



HAL
open science

Reliability Criteria of Two-Dimensional Shear Wave Elastography: Analysis of 4277 Measurements in 788 Patients

Anita Paisant, Sylvain Lemoine, Christophe Cassinotto, Victor de Lédinghen, Maxime Ronot, Marie Irlès-Depé, Valérie Vilgrain, Brigitte Le Bail, Valérie Paradis, Clémence Canivet, et al.

► To cite this version:

Anita Paisant, Sylvain Lemoine, Christophe Cassinotto, Victor de Lédinghen, Maxime Ronot, et al.. Reliability Criteria of Two-Dimensional Shear Wave Elastography: Analysis of 4277 Measurements in 788 Patients. *Clinical Gastroenterology and Hepatology*, 2020, 20 (2), pp.400-408.e10. 10.1016/j.cgh.2020.12.013 . hal-04016375

HAL Id: hal-04016375

<https://univ-angers.hal.science/hal-04016375v1>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Reliability Criteria of Two-Dimensional Shear Wave Elastography: Analysis of 4277 Measurements in 788 Patients

Title page

Short Title: 2D-SWE reliability criteria

Anita Paisant ^{1,2}, Sylvain Lemoine ¹, Christophe Cassinotto ³, Victor de Ledinghen ⁴, Maxime Ronot ⁵, Marie Irles-Depe ⁴, Valerie Vilgrain ⁵, Brigitte Le Bail ⁶, Valérie Paradis ⁷, Clemence M Canivet ^{2,8}, Sophie Michalak ^{2,9}, Marie-Christine Rousselet ^{2,9}, Pierre-Emmanuel Rautou ¹⁰, Jérôme Lebigot ^{1,2}, Gilles Hunault ², Anne Crouan ¹, Christophe Aube ^{1,2}, Jerome Boursier ^{2,8}

- (1) Département de Radiologie, Centre Hospitalier Universitaire d'Angers, Angers, France
- (2) Laboratoire HIFIH, UPRES EA3859, SFR 4208, Université d'Angers, Angers, France
- (3) Département de Radiologie, Hôpital Saint-Eloi Hospital, Centre Hospitalier Universitaire de Montpellier, Montpellier, France
- (4) Service d'Hépatogastroentérologie, Hôpital Haut-Lévêque, Centre Hospitalier Universitaire de Bordeaux, Pessac, France
- (5) Service de Radiologie, HUPNSV, Hôpital Beaujon, Clichy France, INSERM UMR 1149, Université de Paris, Paris, France
- (6) Service d'Anatomopathologie, Hôpital Pellegrin, Centre Hospitalier Universitaire de Bordeaux, Pessac, France
- (7) Service d'Anatomopathologie, HUPNSV, Hôpital Beaujon, Clichy France,
- (8) Service d'Hépatogastroentérologie, Centre Hospitalier Universitaire d'Angers, Angers, France
- (9) Département de Pathologie Cellulaire et Tissulaire, Centre Hospitalier Universitaire d'Angers, Angers, France
- (10) Service de d'Hépatologie, HUPNSV, Hôpital Beaujon, Clichy France,

Grant support: Angers University Hospital

Abbreviations: 2D-SWE: two-dimensional shear wave elastography; AUROC: area under the receiver operating characteristics; CV: coefficient of variation; kPa: kilo Pascal; pSWE: point shear wave elastography; ROI: region of interest; SD: standard deviation of mean stiffness; VCTE: vibration controlled transient elastography

Correspondance: Anita Paisant, MD; Département de Radiologie, Centre Hospitalier Universitaire d'Angers, 4 rue Larrey, 49933 Angers cedex 09, France; Tel: (+33) 2 41 35 42 81; Fax: (+33) 2 41 35 49 38; Email: anita.paisant@chu-angers.fr

Disclosures: Pr Aubé (Guerbet: scientific collaboration; Siemens: scientific collaboration; Bayer: lecture), Pr Boursier (Consultant for Intercept, inventia; Board for BMS, Gilead, Intercept, Pfizer, MSD; Speaker for Gilead, Intercept, Lilly; supporting research for Echosens, Intercept, Siemens), Pr Vilgrain (Speaker fees with Supersonic Imagine).

Word count: 3148

ABSTRACT

Background and aim: Two-dimensional shear wave elastography (2D-SWE) is an accurate method for the non-invasive evaluation of liver fibrosis. We aimed to determine the reliability criteria and the number of necessary reliable measurements for 2D-SWE.

Method: 788 patients with chronic liver disease underwent liver biopsy and 2D-SWE examination in three centers. The 4277 2D-SWE measurements performed were 2:1 randomly divided into derivation (n=2851) and validation (n=1426) sets. Reliability criteria for a 2D-SWE measurement were defined in the derivation set from the intrinsic characteristics given by the device (mean liver stiffness, standard deviation, diameter of the region of interest), with further evaluation in the validation set.

