–75) and 9.1 ± 6.1 kPa (range 3.6–55.7) (paired *t* test: $p < 0.001$), respectively. These were significantly correlated (Pearson $r = 0.932$, $p < 0.001$, ICC 0.850 (0.825–0.872), bias 0.99 ± 4.33 kPa [95% limits of agreement − 9.48 to + 7.49] with proportional error towards higher LSM values). LSM-2D-SWE values significantly increased with TE categories (ANOVA: $p < 0.001$). AUROCs ranged from 0.935 ± 0.010 (95% CI 0.910–0.954) to 0.973 ± 0.009 (95% CI 0.955–0.985), resulting in correct classification of 390/537 (73%) patients. Three 2D-SWE measurements were sufficient for reliable LSMs.

Conclusion LSM using 2D-SWE correlates well with TE. It tends to underestimate higher stages of liver fibrosis but correctly classifies the majority of patients. It may be used in TE-derived algorithms to manage patients.

Key Points

- *Liver stiffness measurement (LSM) by 2D-shear wave elastography (2D-SWE) and transient elastography (TE) are strongly correlated.*
- *2D-SWE shows proportionately lower LSM values compared to TE, particularly with the higher LSM range.*
- *Three individual measurements by 2D-SWE are sufficient to assess LSM reliably.*

Keywords Elasticity imaging technique · Liver diseases · Prospective studies · Portal hypertension · Fibrosis

David Cosgrove had already passed away.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-020-07212-x) contains supplementary material, which is available to authorized users.

✉ Maxime Ronot
maxime.ronot@aphp.fr

[1] Department of Radiology, Beaujon University Hospital, APHP.Nord, Clichy, France

[2] Université de Paris, Paris, France

[3] Department of Clinical Sciences and Infectious Diseases, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

[4] Department of Clinical, Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Pavia, Italy

[5] Department of Hepatology and Imaging, Charité Hospital, University of Berlin, Berlin, Germany

[6] Department of Internal Medicine 1, Division of Gastroenterology and Hepatology, Goethe University Hospital, Frankfurt, Germany

[7] Department of Imaging, Imperial College Healthcare NHS Trust, Charing Cross Hospital, Fulham Palace Road, London W6 8RF, UK

Published online: 09 September 2020

Abbreviations

CI	Confidence interval
HBV	Hepatitis B virus
HCV	Hepatitis C virus
ICC	Intraclass correlation coefficient
kPa	Kilopascal
LOA	Limit of agreement
LSM	Liver stiffness measurement
NAFLD	Non-alcoholic fatty liver disease
ROC	Receiver operating characteristic
SWE	Shear wave elastography
TE	Transient elastography

Introduction

Chronic liver disease is a worldwide epidemic of various causes. In the Western world, the main aetiologies include viral hepatitis (hepatitis B virus (HBV) or hepatitis C virus (HCV)), alcohol-related liver disease (ALD) and non-alcoholic fatty liver disease (NAFLD) [1–3]. All cause progressive deposition of fibrosis that often progresses to cirrhosis after 15–20 years. Occurrence of clinically significant portal hypertension is one of the main complications of chronic liver diseases and is associated with the development of gastric and oesophageal varices and with first clinical decompensation in patients with no varices [4]. The accurate estimation of the stage of liver fibrosis and of portal hypertension is important for ascertaining prognosis, surveillance and treatment. Reference diagnostic methods, namely liver biopsy and catheter-directed measurement of the hepatic venous pressure gradient (HVPG), are minimally invasive and therefore associated with possible morbidity. Moreover, they often require hospitalisation. Finally, HVPG measurement is possible in expert centres only, limiting the availability of the test.

Owing to these reasons, there has been a significant drive by the hepatology and radiology communities to develop the use of non-invasive tools to aid characterisation of chronic liver disease. The most widely used being the liver stiffness measurement (LSM) using transient elastography (TE) that uses an external mechanical push to generate shear waves [5, 6]. LSM by TE has been shown to strongly correlate with the degree of liver fibrosis [7] and with HVPG, and to be useful for detection of CSPH [8, 9]. As a consequence, LSM by TE has been progressively implemented in clinical algorithms for patient triage and monitoring and is now commonly accepted in routine clinical practice. Nevertheless, measurement of LSM by TE needs a dedicated device that lacks the versatility of the ultrasound systems.

Technologies based on the acoustic radiation force impulse (ARFI), i.e. the push-pulse of the ultrasound beam, have also shown much promise. ARFI-based methods, either point shear wave elastography (pSWE) or 2D-shear wave elastography (2D-SWE), have been developed and continually refined to assess LSM and give a handle on liver fibrosis and portal hypertension. 2D-SWE utilises 2D imaging, providing both visual colour maps and quantitative analysis. This technology has been implemented on ultrasound scanners by several manufacturers. Studies have shown that 2D-SWE solutions of different manufacturers may perform differently and provide values that differ from those obtained using TE [5, 6]. Herrmann et al [10] published a large retrospective multicentre study suggesting that 2D-SWE has equivalent to slightly better performance compared with TE for the non-invasive staging of liver fibrosis. Noticeably, only few prospective studies using various ARFI-based technologies have been published to validate this [11–13].

The primary aim of the current study was to prospectively compare liver stiffness measurements provided by Canon 2D-SWE and TE, the latter being considered the reference method. The secondary aim was to assess the influence of the number of LSM measurements on 2D-SWE performance.

Materials and methods

Patient population and inclusion criteria

This prospective study was approved by institutional review boards of all centres involved in the study. The protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki. Patients gave written informed consent. Four European centres were involved in this study (Clichy, France; Pavia Italy; Berlin, Germany; Frankfurt, Germany). All centres were referral centres for liver diseases with long-standing experience using TE and 2D-SWE. Between January 2015 and July 2016, patients were recruited prospectively when referred for a TE study either as part of a routine follow-up in the context of a history of chronic liver disease or for assessment of suspected chronic liver disease. Patients were excluded if they were on antiviral treatments and had a history of cardiac disease or failure or biliary obstruction.