Results: In the whole population of 4277 measurements, AUROC for bridging fibrosis was 0.825 ± 0.006 and AUROC for cirrhosis was 0.880 ± 0.006 . Mean stiffness and coefficient of variation (CV) were independent predictors of bridging fibrosis or cirrhosis. From these two parameters, new criteria were derived to define a reliable 2D-SWE measurement: stiffness < 8.8 kPa, or stiffness between 8.8-11.9 kPa with CV < 0.25 , or stiffness ≥ 12.0 kPa with CV < 0.10 . In the validation set, AUROC for bridging fibrosis was 0.830 ± 0.013 in reliable measurements vs 0.667 ± 0.031 in unreliable measurements ($p < 0.001$). AUROC for cirrhosis was 0.918 ± 0.014 vs 0.714 ± 0.027 , respectively ($p < 0.001$). The best diagnostic accuracy for a 2D-SWE examination was achieved from three reliable measurements.

Conclusion: Reliability of a 2D-SWE measurement relies on the coefficient of variation and the liver stiffness level. A 2D-SWE examination should include three reliable measurements according to our new criteria.

Keywords: Liver-stiffness; 2D-SWE; Liver Fibrosis; Cirrhosis; Reliability

INTRODUCTION

The non-invasive measurement of liver stiffness by ultrasound elastography has an important place in the evaluation and the management of patients with chronic liver disease.¹ Among the different technologies available, two-dimensional shear wave elastography (2D-SWE) displays a real-time stiffness map and allows the measurement of liver stiffness in a region of interest (ROI) placed on the frozen map.² This technique has been reported as an accurate method for the evaluation of liver fibrosis in chronic liver diseases and is now widely used.³

In addition to accuracy, reliability must be also evaluated for a diagnostic tool as it represents a quality criterion for the final diagnosis deduced from the test result. Reliability criteria are now well defined for liver stiffness measurement using vibration controlled transient elastography (VCTE) or point shear wave elastography (pSWE).^{4,5} With these technologies, the result corresponds to the median value of ten measurements, and reliability relies on the interquartile range to median ratio reflecting measurements dispersion. Contrary to VCTE and pSWE, 2D-SWE technology provides the mean and the dispersion of liver stiffness for each single measurement performed in the stiffness map. These results are displayed on the screen of the device, which offers the possibility to control reliability after each single measurement and therefore to adapt the examination in “real-time”. Only one study has explored this approach, including 678 2D-SWE measurements from 142 patients.⁶ Results showed that 2D-SWE measurements with standard deviation of mean stiffness ≤ 1.75 kilo Pascal (kPa) and ROI ≥ 18 mm provided the best diagnostic accuracy for cirrhosis. Authors concluded that these criteria need further

validation, and that a larger study could help to refine them. When compared to other well evaluated elastography techniques (VCTE, pSWE), feasibility of 2D-SWE is considered to be lower because of its higher failure rate and it being more time-consuming⁷ which leads operators to make few measurements. However, the minimal number of measurements required for a 2D-SWE examination is also still under debate.⁸⁻¹²

Therefore, in the present work we aimed to first determine the reliability criteria for a single 2D-SWE measurement, and second to determine the minimal number of reliable measurements required for an accurate 2D-SWE examination. By relying on a large sample of patients and 2D-SWE-measurements, we were able to carry out derivation and validation studies.

MATERIALS AND METHODS

Patients

Patients with chronic liver disease who underwent a liver biopsy between November 2014 and January 2018 at Angers University Hospital, December 2014 and June 2016 at Beaujon University Hospital, and September 2011 and May 2015 at Bordeaux University Hospital were prospectively included. Exclusion criteria were decompensated cirrhosis (liver failure, encephalopathy, ascites, variceal bleeding, systemic infection) and hepatocellular carcinoma. The study protocol conformed to the ethical guidelines of the current Declaration of Helsinki and was approved by local Ethic Committees. All patients gave written informed consent before being included in the study. All authors had access to the study data and reviewed and approved the final manuscript.