Patient demographics including age, gender, body mass index and disease status as well as any other comorbidity were recorded. Available liver biopsy was not mandatory for patient inclusion. Figure 1 shows the flow chart of the study.

Study design and scanning protocol

Before the initiation of the study, a preliminary meeting was set up between all centres to agree on a common protocol. All centres had experience at using TE. All the physicians involved in the study had experience in 2D-SWE with other US systems. They all underwent a short training period for Canon 2D-SWE examination under the supervision of an application specialist. Canon 2D-SWE was utilised for at least 3

months in all centres prior to recruiting patients for the study. All centres had full applications support by the Canon Medical Systems shear wave elastography (SWE) team.

Equipment

All centres used an Aplio 500 Platinum (Canon Medical Systems former Toshiba Medical Systems) with a 6C1 low-frequency probe, version 6.0 software and standardised presets. All centres also used the latest version of TE (FibroScan® [Echosens, 502 touch]) with access to the standard M and XL probes for measurements.

Scanning protocol

After inclusion, all patients underwent both TE and 2D-SWE of the liver on the same day. All patients were fasted for at least 4 h prior to the examination and examined in the supine position with the right hand raised above their head. It was ensured that the studies were performed by different operators for 2D-SWE and for TE at each site where either study could be performed first. Operators were blinded to the results of the other exams. TE and 2D-SWE screens were placed so that patients could not see the results of the LSMs. Patients were also not informed of the results of LSMs from the first examination prior to the second one being performed.

Transient elastography

Transient elastography was performed by physicians or experienced hepatology nurses with more than 500 exams performed before this study to measure LSM in the right lobe of liver with the probe placed within the mid axillary line. The M probe was used when the skin-to-liver capsule distance was ≤ 25 mm; otherwise, the XL probe was used. Reliable LSM measurements were defined as the median value of 10 consecutive measurements and an interquartile range less than 30% of the median liver stiffness measurements. LSM was expressed in kilopascals (kPa).

2D-SWE

Patients were instructed to hold their breath at mid expiration for the 2D-SWE scan acquisition. The right intercostal approach was chosen, and care was taken to ensure that the liver capsule was parallel to the probe to avoid anisotropy. Shear wave measurement scans were acquired within a 'shear wave box' (or elastogram) size of 3 × 3 cm positioned at least 1 cm below the liver capsule and overlaid on B-mode image. The colour spectrum was adapted to avoid the saturation of the values. LSM and shear wave (SW) propagation maps were observed by real-time visualisation, and measurements were not taken until signal stabilisation of the colour image, to ensure the quality of the measurement. For the LSM measurement, a circular ROI with 1 cm diameter was placed inside the shear wave box in an area where the colour signal was stable over time with optimal propagation detection (Supplemental Fig. 1). If the colour box was not filled by > 50% of its surface or if breathing was uncontrolled, the elastogram was discarded and a new acquisition attempted. Ten single LSMs were obtained, and the median in kPa was used for analysis. The LSM was considered reliable if the interquartile range was less than 30% of the median LSMs, as per the recommended TE standard guidelines. The distance from the skin to the liver surface was also recorded.

Data collection and storage

All LSMs by TE and 2D-SWE were documented. All the data from each centre were transferred to a secure central storage and input into an Excel spreadsheet which was later used for statistical analysis.

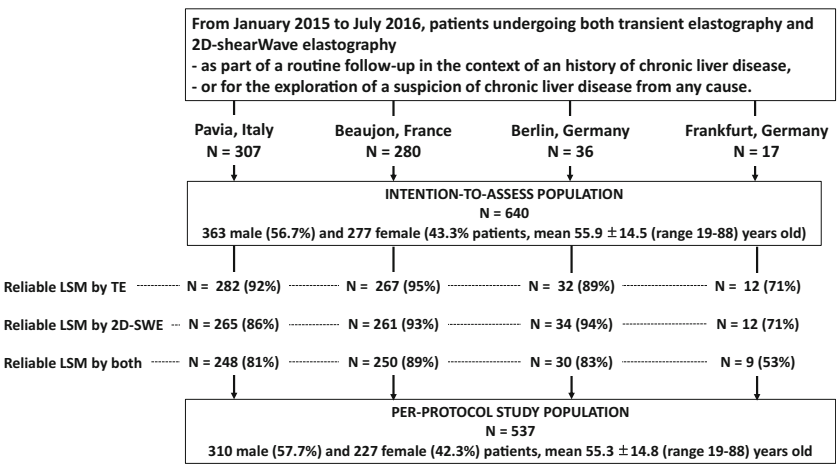

Fig. 1 Flow chart of the study. *LSM* liver stiffness measurement, *TE* transient elastography, *SWE* shear wave elastography

Table 1 Characteristics of the 537 patients with reliable liver stiffness measurement by both transient elastography and 2D-shear wave elastography (2D-SWE)