Histological assessment

Liver biopsy was taken as the reference for the assessment of liver fibrosis. Pathological examinations of liver biopsies were performed in each center by a senior expert specialized in hepatology and blinded for patient data. Liver fibrosis was evaluated according to NASH-CRN staging in patients with non-alcoholic fatty liver disease,¹³ and METAVIR staging in patients with other causes of chronic liver disease.¹⁴ Although the two semi-quantitative scoring systems comprise stages from F0 to F4, they do not completely correspond (**Supplementary Table s1**). For the present study, we defined “no/mild fibrosis” as NASH-CRN F0-2 or METAVIR F0-1, “bridging fibrosis” as NASH-CRN F3 or Metavir F2-3, and “cirrhosis” as NASH-CRN F4 or METAVIR F4. Advanced fibrosis was defined as NASH-CRN F \geq 3 or Metavir F \geq 2 (**Table s1c**). We have already used such methodology in previous works.^{4,15}

2D-SWE examination

2D-SWE examination was performed using an Aixplorer device (Hologic/Supersonic Imagine Aix-en-Provence, France) the day of or no more than three months before or after liver biopsy. The examination was performed by an experienced radiologist blinded for patient data. Patients in fasting condition were lying in dorsal decubitus with the right arm in abduction. The site of measurement was chosen in the right lobe of the liver, in a zone free of large vascular or biliary structures and at least 15mm below the liver capsule. An apnea was mandatory. A homogeneous color mapping of the stiffness was obtained according to the manufacturer's recommendations, and liver stiffness was calculated in a ROI positioned in this color map. As recommended by the last EFSUMB guidelines,² the objective was to target a ROI diameter of at least 15mm. Each patient underwent 10 measurements in Angers center, and 5 measurements in Beaujon and Bordeaux centers. Measurements were judged as failed when no or little signal was obtained in the Aixplorer box. The following intrinsic characteristics were recorded for each measurement: mean stiffness expressed in kPa, standard deviation of mean stiffness (SD), and ROI diameter.

Statistical analysis

Continuous variables were expressed as median with first and third quartile and compared using the Mann-Whitney test. Categorical variables were expressed as percentages and compared using the Fisher's Exact Test. Correlation between continuous variables was evaluated using the Spearman correlation coefficient, and agreement using the intraclass correlation coefficient. Diagnostic accuracy was assessed using the area under the receiver operating characteristics (AUROC) and the rate of correctly classified patients. AUROCs were

compared according to DeLong *et al.*¹⁶ Diagnostic cut-offs were calculated according to the highest Youden index that optimizes sensitivity and specificity.

All measurements performed during the study were randomly divided 2:1 in derivation and validation sets. To identify the potential reliability criteria for a 2D-SWE measurement, we performed multivariate analyses in the derivation set using bridging fibrosis or cirrhosis as dependent variables, and the intrinsic characteristics of 2D-SWE measurements (mean stiffness, SD, and ROI diameter) as independent variables. Indeed, by definition, each variable selected by a multivariate analysis is an independent predictor of the diagnostic target studied. In other words, when selected with 2D-SWE mean stiffness, an independent predictor influences the outcome (diagnostic target) for each fixed level of liver stiffness. Consequently, the multivariate analysis allowed for the identification of the predictors influencing 2D-SWE accuracy, from which reliability criteria were derived. Results obtained were then confirmed in the validation set.

Finally, the minimal number of reliable 2D-SWE measurements for an accurate examination was calculated. For this analysis, we considered the reference for 2D-SWE examination as the mean of the 5 first measurements. We compared this reference to the first measurement, and to the mean of the two, three and four first measurements.

Statistical analyses were performed using SPSS version 18.0 software (IBM, Armonk, NY, USA). This study was reported in accordance with the recently published LiverFibroSTARD statements.¹⁷

RESULTS

Patients

788 patients were included in the study: 205 in the Angers center, 222 in Beaujon and 361 in Bordeaux. Their characteristics are outlined in **Table s2**. 60.6% were male, median age was 55.9 (45.1-63.6) years, and median body mass index was 27.3 (23.6-31.2) kg/m². Non-alcoholic fatty liver disease was the main cause of chronic liver disease (40.7%), 26.4% of patients had chronic viral hepatitis, and 9.9% had alcoholic liver disease. Median biopsy length was 27 (20-33) mm. The prevalence of advanced fibrosis and cirrhosis was 47.2% and 19.0%, respectively.

2D-SWE in the 788 included patients resulted in 4277 measurements. Most of the patients (n=777) had at least 3 measurements, and 478 had at least 5 measurements (**Table s3**). The patient characteristics were not significantly different between the derivation and the validation sets (**Table 1**).

Diagnostic accuracy of 2D-SWE measurements

In the set of 4277 2D-SWE measurements, AUROCs for advanced fibrosis and cirrhosis were 0.825 ± 0.006 and 0.880 ± 0.006 , respectively (**Table 2**). In the derivation set, the best diagnostic cut-off was calculated at 8.8 kPa for advanced fibrosis and 12.0 kPa for cirrhosis. Using these cut-offs in the validation set, the diagnostic accuracy was 74.0% for advanced fibrosis and 77.5% for cirrhosis. AUROCs and the different accuracy indexes were not significantly different between the derivation and the validation sets (**Table 2**).

Development of reliability criteria for a single 2D-SWE measurement (derivation set)

Three intrinsic characteristics were tested in the derivation set: mean stiffness, SD, and ROI diameter. Because SD was highly correlated with mean stiffness (Spearman correlation coefficient: 0.703, $p < 0.001$; **Table s4**), we decided to use the standard deviation to mean stiffness ratio. In statistics, the ratio standard deviation to mean corresponds to the coefficient of variation (CV). By multivariate analysis, mean stiffness and CV but not ROI diameter were independent predictors of advanced fibrosis or cirrhosis. These results suggested that, for a given liver stiffness, CV level influences the likelihood of having advanced fibrosis or cirrhosis. Therefore, we concentrate further analyses on these two parameters.

2D-SWE accuracy decreased as CV increased in the derivation set, and three CV subgroups were identified: <0.10 , $0.10-0.24$, >0.24 (**Table s5**). **Fig. 1** shows the probability of advanced fibrosis and cirrhosis as a function of increasing liver stiffness and these three CV subgroups. CV >0.24 was associated with a decreased probability of advanced fibrosis from stiffness value around the 8.8 kPa cut-off (**Fig. 1a**). CV ≥ 0.10 was associated with a decreased probability of cirrhosis from stiffness value around the 12.0 kPa cut-off (**Fig. 1b**). **Fig. s1** illustrates the impact of CV level for the diagnosis of advanced fibrosis and cirrhosis when using the 8.8 kPa and the 12.0 kPa cut-offs in clinical practice. There was no influence of CV on the negative predictive values for advanced fibrosis or cirrhosis. On the contrary, CV significantly impacted the positive predictive value, which significantly decreased by CV >0.24 for the diagnosis of advanced fibrosis (**Fig. s1a**) and by CV ≥ 0.10 for the diagnosis of cirrhosis (**Fig. s1b**). Finally, 2D-SWE accuracy depended on both CV and liver stiffness levels. Using the previous results, we derived new reliability criteria for a 2D-SWE measurement (**Fig. 2**).

Unreliable measurements had a poor diagnostic accuracy in the derivation set (**Table 3**). Compared to reliable measurements, AUROC for advanced fibrosis was 0.641 ± 0.024 versus 0.840 ± 0.009 ($p < 0.001$) and corresponding diagnostic accuracy using the 8.8 kPa cut-off was 71.6% vs 77.3% ($p = 0.002$). For cirrhosis, AUROC was 0.712 ± 0.019 versus 0.924 ± 0.009 ($p < 0.001$) and diagnostic accuracy using the 12.0 kPa cut-off was 41.4% vs 91.4% ($p < 0.001$).

Validation of 2D-SWE reliability criteria

Validation set

377 of the 1426 measurements (26.4%) were unreliable in the validation set. Results obtained in the validation set confirmed those of the derivation set, with AUROC and diagnostic accuracy for advanced fibrosis or cirrhosis being significantly lower in unreliable compared to reliable measurements (**Table 3**). A sensitivity analysis in the validation set confirmed that reliable measurements were significantly more accurate than unreliable measurements both in the subgroup of patients with steatotic liver diseases (NAFLD/alcohol) and in the subgroup of patients with other causes of chronic liver disease (**Table s6**). The better diagnostic accuracy for reliable measurements was also observed across the three investigating centers (**Table s7**).

Comparison to previously published reliability criteria

Thiele *et al* have recently published 2D-SWE criteria based on $SD \leq 1.75$ kPa and ROI diameter ≥ 18 mm.⁶ In this study, the diagnostic accuracy was best when these two conditions were met (referred to as Thiele criteria #1), and poorer when they were not met (Thiele #4), and it was intermediate when only one of the two conditions was achieved (Thiele #2 and #3). In our whole study population, measurements meeting Thiele #3 and #4 criteria, both

characterized by SD >1.75 kPa, had insufficient diagnostic accuracy and accounted for 33.0% of all measurements (**Table s8**).

Unreliable measurements according to our new definition accounted for 26.0% of all measurements ($p < 0.001$ vs Thiele #3-4), and 92.3% of them also met Thiele #3-4 criteria. Among the 2864 measurements meeting Thiele #1-2 criteria, our new reliability criteria reclassified as unreliable a small subset of 86 measurements with a very low diagnostic accuracy (**Table s9**). More interestingly, among the 1413 measurements meeting the Thiele #3-4 criteria, a substantial number of 387 measurements (27.4%) with very good accuracy were reclassified as reliable thanks to our new criteria.