Variable	Value	Median LSM by 2D-SWE (range)	*p* value
Gender			0.085
Male (%)	310 (57.7)	9.48 (4.15–54.40)	
Female (%)	227 (42.3)	8.60 (3.60–55.70)	
Mean age ± SD (range) years	55.3 ± 14.8 (19–88)		
< 30	25 (4.7)	6.00 (4.95–17.00)	< 0.001
30–49	161 (29.9)	6.10 (4.15–43.55)	
50–69	241 (44.9)	7.48 (3.60–54.40)	
≥ 70	110 (20.5)	9.70 (4.25–55.70)	
Centre			< 0.001
Pavia, Italy	248 (46.1)	10.82 (3.70–55.70)	
Beaujon, France	250 (46.6)	7.60 (3.60–55.70)	
Berlin, Germany	30 (5.6)	7.12 (4.55–19.75)	
Frankfurt, Germany	9 (1.7)	10.71 (5.00–19.85)	
Cause of liver disease*			< 0.001
HCV	372 (69.3)	9.96 (3.60–54.40)	
HBV	109 (20.3)	6.37 (4.15–18.70)	
ALD	3 (0.6)	8.57 (5.75–11.75)	
NAFLD	18 (3.4)	11.13 4.35–55.70)	
Other	12 (2.2)	9.53 (5.00–19.75)	
No known cause	31 (5.7)		
Mean BMI ± SD (range) kg/m^2	24.6 ± 4.1 (17.1–47.3)		
BMI Category			0.109
< 18.5	17 (3.2)	9.40 (3.85–21)	
18.5–24.9	295 (54.9)	8.82 (3.6–54.40)	
25–29.9	163 (30.4)	8.98 (4.15–43.55)	
≥ 30	52 (9.7)	11.04 (3.6–55/70)	
Missing data	10 (1.9)		
Mean skin-liver distance ± SD (range) cm	1.7 ± 0.94 (0.1–5.9)		

Associated liver stiffness measurement values are provided. LSMs are the mean of the median values with range

*The total exceeds 100% because some patients had more than one identified cause

Statistical analysis

For fibrosis estimation, patients were classified using the following set of TE thresholds: ≤ 6.9 kPa, 7.0–9.4 kPa, 9.5–11.9 kPa and ≥ 12.0 kPa, derived from the meta-analysis by Tsochatzis et al [14] for the assessment of liver fibrosis as per the METAVIR system. A second set of LSM thresholds was also applied to facilitate a clinically useful approach to interpreting and utilising LSMs, as recommended for TE by the Baveno VI conference [4]. Extensive statistical analysis protocol is provided as Supplemental Material. A *p* value of 0.05 or less was considered to be significant except in cases of multiple comparisons when the suitable Bonferroni correction was applied. All analyses were performed using the Statistical Package for the Social Sciences (SPSS) software (version 23.0, SPSS Inc.). Graphs were created with Prism (v.7.0, GraphPad).

Results

Patients

During the study period, a total of 640 patients were included (363 male [56.7%] and 277 female [43.3%] patients with a mean age of 55.9 ± 14.5 (range 19–88) years old). Among them, 593 (92.7%) and 572 (89.4%) patients

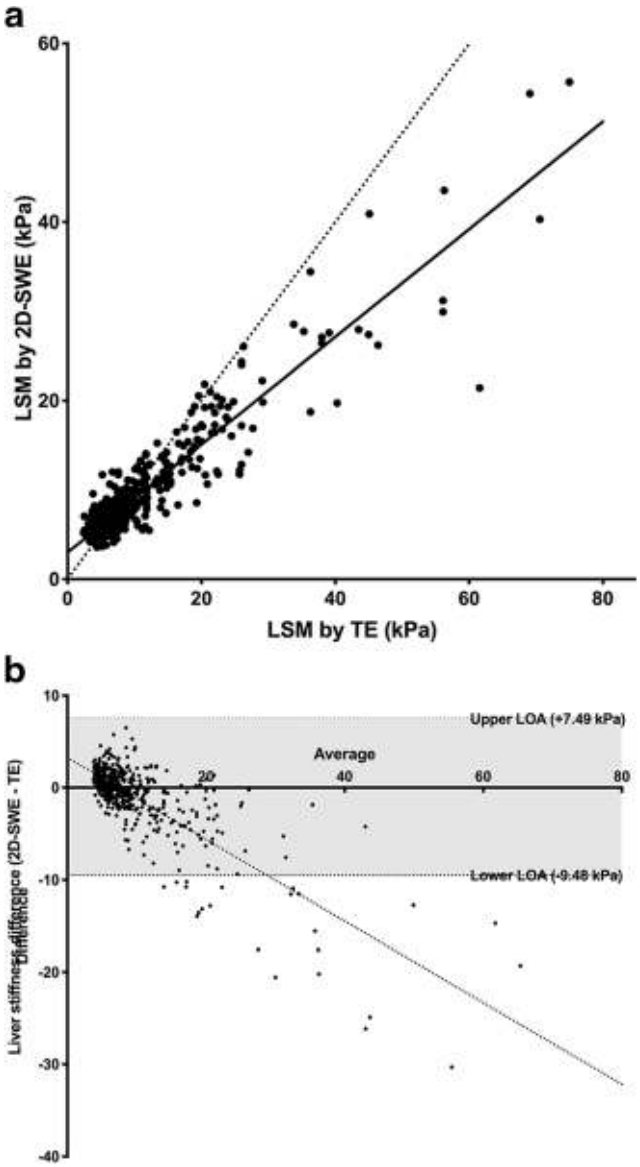

Fig. 2 **a** Plot of liver stiffness measurement (LSM) with transient elastography (TE) and 2D-shear wave elastography (SWE). Both were found to be strongly and significantly correlated (Pearson $r = 0.932$, $p < 0.001$, black line). The dashed line represents the perfect correlation. **b** Bland-Altman plot for the comparison of LSM measured by TE and 2D-SWE. In the entire cohort, the systematic bias was found to be − 0.99 kPa, with limits of agreements (LOAs) from − 9.48 to + 7.49 kPa. The graph shows the proportionate error towards higher LSM values

had reliable LSMs by TE and by 2D-SWE, respectively ($p = 0.05$). Overall, only 537 patients (83.9%) had reliable LSMs with both techniques and thus constituted the per-protocol study population. In this cohort, there were 310 (57.7%) male and 227 (42.3%) female patients, with a mean age of 55.3 ± 14.8 (range 19–88) years old. The main aetiologies of liver disease were chronic HCV and HBV infections with 372 (69.3%) and 109 (20.3%) patients, respectively. Table 1 provides the complete patient characteristics of the study population.

Direct comparison of LSMs using TE and 2D-SWE

The mean values of individual median LSMs using TE and 2D-SWE were 10.1 ± 9.4 (range 2.4–75) kPa and 9.1 ± 6.1 (range 3.6–55.7) kPa, respectively, and were significantly and strongly correlated (Pearson $r = 0.932$, $p < 0.001$) (Fig. 2a). The intraclass correlation coefficient was 0.850 (0.825–0.872).

A paired *t* test comparison showed that in the entire study population, 2D-SWE provided significantly lower LSM values (mean bias of 0.99 ± 4.33 kPa [95% limits of agreement (LOAs) − 9.48 to + 7.49], $p < 0.001$). A more detailed paired analysis showed that 2D-SWE provided lower LSM value compared with TE in patients with TE LSMs < 12 kPa (mean bias + 0.34 ± 1.60 kPa, $p < 0.001$) and subsequently provided higher LSM value compared with TE in patients with TE LSMs ≥ 12 kPa (mean bias − 6.33 ± 6.9 kPa, $p < 0.001$). This proportional error towards higher LSM values was also confirmed by the Bland-Altman plot (Fig. 2b).

Factor influencing 2D-SWE LSMs

LSMs by 2D-SWE was not significantly different according to gender ($p = 0.010$). It was weakly correlated with age (Pearson $r = 0.29$, $p < 0.001$) and with body mass index (Pearson $r = 0.12$, $p = 0.008$), but not with the skin-capsule distance ($p = 0.953$). 2D-LSM values were not significantly different between BMI categories but were shown to increase with age categories (Table 1).

In 485 patients with a BMI < 30 kg/m^2, the Bland-Altman plot showed a bias of 0.83 (95% LOAs − 7.22 to + 8.89). However, in 52 patients with a BMI ≥ 30 kg/m^2, the Bland-Altman plot showed a bias of 2.5 (95% LOAs − 8.94 to + 13.4).

2D-SWE LSMs according to the TE-derived categories

Table 2 and Fig. 3 detail the comparison of LSMs by TE and 2D-SWE using both sets of TE LSM thresholds (i.e. as per Tsochatzis et al [14] and the Baveno VI conference [4]).

Utilising the first set of TE LSM–derived categories by Tsochatzis et al [14], the median 2D-SWE LSMs significantly increased from the first to the last category (ANOVA: $p < 0.001$, all post hoc comparisons ≤ 0.001 except 7.0–9.4 kPa vs. 9.5–11.9 kPa with $p = 0.060$; Fig. 3). AUROCs of LSMs by 2D-SWE ranged from 0.935 ± 0.010 (95% CI 0.910–0.954) to 0.973 ± 0.009 (95% CI 0.955–0.985) (Fig. 4). The Obuchowski measure was 0.925 ± 0.103.

The optimal cut-off values of LSM by 2D-SWE with sensitivities and specificities are provided in Table 3. Applying these sets of thresholds resulted in correct classification of 390/537 (73%) patients (Supplemental Fig. 2).

When the second set of TE thresholds from the Baveno VI conference [4] was applied to the 2D-SWE LSM

Table 2 Comparison of liver stiffness measurement (LSM) by 2D-shear wave elastography (2D-SWE) per transient elastography (TE) categories

LSM by TE (kPa)	No. of patients	LSM by 2D-SWE (kPa)		ANOVA *p* values			
		Median (range)	Overall				
As per Tsochatzis et al [14]				vs. 7.0–9.4	vs. 9.5–11.9	vs. ≥ 12.0	
≤ 6.9	274	6.0 (3.6–12.0)	< 0.001	0.001	< 0.001	< 0.001	
7.0–9.4	93	7.9 (4.55–12.0)			0.060	< 0.001	
9.5–11.9	62	9.6 (5.55–14.1)				< 0.001	
≥ 12.0	108	17.7 (5.5–55.7)					
As per Baveno VI conference [4]				vs. 5.1–10.0	vs. 10.1–15.0	vs. 15.1–20.0	vs. ≥ 20.1
≤ 5.0	130	5.5 (3.6–9.6)	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
5.1–10.0	251	7.1 (3.7–12.0)			< 0.001	< 0.001	< 0.001
10.1–15.0	73	10.6 (5.5–12.0)				< 0.001	< 0.001
15.1–20.0	28	14.3 (8.3–20.6)					< 0.001
≥ 20.1	52	22.9 (10.7–55.7)					

ANOVA was performed to compare LSM by 2D-SWE according to the LSM as per TE-derived categories

dataset, a total of 130 (24.2%) median 2D-SWE LSMs significantly increased from the first to the last category (ANOVA: $p < 0.001$, all post hoc comparisons < 0.001; Fig. 3). AUROCs of 2D-SWE LSMs ranged from 0.885 ± 0.015 (95% CI 0.855–0.911) to 0.983 ± 0.005 (95% CI 0.969–0.992) (Fig. 4). The optimal cut-off values of 2D-SWE LSMs with sensitivities and specificities are provided in Table 3. Applying the second set of thresholds resulted in correctly classifying 360/537 (67%) patients (Supplemental Fig. 2).

Influence of the number of 2D-SWE LSM acquisitions

Table 4 details the influence of the number of acquisitions on 2D-SWE LSMs. Using the set of TE thresholds by Tsochatzis et al [14], the Obuchowski measures of 2D-SWE LSMs were consistent (0.92 ± 0.11) from three to nine acquisitions. The Bland-Altman comparisons between LSMs using TE and 2D-SWE showed systemic biases ranging from 0.86 to 0.98, with limits of agreements ranging from − 6.93 to − 7.77 (lower) and from 8.64 and 9.60 (upper). The ICC ranged from 0.843 (0.816–0.866) to 0.867 (0.844–0.887). Overall, there was no significant difference using three measures/acquisitions or more.