Minimal number of reliable 2D-SWE measurements for an accurate examination

For this analysis, we excluded the 1112 measurements considered unreliable according to our new definition. The 3165 remaining reliable measurements came from 713 patients, of whom 297 with at least 5 measurements available were included in the analysis.

There was an excellent agreement with the reference (mean of first 5 measurements) since the first measurement with intraclass correlation coefficient >0.96 (**Table 4**, see scatter plots in **Fig. s2** and Bland Altman plots in **Fig. s3**). For the diagnosis of advanced fibrosis, the AUROC with one measurement (0.787 ± 0.027 , $p=0.005$) and two measurements (0.813 ± 0.026 , $p=0.070$) were lower compared to the reference (0.829 ± 0.025). Using the 8.8 kPa cut-off, the diagnostic accuracy for advanced fibrosis slightly increased from the first measurement to reach a plateau at the third measurement. For cirrhosis, the diagnostic accuracy was excellent from the first measurement. Taken together, all these results suggested that a 2D-SWE examination should include three reliable examinations.

DISCUSSION

We have previously shown that intrinsic characteristics of liver stiffness measurement with VCTE or pSWE should be carefully controlled to ensure an accurate non-invasive evaluation of liver fibrosis.^{4,5} In the present work, we have identified mean stiffness and CV as two key criteria to define the reliability of a 2D-SWE measurement. By comparison to the Thiele *et al* study,⁶ our reliability criteria were more discriminant, better separating reliable measurements from those with very poor accuracy that should not be used for the evaluation of liver fibrosis in clinical practice.

We have previously shown that the reliability of VCTE and pSWE results relies on the interquartile range to median ratio reflecting the dispersion of the ten measurements performed during the examination.^{4,5} Compared to VCTE and pSWE, 2D-SWE has the particularity to display a real-time stiffness map with several parameters measured in a selected ROI within this map. Our study demonstrates that reliability could be assessed at each single 2D-SWE measurement. Among the parameters measured by the device, we showed that the ratio SD to mean stiffness (corresponds to the coefficient of variation) is a key factor influencing the accuracy for the diagnosis of liver fibrosis from a 2D-SWE measurement. SD reflects the heterogeneity of the signal within the stiffness map making the CV for a single 2D-SWE measurement a kind of equivalent to the interquartile range to median ratio of the ten VCTE or pSWE measurements. So, the homogeneity within the examination, regardless of the technology used, should be very carefully considered to ensure a relevant result of liver stiffness measurement.

The advantage of this real-time stiffness map must nevertheless be balanced with the 26.4% rate of unreliable measurements (377/1426) in the validation set, which is relatively high compared to other elastography techniques. Indeed, the rate of poorly reliable examination with TE and pSWE is less than 10%.^{4,5,18} A study comparing the feasibility of TE and 2D-SWE has shown that the skin-liver capsule distance is of importance for 2D-SWE examination.¹⁹ In this work, the odds of a successful 2D-SWE exam decreased with higher skin-capsule distance (OR = 0.77, 95% CI 0.67-0.98), whereas TE can be accomplished in nearly all patients by use of the FibroScan XL probe and repeated examinations. From our clinical practice, acquisitions with 2D-SWE can be improved by taking the time to explain the neutral breathing position technique (not inspiration or expiration) to the patient, and then test several inter-costal spaces in B-mode to find the best view window.

ROI diameter was not selected as a potential reliability criterion by the multivariate analyses. Compared to the Thiele *et al* study, mean size of the ROI was smaller in our cohort (15 mm versus 19 mm) because we followed the EFSUMB guidelines that recommend a 15 mm diameter.² In a recent study, the repeatability of liver stiffness measurement did not differ between 2D-SWE measurements performed with a 10 mm ROI versus those with 20 mm ROI.²⁰ Moreover, Thiele *et al* reported a 0.6% increase in cirrhosis misclassification for every 1 mm decrease in ROI diameter, which finally shows only slight influence of the ROI diameter on the diagnostic accuracy. To note, the lowest ROI diameter in the Thiele study was 13 mm. Considering all these results and according to our experience, we suggest following the usual recommendation of a 15 mm ROI diameter in clinical practice.