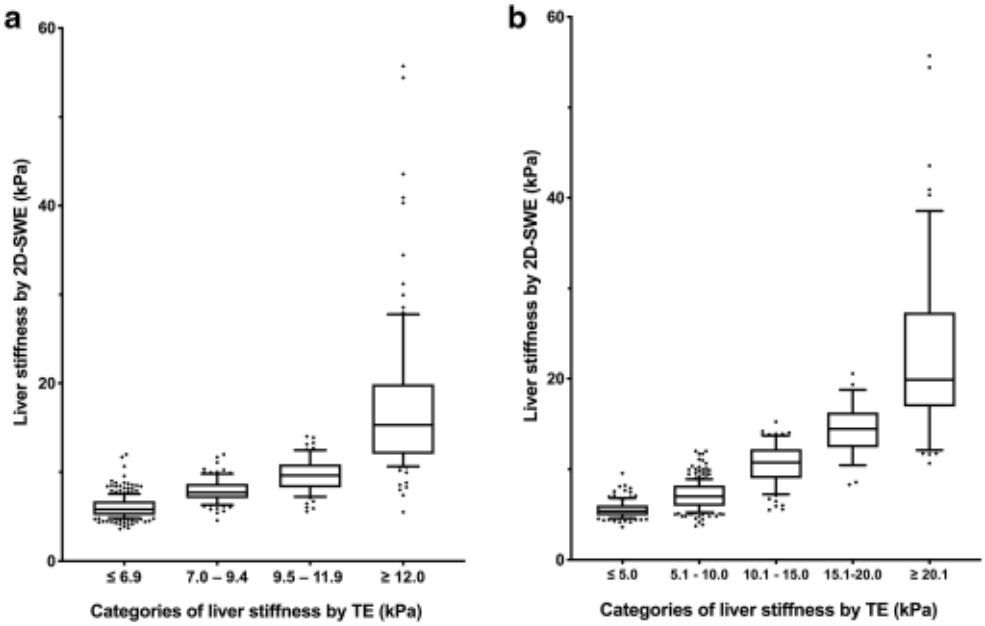

Fig. 3 Box plot representing liver stiffness measurement (LSM) by 2D-shear wave elastography (SWE) according to different LSMs by transient elastography (TE)-derived categories. **a** As per Tsochatzis et al [14]. **b** As per Baveno VI conference [4]. LSM by 2D-SWE progressively and significantly increased from the lowest to the highest TE categories ($p < 0.001$), allowing differentiation between categories. Boxes represent interquartile range, and the bar is the median. Whiskers are 10–90th values, and dots are outliers

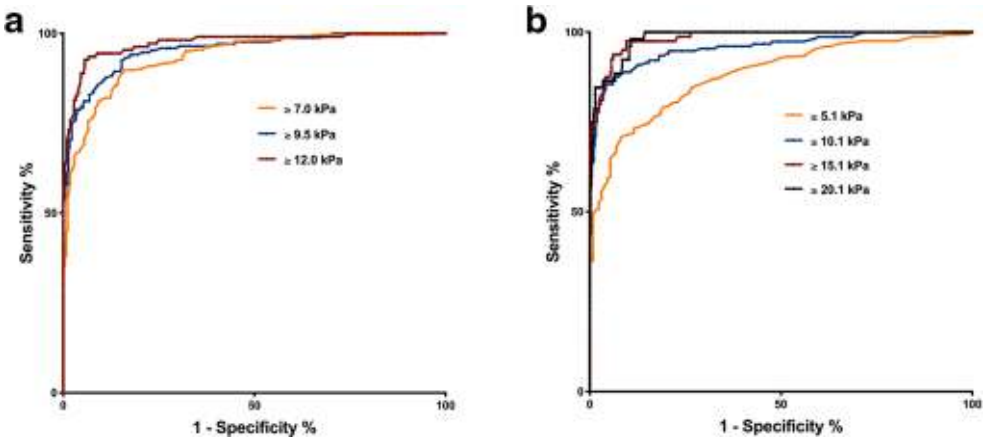

Fig. 4 **a**, **b** Receiver operating characteristic (ROC) curve of liver stiffness by 2D-SWE for the classification of patient according to transient elastography (TE)-derived categories

Discussion

The current study aimed to compare the value of liver stiffness measured by Canon 2D-SWE to that obtained by TE, considered as the reference standard. The rationale for it was that TE has been extensively validated in the non-invasive assessment of chronic liver disease, to the point that it has been included in various diagnostic and management algorithms, thus forming part of the routine clinical practice. This study showed that LSMs using 2D-SWE and TE are strongly correlated. 2D-SWE was shown to underestimate LSMs, especially towards the higher LSM values. Nevertheless, more than ¾ of the patients were correctly classified by 2D-SWE when using LSM TE–derived categories as reference. Finally, at least three individual 2D-SWE acquisitions and measurements were found to be sufficient for a reliable median 2D-SWE LSM.

Overall, LSMs provided by both techniques were strongly correlated. This has been reported by previous studies using 2D-SWE developed by other manufacturers [11, 16–21]. Interestingly, while other researchers reported the influence of gender on LSM values [22], this was not the case in our study. This may be explained by the strict exclusion of unreliable data in our study. Of note, LSM with 2D-SWE values increased with age categories, likely due to unbalanced disease severity. Noticeably, while BMI had no influence of LSM values per se, previous data by Matos et al [21] suggests that a high BMI was associated with more discordance between techniques. In our study, the pair-wise comparison showed that 2D-SWE tended to give lower LSM values compared with TE particularly with the higher stages of fibrosis. One might hypothesise that this could be explained by the SWE system used where previous studies using other elastography systems have not reported such proportionate errors [15–19] although some investigators have [11, 21, 23], especially when using this particular system [16]. Many studies have also previously reported the poor interchangeability between systems when assessing liver stiffness [5, 6, 24, 25]. This suggests that patients should be explored and followed up with the same system to avoid misclassification. This is of notable significance in patients

Table 3 Diagnostic performance of 2D-shear wave elastography (SWE) for liver stiffness measurement (LSM) using transient elastography (TE) categories as the reference

LSM by TE (kPa)	LSM by 2D-SWE optimal cut-off value (kPa)	AUROC (95% CI) of LSM by 2D-SWE	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	*p* value
As per Tsochatzis et al [14]					
< 7.0 vs. ≥ 7.0	≥ 7.40	0.935 (0.915 to 0.955)	89.7 (85.4–93.1)	83.9 (79.0–88.1)	< 0.001
< 9.5 vs. ≥ 9.5	≥ 7.95	0.954 (0.933 to 0.970)	92.4 (87.3–95.9)	84.7 (80.6–88.3)	< 0.001
< 12 vs. ≥ 12.0	≥ 10.50	0.973 (0.955 to 0.985)	92.6 (85.9–96.7)	94.4 (91.8–96.4)	< 0.001
As per Baveno VI conference [4]					
< 5.0 vs. ≥ 5.0	≥ 6.85	0.885 (0.855 to 0.911)	71.3 (66.6–75.7)	91.5 (85.4–95.7)	< 0.001
< 10.0 vs. ≥ 10.0	≥ 9.55	0.957 (0.936 to 0.973)	85.6 (79.0–90.8)	95.8 (93.3–97.6)	< 0.001
< 15.0 vs. ≥ 15.0	≥ 10.60	0.983 (0.968 to 0.992)	97.5 (91.3–99.7)	90.3 (87.2–92.9)	< 0.001
< 20.0 vs. ≥ 20.0	≥ 11.50	0.983 (0.969 to 0.992)	98.1 (89.7–100.0)	89.2 (86.1–91.8)	< 0.001

AUROC area under the receiving operating characteristic curve, *CI* confidence interval

Table 4 Influence of the number of individual liver stiffness measurement (LSM) by 2D-SWE on diagnostic performance (assessed by the Obuchowski measure) and on variability (intraclass correlation coefficient (ICC) and Bland-Altman plot), taking transient elastography as reference

N measures	Obuchowski	ICC	Bias	Lower LOA	Upper LOA
3	0.918 ± 0.110	0.843 (0.816–0.866)	0.91	− 7.77	9.60
4	0.920 ± 0.110	0.867 (0.844–0.887)	0.86	− 6.93	8.64
5	0.920 ± 0.113	0.856 (0.831–0.877)	0.93	− 7.41	9.29
6	0.924 ± 0.112	0.856 (0.831–0.877)	0.95	− 7.39	9.28
7	0.923 ± 0.112	0.859 (0.835–0.880)	0.97	− 7.29	9.22
8	0.925 ± 0.112	0.852 (0.826–0.874)	0.97	− 7.42	9.36
9	0.925 ± 0.112	0.848 (0.822–0.871)	0.98	− 7.49	9.45

Patients were categorised by liver stiffness per transient elastography using the following cut-off values: ≤ 6.9 kPa, 7.0–9.4 kPa, 9.5–11.9 kPa and ≥ 12.0 kPa. The Obuchowski measure was calculated following the fibrosis distribution by Payan et al [15] (F0–F1, 45%; F2, 28%; F3, 14%; F4, 13% [16]) applied to the thresholds above

LOA limit of agreement

with high LSM values by TE (> 12 kPa). However, the discrepant values, especially at the higher end, may have limited clinical consequences in terms of diagnosis, since these patients would still fall within the same LSM category. The consequences may be more significant when LSMs are considered for prognostic purposes. Patient categorisation is indeed a more clinically relevant approach than raw LSM values, since it reflects the way patients are managed in routine clinical practice.

As recommended by the Society of Radiologists in Ultrasound and the European Association for the Study of the Liver, patients should be stratified in groups of increasing risk of liver fibrosis or portal hypertension [26, 27]. In this perspective, LSM values below the lowest threshold are associated with a high probability of a normal liver or minimal fibrosis, and LSM values greater than the highest cut-off value would indicate a high probability of significant fibrosis, cirrhosis and clinically significant portal hypertension [4, 26, 27].

We therefore utilised two comprehensive sets of LSM-TE thresholds to categorise patients and showed that 2D-SWE correctly classified the majority of patients as per the TE categories using these LSM thresholds. As a consequence of the tendency of 2D-SWE to provide higher LSM values compared with TE in patients with high LSM values (i.e. > 12 kPa), low thresholds were close to that of TE while higher ones were significantly lower, as previously proposed by other teams with ARFI-based techniques [11, 21]. Of note, miscategorisations were only limited to one adjacent category and occurred mostly for the intermediate LSM values. Most patients with either very low or very high LSM values were correctly classified. It underlines the importance of adopting a set of LSM thresholds adapted to each particular system when performing LSM measurements.

The TE manufacturer has issued recommendations and guidelines on how to obtain a reliable LSM, and as these criteria have not varied over time, this could be considered as one of the strengths of this technique. Similar sets of recommendations, however, are not yet available for 2D-SWE technologies [27]. Previous studies have advocated the implementation of quality criteria, namely the number of reliable measurements [28, 29], the acquisition depth [28], the heterogeneity of measurements [30] or the variation coefficient of LSMs [28, 31, 32]. We used a two-step approach: first, only elastograms with > 50% colour filling considered stable over time were considered. Then, we applied the same quality criteria as TE to ensure comparability. We showed that TE had a borderline higher reliability rate than 2D-SWE (p = 0.05). The main reason for 2D-SWE unreliability was that elastograms were discarded because of a combination of uncontrolled breathing and absence of sufficient elastogram filling, preventing the operator from obtaining 10 individual acquisitions and measurements.

The 2D-SWE system used in our study provides a propagation map together with the SWE acquisition. The aim is to guide the user to optimal ROI placement for liver stiffness measurement. This sort of quality indicator is becoming more of a feature with advancing 2D-SWE technologies, but whether this influences the accuracy and interobserver variability is still yet to be proven. Previous studies, however, have shown that there is a learning curve with these techniques and that expert users demonstrate higher reproducibility, whatever the elastography technique [23, 25, 33].