Acquisition of the real-time stiffness with 2D-SWE has the advantage of the mapping but is also much more time-consuming than a single measurement with VCTE or pSWE. Acquisition durations of a single 2D-SWE measurement reported in previous works ranged from 22 to 42 seconds, with an increased duration among obese patients.^{21,22} Therefore, it does not seem convenient in clinical practice to perform a high number of acquisitions within a 2D-SWE examination. Previous studies have shown conflicting results concerning the ideal number of 2D-SWE measurements, with some finding only one measurement to be sufficient,¹² whereas others claimed three,^{8,10} five,¹¹ or even six.⁹ This heterogeneity in the results could be explained by the different designs and statistical approaches used across the studies. In our work, we first eliminated unreliable measurements according to our new criteria, and then used both agreement and diagnostic tests to find that a minimum of three reliable examinations is required for an accurate 2D-SWE examination. Our approach simplifies the use of 2D-SWE for physicians in clinical practice: rather than performing a high number of measurements and then evaluating their dispersion at the end to determine if the examination is reliable or not (as for VCTE or pSWE), they check reliability at each measurement and stop the examination as soon as they obtain three reliable 2D-SWE measurements.

Patient characteristics significantly differed between our three investigating centers. Beyond geographical location, this inter-center variability was probably the consequence of different processes for patient recruitment, awareness of liver diseases by non-specialists, as well as local guidelines and procedures in specialized centers. We decided to pool the data of our three populations to catch the different patient profiles in a more representative sample which would provide robust results exportable for clinical practice. The randomization in

derivation and validation sets allowed an internal cross-validation referred as a phase 2a study in the TRIPOD guidelines.²³ We now encourage further phase 4 studies to externally and independently validate our findings, not only with the Aixplorer but also with the other 2D-SWE devices. Indeed, all 2D-SWE technologies are based on identical physical principles between manufacturers. We can therefore suppose that our new reliability criteria developed with the Aixplorer can be extended to other 2D-SWE devices.

CONCLUSIONS

The reliability of 2D-SWE measurement relies on the coefficient of variation and the level of liver stiffness. A 2D-SWE examination should include three reliable measurements as defined by our new reliability criteria. Using these new rules will help physicians to obtain an accurate non-invasive evaluation of liver fibrosis with 2D-SWE in their patients with chronic liver diseases.

REFERENCES

1. 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;44:2419-2440.
2. Dietrich CF, Bamber J, Berzigotti A, et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Liver Ultrasound Elastography, Update 2017 (Long Version). *Ultraschall Med* 2017;38:e16-e47.
3. Herrmann E, de Lédinghen V, Cassinotto C, et al. Assessment of biopsy-proven liver fibrosis by two-dimensional shear wave elastography: An individual patient data-based meta-analysis. *Hepatology* 2018;67:260-272.
4. Boursier J, Cassinotto C, Hunault G, et al. Criteria to Determine Reliability of Noninvasive Assessment of Liver Fibrosis With Virtual Touch Quantification. *Clin Gastroenterol Hepatol* 2019;17:164-171.e5.
5. Boursier J, Zarski J-P, de Lédinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology* 2013;57:1182-1191.
6. Thiele M, Madsen BS, 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 Med* 2017;38:648-654.
7. 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;63:1817-1827.
8. Sporea I, Grădinaru-Tașcău 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;15:268-272.
9. Yoon JH, Lee JM, Han JK, et al. Shear wave elastography for liver stiffness measurement in clinical sonographic examinations: evaluation of intraobserver reproducibility, technical failure, and unreliable stiffness measurements. *J Ultrasound Med* 2014;33:437-447.
10. Shin HJ, Kim M-J, Kim HY, et al. Optimal Acquisition Number for Hepatic Shear Wave Velocity Measurements in Children. *PLoS ONE* 2016;11:e0168758.
11. Choi SH, Jeong WK, Kim Y, et al. How many times should we repeat measuring liver stiffness using shear wave elastography?: 5-repetition versus 10-repetition protocols. *Ultrasonics* 2016;72:158-164.
12. Huang Z-P, Zhang X-L, Zeng J, et al. Study of detection times for liver stiffness evaluation by shear wave elastography. *World J Gastroenterol* 2014;20:9578-9584.
13. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;41:1313-1321.
14. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. *Hepatology* 1994;20:15-20.