Finally, we showed that a minimum of three individual measurements/acquisitions are sufficient to compute a reliable median LSM when using 2D-SWE. This has been a subject of continuous debate, where authors have reported widely varying numbers, ranging from 3 up to 12 measurements [11, 17, 19–22, 28, 31]. In most of these studies, the number of measurements performed was purely arbitrary and few studies have specifically focused on this issue with either 2D-SWE [17, 34] or point SWE [35–37]. All have concluded that less than 10 measurements are sufficient, with most favouring at least five. We have used an extensive statistical analysis, considering both accuracy (AUROC and Obuchowski) and variability (Bland-Altman), which makes our conclusions more solid. Moreover, this threshold has already been suggested by previous report [38]. Importantly, however, the number

of measurements is likely to have less influence on the diagnostic accuracy than data reliability [33, 35, 39, 40].

This current study has limitations. First and foremost, histology was not mandatory for patient inclusion. This was intentional and justified by three main reasons: first, most patients are now routinely diagnosed and followed up by means of non-invasive tests, especially TE which has become the standard of care in most hepatology units, as recommended by hepatology guidelines [27]. As such, it is becoming increasingly more difficult to justify the use of liver biopsy as a gold standard when researching new non-invasive liver fibrosis technologies, owing to potential complications. Second, and as a consequence, including only patients with available liver biopsy would have skewed the population towards atypical cases, rarer diseases or patients included in trials. Therefore, extrapolation of results would have been questionable. Third, we wanted to adopt a 'real life' design to understand the extent which TE and 2D-SWE could be interchanged. Another limitation is the unbalanced inclusion from the four centres, with two being responsible for the majority of the data. Since all centres are located in Europe, and given the similar epidemiology of chronic liver disease in the different countries, we believe this had limited influence of the results. Third, the current results were derived from the SW module of one manufacturer only. It is also well known that LSM values generated by differing ultrasound systems are not interchangeable [16, 24], unlike MR elastography techniques for which the frequency of the generator is now standardised. This current study was undertaken as the clinical validation of a more recently introduced 2D-SWE technique; thus, we did not perform comparisons with other commercially available 2D-SWE systems. Finally, the proposed 2D-SWE LSM thresholds were not validated in another cohort of patients.

In conclusion, LSMs by 2D-SWE and by TE are strongly correlated. 2D-SWE underestimates stiffness especially towards high LSM values, but 2D-SWE can correctly classify more than ¾ of the patients. 2D-SWE may be used in TE-derived algorithms for the non-invasive assessment of chronic liver disease, provided LSM thresholds are adjusted.

Funding This study has received funding by Canon Medical Systems, Japan.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Maxime Ronot.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

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

Ethical approval Institutional review board approval was obtained.

Methodology
- prospective
- diagnostic or prognostic study
- multi-centre study

References

1. Mokdad AA, Lopez AD, Shahraz S et al (2014) Liver cirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. BMC Med 12:145
2. Gower E, Estes C, Blach, Razavi-Shearer K, Razavi H (2014) Global epidemiology and genotype distribution of the hepatitis C virus infection. J Hepatol 61:S45–S57
3. Pimpin L, Cortez-Pinto H, Negro F et al (2018) Burden of liver disease in Europe: epidemiology and analysis of risk factors to identify prevention policies. J Hepatol 69:718–735
4. de Franchis R, Baveno VI Faculty (2015) Expanding consensus in portal hypertension: report of the Baveno VI Consensus Workshop: stratifying risk and individualizing care for portal hypertension. J Hepatol 63:743–752
5. Dietrich CF, Bamber J, Berzigotti A et al (2017) EFSUMB guidelines and recommendations on the clinical use of liver ultrasound elastography, update 2017 (long version). Ultraschall Med 38:e48
6. Ferraioli G, Wong VW, Castera L et al (2018) Liver ultrasound elastography: an update to the World Federation for Ultrasound in Medicine and Biology guidelines and recommendations. Ultrasound Med Biol 44:2419–2440
7. Friedrich-Rust M, Poynard T, Castera L (2016) Critical comparison of elastography methods to assess chronic liver disease. Nat Rev Gastroenterol Hepatol 13:402–411
8. Augustin S, Millan L, Gonzalez A et al (2014) Detection of early portal hypertension with routine data and liver stiffness in patients with asymptomatic liver disease: a prospective study. J Hepatol 60: 561–569
9. Berzigotti A (2017) Non-invasive evaluation of portal hypertension using ultrasound elastography. J Hepatol 67:399–411
10. Herrmann E, de Ledinghen V, Cassinotto C et al (2018) Assessment of biopsy-proven liver fibrosis by two-dimensional shear wave elastography: an individual patient data-based meta-analysis. Hepatology 67:260–272
11. Furlan A, Tublin ME, Yu L, Chopra KB, Lippello A, Behari J (2020) Comparison of 2D shear wave elastography, transient elastography, and MR elastography for the diagnosis of fibrosis in patients with nonalcoholic fatty liver disease. AJR Am J Roentgenol 214:W20–W26
12. Udompap P, Sukonrut K, Suvannarerg V, Pongpaibul A, Charatcharoenwitthaya P (2019) Prospective comparison of transient elastography, point shear wave elastography, APRI and FIB-4 for staging liver fibrosis in chronic viral hepatitis. J Viral Hepat. https://doi.org/10.1111/jvh.13246
13. Lefebvre T, Wartelle-Bladou C, Wong P et al (2019) Prospective comparison of transient, point shear wave, and magnetic resonance elastography for staging liver fibrosis. Eur Radiol 29:6477–6488
14. Tsochatzis EA, Gurusamy KS, Ntaoula S, Cholongitas E, Davidson BR, Burroughs AK (2011) Elastography for the diagnosis of severity of fibrosis in chronic liver disease: a meta-analysis of diagnostic accuracy. J Hepatol 54:650–659

15. Payan C, Roudot-Thoraval F, Marcellin P, et al. (2005) Changing of hepatitis C virus genotype patterns in France at the beginning of the third millenium: The GEMHEP GenoCII Study. J Viral Hepat 12:405–413
16. Iijima H, Tada T, Kumada T et al (2019) Comparison of liver stiffness assessment by transient elastography and shear wave elastography using six ultrasound devices. Hepatol Res 49:676–686
17. Paul SB, Das P, Mahanta M et al (2017) Assessment of liver fibrosis in chronic hepatitis: comparison of shear wave elastography and transient elastography. Abdom Radiol (NY) 42:2864–2873
18. Gerber L, Kasper D, Fitting D et al (2015) Assessment of liver fibrosis with 2-D shear wave elastography in comparison to transient elastography and acoustic radiation force impulse imaging in patients with chronic liver disease. Ultrasound Med Biol 41:2350–2359
19. Leung VY, Shen J, Wong VW et al (2013) Quantitative elastography of liver fibrosis and spleen stiffness in chronic hepatitis B carriers: comparison of shear-wave elastography and transient elastography with liver biopsy correlation. Radiology 269:910–918
20. Zeng J, Zheng J, Huang Z et al (2017) Comparison of 2-D shear wave elastography and transient elastography for assessing liver fibrosis in chronic hepatitis B. Ultrasound Med Biol 43:1563–1570
21. Matos J, Paparo F, Bacigalupo L et al (2019) Noninvasive liver fibrosis assessment in chronic viral hepatitis C: agreement among 1D transient elastography, 2D shear wave elastography, and magnetic resonance elastography. Abdom Radiol (NY) 44:4011–4021
22. Mulabecirovic A, Mjelle AB, Gilja OH, Vesterhus M, Havre RF (2018) Liver elasticity in healthy individuals by two novel shear-wave elastography systems-comparison by age, gender, BMI and number of measurements. PLoS One 13:e0203486
23. Ferraioli G, Tinelli C, Zicchetti M et al (2012) Reproducibility of real-time shear wave elastography in the evaluation of liver elasticity. Eur J Radiol 81:3102–3106
24. Piscaglia F, Salvatore V, Mulazzani L et al (2017) Differences in liver stiffness values obtained with new ultrasound elastography machines and FibroScan: a comparative study. Dig Liver Dis 49:802–808
25. Ferraioli G, De Silvestri A, Lissandrin R et al (2019) Evaluation of inter-system variability in liver stiffness measurements. Ultraschall Med 40:64–75
26. Barr RG, Ferraioli G, Palmeri ML et al (2015) Elastography assessment of liver fibrosis: Society of Radiologists in Ultrasound Consensus Conference statement. Radiology 276:845–861
27. European Association for Study of Liver; Asociacion Latinoamericana para el Estudio del Higado (2015) EASL-ALEH clinical practice guidelines: non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol 63:237–264
28. Procopet B, Berzigotti A, Abraldes JG et al (2015) Real-time shear-wave elastography: applicability, reliability and accuracy for clinically significant portal hypertension. J Hepatol 62:1068–1075
29. Sporea I, Gradinaru-Tascau O, Bota S et al (2013) 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 15:268–272
30. Thiele M, Madsen BS, Procopet B et al (2017) Reliability criteria for liver stiffness measurements with real-time 2D shear wave elastography in different clinical scenarios of chronic liver disease. Ultraschall Med 38:648–654
31. Elkrief L, Rautou PE, Ronot M et al (2015) Prospective comparison of spleen and liver stiffness by using shear-wave and transient elastography for detection of portal hypertension in cirrhosis. Radiology 275:589–598
32. Jansen C, Bogs C, Verlinden W et al (2017) Shear-wave elastography of the liver and spleen identifies clinically significant portal hypertension: a prospective multicentre study. Liver Int 37:396–405
33. Fang C, Konstantatou E, Romanos O, Yusuf GT, Quinlan DJ, Sidhu PS (2017) Reproducibility of 2-dimensional shear wave elastography assessment of the liver: a direct comparison with point shear wave elastography in healthy volunteers. J Ultrasound Med 36:1563–1569
34. Shin HJ, Kim MJ, Kim HY et al (2016) Optimal acquisition number for hepatic shear wave velocity measurements in children. PLoS One 11:e0168758
35. Roccarina D, Iogna Prat L, Buzzetti E et al (2019) Establishing reliability criteria for liver ElastPQ shear wave elastography (ElastPQ-SWE): comparison between 10, 5 and 3 measurements. Ultraschall Med. https://doi.org/10.1055/a-1010-6052
36. Durot I, Akhbardeh A, Rosenberg J et al (2018) Point shear wave elastography for grading liver fibrosis: can the number of measurements be reduced? Ultrasound Med Biol 44:2569–2577
37. Fang C, Jaffer OS, Yusuf GT et al (2018) Reducing the number of measurements in liver point shear-wave elastography: factors that influence the number and reliability of measurements in assessment of liver fibrosis in clinical practice. Radiology 287:844–852
38. Vilgrain VCC, Leung VY-F, Ferraioli G et al (2015) Optimal number of 2D shear wave elastography measurements for diagnosing liver fibrosis severity. ESGAR, Paris
39. Ferraioli G, De Silvestri A, Reiberger T et al (2018) Adherence to quality criteria improves concordance between transient elastography and ElastPQ for liver stiffness assessment-a multicenter retrospective study. Dig Liver Dis 50:1056–1061
40. Ferraioli G, Maiocchi L, Lissandrin R et al (2016) Accuracy of the ElastPQ technique for the assessment of liver fibrosis in patients with chronic hepatitis C: a "real life" single center study. J Gastrointestin Liver Dis 25:331–335