15. Boursier J, de Ledinghen V, Leroy V, et al. A stepwise algorithm using an at-a-glance first-line test for the non-invasive diagnosis of advanced liver fibrosis and cirrhosis. *J Hepatol* 2017;66:1158-1165.
16. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-845.
17. Boursier J, de Ledinghen V, Poynard T, et al. An extension of STARD statements for reporting diagnostic accuracy studies on liver fibrosis tests: the Liver-FibroSTARD standards. *J Hepatol* 2015;62:807-815.
18. Bota S, Sporea I, Sirli R, et al. Factors associated with the impossibility to obtain reliable liver stiffness measurements by means of Acoustic Radiation Force Impulse (ARFI) elastography--analysis of a cohort of 1,031 subjects. *Eur J Radiol* 2014;83:268-272.
19. Staugaard B, Christensen PB, Mössner B, et al. Feasibility of transient elastography versus real-time two-dimensional shear wave elastography in difficult-to-scan patients. *Scand J Gastroenterol* 2016;51:1354-1359.
20. Wang H, Zheng P, Sang L, et al. Does Operator Experience and the Q-Box Diameter Affect the Repeatability of Liver Stiffness Measurements Obtained by 2-Dimensional Shear Wave Elastography? *J Ultrasound Med* 2020;39:741-747.
21. Schellhaas B, Strobel D, Wildner D, et al. Two-dimensional shear-wave elastography: a new method comparable to acoustic radiation force impulse imaging? *Eur J Gastroenterol Hepatol* 2017;29:723-729.
22. Varbobitis IC, Siakavellas SI, Koutsounas IS, et al. Reliability and applicability of two-dimensional shear-wave elastography for the evaluation of liver stiffness. *Eur J Gastroenterol Hepatol* 2016;28:1204-1209.
23. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.

Author names in bold designate shared co-first authorship

FIGURE LEGENDS

Figure 1: Rate of advanced fibrosis (panel 1a) or cirrhosis (panel 1b) as a function of increasing liver stiffness and coefficient of variation (CV) in the derivation set

* $p < 0.05$ between CV > 0.24 and the two others CV subgroups

$p < 0.01$ between CV < 0.10 and the two other groups

Figure 2: New reliability criteria for a 2D-SWE measurement

CV: coefficient of variation

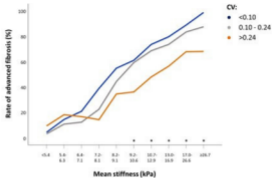
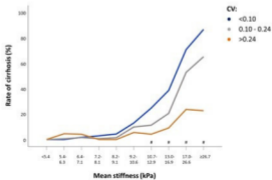
TABLE LEGENDS

Table 1: Characteristics of the 4277 measurements performed with two-dimensional shear wave elastography in the 788 included patients

Table 2: Accuracy of the 4277 2D-SWE measurements

Table 3: Diagnostic accuracy of 2D-SWE measurements according to the new reliability criteria in the derivation and validation sets

Table 4: Diagnostic accuracy of 2D-SWE examination as a function of the number of reliable measurements performed

1a**1b**

Mean stiffness (kPa)

<8.8

8.8 – 11.9

≥12.0

<0.10

Reliable

Reliable

Reliable

CV

0.10 - 0.24

Reliable

Reliable

Unreliable

>0.24

Reliable

Unreliable

Unreliable

Table 1: Characteristics of the 4277 measurements performed with two-dimensional shear wave elastography in the 788 included patients

	All (n=4277)	Derivation (n=2851)	Validation (n=1426)	p
Center (%):				0.989
- Angers	43.8	43.7	44.0	
- Beaujon	23.9	23.9	23.8	
- Bordeaux	32.3	32.4	32.2	
Age (years)	56.3 (45.1-64.2)	56.4 (45.9-64.3)	55.9 (44.3-64.0)	0.408
Male Sex (%)	61.3	60.7	62.4	0.286
Cause of CLD (%):				0.594
- NAFLD	45.8	46.5	44.3	
- Alcohol	9.8	9.8	9.9	
- Viral hepatitis	21.9	21.6	22.5	
- Others	22.5	22.1	23.3	
BMI (kg/m ²)	27.9 (24.2-32.0)	28.1 (24.2-32.1)	27.6 (24.2-31.9)	0.507
Biopsy length (mm)	28 (22-34)	28 (22-33)	28 (22-34)	0.314
Fibrosis (%):				0.710
- No/mild	52.7	52.9	52.2	
- Bridging	30.7	30.3	31.5	
- Cirrhosis	16.6	16.8	16.3	
Mean stiffness (kPa)	9.1 (6.7-14.6)	9.2 (6.7-14.5)	9.1 (6.7-14.9)	0.879
Standard deviation (kPa)	1.1 (0.7-2.3)	1.1 (0.7-2.3)	1.1 (0.7-2.4)	0.704
Region of interest (mm)	15 (12-17)	15 (12-17)	15 (12-17)	0.954

CLD: chronic liver disease; NAFLD: non-alcoholic fatty liver disease; BMI: body mass index

Table 2: Accuracy of the 4277 2D-SWE measurements

Set	Diagnostic target	AUROC	Cut-off (kPa)^a	DA (%)	Se (%)	Spe (%)	NPV (%)	PPV (%)
All	Advanced fibrosis	0.825±0.006	8.8	75.2	80.8	70.2	80.3	70.9
	Cirrhosis	0.880±0.006	12.0	78.2	85.3	76.8	96.3	42.3
Derivation	Advanced fibrosis	0.828±0.008	8.8	75.8	81.8	70.5	81.3	71.2
	Cirrhosis	0.881±0.007	12.0	78.5	85.2	77.2	96.3	43.1
Validation	Advanced fibrosis	0.819±0.011	8.8	74.0	79.0	69.4	78.3	70.2
	Cirrhosis	0.879±0.011	12.0	77.5	85.3	76.0	96.4	40.8

DA: diagnostic accuracy (rate of good-classification); Se: sensitivity; Spe: specificity; NPV: negative predictive value; PPV: positive predictive value

^a calculated in the derivation set using the highest Youden Index

p values for AUROC, DA, Se, and Spe between derivation and validation sets: all not significant

Table 3: Diagnostic accuracy of 2D-SWE measurements according to the new reliability criteria in the derivation and validation sets

Set	Reliability	2D-SWE measurements (n)	AUROC		Diagnostic accuracy (%)		
			Advanced fibrosis	Cirrhosis	Advanced Fibrosis ^a	Cirrhosis ^b	Fibrosis Classification ^c
All	All	4277	0.825±0.006	0.880±0.006	75.2	78.2	61.1
	Reliable	3165	0.836±0.007	0.922±0.008	76.7	91.0	70.0
	Unreliable	1112	0.650±0.019	0.713±0.016	71.1	41.8	35.7
	p	-	<0.001	<0.001	<0.001	<0.001	<0.001
Derivation	All	2851	0.828±0.008	0.881±0.007	75.8	78.5	61.9
	Reliable	2116	0.840±0.009	0.924±0.009	77.3	91.4	71.1
	Unreliable	735	0.641±0.024	0.712±0.019	71.6	41.4	35.4
	p	-	<0.001	<0.001	0.002	<0.001	<0.001
Validation	All	1426	0.819±0.011	0.879±0.011	74.0	77.5	59.5
	Reliable	1049	0.830±0.013	0.918±0.014	75.3	90.0	67.9
	Unreliable	377	0.667±0.031	0.714±0.027	70.3	42.7	36.3
	p	-	<0.001	<0.001	0.065	<0.001	<0.001

^a Using the 8.8 kPa 2D-SWE cut-off

^b Using the 12.0 kPa 2D-SWE cut-off

^c Fibrosis classification: <8.8 kPa: no/mild fibrosis; 8.8-11.9 kPa: bridging fibrosis; ≥12.0 kPa: cirrhosis

p values for AUROC and diagnostic accuracy between derivation and validation sets: all not significant

Table 4: Diagnostic accuracy of 2D-SWE examination as a function of the number of reliable measurements performed

297 patients with at least 5 reliable measurements available were included in the analysis. The reference for 2D-SWE examination was the mean of the 5 first measurements. We compared this reference to the first measurement, and to the mean of the first two, three and four measurements.

	Number of reliable measurements in 2D-SWE examination (n)				
	1	2	3	4	5 (reference)
Mean (kPa)	8.2±4.0	8.3±4.0	8.3±4.0	8.3±4.0	8.3±4.0
p ^a	0.515	0.861	0.429	0.291	-
ICC	0.965	0.988	0.993	0.998	-
p ^a	<0.001	<0.001	<0.001	<0.001	-
AUROC for bridging fibrosis	0.787±0.027	0.813±0.026	0.826±0.025	0.831±0.025	0.829±0.025
p ^a	0.005	0.070	0.617	0.465	-
DA for bridging fibrosis (%) ^b	72.7	75.8	77.8	78.1	76.4
p ^a	0.090	0.824	0.424	0.180	-
AUROC for cirrhosis	0.901±0.035	0.931±0.023	0.932±0.024	0.935±0.023	0.938±0.022
p ^a	0.142	0.313	0.248	0.399	-
DA for cirrhosis (%) ^c	94.3	94.3	93.9	94.3	93.9
p ^a	1.000	1.000	1.000	1.000	-

ICC: intraclass correlation coefficient; DA: diagnostic accuracy (rate of good-classification)

^a compared to the mean of 5 measurements

^b Using the 8.8 kPa 2D-SWE cut-off

^c Using the 12.0 kPa 2D-SWE cut-off

1st STEP

DEFINE RELIABLE MEASUREMENT

CV	> 0.24	RELIABLE	UN RELIABLE	UN RELIABLE
	0.10 0.24	RELIABLE	RELIABLE	UN RELIABLE
	< 0.10	RELIABLE	RELIABLE	RELIABLE
	< 8.8	8.8 11.9	≥ 12	

Mean Stiffness (kPa)

2nd STEP

PERFORM THREE RELIABLE MEASUREMENTS



Clinical Gastroenterology
and Hepatology